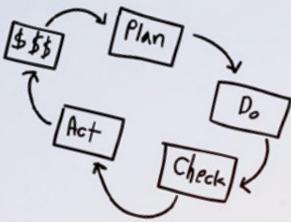
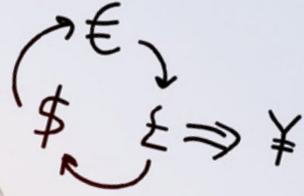




Fundamentos de INTELIGENCIA DE NEGOCIOS

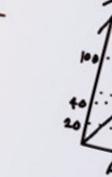
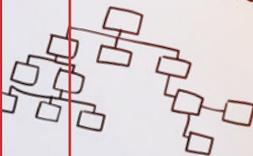
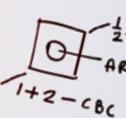


Success



$$A+B = \frac{c^2 + cx^2}{3 \times 6}$$

$$\frac{A+B+C+D}{1+4^2+\frac{3}{2}}$$



Acrobat Reader
Haz Click

Fundamentos de inteligencia de negocios

Primera edición electrónica de un ejemplar
(5 MB) en formato PDF

D.R. © 2022, Universidad Nacional Autónoma de México, Avenida
Universidad 3000, Col. Universidad Nacional Autónoma de México,
Ciudad Universitaria, Delegación Coyoacán, C.P. 04510,
México, CDMX.

FACULTAD DE INGENIERÍA
<http://www.ingenieria.unam.mx/>

Esta edición y sus características son propiedad de la Universidad
Nacional Autónoma de México. Prohibida la reproducción o
transmisión total o parcial por cualquier medio sin la autorización
escrita del titular de los derechos patrimoniales.

Hecho en México.

PREFACIO

En la primera década del siglo XXI, hemos estado inmersos en una serie de cambios tecnológicos que han modificado de manera radical nuestra forma de vida. Algunos de estos cambios son: la producción en masa de equipos de cómputo personal y su uso en tareas más cotidianas, la creación de la hoja de cálculo, la conformación y consolidación de la red de datos, los avances de la telefonía celular con los denominados teléfonos inteligentes, la creación de las tabletas electrónicas, y la unión de las disciplinas mecánica y eléctrica-electrónica para conformar la mecatrónica, todos estos cambios han estado encaminados a revolucionar la generación y recopilación de datos en el mundo actual. Estos cambios sin precedentes en la historia de la humanidad, debido a la velocidad con la que han venido ocurriendo, han conformado lo que se ha denominado la Cuarta Revolución Industrial; que se centró de manera inicial en la denominada Industria 4.0, pero que con el paso del tiempo ha llegado a otros campos como Logística 4.0.

El término Industria 4.0 fue usado por primera vez en la Feria Tecnológica de Hannover-Messe en el año 2011, y fue en este mismo evento en el año 2013 que el gobierno alemán presentó el documento fundacional que tiene por título “Recommendations for implementing the strategic initiative Industrie 4.0”, que fue elaborado por la Academia Nacional de Ciencia e Ingeniería de Alemania (ACATECH) y el Centro de Investigación alemán para la Inteligencia Artificial (DFKI). Tuvieron que pasar tres años para que evolucionara un poco más el concepto y, finalmente, en el marco del Foro Económico Mundial de 2016 (World Economic Forum) el tema central fue

“La Cuarta Revolución Industrial”. Esta Cuarta Revolución Industrial se refiere a la digitalización de sistemas y procesos industriales y su interconexión mediante el internet de las cosas para conseguir una mayor flexibilidad e individualización de los procesos productivos, lo cual trae como consecuencia la recolección de un gran volumen de datos (Big Data) que pueden ser usados para ser analizados y, de esta forma, convertirse en información que soporte la toma de decisiones en las organizaciones.

En el plan de estudio de la licenciatura en Ingeniería Industrial de la Facultad de Ingeniería de la UNAM, aprobado en el año 2016, se incluyó la asignatura de Inteligencia de Negocios, como una asignatura optativa dentro del módulo terminal en Gestión de la Cadena de Suministros. Este trabajo surge de la necesidad de establecer un marco básico de los conceptos y temas que sirvan como referencia para dicha asignatura, la cual tiene como objetivo que el estudiante explique la importancia de la Inteligencia de Negocios y que obtenga una visión integral de la repercusión que tienen los temas de recolección y análisis de datos en el entorno actual y su utilización como soporte en la toma de decisiones para alcanzar los objetivos de la organización. Este trabajo fue realizado con el apoyo del Programa UNAM-DGAPA-PAPIME PE102218 “Elaboración de Material didáctico (notas, presentaciones, series de ejercicio y estudios de caso) y cursos en línea respecto a Simulación de Sistemas Complejos e Inteligencia de Negocios (Business Intelligence).”

INTRODUCCIÓN

Un tema que es un hecho en la vida cotidiana, actualmente, es la generación de grandes volúmenes de datos (Big Data), para Fuentes (2017) esto es el resultado del avance de la computación y las telecomunicaciones. Lo cierto es que la generación de datos comenzó a cristalizarse cuando se comprendió la necesidad de estos en las organizaciones para hacer un uso más efectivo de sus recursos, un ejemplo de lo anterior se observa en el área de producción de las empresas con el surgimiento de los sistemas para la Planeación de Requerimientos de Materiales (Material Requirement Planning-MRP) en la década de 1960.

Posteriormente, a medida que la tecnología de cómputo avanzaba entre la década de 1970 y 1980, surge un sistema más robusto denominado Planeación de Requerimientos de Manufactura o Fabricación (Manufacturing Requirement Planning- MRP II), las funciones que se incluían en este sistema eran la compra, las listas de materiales, la planeación y administración de contratos, lo que permitió tener un adelanto de los beneficios que se tenían al recolectar, integrar y compartir los datos de la empresa, y que tuvo como consecuencia una disminución en los inventarios y la eliminación de desperdicios. Durante la década de 1980, la experiencia de recolectar datos en el área de producción se fue trasladando a otras funciones de las organizaciones como en la administración de las relaciones con los clientes, en la parte contable y financiera, y en el área de recursos humanos, lo que propició que en la década de 1990 todos esos sistemas comenzaran a integrarse de manera natural, iniciando así la Planeación de Recursos Empresariales (Enterprise Resources Planning).

Los datos recolectados mediante estos sistemas reflejan las interacciones entre las funciones y áreas de las organizaciones, sin embargo, son los datos que las organizaciones recolectan de la interacción con su medio ambiente los que han favorecido la creación de los grandes volúmenes de datos. Según datos del Banco Mundial (2021), el número de personas que usan internet como un porcentaje total de la población ha pasado de ser 0.681% en 1995 a 50.763% en el año 2018. Otro dato relevante proporcionado por el Banco Mundial (2021) es el número de suscripciones a telefonía celular móvil que pasó de 90.74 millones en 1995 a 8,196 millones en el año 2019.

Es a finales de la década de 1990 y en la primera década del siglo XXI que surgen de manera conceptual algunas de las organizaciones e ideas que han cambiado los paradigmas de comunicación e interacción de la humanidad. Facebook surge en el año 2004, Twitter en el año 2006 y WhatsApp en 2009. Estas organizaciones, que han cambiado el paradigma de las comunicaciones y relaciones humanas, han hecho que los canales tradicionales de promoción y publicidad como lo son la radio y la televisión sean complementados y, en algunos casos, casi sustituidos con estas denominadas redes sociales.

La industria del entretenimiento y los canales tradicionales de comercialización también han ido cambiando debido a los avances tecnológicos, en particular, el surgimiento y crecimiento de internet. Un ejemplo de esto es Netflix que inicia sus operaciones en 1997, pero no es hasta 2007 que en Estados Unidos comienza su servicio de video sobre demanda y hasta el año 2011 que comienza su internacionalización.

Por otra parte, a inicios de la segunda década del siglo XXI, Amazon es considerada la cadena minorista más grande del mundo a pesar de no contar con puntos de venta físicos, esta fue fundada en 1994, pero no es hasta la primera década del siglo XXI que consolida su presencia basándose en el incremento exponencial de usuarios de internet. En 2007, Apple, de la mano de su fundador y director general Steve Jobs, lanza su concepto de teléfono móvil denominado Iphone® lo que da pie al inicio de una masificación de las comunicaciones a través de estos dispositivos.

Estos cambios han generado la necesidad de contar con nuevas técnicas y metodologías que permitan usar los grandes volúmenes de datos para la toma de decisiones, de esta forma surge la Analítica de Negocios (Business Analytics) o simplemente Analítica (Analytics). El Instituto para la Investigación de Operaciones y la Ciencia Directiva (Institute For Operation Research and Management Science-INFORMS) define la Analítica o Analítica de Negocios (Analytics/Business Analytics) como: “el proceso científico de transformación de datos en hallazgos y conocimiento con el fin de tomar mejores decisiones”. La analítica se considera la toma de decisiones basada en hechos apoyados por datos, y es frecuentemente vista como más objetiva que otros enfoques para la toma de decisiones. El mismo instituto conceptualiza la analítica como una combinación de la denominada ciencia de datos y la investigación de operaciones. Este enfoque de la analítica de usar datos para identificar hallazgos y generar información para resolver problemas en las organizaciones en sus diferentes funciones: mercadotecnia, finanzas y operaciones, fue el motivo de un estudio conducido por el Instituto Tecnológico de Massachusetts (Massachusetts Institute of Technology-MIT) a través de su publicación MIT Sloan Management Review en asociación con el Instituto para el valor empresarial de la empresa IBM (IBM Institute for Business Value) en octubre de 2010, mediante una encuesta global a casi 3,000 ejecutivos, gerentes y analistas en 30 tipos de industria y en alrededor de 100 países. El estudio encontró que las organizaciones con alto desempeño utilizan la analítica cinco veces más que las empresas con menor desempeño, y que la mejora de la información y el análisis de los datos es una prioridad en las empresas encuestadas.

En los inicios de la segunda década del siglo XXI, el análisis de los grandes volúmenes de datos (Big Data), se convertirá en una necesidad para todas las empresas desde la micro, pequeña y mediana empresa hasta las grandes corporaciones globales. Por lo que es necesario contar con documentos que busquen conceptualizar de la forma más clara posible estas tendencias.

Este libro está conformado por cinco capítulos o secciones; en el capítulo 1 se inicia con la evolución de la Inteligencia de Negocios desde sus orígenes en la década de 1970 con los sistemas de soporte de decisión empresarial hasta llegar al surgimiento de los conceptos de analítica y ciencia de datos; lo cual implica tener una esquema conceptual que permita entender estos elementos y cómo interactúan entre sí para conformar lo que es la Inteligencia de Negocios y su relación con la transformación digital en las organizaciones.

En el capítulo 2 se analizan los factores tecnológicos que intervienen en la transformación digital en las organizaciones, se abordan de manera más detallada los conceptos de grandes volúmenes de datos (Big Data), la analítica de negocios (Analytics/Business Analytics) y la ciencia de datos (Data Science), para que en el capítulo 3 se profundice en la forma en que los datos son recolectados y almacenados en los denominados almacenes de datos (Data Warehousing), una vez que los datos están almacenados se debe proceder a su uso.

En el capítulo 4 se aborda uno de los tipos de analítica que se refiere a la descripción y entendimiento de los datos para utilizarlos en la comprensión de los sistemas, mediante hallazgos (insights) que conduzcan a una toma de decisiones. Finalmente, es en el capítulo 5 en donde se liga el uso de la Inteligencia de Negocios con la dirección estratégica de las organizaciones.

ÍNDICE

PREFACIO
 INTRODUCCIÓN

CAPÍTULO 1. Inteligencia de negocios y la cuarta revolución industrial en las organizaciones

1.1 Evolución de los Sistemas de Soporte de Decisión a la Ciencia de Datos
 1.2 Definición y esquema de la Inteligencia Analítica de Negocios.....
 1.3 La transformación digital en las empresas y organizaciones
 1.3.1 Definición de transformación digital
 1.3.2 Tendencias de la Transformación Digital.....
 1.3.3 Soportes, conceptos, metodologías y tecnologías para la Transformación Digital.....
 1.3.4 Metodología para la transformación digital

Referencias.....

CAPÍTULO 2. Factores tecnológicos en la transformación digital de las organizaciones.....

2.1 Grandes volúmenes de datos (Big Data).....
 2.1.1 Grandes volúmenes de datos (Big Data).....
 2.1.2 Tipos de datos.....
 2.1.3 Características de los Grandes volúmenes de datos (Big Data).....
 2.1.4 Tipos de fuentes de los Grandes volúmenes de datos (Big Data).....
 2.1.5 Iniciativa de datos abiertos
 2.2 Analítica de Negocios.....
 2.2.1 Definición de analítica
 2.2.2 Niveles o tipos de analítica de negocios

| | | |
|-------|--|----|
| 2.2.3 | Categorías de analítica de negocios con base en la fuente de los datos | |
| 2.3 | Ciencia de Datos. | 34 |
| 2.3.1 | Definición de Ciencia de Datos. | 32 |
| 2.3.2 | Herramientas para Ciencia de Datos. | 36 |
| 2.3.3 | Proceso de Ciencia de Datos | 40 |
| 2.3.4 | Certificaciones profesionales en Ciencia de Datos | 41 |
| | Referencias | 41 |

CAPÍTULO 3. Almacenamiento de datos (Data Warehousing) 44

| | | |
|-----|---|----|
| 3.1 | Definición y conceptos de almacenamiento de datos (Data Warehousing) | 44 |
| 3.2 | Proceso General de Almacenamiento de Datos | 53 |
| 3.3 | Arquitecturas de Almacenamiento de Datos | 55 |
| 3.4 | Integración, Extracción, Transformación y Carga de Datos (EII, ETL) | 65 |
| 3.5 | Desarrollo y problemas en la implementación del Almacenamiento de Datos | 70 |
| 3.6 | Problemas de almacenamiento de datos-implementación | 80 |
| | Referencias | 84 |

CAPÍTULO 4. Analítica descriptiva 86

| | | |
|-----|--|-----|
| 4.1 | Solución de problemas de análisis de datos | 86 |
| 4.2 | Datos para analítica de negocios | 91 |
| 4.3 | Herramientas para la analítica de datos. | 95 |
| 4.4 | Importar datos | 96 |
| 4.5 | Ordenar los datos. | 100 |
| 4.6 | Transformar los datos. | 107 |
| 4.7 | Visualizar los datos | 115 |
| 4.8 | Modelar los datos. | 134 |
| 4.9 | Comunicar los datos | 148 |

Referencias 157

**CAPÍTULO 5. Enfoque directivo y estratégico
de la inteligencia analítica de negocios 159**

5.1 Definiciones de dirección y estrategia 159

5.2 Metodología dirección estratégica 162

5.3 Inteligencia analítica de negocios y la dirección estratégica . . . 166

Referencias 170

Conclusiones, reflexiones y trabajo futuro 171

CAPÍTULO 1

Inteligencia de negocios y la cuarta revolución industrial en las organizaciones

1.1 Evolución de los sistemas de soporte de decisión a la ciencia de datos

En la última década la globalización de las operaciones de las empresas y las economías ha generado que estas se vean envueltas por un entorno más complejo, ya que el número de clientes, proveedores, productos y servicios se ha incrementado, a esto hay que sumar los distintos aspectos culturales donde las empresas tengan presencia.

Adicionalmente, la evolución de la tecnología en lo referente a la generación, almacenamiento y transmisión de datos ha generado que los directores y gerentes tengan la posibilidad de hacer uso de estos elementos como un soporte en su toma de decisiones.

Para Sharda, Denle y Turban (2018), existen algunas condiciones que han permitido el crecimiento del soporte de la toma de decisiones, los cuales se enlistan a continuación:

- **Comunicación grupal y colaboración.** Las decisiones actualmente son hechas por grupos, cuyos miembros pueden estar ubicados en diferentes lugares. Estos grupos se comunican y colaboran de manera sencilla usando herramientas de colaboración como los teléfonos inteligentes y aplicaciones de videoconferencias.
- **Mejor gestión de los datos.** Muchas decisiones implican el uso de técnicas que requieren cálculos complejos. Los datos usados por dichas técnicas son almacenados en diferentes bases de datos en cualquier parte o proceso de la organización, e incluso fuera de esta. Estos datos pueden incluir texto, sonido, gráficas y video, e incluso estar en diferentes lenguajes.
- **Gestión de grandes almacenes de datos.** Actualmente, las grandes empresas y organizaciones generan grandes almacenes de datos que contienen enormes cantidades de datos. Los costos relacionados a la

gestión (almacenamiento y minería) de estos datos han disminuido drásticamente en la última década.

- **Soporte analítico.** Al contar con una mayor cantidad de datos, técnicas y tecnologías, la cantidad de análisis y alternativas de solución que pueden ser evaluadas por los expertos y/o dueños de los procesos se realizan en menor tiempo y costo, e incluso en diferentes regiones del mundo.
- **Superar los límites cognitivos en el procesamiento y almacenamiento de información.** Para Simón (1977), la mente humana posee una capacidad limitada para procesar y almacenar información. Los sistemas de cómputo han hecho posible que ese límite sea mayor gracias al rápido acceso y análisis de datos que son transformados en información para la toma de decisiones.
- **Gestión del conocimiento.** Las organizaciones tienen recolectada y almacenada una gran cantidad de datos e información de sus operaciones, clientes, procesos, empleados. Dicho conocimiento es comunicado entre los diferentes dueños de los procesos de manera estructurada y no estructurada.
- **Soporte en cualquier lugar y en cualquier momento.** Al tener acceso a telecomunicaciones como la tecnología inalámbrica, los tomadores de decisión pueden acceder a información a cualquier hora y desde cualquier lugar; analizar e interpretar esta y comunicarse con los dueños de los procesos.

Lo que ahora se denomina ‘grandes volúmenes de datos’ y ‘analítica de los negocios’ (*Big Data & Analytics*), conceptualmente tuvo sus inicios en la década de 1970, cuando comenzaron a surgir sistemas de información que servían para que los tomadores de decisiones los usarán como un soporte para la toma de decisiones. Una definición de lo que son los ‘sistemas de soporte de decisión’ hecha por Keen y Scott-Morton (1978) es “los Sistemas

de Soporte de Decisión son los que combinan los recursos intelectuales de las personas con las capacidades de la computadora para mejorar la calidad de las decisiones”.

El desarrollo de los sistemas de soporte de decisión (*Decision Support Systems DSS*), ha venido ocurriendo desde la década de 1970 hasta la actualidad. Sharda, Denle y Turban (2018) identifican cinco etapas las cuales se enlistan a continuación:

- Sistemas de soporte de decisión (1970-1980)
- Sistemas de información empresarial/ejecutivos (1980-1990)
- Inteligencia de negocios (1990-2006)
- Analítica de negocios (2006-2015)
- Grandes volúmenes de datos (2015 a la fecha)

Respecto a las últimas dos etapas, si bien Sharda, Denle y Turban (2018) buscan establecer una diferencia, en realidad debe ser considerada una sola etapa que debe denominarse ‘grandes volúmenes de datos y analítica de los negocios’ (*Big Data & Analytics*). A continuación, se desarrollarán las características de cada una de estas etapas.

- **Sistemas de soporte de decisión (1970-1980)**

Los primeros pasos respecto a la generación de los sistemas de soporte de decisión fue elaborar reportes periódicos estructurados, para que estos fueran utilizados como soporte para la toma de decisiones. Respecto a los datos, estos eran obtenidos mediante la consulta de manuales de los procesos y la consulta a expertos, los cuales eran capturados en modelos para resolver modelos de optimización con ciertas restricciones; es decir, los modelos que usaban para soportar las decisiones eran los modelos de investigación de operaciones. A finales de la década de 1970 y comienzos de la década de 1980, surgió el modelado de sistemas expertos basados en reglas, cuya premisa es capturar el conocimiento de los expertos de tal

forma que las computadoras lo puedan procesar (mediante un grupo de reglas del tipo, si-entonces-de otro modo o bien heurísticas), y así estos modelos se podrían utilizar para consultarlos de la misma forma que lo usan los expertos.

- **Sistemas de información empresarial/ ejecutivos (1980-1990)**

Al iniciar la década de 1980, la forma en que las organizaciones capturaban los datos relacionados con su operación sufrió un cambio. Se pasó de tener muchos sistemas de información sin conexión y hechos para capturar la información requerida de cada una de las áreas o funciones de la organización (contabilidad, ventas y mercadotecnia, finanzas y manufactura) a integrar todos esos sistemas en un solo sistema de información a nivel de toda la empresa que se denominó 'sistema de planeación de recursos empresariales' (*Enterprise Resources Planning-ERP*).

El cambio más importante fue el ver a la empresa como un conjunto de áreas y funciones inconexas para la recolección de datos y medición del desempeño, lo que se tradujo en una recolección de datos y medición del desempeño por área, a una visión sistémica de la organización. En esta, el comportamiento de un área se relaciona y afecta directamente el desempeño de la empresa u organización en su conjunto, lo que se traduce en que los datos son recolectados de cada una de las áreas o funciones de la empresa y son integrados en un solo esquema, a este se le conoce como 'Gestión de bases de datos relacionales' (*Relational Database Management-RDBM*). El resultado fue que cada una de las funciones de la organización tuvieran acceso a los mismos datos e indicadores de desempeño cuándo y en dónde se necesitaran.

- **Inteligencia de negocios (1990-2006)**

Una vez que se contó con un sistema que interconectaba a las diferentes áreas, las organizaciones se centraron en la necesidad de contar con reportes más dinámicos y configurables, lo que condujo al desarrollo de los ‘Sistemas ejecutivos de información’ (*Executive Information Systems-EIS*). Estos sistemas fueron diseñados como tableros y gráficas que sintetizaban la información de los principales indicadores de desempeño de las organizaciones de forma tal, que los tomadores de decisiones pudieran monitorearlos.

Para lograr elaborar reportes más versátiles y al mismo tiempo mantener intacta la integridad transaccional de los sistemas de información empresarial, fue necesario crear un nivel de datos intermedio. Este repositorio de datos soporta específicamente la generación de reportes y la toma de decisiones y es denominado ‘Almacén de datos’ (*Data Warehouse-DW*). En el inicio de la primera década del siglo XXI, los sistemas de soporte de decisión impulsados por los almacenes de datos fueron denominados ‘sistemas de inteligencia de negocios’.

Los datos en los almacenes de datos al inicio se actualizaban periódicamente, de forma regular en periodos diarios o semanales, por lo que la toma de decisiones se hacía con base en los últimos datos actualizados. Para que los datos fueran recolectados con una mayor frecuencia se creó el almacenaje de datos en tiempo real (*real-time data warehousing*), cuyo concepto más acertado es el almacenaje de datos en tiempo correcto (*right-time data warehousing*) donde la base es el adoptar una política de actualización de datos basándose en los datos y su necesidad respecto a su actualización, esto es, analizar los procesos que generan dichos datos y determinar los periodos de tiempo en que estos deben ser actualizados para que contribuyan en la toma de decisiones.

Al incrementarse la cantidad de datos almacenados debido al aumento en su frecuencia de recolección, fue necesario desarrollar dos campos, el

primero de ellos es la capacidad de almacenaje de las organizaciones y el segundo, la capacidad de “minar” los datos corporativos para convertirlos en información que permita mejorar los procesos y prácticas de las organizaciones, de ahí los términos ‘minería de datos y minería de textos’.

- **Grandes volúmenes de datos y analítica de los negocios (Big Data & Analytics) (desde 2006 a la actualidad)**

Respecto a la evolución de los ‘sistemas de soporte de decisión’, el cambio conceptual más reciente se dio en el año 2006 cuando Davenport (2006) publicó un artículo respecto a la competencia mediante lo que él denominó ‘analítica’ (Analytics), que en palabras del autor es: “el uso exhaustivo de datos, análisis estadísticos, modelos explicativos y de predicción, y dirección y gestión basados en hechos como motor de la toma de decisiones y de las acciones emprendidas”.

Adicionalmente al cambio conceptual de la forma en que las empresas comenzaron a tomar decisiones, en la década que inició en el año 2010 comenzó un cambio en el paradigma de la forma en que los datos comenzaron a ser recolectados, ligado a que se comenzó a tener un mayor acceso y uso de la red de datos. De esta forma, los registros web, dispositivos domésticos “inteligentes”, los dispositivos portables para monitorear la salud y lo que ha representado un cambio en la forma de vida de la humanidad: las denominadas redes sociales (Facebook, Twitter, Instagram, LinkedIn, entre otras), se han convertido en una fuente de datos que permite caracterizar a los usuarios y clientes, y que al analizarlos se convierten en información para las organizaciones. Las redes sociales capturan una gran cantidad de datos no estructurados y extraer gran cantidad de información valiosa de estos datos ha representado un desafío para los sistemas de cómputo. El término Big Data surge para referirse a la generación y almacenamiento de grandes volúmenes de datos provenientes de las diferentes fuentes antes mencionadas.

La forma en que las empresas están tomando sus decisiones actualmente está relacionada con la capacidad de las organizaciones para recolectar, almacenar y analizar sus datos, por lo que es indispensable contar con personal con el conocimiento necesario para efectuar una toma de decisiones eficaz y eficiente, así como con la capacidad de implantar o llevar las acciones necesarias resultado de esta toma de decisiones. Un concepto que se ha retomado con más fuerza en la segunda mitad de la década 2010, es el de ciencia de datos el cual tiene una relación directa con la capacidad de las organizaciones para la toma de decisiones.

En la figura 1 se muestra un resumen de la evolución de los sistemas de soporte de decisión y las características de cada una de sus etapas.

| | Sistemas de Soporte de Decisión | Sistemas de Información Empresarial/ Ejecutivos | Inteligencia de Negocios | Grandes volúmenes de datos y Analítica de los negocios (Big Data & Analytics) |
|-----------------|--|--|---|---|
| | 1970-1980 | 1980-1990 | 1990-2006 | 2006-ACTUALIDAD |
| Características | <ul style="list-style-type: none"> • Reporte periódicos estructurados. • Datos de manuales de procesos y consulta a expertos. • Modelos de optimización. • Sistemas expertos/ Inteligencia Artificial. | <ul style="list-style-type: none"> • Sistema de Planeación de Recursos Empresariales (<i>Enterprise Resources Planning-ERP</i>). • Gestión de bases de datos relacionales (<i>Relational Database Management-RDBM</i>) | <ul style="list-style-type: none"> • Almacenamiento de Datos (<i>Data Warehousing-DW</i>). • Tableros y gráficas de los principales indicadores de desempeño). • Sistemas Ejecutivos de Información (<i>Executive Information Systems-EIS</i>). • Almacenaje de datos en tiempo real (<i>real-time data warehousing</i>). • Almacenamiento en la nube. • Minería de datos y de texto. | <ul style="list-style-type: none"> • Inteligencia Analítica (Davenport-2006) • Internet de las cosas (<i>Internet of Things IoT</i>). Teléfonos inteligentes, dispositivos domésticos “inteligentes”, los dispositivos portables para monitorear la salud. • Redes sociales (Facebook, Twitter, Instagram, LinkedIn, entre otras). |

Figura 1. Características de la evolución de los sistemas de soporte de decisión.

Fuente: Elaboración propia adaptado de Sharda, Denle y Turban (2018)

1.2 Definición y esquema de la *inteligencia analítica* de negocios

Una vez que hemos visto la forma en que los sistemas de soporte de decisión han surgido y evolucionado, es necesario establecer la definición de lo que es inteligencia de negocios, aunque un término más adecuado sería inteligencia analítica de negocios. Sharda, Denle y Turban (2018) dan la siguiente definición de inteligencia de negocios:

“La Inteligencia de Negocios es un término general que se refiere a la combinación de arquitecturas, herramientas, bases de datos, herramientas analíticas, aplicaciones y metodologías”¹.

Esta definición carece de un vínculo a lo que es la aplicación de la inteligencia de negocios como un soporte de la toma de decisiones en las organizaciones. Por lo que revisaremos otra definición de este término, la cual fue realizada por Gartner, una empresa de consultoría fundada en 1979, cuyo enfoque principal en sus inicios estaba ligado a la tecnología, esta empresa en su glosario define lo que es analítica e inteligencia de negocios como sigue:

“Analítica e Inteligencia de Negocios (*Analytics and Business Intelligence-ABI*) es un término general que incluye las aplicaciones, la infraestructura y las herramientas, y las mejores prácticas que permiten el acceso y el análisis de la información para mejorar y optimizar las decisiones y el rendimiento”².

-
1. Sharda Ramesh, Delen Dursun y Turban Efraim. (2018). *Business Intelligence, analytics and data science: a managerial perspective*. New York: Pearson, p.16.
 2. Gartner (2019). Gartner Glossary. <https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi>. Consultada el 20 de diciembre de 2019.

Esta definición contiene el concepto de información y cómo este se aplica a la toma de decisiones y con qué objetivo, por lo que complementa la definición previa propuesta por Sharda, Denle y Turban (2018). Una última definición que nos servirá para agregar más conceptos que permiten clarificar el concepto de inteligencia analítica de negocios es la propuesta por Eckerson (2002), la cual dice:

“Inteligencia de Negocios, son los procesos, las tecnologías y las herramientas necesarias para convertir los datos en información, la información en conocimiento y el conocimiento en planes que impulsan una acción comercial rentable. La inteligencia empresarial abarca el almacenamiento de datos, las herramientas de análisis empresarial y la gestión de contenido / conocimiento”³.

De esta última definición existen cuatro conceptos que son importantes para definir lo que es la inteligencia analítica de negocios, estos conceptos son: procesos, tecnología, datos y conocimiento. Para efectos de poder contar con una definición lo más completa utilizaremos la siguiente definición:

“La Inteligencia Analítica de Negocios es una arquitectura que considera los procesos, infraestructura tecnológica (base de datos, ERP, CRM, WMS, aplicaciones, ECOM) y herramientas analíticas que permiten acceder y analizar los datos para convertirlos en información, la cual es usada como soporte para la toma de decisiones que conduzcan la implantación de acciones que generen una ventaja competitiva y la adquisición de conocimiento para la organización”⁴.

-
3. Eckerson Wayne W. (2002). The Rise of Analytic Applications: Build or Buy?, The Data Warehouse Institute http://download.101com.com/tdwi/research_report/0402aa-report.pdf Consultada el 21 de diciembre de 2019.
 4. Es una definición que elaboré a partir de la lectura de definiciones previas por lo que es de mi autoría.

Para Sharda, Denle y Turban (2018), un sistema de inteligencia analítica de negocios tiene cuatro grandes componentes o subsistemas:

1. **Un subsistema de almacén de datos.** Está conformado por todas las posibles fuentes de datos y los almacenes de datos (*Data Warehouses*) de las organizaciones.

Las posibles fuentes de datos pueden incluir: sistemas de planeación de recursos empresariales (*Enterprise Resources Planning*), sistemas de gestión de relaciones con los clientes (*Customer Relationship Management*), portales con proveedores (*Electronic Data Interfase*), portales de comercio electrónico (*Ecommerce*), redes sociales, y programas informáticos autónomos (*bots*).

Además, se debe considerar al personal técnico que dará soporte a todos los sistemas que, generalmente, es el área de tecnologías de la información.

2. **Un subsistema de analítica de negocios.** Consiste en todas las áreas que integran la organización y que usan los datos recolectados en los almacenes de datos (*Data Warehouses*). Para algunas organizaciones se han comenzado a formar áreas específicas, aunque dichas áreas han sido nombradas o relacionadas con la denominada ciencia de datos (*Data Science*). El concepto de analítica se divide en tres niveles: la analítica descriptiva, la analítica predictiva y la analítica prescriptiva.
3. **Un subsistema de modelo de negocios/estrategia.** Está conformado por todos los directores y gerentes que deben vincular el uso de los datos para obtener la información que se necesita para soportar la toma de decisiones y ganar conocimiento para integrar a la inteligencia analítica de negocios dentro de la estrategia de la organización. Para conseguir esto, lo más adecuado es partir de un conocimiento integral de los procesos y de los dueños de los procesos presentes en la organización.

4. **Un subsistema de reporte de indicadores de desempeño.** Este subsistema es con el que la mayor parte de las personas están familiarizadas al hablar de inteligencia de negocios, y consiste en todos los tableros de indicadores de desempeño que existen en una organización, los cuales son utilizados para medir el efecto de la implementación de las estrategias dentro de la organización.

En la figura 2 se muestra los subsistemas que componen la arquitectura de un Sistema de Inteligencia de negocios.

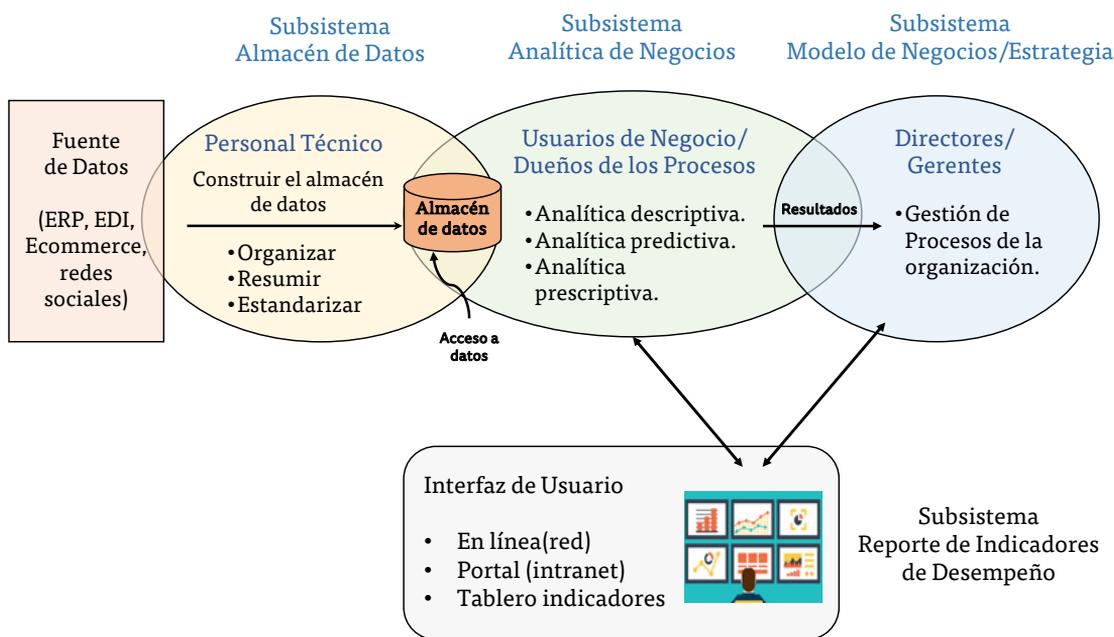


Figura 2. Subsistemas que componen la arquitectura de un sistema de Inteligencia Analítica de Negocios. Fuente: Elaboración propia adaptado de Sharda, Denle y Turban (2018)

El que una organización invierta en un sistema de inteligencia analítica de negocios, se debe generalmente a la necesidad de esta de trabajar de manera más inteligente (efectiva). Gartner (2004) propone la creación de un Centro de Competencia en Inteligencia Analítica de Negocios, dicho centro podría tener las siguientes funciones:

- Vincular los proyectos de inteligencia analítica de negocios a la planeación y ejecución de la estrategia en la organización.
- Documentar y promover las mejores prácticas relacionadas con la inteligencia analítica de negocios dentro de las diferentes áreas y procesos en la organización.
- Dirigir el diseño de la arquitectura de la inteligencia analítica de negocios o sus modificaciones junto con los niveles directivos.

1.3 La transformación digital en las empresas y organizaciones

1.3.1 Definición de transformación digital

Para analizar lo que es la transformación digital de las empresas y las organizaciones (*DX-Digital Transformation*), analizaremos algunas de las definiciones que se tienen al respecto.

Para Rouse (2015), “la transformación digital es la reinención de una organización a través de la utilización de la tecnología digital para mejorar la forma en que la organización se desempeña y sirve a quienes la constituyen. Digital se refiere al uso de la tecnología que genera, almacena y procesa los datos”⁵.

En esta definición destacan las siguientes palabras o conceptos:

- Reinención, esta palabra hace referencia a una organización que está operando y que requiere gestionar un cambio de sus procesos.
- Tecnología digital, hace referencia a la tecnología que genera, almacena y procesa datos.
- Mejorar el desempeño y servir a quienes la constituyen.

5. Rouse Margaret. *Transformación digital*. TechTarget. <https://searchdatacenter.techtarget.com/es/definicion/Transformacion-digital> Consultada el 21 de diciembre de 2019.

Otra definición que vale la pena revisar es la que hace Lombardero (2015), “la transformación digital se basa fundamentalmente en la hibridación de productos físicos y virtuales o en la conexión de productos, servicios, cosas y personas de forma inteligente, de manera que modifica la naturaleza de los productos y servicios”⁶.

Esta segunda definición hace referencia a un cambio de naturaleza de los productos y servicios, vinculándolo con el concepto de hibridación, el cual hace referencia a que la naturaleza de los productos y/o servicios que las organizaciones ofrecen en la actualidad deben considerar el uso de las nuevas tecnologías que han hecho que la población necesite nuevos modelos de negocio. Por ejemplo, respecto a la necesidad de entretenimiento de la sociedad actual, las nuevas generaciones siguen necesitando ver contenido como películas, series de entretenimiento y música, lo que ha cambiado es la forma en la que las compañías les proporcionan estos productos, en la actualidad cadenas de televisión que producían dichos contenidos y los hacían llegar a los clientes mediante medios tradicionales como radio y televisión, ofrecen sus contenidos de forma directa mediante el uso de las diferentes plataformas en sitios de internet que te permiten por un costo poder ver series, películas, música sin la necesidad de hacerlo en una televisión, ir a una sala de cine, escuchar la radio o tener un reproductor de música.

En resumen, los avances tecnológicos de la última década han afectado la forma en que las empresas y organizaciones ejecutan sus procesos, así como los productos y servicios que ofrecen. Para efectos de tener una definición de lo que se denomina transformación digital (*DX-Digital Transformation*) de las empresas y organizaciones tenemos:

“La transformación digital (*DX-Digital Transformation*) de una empresa y organización, es el proceso de diseñar y/o adaptar sus procesos, productos y/o servicios, para incluir las nuevas tecnologías (internet de las cosas,

6. Lombardero, Luis. *Trabajar en la era digital. Tecnología y competencias para la Transformación Digital*. Madrid: LID Editorial, 2015, p. 135.

almacenamiento virtual-*cloud*, cadenas de bloque-*blockchain* e inteligencia artificial, entre otras) que permiten la recolección y análisis de datos y así mejorar su desempeño, causando un impacto en sus integrantes (empleados y accionistas), clientes y proveedores, brindándole una ventaja competitiva en el mercado.”

1.3.2 Tendencias de la transformación digital

En los últimos años se han consolidado cambios tecnológicos encaminados a la transformación digital de las empresas y organizaciones, debido a esto, la revista Forbes ha venido publicando anualmente las principales tendencias en lo referente a la transformación digital. Las tendencias propuestas por Newman (2018-2019) son:

1. **5G como una realidad.** Las empresas más importantes en telecomunicaciones, como Qualcomm, AT&T, Verizon, Nokia, Ericsson y Huawei, tienen dentro de sus objetivos que las implementaciones globales de 5G se mantengan al ritmo. La red 5G, incrementará la velocidad y confiabilidad de la transmisión de datos y dentro de sus principales aplicaciones están las diferentes aplicaciones para teléfonos inteligentes, ciudades inteligentes, vehículos inteligentes y la industria 4.0.
2. **Chatbots e inteligencia artificial conversacional.** Los programas que simulan mantener una conversación con una persona al proveer respuestas automáticas a entradas hechas por el usuario (chatbots), junto con los asistentes conversacionales como Siri, Cortana y Alexa, se han convertido en una forma en el que las empresas procuran tener un servicio de atención a los usuarios que cumpla con dos aspectos fundamentales: la atención a los clientes las 24 horas del día todos los días del año y la recolección de datos de los clientes de forma digital. Para esto se tienen que desarrollar aún más el procesamiento del lenguaje natural (*Natural Language Processing-NLP*) y el análisis de sentimientos (*Sentimental Analysis-SA*).

3. **Cloud y multicloud (Nubes conectadas-públicas, privadas y/o híbridas).** “Las empresas se están dando cuenta de que ir a la nube pública, a la nube privada o al centro de datos no es la mejor opción. A veces, necesitan una mezcla de todos o ambos. Por lo tanto, las nubes conectadas continúan desarrollándose para satisfacer las necesidades cambiantes de las empresas, ya sea que deseen almacenar en la nube el almacenamiento, las redes, la seguridad o la implementación de aplicaciones”⁷.
4. **Un WiFi más rápido para un mundo más rápido.** “WiFi 6 y 5G son tecnologías completamente diferentes, pero ambos nos traerán velocidades de procesamiento y conexión inalámbrica mucho más rápidas en 2020. 5G y WiFi 6 trabajando en concierto crearán la combinación perfecta de extremo a extremo de conectividad ultrarrápida para el hogar y la oficina. Las velocidades de descarga serán hasta 3 veces más rápidas que las que se pueden lograr con WiFi 5, pero esa no es la mejor medida del valor del nuevo estándar. El valor real de WiFi 6 será su capacidad de extender velocidades de datos más rápidas a muchos más dispositivos de los que WiFi 5 pudo administrar. Esto es importante por dos razones principales: la primera es que se espera que el número de dispositivos conectados en la red WiFi promedio aumente de aproximadamente 10 a 50 en los próximos años, y ese aumento requerirá una mayor velocidad, más eficiente y capacidades WiFi más inteligentes.

La segunda razón es que la calidad y el volumen de datos consumidos a través de las redes WiFi también están aumentando, y WiFi 6 podrá sostener dicho volumen”⁸.

-
7. Newman Daniel. (2018). Top 10 Digital Transformation Trends For 2019. <https://www.forbes.com/sites/danielnewman/2018/09/11/top-10-digital-transformation-trends-for-2019/#575e1ace3c30> Consultada el 21 de diciembre de 2019.
 8. Newman Daniel. (2019). Top 10 Digital Transformation Trends For 2020. <https://www.forbes.com/sites/danielnewman/2019/07/14/top-10-digital-transformation-trends-for-2020/#4c58dbcd76be> Consultada el 21 diciembre de 2019.

5. **Blockchain va más allá de Crypto.** El acceso a la tecnología de cadena de bloques comienza a masificarse debido al soporte de empresas como Amazon, Samsung, Microsoft, IBM y Alibaba de China. Además, se está comenzando a ver casos de uso real de la tecnología más allá de los pagos y las criptomonedas tales como: transporte de bienes y servicios, seguridad alimentaria, propiedad intelectual y regalías, y gestión de bienes inmuebles/activos.
6. **Analítica como el centro de la ventaja competitiva.** El análisis de datos como elemento fundamental para la dirección estratégica de las empresas parte de que existen datos valiosos respecto a los clientes que deben ser recolectados, analizados y convertidos en estrategias, que permitan tener una ventaja competitiva sobre la competencia. Parte de la preparación del impulso futuro se ve en adquisiciones como la hecha por Salesforce de Tableau; empresas como Alteryx, Splunk, SAS y Qlik y Tibco continuarán con la adquisición, inversión y crecimiento, mediante adquisiciones, empresas conjuntas y/o asociaciones estratégicas.
7. **Inteligencia artificial (*Artificial Intelligence-AI*) y aprendizaje automático (*Machine Learning*) como soporte en el crecimiento del análisis de datos.** Newman (2018) menciona que “los datos recolectados recientemente han demostrado que hemos creado el 90 % de los datos mundiales en el último año, dicha investigación también muestra que solo estamos utilizando el 1 % de los datos de manera efectiva”⁹. Para Newman (2019) “la inteligencia artificial y el aprendizaje automático se convierten en multiplicadores de fuerza para el análisis de datos. El valor de la inteligencia artificial y el aprendizaje automático para el análisis de datos se puede resumir en tres propuestas de valor separadas: velocidad, escala y conveniencia. La velocidad y la escala hablan de la ventaja

9. Newman Daniel. (2018). Top 10 Digital Transformation Trends For 2019. <https://www.forbes.com/sites/danielnewman/2018/09/11/top-10-digital-transformation-trends-for-2019/#575e1ace3c30> Consultada el 21 diciembre de 2018.

de automatizar el análisis de conjuntos de datos masivos en lugar de asignar analistas de datos humanos a la tarea. Gracias a la inteligencia artificial y al aprendizaje automático, los conjuntos de datos complejos ahora se pueden analizar en una fracción del tiempo que solía tomar hace solo dos años. Esto no se debe a que las computadoras se hayan vuelto más rápidas o mejores, sino a que los algoritmos de inteligencia artificial y aprendizaje automático se han vuelto extremadamente buenos en el análisis de datos, y porque ese análisis se puede escalar fácilmente en la nube. Por el lado de la conveniencia, a diferencia de las herramientas de análisis de datos de antaño, la adición de inteligencia artificial y aprendizaje automático a las herramientas de análisis las ha hecho intuitivas, fáciles de usar y mucho más confiables¹⁰.

8. **Realidad aumentada para capacitación (*Augmented Reality-AR*)**. Para Newman (2018), la realidad aumentada (AR) ha encontrado una aplicación en su uso para la capacitación de la fuerza laboral en las empresas.
9. **Automatización robótica de procesos (*Robotic Process Automation-RPA*)**. Esta tecnología consiste en configurar un programa (robot) que emule e integre las acciones de una interacción humana con sistemas digitales para ejecutar un proceso.
10. **Internet de las cosas (*Internet of Things-IoT*), vehículos conectados, drones autónomos y ciudades inteligentes (*Smart Cities*)**. Para Newman (2018-2019), la cantidad de dispositivos conectados se incrementará debido a que conceptos como vehículos conectados y autónomos, el uso de drones autónomos para realizar entregas de paquetería en la última milla y, en general, la búsqueda por implantar el concepto de ciudades inteligentes (*Smart cities*). Para lograr esto, se necesitará un mayor

10. Newman Daniel. (2019). Top 10 Digital Transformation Trends For 2020. <https://www.forbes.com/sites/danielnewman/2019/07/14/top-10-digital-transformation-trends-for-2020/#4c58dbcd76be> Consultada el 21 diciembre de 2019.

espacio de almacenamiento de datos, y el procesamiento de estos en “tiempo real”, cosa que buscará hacerse mediante *Edge to Core*.

11. **Las computadoras que siempre están conectadas (Always Connected Personal Computers-ACPC’S).** El surgimiento de computadoras compatibles con la red 5G y una extraordinaria duración de la batería, es algo que deberá desarrollarse en el corto plazo.
12. **Dirección general como cabeza de la transformación digital.** Generalmente, los proyectos de transformación digital en las empresas son delegados en las áreas de tecnologías de información, recursos humanos y/o mercadotecnia, sin embargo, cada vez comenzaremos a ver más que dichos proyectos sean implantados como parte de la estrategia de las organizaciones por parte de los directores generales (*Chief Executive Officers-CEO’S*).
13. **Servicios de tecnologías de información basados en la estrategia.** Las empresas y organizaciones están seleccionando sus proyectos para implantación de servicios de tecnologías de la información con base en sus necesidades estratégicas.
14. **Todo como servicio (XaaS), experiencia de usuario/cliente (UX/CX) y privacidad digital.** La tendencia a colocar la mayor cantidad de programas como un servicio en la nube (todo como servicio-XaaS) es algo que disminuye costos y hace asequible la tecnología a las pequeñas y medianas empresas. Por otro lado, el éxito de la transformación digital está íntimamente relacionado con la experiencia del usuario y del cliente (UX/CX) y un factor diferenciador para los clientes y usuarios será la privacidad digital, por lo que veremos que más compañías finalmente se vuelven más profesionales sobre cuestiones de privacidad y seguridad de datos.

1.3.3 Soportes, conceptos, metodologías y tecnologías para la transformación Digital

En un estudio realizado por Roland Berger, que es una empresa de consultoría financiada por Siemens en el año 2016 respecto al reto de la transformación digital en España, se identificaron los soportes, conceptos/metodologías y tecnologías que están conduciendo la transformación digital de las empresas. La idea central es que un soporte contiene una serie de conceptos y metodologías que son desarrollados mediante diferentes tecnologías. En la figura 3 se muestra un esquema de los soportes, conceptos, metodologías y tecnologías para la transformación digital, y a continuación se describen:

1. **Acceso digital al cliente.** Los clientes y consumidores comienzan a ser nativos digitales, lo que ha dado paso a un cambio de los modelos de negocio, los más representativos son el entretenimiento (servicios de streaming: Netflix, Prime Video, HBO, Disney); transporte (UBER, Cabify, Didi); turismo (Airbnb); comercio al menudeo (Amazon). De tal forma que los conceptos de comercio electrónico (Ecommerce), plataformas de entretenimiento bajo demanda (Servicios *Over The Top*-OTT) y Logística 4.0, han cobrado relevancia, basados en los teléfonos inteligentes y sus aplicaciones (Android o iOS), y las redes sociales (Facebook, Twitter, Instagram).
2. **Conectividad.** La interconexión de toda la cadena de suministro con el uso de la red de datos permite sincronizar las diferentes logísticas de sus integrantes para disminuir plazos de entrega y ciclos de innovación. Asimismo, permite la conexión de todos los dueños de proceso en el ámbito interno de la empresa, lo que permite agilizar procesos y disminuir plazos de entrega.

Existen dos metodologías relacionadas con la conectividad de la empresa, estas son: los procesos de colaboración para planeación de la demanda y el abastecimiento (*Collaborative Planning Forecast Replenish-*

ment-CPFR), y la planeación de las ventas y las operaciones (*Sales and Operations Planning-S&OP*); mientras que existe un concepto que se relaciona con la conectividad que es la denominada manufactura en la nube (*Cloud Manufacturing*). Las tecnologías en las que se basan estas metodologías y conceptos son el almacenamiento en la nube y la red 5G.

3. **Información digital.** En el diseño de los procesos para las empresas y organizaciones debe considerarse la recolección y análisis de los datos para de esta forma encontrar la información. El contar con acceso a esta información en forma instantánea, accesible y masiva condiciona todo el proceso de toma de decisiones ya que debe ser correctamente analizada para asegurar el éxito y adaptarse a las tendencias que condicionan el mercado. Otro punto por considerar es la medición constante del desempeño de la empresa u organización.

Las metodologías que se relacionan con la información digital son la gestión de los procesos de negocio (*Business Process Management*), la inteligencia analítica de los negocios (*Business Analytics*) y el concepto de grandes volúmenes de datos (*Big Data*). Algunas de las tecnologías en las que se basa son el internet de las cosas (*Internet of Things-IoT*), los dispositivos ponibles (*wearables*), la realidad aumentada (*Augmented Reality-AR*) y la planificación de recursos empresariales (*Enterprise Resources Planning-ERP*).

4. **Automatización de procesos.** La automatización de procesos de manufactura y administrativos le permite a las empresas u organizaciones ganar en autonomía y flexibilidad en el funcionamiento interno y ofrecer un servicio y un producto flexible, adaptado y escalable a las necesidades de los clientes o usuarios. Conceptualmente se habla de los denominados sistemas ciberfísicos (*Ciber Physical Systems-CPS*), las tecnologías con una relación directa son los drones, la robótica avanzada, la simulación, el aprendizaje automático (*Machine Learning/Deep Learning*) y los bots.

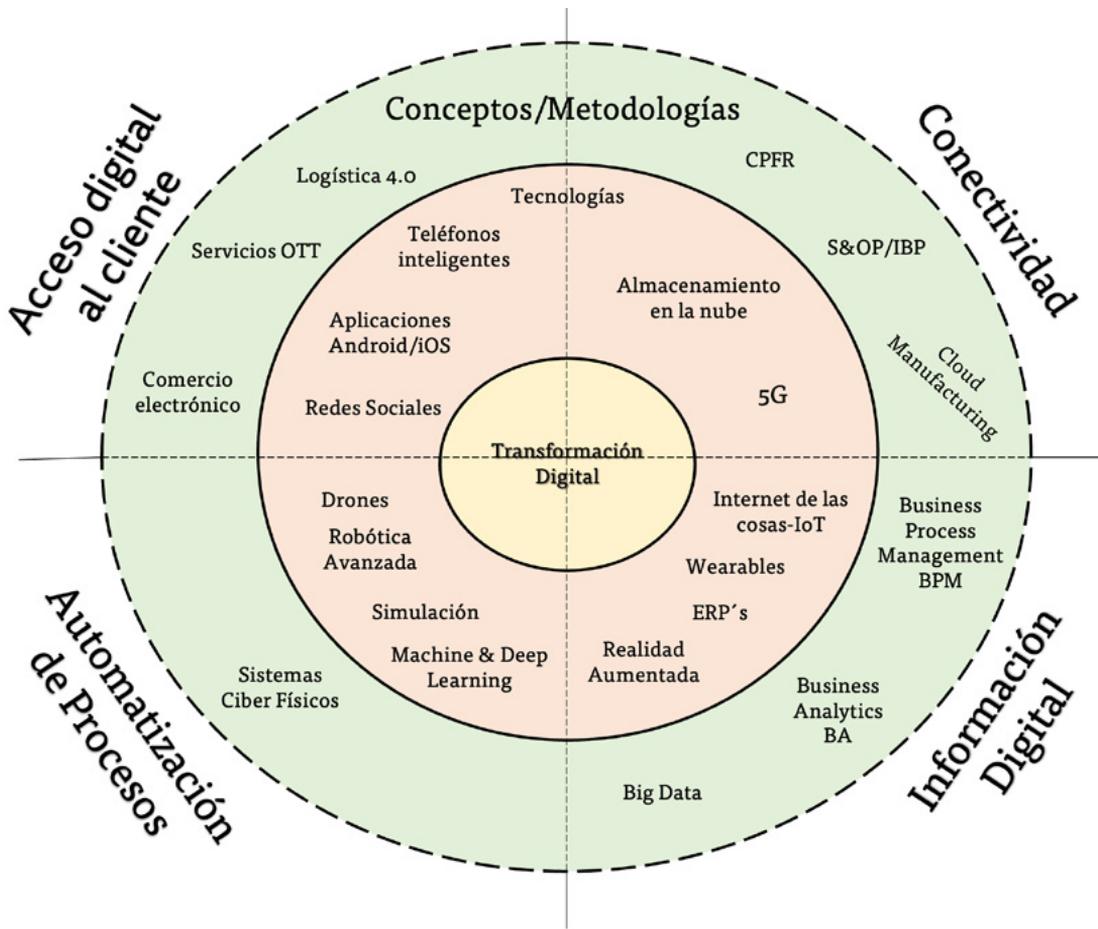


Figura 3. Esquema con los soportes, conceptos, metodologías y tecnologías para la transformación digital. Fuente: Elaboración propia adaptado de Roland Berger/Siemens (2016).

1.3.4 Metodología para la transformación digital

Una de las metodologías propuestas para la transformación digital es la que efectuó la unidad de inteligencia de *The Economist* en el año 2016, la cual consta de tres etapas que consisten en establecer una visión, ver en qué parte se encuentra la empresa u organización y definir la forma en que se llegará del estado actual al estado deseado. En la tabla 1 se muestran las fases y actividades propuestas para la transformación digital, a partir de la propuesta realizada por *The Economist*. A continuación, se desarrollarán cada una de estas de manera más detallada:

Etapa 1. Diseñar un modelo de negocio disruptivo y seleccionar las estrategias de implantación de este.

- **Actividad 1. Diseñar el modelo de negocio.** Consiste en diseñar el modelo de negocio disruptivo al que se aspira llegar.
- **Actividad 2. Evaluación de fortalezas y debilidades.** Consiste en evaluar la situación interna de la empresa u organización con respecto al modelo de negocio que se desea adoptar.
- **Actividad 3. Selección de la estrategia.** Una vez que se conoce el punto de partida es conveniente evaluar las posibles estrategias para seleccionar la(s) estrategia(s) para establecer el modelo de negocio seleccionado al inicio de esta etapa.

Etapa 2. Diseñar e implantar procesos y arquitectura tecnológica.

- **Actividad 4. Mejorar o rediseñar procesos.** Una vez que el modelo de negocio y la estrategia han sido seleccionados es necesario mejorar o rediseñar los procesos de la empresa u organización, así como ubicar la arquitectura tecnológica que se requiere para hacer más eficientes a esos procesos.
- **Actividad 5. Evaluar y seleccionar las diferentes soluciones tecnológicas.** Una vez que se conocen las diferentes necesidades tecnológicas para la ejecución de los procesos, es necesario evaluar y seleccionar cuáles de estas serán implantadas en la compañía.
- **Actividad 6. Conducir proyectos de implantación de soluciones tecnológicas.** La parte más compleja de la etapa 2 consiste en implantar las soluciones tecnológicas seleccionadas, mediante diferentes proyectos que pueden hacerse en paralelo. Si el proceso fue realizado de manera colaborativa con los dueños de los procesos, la resistencia al cambio probablemente sea menor.

Etapa 3. Garantizar la seguridad de los datos.

- **Actividad 7. Diseñar la estrategia de seguridad de datos.** Se debe definir un horizonte temporal para diseñar la estrategia de seguridad de datos que se seguirá.

- **Actividad 8. Evaluación de las diferentes estrategias de seguridad de datos.** Al momento de evaluar las estrategias de seguridad deben considerarse la flexibilidad y escalabilidad de dichas estrategias, así como los costos, tiempos y recursos necesarios en su implantación.
- **Actividad 9. Implantación y control de la seguridad de los datos.** La última actividad de la etapa que permite garantizar la seguridad de los datos consiste en la implantación y el control de la(s) estrategia(s) seleccionadas.

Aunque en el estudio se menciona que el tiempo requerido para ejecutar una transformación digital es variable y depende de las condiciones y características de cada empresa u organización, se debe considerar que la transformación digital puede establecer una ventaja competitiva y se sugiere que se realice entre 6 y 12 meses.

Tabla 1. Metodología propuesta para la transformación digital.
Fuente: Elaboración propia adaptado de Intelligence Unit-The economist (2016).

| Etapa | Diseño | Evaluación | Selección y/o implantación |
|--|---|---|--|
| Modelo de negocio disruptivo y estrategia | Diseñar el modelo de negocio y estrategias. | Evaluar la situación interna de la empresa u organización. | Seleccionar la(s) estrategia(s) para establecer el modelo de negocio seleccionado. |
| Procesos y arquitectura tecnológica | Mejorar o rediseñar los procesos, así como identificar la arquitectura tecnológica. | Evaluar y seleccionar las diferentes soluciones tecnológicas. | Implantación mediante diferentes proyectos de las soluciones tecnológicas. |
| Seguridad de los datos y la información | Diseñar la estrategia de seguridad de datos. | Evaluar las estrategias de seguridad. | Implantar y controlar de la(s) estrategia(s) seleccionadas. |

REFERENCIAS

ECKERSON, W. W., (2002). *The Rise of Analytic Applications: Build or Buy*, The Data Warehouse Institute.

<http://download.101com.com/tdwi/research-report/0402aareport.pdf>

Consultada el 21 de diciembre de 2019.

GARTNER, (2019). *Gartner Glossary*.

<https://www.gartner.com/en/information-technology/glossary/business-intelligence-bi>

Consultada el 20 de diciembre de 2019.

LOMBARDERO, L., *Trabajar en la era digital. Tecnología y competencias para la Transformación Digital*. Madrid: LID Editorial, 2015.

INTELLIGENCE UNIT-THE ECONOMIST, (2016). *The path to self disruption: Nine steps of a digital transformation journey*. The economist/Sponsored by Hewlett Packard.

<https://eiuperspectives.economist.com/technology-innovation/path-self-disruption/article/path-self-disruption-nine-steps-digital-transformation-journey>

Consultada el 21 de diciembre de 2019.

NEWMAN, D., (2018). *Top 10 Digital Transformation Trends For 2019*.

<https://www.forbes.com/sites/danielnewman/2018/09/11/top-10-digital-transformation-trends-for-2019/#575e1ace3c30>

Consultada el 21 de diciembre de 2019.

_____, (2019). *Top 10 Digital Transformation Trends For 2020*.

<https://www.forbes.com/sites/danielnewman/2019/07/14/top-10-digital-transformation-trends-for-2020/#4c58dbcd76be>

Consultada el 21 de diciembre de 2019.

ROLAND BERGER/SIEMENS, (2016). *España 4.0-El reto de la transformación digital de la economía*.

<https://assets.new.siemens.com/siemens/assets/public.1515407804.4fe796280dd1d58ab6eb71e-51f14e13a546c3948.estudio-digitalizacion-espa-a-40.pdf>

Consultada el 21 de diciembre de 2019.

ROUSE, M., (2015). *Transformación digital*.

<https://searchdatacenter.techtarget.com/es/definicion/Transformacion-digital>

Consultada el 21 de diciembre de 2019.

SHARDA, R., Delen, D. & Turban, E., *Business Intelligence, analytics and data science: a managerial perspective*. New York: Pearson, 2018.

SIMON, H., *The new science of management decision*. Englewood Cliffs, NJ: Prentice Hall, 1977.

CAPÍTULO 2

Factores tecnológicos en
la transformación digital
de las organizaciones

2.1 Grandes volúmenes de datos (Big Data)

2.1.1 Definiciones de big data

Un término que ha cobrado relevancia es Big Data, dicho concepto no tiene una traducción directa al español, por lo que el término usado para este texto será *grandes volúmenes de datos*. Para poder entender más el concepto de lo que son los grandes volúmenes de datos (big data), analizaremos un par de definiciones. La primera definición, proviene de la consultora Gartner quien define los grandes volúmenes de datos de la siguiente manera:

“Los grandes volúmenes de datos (*Big Data*) son activos de información de gran volumen, alta velocidad y/o gran variedad que exigen formas rentables e innovadoras de procesamiento de información que permitan una mejor comprensión, toma de decisiones y automatización de procesos”¹.

En esta definición se habla de información y no de datos, que son conceptos diferentes, considerando que para Evans (2017) los datos son hechos numéricos y cifras que se recopilan a través de algún tipo de proceso de medición y la información proviene del análisis de datos, es decir, de extraer el significado de los datos para respaldar la evaluación y la toma de decisiones. Otro punto por considerar en la definición es que se usa la palabra activos, esto muestra la importancia que tiene en las empresas y organizaciones los datos y la información; asimismo, habla de las características que son volumen, velocidad y variedad las cuales se verán con más detalle más adelante. Por último, hace referencia al procesamiento de dicha información y menciona que las formas de procesamiento deben ser rentables (al menor costo) e innovadoras.

1. Gartner. (2019). Gartner Glossary. <https://www.gartner.com/en/information-technology/glossary/big-data> Consultada el 22 de diciembre de 2019.

Otra definición que analizaremos es la hecha por Mckinsey Global Institute en 2011 (Manyika, et al. 2011), la cual define a los grandes volúmenes de datos (big data) como sigue:

“Los grandes volúmenes de datos (Big Data) se refiere a conjuntos de datos cuyo tamaño está más allá de la capacidad de las herramientas típicas de los programas de base de datos para capturar, almacenar, gestionar y analizar”².

De esta definición destacamos que se menciona la característica de volumen, refiriéndose al tamaño del conjunto de datos; otro punto a destacar es que desglosa las actividades (capturar, almacenar, gestionar y analizar) que se ejecutan sobre el conjunto de datos hasta convertir esos datos en información.

Haciendo uso de estas definiciones podríamos proponer la siguiente definición:

Los grandes volúmenes de datos (big data) se refiere a conjuntos de datos que se generan de diferentes fuentes (variedad) en breves lapsos de tiempo (velocidad) y cuyo tamaño (volumen) requiere el uso de formas de procesamiento (recolección almacenamiento, gestión y análisis) especializadas y que sean rentables.

2.1.2 Tipos de datos

Una forma de clasificar los datos consiste en considerar la forma en que estos son recolectados y almacenados, de esta forma se tienen los siguientes tipos de datos:

2. Manyika James, et. al. (2011). Big data: The next frontier for innovation, competition, and productivity. Mckinsey Global Institute. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#> Consultada el 22 de diciembre de 2019.

- **Estructurados (datos tradicionales)**

La mayoría de las fuentes de datos tradicionales son datos estructurados, datos con formato o esquema fijo que poseen campos fijos. En estas fuentes, los datos vienen en un formato bien definido que se especifica en detalle, y que conforma las bases de datos relacionales. Fundamentalmente, son los datos de las bases de datos relacionales, las hojas de cálculo y los archivos.

La gestión y búsqueda de los datos estructurados en las bases de datos relacionales se realizan con el lenguaje de programación estándar SQL –lenguaje creado por IBM en la década de los setenta– y que todavía sigue en vigor y soporta a la mayoría de las bases de datos establecidas en organizaciones y empresas. Sin embargo, las bases de datos relacionales tienen el inconveniente de manejar con poca facilidad datos no estructurados.

- **No estructurados**

Los datos no estructurados (sin estructurar) son datos sin tipos predefinidos. Se almacenan como “documentos” u “objetos” sin estructura uniforme, y se tiene poco o ningún control sobre ellos. Tienen un formato que no puede ser gestionado fácilmente en tablas de bases de datos relacionados. Por ejemplo: videos, audio, imágenes, fotografías, datos de texto, correos electrónicos, mensajes de textos y publicaciones en redes sociales. Algunas herramientas para su manipulación son MapReduce, Hadoop o bases de datos NoSQL.

- **Semi estructurados**

Los datos semiestructurados tienen propiedades de datos estructurados y no estructurados, y pueden tener algún tipo específico de estructura que se puede utilizar en un análisis de datos, pero no contienen la estructura

de un modelo de datos. Asimismo, poseen un flujo lógico y un formato que puede ser definido, pero no es fácil su compresión para el usuario. Son datos que no tienen formatos fijos, pero sí etiquetas y otros marcadores que permiten separar los elementos datos. Por ejemplo, documentos XML de páginas web, contenidos blogs, software de tratamiento de textos, lenguajes de marca de hipertexto extensibles.

2.1.3 Características de los grandes volúmenes de datos (Big Data)

La primera referencia a las características de volumen, velocidad y variedad referida a los datos es de Laney (2001); posteriormente, la consultora Gartner (2011) y Russom (2011) retomaron estas características referidas a lo que son los grandes volúmenes de datos (big data). Las definiciones que se presentan a continuación son las de la consultora Gartner (2011):

- **Volumen:** El aumento en los volúmenes de datos dentro de los sistemas empresariales es causado por los volúmenes de transacciones y otros tipos de datos tradicionales, así como, por nuevos tipos de datos. Demasiado volumen es un problema de almacenamiento, pero demasiados datos también son un problema de análisis masivo.
- **Variedad:** Con la llegada de las redes sociales y los dispositivos móviles, hay más tipos de datos para analizar, estos incluyen datos tabulares (bases de datos), datos jerárquicos, documentos, correo electrónico, datos de medición, video, imágenes fijas, audio, datos de cotizaciones bursátiles, transacciones financieras y más.
- **Velocidad:** Se refiere a los flujos de datos, creación de registros estructurados y disponibilidad para acceso y entrega. Velocidad significa tanto qué tan rápido se producen los datos y qué tan rápido se deben procesar los datos para satisfacer la demanda.

Posteriormente, se agregaron las características de veracidad y valor. A continuación, presentamos dichas definiciones, según Marr (2014):

- **Veracidad:** Se refiere a la confiabilidad de los datos. Con muchos de los datos recolectados, la calidad y la precisión son menos controlables (solo piense en las publicaciones de Twitter con etiquetas hash, abreviaturas, errores tipográficos y habla coloquial, así como la confiabilidad y precisión del contenido), pero la tecnología actual nos permite trabajar con este tipo de datos. Los volúmenes a menudo compensan la falta de calidad o precisión.
- **Valor:** Se refiere a que la recolección de grandes volúmenes de datos debe estar enfocado a generar un beneficio financiero (costo-beneficio) o ventaja competitiva a la empresa u organización.

Conforme han pasado los años se han agregado algunas características, Joyanes (2018) menciona dos características más: viabilidad y visualización

- **Viabilidad:** Se refiere a la capacidad que tienen las empresas de hacer un uso eficaz del gran volumen de datos que manejan.
- **Visualización:** Se refiere al modo en que los datos se presentan para encontrar patrones y claves que permiten la obtención de resultados para una toma de decisión eficiente.

En la tabla 2, se muestra un resumen de la evolución de las características de los grandes volúmenes de datos (big data), en esta tabla se aprecia el o los autores, el año y las características de los datos, las características que se agregan aparecen subrayadas.

Tabla 2. Resumen de la evolución de las características de los grandes volúmenes de datos (big data). Fuente: elaborado por Susana C. Téllez y Ricardo Torres.

| Modelos de las 3 V | Modelos de las 5 V | Modelos de las 7 V |
|--|--|---|
| Doug Laney (2001) Mark Beyer (2011) Philip Russom (2011) | Bernard Marr (2014) IBM (2014) | Luis Joyanes (2017) |
| <u>Volumen</u> <u>Velocidad</u> <u>Variedad</u> | Volumen Velocidad Variedad <u>Valor</u> <u>Veracidad</u> | Volumen Velocidad Variedad Valor Veracidad <u>Visualización</u> <u>Viabilidad</u> |

2.1.4 Tipos de fuentes de los grandes volúmenes de datos (Big Data)

La característica de variedad de los grandes volúmenes de datos ha generado que se generen clasificaciones con base en la fuente de los datos, así, una de las clasificaciones más referenciadas es la generada por Soares (2012), la cual se desglosa a continuación:

- **Red de datos y redes sociales.** Se incluyen todos los datos obtenidos de la red de datos (Internet-Web) y de las redes sociales como Facebook, Twitter, Instagram, LinkedIn, Pinterest, entre otras.
- **Maquina a máquina (Machine to Machine-M2M)/Internet de las cosas.** Se refiere a las tecnologías que permiten la conexión entre diferentes dispositivos, máquinas o robots, mediante el uso de sensores, medidores que capturan datos como: humedad, velocidad, temperatura, presión, salinidad), y también se usan contadores inteligentes (consumo de electricidad).

- **Datos transaccionales.** Son los datos que proceden de operaciones normales de transacciones de todo tipo. Incluyen registros de facturación (*Point Of Sales-POS*), datos de operaciones de almacenes (*Warehouse Management Systems-WMS*), líneas de fabricación (*Manufacturing Requirement Planning- MRP II*) y de todos los módulos que constituyen las gestiones dentro de las organizaciones (Sistemas de Planeación de Recursos Empresariales -*Enterprise Resources Planning-ERP*).
- **Biometría.** Se refiere a la identificación automática de una persona basada en sus características anatómicas o trazos personales, tales como información procedente del cuerpo humano y actividad física (huellas digitales, reconocimiento facial, escaneo de la retina, genética). Los datos anatómicos se crean a partir del aspecto físico de una persona, incluyendo huellas digitales, iris, escaneo de la retina, reconocimiento facial, genética, ADN, reconocimiento de voz, incluso olor corporal.
- **Datos generados por las personas.** Las personas generan enormes y diversas cantidades de datos como son la información que guarda un centro de llamados telefónicos (*call center*) al establecerlas, notas de voz, correos electrónicos, documentos electrónicos, estudios y registros médicos electrónicos, recetas médicas, documentos en papel y faxes.

2.1.5 Iniciativa de datos abiertos

La iniciativa de datos abiertos nació en 2009 en Washington y se refiere a la posibilidad de que el ciudadano acceda a los datos del gobierno que antes solo eran analizados en el interior de las administraciones públicas. Las iniciativas de datos abiertos están creciendo a lo largo del mundo y algunos de los proyectos más innovadores están en Estados Unidos de Norteamérica y Gran Bretaña. En las figuras 4 y 5 se muestra una vista parcial del portal de datos abiertos del gobierno de EUA y del Reino Unido, respectivamente. En los enlaces que se muestran en el pie de figura se pueden consultar los portales completos.

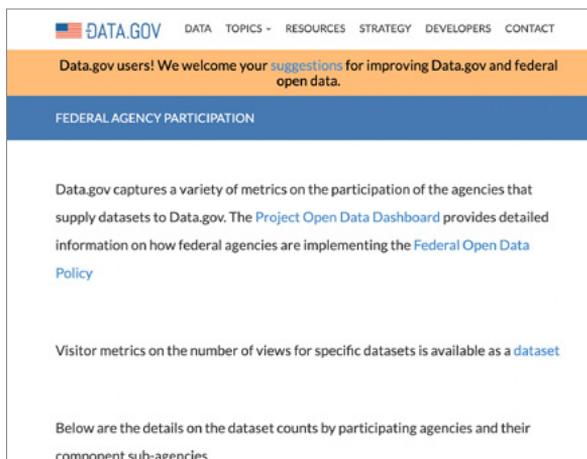


Figura 4. Portal de datos abiertos del gobierno de Estados Unidos de América, extraído el 23 de diciembre 2019 de <https://www.data.gov/metrics>

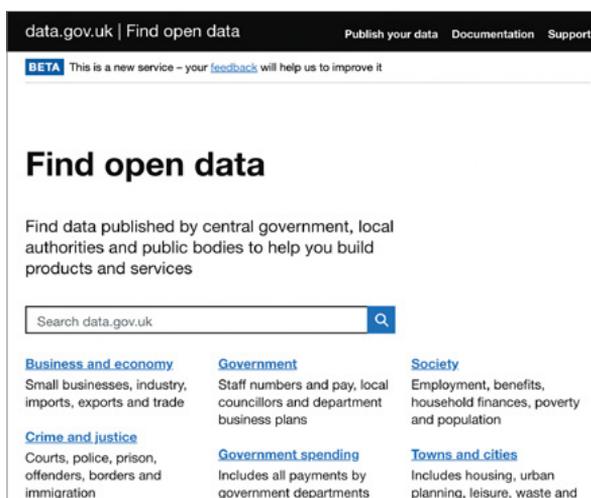


Figura 5. Portal de datos abiertos del gobierno del Reino Unido de la Gran Bretaña, extraído el 23 de diciembre 2019 de <https://data.gov.uk/>

Existen dos organizaciones que son las que están conduciendo esta iniciativa de datos abiertos a nivel mundial, estas son:

- **Alianza Central de Datos Abiertos (Open Data Center Alliance).** Esta organización fue constituida en 2010 y su objetivo principal es trabajar la configuración futura de la computación en la nube (cloud

computing), su página de internet es la siguiente <https://openpower-foundation.org/>.

- **Instituto de Datos Abiertos (Open Data Institute).** Esta organización se constituyó en 2012 por Tim Berners-Lee, creador de internet(web), su objetivo es canalizar la evolución hacia una cultura de datos abiertos para crear valor económico, ambiental y social.

2.2 Analítica de negocios

2.2.1 Definición de analítica

El Instituto de Investigación de Operaciones y Ciencias Directivas (*Institute for Operations Research and the Management Sciences-INFORMS*), define Analítica de Datos (*Analytics*) de la siguiente forma:

“La analítica es el proceso científico de transformación de datos en hallazgos con el fin de tomar mejores decisiones. La analítica es siempre un enfoque impulsado por la acción. Siempre hay que tomar una decisión cuando consideramos hacer análisis. Procedente de un fondo de ciencia de datos y trabajando con muchos estadísticos, a los científicos de datos les encanta analizar datos solo por analizarlos. Sin embargo, es importante asegurarse de que nuestro análisis impulse la acción comercial. En definitiva, queremos que la analítica fortalezca la visión de una organización”³.

3. INFORMS. (2019). Best definition of analytics. <https://www.informs.org/About-INFORMS/News-Room/O.R.-and-Analytics-in-the-News/Best-definition-of-analytics>. Consultada el 23 de diciembre de 2019.

2.2.2 Niveles o tipos de analítica de negocios

INFORMS considera que los datos son la base para comprender lo que está sucediendo, lo que sucederá y cómo aprovecharlo, para esto propone tres niveles de analítica de negocios. Estos tres niveles se identifican como:

- Analítica descriptiva
- Analítica predictiva
- Analítica prescriptiva

- **Analítica descriptiva**

Para Sharda, Delen y Turban (2018), la analítica descriptiva (o de informes) se refiere a identificar la situación de una empresa u organización, mediante el análisis de los datos (series de tiempo) y comprender algunos patrones, tendencias y causas subyacentes de las variables que dichos datos representan. Las actividades que deben realizarse son:

1. Consolidación de las fuentes de datos.
2. Disposición de los datos en un solo sitio donde toda la organización pueda tener disponibilidad (desarrollar un almacén de datos-Data Warehouse).
3. Desarrollar informes, consultas, alertas y tendencias apropiadas utilizando diversas herramientas y técnicas de informes.
4. Desarrollar una visualización de los datos mediante alguna de las herramientas de visualización en el mercado.

- **Analítica predictiva**

Para Sharda, Delen y Turban (2018), la analítica predictiva tiene como objetivo determinar qué sucederá en el futuro asociado a una probabilidad de

ocurrencia o bien con determinada exactitud. Este análisis se basa en técnicas estadísticas, así como en otras técnicas desarrolladas más recientemente que se incluyen en lo que se denomina, de manera general, minería de datos. Algunas de las posibles aplicaciones son: predecir si es probable que el cliente cambie a un competidor (“abandono”), lo que el cliente probablemente compraría a continuación y en qué cantidad (pronóstico), a qué promociones respondería un cliente, entre algunas otras. Se utilizan varias técnicas para desarrollar aplicaciones analíticas predictivas, incluidos varios algoritmos de clasificación. Recientemente han surgido los términos de aprendizaje automático (*Machine Learning*) y aprendizaje profundo (*Deep Learning*), una de sus aplicaciones es la elaboración de pronósticos.

En resumen, la analítica predictiva trata de extraer conocimiento de los datos en forma de patrones, tendencias o modelos que nos proporcione una cierta tendencia sobre potenciales situaciones futuras o el resultado de diferentes alternativas sobre la que tomar una decisión. El objetivo clave es pronosticar lo que sucederá en el futuro a partir del análisis de datos históricos; es decir, la analítica predictiva se construye sobre la analítica descriptiva y usa modelos estadísticos avanzados de minería de datos y aprendizaje automático.

- **Analítica prescriptiva**

Sharda, Delen y Turban (2018) mencionan que el objetivo de la analítica prescriptiva es reconocer lo que está sucediendo, a partir de conocer las variables de decisión, entrada y de desempeño, mediante el modelado de sistemas que permiten soportar la toma de decisiones para lograr el mejor rendimiento posible. Históricamente, este grupo de técnicas de modelado ha sido estudiado por la Investigación de Operaciones y/o Ciencias Directivas y, en general, están destinadas a mejorar u optimizar el rendimiento de un sistema. El objetivo aquí es sugerir una decisión que resulte en una acción específica, las decisiones pueden ser tan simples como el decidir sí o no se ejecutan ciertas actividades, definir cantidades específicas

respecto a las variables de decisión (por ejemplo, el precio de un artículo específico o tarifa aérea por cobrar) o un incluso tan complejas como elaborar un conjunto completo de planes de producción.

La decisión debe tomarse por los directores, gerentes o un tomador de decisiones, por lo que debe elaborarse y presentarse un informe para estos, o bien, los resultados de las técnicas pueden usarse para alimentar directamente un sistema automatizado de reglas de decisión (por ejemplo, en sistemas de precios de aerolíneas).

En la figura 6 se muestra un resumen de lo que es analítica de negocios y los tres tipos de analítica que existen, en ella se muestran las preguntas que cada una de las analíticas ayuda a responder, cuáles son las metodologías, técnicas y herramientas que son usadas en cada analítica y qué resultados se esperan para cada una de ellas.

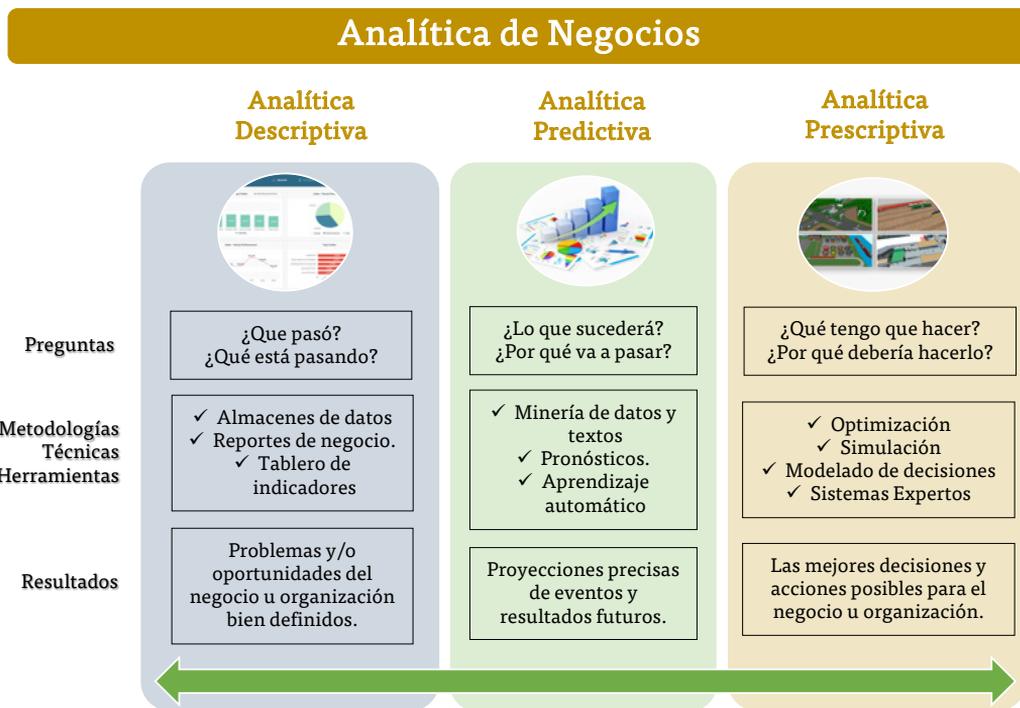


Figura 6. Tres tipos de analítica de negocios. Fuente: Elaboración propia con base en Sharda, Delen y Turban (2018)

2.2.3 Categorías de analítica de negocios con base en la fuente de los datos

Joyanes (2018) propone otra clasificación de analítica de datos en función de los dispositivos y plataformas a utilizar en el proceso de recolección de los datos, dicha clasificación se desarrolla a continuación:

- **Analítica de datos tradicionales.** Hace referencia a la analítica de negocios con base en el análisis de datos tradicionales: transaccionales y operacionales en organizaciones y empresas.
- **Analítica web o analítica de tráfico de datos en un sitio web.** Tecnologías y herramientas de análisis web, utilizadas en servidores, PC, laptop, etc., que han conformado los componentes de los sistemas de información tradicional, base y fundamento de la disciplina de analítica web.
- **Analítica social.** Análisis de datos de los medios sociales (blogs, wikis, redes sociales, RSS, etc.) A medida que los medios sociales comienzan a implantarse en organizaciones y empresas como servicios, al igual que cualquier otro software o servicios, aparece la necesidad de analizar y gestionar los datos procedentes de estas aplicaciones. Surgen nuevas profesiones asociadas a los medios sociales como especialistas en SMO (Social Media Optimization), especialistas en gestión de comunidades conocido popularmente como Community Manager y el SMM (Social Media Manager), gestor o administrador de medios sociales – similar al Community Manager, pero con tareas de administración o dirección–.
- **Analítica móvil.** Aplicada en dispositivos móviles con el objeto de analizar los datos que envían, reciben o transitan desde dichos dispositivos. El despliegue imparable, primero de teléfonos inteligentes (smartphones) y, posteriormente, de las tabletas (tablets), en particular desde el lanzamiento de la tableta iPad de Apple en 2010, ha creado la necesidad de analizar los datos producidos por estos dispositivos

móviles, dado que se están convirtiendo en los puntos de acceso a Internet más empleados en organizaciones y empresas, y también por particulares.

- **Analítica de big data o analítica de los grandes volúmenes de datos.** La gran tendencia que se está produciendo en estos dos y tres últimos años es la explosión de los big data. Esta creciente tendencia ha traído de modo irreversible la necesidad de realizar estudios de analítica web sobre el inmenso tráfico de grandes datos que se están produciendo día a día.

2.3 Ciencia de datos

2.3.1 Definición de ciencia de datos

Para poder entender lo que es ciencia de datos, utilizaremos en primera instancia el diagrama de Venn propuesto por Conway (2010), en dicho diagrama se mencionan las tres disciplinas esenciales que conforman la ciencia de datos son (véase la figura 7), las cuales se describen a continuación:

- **Extracción de datos y generación de códigos (*Hacking skills*).** La mayoría de los datos procederán de fuentes heterogéneas de datos; en consecuencia, deberán tener las habilidades necesarias para poder extraer, ordenar, procesar, analizar y visualizar estos datos. Debe crear los algoritmos necesarios utilizando distintos lenguajes de programación para cada caso concreto.
- **Estadística y matemática (*Math and Statistics Knowledge*).** Una vez extraídos los datos, el científico de datos deberá tener los conocimientos matemáticos necesarios para poder interpretarlos y procesarlos mediante las herramientas más adecuadas. La formación matemática es de gran importancia en la ciencia de datos.

- **Conocimiento del dominio o experiencia en el entorno (*Substantive Expertise*)**. Para poder diseñar y desarrollar el análisis masivo de datos a diferentes casos de uso y aplicación, es necesario conocer el contexto. El científico de datos debe tener un alto conocimiento del entorno que le motive a plantear nuevos escenarios y crear nuevas hipótesis en las que trabajar, siempre cuidando la calidad de los datos, todo ello con la intención de obtener resultados que terminen incrementando el conocimiento del área de trabajo.

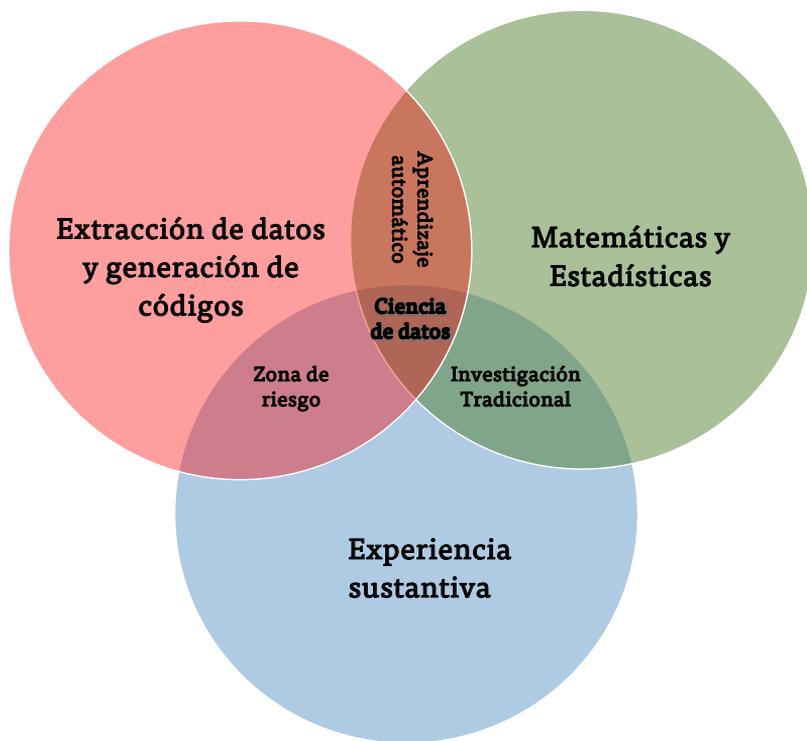


Figura 7. Disciplinas de la ciencia de datos (licencia Creative Commons), 2010. Fuente: Conway Drew (2010)

Posterior a este acercamiento por parte de Conway, IBM (2019) propone la siguiente definición para ciencia de datos:

“Es el proceso por el que se descubren conocimientos ocultos de enormes cantidades de datos estructurados y no estructurados, mediante

la utilización de métodos como la estadística, el aprendizaje automatizado, la minería de datos y el análisis predictivo”⁴.

Finalmente, Eubanks (2016) propone una versión más actualizada del diagrama de Venn propuesto por Conway (2010), cuya aplicación se relaciona directamente con Mercadotecnia, en la figura 8 observamos dicho diagrama. De este diagrama, se destacan el cambio del concepto de zona de riesgo propuesto por Conway por el concepto de ingeniero de datos en el cruce de ‘experiencia sustantiva’ y la ‘extracción de datos y generación de códigos’.

Para Conway es en esta área donde las personas que son perfectamente capaces de extraer y estructurar datos, probablemente relacionadas con un campo del que conocen bastante e incluso dominan alguna herramienta para ejecutar una regresión lineal e informar los coeficientes, carecen de una comprensión de lo que significan esos coeficientes. Es decir, que ya sea por ignorancia o malicia, esta superposición de habilidades les da a las personas la capacidad de crear lo que parece ser un análisis legítimo sin comprender cómo llegaron allí o lo que han creado. En su publicación Eubanks (2016), no hace referencia del porqué denominar *ingeniero de datos* a las personas que cuentan con ambas habilidades, independientemente de esto el riesgo propuesto por Conway sigue vigente no importando el término usado para referirse a dichas personas.

4. IBM. (2019). Ciencia de datos. <https://www.ibm.com/analytics/mx/es/technology/data-science/> Consultada el 22 de diciembre de 2019.

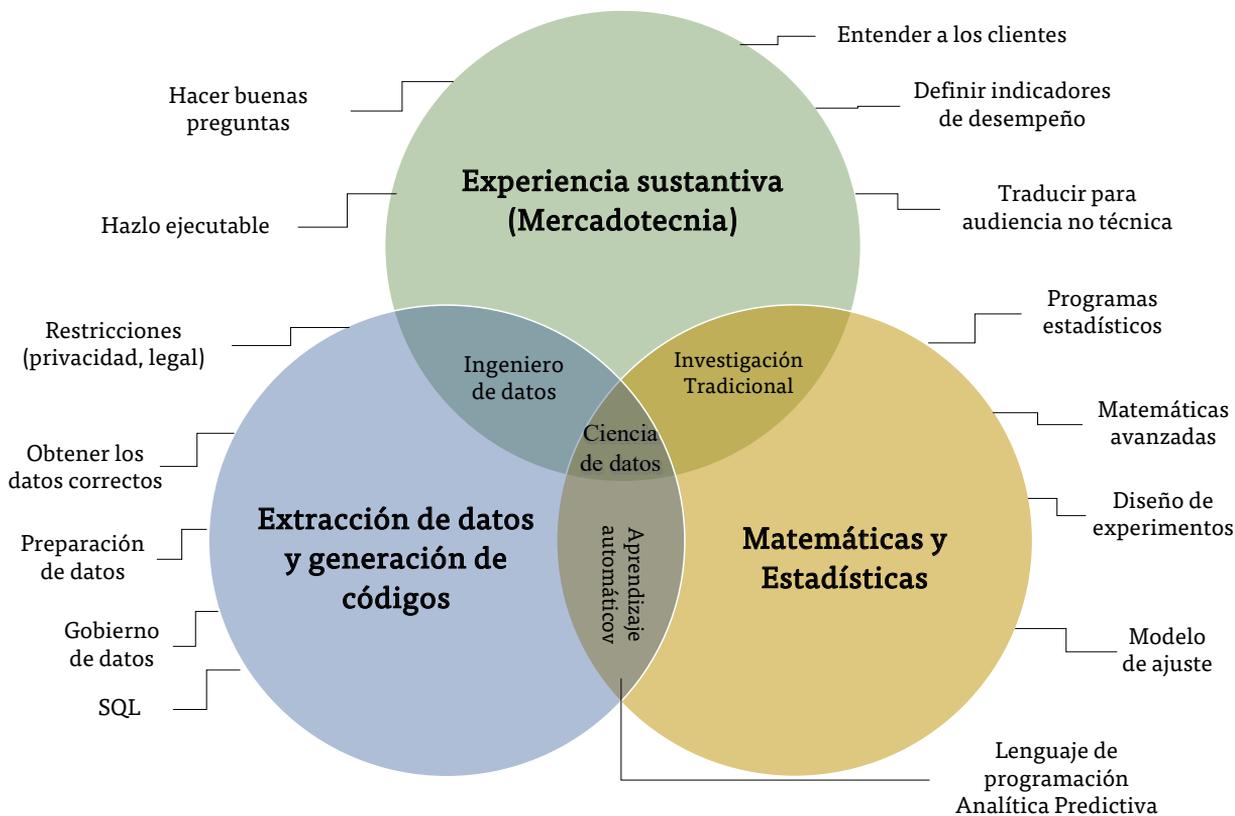


Figura 8. Disciplinas de la ciencia de datos con aplicación a mercadotecnia. **Fuente:** extraído de Eubanks (2016)

2.3.2 Herramientas para ciencia de datos

Existe un gran número de herramientas (programas especializados y lenguajes de programación) que pueden ser empleadas para lo que se ha denominado ciencia de datos. Para ofrecer un panorama general de lo que está sucediendo al respecto, revisaremos los resultados de dos fuentes diferentes, la primera es KDnuggets, que es un portal que se dedica a publicar temas de inteligencia artificial (*Artificial Intelligence-AI*), analítica, grandes volúmenes de datos (*Big Data*), minería de datos (*Data Mining*), ciencia de datos (*Data Science*) y aprendizaje automático (*Machine Learning*), cuyos editores son Gregory Piatetsky-Shapiro y Matthew Mayo; la segunda fuente es el cuadrante mágico de la consultora Gartner.

Primero analizaremos los resultados de la 20ª Encuesta Anual de Programas de KDnuggets que se realizó en 2019, dicha encuesta tuvo 1800 participantes, los cuales en promedio mencionaron 6.1 herramientas diferentes. Para el año 2019, la participación por región fue: para Estados Unidos de América (USA)/Canadá de 40.6, de Europa de 33.3%, de Asia de 12.1%, de América Latina, 7.3 %, de África/Medio Oriente de 4.5% y de Australia/Nueva Zelanda de 2.3%. En la tabla 3 se muestran los resultados de dicha encuesta de los años 2017, 2018 y 2019. Es importante mencionar que para la encuesta no se incluye RStudio y que Piatetsky menciona que para el año 2020 esta herramienta será incluida⁵.

Tabla 3. Principales Programas y Lenguajes de Programación para Analítica/Ciencia de Datos y Aprendizaje Automático de los años 2017, 2018 y 2019. Fuente: Piatetsky (2019)

| Programas y lenguajes de programación | %Participación 2019 | %Participación 2018 | %Participación 2017 |
|---------------------------------------|---------------------|---------------------|---------------------|
| Python | 65.8% | 65.6% | 59.0% |
| RapidMiner | 51.2% | 52.7% | 59.0% |
| Lenguaje R | 46.6% | 48.5% | 56.6% |
| Excel® | 34.8% | 39.1% | 31.5% |
| Anaconda | 33.9% | 33.4% | 24.3% |
| Lenguaje SQL | 32.8% | 39.6% | 39.2% |
| Tensorflow | 31.7% | 29.9% | 22.7% |
| Keras | 26.6% | 22.2% | 10.7% |
| Scikit-learn | 25.5% | 24.4% | 21.9% |
| Tableau | 22.1% | 26.4% | 21.8% |
| Apache Spark | 21.0% | 21.5% | 25.5% |

⁵ No se ha realizado ningún estudio posterior.

Para hacer referencia al cuadrante mágico de la consultora Gartner, haremos uso de las definiciones hechas por Ingeniería, Servicios y Comunicaciones S.A. (2019) la cual menciona que: “El Cuadrante Mágico de Gartner permite a las empresas que contratan servicios y soluciones en TI tener una visión de conjunto de una determinada área de productos o servicios tecnológicos, y poder tomar las mejores decisiones en sus procesos de transformación digital.”

Asimismo, menciona que: “Este cuadrante nos ayuda a saber qué posición ocupan los proveedores y en qué punto de desarrollo se encuentran en cuanto a visión de mercado y poder de implementación, siendo definido para cada sector de Tecnologías de la Información de acuerdo a dos criterios clave: la amplitud de la visión y la capacidad para la ejecución. Estos dos criterios se convierten en los ejes del gráfico, los cuales generan 4 zonas diferenciadas de calificación para definir a las empresas de TI, las cuales son:

- **Los aspirantes (Challengers).** Según la web de Gartner: “tienen una buena ejecución del negocio actualmente y son capaces de dominar un gran segmento de mercado, pero aún no demuestran un real entendimiento de la dirección en que va el mercado”.
- **Los líderes (Leaders).** Desarrollan bien su negocio de acuerdo con una adecuada visión actual del mercado y están bien posicionados para el futuro.
- **Los jugadores de nichos específicos (Niche Players).** Se enfocan con éxito en un segmento de mercado específico, pero muchos de ellos no adquieren en su desempeño una visión global y no se caracterizan por hacer grandes innovaciones o por superar a sus competidores.
- **Los visionarios (Visionaires).** Entienden hacia dónde van los mercados e incluso pueden tener una idea para cambiar las reglas y paradigmas, pero aún no son capaces de llevar a cabo estas ideas por completo o con éxito.

En febrero de 2020, Gartner publicó su cuadrante mágico para plataformas de Ciencia de Datos y Aprendizaje Automático (*Data Science and Machine Learning-DSML platforms*). Para Gartner (2020), una plataforma DSML es “un producto central y una cartera de soporte de productos, componentes, bibliotecas y marcos integrados coherentemente (incluidos los de propiedad, sociedad y código abierto). Sus principales usuarios son profesionales de la ciencia de datos. Estos incluyen científicos expertos en datos, científicos de datos ciudadanos, ingenieros de datos e ingenieros/especialistas en aprendizaje automático (ML)”. En la figura 9 se muestra el cuadrante mágico de Gartner del año 2020 para plataformas de Ciencia de Datos y Aprendizaje Automático (*Data Science and Machine Learning-DSML platforms*).

Para Joyanes (2018), las plataformas para ciencia de datos cuentan con las siguientes características que les permiten ser una opción en el análisis de la información.

- **SQL:** Es el lenguaje de programación estándar de las bases de datos relacionales y que debe conocerse profundamente para realizar la integración con los lenguajes especializados en estadística, aprendizaje automático, lenguaje de procesamiento natural y reconocimiento de voz, entre otros. Es un lenguaje normalizado de código abierto que se utiliza mucho en ciencia de datos, especialmente integrado con otras plataformas y lenguajes de programación. Además, han aparecido versiones de SQL que están comenzado a ser empleados en el desarrollo de big data.
- **R:** Es el lenguaje estadístico por excelencia, y existen programas y paquetes de software para casi todas las materias utilizadas en organizaciones y empresas. Es una herramienta que se emplea mucho en aplicaciones de ciencias de datos y en gestión y desarrollo de big data, y cada vez más en aplicaciones de aprendizajes automático y profundo.
- **Python:** Es la otra gran herramienta de ciencia de datos, con la plataforma R. Una de sus ventajas más importante es la gran cantidad de bibliotecas y aplicaciones estadísticas. También es un lenguaje de programación fácil de aprendizaje que unido a sus grandes bibliotecas lo hace idóneo para desarrollo de aplicaciones de analítica de datos.

- **KNIME** (*Konstanz Information Miner*, www.knime.org). Es una plataforma de minería de datos que se usa como herramienta de inteligencia de negocios y para ciencia de datos. Se creó en la Universidad de Constanza (Alemania) y, posteriormente, se convirtió en una empresa con sede en Zurich (Suiza) donde ofrece servicios generales de desarrollo de la herramienta, formación y consultoría.

2.3.3 Proceso de ciencia de datos

Para Joyanes (2018), el proceso de ciencia de datos se compone de una serie de etapas que tienen cierta semejanza con procesos similares de minería de datos o procesos de inteligencia de negocios e, incluso, ingeniería de software avanzada, véase la figura 9.



Figura 9. Cuadrante mágico para plataformas de ciencia de datos y aprendizaje automático. Fuente: Tomado de Krensky (2020)

Las etapas de un proceso de datos son:

- a. Recolección de los datos en bruto, extraídos del mundo real.
- b. Procesamiento de datos (herramientas como R, Python, SQL, se utilizan en esta etapa en sincronización con las etapas a y b).
- c. Limpieza de datos.
- d. Una vez realizada la limpieza de datos, se deberán hacer varias tareas de modo secuencial o en paralelo.
 - Análisis exploratorio de datos (*EDA-Exploratory Data Analysis*).
 - Realización de algoritmos de aprendizaje automático y creación de modelos estadísticos. Construcción de prototipos teniendo presente datos del mundo real con las realimentaciones necesarias. Construir productos dato.
 - Comunicar, realizar visualizaciones y presentación de informes.
- e. Toma de decisiones.

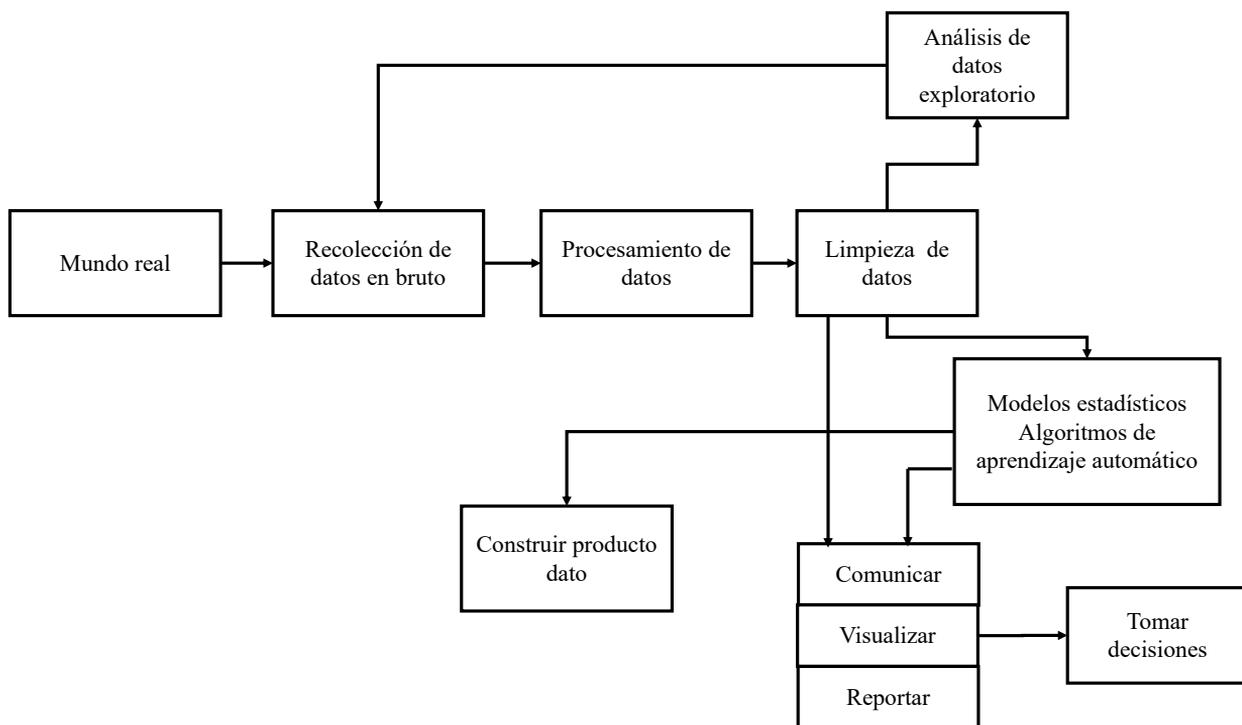


Figura 10. Proceso de ciencia de datos.

Fuente: (Schutt y O’Neill:2014, pag. 41)

2.3.4 Certificaciones profesionales en ciencia de datos

Joyanes (2018) hace referencia a que para las certificaciones profesionales en ciencia de datos (*Data Science*) hay usuarios personales y empresas, y que existen algunos cursos gratuitos en plataformas MOOC (cursos masivos en línea y abiertos) donde es posible formarse y obtener diplomas y certificaciones que acrediten una formación avanzada. Las plataformas más populares y acreditadas son: Coursera, edX y MiriadaX.

Las empresas distribuidoras de big data cuentan con certificaciones profesionales acreditadas por el prestigio de la empresa correspondiente. Como cursos impartidos por empresas como SAP, IBM, Microsoft, HP, SAS, EMC y Coursera.

Otras certificaciones profesionales ya acreditadas son:

- Certified Analytics Professional (CAP)
- Cloudera Certified Professional: Data Scientist (CCP: DS)
- EMC Data Science Associate (EMDCSA)

REFERENCIAS

CONWAY, D., (2010). *The Data Science Venn Diagram*. Drew Conway Data Consulting.

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

Consultada el 23 diciembre de 2019.

EUBANKS, C., (2016). *Three Lessons CrossFit Taught Me About Data Science*. Gartner.

<https://blogs.gartner.com/christi-eubanks/three-lessons-crossfit-taught-data-science/>

Consultada el 23 diciembre de 2019.

EVANS, J., *Business Analytics*. Pearson, 2a ed., University of Cincinnati, 2017.

GARTNER, (2011). *'Big Data' Challenge Involves More Than Just Managing Volumes of Data*.

<https://www.businesswire.com/news/home/20110627005655/en/Gartner-Solving-Big-Data-Challenge-Involves-Managing>

Consultada el 22 diciembre de 2019.

_____, (2019). *Gartner Glossary*.

<https://www.gartner.com/en/information-technology/glossary/big-data>

Consultada el 22 diciembre de 2019.

IBM, (2019). *Ciencia de Datos*.

<https://www.ibm.com/analytics/mx/es/technology/data-science/>

Consultada el 23 diciembre de 2019.

INFORMS, (2019). *Best definition of analytics*.

<https://www.informs.org/About-INFORMS/News-Room/O.R.-and-Analytics-in-the-News/Best-definition-of-analytics>.

Consultada el 23 diciembre de 2019.

INGENIERÍA, SERVICIOS Y COMUNICACIONES S. A., (2019). *¿Qué es el cuadrante Mágico de Gartner y para qué sirve en transformación digital?*

<https://www.isc.cl/que-es-el-cuadrante-magico-de-gartner-transformacion-digital/>

Consultada el 23 diciembre de 2019.

JOYANES, A. L., *Industria 4.0. La cuarta revolución industrial*. México: Alfaomega, 2018.

KHAN, M. A., Uddin, M. F. & Gupta, N., “Seven V’s of Big Data understanding Big Data to extract value,” *Proceedings of the 2014 Zone 1, Conference of the American Society for Engineering Education*, Bridgeport, CT, 2014, pp. 1-5.

KRENSKY, P., et al., (2020). *Magic Quadrant for Data Science and Machine Learning Platforms*, Gartner.

https://www.gartner.com/doc/reprints?id=11YCTPMUL&c-t=200213&st=sb&utm_medium=Website

Consultada el 20 febrero de 2020.

LANEY, D., (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety*. Meta Group (Gartner).

<https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>

Consultada el 22 diciembre de 2019.

MANYIKA, J., et. al., (2011). *Big data: The next frontier for innovation, competition, and productivity*. Mckinsey Global Institute.

<https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#>

Consultada el 22 diciembre de 2019.

MARR, B., (2014). *Big Data: The 5 Vs Everyone Must Know*. LinkedIn.

<https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know/>

Consultada el 22 diciembre de 2019.

PIATETSKY, G., (2019). *Python leads the 11 top Data Science, Machine Learning platforms: Trends and Analysis*. KDnuggets.

<https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html>

Consultada el 23 diciembre de 2019.

RUSSOM, P., (2011). *Big Data Analytics*. The Data Warehouse Institute (TDWI) sponsored by IBM.

https://tdwi.org/research/2011/09/~ /media/TDWI/TDWI/Research/BPR/2011/TDWI_BPReport_Q411_Big_Data_Analytics_Web/TDWI_BPReport_Q411_Big%20Data_ExecSummary.ashx

Consultada el 22 diciembre de 2019.

SCHUTT, R. & O'neil, C., *Doing Data Science*, Sebastopol: O'Reilly, 2014.

CAPÍTULO 3

Almacenamiento de datos (Data Warehousing)

3.1 Definición y conceptos de almacenamiento de datos (Data Warehousing)

El uso del almacenamiento de datos en tiempo real junto con el sistema de apoyo a la toma de decisiones (DSS) y las herramientas de inteligencia de negocios (BI) es una forma importante de llevar a cabo los procesos empresariales. El almacén de datos apoya la toma de decisiones mediante el análisis de grandes cantidades de datos de varias fuentes para proporcionar resultados rápidos y, de esta forma, dar soporte a procesos críticos; con flujos de datos en tiempo real y con el objetivo de ver el estado actual del negocio e identificar rápidamente los problemas, que es el primer y principal paso hacia la resolución analítica.

Los responsables de la toma de decisiones requieren información concisa y confiable sobre las operaciones, tendencias y cambios actuales. Los datos a menudo se fragmentan en sistemas operativos distintos, por lo que los gerentes con frecuencia toman decisiones con información parcial, en el mejor de los casos. El almacenamiento de datos atraviesa este obstáculo accediendo, integrando y organizando datos operativos clave en una forma consistente, confiable, oportuna y fácilmente disponible, dónde y cuándo sea necesario.

Un almacén de datos (DW) es un grupo de datos producidos para apoyar la toma de decisiones; también es un repositorio de datos actuales e históricos de interés potencial para los gerentes en toda la organización. Por lo general, los datos están estructurados para estar disponibles en un formulario listo para actividades de procesamiento analítico (es decir, procesamiento analítico en línea [OLAP], minería de datos, consultas, informes y otras aplicaciones de apoyo a la toma de decisiones). Un almacén de datos es una recopilación de datos no volátil orientada al sujeto, integrada, variante del tiempo y no volátil en apoyo al proceso de toma de decisiones de la administración.

A pesar de que el almacenamiento de datos es un término relativamente nuevo en tecnología de la información, sus raíces se pueden rastrear en el tiempo, incluso antes de que las computadoras fueran ampliamente utilizadas. A principios de la década de 1900, las personas utilizaban datos (aunque sobre todo a través de métodos manuales) para formular tendencias para ayudar a los usuarios empresariales a tomar decisiones informadas, que es el propósito más predominante del almacenamiento de datos.

Las motivaciones que llevaron al desarrollo de tecnologías de almacenamiento de datos se remontan a la década de 1970, cuando el mundo de la computación estaba dominado por una unidad central (o mainframes). Las aplicaciones de procesamiento de datos empresariales, las que se ejecutan en los mainframes corporativos, tenían estructuras de archivos complicadas utilizando bases de datos de generación temprana (no las bases de datos relacionales orientadas a tablas que la mayoría de las aplicaciones utilizan hoy en día) en las que almacenaban datos. Aunque estas aplicaciones hicieron un trabajo decente al realizar funciones de procesamiento de datos transaccionales de rutina, los datos creados como resultado de estas funciones (como información sobre los clientes, los productos que ordenaron y cuánto dinero gastaron) se quedaron en las profundidades de los archivos y bases de datos. Cuando se necesitaba información agregada, como las tendencias de ventas por región y por tipo de producto, se tenía que solicitar formalmente al departamento de procesamiento de datos, donde se ponía en una lista de espera con un par de cientos de otras solicitudes de informe. A pesar de que existía la necesidad de información y los datos utilizados para generarla, la tecnología de la base de datos no estaba allí para satisfacerla. La tabla 4 muestra algunos de los eventos significativos que llevaron al desarrollo del almacenamiento de datos.

Tabla 4. Eventos característicos del desarrollo del almacenamiento de datos

Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

| | |
|-----------------------|--|
| Década de 1970 | <ul style="list-style-type: none"> • Computadoras mainframe • Entrada simple de datos • Informes de rutina • Estructuras de bases de datos primitivas • Teradata incorporado |
| 1980 | <ul style="list-style-type: none"> • Computadoras personales (PC) • Aplicaciones empresariales para PC • Distribución DBMS • DBMS relacional • Teradata envía datos de datos por cable comerciales • Software de código abierto SaaS, PaaS, Cloud Computing |
| 1990 | <ul style="list-style-type: none"> • Almacenamiento centralizado de datos • Nace el almacenamiento de datos • Bill Inmon escribe el libro “Construyendo el almacén de datos” • Ralph Kimball escribe el libro “El kit de herramientas de almacenamiento de datos” • Diseño de arquitectura EDW |
| 2000 | <ul style="list-style-type: none"> • Datos de datos, en aumento exponencialmente datos web • Consolidación de la industria DW/BI • Surgieron aplicaciones de almacenamiento de datos • Inteligencia de negocios apuntalada • Modelado predictivo y de desplazamiento de datos • Almacén de datos empresariales |
| 2010 | <ul style="list-style-type: none"> • Análisis de big data • Analíticos en las redes sociales • Análisis de texto y Web • Hadoop, MapReduce, NoSQL • En memoria, en la base de datos |

Más tarde en el siglo pasado, las empresas comerciales de hardware y software comenzaron a surgir con soluciones a este problema. Entre 1976 y 1979, el concepto de una nueva empresa, Teradata, surgió de la investigación

en el Instituto de Tecnología de California (Caltech), impulsado de las discusiones con el grupo de tecnología avanzada de Citibank. Los fundadores trabajaron para diseñar un sistema de gestión de bases de datos para el procesamiento paralelo con múltiples microprocesadores, dirigido específicamente para el apoyo a la toma de decisiones. Teradata se incorporó el 13 de julio de 1979, y comenzó en un garaje en Brentwood, California. El nombre Teradata se eligió para simbolizar la capacidad de administrar terabytes (billones de bytes) de datos.

La década de 1980 fue la década de las computadoras personales y mini-computadoras. Las aplicaciones informáticas reales ya no solo estaban en mainframes; estaban por todas partes, dondequiera que miraras en una organización. Eso condujo a un problema llamado islas de datos. La solución a este problema llevó a un nuevo tipo de software, llamado sistema de gestión de bases de datos distribuidas, que extraería mágicamente los datos solicitados de las bases de datos en toda la organización, traería todos los datos de vuelta al mismo lugar, y luego los consolidaría, ordenaría y haría cualquier otra cosa que fuera necesaria para responder a la pregunta del usuario. Aunque el concepto era bueno y los primeros resultados de la investigación eran prometedores, los resultados fueron simples. Lamentablemente no funcionaban eficientemente y el problema de las islas de datos todavía existía.

Mientras tanto, Teradata comenzó a enviar productos comerciales para resolver este problema. Wells Fargo Bank recibió el primer sistema de pruebas Teradata en 1983, un RDBMS paralelo (sistema de administración de bases de datos relacionales) para el apoyo a la toma de decisiones, el primero del mundo. En 1984, Teradata lanzó una versión de producción de su producto y, en 1986, la revista Fortune nombró a Teradata “Producto del Año”. Teradata, que todavía existe hoy en día, construyó el primer dispositivo de almacenamiento de datos, una combinación de hardware y software para resolver las necesidades de almacenamiento de datos de muchos. Otras empresas comenzaron a formular sus estrategias, también.

Durante la década de 1980 ocurrieron otros acontecimientos, convirtiéndolo colectivamente en la década de la innovación en el almacenamiento de datos. Por ejemplo, Ralph Kimball fundó Red Brick Systems en 1986. Red Brick comenzó a emerger como una empresa de software visionaria discutiendo cómo mejorar el acceso a los datos; en 1988, Barry Devlin y Paul Murphy de IBM Ireland introdujeron el término almacén de datos empresariales como un componente clave de los sistemas de información empresarial.

En la década de 1990 surgió un nuevo enfoque para resolver el problema de las islas de datos. Si el enfoque de la década de 1980 de llegar y acceder a los datos directamente desde los archivos y bases de datos no funcionó, la filosofía de la década de 1990 implicaba volver a la década de 1970, en la que los datos de esos lugares se copiaban a otra ubicación, solo haciéndolo bien esta vez; por lo tanto, nació el almacenamiento de datos. Muchas personas reconocen a Bill como el padre del almacenamiento de datos por su libro “Building the Data Warehouse”. Surgieron publicaciones adicionales, incluyendo el libro de 1996 de Ralph Kimball, “The Data Warehouse Toolkit”, que discutió técnicas de diseño dimensional de propósito general para mejorar la arquitectura de datos para sistemas de apoyo a la toma de decisiones centrados en consultas.

En el año 2000, la popularidad como la cantidad de datos continuaron creciendo. La comunidad de proveedores y las opciones comenzaron a consolidarse. En 2006, Microsoft adquirió ProClarity, saltando al mercado de almacenamiento de datos. En 2007, Oracle compró Hyperion, SAP adquirió Business Objects e IBM se fusionó con Cognos. Los líderes de almacenamiento de datos de la década de 1990 han sido absorbidos por algunos de los mayores proveedores de soluciones de sistemas de información en el mundo. Durante este tiempo, surgieron otras innovaciones, incluyendo dispositivos de data warehouse de proveedores como Netezza (adquirido por IBM), Greenplum (adquirido por EMC), DATAlegro (adquirido por Microsoft) y dispositivos de administración de desempeño que permitieron supervisión del rendimiento. Estas soluciones innovadoras proporcionaban ahorros de costos porque eran compatibles con conectores para soluciones de almacenamiento de datos heredadas.

En la década de 2010 surge el Big Data. Muchos creen que el big data va a tener un impacto en el almacenamiento de datos tal como lo conocemos. O bien, encontrarán una manera de coexistir (que parece ser el caso más probable, al menos durante varios años) o big data (y las tecnologías que vienen con él) harán que el almacenamiento de datos tradicional sea obsoleto. Las tecnologías que venían con big data incluyen Hadoop, MapReduce, NoSQL y Hive.

El almacenamiento de datos depende de sus características, principalmente se puede diferenciar en las siguientes:

- **Orientado al sujeto.** Los datos están organizados por temas detallados, como ventas, productos o clientes, que contienen solamente información relevante para el apoyo a la toma de decisiones. La orientación del asunto permite a los usuarios determinar no solo el rendimiento de su negocio, sino también por qué. Un almacenamiento de datos difiere de una base de datos operativa en que la mayoría de las bases de datos operativas tienen una orientación de producto y se ajustan para controlar las transacciones que actualizan la base de datos. La orientación del sujeto proporciona una visión más completa de la organización.
- **Integrado.** La integración está estrechamente relacionada con la orientación del sujeto. Los almacenes de datos deben colocar datos de orígenes diferentes en un formato coherente. Para ello, deben ocuparse de los conflictos de nomenclatura y las discrepancias entre las unidades de medida. Se presume que un almacén de datos está totalmente integrado.
- **Variante de tiempo (serie temporal).** Un almacén mantiene datos históricos. Los datos no proporcionan necesariamente el estado actual (excepto en sistemas en tiempo real). Detectan tendencias, desviaciones y relaciones a largo plazo para la previsión y las comparaciones, lo que lleva a la toma de decisiones. Cada almacén de datos tiene una calidad temporal. El tiempo es la única dimensión importante que todos los almacenes de datos deben admitir. Los datos para el análisis de

varias fuentes contienen varios puntos de tiempo (por ejemplo, vistas diarias, semanales y mensuales).

- **Permanente.** Una vez introducidos los datos en un almacén de datos, los usuarios no pueden cambiar ni actualizar los datos. Los datos obsoletos se descartan y los cambios se registran como nuevos datos.

Estas características permiten ajustar los almacenes de datos casi exclusivamente para el acceso a los datos. Algunas características adicionales pueden incluir lo siguiente:

- **Basado en Web.** Los almacenes de datos suelen estar diseñados para proporcionar un entorno informático eficaz para aplicaciones basadas en Web.
- **Relacional/multidimensional.** Un almacén de datos utiliza una estructura relacional o una estructura multidimensional.
- **Cliente/servidor.** Un almacén de datos utiliza la arquitectura cliente/servidor para proporcionar un acceso sencillo a los usuarios finales.
- **En tiempo real.** Los almacenes de datos más recientes proporcionan capacidades de análisis y acceso a datos en tiempo real o activos.
- **Incluir metadatos.** Un almacenamiento de datos contiene metadatos (datos sobre datos) sobre cómo se organizan los datos y cómo usarlos eficazmente.

Mientras que un almacén de datos es un repositorio de datos, el almacenamiento de datos es literalmente todo el proceso. El almacenamiento de datos es una disciplina que da como resultado aplicaciones que proporcionan capacidad de apoyo a la toma de decisiones, permiten un acceso rápido a la información empresarial y crean información empresarial.

Los tres tipos principales de almacenes de datos son los Data Marts, los almacenes de datos operativos (ODS) y los almacenes de datos empresariales (EDW). A continuación, se describen estos tres tipos de almacenes, también se describen los metadatos.

Data Mart

Mientras que un almacén de datos combina bases de datos en toda una empresa, un data mart suele ser más pequeño y se centra en un tema o departamento en particular. Un data mart es un subconjunto de un almacén de datos, que normalmente consta de una única área de asunto (por ejemplo, marketing, operaciones). Un data mart puede ser dependiente o independiente. Un data mart dependiente es un subconjunto que se crea directamente desde el almacenamiento de datos. Tiene las ventajas de usar un modelo de datos coherente y proporcionar datos de calidad. Los data marts dependientes admiten el concepto de un único modelo de datos de toda la empresa, pero el almacenamiento de datos debe construirse primero. Un data mart dependiente garantiza que el usuario final esté viendo la misma versión de los datos a los que tienen acceso todos los demás usuarios del almacenamiento de datos. El alto costo de los almacenes de datos limita su uso a las grandes empresas. Como alternativa, muchas empresas utilizan una versión reducida de menor costo de un almacén de datos denominado data mart independiente, el cual es un pequeño almacén diseñado para una unidad de negocio estratégica (SBU) o un departamento, pero su fuente no es un EDW.

Almacenes de datos operativos (ODS)

Un almacén de datos operativo (*Operational Data Stores* -ODS) proporciona una forma bastante reciente de archivo de información del cliente (CIF). Este tipo de base de datos se utiliza a menudo como un área de ensayo provisional para un almacenamiento de datos. A diferencia del contenido estático de un almacén de datos, el contenido de un ODS se actualiza a lo largo de las operaciones comerciales. Un ODS se utiliza para decisiones a corto plazo que implican aplicaciones de misión crítica en lugar de las decisiones a medio y largo plazo asociadas con una EDW. Un ODS es similar a la memoria a corto plazo ya que almacena solo información

muy reciente. En comparación, un almacenamiento de datos es como la memoria a largo plazo porque almacena información permanente. Un ODS consolida los datos de varios sistemas de origen y proporciona una vista integrada casi en tiempo real de los datos volátiles y actuales. Los procesos de intercambio, transferencia y carga (ETL) (discutidos más adelante en este capítulo) para un ODS son idénticos a los de un almacén de datos. Por último, los open marts (véase Imhoff, 2001) se crean cuando los datos operativos deben analizarse de forma multidimensional. Los datos de un open mart provienen de un ODS.

Almacén de datos empresariales (EDW)

Un almacén de datos empresariales (*Enterprise Data Warehouses* -EDW) es un almacén de datos a gran escala que se utiliza en toda la empresa para el soporte de decisiones. La naturaleza a gran escala proporciona la integración de datos de muchas fuentes en un formato estándar para aplicaciones eficaces de soporte de decisiones e inteligencia de negocios. EDW se utiliza para proporcionar datos para muchos tipos de sistemas de soporte de decisiones, incluyendo la gestión de proveedores (CRM), gestión de la cadena de suministro (SCM), gestión del rendimiento empresarial (BPM), monitoreo de la actividad empresarial (BAM), gestión del ciclo de vida del producto (PLM), gestión de ingresos y, a veces, incluso conocimiento de sistemas de gestión (KMS).

Metadatos

Los metadatos se pueden describir como datos sobre datos. Los metadatos describen la estructura y cierto significado de los datos, contribuyendo así a su uso eficaz o ineficaz. Los metadatos se definen, generalmente, en términos de uso como metadatos técnicos o empresariales. Se puede diferenciar entre metadatos sintácticos (es decir, datos que describen la sintaxis de los datos), metadatos estructurales (es decir, datos que describen la

estructura de los datos) y metadatos semánticos (es decir, datos que describen el significado de los datos en un dominio específico).

A continuación, explicamos los patrones de metadatos tradicionales y la información sobre cómo implementar una estrategia de metadatos eficaz a través de un enfoque holístico para la integración de metadatos empresariales. El enfoque incluye registros de ontología y metadatos; integración de información empresarial (EII); extracción, transformación y carga (ETL); y arquitecturas orientadas al servicio (SOA). Eficacia, extensibilidad, reutilizabilidad, interoperabilidad, eficiencia y rendimiento, evolución, derecho, flexibilidad, segregación, interfaz de usuario, control de versiones, versatilidad y bajo costo de mantenimiento son algunos de los requisitos clave para construir una empresa exitosa basada en metadatos.

Según Kassam (2002), los metadatos empresariales comprenden información que aumenta nuestra comprensión de los datos tradicionales (es decir, estructurados). El propósito principal de los metadatos debe ser proporcionar contexto a los datos notificados; es decir, suministrar información enriquecedora que conduzca a la creación de conocimiento. Los metadatos empresariales, aunque difíciles de proporcionar de forma eficiente, liberan más del potencial de los datos estructurados. El contexto no tiene por qué ser el mismo para todos los usuarios. En muchos sentidos, los metadatos ayudan a la conversión de datos e información en conocimiento.

El diseño, la creación y el uso de metadatos (datos descriptivos o resúmenes sobre los datos) y sus estándares de acompañamiento pueden implicar problemas éticos. Hay consideraciones éticas involucradas en la recopilación y propiedad de la información contenida en los metadatos, incluidos los problemas de privacidad y propiedad intelectual que surgen en las etapas de diseño, recopilación y difusión.

3.2 Proceso general de almacenamiento de datos

Las organizaciones, privadas y públicas recopilan continuamente datos, información y conocimientos a un ritmo cada vez más acelerado y los almacenan en sistemas computarizados. Mantener y utilizar estos datos e información se vuelve extremadamente complejo, especialmente, a medida que surgen problemas de escalabilidad. Además, el número de usuarios que necesitan acceder a la información sigue aumentando como resultado de la mejora de la fiabilidad y disponibilidad del acceso a la red, especialmente Internet. Trabajar con múltiples bases de datos, integradas en un almacén de datos o no, se ha convertido en una tarea extremadamente difícil que requiere una considerable experiencia, pero puede proporcionar enormes beneficios que superan con creces su costo. Como ejemplo ilustrativo, la figura 11 muestra los beneficios empresariales del almacén de datos empresarial creado por Teradata para un importante fabricante de automóviles.

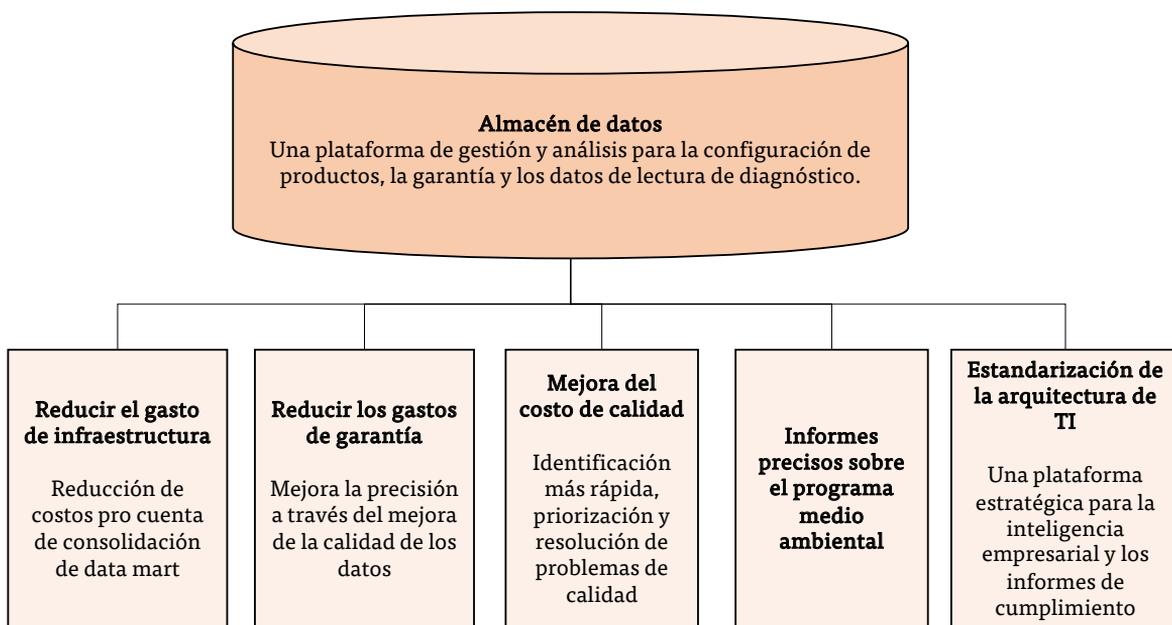


Figura 11. Toma de decisiones basada en datos: beneficios empresariales del almacenamiento de datos

Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

Muchas organizaciones necesitan crear almacenes de datos: almacenes masivos de datos de series temporales para el apoyo a la toma de decisiones. Los datos se importan de diversos recursos externos e internos, y se limpian y organizan de manera coherente con las necesidades de la organización (véase la figura 12). Después de rellenar los datos en el almacenamiento de datos, los data marts se pueden cargar para un área o departamento específico. Alternativamente, los data marts se pueden crear primero, sin embargo, a menudo, los data marts no se desarrollan, sino que los datos simplemente se cargan en los PC o se dejan en su estado original para su manipulación directa mediante herramientas de BI.

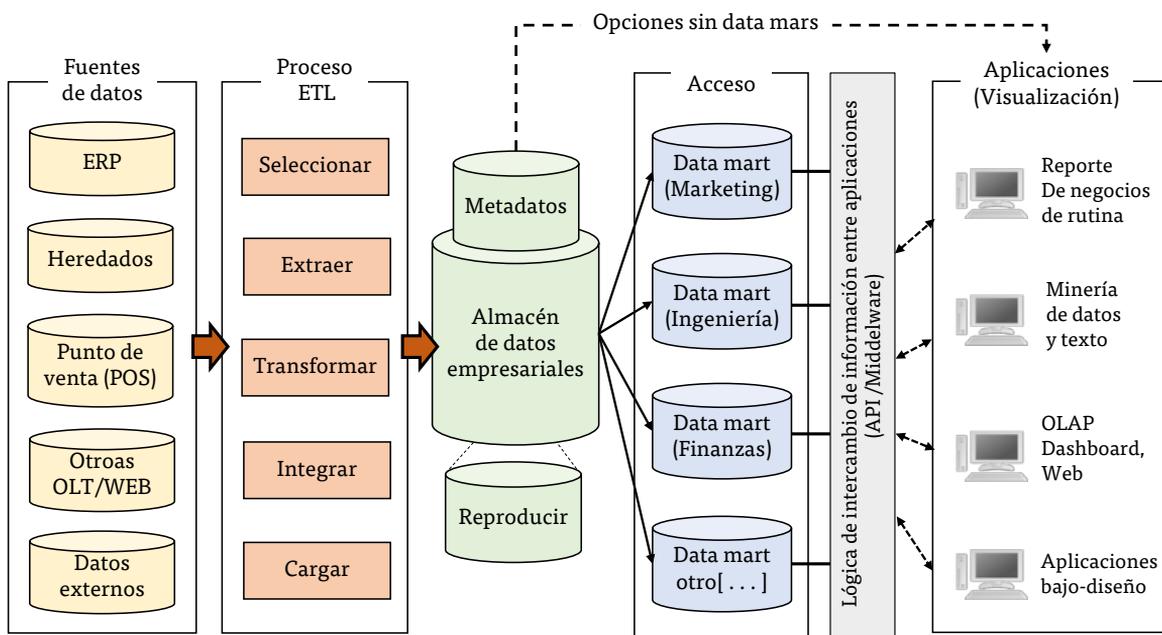


Figura 12. Ambiente y vista del almacén de datos
Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

- **Fuentes de datos.** Los datos proceden de múltiples sistemas operativos independientes “heredados” y posiblemente de proveedores de datos externos. Los datos también pueden provenir de un sistema OLTP o ERP. Datos externos en forma de registros web.

- **Extracción y transformación de datos.** Los datos se extraen y transforman correctamente utilizando software comercial o escrito a medida llamado ETL.
- **Carga de datos.** Los datos se cargan en un área de ensayo, donde se transforman y limpian. A continuación, los datos están listos para cargarse en el almacenamiento de datos y/o en los data marts.
- **Base de datos completa.** Esencialmente, esta es la EDW para apoyar todo el análisis de decisiones proporcionando información resumida y detallada relevante procedente de muchas fuentes diferentes.
- **Metadatos.** Los metadatos se mantienen para que puedan ser evaluados por el personal de tecnologías de la información y los usuarios. Los metadatos incluyen programas de software sobre datos y reglas para organizar resúmenes de datos que son fáciles de indexar y buscar, especialmente con herramientas web.
- **Herramientas de middleware.** Las herramientas de middleware permiten el acceso al almacenamiento de datos. Los usuarios avanzados, como los analistas, pueden escribir sus propias consultas SQL. Otros pueden emplear un entorno de consulta administrado, como objetos de negocio, para tener acceso a los datos. Hay muchas aplicaciones de interfaz que los usuarios empresariales pueden usar para interactuar con los datos almacenados en los repositorios de datos, incluida la minería de datos, OLAP, herramientas de informes y herramientas de visualización de datos.

3.3 Arquitecturas de almacenamiento de datos

Hay varias arquitecturas básicas del sistema de información que se pueden utilizar para el almacenamiento de datos (*Data Warehousing*). En términos generales, estas arquitecturas se denominan comúnmente arquitecturas cliente/servidor o *n* niveles, de las cuales las arquitecturas de dos y tres niveles son las más comunes, pero a veces simplemente hay un nivel. Se sabe que estos tipos de arquitecturas de varios niveles son capaces de satisfacer las necesidades de sistemas de información a gran escala y exigentes

para el rendimiento, como los almacenes de datos. Refiriéndose al uso de arquitecturas de n niveles para el almacenamiento de datos, Hoffer et al. distinguieron entre estas arquitecturas dividiendo el almacenamiento de datos en tres partes:

1. El propio almacén de datos, que contiene los datos y el software asociado.
2. Software de adquisición de datos (back-end), que extrae datos de sistemas heredados y fuentes externas, los consolida y los resume, y los carga en el almacén de datos.
3. Software de cliente (front-end), que permite a los usuarios acceder y analizar datos desde el almacén (un motor DSS/BI/business analytics).

En una arquitectura de dos niveles (véase figura 13), el motor DSS se ejecuta físicamente en la misma plataforma de hardware que el almacenamiento de datos, por lo tanto, es más económico que la estructura de tres niveles. La arquitectura de dos niveles puede tener problemas de rendimiento para grandes almacenes de datos que funcionan con aplicaciones de uso intensivo de datos para el soporte de decisiones.

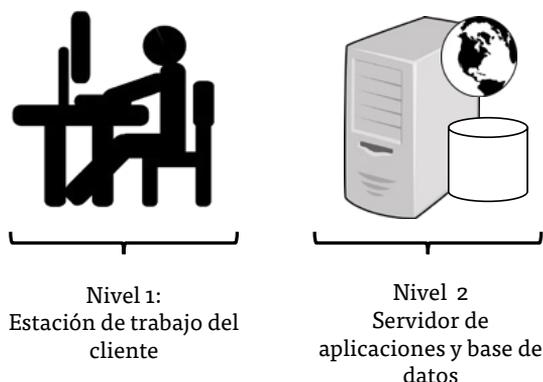


Figura 13. Arquitectura de un almacén de datos de dos niveles

Fuente: Sharda, Delen, & Turban, 2014.

En una arquitectura de tres niveles, los sistemas operativos contienen los datos y el software para la adquisición de datos en un nivel (es decir, el servidor), el almacenamiento de datos es otro nivel y el tercer nivel incluye el motor DSS/BI/BA (es decir, el servidor de aplicaciones) y el cliente. Los

datos del almacén se procesan dos veces y se depositan en una base de datos multinivel adicional, se organizan para facilitar el análisis y la presentación multidimensionales, o se replican en data marts. La ventaja de la arquitectura de tres niveles (véase la figura 14) es su separación de las funciones del almacenamiento de datos, lo que elimina las restricciones de recursos y permite crear fácilmente data marts.

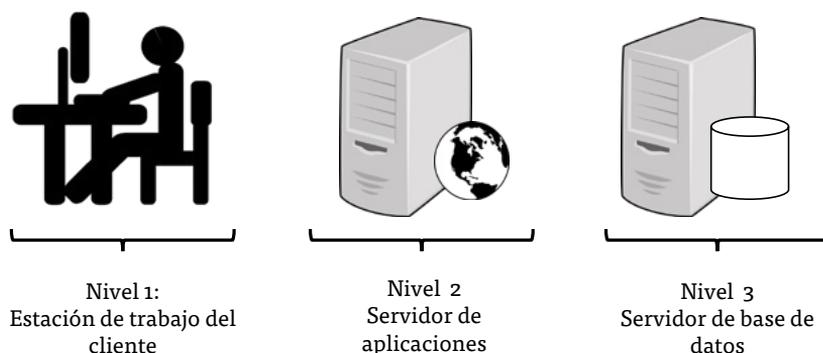


Figura 14. Arquitectura de un almacén de datos de tres niveles
Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

Consultores y proveedores de software se centran en una parte de la arquitectura, limitando así su capacidad y motivación para ayudar a una organización a través de las opciones basadas en sus necesidades. Pero estos aspectos están siendo cuestionados y analizados. Por ejemplo, Ball (2005) proporcionó criterios de decisión para las organizaciones que planean implementar una aplicación de BI y ya han determinado su necesidad de data marts multidimensionales, pero necesitan ayuda para determinar la arquitectura en niveles adecuada. Sus criterios giran en torno a las necesidades de previsión de espacio y velocidad de acceso.

El almacenamiento de datos e Internet son dos tecnologías clave que ofrecen soluciones importantes para la gestión de datos corporativos. La integración de estas dos tecnologías produce almacenamiento de datos basado en Web. En la figura 15, se muestra la arquitectura del almacenamiento de datos basado en Web. La arquitectura es de tres niveles e incluye el cliente de PC, el servidor web y el servidor de aplicaciones. En el lado

del cliente, el usuario necesita una conexión a internet y un navegador web (preferiblemente habilitado para Java) a través de la interfaz gráfica de usuario (GUI) familiar. Internet/intranet/extranet es el medio de comunicación entre el cliente y los servidores. En el lado del servidor, se utiliza un servidor web para administrar la entrada y la salida de información entre el cliente y el servidor. Está respaldado por un almacén de datos y un servidor de aplicaciones. El almacenamiento de datos basado en la web ofrece varias ventajas convincentes, incluida la facilidad de acceso, la independencia de la plataforma y un menor costo.

Las arquitecturas web para el almacenamiento de datos son similares en estructura a otras arquitecturas de almacenamiento de datos, lo que requiere una opción de diseño para alojar el almacén de datos web con el servidor de transacciones o como un servidor o servidores independientes. La velocidad de carga de páginas es una consideración importante en el diseño de aplicaciones basadas en Web; por lo tanto, la capacidad del servidor debe planificarse cuidadosamente.

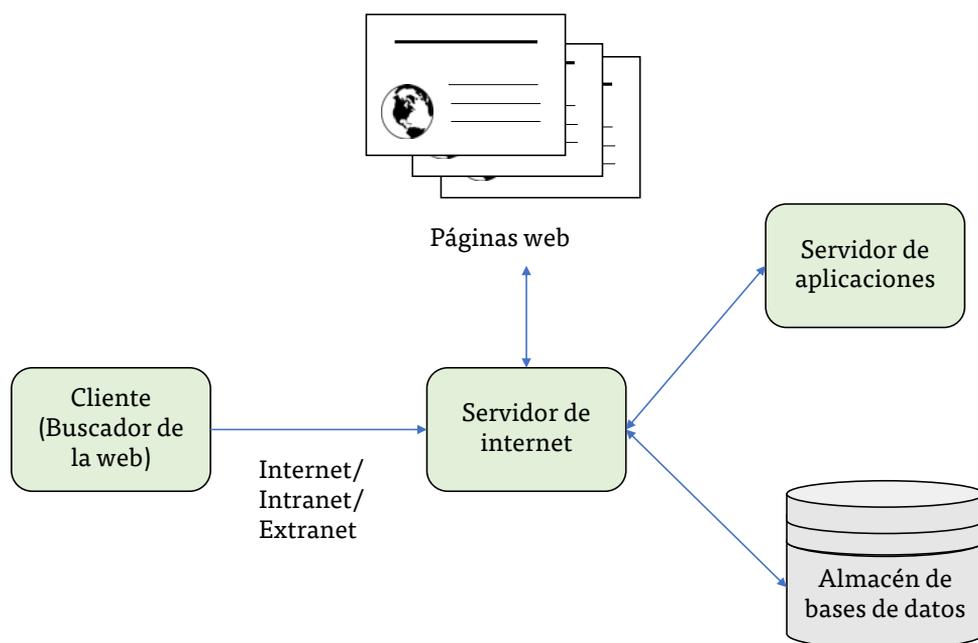


Figura 15. Arquitectura del almacenamiento de datos basado en web

Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

Se deben tener en cuenta varios problemas a la hora de decidir qué arquitectura utilizar. Entre ellos se encuentran los siguientes:

- **El tipo de sistema de gestión de bases de datos (DBMS) a utilizar.** La mayoría de los almacenes de datos se crean mediante sistemas de administración de bases de datos relacionales (RDBMS). Cada uno de estos productos admite arquitecturas basadas en cliente/servidor y web.
 - ✓ Oracle (Oracle Corporation, oracle.com),
 - ✓ SQL Server (Microsoft Corporation, microsoft.com/sql) y
 - ✓ DB2 (IBM Corporation, <http://www-01.ibm.com/software/data/db2>) son los más utilizados.
- **Tipo de procesamiento paralelo y/o el particionamiento**
 - ✓ El procesamiento paralelo permite que varias CPU procesen solicitudes de consulta de almacenamiento de datos simultáneamente y proporciona escalabilidad.
 - ✓ Los diseñadores de almacenamiento de datos deben decidir si las tablas de base de datos se particionarán (es decir, se dividirán en tablas más pequeñas) para la eficiencia del acceso y cuáles serán los criterios. Esta es una consideración importante que es necesaria por las grandes cantidades de datos contenidos en un almacén de datos típico.
- **Tipo de herramientas de migración de datos para cargar el almacenamiento de datos**

Mover datos de un sistema existente a un almacén de datos es una tarea tediosa y laboriosa. Dependiendo de la diversidad y la ubicación de los activos de datos, la migración puede ser un procedimiento relativamente simple o (por el contrario) un proyecto de meses de duración. Los resultados de una evaluación exhaustiva de los activos de datos existentes deben utilizarse para determinar si se deben utilizar

herramientas de migración y, en caso afirmativo, qué capacidades buscar en esas herramientas comerciales.

- **Las herramientas que se utilizarán para apoyar la recuperación y el análisis de datos**

A menudo es necesario utilizar herramientas especializadas para localizar, acceder, analizar, extraer, transformar y cargar periódicamente los datos necesarios en un almacén de datos. Se debe tomar una decisión sobre:

1. Desarrollar las herramientas de migración internamente,
2. Comprarlas a un proveedor externo, o
3. Utilizar las que se proporcionan con el sistema de almacenamiento de datos. Las migraciones en tiempo real excesivamente complejas justifican herramientas ETL especializadas de terceros.

Arquitecturas de almacenamiento de datos entrelazadas

El diseño de arquitectura de almacenamiento de datos se puede clasificar en diseño de almacenamiento de datos (EDW) de toda la empresa y diseño de data mart (DM). A continuación, se describirán algunas alternativas a los tipos de diseño arquitectónico básico que no son EDW puro ni DM puro, sino entre o más allá de las estructuras arquitectónicas tradicionales. Las nuevas y notables incluyen arquitecturas especializadas y federadas. Las cinco arquitecturas que se muestran son propuestas por Ariyachandra & Watson, (2005).

a. Arquitectura independiente de data marts

Se puede decir que esta es la alternativa de arquitectura más sencilla y menos costosa. Los data marts se desarrollan para operar independientemente

unos de otros para satisfacer las necesidades de las unidades organizativas individuales. Debido a su independencia, pueden tener definiciones de datos inconsistentes y diferentes dimensiones y medidas, lo que dificulta el análisis de datos en los data marts (es decir, es difícil, si no imposible, llegar a la “una versión de la verdad”), véase la figura 16.



Figura 16. Arquitectura independiente de data marts
Fuente: Elaborado con base en Ariyachandra & Watson, 2006.

b. Arquitectura de bus de data mart

Esta arquitectura es una alternativa viable a los data marts independientes donde los marts individuales están vinculados entre sí a través de algún tipo de middleware. Dado que los datos están vinculados entre los marts individuales, hay más posibilidades de mantener la coherencia de los datos en toda la empresa (al menos en el nivel de metadatos). Aunque permite consultas de datos complejas en data marts, el rendimiento de estos tipos de análisis puede no estar en un nivel satisfactorio, véase la figura 17.

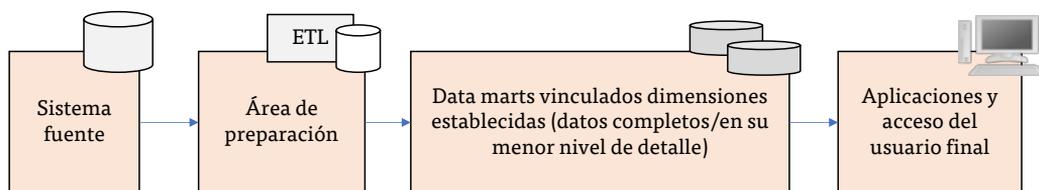


Figura 17. Arquitectura de bus data Mart con Data Marts dimensionales vinculados
Fuente: Elaborado con base en Ariyachandra & Watson, 2006.

c. Arquitectura Hub-and-spoke, fábrica de información corporativa

Esta es quizás la arquitectura de almacenamiento de datos más famosa hoy en día. Aquí la atención se centra en la construcción de una infraestructura escalable y mantenible (a menudo desarrollada de manera iterativa, área de tema por área temática) que incluye un almacén de datos centralizado y varios data marts dependientes (cada uno para una unidad organizativa). Esta arquitectura permite una personalización de interfaces de usuario e informes. En el lado negativo, esta arquitectura carece de la vista empresarial holística y puede dar lugar a redundancia de datos y latencia de datos, véase la figura 18.

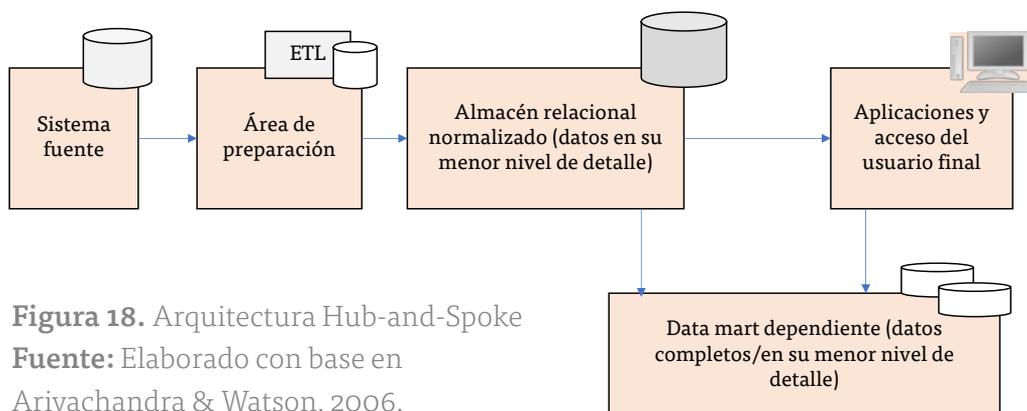


Figura 18. Arquitectura Hub-and-Spoke

Fuente: Elaborado con base en Ariyachandra & Watson, 2006.

d. Almacenamiento centralizado de datos

La arquitectura de almacenamiento de datos centralizada es similar a la arquitectura de concentrador y radio, excepto que no hay data marts dependientes; en su lugar, hay un gigantesco almacén de datos empresariales que atiende las necesidades de todas las unidades organizativas. Este enfoque centralizado proporciona a los usuarios acceso a todos los datos del almacenamiento de datos en lugar de limitarlos a data marts. Además, reduce la cantidad de datos que el equipo técnico tiene que

transferir o cambiar, lo que simplifica la gestión y administración de datos. Si se diseña e implementa correctamente, esta arquitectura proporciona una visión oportuna y holística de la empresa a quien sea, cuando y dondequiera que esté dentro de la organización. La arquitectura central de almacenamiento de datos, que es defendida principalmente por Tera-data Corp., aconseja el uso de almacenes de datos sin ningún data marts, consulte la figura 19.

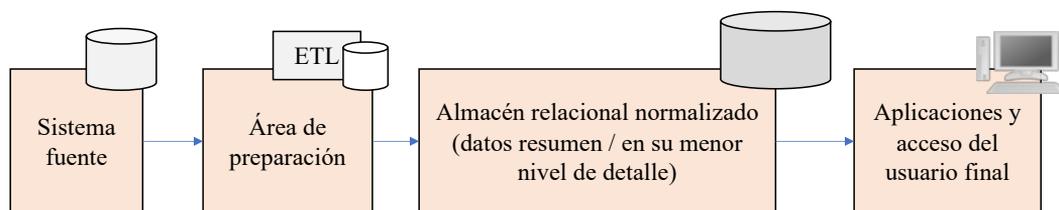


Figura 19. Arquitectura centralizada de almacenamiento de datos

Fuente: Elaborado con base en Ariyachandra & Watson, 2006.

e. Almacén de datos federado

El enfoque federado es una concesión a las fuerzas naturales que socavan los mejores planes para desarrollar un sistema perfecto. Utiliza todos los medios posibles para integrar recursos analíticos de múltiples fuentes para satisfacer las necesidades cambiantes o las condiciones de negocio. Esencialmente, el enfoque federado implica la integración de sistemas dispares. En una arquitectura federada, las estructuras de soporte de decisión existentes se dejan en su lugar y se accede a los datos de esos orígenes, según sea necesario. El enfoque federado es compatible con proveedores de middleware que plantean capacidades de combinación y consulta distribuida. Estas herramientas basadas en lenguaje con extensión XML ofrecen a los usuarios una vista global de los orígenes de datos distribuidos, incluidos almacenes de datos, data marts, sitios web, documentos y sistemas operativos, véase la figura 20.

Cuando los usuarios eligen objetos de consulta de esta vista y presionan el botón enviar, la herramienta consulta automáticamente los orígenes distribuidos, une los resultados y los presenta al usuario. Debido a problemas de rendimiento y calidad de datos, la mayoría de los expertos coinciden en que los enfoques federados funcionan bien para complementar los almacenes de datos, no para reemplazarlos.

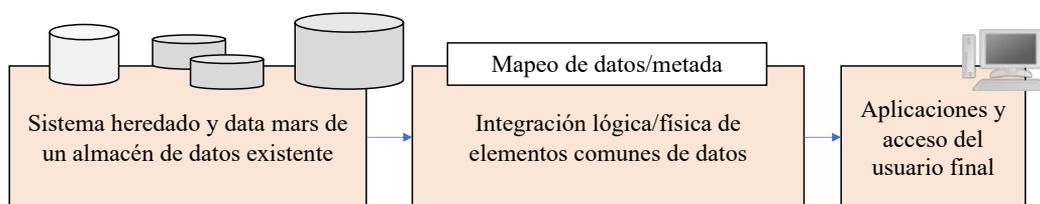


Figura 20. Arquitectura de datos federado

Fuente: Elaborado con base en Ariyachandra & Watson, 2006.

Ariyachandra y Watson identificaron 10 factores que pueden afectar a la decisión de selección de arquitectura:

1. Interdependencia de la información entre unidades organizativas
2. Las necesidades de información de la alta dirección
3. Urgencia de necesidad de un almacén de datos
4. Naturaleza de las tareas del usuario final
5. Restricciones en los recursos
6. Vista estratégica de data warehouse antes de la implementación
7. Compatibilidad con sistemas existentes
8. Capacidad percibida del personal de TI interno
9. Problemas técnicos
10. Factores sociales/políticos

Las cuestiones técnicas, más allá de proporcionar tecnología que sea factible, es importante para satisfacer las necesidades de información de la alta gerencia y la participación de los usuarios en el proceso de desarrollo (un factor social/político). Cada arquitectura de almacenamiento de datos

tiene aplicaciones específicas para las que es eficaz y, por lo tanto, proporciona beneficios máximos a la organización. Sin embargo, en general, la estructura de data mart parece ser la menos eficaz en la práctica. En la figura 21 se muestra un almacén de datos empresariales propuesto por Teradata Corporation, con sus diferentes tipos de usuarios.

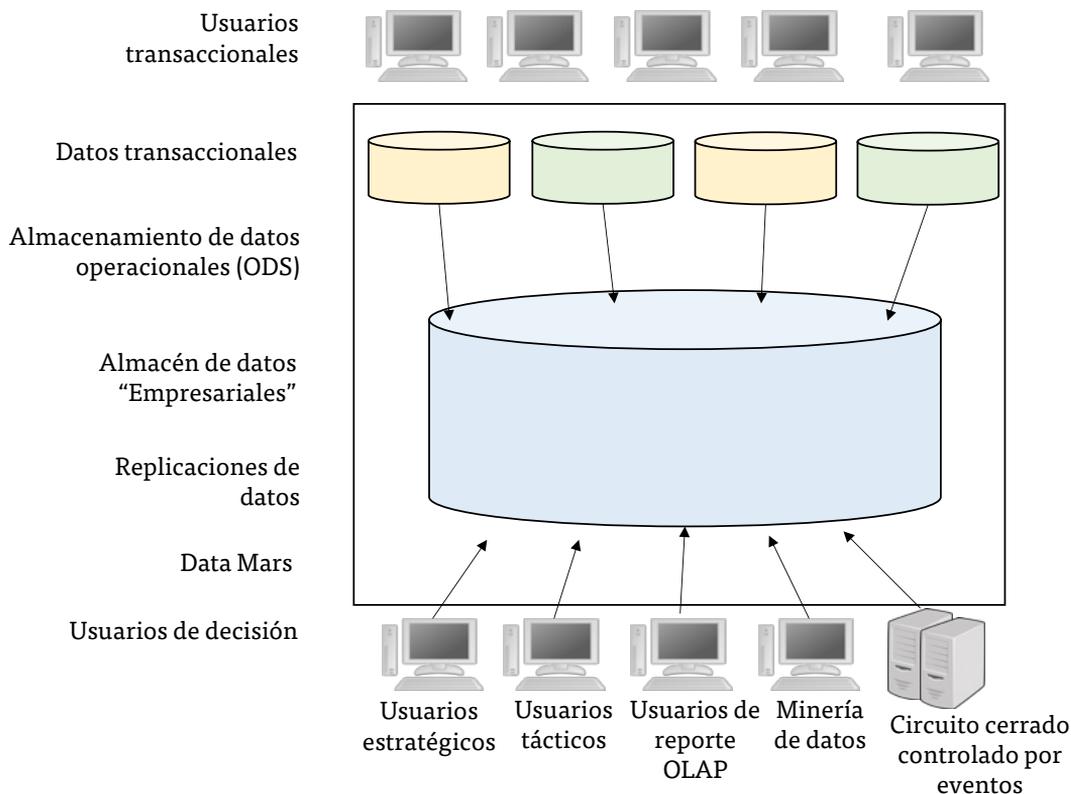


Figura 21. Almacén de datos empresariales de Teradata Corporation

Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

Desde que el almacenamiento de datos se convirtió en una parte crítica de las empresas modernas, la cuestión de qué arquitectura de almacenamiento de datos es la mejor ha sido un tema de discusión. Los dos gurús del campo del almacenamiento de datos, Bill Inmon y Ralph Kimball, lo han abordado. Inmon aboga por la arquitectura Hub-and-spoke (Fábrica de Información Corporativa), mientras que Kimball promueve la arquitectura de bus data mart con dimensiones conformadas. Otras arquitecturas

son posibles, pero estas dos opciones son enfoques fundamentalmente diferentes, y cada una tiene defensores fuertes. Para arrojar luz sobre esta polémica pregunta, Ariyachandra y Watson (2006b) llevaron a cabo un estudio empírico. Para recopilar los datos, usaron una encuesta basada en Web dirigida a personas involucradas en implementaciones de almacenamiento de datos. Su encuesta incluyó preguntas sobre el encuestado, la empresa del encuestado, el almacenamiento de datos de la empresa y el éxito de la arquitectura de almacenamiento de datos.

En total, 454 encuestados proporcionaron información utilizable. Las empresas encuestadas oscilaron entre pequeñas (menos de 10 millones de dólares en ingresos) hasta grandes (más de 10,000 millones de dólares). La mayoría de las empresas se encontraban en los Estados Unidos (60%) y representaba una variedad de industrias, la industria de servicios financieros proporcionó la mayoría de las respuestas (15%). La arquitectura predominante fue la arquitectura hub-and-spoke (39%), seguida de la arquitectura de bus (26%), la arquitectura centralizada (17%), los data marts independientes (12%) y la arquitectura de datos federado (4%). La plataforma más común para alojar los almacenes de datos fue Oracle (41%), seguido de Microsoft (19%) e IBM (18%). Los ingresos brutos promedio (medios) variaron de 3.700 millones de dólares para data marts independientes a 6.000 millones de dólares para la arquitectura federada.

Utilizaron cuatro medidas para evaluar el éxito de las arquitecturas: 1) calidad de la información, (2) calidad del sistema, (3) impactos individuales y (4) impactos organizativos. Las preguntas utilizaban una escala de siete puntos, con la puntuación más alta que indicaba una arquitectura más exitosa. La tabla 5 muestra las puntuaciones medias de las medidas en todas las arquitecturas.

Tabla 5. Puntuaciones medias de evaluación para el éxito de las arquitecturas**Fuente:** Elaborado con base en Sharda, Delen, & Turban, 2014.

| | Arquitectura independiente | Arquitectura de bus | Arquitectura de Hub and spoke | Centralizado de datos | Arquitectura federada |
|---------------------------|----------------------------|---------------------|-------------------------------|-----------------------|-----------------------|
| Calidad de la información | 4.42 | 5.16 | 5.35 | 5.23 | 4.73 |
| Calidad del sistema | 4.59 | 5.60 | 5.56 | 5.41 | 4.69 |
| Impactos individuales | 5.08 | 5.80 | 5.62 | 5.64 | 5.15 |
| Impactos organizacionales | 4.66 | 5.34 | 5.24 | 5.30 | 4.77 |

Como indican los resultados del estudio, los data marts independientes obtuvieron la puntuación más baja en todas las medidas. Este hallazgo confirma que los data marts independientes son una solución arquitectónica deficiente. El siguiente más bajo en todas las medidas fue la arquitectura de datos federado. Las empresas a veces tienen plataformas de apoyo a la toma de decisiones dispares resultantes de fusiones y adquisiciones, y pueden elegir un enfoque federado, al menos a corto plazo. Los hallazgos sugieren que la arquitectura federada no es una solución óptima a largo plazo. Lo que es interesante, sin embargo, es la similitud de los promedios para el bus, hub-and-spoke, y arquitecturas centralizadas. Las diferencias son lo suficientemente pequeñas como para que no se pueda afirmar la superioridad de una arquitectura en particular sobre las demás, al menos sobre la base de una simple comparación de estas medidas de éxito.

También recopilamos datos sobre el dominio (por ejemplo, que varían de una subunidad a toda la empresa) y el tamaño (es decir, la cantidad de datos almacenados) de los almacenes. Encontraron que la arquitectura de hub-and spoke se utiliza normalmente con implementaciones más grandes en toda la empresa y almacenes más grandes. También investigaron el costo y el tiempo necesarios para implementar las diferentes arquitecturas. En general, la arquitectura de hub-and spoke era la más costosa y lenta de implementar.

3.4 Integración, extracción, transformación y carga de datos (EII, ETL)

Integración de datos

La integración de datos comprende tres procesos principales que, cuando se implementan correctamente, permiten acceder a los datos y ponerlos a disposición de una serie de ETL y herramientas de análisis y del entorno de almacenamiento de datos: acceso a datos (es decir, la capacidad de acceder y extraer datos de cualquier origen de datos), la arquitectura de datos (es decir, la integración de vistas empresariales en varios almacenes de datos) y la captura de cambios (basada en la identificación, captura y entrega de los cambios realizados en orígenes de datos empresariales). Algunos proveedores, como SAS Institute, Inc., han desarrollado sólidas herramientas de integración de datos. El servidor de integración de datos empresariales SAS incluye herramientas de integración de datos de clientes que mejoran la calidad de los datos en el proceso de integración.

Un propósito principal de un almacén de datos es integrar datos de varios sistemas. Varias tecnologías de integración permiten la integración de datos y metadatos:

- Integración de aplicaciones empresariales (EAI)
- Arquitectura orientada a servicios (SOA)
- Integración de información empresarial (EII)

La integración de aplicaciones empresariales (EAI), de extracción, transformación y carga (ETL) proporciona un vehículo para insertar datos de los sistemas de origen en el almacenamiento de datos. Implica la integración de la funcionalidad de la aplicación y se centra en compartir la funcionalidad (en lugar de los datos) entre sistemas, lo que permite flexibilidad y reutilización. Tradicionalmente, las soluciones EAI se han centrado en permitir la reutilización de aplicaciones en el nivel de interfaz de programación de aplicaciones (API). Recientemente, EAI se logra mediante el uso

de servicios de grano grueso SOA (una colección de procesos de negocio o funciones) que están bien definidos y documentados. El uso de servicios web es una forma especializada de implementar un SOA. EAI se puede utilizar para facilitar la adquisición de datos directamente en un almacén de datos casi en tiempo real o para tomar decisiones a los sistemas OLTP. Existen muchos enfoques y herramientas diferentes para la implementación de EAI.

Enterprise-Information-Integration (EII) es un espacio de herramientas en evolución que promete integración de datos en tiempo real de una variedad de fuentes, como bases de datos relacionales, servicios web y bases de datos multidimensionales. Es un mecanismo para extraer datos de los sistemas de origen para satisfacer una solicitud de información. Las herramientas EII utilizan metadatos predefinidos para rellenar vistas que hacen que los datos integrados parezcan relacionales para los usuarios finales. XML puede ser el aspecto más importante de EII porque XML permite etiquetar los datos en el momento de la creación o posterior. Estas etiquetas se pueden ampliar y modificar para adaptarse a casi cualquier área de conocimiento. La integración de datos físicos ha sido convencionalmente el principal mecanismo para crear una vista integrada con almacenes de datos y data marts. Con la llegada de las herramientas EII, los nuevos patrones de integración de datos virtuales son factibles.

Extracción, transformación y carga (ETL)

En el corazón de la parte técnica del proceso de almacenamiento de datos está la extracción, transformación y carga (ETL). Las tecnologías ETL, que han existido durante algún tiempo, son fundamentales en el proceso y el uso de los almacenes de datos. El proceso ETL es un componente integral en cualquier proyecto centrado en datos. Los administradores de TI a menudo se enfrentan a desafíos porque el proceso ETL normalmente consume el 70 por ciento del tiempo en un proyecto centrado en datos.

El proceso ETL consiste en la extracción (es decir, la lectura de datos de una o más bases de datos), la transformación (es decir, la conversión de los datos extraídos de su formulario anterior en la forma en la que debe ser para que puedan ser colocados en un almacén de datos o simplemente en otro base de datos) y la carga (es decir, colocar los datos en el almacenamiento de datos). La transformación se produce mediante reglas o tablas de búsqueda o combinando los datos con otros datos. Las tres funciones de base de datos se integran en una herramienta para extraer datos de una o más bases de datos y colocarlos en otra base de datos consolidada o en un almacén de datos.

Las herramientas ETL también transportan datos entre orígenes y destinos, documentan cómo cambian los elementos de datos (por ejemplo, metadatos) a medida que se mueven entre el origen y el destino, intercambian metadatos con otras aplicaciones según sea necesario y administran todos los procesos y operaciones en tiempo de ejecución (por ejemplo, programación, gestión de errores, registros de auditoría, estadísticas). ETL es extremadamente importante para la integración de datos, así como para el almacenamiento de datos. El propósito del proceso ETL es cargar el almacén con datos integrados y limpios. Los datos utilizados en los procesos ETL pueden provenir de cualquier fuente: una aplicación de mainframe, una aplicación ERP, una herramienta CRM, un archivo sin formato, una hoja de cálculo de Excel o incluso una cola de mensajes. En la figura 22, se describe el proceso ETL.

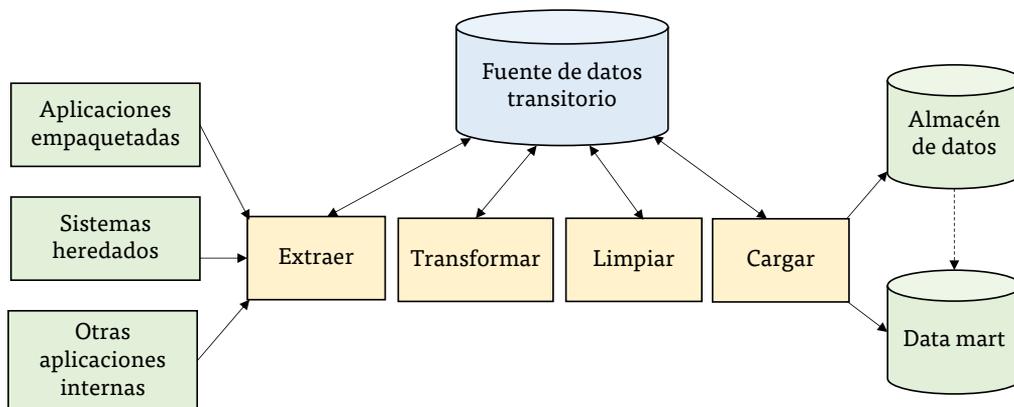


Figura 22. Proceso ETL

Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

El proceso de migración de datos a un almacén de datos implica la extracción de datos de todos los orígenes relevantes. Los orígenes de datos pueden consistir en archivos extraídos de bases de datos OLTP, hojas de cálculo, bases de datos personales (por ejemplo, Microsoft Access) o archivos externos. Normalmente, todos los archivos de entrada se escriben en un conjunto de tablas de ensayo, que están diseñadas para facilitar el proceso de carga. Un almacén de datos contiene numerosas reglas de negocio que definen aspectos de cómo se usarán los datos, las reglas de integración, la estandarización de atributos codificados y las reglas de cálculo. Cualquier problema de calidad de datos relacionado con los archivos de origen debe corregirse antes de cargar los datos en el almacenamiento de datos. Una de las ventajas de un almacén de datos bien diseñado es que estas reglas se pueden almacenar en un repositorio de metadatos y se aplica al almacenamiento de datos de forma centralizada. Esto difiere de un enfoque OLTP, que normalmente tiene datos y reglas de negocio dispersas por todo el sistema.

El proceso de carga de datos en un almacén de datos se puede realizar a través de herramientas de transformación de datos que proporcionan una interfaz gráfica de usuario para ayudar en el desarrollo y mantenimiento de reglas de negocio o a través de métodos más tradicionales, como el desarrollo de programas o utilidades para cargar el almacenamiento de datos, utilizando lenguajes de programación como PL/SQL, C++, Java o lenguajes de .NET Framework. Esta decisión no es fácil para las organizaciones. Varios problemas afectan si una organización comprará herramientas de transformación de datos o creará el propio proceso de transformación:

- Las herramientas de transformación de datos son caras.
- Las herramientas de transformación de datos tienen una curva de aprendizaje larga.
- Es difícil medir cómo hacer que la organización aprenda a utilizar las herramientas de transformación de datos.

A largo plazo, un enfoque de herramienta de transformación debe simplificar el mantenimiento del almacenamiento de datos de una organización. Las herramientas de transformación también pueden ser eficaces para detectar y eliminar (es decir, eliminar cualquier anomalía en los datos). OLAP y las herramientas de minería de datos se basan en la transformación de los datos.

Como ejemplo de ETL eficaz, Motorola, Inc. utiliza ETL para alimentar sus almacenes de datos. Motorola recopila información de 30 sistemas de compras diferentes y los envía a su almacén global de datos SCM para el análisis del gasto total de la empresa.

Solomon clasificó las tecnologías ETL en cuatro categorías: sofisticadas, facilitadoras, sencillas y rudimentarias. Por lo general, se reconoce que las herramientas de la categoría sofisticada darán como resultado que el proceso ETL se documente mejor y se administre con mayor precisión a medida que evolucione el proyecto de almacenamiento de datos.

A pesar de que es posible que los programadores desarrollen software para ETL, es más sencillo utilizar una herramienta ETL existente. Los siguientes son algunos de los criterios importantes para seleccionar una herramienta ETL:

- Habilidad de leer la información que tiene una arquitectura ilimitada de recursos.
- Captura automática y entrega de metadatos.
- Una historia de estándares abiertos de conformidad.
- Una interfaz fácil de usar para el desarrollador y el usuario funcional.

La realización de una amplia ETL puede ser un signo de datos mal administrados y una falta fundamental de una estrategia coherente de gestión de datos. Cuando los datos se administran correctamente como un activo empresarial, los esfuerzos de ETL se reducen significativamente y los datos redundantes se eliminan por completo. Esto conduce a enormes

ahorros en mantenimiento y una mayor eficiencia en el nuevo desarrollo, al mismo tiempo que mejora la calidad de los datos. Los procesos ETL mal diseñados son costosos de mantener, cambiar y actualizar. En consecuencia, es crucial tomar las decisiones adecuadas en términos de la tecnología y las herramientas que se utilizarán para desarrollar y mantener el proceso ETL.

Existen varias herramientas ETL empaquetadas disponibles. Actualmente, los proveedores de bases de datos ofrecen capacidades de ETL que mejoran y compiten con herramientas ETL independientes. SAS reconoce la importancia de la calidad de los datos y ofrece la primera solución totalmente integrada de la industria que combina ETL y calidad de datos para transformar los datos en activos estratégicos y valiosos.

3.5 Desarrollo y problemas en la implementación del almacenamiento de datos

Muchas organizaciones necesitan crear los almacenes de datos utilizados para el apoyo a la toma de decisiones. Se emplean dos enfoques que compiten. El primer enfoque es el de Bill Inmon, a quien a menudo se le llama “el padre del almacenamiento de datos”. Inmon admite un enfoque de desarrollo descendente que adapta las herramientas de base de datos relacionales tradicionales a las necesidades de desarrollo de un almacén de datos de toda la empresa, también conocido como el enfoque EDW. El segundo enfoque es el de Ralph Kimball, quien propone un enfoque de abajo hacia arriba que emplea el modelado dimensional, también conocido como el enfoque data mart.

Tabla 6. Contrastes entre los enfoques de desarrollo de Data Mart y EDW

Fuente: Elaborado con base en Ariyachandra & Watson, ¿Qué arquitectura de almacenamiento de datos es más exitosa?, 2006.

| Esfuerzo | Enfoque Data Mart | Enfoque EDW |
|--|---|--|
| Alcance | Un tema área | Varias áreas temáticas |
| Tiempo de desarrollo | Meses | Años |
| Costo de desarrollo | \$10,000 a \$100,000+ | \$1,000,000+ |
| Desarrollo "dificultad" | Bajo a medio | Alto |
| Requisito previo de datos para compartir | Común (dentro del área de negocio) | Común (en toda la empresa) |
| Fuentes | Sólo algunos sistemas operativos y externos | Muchos sistemas operativos y externos |
| Tamaño | Megabytes a varios gigabytes | Gigabytes a petabytes |
| Horizonte temporal | Datos históricos y casi actuales | Datos históricos |
| Transformaciones de datos | Bajo a medio | Alto |
| Frecuencia de actualización | Horario, diario, semanal | Semanal, mensual |
| Tecnología Hardware | Estaciones de trabajo y servidores departamentales | Servidores empresariales y ordenadores mainframe |
| Sistema operativo | Windows y Linux | Unix, Z/OS, OS/390 |
| Bases | Servidores de bases de datos estándar o de grupo de trabajo | Servidores de bases de datos empresariales |
| Número de uso de usuarios simultáneos | 10s | 100s a 1000s |
| Tipos de usuario | Analistas y gerentes del área de negocio | Analistas empresariales y altos ejecutivos |
| Foco de negocios | Optimización de actividades dentro del área de negocio | Optimización multifuncional y toma de decisiones |

Saber cómo estos dos modelos son iguales y cómo difieren nos ayuda a comprender los conceptos básicos de almacenamiento de datos. En la tabla 6 se comparan los dos enfoques.

El enfoque de Inmon hace hincapié en el desarrollo descendente, empleando metodologías y herramientas de desarrollo de bases de datos establecidas, como diagramas de relación de entidades (ERD), y un ajuste del enfoque de desarrollo en espiral. El enfoque EDW no excluye la creación de data marts. El EDW es el ideal en este enfoque porque proporciona una visión coherente y completa de la empresa. Murtaza (1998) presentó un marco para el desarrollo de EDW.

La estrategia de data mart de Kimball es un enfoque de “planificar grandes y construir pequeños”. Un data mart es un almacén de datos orientado al tema o al departamento. Es una versión reducida de un almacén de datos que se centra en las solicitudes de un departamento específico, como marketing o ventas. Este modelo aplica el modelado de datos dimensionales, que comienza con tablas. Kimball abogó por una metodología de desarrollo que implica un enfoque de abajo hacia arriba, que en el caso de los almacenes de datos significa construir un data mart a la vez.

No existe una estrategia única para el almacenamiento de datos. La estrategia de almacenamiento de datos de una empresa puede evolucionar de un simple data mart a un almacén de datos complejo en respuesta a las demandas de los usuarios, los requisitos empresariales de la empresa y la madurez de la empresa en la administración de sus recursos de datos. Para muchas empresas, un data mart es con frecuencia un primer paso conveniente para adquirir experiencia en la construcción y administración de un almacenamiento de datos, al tiempo que presenta a los usuarios empresariales los beneficios de un mejor acceso a sus datos; además, un data mart indica comúnmente el valor comercial del almacenamiento de datos. En última instancia, la ingeniería de un EDW que consolida los data marts antiguos y los almacenes de datos es la solución ideal. Sin embargo, el desarrollo de data marts individuales a menudo puede proporcionar

muchos beneficios en el camino hacia el desarrollo de un EDW, especialmente si la organización no puede o no está dispuesta a invertir en un proyecto a gran escala. Los data marts también pueden demostrar viabilidad y éxito al proporcionar beneficios. Esto podría conducir potencialmente a una inversión en una EDW. En la tabla 7 se resumen las diferencias características más esenciales entre los dos modelos.

El acceso acelerado a los datos permite, por ejemplo, que todas las propiedades de Starwood obtengan los mismos datos actualizados necesarios para sus informes a nivel mundial. Anteriormente, los gerentes de hoteles en algunas áreas no podían hacer análisis en el mismo día o al día siguiente. Había algunas ubicaciones que tenían datos nuevos y otras que tenían datos más antiguos. Los gerentes de hoteles, en todo el mundo, ahora tienen datos actualizados para sus hoteles, aumentando la eficiencia y la rentabilidad, mejorando el servicio al cliente asegurándose de que las habitaciones estén disponibles para los clientes de primer nivel y mejorando la capacidad de la empresa para gestionar la ocupación de las habitaciones y sus tarifas.

Tabla 7. Diferencias esenciales entre los enfoques de Inmon y Kimball

Fuente: Elaborado con base en Ariyachandra & Watson, ¿Qué arquitectura de almacenamiento de datos es más exitosa?, 2006.

| Característica | Inmon | Kimball |
|-----------------------------------|---|---|
| Metodología y Arquitectura | | |
| Enfoque general | De arriba hacia abajo | De abajo hacia arriba |
| Estructura arquitectónica | Datos de toda la empresa (atómicas) almacén “alimenta” bases de datos departamentales | Los data marts modelan un único proceso empresarial, y la coherencia empresarial se logra a través de un bus de datos y dimensiones conformadas |
| Complejidad del método | Bastante complejo | Bastante simple |

| Característica | Inmon | Kimball |
|--|--|---|
| Comparación con metodologías de desarrollo establecidas | Derivado de la metodología espiral | Proceso de cuatro pasos; una desviación de los métodos del sistema de gestión de bases de datos relacionales (RDBMS) |
| Discusión del diseño físico | Bastante exhaustivo | Bastante ligero |
| Modelado de datos | | |
| Orientación de datos | Sujeto o impulsado por datos | Orientado a procesos |
| Herramientas | Tradicional (diagramas de relación de entidad [ERD], diagramas de flujo de datos [DFD]) | Modelado dimensional; una desviación del modelado relacional |
| Accesibilidad del usuario final | Bajo | Alto |
| Filosofía | | |
| Audiencia primaria | Profesionales de TI | Usuarios finales |
| Colocar en la organización | Parte integral de la fábrica de información corporativa | Transformador y retenedor de datos operativos |
| Objetivo | Ofrezca una solución técnica sólida basada en métodos y tecnologías de bases de datos probados | Ofrezca una solución que facilite a los usuarios finales consultar directamente los datos y obtener tiempos de respuesta razonables |

Las herramientas de generación de informes adicionales, como las utilizadas para CRM, ventas y catering, también se beneficiaron de la mejora del procesamiento. Otros informes críticos también se han beneficiado. La gestión de campañas de marketing también es más eficiente ahora que los gerentes pueden analizar los resultados en días o semanas en lugar de meses.

Algunas organizaciones quieren externalizar completamente sus esfuerzos de almacenamiento de datos. Simplemente no quieren lidiar con las adquisiciones de software y hardware, y no quieren administrar sus sistemas de información. Una alternativa es usar almacenes de datos hospedados. En este escenario, otra empresa, idealmente, una que tiene mucha experiencia y experiencia, desarrolla y mantiene el almacenamiento de datos. Sin embargo, hay problemas de seguridad y privacidad con este enfoque.

Representación de datos en el almacenamiento de datos

Muchas variaciones de la arquitectura de almacenamiento de datos son posibles. No importa cuál fuera la arquitectura, el diseño de la representación de datos en el almacenamiento de datos siempre se ha basado en el concepto de modelado dimensional, el cual es un sistema basado en recuperación que admite el acceso a consultas de gran volumen. La representación y el almacenamiento de datos en un almacén de datos deben diseñarse de tal manera que no solo se acomoden, sino que también aumente el procesamiento de consultas multidimensionales complejas. A menudo, el esquema de estrella y el esquema de copos de nieve son el medio por el que se implementa el modelado dimensional en los almacenes de datos.

El esquema de estrella (a veces denominado esquema de combinación de estrellas) es el estilo más utilizado y el más simple de modelado dimensional. Un esquema de estrella contiene una tabla de hechos central rodeada y conectada a varias tablas de dimensiones. La tabla de hechos contiene un gran número de filas que corresponden a hechos observados y vínculos externos (es decir, claves externas). Una tabla de hechos contiene los atributos descriptivos necesarios para realizar análisis de decisiones e informes de consultas, y las claves externas se utilizan para vincular a tablas de dimensiones. Los atributos de análisis de decisiones consisten en medidas de rendimiento, métricas operativas, medidas agregadas (por ejemplo, volúmenes de ventas, tasas de retención de clientes, márgenes de beneficio, costos de producción, tasas de desperdicios, etc.) y todas las

demás métricas necesarias para analizar el rendimiento de la organización. En otras palabras, la tabla de hechos aborda principalmente lo que el almacenamiento de datos admite para el análisis de decisiones.

Rodeando las tablas de hechos centrales (y vinculadas a través de claves externas) hay tablas de dimensiones. Estas tablas de dimensiones contienen información de clasificación y agregación sobre las filas de hechos centrales. Además, contienen atributos que describen los datos contenidos en la tabla de hechos; abordan cómo se analizarán y resumirán los datos. Las tablas de dimensiones tienen una relación de uno a varios, con las filas de la tabla de hechos central. En las consultas, las dimensiones se utilizan para segmentar y cortar los valores numéricos de la tabla de hechos para satisfacer los requisitos de una necesidad de información ad hoc. El esquema de estrella está diseñado para proporcionar un tiempo de respuesta de consulta rápido, simplicidad y facilidad de mantenimiento para estructuras de base de datos de solo lectura. En la figura 23 se muestra un esquema de estrella simple. El esquema de estrella se considera un caso especial del esquema de copo de nieve.

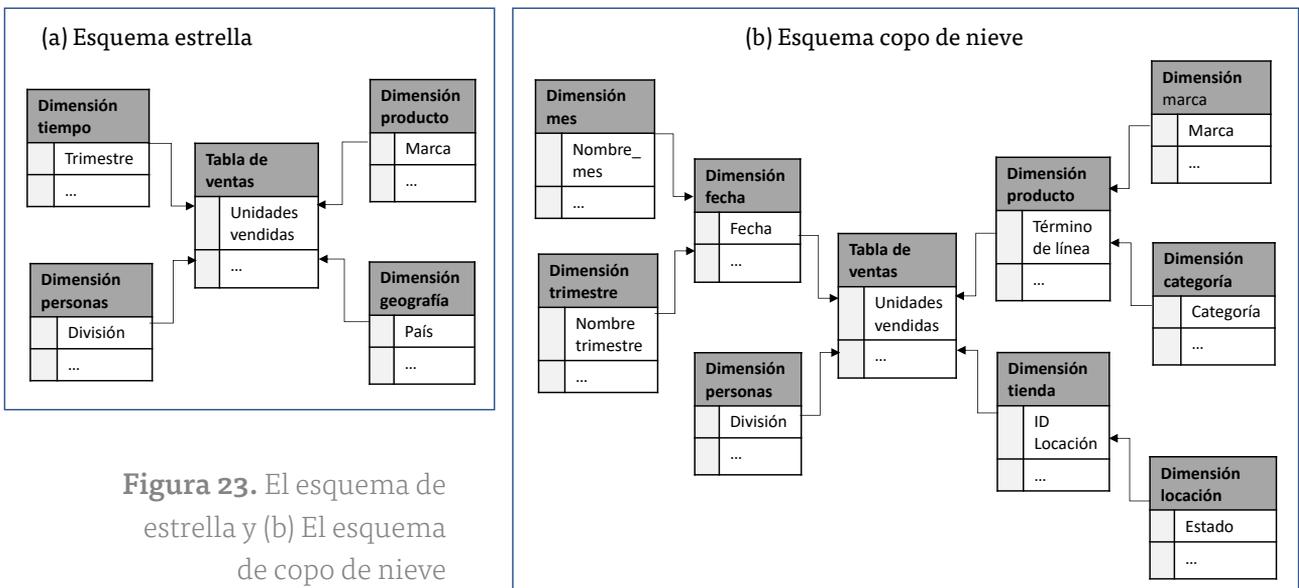


Figura 23. El esquema de estrella y (b) El esquema de copo de nieve

Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

El esquema de copo de nieve es una disposición lógica de tablas en una base de datos multidimensional, de tal manera que el diagrama de relación de entidad se asemeja a un copo de nieve por su forma. Muy relacionado con el esquema de estrella, el esquema de copo de nieve se representa mediante tablas de hechos centralizadas (normalmente solo una), que están conectadas a varias dimensiones. En el esquema de copo de nieve, sin embargo, las dimensiones se normalizan en varias tablas relacionadas, mientras que las dimensiones del esquema de estrella se desnormalizan con cada dimensión representada por una sola tabla.

Análisis de datos en el almacenamiento de datos

Una vez que los datos se almacenan correctamente en un almacén de datos, se pueden utilizar de varias maneras para apoyar la toma de decisiones de la organización. OLAP (procesamiento analítico en línea) es posiblemente la técnica de análisis de datos más utilizada en los almacenes de datos, y ha ido creciendo en popularidad debido al aumento exponencial de los volúmenes de datos y el reconocimiento del valor empresarial de los datos. Simplemente, OLAP es un enfoque para responder rápidamente a preguntas ad hoc mediante la ejecución de consultas analíticas multidimensionales en repositorios de datos de la organización (es decir, almacenes de datos, data marts).

OLAP contra OLTP

OLTP (sistema de procesamiento de transacciones en línea) es un término utilizado para un sistema de transacciones que es el principal responsable de capturar y almacenar datos relacionados con funciones comerciales cotidianas como ERP, CRM, SCM, POS, etc. Un sistema OLTP aborda una necesidad empresarial crítica, automatizando las transacciones comerciales diarias y ejecutando informes en tiempo real y análisis rutinarios. Pero estos sistemas no están diseñados para análisis ad hoc y consultas

complejas que tratan con una serie de elementos de datos. OLAP, por otro lado, está diseñado para abordar esta necesidad proporcionando un análisis ad hoc de los datos de la organización de manera mucho más eficaz y eficiente. OLAP y OLTP dependen en gran medida unos de otros: OLAP utiliza las capturas de datos de OLTP y OLTP automatiza los procesos empresariales que se administran mediante decisiones admitidas por OLAP. La tabla 8 proporciona una comparación multicriterio entre OLTP y OLAP.

La estructura operativa principal en OLAP se basa en un concepto llamado cubo. Un cubo en OLAP es una estructura de datos multidimensional (real o virtual) que permite un análisis rápido de los datos. También se puede definir como la capacidad de manipular y analizar datos de manera eficiente desde múltiples perspectivas. La disposición de los datos en cubos tiene como objetivo superar una limitación de las bases de datos relacionales: Las bases de datos relacionales no son adecuadas para el análisis casi instantáneo de grandes cantidades de datos. En su lugar, son más adecuados para manipular registros (agregar, eliminar y actualizar datos) que representan una serie de transacciones. Aunque existen muchas herramientas de escritura de informes para bases de datos relacionales, estas herramientas son lentas cuando es necesario ejecutar una consulta multidimensional que abarca muchas tablas de base de datos.

Tabla 8. Comparación entre OLTP y OLAP

| Crterios | OLTP | OLAP |
|------------------------|---|--|
| Propósito | Para llevar a cabo las funciones empresariales diarias | Apoyar la toma de decisiones y proporcionar respuestas a las consultas empresariales y de gestión |
| Fuente de datos | Base de datos de transacciones (un repositorio de datos normalizado centrado principalmente en la eficiencia y la coherencia) | Almacén de datos o data mart (un repositorio de datos no normalizado centrado principalmente en la precisión e integridad) |

| Crterios | OLTP | OLAP |
|-------------------------------|---|---|
| Informes | Informes rutinarios, periódicos y estrechamente enfocados | Informes y consultas ad hoc, multidimensionales y ampliamente enfocados |
| Requisitos de recursos | Bases de datos relacionales ordinarias | Bases de datos especializadas, multiprocesador, de gran capacidad |
| Velocidad de ejecución | Rápido (registro de transacciones comerciales e informes de rutina) | Lento (consultas de uso intensivo de recursos, complejas y a gran escala) |

Con OLAP, un analista puede navegar por la base de datos y la pantalla de un subconjunto determinado de los datos (y su progresión a lo largo del tiempo) cambiando las orientaciones de los datos y definiendo cálculos analíticos. Estos tipos de navegación iniciada por el usuario de datos a través de la especificación de sectores (a través de rotaciones) y desglose/arriba (a través de la agregación y la desagregación) a veces se denomina “cortar y dados”. Las operaciones OLAP de uso común incluyen cortes y dados, taladrar, enrollar y pivotar.

- **Rebanada.** Un sector es un subconjunto de una matriz multidimensional (normalmente una representación bidimensional) correspondiente a un único conjunto de valores para una (o más) de las dimensiones que no están en el subconjunto. En la figura 24 se muestra una simple operación de corte en un cubo tridimensional.
- **Dados.** La operación de dados es un sector en más de dos dimensiones de un cubo de datos.
- **Profundice hacia Abajo/Arriba.** La perforación hacia abajo o hacia arriba es una técnica OLAP específica mediante la que el usuario navega entre niveles de datos que van desde el más resumido (arriba) hasta el más detallado (abajo).
- **Un paquete.** Un resumen implica calcular todas las relaciones de datos para una o más dimensiones. Para ello, se puede definir una relación computacional o una fórmula.

- **Pivote.** Esto se utiliza para cambiar la orientación dimensional de un informe o una visualización de página de consulta *ad hoc*.

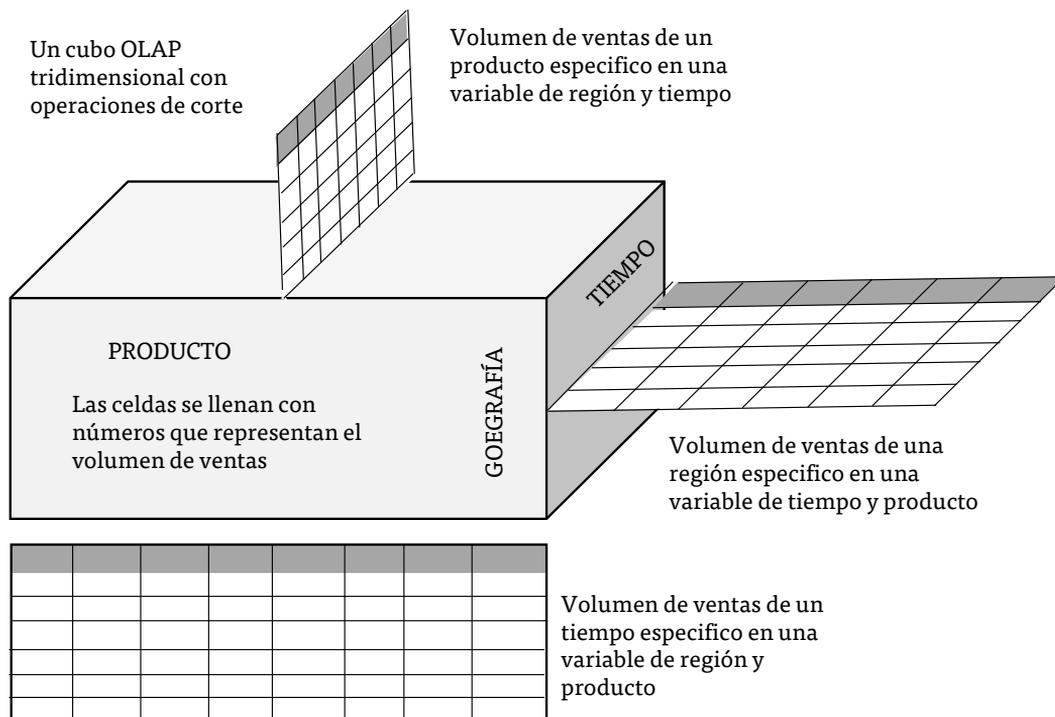


Figura 24. Operaciones de corte en un cubo de datos tridimensional simple
Fuente: Elaborado con base en Sharda, Delen, & Turban, 2014.

Existen algunas variaciones de OLAP, entre ellas, ROLAP, MOLAP y HOLAP son las más comunes.

ROLAP significa procesamiento analítico relacional en línea. ROLAP es una alternativa a la tecnología MOLAP (Multidimensional OLAP). Aunque las herramientas analíticas ROLAP y MOLAP están diseñadas para permitir el análisis de datos mediante el uso de un modelo de datos multidimensional, ROLAP difiere significativamente en que no requiere el precálculo y almacenamiento de información. En su lugar, las herramientas ROLAP acceden a los datos de una base de datos relacional y generan consultas SQL para calcular la información en el nivel adecuado cuando un usuario final lo solicita. Con ROLAP, es posible crear tablas de base de datos adicionales

(tablas de resumen o agregaciones) que resumen los datos en cualquier combinación deseada de dimensiones. Mientras que ROLAP utiliza un origen de base de datos relacional, generalmente, la base de datos debe diseñarse cuidadosamente para el uso de ROLAP. Una base de datos diseñada para OLTP no funcionará bien como una base de datos ROLAP. Por lo tanto, ROLAP todavía implica la creación de una copia adicional de los datos.

MOLAP es una alternativa a la tecnología ROLAP. MOLAP difiere significativamente de ROLAP en que requiere el precálculo y el almacenamiento de información en el cubo, la operación conocida como preprocesamiento. MOLAP almacena estos datos en un almacenamiento de matriz multidimensional optimizado, en lugar de en una base de datos relacional (que suele ser el caso de ROLAP).

El equilibrio indeseable entre ROLAP y MOLAP con respecto al costo adicional de ETL (extracción, transformación y carga) y el lento rendimiento de las consultas ha dado lugar a consultas para obtener mejores enfoques en los que se optimizan los pros y los contras de estos dos enfoques. Estas consultas dieron como resultado HOLAP (Hybrid Online Analytical Processing), que es una combinación de ROLAP y MOLAP. HOLAP permite almacenar parte de los datos en un almacén MOLAP y otra parte de los datos en un almacén ROLAP. El grado de control que tiene el diseñador de cubos sobre esta partición varía de un producto a otra.

3.6 Problemas de almacenamiento de datos-implementación

La implementación de un almacenamiento de datos es generalmente un esfuerzo masivo que debe planificarse y ejecutarse de acuerdo con los métodos establecidos. Sin embargo, el ciclo de vida del proyecto tiene muchas facetas y ninguna persona puede ser un experto en cada área. Aquí discutimos ideas y problemas específicos en relación con el almacenamiento de datos.

La gente quiere saber qué tan exitosas son sus iniciativas de BI y almacenamiento de datos en comparación con las de otras empresas. Ariyachandra y Watson propusieron algunos puntos de referencia para el éxito de BI y data warehousing. Investigaron fallas en el almacenamiento de datos y sus resultados mostraron que la gente define un “fracaso” de diferentes maneras. El Data Warehousing Institute (tdwi.org) ha desarrollado un modelo de madurez de almacenamiento de datos que una empresa puede aplicar para comparar su evolución. El modelo ofrece un medio rápido para medir dónde está ahora la iniciativa de almacenamiento de datos de la organización y dónde debe ir a continuación. El modelo de madurez consta de seis etapas: prenatal, infantil, niño, adolescente, adulto y salvia. El valor empresarial aumenta a medida que el almacenamiento de datos progresa a través de cada etapa siguiente. Las etapas se identifican por una serie de características, incluyendo alcance, estructura analítica, percepciones ejecutivas, tipos de análisis, administración, financiación, plataforma tecnológica, gestión del cambio y administración.

Los proyectos de almacenamiento de datos tienen muchos riesgos. La mayoría de ellos también se encuentran en otros proyectos de TI, pero los riesgos de almacenamiento de datos son más graves porque los almacenes de datos son proyectos caros, exigentes en tiempo y recursos. Cada riesgo debe evaluarse al inicio del proyecto. Al desarrollar un almacenamiento de datos exitoso, es importante considerar cuidadosamente varios riesgos y evitar los siguientes problemas:

- **Empezando por la cadena de patrocinio equivocada.** Necesita un patrocinador ejecutivo que tenga influencia sobre los recursos necesarios para apoyar e invertir en el almacenamiento de datos. También necesita un conductor de proyecto ejecutivo, alguien que se ha ganado el respeto de otros ejecutivos, tiene un sano escepticismo sobre la tecnología y es decisivo pero flexible. También necesita un administrador de IS/IT para dirigir el proyecto.

- **Establecer expectativas que no puede cumplir.** No quieres frustrar a los ejecutivos en el momento de la verdad. Cada proyecto de almacenamiento de datos tiene dos fases: la fase 1 es la de venta, en la que se comercializa internamente el proyecto mediante la venta de los beneficios a aquellos que tienen acceso a los recursos necesarios. La fase 2 es la lucha para cumplir con las expectativas descritas en la fase 1.
- **Participar en comportamientos políticamente ingenuos.** No indique simplemente que un almacenamiento de datos ayudará a los administradores a tomar mejores decisiones. Esto puede implicar que usted siente que han estado tomando malas decisiones hasta ahora. Vender la idea de que serán capaces de obtener la información que necesitan para ayudar en la toma de decisiones.
- **Cargar el almacén con información solo porque está disponible.** No permita que el almacenamiento de datos se convierta en un vertedero de datos. Esto ralentizaría innecesariamente el uso del sistema. Hay una tendencia hacia la computación y el análisis en tiempo real. Los almacenes de datos deben cerrarse para cargar datos de forma oportuna.
- **Crear que el diseño de la base de datos de almacenamiento de datos es el mismo que el diseño de base de datos transaccional.** En general, no lo es. El objetivo del almacenamiento de datos es acceder a agregados en lugar de a uno o unos pocos registros, como en los sistemas de procesamiento de transacciones. El contenido también es diferente, como es evidente en la forma en que se organizan los datos. DBMS tiende a ser no redundante, normalizado y relacional, mientras que los almacenes de datos son redundantes, no normalizados y multidimensionales.
- **Elegir un gestor de almacenamiento de datos orientado a la tecnología en lugar de orientado al usuario.** Una clave para el éxito del almacenamiento de datos es entender que los usuarios deben obtener lo que necesitan, no la tecnología avanzada por el bien de la tecnología.

- **Centrándose en los datos internos tradicionales orientados a registros e ignorando el valor de los datos externos y del texto, las imágenes y, tal vez, el sonido y el vídeo.** Los datos vienen en muchos formatos y deben ser accesibles a las personas adecuadas en el momento adecuado y en el formato correcto. Deben catalogarse correctamente.
- **Entrega de datos con definiciones superpuestas y confusas.** La limpieza de datos es un aspecto crítico del almacenamiento de datos. Incluye la conciliación de definiciones de datos en conflicto y formatos en toda la organización. Políticamente, esto puede ser difícil porque implica un cambio, por lo general a nivel ejecutivo.
- **Promesas de rendimiento, capacidad y escalabilidad.** Los almacenes de datos generalmente requieren más capacidad y velocidad de la que se ha presupuestado originalmente. Debe planificar con anticipación para escalar verticalmente.
- **Crear que sus problemas han terminado cuando el almacenamiento de datos está en funcionamiento.** Los proyectos DSS/BI tienden a evolucionar continuamente. Cada implementación es una iteración del proceso de creación de prototipos. Siempre será necesario agregar más y diferentes conjuntos de datos al almacenamiento de datos, así como herramientas analíticas adicionales para grupos existentes y adicionales de tomadores de decisiones. Se deben planificar presupuestos de alta energía y anuales porque el éxito genera éxito. Considerar el almacenamiento de datos como un proceso continuo.
- **Centrarse en la minería de datos *ad hoc* y los informes periódicos en lugar de alertas.** La progresión natural de la información en un almacén de datos es (1) extraer los datos de los sistemas heredados, limpiarlos y alimentarlos al almacén; (2) apoyar los informes *ad hoc* hasta que se entere de lo que la gente quiere; y 3) convertir los informes *ad hoc* en informes regulares. Este proceso de aprender lo que la gente quiere para proporcionarlo parece natural, pero no es óptimo ni

quiera práctico. Los gerentes están ocupados y necesitan tiempo para leer informes. Los sistemas de alertas son mejores que los sistemas de informes periódicos y pueden hacer que una misión de almacenamiento de datos sea crítica. Los sistemas de alertas supervisan los datos que fluyen en el almacén e informan a todas las personas clave que tienen la necesidad de saber tan pronto como se produce un evento crítico.

En muchas organizaciones, un almacenamiento de datos tendrá éxito solo si hay un fuerte soporte de alta dirección para su desarrollo y si hay un campeón del proyecto que está en lo alto en el organigrama. Aunque esto probablemente sería cierto para cualquier proyecto de TI a gran escala, es especialmente importante para la realización de un almacenamiento de datos. La implementación exitosa de un almacén de datos da como resultado el establecimiento de un marco arquitectónico que puede permitir el análisis de decisiones en toda una organización y, en algunos casos, también proporciona SCM completo al conceder acceso a la información sobre los clientes y proveedores de una organización. La implementación de almacenes de datos basados en Web (a veces denominados webhousing) ha facilitado el acceso a grandes cantidades de datos, pero es difícil determinar los beneficios duros asociados con un almacenamiento de datos. Los beneficios duros se definen como beneficios para una organización que se pueden expresar en términos monetarios. Muchas organizaciones tienen recursos de TI limitados y deben dar prioridad a los proyectos. El soporte de administración y un sólido defensor del proyecto pueden ayudar a garantizar que un proyecto de almacenamiento de datos reciba los recursos necesarios para una implementación exitosa. Los recursos de almacenamiento de datos pueden suponer un costo significativo, en algunos casos requieren procesadores de gama alta y grandes aumentos en los dispositivos de almacenamiento de acceso directo (DASD). Los almacenes de datos basados en Web también pueden tener requisitos de seguridad especiales para garantizar que solo los usuarios autorizados tengan acceso a los datos.

La participación de los usuarios en el desarrollo de datos y modelado de acceso es un factor crítico de éxito en el desarrollo del almacenamiento de

datos. Durante el modelado de datos se requiere experiencia para determinar qué datos se necesitan, definir reglas de negocio asociadas con los datos y decidir qué agregaciones y otros cálculos pueden ser necesarios. El modelado de acceso es necesario para determinar cómo se van a recuperar los datos de un almacén de datos y ayuda en la definición física del almacén ayudando a definir qué datos requieren indexación. También puede indicar si se necesitan datos dependientes para facilitar la recuperación de información. Se necesitan habilidades de equipo para desarrollar e implementar un almacén de datos, incluido un conocimiento profundo de la tecnología de la base de datos y las herramientas de desarrollo utilizadas. Los sistemas de origen y la tecnología de desarrollo, como se mencionó anteriormente, hacen referencia a las numerosas entradas y a los procesos utilizados para cargar y mantener un almacén de datos.

REFERENCIAS

ARIYACHANDRA, T. & Watson, H., “Which DataWarehouse Architecture Is Most Successful?”, *Business Intelligence Journal*, Vol. 11, No. 1, 1er trimestre, Chatsworth, CA, 2006, pp. 4-6.

_____, “Key factors in selecting a data warehouse architecture”, *Business Intelligence Journal*, 10 (2), Chatsworth, CA, 2005, pp. 19-26.

HAMMERGREN, T., *Data Warehousing for Dummies*, Hoboken, NJ: John Wiley & Sons, 2009.

HOFFER, J., Prescott, M., & Mcfadden, F., *Modern database management*, 8a ed., Upper Saddle River, NJ: Pearson Prentice Hall, 2009.

INMON, W., *Building the Data Warehouse*. 2a ed, The University of Michigan QED Technical Publishing Group, 1992.

_____, *Building the Data Warehouse*, 3a ed, NJ: John Wiley & Sons, 2005.

KIMBALL, R., & Ross, M., *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3er ed, Indianapolis, IN: John Wiley & Sons, 2013.

SHARDA, R., Delen, D. & Turban, E. *Business Intelligence: A Managerial Perspective on Analytics*. 3er ed., Harlow, Essex, Pearson Education Limited, 2014.

SOLOMON, M., “Ensuring a Successful Data Warehouse Initiative.”, *Information Systems Management*, 22(1), United States: CIGNA, Inc., 2005, pp. 26-36.

CAPÍTULO 4

Analítica descriptiva

4.1 Solución de problemas de análisis de datos

Los gerentes de diferentes niveles organizativos se enfrentan a diferentes tipos de problemas. En una empresa de fabricación, por ejemplo, los altos directivos confrontan las decisiones de asignar recursos financieros, construir o ampliar instalaciones, determinar la combinación de productos y abastecer estratégicamente la producción. Los gerentes intermedios en las operaciones desarrollan planes de distribución, planes de producción e inventario, y planes de personal. Los gerentes de finanzas analizan los riesgos, determinan las estrategias de inversión y toman decisiones de precios. Los gerentes de marketing desarrollan planes publicitarios y establecen la asignación de fuerza de ventas. En las operaciones de fabricación, los problemas implican el tamaño de las ejecuciones diarias de producción, las programaciones de máquinas individuales y las asignaciones de trabajadores. Cualquiera que sea el problema, el primer paso es darse cuenta de que existe.

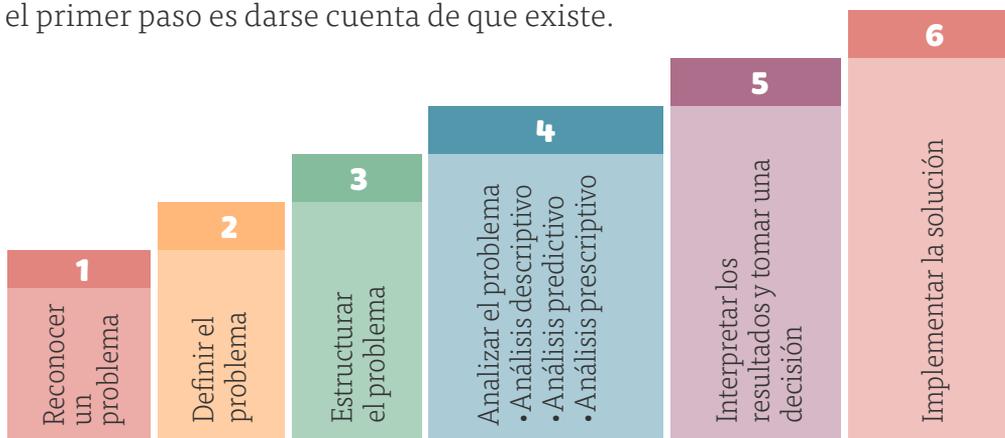


Figura 25. Fases de la solución de problemas de análisis de datos

Fuente: Elaboración propia

Paso 1. Reconocer el problema

Los problemas existen cuando hay una brecha entre lo que está sucediendo y lo que creemos que debería estar sucediendo. Por ejemplo, un administrador de productos de consumo podría sentir que los costos de

distribución son demasiado altos. Este reconocimiento podría ser el resultado de comparar el rendimiento con un competidor, observando una tendencia creciente en comparación con años anteriores.

Paso 2. Definición del problema

El segundo paso en el proceso de resolución de problemas es definir claramente el problema. Encontrar el problema real y distinguirlo de los síntomas que se observan es un paso crítico. Por ejemplo, los altos costos de distribución podrían derivar de ineficiencias en el enrutamiento de camiones, mala ubicación de los centros de distribución o factores externos como el aumento de los costos de combustible. El problema podría definirse como mejorar el proceso de enrutamiento, rediseñar todo el sistema de distribución o cubrir de manera óptima las compras de combustible, por ejemplo.

Definir problemas no es una tarea trivial. La complejidad de un problema aumenta cuando ocurre lo siguiente:

- El número de posibles cursos de acción es grande.
- El problema pertenece a un grupo en lugar de a un individuo.
- El solucionador de problemas tiene varios objetivos de la competencia.
- Los grupos o individuos externos se ven afectados por el problema.
- El solucionador de problemas y el verdadero propietario del problema —la persona que experimenta el problema y es responsable de resolverlo— no son lo mismo.
- Las limitaciones de tiempo son importantes.

Estos factores dificultan el desarrollo de objetivos significativos y caracterizan la gama de decisiones potenciales. Al definir los problemas, es importante involucrar a todas las personas que toman las decisiones o que pueden verse afectadas por ellos.

Paso 3. Estructuración del problema

Por lo general, esto implica indicar metas y objetivos, caracterizar las posibles decisiones e identificar cualquier restricción. Por ejemplo, si el problema es rediseñar un sistema de distribución, las decisiones podrían implicar nuevas ubicaciones para plantas de fabricación y almacenes (¿dónde?), nuevas asignaciones de productos a plantas (¿cuáles?) y la cantidad de cada producto para enviar desde diferentes almacenes a los clientes (¿cuánto?). El objetivo de la reducción de costos podría medirse por el costo total de entrega del producto. El gerente probablemente desearía asegurarse de que un nivel específico de servicio al cliente, por ejemplo, ser capaz de entregar pedidos dentro de 48 horas, se logra con el rediseño. Este es un ejemplo de una restricción. Estructurar un problema a menudo implica desarrollar un modelo formal.

Paso 4. Analizar el problema

Aquí es donde el análisis juega un papel importante. El análisis implica algún tipo de experimento o proceso de solución, como evaluar diferentes escenarios, analizar los riesgos asociados con varias alternativas de decisión, encontrar una solución que cumpla ciertos objetivos o determinar una solución óptima. Los profesionales de análisis han pasado décadas desarrollando y perfeccionando una variedad de enfoques para abordar diferentes tipos de problemas.

El análisis empresarial comienza con la recopilación, organización y manipulación de datos y está respaldado por tres componentes principales:¹

1. Partes de esta sección están adaptadas de Lustig, Dietric, Johnson y Dziekan, 2010.

- **Análisis descriptivo**

La mayoría de las empresas comienzan con análisis descriptivos: el uso de datos para comprender el rendimiento empresarial pasado y actual, y tomar decisiones informadas. El análisis descriptivo es el tipo de análisis más utilizado y más bien entendido. Estas técnicas clasifican, caracterizan, consolidan los datos para convertirlos en información útil con el fin de comprender y analizar el rendimiento del negocio. El análisis descriptivo resume los datos en gráficos e informes significativos; por ejemplo, sobre presupuestos, ventas, ingresos o costos. Este proceso permite a los gerentes obtener informes estándar y personalizados, y luego profundizar en los datos y realizar consultas para comprender el impacto de una campaña publicitaria; por ejemplo, revisar el rendimiento del negocio para encontrar problemas o áreas de oportunidad, e identificar patrones y tendencias en los datos. Las preguntas típicas que el análisis descriptivo ayuda a responder son “¿Cuánto vendimos en cada región?”, «¿Cuáles fueron nuestros ingresos y ganancias el último trimestre?”, «¿Cuántas y qué tipos de quejas por resolver?”, «¿Qué fábrica tiene la menor productividad?». El análisis descriptivo también ayuda a las empresas a clasificar a los clientes en diferentes segmentos, lo que les permite desarrollar campañas de marketing específicas y estrategias publicitarias.

- **Análisis predictivo**

El análisis predictivo busca predecir el futuro examinando los datos históricos, detectando patrones o relaciones en estos datos y luego extrapolando estas relaciones en el tiempo. Por ejemplo, un vendedor podría desear inferir la respuesta de diferentes segmentos de clientes a una campaña publicitaria, un comerciante de materias primas podría desear conocer los movimientos a corto plazo en los precios de las materias primas, o un fabricante de ropa de esquí podría querer pronosticar la demanda de ropa de esquí de un color y tamaño específicos. Los análisis predictivos pueden predecir el riesgo y encontrar relaciones en datos que no son evidentes con los análisis tradicionales. Mediante técnicas avanzadas, el análisis predictivo puede ayudar a detectar pa-

trones ocultos en grandes cantidades de datos para segmentar y agrupar datos en conjuntos coherentes para prever el comportamiento y detectar tendencias. Por ejemplo, un gerente de banco podría querer identificar a los clientes más rentables o estimar las posibilidades de que un solicitante de préstamo incumpla, o alertar a un cliente de tarjeta de crédito de un posible cargo fraudulento. El análisis predictivo ayuda a responder preguntas como “¿Qué sucederá si la demanda cae un 10% o si los precios de los proveedores suben un 5%?”, “¿Qué esperamos pagar por el combustible en los próximos meses?”, “¿Cuál es el riesgo de perder dinero en una nueva empresa?”

- **Análisis prescriptivo**

Muchos problemas como la programación de aeronaves y asignación de empleados en el diseño de la cadena de suministro, por ejemplo, simplemente implican demasiadas opciones y alternativas. El análisis prescriptivo utiliza la optimización para identificar la mejor alternativa para minimizar o maximizar algún objetivo. El análisis prescriptivo se utiliza en muchas áreas de negocio, incluyendo operaciones, marketing y finanzas. Por ejemplo, podemos determinar la mejor estrategia de precios y publicidad para maximizar los ingresos, la cantidad óptima de efectivo para almacenar en cajeros automáticos o la mejor combinación de inversiones en una cartera de jubilación para gestionar el riesgo. Las técnicas matemáticas y estadísticas de análisis predictivo también se pueden combinar con la optimización para tomar decisiones y que, además, tienen en cuenta la incertidumbre en los datos. El análisis prescriptivo aborda preguntas como “¿Cuánto debemos producir para maximizar el beneficio?”, “¿Cuál es la mejor manera de enviar mercancías desde nuestras fábricas para minimizar los costos?”, «¿Deberíamos cambiar nuestros planes si un desastre natural cierra la fábrica de un proveedor, si es así, por cuánto?»

Paso 5. Interpretar los resultados y tomar una decisión

Interpretar los resultados de la fase de análisis es crucial para tomar buenas decisiones. Los modelos no pueden capturar cada detalle del problema real y los gerentes deben entender las limitaciones de los modelos y sus supuestos subyacentes, y a menudo incorporar el juicio en la toma de una decisión. Por ejemplo, al localizar una instalación, podríamos utilizar un procedimiento analítico para encontrar una ubicación “central”; sin embargo, muchas otras consideraciones deben incluirse en la decisión: como el acceso a las carreteras, el suministro de mano de obra y el costo de las instalaciones. Por lo tanto, la ubicación especificada por una solución analítica podría no ser la ubicación exacta que la empresa realmente elige.

Paso 6. Implementación de la solución

Implementar la solución significa hacer que funcione en la organización, o traducir los resultados de un modelo al mundo real. Por lo general, esto requiere proporcionar recursos adecuados, motivar a los empleados, eliminar la resistencia al cambio, modificar las políticas organizativas y desarrollar la confianza. Los problemas y sus soluciones afectan a las personas: clientes, proveedores y empleados. Todos deben ser una parte importante del proceso de resolución de problemas. La sensibilidad a las cuestiones políticas y organizativas es una habilidad importante que tanto los gerentes como los profesionales analíticos deben poseer a la hora de resolver problemas.

En cada uno de estos pasos, una buena comunicación es vital. Los profesionales de análisis deben poder comunicarse con los gerentes y los clientes para comprender el contexto empresarial del problema y explicar los resultados de forma clara y eficaz. Tales habilidades como la construcción de buenos gráficos visuales y hojas de cálculo que sean fáciles de entender son vitales para los usuarios de análisis.

4.2 Datos para analítica de negocios

Un conjunto de datos es simplemente una colección de datos. Las respuestas de las encuestas de marketing, una tabla de precios históricos de las acciones y una colección de medidas de dimensiones de un artículo fabricado son ejemplos de conjuntos de datos. Una base de datos es una colección de archivos relacionados que contienen registros sobre personas, lugares o cosas. Las personas, lugares o cosas para las que almacenamos y mantenemos información se denominan entidades². Una base de datos para un minorista en línea que vende libros de fitness instructivos y DVD, por ejemplo, podría consistir en un archivo para tres entidades: editores desde los que se compran bienes, transacciones de ventas de clientes e inventario de productos. Un archivo de base de datos normalmente se organiza en una tabla bidimensional, donde las columnas corresponden a cada elemento individual de datos (denominados campos o atributos) y las filas representan registros de elementos de datos relacionados. Una característica clave de las bases de datos computarizadas es la capacidad de relacionar rápidamente un conjunto de archivos con otro.

Una métrica es una unidad de medida que proporciona una manera de cuantificar objetivamente el rendimiento. Por ejemplo, los gerentes senior podrían evaluar el rendimiento general del negocio utilizando métricas tales como el beneficio neto, el retorno de la inversión, la cuota de mercado y la satisfacción del cliente. Un gerente de planta puede supervisar métricas como la proporción de piezas defectuosas producidas o el número de turnos de inventario cada mes. Para un minorista basado en Web, algunas métricas útiles son el porcentaje de pedidos completados con precisión y el tiempo necesario para completar el pedido de un cliente. La medición es el acto de obtener datos asociados a una métrica. Las medidas son valores numéricos asociados a una métrica.

2. Laudon & Laudon, 2011.

Las métricas pueden ser discretas o continuas.

- Una métrica discreta se deriva de contar algo. Por ejemplo, una entrega es a tiempo o no; un pedido es completo o incompleto; o una factura puede tener uno, dos, tres o cualquier número de errores. Algunas métricas discretas asociadas con estos ejemplos serían la proporción de entregas puntuales, el número de pedidos incompletos cada día y el número de errores por factura.
- Las métricas continuas se basan en una escala de medición continua. Cualquier métrica que implique dólares, duración, tiempo, volumen o peso.

Otra clasificación de datos es por el tipo de escala de medición. Los datos pueden clasificarse en cuatro grupos:

- 1. Datos categóricos (nominales).** Se clasifican en categorías, según las características especificadas. Por ejemplo, los clientes de una empresa podrían ser clasificados por su región geográfica (América del Norte, América del Sur, Europa y Pacífico); empleados podrían ser clasificados como gerentes, supervisores y asociados. Las categorías no tienen ninguna relación cuantitativa entre sí, pero normalmente asignamos un número arbitrario a cada categoría para facilitar el proceso de gestión de los datos y las estadísticas informáticas. Los datos categóricos generalmente se cuentan o expresan como proporciones o porcentajes.
- 2. Datos ordinales.** Se pueden ordenar o clasificar de acuerdo con alguna relación entre sí. Las clasificaciones de fútbol universitario o baloncesto son ordinales; una clasificación más alta significa un equipo más fuerte, pero no especifica ninguna medida numérica de fuerza. Los datos ordinales son más significativos que los datos categóricos porque los datos se pueden comparar entre sí. Un ejemplo común en el negocio son los datos de las escalas de encuestas, por ejemplo, calificar un servicio como pobre, promedio, bueno, muy bueno o excelente. Estos datos son categóricos, pero también tienen un orden natural

(excelente es mejor que muy bueno) y, en consecuencia, son ordinales. Sin embargo, los datos ordinales no tienen unidades de medida fijas, por lo que no podemos hacer declaraciones numéricas significativas sobre las diferencias entre categorías. Por lo tanto, no podemos decir que la diferencia entre excelente y muy bueno es la misma que entre bueno y medio, por ejemplo. Del mismo modo, un equipo clasificado número 1 puede ser muy superior al equipo número 2, mientras que puede haber poca diferencia entre los equipos clasificados en el noveno y el décimo puesto.

- 3. Datos de intervalo.** Son ordinales, pero tienen diferencias constantes entre las observaciones y tienen puntos cero arbitrarios. Ejemplos comunes son el tiempo y la temperatura. El tiempo es relativo a la ubicación global, mientras que los calendarios tienen fechas de inicio arbitrarias (comparar, por ejemplo, el calendario gregoriano estándar con el calendario chino). Las escalas Fahrenheit y Celsius representan una medida especificada de distancia (grados), pero tienen cero puntos arbitrarios. Por lo tanto, no podemos tomar proporciones significativas; por ejemplo, no podemos decir que 50 grados es el doble de caliente que 25 grados. Sin embargo, podemos comparar diferencias. Otro ejemplo son las puntuaciones SAT o GMAT. Las puntuaciones se pueden utilizar para clasificar a los estudiantes, pero solo las diferencias entre las puntuaciones proporcionan información sobre cuánto mejor se desempeñó un alumno sobre otro; relaciones que tienen poco sentido. A diferencia de los datos ordinales, los datos de intervalo permiten una comparación significativa de rangos, promedios y otras estadísticas.

En el negocio, los datos de las escalas de la encuesta, aunque técnicamente ordinales, a menudo se tratan como datos de intervalo cuando las escalas numéricas están asociadas a las categorías (por ejemplo, 1-pobre, 2-promedio, 3-bueno, 4-muy bueno, 5-excelente). Estrictamente hablando, esto no es correcto porque la “distancia” entre categorías puede no ser percibida como la misma (los encuestados podrían perci-

bir una brecha mayor entre pobre y promedio que entre bueno y muy bueno, por ejemplo). Sin embargo, muchos usuarios de datos de encuestas los tratan como un intervalo al analizar los datos, especialmente cuando solo se utiliza una escala numérica sin etiquetas descriptivas.

- 4. Datos de razón.** Son continuos y tienen un cero natural. La mayoría de los datos empresariales y económicos, como los dólares y el tiempo, entran en esta categoría. Por ejemplo, los dólares de medida tienen un cero absoluto. Las proporciones de las cifras del dólar son significativas. Por ejemplo, saber que la región de Seattle vendió 12 millones de dólares en marzo, mientras que la región de Tampa vendió 6 millones de dólares significa que Seattle vendió el doble que Tampa.

Esta clasificación es jerárquica en la que cada nivel incluye todo el contenido de información del que lo precede. Por ejemplo, los datos ordinales también son categóricos y la información de relación se puede convertir a cualquiera de los otros tipos de datos. La información de intervalo se puede convertir en datos ordinales o categóricos, pero no se puede convertir en datos de relación sin el conocimiento del punto cero absolutos. Por lo tanto, una escala de relación es la forma más fuerte de medición.

Otras características importantes para considerar en los datos son su fiabilidad y validez. Los datos deficientes pueden dar lugar a malas decisiones. Por ejemplo, en cierta situación, un modelo de diseño de sistema de distribución se basaba en datos obtenidos del departamento de finanzas corporativas. Los costos de transporte se determinaron utilizando una fórmula basada en la latitud y longitud de las ubicaciones de las plantas y los clientes. Pero cuando la solución se representó en un programa de mapeo del sistema de información geográfica (SIG), resultó que uno de los clientes estaba en el Océano Atlántico.

Por lo tanto, los datos utilizados en las decisiones empresariales deben ser fiables y válidos. La fiabilidad significa que los datos son precisos y coherentes. Validez significa que los datos miden correctamente lo que se

supone que deben medir. Por ejemplo, un manómetro de neumáticos que lee constantemente varias libras de presión por debajo del valor real no es confiable, aunque es válido porque mide la presión de los neumáticos. El número de llamadas a un servicio de atención al cliente puede contarse correctamente cada día (y, por lo tanto, es una medida confiable), pero no es válido si se utiliza para evaluar la insatisfacción del cliente, ya que muchas llamadas pueden ser consultas simples. Por último, una pregunta de encuesta que pide al cliente que califique la calidad de los alimentos en un restaurante no puede ser confiable (porque diferentes clientes pueden tener percepciones contradictorias) ni válida (si la intención es medir la satisfacción del cliente, como satisfacción del cliente generalmente incluye otros elementos de servicio además de la comida).

4.3 Herramientas para la analítica de datos

Las hojas de cálculo son una plataforma eficaz para manipular datos, y desarrollar y resolver modelos; soportan poderosos complementos comerciales y facilitan la comunicación de resultados. Las hojas de cálculo proporcionan un entorno de modelado flexible y son especialmente útiles cuando el usuario final no es el diseñador del modelo.

A continuación, se agrupan las funciones de Excel® de Microsoft desde las principales actividades de la ciencia de datos, véase la figura 26.

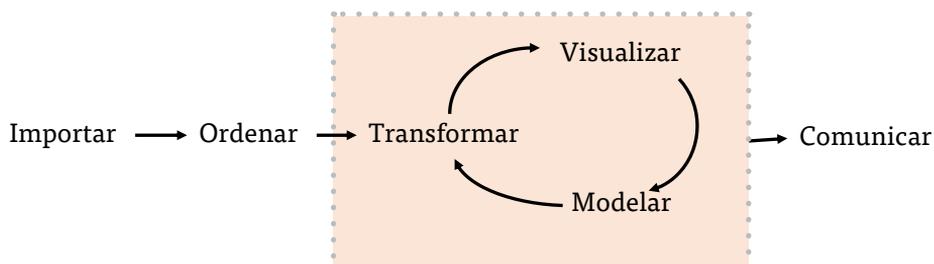


Figura 26. Herramientas para la analítica de datos
Elaborado con base en Wickham y Grolemund, 2017

1. Importar, primero se deben importar los datos a Excel®. Esto generalmente significa que se toman los datos almacenados en un archivo o base de datos y se cargan en Microsoft Excel®.
2. Poner en orden los datos significa almacenarlos en una forma consistente que coincida con la semántica del conjunto de datos con la forma en que se almacenan. En resumen, cuando sus datos están ordenados, cada columna es una variable y cada fila es una observación. Los datos ordenados son importantes porque la estructura coherente le permite centrar su lucha en las preguntas sobre los datos, no en luchar para obtener los datos en la forma correcta para diferentes funciones.
3. Después de ordenar los datos, el primer paso común es transformarlos. La transformación incluye reducir las observaciones de interés (como todas las personas en una ciudad, o todos los datos del año pasado), crear nuevas variables que sean funciones de variables existentes (como calcular la velocidad a partir de la distancia y el tiempo) y calcular un conjunto de resumen de estadísticas (como recuentos o medios).
4. Una vez que se tengan los datos ordenados con las variables que se necesitan, hay dos motores principales de generación de conocimiento: visualización y modelado. Estos tienen fortalezas y debilidades complementarias, por lo que cualquier análisis real se repetirá entre ellos muchas veces.

La visualización es una actividad fundamentalmente humana. Una buena visualización mostrará cosas que no se esperaban o generará nuevas preguntas sobre los datos. Asimismo, puede indicar que se está haciendo una pregunta incorrecta o que necesita recopilar datos diferentes. Las visualizaciones pueden sorprenderte, se requiere que un humano las interprete.

5. Los modelos son herramientas complementarias para la visualización. Una vez que se hayan hecho las preguntas lo suficientemente precisas, se puede usar un modelo para responderlas. Los modelos son una herramienta fundamentalmente matemática o computacional, por lo que generalmente se escalan bien. Incluso cuando no lo hacen, ¡generalmente es más barato comprar más computadoras que comprar más cerebros! Pero cada modelo hace suposiciones y, por su propia naturaleza,

un modelo no puede cuestionar sus propias suposiciones. Eso significa que un modelo no puede sorprender, fundamentalmente.

6. El último paso de la ciencia de datos es la comunicación, una parte absolutamente crítica de cualquier proyecto de análisis de datos. No importa qué tan bien los modelos y visualización se hayan realizado para comprender los datos, lo importante es que también se puedan comunicar esos resultados a otros.

4.4 Importar datos

Para poder aplicar el análisis de los datos se debe estar familiarizado con los conceptos y procedimientos de hoja de cálculo más elementales, como abrir, guardar e imprimir archivos; seleccionar celdas y rangos; insertar/eliminar filas y columnas; así como, introducir y editar texto, datos numéricos y fórmulas en celdas.

Los menús y comandos en Excel® 2013 residen en la imagen como se muestra en la figura 27. Los menús y comandos se organizan en grupos lógicos en diferentes pestañas (Archivo, Inicio, Insertar, etc.); los pequeños triángulos que apuntan hacia abajo indican menús de opciones adicionales. A menudo nos referimos a ciertos comandos u opciones y dónde se pueden encontrar en la cinta de opciones.



Figura 27. Imagen de Excel®

Copiar fórmulas

Excel® proporciona varias formas de copiar fórmulas en diferentes celdas. Esto es extremadamente útil en la creación de modelos de decisión, porque

muchos modelos requieren la replicación de fórmulas para diferentes períodos de tiempo, productos similares, etc. Una manera es seleccionar la celda con la fórmula que se va a copiar, hacer clic en copiar botón desde la pestaña principal “Inicio”, subpestaña “Portapapeles” (o simplemente presionar Ctrl-C en el teclado), hacer clic en la celda que desea copiar, y luego hacer clic en Pegar botón (o presionar Ctrl-V). También se puede ingresar una fórmula directamente en un rango de celdas sin copiar y pegar seleccionando el rango, escribiendo la fórmula y presionando Ctrl-Enter.

Para copiar una fórmula de una sola celda o rango de celdas en una columna o a través de una fila, primero se selecciona la celda o rango, se hace clic y se mantiene presionado el ratón en el cuadrado pequeño en la esquina inferior derecha de la celda (el “controlador de relleno”), y se arrastra la fórmula a las celdas “objetivo” que se desean copiar.

Otros consejos útiles de Excel®

Pantalla dividida. Se puede dividir la hoja de trabajo horizontal y/o verticalmente para ver diferentes partes de la hoja de trabajo al mismo tiempo. La barra divisora vertical está justo a la derecha de la barra de desplazamiento inferior, y la barra divisora horizontal está justo encima de la barra de desplazamiento derecha. Coloque el cursor sobre uno de estos hasta que cambie de forma, haga clic y arrastre la barra divisora hacia la izquierda o hacia abajo (véase la figura 28).

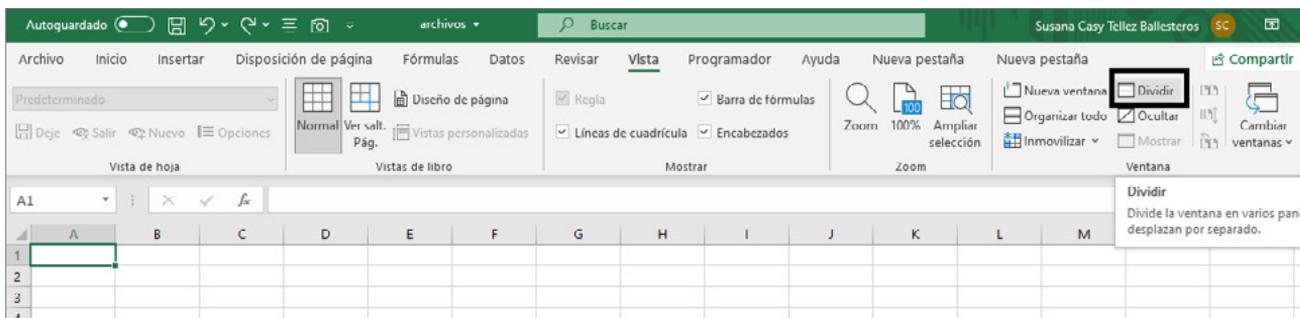


Figura 28. Función división de una hoja de Excel® Fuente: Microsoft Excel®

Pegar Especial. Cuando normalmente se copian (una o más) celdas y se pegan en una hoja de trabajo, Excel® coloca una copia exacta de las fórmulas o datos en las celdas (excepto para el direccionamiento relativo). A menudo simplemente se desea el resultado de las fórmulas, por lo que los datos permanecerán constantes incluso si cambian otros parámetros utilizados en las fórmulas. Para ello, se utiliza la opción Pegar especial que se encuentra en el menú Pegar del grupo Portapapeles en la pestaña Inicio en lugar del comando Pegar. Al elegir Pegar valores se pegará el resultado de las fórmulas a partir de las cuales se calcularon los datos, véanse las figuras 29 y 30.

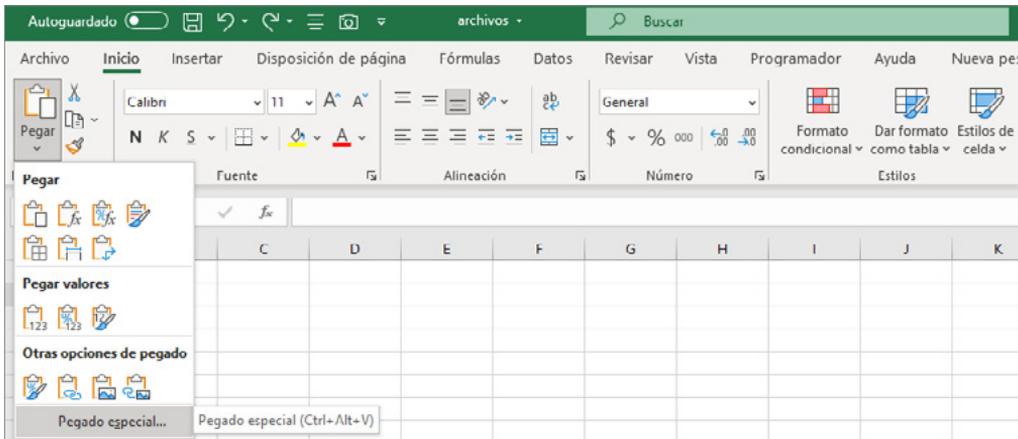


Figura 29. Función Pegado especial de una hoja de Excel®

Fuente: Microsoft Excel®

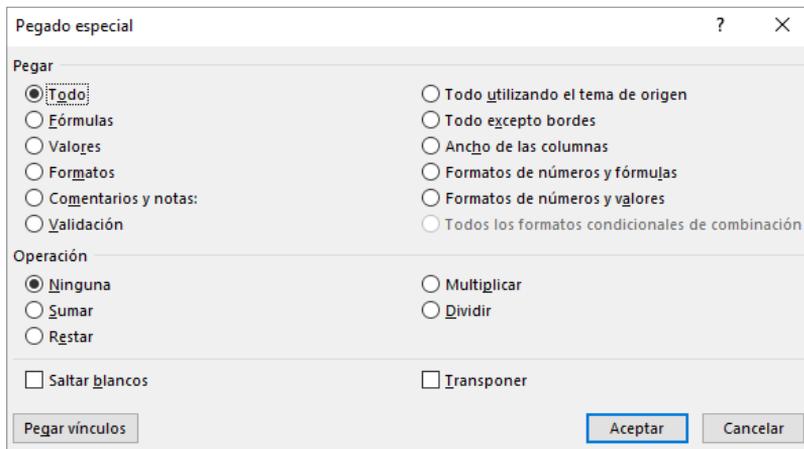


Figura 30. Menú Pegado especial de una hoja de Excel®

Fuente: Microsoft Excel®

Anchos de columna y fila. Muchas veces una celda contiene un número que es demasiado grande para mostrarse correctamente porque el ancho de la columna es demasiado pequeño. Se puede cambiar el ancho de columna para que se ajuste al valor más grande o a la cadena de texto en cualquier parte de la columna, colocando el cursor a la derecha de la etiqueta de columna para que cambie a una cruz con flechas horizontales y, a continuación, hacer doble clic. También se puede mover la flecha a la izquierda o a la derecha para cambiar manualmente el ancho de la columna. Se pueden cambiar las alturas de fila de forma similar, moviendo el cursor debajo de la etiqueta de número de fila. Esto puede ser especialmente útil si se tiene una fórmula muy larga para mostrar. Para romper una fórmula dentro de una celda, colocar el cursor en el punto de interrupción de la barra de fórmulas y presionar Alt-Enter.

Visualización de fórmulas en hojas de trabajo. Elegir Mostrar fórmulas en el grupo Auditoría de fórmulas en la ficha Fórmulas. A menudo es necesario cambiar el ancho de la columna para mostrar las fórmulas correctamente, véase la figura 31.

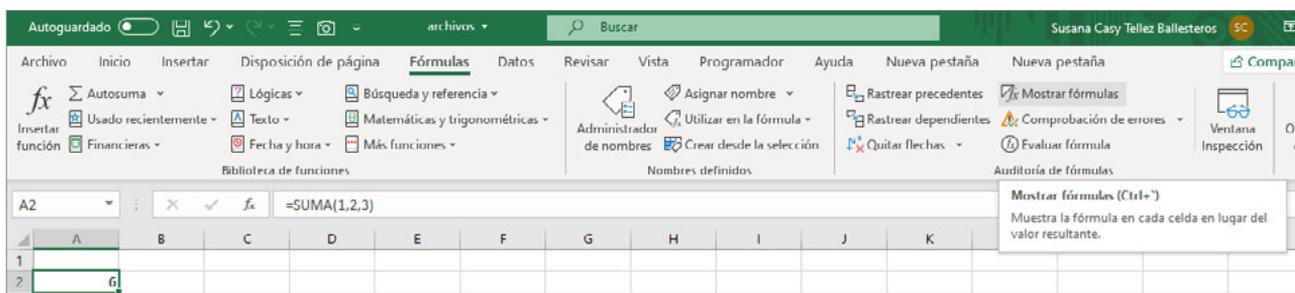


Figura 31. Función Mostrar fórmulas de una hoja de Excel®

Fuente: Microsoft Excel®

Visualización de líneas de cuadrícula y encabezados de fila y columna para la impresión. Seleccionar del menú “Vista”, la opción “Diseño de página”, el submenú mostrar: seleccionar las casillas Regla, Barra de fórmulas, Líneas de cuadrícula y Encabezados, para modificar el formato de impresión de la página, véase la figura 32.

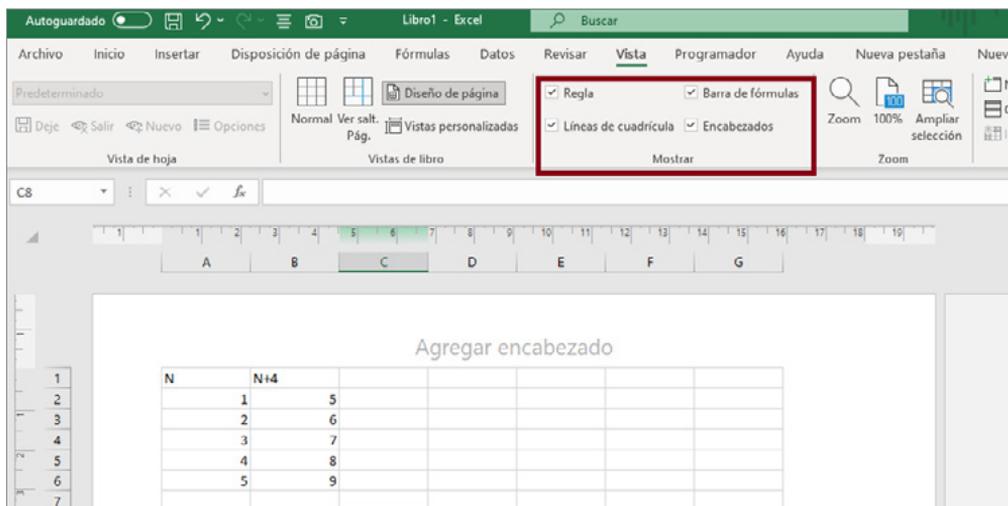


Figura 32. Función Diseño de página de una hoja de Excel®
Fuente: Microsoft Excel®

Llenar un rango con una serie de números. Supongamos que se desea crear una hoja de cálculo para introducir 100 valores de datos. Sería tedioso tener que introducir los números del 1 al 100 uno a la vez. Simplemente se rellenan los primeros valores de la serie y se resaltan. A continuación, se hace clic y se arrastra el pequeño cuadrado (controlador de relleno) en la esquina inferior derecha hacia abajo (Excel® mostrará una pequeña ventana emergente que indica el último valor en el rango) hasta que se haya rellenado la columna a 100; a continuación, se suelta el ratón, véanse las figuras 33 y 34.

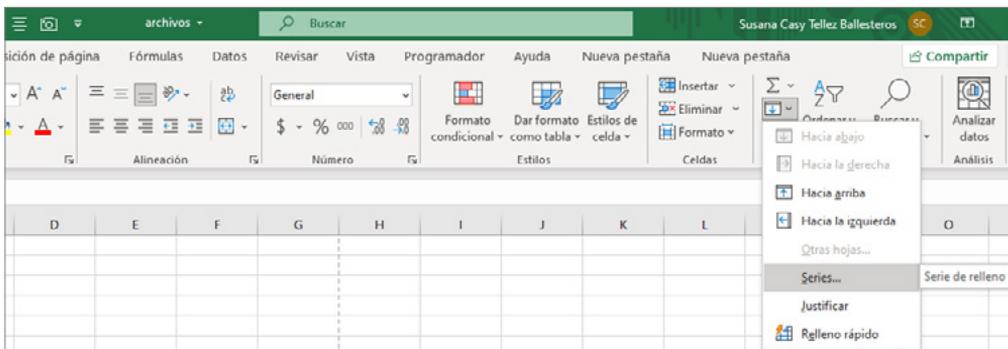


Figura 33. Función ingreso de Serie en una página de una hoja de Excel®
Fuente: Microsoft Excel®

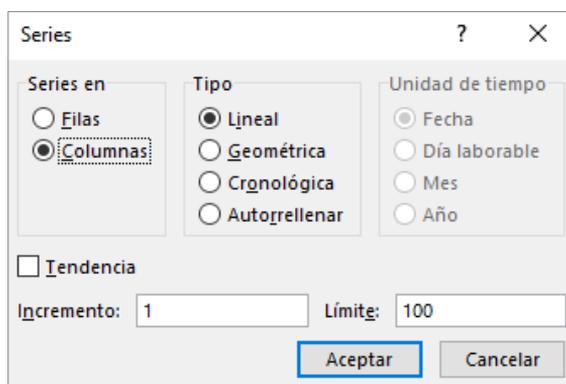


Figura 34. Menú de función ingreso de Serie en una página de una hoja de Excel® **Fuente:** Microsoft Excel®

4.5 Ordenar los datos

Para extraer información clave de una base de datos para apoyar a los representantes de servicio al cliente, soporte técnico, fabricación y otras necesidades, Excel® proporciona algunas funciones útiles para encontrar esos datos específicos en una hoja de cálculo. En la tabla 9, se muestran algunas de estas funciones:

Tabla 9. Funciones búsqueda de datos. **Fuente:** Microsoft Excel®

| Función en español | Función en inglés | Descripción |
|--|--|---|
| BUSCARV(valor_buscado, buscar_en_matriz, num_indicador_columnas, [ordenado]) | VLOOKUP(lookup_value, table_array, col_index_num, [búsqueda de rango]) | Busca un valor en la columna más a la izquierda de una tabla (especificada por el table_array) y devuelve un valor en la misma fila de una columna que especifique (col_index_num). |

| Función en español | Función en inglés | Descripción |
|---|--|--|
| BUSCARH(valor_buscado, buscar_en_matriz, num_indicador_filas, [ordenado]) | HLOOKUP(lookup_value, table_array, row_index_num, [búsqueda de rango]) | Busca un valor en la fila superior de una tabla y devuelve un valor en la misma columna de una fila que especifique. |
| INDICE(matriz; núm_fila; [núm_columna]) | INDEX(array, row_num, col_num) | Devuelve un valor o una referencia de la celda en la intersección de una fila y columna determinada en un intervalo determinado. |
| COINCIDIR(valor_buscado, matriz_buscada, [tipo_de_coincidencia]) | MATCH(lookup_value, lookup_array, match_type) | Devuelve la posición relativa de un elemento en una matriz que coincide con un valor especificado en un orden especificado. |

En las funciones VLOOKUP y HLOOKUP, la búsqueda de rango es opcional. Si se omite o se establece como True, la primera columna de la tabla debe ordenarse en orden numérico ascendente. Si se encuentra una coincidencia exacta para el valor de búsqueda en la primera columna, Excel® devolverá el valor que dé el número de esa fila o columna. Si no se encuentra una coincidencia exacta, Excel® elegirá la fila con el valor más grande en la primera columna que es menor que el valor buscado. Si la búsqueda de rango es false, Excel® buscará una coincidencia exacta en la primera columna del rango de tabla. Si no se encuentra ninguna coincidencia exacta, Excel® devolverá #N/A (no disponible). Se recomienda especificar la búsqueda de intervalo para evitar errores.



Ejemplo 1. Uso de la función BUSQUEDAV

Considere la base de datos de transacciones de ventas para una empresa que vende libros y DVD. La base de datos se ordena por el identificador de cliente y una parte de ella se muestra en la figura 35. Supongamos que un cliente llama a un representante sobre un problema de pago. El

representante encuentra el ID de cliente (por ejemplo, 10007) y necesita buscar el tipo de código de pago y transacción. Podemos usar la función VLOOKUP para hacer esto. En la función BUSCARV (valor_a_buscar, tabla_de_búsqueda, número_de_columna), valor_a_buscar representa el ID de cliente. El valor_a_buscar es el rango de los datos en la hoja de cálculo; en este caso, es el rango A4: H475. El valor de valor_a_buscar representa la columna del intervalo de tablas que desea recuperar. Para el tipo de pago, esta es la columna 3; para el código de transacción, esta es la columna 4. Tenga en cuenta que la primera columna ya está ordenada en orden numérico ascendente, por lo que podemos omitir el argumento de búsqueda de rango o establecerlo como true. Por lo tanto, si ingresamos la fórmula a continuación en cualquier celda en blanco de la hoja de cálculo:

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|----|----------------------------------|---------------|-------------|---------------------------|----------------|-----------------|-----------------|---------------------|---|----------|-----------------------------------|---|---|
| 1 | Transacciones de venta: Julio 14 | | | | | | | | | | | | |
| 2 | | | | | | | | | | | | | |
| 3 | ID cliente | Region | Pago | Código Transacción | Recurso | Cantidad | Producto | Hora del día | | | | | |
| 4 | 10001 | Este | Paypal | 93816545 | Web | \$20.19 | DVD | 22:19 | | | | | |
| 5 | 10002 | Oeste | Credito | 74083490 | Web | \$17.85 | DVD | 13:27 | | Credito | =BUSCARV(10007,\$A\$4:\$H\$475,3) | | |
| 6 | 10003 | Norte | Credito | 64942368 | Web | \$23.98 | DVD | 14:27 | | 80103311 | =BUSCARV(10007,\$A\$4:\$H\$475,4) | | |
| 7 | 10004 | Oeste | Paypal | 70560957 | Corre elec | \$23.51 | Libro | 15:38 | | | | | |
| 8 | 10005 | Sur | Credito | 35208817 | Web | \$15.33 | Libro | 15:21 | | | | | |
| 9 | 10006 | Oeste | Paypal | 20978903 | Corre elec | \$17.30 | DVD | 13:11 | | | | | |
| 10 | 10007 | Este | Credito | 80103311 | Web | \$177.72 | Libro | 21:59 | | | | | |
| 11 | 10008 | Oeste | Credito | 14132683 | Web | \$21.76 | Libro | 04:04 | | | | | |
| 12 | 10009 | Oeste | Paypal | 40128225 | Web | \$15.92 | DVD | 19:35 | | | | | |
| 13 | 10010 | Sur | Paypal | 49073721 | Web | \$23.39 | DVD | 13:26 | | | | | |
| 14 | 10011 | Sur | Paypal | 57398827 | Corre elec | \$24.45 | Libro | 14:17 | | | | | |
| 15 | 10012 | Este | Credito | 34400661 | Web | \$20.39 | Libro | 01:01 | | | | | |
| 16 | 10013 | Norte | Paypal | 54242587 | Web | \$19.54 | DVD | 10:04 | | | | | |

Figura 35. Base de datos transacciones de venta

=BUSCARV(10007,\$A\$4:\$H\$475,3)
 devuelve el tipo de pago, Crédito. Si usamos la siguiente fórmula:

=BUSCARV(10007,\$A\$4:\$H\$475,4)
 la función devuelve el código de transacción, 80103311.

Ahora, supongamos que la base de datos se ha ordenado por código de transacción para que la columna de identificador de cliente ya no esté en orden numérico ascendente, como se muestra en la figura 36. Si usamos la función “BUSCARV(10007,\$A\$4:\$H\$475,4, VERDADERO)”, Excel® devuelve

#N/A. Sin embargo, si cambiamos el argumento de búsqueda de rango a Falso, la función devuelve el valor correcto del código de transacción.

| | A | B | C | D | E | F | G | H |
|----|---|---------------|-------------|---------------------------|----------------|-----------------|-----------------|---------------------|
| 1 | Transacciones de venta: Julio 14 | | | | | | | |
| 2 | | | | | | | | |
| 3 | ID cliente | Region | Pago | Código Transacción | Recurso | Cantidad | Producto | Hora del día |
| 4 | 10001 | Este | Paypal | 93816545 | Web | \$20.19 | DVD | 22:19 |
| 5 | 10002 | Oeste | Credito | 74083490 | Web | \$17.85 | DVD | 13:27 |
| 6 | 10003 | Norte | Credito | 64942368 | Web | \$23.98 | DVD | 14:27 |
| 7 | 10004 | Oeste | Paypal | 70560957 | Corre elec | \$23.51 | Libro | 15:38 |
| 8 | 10005 | Sur | Credito | 35208817 | Web | \$15.33 | Libro | 15:21 |
| 9 | 10006 | Oeste | Paypal | 20978903 | Corre elec | \$17.30 | DVD | 13:11 |
| 10 | 10007 | Este | Credito | 80103311 | Web | \$177.72 | Libro | 21:59 |
| 11 | 10008 | Oeste | Credito | 14132683 | Web | \$21.76 | Libro | 04:04 |
| 12 | 10009 | Oeste | Paypal | 40128225 | Web | \$15.92 | DVD | 19:35 |
| 13 | 10010 | Sur | Paypal | 49073721 | Web | \$23.39 | DVD | 13:26 |

Figura 36. Porción de datos de transacciones de ventas ordenados por ID de cliente

La función BUSCARH funciona de forma similar. Para la mayoría de las bases de datos de hojas de cálculo, normalmente tendríamos que utilizar la función BUSCARV. En algunas situaciones de modelado, sin embargo, la función BUSCARH puede ser útil si los datos están organizados columna por columna en lugar de fila por fila.

La función INDICE funciona como un procedimiento de búsqueda devolviendo el valor de una fila y columna determinada de una matriz. Por ejemplo, en la base de datos Transacciones de ventas, INDICE(\$A\$4:\$H\$475, 7, 4) recuperaría el código de transacción, 80103311 que se encuentra en la 7a fila y 4a columna de la matriz de datos, como hizo la función BUSCARV en el ejemplo 1. La diferencia es que se basa en el número de fila en lugar del valor real del identificador de cliente.

En la función COINCIDIR(valor_buscado) es el valor que desea que coincida en la tabla, que es el rango de celdas que se están buscando. El tipo de coincidencia es -1, 0 o 1. El valor predeterminado es 1. Si el tipo de coincidencia es 1, la función encuentra el valor más grande que es menor o igual que el valor buscado. Los valores del arreglo buscado deben colocarse en orden

ascendente. Si el tipo de coincidencia es 0, la función COINCIDIR busca el primer valor que es exactamente igual al valor buscado. Los valores del arreglo buscado pueden estar en cualquier orden. Si el tipo de coincidencia es -1, la función encuentra el valor más pequeño que es mayor o igual que el valor buscado. Los valores del arreglo buscado deben colocarse en orden descendente. El ejemplo 2 muestra cómo se pueden utilizar las funciones INDICE y COINCIDIR.

La función BUSCARV no funcionará si se desea buscar algo a la izquierda de un intervalo especificado (porque utiliza la primera columna del intervalo para buscar el valor de búsqueda). Sin embargo, podemos usar la función INDICE y COINCIDIR fácilmente para hacer esto, así como se muestra en el ejemplo 3.

Ejemplo 2. Uso de las funciones INDICE y COINCIDIR para consultas de bases de datos

La figura 37 muestra los datos del archivo de Excel® Consultas mensuales de ventas de productos. Supongamos que se desea diseñar una aplicación de consulta simple para introducir el mes y el nombre del producto, y recuperar las ventas correspondientes. Las tres hojas de trabajo adicionales del libro muestran cómo hacerlo de tres maneras diferentes.

| | A | B | C | D | E | F |
|----|--------------------------|----------|----------|-----------------|----------|----------|
| 1 | Unidades vendidas | | | | | |
| 2 | | | | Producto | | |
| 3 | Mes | A | B | C | D | E |
| 4 | Enero | 7,792 | 5,554 | 3,105 | 3,168 | 10,350 |
| 5 | Febrero | 7,268 | 3,024 | 3,228 | 3,751 | 8,965 |
| 6 | Marzo | 7,049 | 5,543 | 2,147 | 3,319 | 6,827 |
| 7 | Abril | 7,560 | 5,232 | 2,636 | 4,057 | 8,544 |
| 8 | Mayo | 8,233 | 5,450 | 2,726 | 3,837 | 7,535 |
| 9 | Junio | 8,629 | 3,943 | 2,705 | 4,664 | 9,070 |
| 10 | Julio | 8,702 | 5,991 | 2,891 | 5,418 | 8,389 |
| 11 | Agosto | 9,215 | 3,920 | 2,782 | 4,085 | 7,367 |
| 12 | Septiembre | 8,986 | 4,753 | 2,524 | 5,575 | 5,377 |
| 13 | Octubre | 8,654 | 4,746 | 3,258 | 5,333 | 7,645 |
| 14 | Noviembre | 8,315 | 3,566 | 2,144 | 4,924 | 8,173 |
| 15 | Diciembre | 7,978 | 5,670 | 3,071 | 6,563 | 6,088 |

Figura 37. Libro de consultas mensuales de ventas de productos

La hoja de cálculo Unidades vendidas (consulte la figura 38) utiliza la función BUSCARV con instrucciones SI incrustadas. Las fórmulas de la celda I8 son:

=BUSCARV(I5,A4:F15,SI(I6="A",2,SI(I6="B",3,SI(I6="C",4,SI(I6="D",5,SI(I6="E",6))))),FALSO)

| 1 | Unidades vendidas | | | | | | Usando BUSCARV + SI | |
|----|-------------------|-------|-------|-------|-------|--------|---------------------|-------|
| 2 | Producto | | | | | | | |
| 3 | Mes | A | B | C | D | E | Busqueda de ventas | |
| 4 | Enero | 7,792 | 5,554 | 3,105 | 3,168 | 10,350 | | |
| 5 | Febrero | 7,268 | 3,024 | 3,228 | 3,751 | 8,965 | Mes | Abril |
| 6 | Marzo | 7,049 | 5,543 | 2,147 | 3,319 | 6,827 | Producto | E |
| 7 | Abril | 7,560 | 5,232 | 2,636 | 4,057 | 8,544 | | |
| 8 | Mayo | 8,233 | 5,450 | 2,726 | 3,837 | 7,535 | Ventas | 8,544 |
| 9 | Junio | 8,629 | 3,943 | 2,705 | 4,664 | 9,070 | | |
| 10 | Julio | 8,702 | 5,991 | 2,891 | 5,418 | 8,389 | | |
| 11 | Agosto | 9,215 | 3,920 | 2,782 | 4,085 | 7,367 | | |
| 12 | Septiembre | 8,986 | 4,753 | 2,524 | 5,575 | 5,377 | | |
| 13 | Octubre | 8,654 | 4,746 | 3,258 | 5,333 | 7,645 | | |
| 14 | Noviembre | 8,315 | 3,566 | 2,144 | 4,924 | 8,173 | | |
| 15 | Diciembre | 7,978 | 5,670 | 3,071 | 6,563 | 6,088 | | |

Figura 38. Consulta la hoja de trabajo de Ventas de productos mensuales. Uso de la función BUSCARV +SI

Fuente: Excel®

Las funciones SI se utilizan para determinar la columna de la tabla de búsqueda que se va a usar y, como se puede ver, es algo compleja, especialmente si la tabla era mucho más grande.

En la hoja de cálculo se usan las funciones BUSCARV y COINCIDIR en la celda I8 (véase la figura 39). La fórmula en la celda I8 es:

=BUSCARV(I5,A4:F15,COINCIDIR(I6,B3:F3,0)+1,FALSO)

| | A | B | C | D | E | F | G | H | I |
|----|--------------------------|----------|----------|-----------------|----------|----------|-----------------|---------------------------|-----------------------------------|
| 1 | Unidades vendidas | | | | | | | | Usando BUSCARV + COINCIDIR |
| 2 | | | | Producto | | | | | |
| 3 | Mes | A | B | C | D | E | | Busqueda de ventas | |
| 4 | Enero | 7,792 | 5,554 | 3,105 | 3,168 | 10,350 | | | |
| 5 | Febrero | 7,268 | 3,024 | 3,228 | 3,751 | 8,965 | Mes | Septiembre | |
| 6 | Marzo | 7,049 | 5,543 | 2,147 | 3,319 | 6,827 | Producto | A | |
| 7 | Abril | 7,560 | 5,232 | 2,636 | 4,057 | 8,544 | | | |
| 8 | Mayo | 8,233 | 5,450 | 2,726 | 3,837 | 7,535 | Ventas | 8,986 | |
| 9 | Junio | 8,629 | 3,943 | 2,705 | 4,664 | 9,070 | | | |
| 10 | Julio | 8,702 | 5,991 | 2,891 | 5,418 | 8,389 | | | |
| 11 | Agosto | 9,215 | 3,920 | 2,782 | 4,085 | 7,367 | | | |
| 12 | Septiembre | 8,986 | 4,753 | 2,524 | 5,575 | 5,377 | | | |
| 13 | Octubre | 8,654 | 4,746 | 3,258 | 5,333 | 7,645 | | | |
| 14 | Noviembre | 8,315 | 3,566 | 2,144 | 4,924 | 8,173 | | | |
| 15 | Diciembre | 7,978 | 5,670 | 3,071 | 6,563 | 6,088 | | | |

Figura 39. Uso de las funciones BUSCARV + COINCIDIR
Fuente: Excel®

En este caso, la función COINCIDIR se utiliza para identificar la columna de la tabla correspondiente al nombre del producto en la celda I6. Tenga en cuenta el uso de “+1” para desplazar el número de columna relativo del producto al número de columna correcto en la tabla de búsqueda.

Por último, la hoja de cálculo solo utiliza las funciones INDICE y COINCIDIR en la celda I8 (véase la figura 40). La fórmula en la celda I8 es:

$$=INDICE(A4:F15,COINCIDIR(I5,A4:A15,0),COINCIDIR(I6,A3:F3,0))$$

| I8 =INDICE(A4:F15,COINCIDIR(I5,A4:A15,0),COINCIDIR(I6,A3:F3,0)) | | | | | | | | | |
|--|--------------------------|----------|----------|-----------------|----------|----------|-----------------|---------------------------|----------------------------------|
| | A | B | C | D | E | F | G | H | I |
| 1 | Unidades vendidas | | | | | | | | Usando INDICE + COINCIDIR |
| 2 | | | | Producto | | | | | |
| 3 | Mes | A | B | C | D | E | | Busqueda de ventas | |
| 4 | Enero | 7,792 | 5,554 | 3,105 | 3,168 | 10,350 | | | |
| 5 | Febrero | 7,268 | 3,024 | 3,228 | 3,751 | 8,965 | Mes | Agosto | |
| 6 | Marzo | 7,049 | 5,543 | 2,147 | 3,319 | 6,827 | Producto | B | |
| 7 | Abril | 7,560 | 5,232 | 2,636 | 4,057 | 8,544 | | | |
| 8 | Mayo | 8,233 | 5,450 | 2,726 | 3,837 | 7,535 | Sales | 3,920 | |
| 9 | Junio | 8,629 | 3,943 | 2,705 | 4,664 | 9,070 | | | |
| 10 | Julio | 8,702 | 5,991 | 2,891 | 5,418 | 8,389 | | | |
| 11 | Agosto | 9,215 | 3,920 | 2,782 | 4,085 | 7,367 | | | |
| 12 | Septiembre | 8,986 | 4,753 | 2,524 | 5,575 | 5,377 | | | |
| 13 | Octubre | 8,654 | 4,746 | 3,258 | 5,333 | 7,645 | | | |
| 14 | Noviembre | 8,315 | 3,566 | 2,144 | 4,924 | 8,173 | | | |
| 15 | Diciembre | 7,978 | 5,670 | 3,071 | 6,563 | 6,088 | | | |

Figura 40. Uso de las funciones INDICE + COINCIDIR
Fuente: Excel®

Las funciones COINCIDIR se utilizan como argumentos en la función INDICE para identificar los números de fila y columna de la tabla en función del mes y el nombre del producto. A continuación, la función INDICE recupera el valor de la fila y columna correspondientes. Esta es quizás la fórmula más limpia de los tres.

Ejemplo 3. Uso de INDICE y COINCIDIR para una búsqueda de tabla izquierda

Supongamos que, en la base de datos Transacciones de ventas se desea encontrar el identificador de cliente asociado a un código de transacción específico, y supongamos, además, que ingresamos el código de transacción en la celda K2 y queremos mostrar el ID de cliente en la celda K4.

Utilice la fórmula en la celda

$K4 = \text{INDICE}(A4:A475, \text{COINCIDIR}(K2, D4:D475, 0), 1)$

Aquí, la función COINCIDIR se utiliza para identificar el número de fila en el intervalo de tablas que coincide exactamente con el código de transacción, y la función INDICE utiliza este número de fila y la columna 1 para identificar el identificador de cliente asociado (véase la figura 41).

| | A | B | C | D | E | F | G | H | I | J | K |
|----|----------------------------------|--------|---------|--------------------|-------------------|----------|----------|--------------|---|--------------|----------|
| 1 | Transacciones de venta: Julio 14 | | | | | | | | | | |
| 2 | | | | | | | | | | Código Trans | 35208817 |
| 3 | ID cliente | Region | Pago | Código Transacción | Recurso | Cantidad | Producto | Hora del día | | ID cliente | 10005 |
| 4 | 10001 | Este | Paypal | 93816545 | Web | \$20.19 | DVD | 22:19 | | | |
| 5 | 10002 | Oeste | Credito | 74083490 | Web | \$17.85 | DVD | 13:27 | | | |
| 6 | 10003 | Norte | Credito | 64942368 | Web | \$23.98 | DVD | 14:27 | | | |
| 7 | 10004 | Oeste | Paypal | 70560957 | Corre electrónico | \$23.51 | Libro | 15:38 | | | |
| 8 | 10005 | Sur | Credito | 35208817 | Web | \$15.33 | Libro | 15:21 | | | |
| 9 | 10006 | Oeste | Paypal | 20978903 | Corre electrónico | \$17.30 | DVD | 13:11 | | | |
| 10 | 10007 | Este | Credito | 80103311 | Web | \$177.72 | Libro | 21:59 | | | |
| 11 | 10008 | Oeste | Credito | 14132683 | Web | \$21.76 | Libro | 04:04 | | | |
| 12 | 10009 | Oeste | Paypal | 40128225 | Web | \$15.92 | DVD | 19:35 | | | |
| 13 | 10010 | Sur | Paypal | 49073721 | Web | \$23.39 | DVD | 13:26 | | | |
| 14 | 10011 | Sur | Paypal | 57398827 | Corre electrónico | \$24.45 | Libro | 14:17 | | | |

Figura 41. Uso de las funciones INDICE + COINCIDIR en el ejemplo 3

Filtrar datos

Para archivos de datos de gran tamaño, encontrar un subconjunto determinado de registros que cumplan ciertas características mediante la ordenación puede ser tedioso. Excel® proporciona dos herramientas de filtrado: Autofiltro para criterios simples y Filtro avanzado para criterios más complejos. Estas herramientas se entienden mejor trabajando a través de algunos ejemplos.

Ejemplo 4. Filtrar registros por descripción del artículo

En la base de datos de Órdenes de compra, supongamos que estamos interesados en extraer todos los registros correspondientes al paquete de tuerca. En primer lugar, seleccione cualquier celda dentro de la base de datos. A continuación, en la pestaña Datos de Excel®, haga clic en Filtrar en el grupo Ordenar y filtrar. A continuación, se mostrará una flecha desplegable en el lado derecho de cada columna de encabezado. Al hacer clic en uno de estos, se mostrará un cuadro desplegable. Estas son las opciones para filtrar en esa columna de datos. Haga clic en el encabezado Descripción del artículo. Desmarque la casilla para Seleccionar todo y después marque la casilla correspondiente al Paquete de tuerca, tal y como se muestra en la figura 42. Haga clic en el botón “ACEPTAR”, y la herramienta Filtro mostrará solo los pedidos para este elemento (véase la figura 43). En realidad, la herramienta de filtro no extrae los registros; simplemente oculta los registros que no coinciden con los criterios. Sin embargo, se pueden copiar y pegar los datos en otra hoja de cálculo de Excel®, documento de Microsoft Word o una presentación de PowerPoint, por ejemplo. Para restaurar el archivo de datos original, haga clic en la flecha desplegable de nuevo y luego en Borrar filtro de “Descripción del artículo.”

| A | B | C | D | E | F | G | H | I | J | | |
|----|-----------------------|-----------|---------|-------------------|-----------|----------|---------------|-------------------|----------|----------|-------|
| 1 | Ordenes de compra | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | Proveedor | No. Orden | No. Pri | Descripción del a | Costo d | Cantidad | Costo por or | A/P terminos (mes | Fecha de | Fecha de | Libro |
| 4 | Sujetadores Hulkey | Au | | | \$ 4.25 | 19,500 | \$ 82,875.00 | 30 | 08/05/11 | 08/13/11 | |
| 5 | Láminas Alum | Au | | | \$ 4.25 | 10,000 | \$ 42,500.00 | 30 | 08/08/11 | 08/14/11 | |
| 6 | Aeroespacial Fast-Tie | Au | | | \$ 1.05 | 23,000 | \$ 24,150.00 | 30 | 08/10/11 | 08/15/11 | |
| 7 | Aeroespacial Fast-Tie | Au | | | \$ 1.05 | 21,500 | \$ 22,575.00 | 30 | 08/15/11 | 08/22/11 | |
| 8 | Steelpin Co. | Au | | | \$ 1.10 | 17,500 | \$ 19,250.00 | 30 | 08/20/11 | 08/31/11 | |
| 9 | Aeroespacial Fast-Tie | Au | | | \$ 1.05 | 22,500 | \$ 23,625.00 | 30 | 08/20/11 | 08/26/11 | |
| 10 | Steelpin Co. | Au | | | \$ 3.75 | 4,250 | \$ 15,937.50 | 30 | 08/25/11 | 09/01/11 | |
| 11 | Productos Durrable | Au | | | \$ 90.00 | 100 | \$ 9,000.00 | 45 | 08/25/11 | 08/28/11 | |
| 12 | Aeroespacial Fast-Tie | Au | | | \$ 2.45 | 1,300 | \$ 3,185.00 | 30 | 08/25/11 | 09/04/11 | |
| 13 | Aeroespacial Fast-Tie | Au | | | \$ 1.05 | 22,500 | \$ 23,625.00 | 30 | 08/25/11 | 09/02/11 | |
| 14 | Steelpin Co. | Au | | | \$ 1.10 | 18,100 | \$ 19,910.00 | 30 | 08/25/11 | 09/05/11 | |
| 15 | Sujetadores Hulkey | Au | | | \$ 1.25 | 5,600 | \$ 7,000.00 | 30 | 08/25/11 | 08/29/11 | |
| 16 | Sujetadores Hulkey | Au | | | \$ 0.75 | 500 | \$ 375.00 | 30 | 08/25/11 | 08/31/11 | |
| 17 | Steelpin Co. | Au | | | \$ 1.65 | 4,500 | \$ 7,425.00 | 30 | 08/28/11 | 09/05/11 | |
| 18 | Steelpin Co. | Se | | | \$ 3.75 | 4,200 | \$ 15,750.00 | 30 | 09/01/11 | 09/10/11 | |
| 19 | Láminas Alum | Se | | | \$ 255.00 | 406 | \$ 103,530.00 | 30 | 09/01/11 | 09/10/11 | |
| 20 | Sujetadores Hulkey | Se | | | \$ 1.25 | 5,500 | \$ 6,875.00 | 30 | 09/01/11 | 09/06/11 | |
| 21 | Steelpin Co. | Se | | | \$ 1.65 | 4,850 | \$ 8,002.50 | 30 | 09/02/11 | 09/11/11 | |
| 22 | Steelpin Co. | Se | | | \$ 3.75 | 4,150 | \$ 15,562.50 | 30 | 09/03/11 | 09/11/11 | |
| 23 | Sujetadores Hulkey | Se | | | \$ 4.25 | 15,500 | \$ 65,875.00 | 30 | 09/04/11 | 09/12/11 | |
| 24 | Tecnologías Spacetime | Se | | | \$ 3.55 | 4,800 | \$ 17,040.00 | 25 | 09/05/11 | 09/20/11 | |

Figura 42. Selección de registros para el Paquete de tuerca. Fuente: Excel®

| A | B | C | D | E | F | G | H | I | J | | |
|----|-----------------------|-----------|---------|-------------------|---------|----------|--------------|-------------------|----------|----------|-------|
| 1 | Ordenes de compra | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | Proveedor | No. Orden | No. Pri | Descripción del a | Costo d | Cantidad | Costo por or | A/P terminos (mes | Fecha de | Fecha de | Libro |
| 10 | Steelpin Co. | Aug11007 | 4312 | Paquete de tuerca | \$ 3.75 | 4,250 | \$ 15,937.50 | 30 | 08/25/11 | 09/01/11 | |
| 18 | Steelpin Co. | Sep11001 | 4312 | Paquete de tuerca | \$ 3.75 | 4,200 | \$ 15,750.00 | 30 | 09/01/11 | 09/10/11 | |
| 22 | Steelpin Co. | Sep11005 | 4312 | Paquete de tuerca | \$ 3.75 | 4,150 | \$ 15,562.50 | 30 | 09/03/11 | 09/11/11 | |
| 24 | Tecnologías Spacetime | Sep11007 | 4111 | Paquete de tuerca | \$ 3.55 | 4,800 | \$ 17,040.00 | 25 | 09/05/11 | 09/20/11 | |
| 32 | Tecnologías Spacetime | Sep11015 | 4111 | Paquete de tuerca | \$ 3.55 | 4,585 | \$ 16,276.75 | 25 | 09/10/11 | 09/30/11 | |
| 36 | Tecnologías Spacetime | Sep11019 | 4111 | Paquete de tuerca | \$ 3.55 | 4,200 | \$ 14,910.00 | 25 | 09/15/11 | 10/15/11 | |
| 39 | Tecnologías Spacetime | Sep11022 | 4111 | Paquete de tuerca | \$ 3.55 | 4,250 | \$ 15,087.50 | 25 | 09/20/11 | 10/10/11 | |
| 43 | Tecnologías Spacetime | Sep11026 | 4111 | Paquete de tuerca | \$ 3.55 | 4,200 | \$ 14,910.00 | 25 | 09/25/11 | 10/25/11 | |
| 61 | Tecnologías Spacetime | Oct11010 | 4111 | Paquete de tuerca | \$ 3.55 | 4,600 | \$ 16,330.00 | 25 | 10/05/11 | 10/19/11 | |
| 62 | Productos Durrable | Oct11011 | 4569 | Paquete de tuerca | \$ 3.50 | 3,900 | \$ 13,650.00 | 45 | 10/05/11 | 10/10/11 | |
| 73 | Láminas Alum | Oct11022 | 4224 | Paquete de tuerca | \$ 3.95 | 4,500 | \$ 17,775.00 | 30 | 10/15/11 | 10/20/11 | |
| 98 | | | | | | | | | | | |
| 99 | | | | | | | | | | | |

Figura 43. Resultados del filtro para el paquete de tuercas. Fuente: Excel®



Ejemplo 5. Filtrar registros por costo de artículo

En este ejemplo, supongamos que se desea identificar todos los registros de la base de datos de pedidos de compra cuyo costo del artículo es de al menos 200. En primer lugar, haga clic en la flecha desplegable de la columna Costo de artículo y coloque el cursor sobre Filtro de números. Esto muestra una lista de opciones, como se muestra en la figura 44, trae Mayor_o_igual_que. . . de la lista. Este cuadro de diálogo Autofiltro personalizado (véase figura 20) permite especificar hasta dos criterios específicos mediante la lógica “y” y “o”. Ingrese 200 en el cuadro como se muestra, y después haga clic en Aceptar. La herramienta mostrará todos los registros con un costo de artículo de \$200 o más.

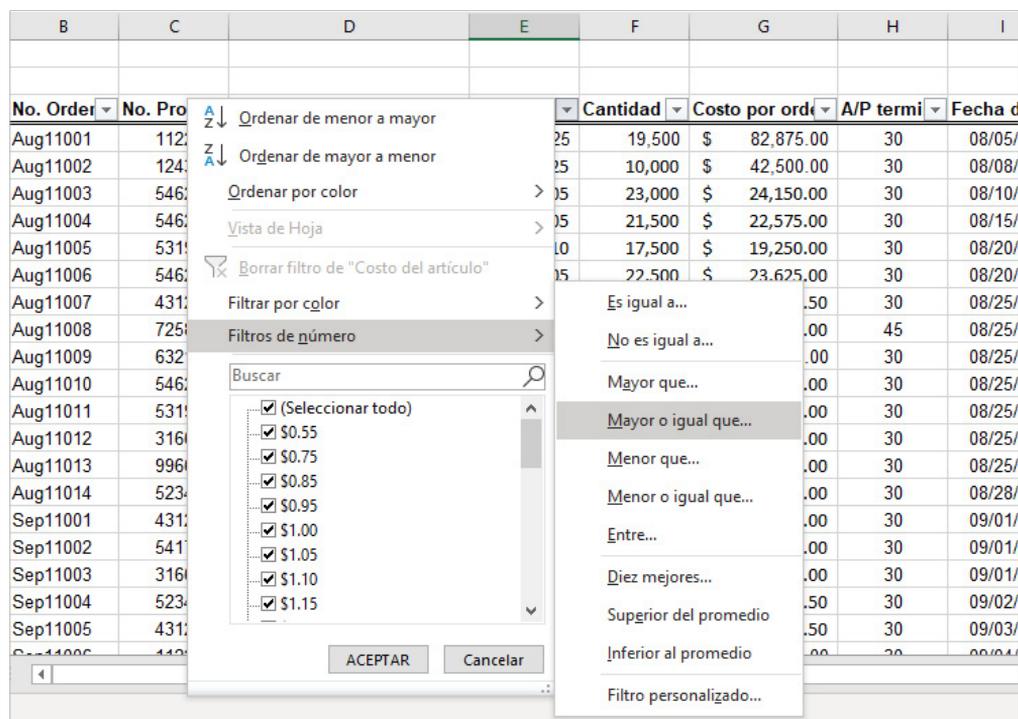


Figura 44. Opciones de menú de filtro. Fuente: Excel®

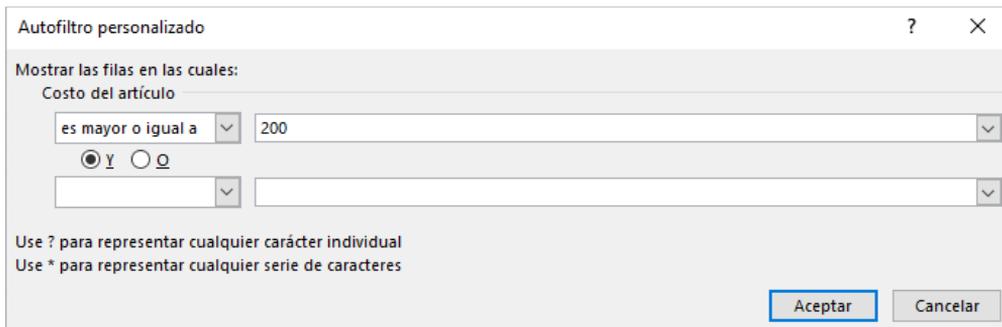


Figura 45. Opciones de menú de Autofiltro personalizado

| | A | B | C | D | E | F | G | H | I | J |
|----|-------------------|-----------|-----------|--------------------------|-----------|----------|---------------|-----------|----------|----------|
| 1 | Órdenes de compra | | | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | Proveedor | No. Order | No. Produ | Descripción del artículo | Costo del | Cantidad | Costo por ord | A/P termi | Fecha de | Fecha de |
| 19 | Láminas Alum | Sep11002 | 5417 | Control Panel | \$ 255.00 | 406 | \$ 103,530.00 | 30 | 09/01/11 | 09/10/11 |
| 65 | Productos Durable | Oct11014 | 5454 | Control Panel | \$ 220.00 | 550 | \$ 121,000.00 | 45 | 10/09/11 | 10/14/11 |
| 68 | Steelpin Co. | Oct11017 | 8008 | Valvula mecanizada | \$ 645.00 | 100 | \$ 64,500.00 | 30 | 10/10/11 | 10/21/11 |
| 70 | Valvulas Manley | Oct11019 | 8148 | Valvula mecanizada | \$ 655.50 | 125 | \$ 81,937.50 | 30 | 10/10/11 | 10/17/11 |
| 74 | Productos Durable | Oct11023 | 5454 | Control Panel | \$ 220.00 | 500 | \$ 110,000.00 | 45 | 10/15/11 | 10/20/11 |
| 76 | Steelpin Co. | Oct11025 | 8008 | Valvula mecanizada | \$ 645.00 | 150 | \$ 96,750.00 | 30 | 10/15/11 | 10/26/11 |
| 77 | Láminas Alum | Oct11026 | 5417 | Control Panel | \$ 255.00 | 500 | \$ 127,500.00 | 30 | 10/20/11 | 10/27/11 |
| 85 | Steelpin Co. | Oct11034 | 8008 | Valvula mecanizada | \$ 645.00 | 120 | \$ 77,400.00 | 30 | 10/28/11 | 11/04/11 |

Figura 46. Visualización de resultado de filtro

AutoFiltro crea criterios de filtrado basados en el tipo de datos que se filtran. Por ejemplo, en la figura 44 se observa que la lista de menús Filtros numéricos incluye criterios numéricos como “igual”, “no es igual”, etc. Si se elige filtrar por Fecha de pedido o Fecha de llegada, las herramientas de Autofiltro mostrarán una lista de menús de Filtros de fecha diferente para filtrar, que incluye “mañana”, “próxima semana”, “año hasta la fecha”, etc. El Autofiltro se puede utilizar secuencialmente para “profundizar” en los datos. Por ejemplo, después de filtrar los resultados por Paquete de tuerca, se podría filtrar por fecha de pedido y seleccionar todos los pedidos procesados en septiembre.

4.6 Transformar los datos

Las fórmulas en Excel® utilizan operadores matemáticos comunes: adición (+), resta (-); multiplicación (*); división (/); la exponenciación utiliza el símbolo de la clase; por ejemplo, 25 se escribe como 2^5 en una fórmula

de Excel®. Las referencias de celda en fórmulas se pueden escribir con direcciones relativas o direcciones absolutas. Una dirección relativa utiliza solo la etiqueta de fila y columna en la referencia de celda (por ejemplo, A4 o C21). Una dirección absoluta utiliza un signo de dólar (\$) antes de la etiqueta de fila o columna o ambos (por ejemplo: \$A2, C\$21 o \$B\$15).

Cuando se utiliza el direccionamiento relativo, al copiar una fórmula a otra celda se cambia la celda a las referencias por el número de filas o columnas en la dirección en la que se copia la fórmula. Por lo tanto, por ejemplo, si usamos una fórmula en la celda B8, $B4-B5*A8$, y la copiamos a la celda C9 (una columna a la derecha y una fila hacia abajo), todas las referencias de celda se incrementan en una y la fórmula se cambiaría a $=C5-C6*B9$.

El uso de un signo \$ antes de una etiqueta de fila (por ejemplo, B\$4) mantiene la referencia fija a la fila 4, pero permite que la referencia de columna cambie si la fórmula se copia en otra celda. De forma similar, el uso de un signo \$ antes de una etiqueta de columna (por ejemplo, \$B4) mantiene fija la referencia a la columna B, pero permite que la referencia de fila cambie. Por último, el uso de un signo \$ antes de las etiquetas de fila y columna (por ejemplo, \$B\$4) mantiene la referencia a la celda B4 fija sin importar dónde se copie la fórmula. Se debe tener mucho cuidado al usar el direccionamiento relativo y absoluto adecuadamente en los modelos, especialmente al copiar fórmulas.

Ejemplo 6. Implementación de modelos de precio-demanda en Excel®

A continuación, se muestran dos modelos para predecir la demanda en función del precio:

$$D = a - bP \quad \text{y} \quad D = cP^{-d}$$

La figura 47 muestra una hoja de cálculo (modelos de predicción de demanda de archivo Excel®) para calcular la demanda de precios diferentes utilizando cada uno de estos modelos. Por ejemplo, para calcular la demanda en la celda B8 para el modelo lineal, se utiliza la fórmula

$$= \$B\$4 - \$B\$5 * A8$$

Para calcular la demanda en la celda E8 para el modelo no lineal, se emplea la fórmula

$$= \$E\$4 * D8 ^ - \$E\$5$$

Observe cómo se utilizan las direcciones absolutas para que, a medida que se copian estas fórmulas, la demanda se calcule correctamente.

| | A | B | C | D | E |
|----|----------------------------------|---------|---|------------------|------------|
| 1 | Modelos de predicción de demanda | | | | |
| 2 | | | | | |
| 3 | Modelo lineal | | | Modelo no lineal | |
| 4 | a | 20000 | | c | 20000 |
| 5 | b | 10 | | d | 0.011132 |
| 6 | | | | | |
| 7 | Precio | Demanda | | Precio | Demanda |
| 8 | 80 | 19200 | | 70 | 19076.134 |
| 9 | 90 | 19100 | | 80 | 19047.7989 |
| 10 | 100 | 19000 | | 90 | 19022.8406 |
| 11 | 110 | 18900 | | 100 | 19000.5423 |
| 12 | 120 | 18800 | | 110 | 18980.3935 |
| 13 | | | | 120 | 18962.0178 |
| 14 | | | | 130 | 18945.1295 |

Figura 47. Modelos de Excel® para la predicción de la demanda

Las funciones se utilizan para realizar cálculos especiales en celdas y se utilizan ampliamente en aplicaciones de análisis empresarial. Todas las funciones de Excel® requieren un signo igual y un nombre de función seguido de paréntesis, en los que se especifican argumentos para la función.

Funciones básicas de Excel®

Algunas de las funciones más comunes que utilizaremos en las aplicaciones se muestran en la tabla 10:

Tabla 10. Funciones de Excel®

| Nombre función en español | Nombre función en inglés | Descripción |
|---|---|--|
| MIN(rango) | MIN(rango) | Encuentra el valor más pequeño en un rango de celdas |
| MAX(rango) | MAX(rango) | Encuentra el valor más grande en un rango de celdas |
| SUMA(rango) | SUM(rango) | Encuentra la suma de valores en un rango de celdas |
| SUBTOTALES(9,rango) | | Encuentra la suma de valores en un rango de celdas con filtro |
| PROMEDIO(rango) | AVERAGE(rango) | Busca el promedio de los valores en un rango de celdas |
| CONTAR(rango) | COUNT(rango) | Busca el número de celdas en un rango que contienen números |
| CONTARA(rango) | COUNTA(rango) | Busca el número de celdas en un rango que contienen valores alfanuméricos |
| CONTAR.BLANCO(rango) | COUNTBLANK(rango) | Cuenta el número de celdas en blanco en un rango |
| =CONTAR.SI(rango, criterio) | COUNTIF(rango, criterios) | Busca el número de celdas dentro de un rango que cumplen un criterio especificado. |
| =CONTAR.SI.CONJUNTO(rango1, criterio1, rango2, criterio2) | COUNTIFS(range1, criterion1, range2, criterion2,... range_n, criterion_n) | Encuentra el número de celdas dentro de varios rangos que cumplen criterios específicos para cada rango. |

La función CONTAR.SI cuenta el número de celdas dentro de un rango que cumplen un criterio que se especifique. Por ejemplo, puede contar todas las celdas que comienzan con una letra determinada, o puede contar todas las celdas que contienen un número mayor o menor que un número que se especifique. Ejemplos de criterios son 100, ">100", una referencia de celda como A4, una cadena de texto como "libro." Tenga en cuenta que el texto y las fórmulas lógicas deben incluirse entre comillas.

Ejemplo 7. Uso de funciones básicas de Excel®

En el conjunto de datos de órdenes de compra, se encuentran los resultados para calcular la cantidad más pequeña y grande de cualquier artículo pedido. Utilizamos las funciones MIN y MAX para los datos de la columna F. (Se emplea la función en el menú Vista, Ventana, dividir para ver la lista de datos y sus cálculos).

Por lo tanto, la fórmula en la celda B99 es =MIN(F4: F97) y la fórmula en la celda B100 es =MAX(F4: F97).

Para encontrar los costos totales de la orden, se suman los datos de la columna G utilizando la función =SUMA(G4:G97); esta es la fórmula en la celda B101.

Para encontrar el número medio de meses de Proveedores, se utiliza la función PROMEDIO para los datos de la columna H. La fórmula en la celda B102 es =PROMEDIO(H4: H97).

Para encontrar el número de pedidos de compra realizados, utilice la función CONTAR. Tenga en cuenta que la función CONTAR cuenta solo el número de celdas de un rango que contienen números, por lo que no se puede usar en las columnas A, B o D que contienen texto. Si se usan los elementos en la columna C, la fórmula en la celda B103 es CONTAR(C4: C97).

Para encontrar el número de pedidos realizados para el producto “Empaque oring”, utilizamos la función CONTAR.SI. Para este ejemplo, la fórmula utilizada en la celda B104 es =CONTAR.SI(D4:D97,”Empaque oring”). También se podría haber utilizado la referencia de celda para cualquier celda que contenga el texto Empaque oring, como por ejemplo, =CONTAR.SI (D4:D97,D12).

Para encontrar el número de pedidos con términos de A/P inferiores a 30 meses, utilice la fórmula =CONTAR.SI(H4:H97,”<30”) en la celda B105. Por último, para el número de pedidos con costo de artículo menor a 100 y términos A/O menor e igual a 30, se emplea =CONTAR.SI.CONJUNTO(H4:H97,”<=30”, E4:E97,”<=100”)

| | A | B | C | D | E | F | G | |
|-----|---|------------------|-----------------|--|----------------|---------------|---------------------|------------|
| 1 | Ordenes de compra | | | | | | | |
| 2 | | | | | | | | |
| 3 | Proveedor | No. Orden | No. Pr | Descripción del a | Costo d | Cantid | Costo por or | A/P |
| 4 | Sujetadores Hulkey | Aug11001 | 1122 | Fijaciones de fuselaje | \$ 4.25 | 19,500 | \$ 82,875.00 | |
| 5 | Láminas Alum | Aug11002 | 1243 | Fijaciones de fuselaje | \$ 4.25 | 10,000 | \$ 42,500.00 | |
| 6 | Aeroespacial Fast-Tie | Aug11003 | 5462 | Cable blindado | \$ 1.05 | 23,000 | \$ 24,150.00 | |
| 7 | Aeroespacial Fast-Tie | Aug11004 | 5462 | Cable blindado | \$ 1.05 | 21,500 | \$ 22,575.00 | |
| 8 | Steelpin Co. | Aug11005 | 5319 | Cable blindado | \$ 1.10 | 17,500 | \$ 19,250.00 | |
| 9 | Aeroespacial Fast-Tie | Aug11006 | 5462 | Cable blindado | \$ 1.05 | 22,500 | \$ 23,625.00 | |
| 10 | Steelpin Co. | Aug11007 | 4312 | Paquete de tuerca | \$ 3.75 | 4,250 | \$ 15,937.50 | |
| 11 | Productos Durable | Aug11008 | 7258 | Manómetro | \$ 90.00 | 100 | \$ 9,000.00 | |
| 12 | Aeroespacial Fast-Tie | Aug11009 | 6321 | Empaque oring | \$ 2.45 | 1,300 | \$ 3,185.00 | |
| 13 | Aeroespacial Fast-Tie | Aug11010 | 5462 | Cable blindado | \$ 1.05 | 22,500 | \$ 23,625.00 | |
| 14 | Steelpin Co. | Aug11011 | 5319 | Cable blindado | \$ 1.10 | 18,100 | \$ 19,910.00 | |
| 15 | | | | | | | | |
| 98 | | | | | | | | |
| 99 | Cantidad más pequeña de cualquier artículo ordenado | | 90 | =MIN(F4:F97) | | | | |
| 100 | Cantidad más grande de cualquier artículo ordenado | | 25,000 | =MAX(F4:F97) | | | | |
| 101 | Costos totales de la orden | | \$ 2,471,760.00 | =SUMA(G4:G97) | | | | |
| 102 | Cantidad promedio de meses por orden para cuentas por pagar | | 30.63829787 | =PROMEDIO(H4:H97) | | | | |
| 103 | Cantidad de órdenes de compra realizadas | | 94 | =CONTAR(C4:C97) | | | | |
| 104 | Número de pedidos realizados para Empaque oring | | 12 | =CONTAR.SI(D4:D97,"Empaque oring") | | | | |
| 105 | Número de pedidos con plazos A / P inferiores a 30 meses | | 17 | =CONTAR.SI(H4:H97,"<30") | | | | |
| 106 | Número de pedidos con costo de artículo menor a 100 y terminos A/O menor e igual a 30 | | 64 | =CONTAR.SI.CONJUNTO(H4:H97,"<=30", E4:E97,"<=100") | | | | |

Figura 48. Uso de funciones de Excel®

Las funciones de tipo SI también están disponibles para otros cálculos. Por ejemplo, las funciones SUMAR.SI, PROMEDIO.SI, SUMA.SI.CONJUNTO y PROMEDIO.SI.CONJUNTO se pueden utilizar para incrustar lógica condicional SI dentro de funciones matemáticas.

Excel® tiene una amplia variedad de otras funciones para aplicaciones estadísticas, financieras y otras. Algunos modelos financieros requieren el cálculo del valor actual neto (VPN) de una inversión o flujo de efectivo neto. El valor actual neto mide el valor de un conjunto de flujos de efectivo, teniendo en cuenta el valor del dinero a través del tiempo, donde i es la tasa de descuento. La tasa de descuento refleja los costos de oportunidad de gastar fondos ahora, en comparación con lograr un retorno a través de otra inversión, así como los riesgos asociados con no recibir rendimientos hasta un momento posterior. La suma del valor actual de todos los flujos de efectivo a lo largo de un horizonte de tiempo declarado es el valor actual neto:

$$\text{VPN} = \sum_{t=1}^n \frac{F_t}{(1+i)^t}$$

donde F_t es el flujo de efectivo en el período t . Un VPN positivo significa que la inversión proporcionará valor agregado porque el rendimiento proyectado supera la tasa de descuento.

La función de Excel® `VNA(tasa, valor1, valor2, ...)` calcula el valor presente neto de una inversión utilizando una tasa de descuento y una serie de pagos futuros (valores negativos) e ingresos (valores positivos). La tarifa es el valor de la tasa de descuento i sobre la longitud de un período, y `valor1`, `valor2`, ... son 1 a 29 argumentos que representan los pagos y los ingresos de cada período. Los valores deben estar igualmente espaciados en el tiempo y se supone que se producen al final de cada período. La inversión VPN comienza un período antes de la fecha del flujo de caja y termina con el último flujo de efectivo en la lista. El cálculo del VPN se basa en flujos de efectivo futuros. Si el primer flujo de caja (como una inversión inicial o un costo fijo) se produce al principio del primer período, debe añadirse al resultado del VPN y no incluirse en los argumentos de función.

Ejemplo 8. Uso de la función VNA

Una empresa está introduciendo un nuevo producto. El costo fijo para marketing y distribución es de \$25,000 y se incurre justo antes del lanzamiento. Los ingresos netos de ventas previstos para los primeros seis meses se muestran en la figura 49. La fórmula de la celda B8 calcula el valor actual neto de estos flujos de efectivo como $VNA(B6, C4: H4)$ y B5. Tenga en cuenta que el costo inicial no es un flujo de caja futuro y no se incluye en los argumentos de la función VNA.

| | | =B2+VNA(B3,B5:G5) | | | | | |
|---|---------------------|-------------------|-------------|-------------|-------------|-------------|-------------|
| | A | B | C | D | E | F | G |
| 1 | Valor presente neto | | | | | | |
| 2 | Costo inicial | \$ 25,000.00 | | | | | |
| 3 | Tasa de desc | 0.03 | | | | | |
| 4 | Mes | Enero | Febrero | Marzo | Abril | Mayo | Junio |
| 5 | Pronósticos | \$ 2,500.00 | \$ 4,000.00 | \$ 5,000.00 | \$ 8,000.00 | \$10,000.00 | \$12,500.00 |
| 6 | | | | | | | |
| 7 | Valor preser | \$11,975.81 | | | | | |

Figura 49. Cálculo del valor actual neto

Función de inserción. La forma más fácil de localizar una función en particular es seleccionar una celda y hacer clic en el botón insertar función $[f_x]$, que se encuentra debajo de la cinta de opciones junto a la barra de fórmulas y también en el grupo Biblioteca de funciones en la pestaña Fórmulas. Puede escribir una descripción en el campo de búsqueda, como “valor presente neto”, o seleccionar una categoría, como “Financiero”, en el cuadro desplegable.

Esta característica es especialmente útil si se sabe qué función utilizar, pero no se está seguro de qué argumentos introducir, ya que le guiará en la introducción de los datos adecuados para los argumentos de función. La figura 50 muestra el cuadro de diálogo desde el que se puede seleccionar la función que se desea utilizar. Por ejemplo, si se elige la función CONTAR.SI, aparece el cuadro de diálogo de la figura 51. Al hacer clic en una

celda de entrada, se muestra una descripción del argumento. Por lo tanto, si no se está seguro de qué introducir para el rango, la explicación de la figura 51 le ayudará. Para obtener más información, se puede hacer clic en el botón Ayuda en la esquina inferior izquierda.

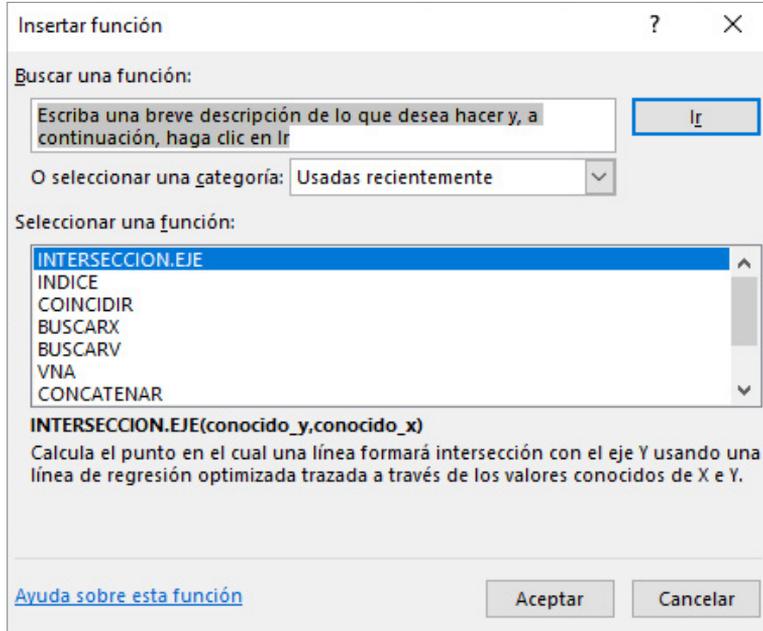


Figura 50. Insertar diálogo de función

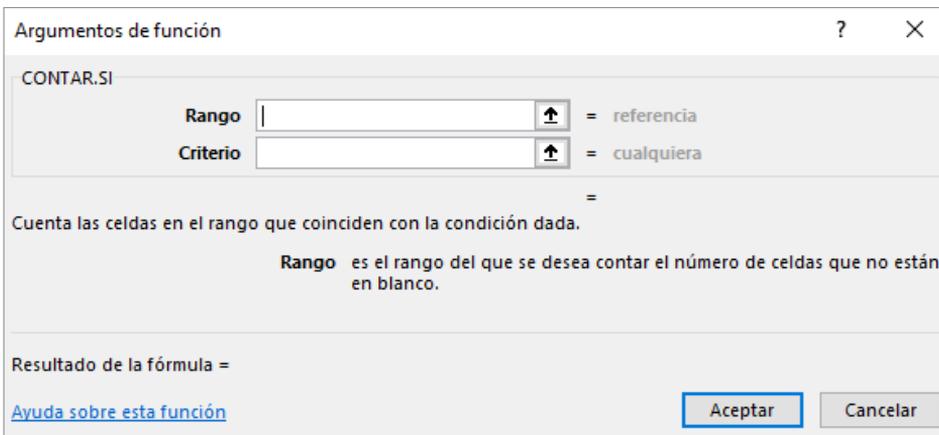


Figura 51. Venta de menú función “CONTAR.SI”

Funciones lógicas

Las funciones lógicas devuelven solo uno de dos valores: VERDADERO o FALSO. Tres funciones lógicas útiles en las aplicaciones de análisis empresarial son:

SI(condición, valor si es verdadero, valor si es falso): una función lógica que devuelve un valor si la condición es verdadera y otra si la condición es falsa. Esta función permite elegir uno de los dos valores para entrar en una celda. Si la condición especificada es verdadera, el valor verdadero se colocará en la celda. Si la condición es falsa, el valor que se introducirá será el correspondiente al falso. El valor falso y verdadero pueden ser un número o una cadena de texto entre comillas. Si se utiliza un espacio en blanco o entre comillas (" "), entonces el resultado será simplemente una celda en blanco. Esto suele ser útil para crear una hoja de cálculo limpia.

Por ejemplo, si la celda C2 contiene la función =SI(A8=2,7,12). Indica que, si el valor de la celda A8 es igual al número 2, se asignará el valor del número 7 a la celda C2; si el valor de la celda A8 no es igual al número 2, se asignará el valor del número 12 a la celda C2.

Y(condición 1, condición 2...) -una función lógica que devuelve VERDADERO si todas las condiciones son verdaderas y FALSO si no.

O(condición 1, condición 2...) -una función lógica que devuelve VERDADERO si alguna condición es verdadera y FALSO si no.

Las condiciones pueden incluir lo siguiente: "=" igual a, ">" mayor que, "<" menos de, ">=" mayor o igual que, "<=" menor o igual que, "<>" diferente a. Puede "anidar" hasta siete funciones SI, reemplazando valor_si_verdadero o valor_si_falso en una función SI por otra función SI:

=SI(A8=2,(SI(B3=5,"Si","")),15)

Esto significa que, si la celda A8 es igual a 2, entonces verifique el contenido de la celda B3. Si la celda B3 es 5, el valor de la función es la cadena de texto “Si”; si no, es un espacio en blanco (representado por comillas sin nada en el medio). Sin embargo, si la celda A8 no es 2, el valor de la función es 15 sin importar qué celda sea B3.

Las funciones “Y” y “O” simplemente devuelven los valores de verdadero o falso si se cumplen todas o al menos una de varias condiciones, respectivamente. Se puede utilizar las funciones “Y” y “O” como condición dentro de una función SI; por ejemplo, =SI(Y(B1=3, C1=5),12,22). Aquí, si la celda B1 es igual al número 3 y la celda C1 es igual al número 5, entonces el valor de la función es 12, de lo contrario es 22.

Ejemplo 9. Uso de la función SI

Supongamos que dentro de la base de datos Ordenes de compra se considera que cualquier orden de 10,000 unidades o más es grande, mientras que cualquier otro tamaño de pedido se considera pequeño. Podemos usar la función SI para clasificar los órdenes.

En primer lugar, cree una nueva columna en la hoja de cálculo para realizar la clasificación por el tamaño del pedido, por ejemplo, la columna K. En la celda K4, use la fórmula =SI(F4>10000, “Grande”, “Pequeño”).

Esta función devolverá el valor “Grande” en la celda K4 si el tamaño de la orden en la celda F4 es 10,000 o más; de lo contrario, devuelve el valor “Pequeño”. Además, supongamos que los pedidos grandes con un costo total de al menos \$25,000 se consideran críticos. Podemos marcar estas órdenes como críticas mediante el uso de la función en la celda L4: =SI(Y(K4=“Grande”, G4>=25000), “Crítica”, “ ”)

Después de copiar estas fórmulas en las columnas, en la figura 52 se muestra una parte de los resultados.

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|-------------------|------------------|--------------------|--------------------|---------------------|-----------------|------------------------|--------------------|--------------------|--------------------|-------------------|-------------|
| 1 | Ordenes de compra | | | | | | | | | | | |
| 2 | | | | | | | | | | | | |
| 3 | Proveedor | No. Orden | No. Product | Descripción | Costo del ar | Cantidad | Costo por orden | A/P termino | Fecha de or | Fecha de re | Tamaño ord | Tipo |
| 4 | Sujetadores | Aug11001 | 1122 | Fijaciones de | \$ 4.25 | 19,500 | \$ 82,875.00 | 30 | 08/05/11 | 08/13/11 | Grande | Critica |
| 5 | Láminas Alur | Aug11002 | 1243 | Fijaciones de | \$ 4.25 | 10,000 | \$ 42,500.00 | 30 | 08/08/11 | 08/14/11 | Grande | Critica |
| 6 | Aeroespacial | Aug11003 | 5462 | Cable blindad | \$ 1.05 | 23,000 | \$ 24,150.00 | 30 | 08/10/11 | 08/15/11 | Grande | |
| 7 | Aeroespacial | Aug11004 | 5462 | Cable blindad | \$ 1.05 | 21,500 | \$ 22,575.00 | 30 | 08/15/11 | 08/22/11 | Grande | |
| 8 | Steelpin Co. | Aug11005 | 5319 | Cable blindad | \$ 1.10 | 17,500 | \$ 19,250.00 | 30 | 08/20/11 | 08/31/11 | Grande | |
| 9 | Aeroespacial | Aug11006 | 5462 | Cable blindad | \$ 1.05 | 22,500 | \$ 23,625.00 | 30 | 08/20/11 | 08/26/11 | Grande | |
| 10 | Steelpin Co. | Aug11007 | 4342 | Resorte de | \$ 3.75 | 4,350 | \$ 16,337.50 | 30 | 08/25/11 | 09/04/11 | Pequeño | |

Figura 52. Clasificación de tamaños de orden mediante la función SI

4.7 Visualizar los datos

Microsoft Excel® proporciona gráficos con muchas características. Con un poco de experimentación, se pueden crear gráficos muy profesionales para análisis de negocios y presentaciones. Estos incluyen gráficos de barras verticales y horizontales, gráficos de líneas, gráficos circulares, gráficos de área, gráficos de dispersión y muchos otros tipos especiales de gráficos.

Ciertos gráficos funcionan mejor para ciertos tipos de datos, y el uso del gráfico incorrecto puede dificultar su uso e interpretación. Se recomienda que mantenga los gráficos simples para cumplir con lo anterior.

Para crear un gráfico en Excel®, es mejor resaltar primero el rango de los datos que desea trazar. Los archivos de Ayuda de Excel® proporcionan instrucciones sobre cómo dar formato a los datos para un tipo determinado de gráfico. Se hace clic en la pestaña Insertar del menú de Excel® (véase la figura 53). En el grupo Gráficos, se hace clic en el tipo de gráfico y, a continuación, en un subtipo de gráfico que se desee utilizar. Una vez creado un gráfico básico, se pueden utilizar las opciones de las pestañas Diseño y Formato de las pestañas Herramientas de gráfico para personalizar el gráfico (figura 54). En la pestaña Diseño, se puede cambiar el tipo de gráfico, los datos incluidos en el gráfico, el diseño del gráfico y los estilos. La pestaña Formato proporciona varias opciones de formato. También se pueden personalizar gráficos fácilmente al hacer clic con el botón derecho en

los elementos del gráfico o utilizando las opciones de Diseño rápido en el grupo Diseño de gráfico dentro de la pestaña Diseño de herramientas de gráfico.

Los colores predeterminados de Excel® usan rojo, se sugieren usar paletas azul-naranja para aplicaciones profesionales y comerciales.

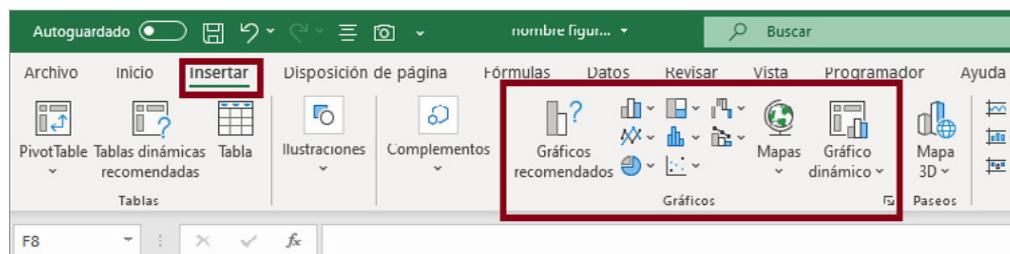


Figura 53. Pestaña Insertar Excel®



Figura 54. Herramientas de gráficos de Excel®

Excel® distingue entre gráficos de barras verticales y horizontales, los gráficos de barras verticales se conocen como gráficos de columnas y los gráficos de barras horizontales como gráficos de barras. Un gráfico de columnas agrupadas compara los valores entre categorías mediante rectángulos verticales; un gráfico de columnas apiladas muestra la contribución de cada valor al total apilando los rectángulos; y un gráfico de columnas apiladas al 100% compara el porcentaje que cada valor contribuye a un total. Los gráficos de columnas y barras son útiles para comparar datos categóricos u ordinales, para ilustrar diferencias entre conjuntos de valores y para mostrar proporciones o porcentajes de un todo.



Ejemplo 10. Creación de gráficos de columnas

El archivo de Excel® de la Tasa de desocupación parcial y desocupación nacional proporciona datos sobre la tasa de desocupación por género (véase la figura 55). Vamos a construir un gráfico de columnas simple para el horizonte de tiempo. En primer lugar, se resalta el rango C6:E21, que incluye los datos de cada categoría. Haga clic en el botón Gráfico de columnas y, a continuación, seleccione el primer tipo de gráfico de la lista (un gráfico de columnas agrupadas). Para agregar un título, haga clic en el botón Agregar elementos de gráfico de la cinta de opciones de la ficha Diseño. Después, haga clic en “Título del gráfico” en el gráfico y se cambia a “Tasa de desocupación parcial y desocupación nacional trimestral”. Los nombres de la serie de datos se pueden cambiar haciendo clic en el botón Seleccionar datos del grupo Datos de la pestaña Diseño. En el cuadro de diálogo Seleccionar origen de datos (consulte la figura 56), haga clic en “Series1” y, a continuación, en el botón Editar. Se introducen las celdas A6:C21. También se puede cambiar el orden en el que se muestran las series de datos en el gráfico utilizando los botones arriba y abajo. El gráfico final se muestra en la figura 57.

Se debe tener cuidado al cambiar la escala del eje numérico. Las alturas o longitudes de las barras solo reflejan con precisión los valores de datos si el eje comienza en cero. Si no es así, los tamaños relativos pueden pintar una imagen engañosa de los valores relativos de los datos.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|--|--------------|----------------|---------------|----------|---------------------|----------------------|--------------------------|------------|----------|--------------------------|
| 1 | Informe de la Comisión de Igualdad de Oportunidades en el Empleo - Número de empleados en el estado de Alabama, 2006 | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | Grupo Racial/Etnico y | Empleo total | Administrativo | Profesionales | Técnicos | Trabajador de venta | Empleados de oficina | Trabajadores artesanales | Operativos | Empleado | Trabajadores de servicio |
| 4 | TODOS LOS | 632.329 | 60.258 | 80.733 | 39.868 | 62.019 | 67.014 | 61.322 | 120.810 | 68.752 | 71.553 |
| 5 | Hombres | 349.353 | 41.777 | 39.792 | 19.848 | 23.727 | 11.293 | 55.853 | 84.724 | 44.736 | 27.603 |
| 6 | Mujeres | 282.976 | 18.481 | 40.941 | 20.020 | 38.292 | 55.721 | 5.469 | 36.086 | 24.016 | 43.950 |
| 7 | | | | | | | | | | | |

Figura 55. Parte de los datos del informe sobre el empleo de la EEO

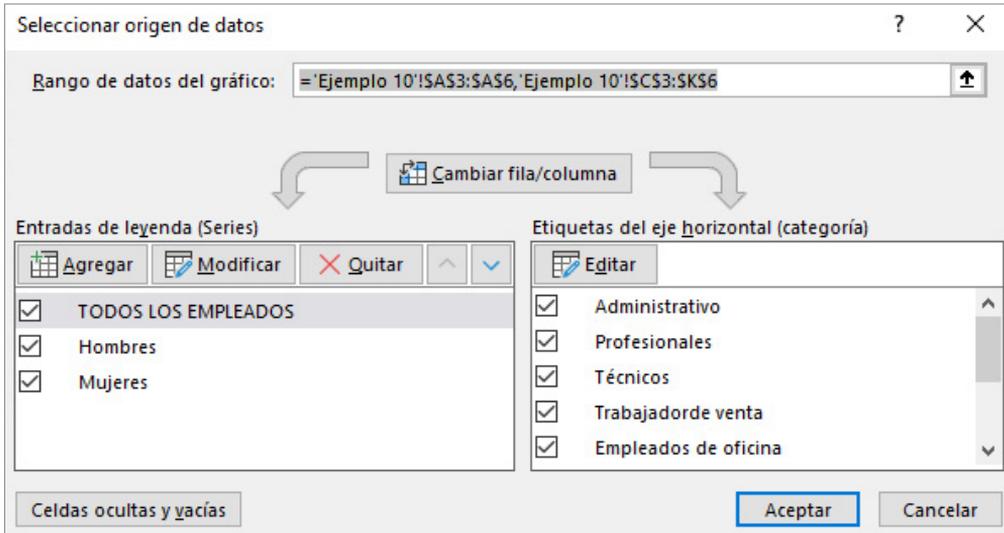


Figura 56. Seleccionar cuadro de diálogo. Origen de datos

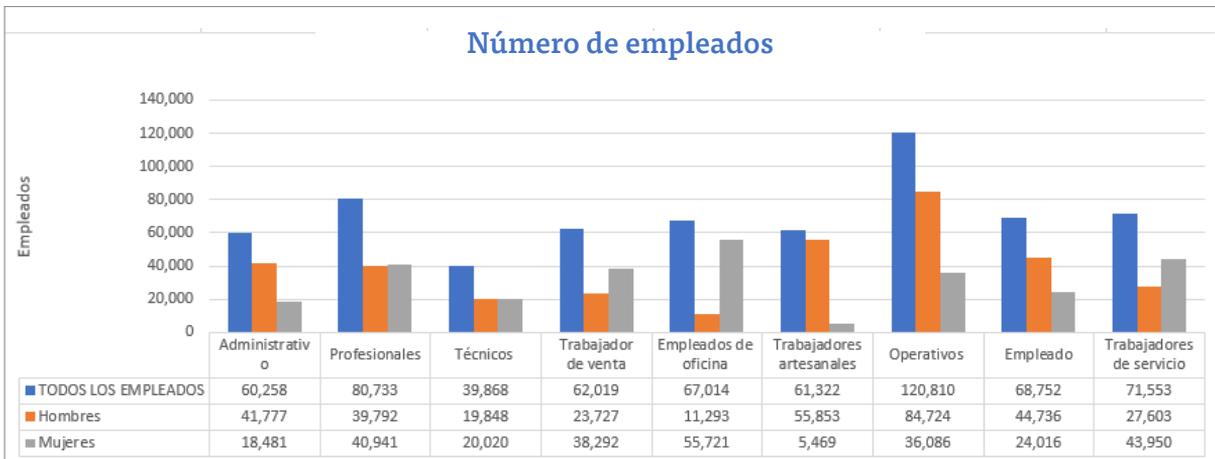


Figura 57. Formato de gráfico de columna alternativo.

Opciones del gráfico Etiquetas de datos y Tablas de datos

Excel® proporciona opciones para incluir los datos numéricos en los que se basan los gráficos dentro de los gráficos. Las etiquetas de datos se pueden agregar a los elementos del gráfico para mostrar el valor real de las barras, por ejemplo. También se pueden agregar tablas de datos; estos son generalmente mejores que las etiquetas de datos, que pueden llegar a ser bastante

desordenados. Ambos se pueden agregar desde el botón Agregar elemento de gráfico en la pestaña Diseño de herramientas de gráfico, o también desde el botón Diseño rápido, que proporciona opciones de diseño estándar.

Gráficos de líneas

Los gráficos de líneas proporcionan un medio útil para mostrar datos a lo largo del tiempo, como se muestra en el ejemplo 11. Se pueden trazar varias series de datos en gráficos de líneas; sin embargo, pueden ser difíciles de interpretar si la magnitud de los valores de datos difiere mucho. En ese caso, se aconseja crear gráficos separados para cada serie de datos.

Ejemplo 11. Gráfico de líneas

En la figura 58 se muestra un gráfico de líneas que da la cantidad de Exportaciones trimestrales de fabricación de equipo de transporte (SCIAN 336) en miles de dólares. En el gráfico se muestra claramente un aumento significativo y una reducción sustancial en el año 2020.



Figura 58. Gráfico de líneas

Gráficos circulares

Para muchos tipos de datos, es de interés comprender la proporción relativa de cada origen de datos con respecto al total. Un gráfico circular muestra esto particionando un círculo en áreas en forma de tarta que muestran la proporción relativa. En el ejemplo 12 se proporciona una aplicación.



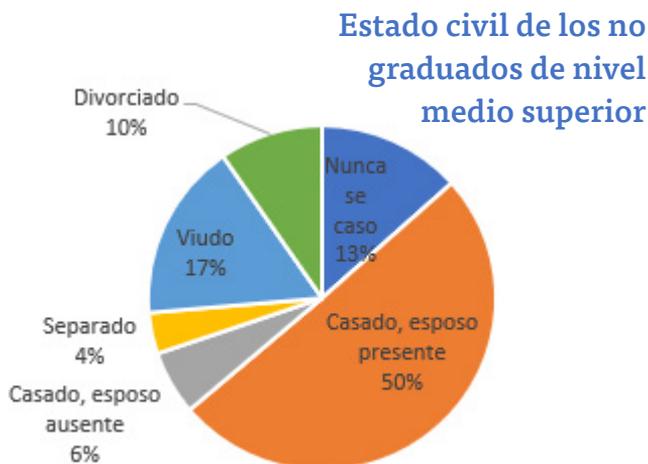
Ejemplo 12. Gráfico circular para los datos del censo

Considere la población por condiciones de pobreza de los Estados Unidos Mexicanos en el año 2018, como se muestra en el archivo, una parte de la cual se muestra en la figura 59. Para mostrar la proporción relativa se usa un gráfico circular, como se observa en la figura 60. Este gráfico utiliza una opción de diseño que muestra las etiquetas asociadas a los datos, así como las proporciones reales como porcentajes. También se puede elegir un diseño diferente que muestre los valores y/o las proporciones.

| | A | B | C | D | E | F | G |
|----|------------------------------------|---|---|---|----------------|---------------------|-------------------------|
| 1 | Censo de datos de educación | | | | | | |
| 2 | | No es graduado de nivel medio superior | Graduado de nivel medio superior | Estudios universitarios sin título | Pasante | Licenciatura | Grados avanzados |
| 3 | Total de personas | 29,620,292 | 58,207,318 | 29,620,292 | 12,915,825 | 28,242,604 | 13,604,669 |
| 19 | Nunca se caso | 4,120,320 | 7,777,104 | 4,789,872 | 1,828,392 | 5,124,648 | 2,137,416 |
| 20 | Casado, esposo presente | 15,516,160 | 36,382,720 | 18,084,352 | 8,346,624 | 19,154,432 | 9,523,712 |
| 21 | Casado, esposo ausente | 1,847,880 | 2,368,024 | 1,184,012 | 465,392 | 670,712 | 301,136 |
| 22 | Separado | 1,188,090 | 1,667,010 | 842,715 | 336,165 | 405,240 | 165,780 |
| 23 | Viudo | 5,145,683 | 4,670,488 | 1,765,010 | 556,657 | 977,544 | 475,195 |
| 24 | Divorciado | 2,968,680 | 7,003,040 | 3,806,000 | 1,674,640 | 2,340,690 | 1,217,920 |

Figura 59. Porción de los datos de educación del Censo

Figura 60. Gráfico circular para el estado civil



Los profesionales de visualización de datos recomiendan el uso de gráficos circulares a un número pequeño de categorías. Cuando los números a representar suman al cien por ciento de la población, se recomienda usar etiquetas para mostrar los nombres de grupo y los porcentajes reales.

Gráficos de área

Un gráfico de área combina las características de un gráfico circular con las de los gráficos de líneas. Los gráficos de área presentan más información que los gráficos circulares o de solo líneas, pero pueden saturar la mente del observador con demasiados detalles si se utilizan demasiadas series de datos; por lo tanto, deben ser utilizados con cuidado.

Ejemplo 13. Gráfico de área

En la figura 61 se muestran las exportaciones trimestrales por subsector de actividad total nacional y en color azul, en el estado de Chihuahua (en miles de dólares). De todos los estados que se dedican a la fabricación de equipo de transporte, Chihuahua es el que muestra mayor volumen de fabricación. En este gráfico se observa que el volumen de fabricación ha aumentado en el país, sin embargo, no precisamente lo ha llevado a cabo de manera semejante en el estado de Chihuahua.

Consumo total de energía vs. producción de combustibles fósiles

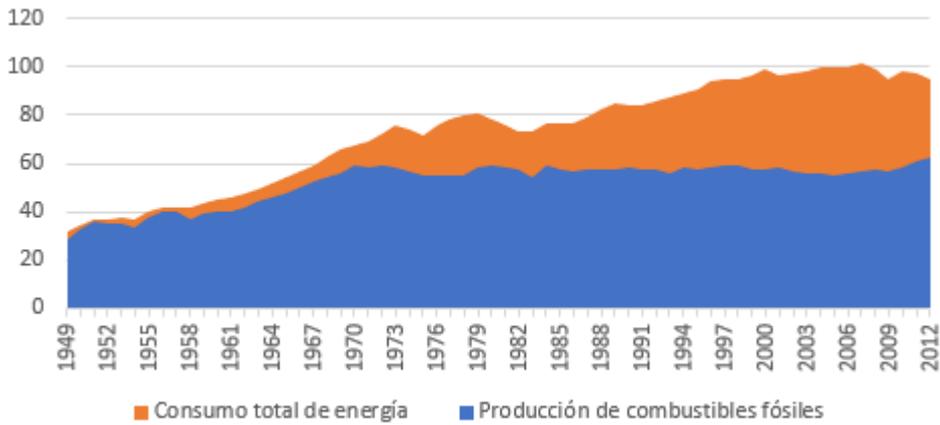


Figura 61. Gráfico de área para exportaciones

Gráfico de dispersión

Los gráficos de dispersión muestran la relación entre dos variables. Para construir un gráfico de dispersión, se necesitan observaciones que consistan en pares de variables. Por ejemplo, los alumnos de una clase podrían tener calificaciones tanto para un examen de mitad de período como para un examen final. Un gráfico de dispersión mostraría si las calificaciones altas o bajas en el semestre corresponden fuertemente a las calificaciones altas o bajas en el examen final, o si la relación es débil o inexistente.

Ejemplo 14. Gráfico de dispersión

En la figura 62 se muestra un gráfico de dispersión del tamaño de una casa (en metros cuadrados) frente al valor del mercado interno del archivo de Excel® Valor de mercado interno. Los datos sugieren claramente que los valores más altos del mercado están asociados con hogares más grandes.

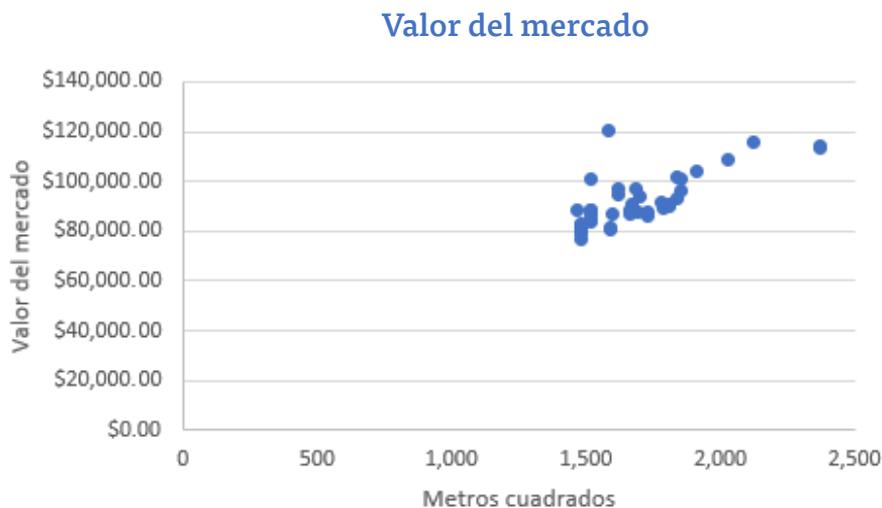


Figura 62. Gráfico de dispersión

Gráficos de burbujas

Un gráfico de burbujas es un tipo de gráfico de dispersión en el que el tamaño del marcador de datos corresponde al valor de una tercera variable; en consecuencia, es una forma de trazar tres variables en dos dimensiones.

Ejemplo 15. Gráfico de burbujas para comparar características de stock

En la figura 63 se muestra un gráfico de burbujas para mostrar el precio, la relación P/E (precio/ganancias) y la capitalización de mercado para el gráfico de cinco acciones diferentes en un día determinado en el archivo Comparaciones de acciones. La posición en el gráfico muestra el precio y la P/E, además, el tamaño de la burbuja representa la capitalización de mercado en miles de millones de dólares.

Comparación de acciones

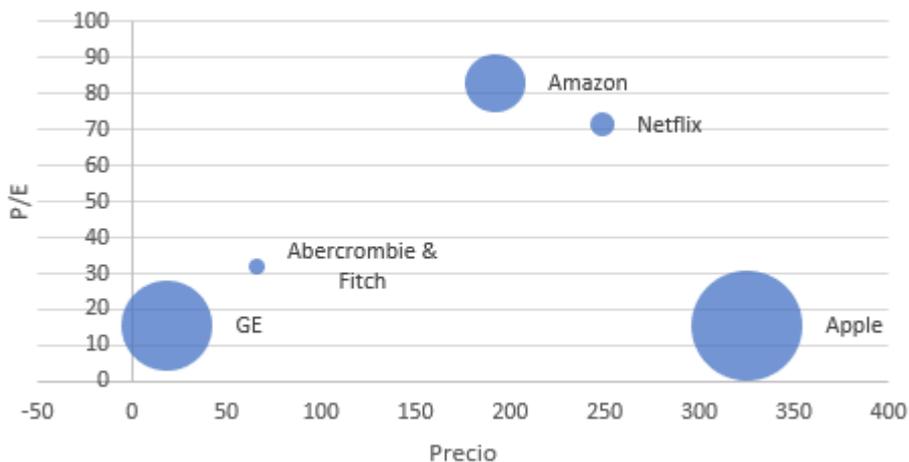


Figura 63. Gráfico de burbujas de Acciones

Barras de datos, escalas de color y conjuntos de iconos

Estas opciones forman parte de las reglas de formato condicional de Excel® que permiten visualizar diferentes valores numéricos mediante el uso de colores y símbolos. Excel® tiene una variedad de plantillas estándar para usar, pero también se pueden personalizar las reglas para que se cumplan las propias condiciones y estilos. Se sugiere experimentar con estas herramientas.

Ejemplo 16. Uso de formato condicional

Para la hoja de datos Ventas mensuales de productos, se resaltan los datos de cada columna, se hace clic en el botón Formato condicional del grupo Estilos dentro de la pestaña Inicio, además, se selecciona Barras de datos y se elige la opción de relleno y el color.

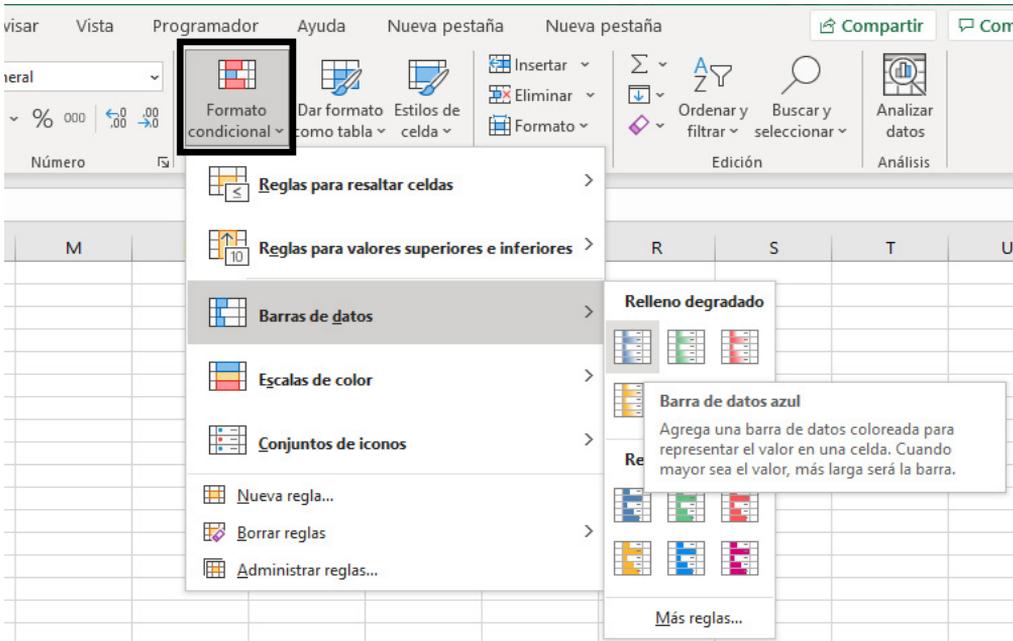


Figura 64. Menú para seleccionar el formato condicional en Excel®

| | A | B | C | D | E | F |
|----|--------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 1 | Ventas mensuales de productos | | | | | |
| 2 | | | | | | |
| 3 | Mes | Producto A | Producto B | Producto C | Producto D | Producto E |
| 4 | Enero | 7792 | 5554 | 3105 | 3168 | 10350 |
| 5 | Febrero | 7268 | 3024 | 3228 | 3751 | 8965 |
| 6 | Marzo | 7049 | 5543 | 2147 | 3319 | 6827 |
| 7 | Abril | 7560 | 5232 | 2636 | 4057 | 8544 |
| 8 | Mayo | 8233 | 5450 | 2726 | 3837 | 7535 |
| 9 | Junio | 8629 | 3943 | 2705 | 4664 | 9070 |
| 10 | Julio | 8702 | 5991 | 2891 | 5418 | 8389 |
| 11 | Agosto | 9215 | 3920 | 2782 | 4085 | 7367 |
| 12 | Septiembre | 8986 | 4753 | 2524 | 5575 | 5377 |
| 13 | Octubre | 8654 | 4746 | 3258 | 5333 | 7645 |
| 14 | Noviembre | 8315 | 3566 | 2144 | 4924 | 8173 |
| 15 | Diciembre | 7978 | 5670 | 3071 | 6563 | 6088 |

Figura 65. Ejemplo de barras de datos

Las escalas de color somborean las celdas en función de su valor numérico mediante una paleta de colores. Esta es otra opción en el menú Formato condicional. Por ejemplo, en la figura 66 se usa una escala de color verde-amarillo-rojo, que resalta las celdas que contienen valores grandes en verde, valores pequeños en rojo y valores medios en amarillo. Cuanto más oscuro sea el verde, mayor será el valor; cuanto más oscuro sea el rojo, menor será el

valor. Para los valores intermedios, se puede ver que los colores se mezclan. Esto proporciona una forma rápida de identificar los valores de ventas de productos y meses más grandes y pequeños. La codificación de color de los datos cuantitativos se denomina comúnmente un mapa de calor.

| | A | B | C | D | E | F |
|----|--------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 1 | Ventas mensuales de productos | | | | | |
| 2 | | | | | | |
| 3 | Mes | Producto A | Producto B | Producto C | Producto D | Producto E |
| 4 | Enero | 7792 | 5554 | 3105 | 3168 | 10350 |
| 5 | Febrero | 7268 | 3024 | 3228 | 3751 | 8965 |
| 6 | Marzo | 7049 | 5543 | 2147 | 3319 | 6827 |
| 7 | Abril | 7560 | 5232 | 2636 | 4057 | 8544 |
| 8 | Mayo | 8233 | 5450 | 2726 | 3837 | 7535 |
| 9 | Junio | 8629 | 3943 | 2705 | 4664 | 9070 |
| 10 | Julio | 8702 | 5991 | 2891 | 5418 | 8389 |
| 11 | Agosto | 9215 | 3920 | 2782 | 4085 | 7367 |
| 12 | Septiembre | 8986 | 4753 | 2524 | 5575 | 5377 |
| 13 | Octubre | 8654 | 4746 | 3258 | 5333 | 7645 |
| 14 | Noviembre | 8315 | 3566 | 2144 | 4924 | 8173 |
| 15 | Diciembre | 7978 | 5670 | 3071 | 6563 | 6088 |

Figura 66. Ejemplo de escalas de color

Por último, los conjuntos de iconos proporcionan información similar utilizando varios símbolos, como flechas o colores de semáforo. En la figura 67 se muestra un ejemplo.

| | A | B | C | D | E | F |
|----|--------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 1 | Ventas mensuales de productos | | | | | |
| 2 | | | | | | |
| 3 | Mes | Producto A | Producto B | Producto C | Producto D | Producto E |
| 4 | Enero | ↑ 7792 | ⇒ 5554 | ↓ 3105 | ↓ 3168 | ↑ 10350 |
| 5 | Febrero | ⇒ 7268 | ↓ 3024 | ↓ 3228 | ↓ 3751 | ↑ 8965 |
| 6 | Marzo | ⇒ 7049 | ⇒ 5543 | ↓ 2147 | ↓ 3319 | ⇒ 6827 |
| 7 | Abril | ⇒ 7560 | ⇒ 5232 | ↓ 2636 | ↓ 4057 | ↑ 8544 |
| 8 | Mayo | ↑ 8233 | ⇒ 5450 | ↓ 2726 | ↓ 3837 | ⇒ 7535 |
| 9 | Junio | ↑ 8629 | ↓ 3943 | ↓ 2705 | ↓ 4664 | ↑ 9070 |
| 10 | Julio | ↑ 8702 | ⇒ 5991 | ↓ 2891 | ⇒ 5418 | ↑ 8389 |
| 11 | Agosto | ↑ 9215 | ↓ 3920 | ↓ 2782 | ↓ 4085 | ⇒ 7367 |
| 12 | Septiembre | ↑ 8986 | ↓ 4753 | ↓ 2524 | ⇒ 5575 | ⇒ 5377 |
| 13 | Octubre | ↑ 8654 | ↓ 4746 | ↓ 3258 | ⇒ 5333 | ↑ 7645 |
| 14 | Noviembre | ↑ 8315 | ↓ 3566 | ↓ 2144 | ⇒ 4924 | ↑ 8173 |
| 15 | Diciembre | ↑ 7978 | ⇒ 5670 | ↓ 3071 | ⇒ 6563 | ⇒ 6088 |

Figura 67. Ejemplo de conjuntos de iconos

Minigráficos

Los minigráficos son gráficos que resumen una fila o columna de datos en una sola celda. Los destellos fueron introducidos por Edward Tufte, un famoso experto en presentación visual de datos, quien describió los minigráficos como “gráficos de datos intensos, simples de diseño y tamaño de palabra”. Excel® tiene tres tipos de minigráficos: línea, columna y ganar/perder. Los de línea son claramente útiles para los datos de series temporales, mientras que los de columna son más adecuados para los datos categóricos. Los minigráficos de pérdidas y ganancias son útiles para los datos que suben o bajan con el tiempo. Se encuentran en el grupo Sparklines en el menú Insertar de la cinta de opciones.

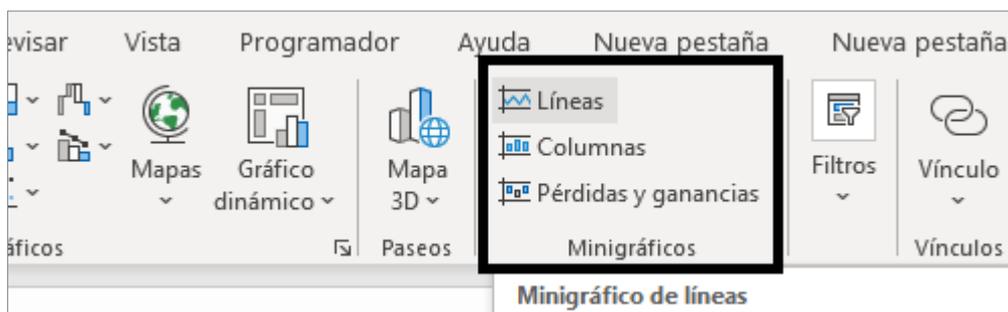


Figura 68. Menú para seleccionar Minigráficos en Excel®

Ejemplo 17. Ejemplos de Minigráficos

Volveremos a usar los datos mensuales de ventas de productos. En la figura 69 se muestran los minigráficos de línea en la fila 14 para cada producto. En la columna G se muestran los minigráficos de columna, que son esencialmente gráficos de columnas pequeños. Por lo general, se expanden los anchos de fila o columna para mostrarlos eficazmente. Se debe tener en cuenta, sin embargo, que las longitudes de las barras no se escalan correctamente a los datos; por ejemplo, en el primero, los productos D y E son aproximadamente un tercio del valor del producto E, pero las barras no se escalan correctamente. Se recomienda tener cuidado al utilizarlos.

| | A | B | C | D | E | F | G |
|----|--------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|---|
| 1 | Ventas mensuales de productos | | | | | | |
| 2 | | | | | | | |
| 3 | Mes | Producto A | Producto B | Producto C | Producto D | Producto E | |
| 4 | Enero | 7792 | 5554 | 3105 | 3168 | 10350 | |
| 5 | Febrero | 7268 | 3024 | 3228 | 3751 | 8965 | |
| 6 | Marzo | 7049 | 5543 | 2147 | 3319 | 6827 | |
| 7 | Abril | 7560 | 5232 | 2636 | 4057 | 8544 | |
| 8 | Mayo | 8233 | 5450 | 2726 | 3837 | 7535 | |
| 9 | Junio | 8629 | 3943 | 2705 | 4664 | 9070 | |
| 10 | Julio | 8702 | 5991 | 2891 | 5418 | 8389 | |
| 11 | Agosto | 9215 | 3920 | 2782 | 4085 | 7367 | |
| 12 | Septiembre | 8986 | 4753 | 2524 | 5575 | 5377 | |
| 13 | Octubre | 8654 | 4746 | 3258 | 5333 | 7645 | |
| 14 | Noviembre | 8315 | 3566 | 2144 | 4924 | 8173 | |
| 15 | Diciembre | 7978 | 5670 | 3071 | 6563 | 6088 | |
| 16 | | | | | | | |

Figura 69. Minigráficos de líneas y columnas

En la figura 70 se muestra una hoja de trabajo modificada en la que calculamos el cambio porcentual de un mes al siguiente para los productos A y B. Los minigráficos de pérdidas y ganancias de la fila 14 muestran que los patrones de ventas aumentan y disminuyen, lo que sugiere que el producto A tiene un patrón cíclico, mientras que el producto B cambió de una manera más aleatoria. Si se hace clic en cualquier celda que contenga un minigráfico, aparecerá la pestaña Diseño de herramientas de Sparkline, que permite personalizar colores y otras opciones.

| | A | B | C | D | E | F | G | H | I | J | K |
|----|--------------------------------------|-------------------|--------------------------|-------------------|--------------------------|-------------------|--------------------------|-------------------|--------------------------|-------------------|--------------------------|
| 1 | Ventas mensuales de productos | | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | Mes | Producto A | Cambio porcentual | Producto B | Cambio porcentual | Producto C | Cambio porcentual | Producto D | Cambio porcentual | Producto E | Cambio porcentual |
| 4 | Enero | 7792 | | 5554 | | 3105 | | 3168 | | 10350 | |
| 5 | Febrero | 7268 | -6.72% | 3024 | -45.55% | 3228 | 3.96% | 3751 | 18.40% | 8965 | -13.38% |
| 6 | Marzo | 7049 | -3.01% | 5543 | 83.30% | 2147 | -33.49% | 3319 | -11.52% | 6027 | -23.05% |
| 7 | Abril | 7560 | 7.25% | 5232 | -5.61% | 2636 | 22.78% | 4057 | 22.24% | 8544 | 25.15% |
| 8 | Mayo | 8233 | 8.90% | 5450 | 4.17% | 2726 | 3.41% | 3837 | -5.42% | 7535 | -11.81% |
| 9 | Junio | 8629 | 4.81% | 3943 | -27.65% | 2705 | -0.77% | 4664 | 21.55% | 9070 | 20.37% |
| 10 | Julio | 8702 | 0.85% | 5991 | 51.94% | 2891 | 6.88% | 5418 | 16.17% | 8389 | -7.51% |
| 11 | Agosto | 9215 | 5.90% | 3920 | -34.57% | 2782 | -3.77% | 4085 | -24.60% | 7367 | -12.18% |
| 12 | Septiembre | 8986 | -2.19% | 4753 | 21.25% | 2524 | -9.27% | 5575 | 36.17% | 5377 | -27.01% |
| 13 | Octubre | 8654 | -3.69% | 4746 | -0.15% | 3258 | 29.08% | 5333 | -4.34% | 7645 | 42.18% |
| 14 | Noviembre | 8315 | -3.92% | 3566 | -24.86% | 2144 | -34.19% | 4924 | -7.67% | 8173 | 6.91% |
| 15 | Diciembre | 7978 | -4.05% | 5670 | 59.00% | 3071 | 43.24% | 6563 | 33.29% | 6088 | -25.51% |
| 16 | | | | | | | | | | | |

Figura 70. Minigráficos de pérdidas y ganancias

Distribuciones de frecuencia para datos categóricos

Una distribución de frecuencia es una tabla que muestra el número de observaciones en cada uno de varios grupos no superpuestos. Las variables categóricas definen naturalmente los grupos en una distribución de frecuencia. Por ejemplo, en la base de datos pedidos de compra, se realizaron pedidos para los siguientes artículos: cable blindado, conector eléctrico, control panel, empaque oring, empaquetadura, etiqueta de puerta, etiqueta de sombreado, fijaciones de fuselaje, manómetro, panel Decal, panel lateral, paquete de tuerca, válvula mecanizada.

Para construir una distribución de frecuencia, solo se necesita contar el número de observaciones que aparecen en cada categoría. Esto se puede hacer mediante la función CONTAR.SI de Excel®.

Ejemplo 18. Construir una distribución de frecuencia para artículos en la base de datos de órdenes de compra

En primer lugar, enumeramos los nombres de los elementos en una columna de la hoja de cálculo. Se usa la columna A, comenzando en la celda A100, debajo de la matriz de datos existente. Es importante utilizar los nombres exactos utilizados en el archivo de datos. Para contar el número de pedidos realizados para cada elemento, utilizamos la función =CONTAR.SI(\$D\$4:\$D\$97, A100), donde A100 es la celda que contiene el nombre del elemento. Esto se muestra en la figura 71. La distribución de frecuencia resultante para los elementos se muestra en la figura 72. Por lo tanto, la compañía realizó 14 pedidos para los fijadores de fuselaje y 11 pedidos para el paquete de tuercas.

Es posible que deseemos ordenarlos usando el análisis de Pareto para obtener más información sobre la frecuencia de la orden.

| | A | B | C | D | E |
|-----|--------------------------|------------|----------------------------------|---|---|
| 99 | Descripción del producto | Frecuencia | | | |
| 100 | Cable blindado | 11 | =CONTAR.SI(\$D\$4:\$D\$97, B100) | | |
| 101 | Conector electrico | 8 | =CONTAR.SI(\$D\$4:\$D\$97, B101) | | |
| 102 | Control Panel | 4 | =CONTAR.SI(\$D\$4:\$D\$97, B102) | | |
| 103 | Empaque oring | 12 | =CONTAR.SI(\$D\$4:\$D\$97, B103) | | |
| 104 | Empaquetadura | 10 | =CONTAR.SI(\$D\$4:\$D\$97, B104) | | |
| 105 | Etiqueta de puerta | 2 | =CONTAR.SI(\$D\$4:\$D\$97, B105) | | |
| 106 | Etiqueta de sombreado | 2 | =CONTAR.SI(\$D\$4:\$D\$97, B106) | | |
| 107 | Fijaciones de fuselaje | 14 | =CONTAR.SI(\$D\$4:\$D\$97, B107) | | |
| 108 | Manómetro | 7 | =CONTAR.SI(\$D\$4:\$D\$97, B108) | | |
| 109 | Panel Decal | 1 | =CONTAR.SI(\$D\$4:\$D\$97, B109) | | |
| 110 | Panel lateral | 8 | =CONTAR.SI(\$D\$4:\$D\$97, B110) | | |
| 111 | Paquete de tuerca | 11 | =CONTAR.SI(\$D\$4:\$D\$97, B111) | | |
| 112 | Valvula mecanizada | 4 | =CONTAR.SI(\$D\$4:\$D\$97, B112) | | |

Figura 71. Uso de la función CONTAR.SI para construir una distribución de frecuencia

Distribución de frecuencia para las órdenes de compra

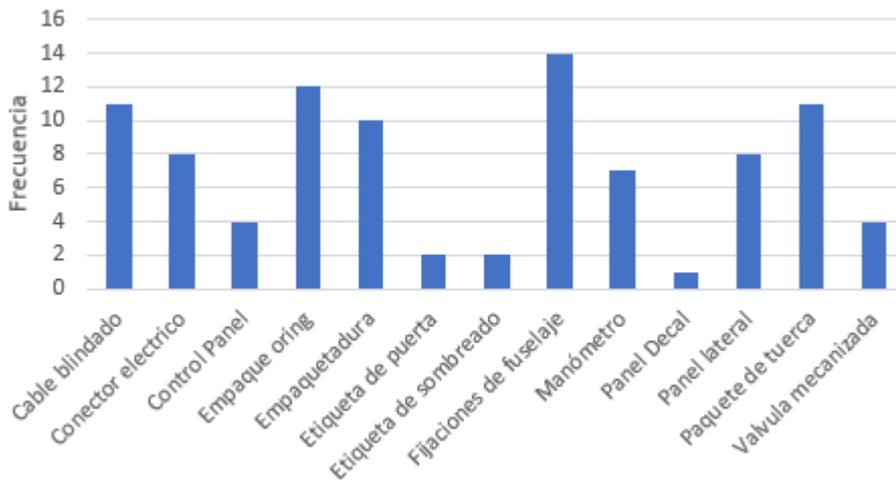


Figura 72. Gráfico de columnas para la distribución de frecuencia de los artículos comprados

Distribuciones de frecuencia relativas

Podemos expresar las frecuencias como una fracción, o proporción, del total; esto se llama la frecuencia relativa. Si un conjunto de datos tiene n observaciones, la frecuencia relativa de la categoría i se calcula como

$$\text{frecuencia relativa de la categoría } i = \frac{\text{frecuencia de la categoría } i}{n}$$

A menudo multiplicamos las frecuencias relativas por 100 para expresarlas como porcentajes. Una distribución de frecuencia relativa es un resumen tabular de las frecuencias relativas de todas las categorías.



Ejemplo 19. Construir una distribución de frecuencia relativa para artículos en la base de datos de órdenes de compra

Los cálculos para las frecuencias relativas son simples. En primer lugar, se suman las frecuencias para encontrar el número total (tener en cuenta que la suma de las frecuencias debe ser la misma que el número total de observaciones, n). A continuación, se divide la frecuencia de cada categoría por este valor. La figura 73 muestra la distribución de frecuencia relativa para los artículos de pedido de compra. La fórmula en la celda C101, por ejemplo, es B101/\$B\$113

Figura 73.
Distribución de frecuencia relativa para los artículos comprados

| | A | B | C | D |
|-----|--------------------------|------------|---------------------|----------------|
| 99 | Descripción del producto | Frecuencia | Frecuencia relativa | |
| 100 | Cable blindado | 11 | 0.11702128 | =C100/\$B\$113 |
| 101 | Conector electrico | 8 | 0.08510638 | =C101/\$B\$113 |
| 102 | Control Panel | 4 | 0.04255319 | =C102/\$B\$113 |
| 103 | Empaque oring | 12 | 0.12765957 | =C103/\$B\$113 |
| 104 | Empaquetadura | 10 | 0.10638298 | =C104/\$B\$113 |
| 105 | Etiqueta de puerta | 2 | 0.0212766 | =C105/\$B\$113 |
| 106 | Etiqueta de sombreado | 2 | 0.0212766 | =C106/\$B\$113 |
| 107 | Fijaciones de fuselaje | 14 | 0.14893617 | =C107/\$B\$113 |
| 108 | Manómetro | 7 | 0.07446809 | =C108/\$B\$113 |
| 109 | Panel Decal | 1 | 0.0106383 | =C109/\$B\$113 |
| 110 | Panel lateral | 8 | 0.08510638 | =C110/\$B\$113 |
| 111 | Paquete de tuerca | 11 | 0.11702128 | =C111/\$B\$113 |
| 112 | Valvula mecanizada | 4 | 0.04255319 | =C112/\$B\$113 |
| 113 | Total | 94 | 1 | |

A continuación, se copia esta fórmula en la columna para calcular las otras frecuencias relativas. Tenga en cuenta que la suma de las frecuencias relativas es igual a 1.0. Un gráfico circular de las frecuencias a veces se utiliza para mostrar estas proporciones visualmente, aunque es más atractivo para un número menor de categorías. Para un gran número de categorías, un gráfico de columnas o barras funcionaría mejor.



Ejemplo 20. Distribución de frecuencia relativa y frecuencia para términos A/P

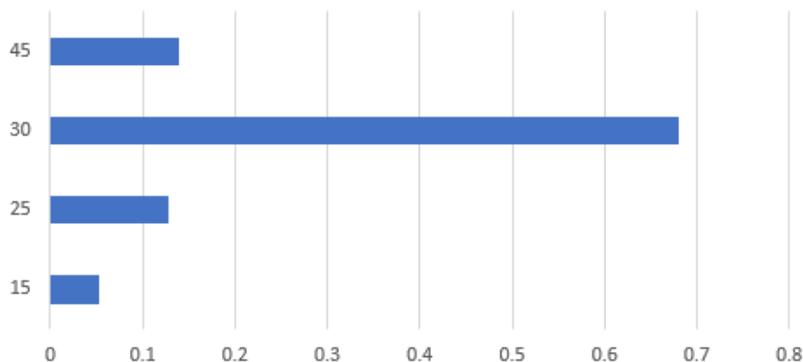
En los datos de las órdenes de compra, los términos de proveedores son del gráfico de barras que muestra las proporciones o los números de frecuencia relativos 15, 25, 30 y 45. Una frecuencia y frecuencias relativas muestran claramente que la mayoría de los pedidos tenían plazos de cuentas por pagar de 30 meses, como puede observarse en la figura 74.

Figura 74.
Distribución de frecuencia y frecuencia relativa para términos de A/P

| | A | B | C |
|-----|-----------------------------|-------------------|----------------------------|
| 99 | A/P terminos (meses) | Frecuencia | Frecuencia relativa |
| 100 | 15 | 5 | 0.053191489 |
| 101 | 25 | 12 | 0.127659574 |
| 102 | 30 | 64 | 0.680851064 |
| 103 | 45 | 13 | 0.138297872 |
| 104 | Total | 94 | 1.000000000 |

Frecuencia relativa de término de pago A/P (meses)

Figura 75.
Gráfico de barras de frecuencias relativas de los términos de A/P



Herramienta Histograma de Excel®

Una representación gráfica de una distribución de frecuencia para datos numéricos en forma de un gráfico de columnas se denomina histograma. Para ello, se hace clic en el botón Herramientas de análisis de datos en el grupo de análisis de la pestaña Datos de la barra de menús de Excel®, y se selecciona Histograma en la lista. En el cuadro de diálogo (consulte la figura 76), se especifica el rango de entrada correspondiente a los datos. Si se incluye el encabezado de columna, también se activa la casilla Etiquetas para que Excel® sepa que el rango contiene una etiqueta. El rango de ubicación define los grupos (Excel® llama a estos “bins”) utilizados para la distribución de frecuencia. Si no se especifica un rango de ubicación, Excel® determinará automáticamente los valores de ubicación para la distribución de frecuencia y el histograma, lo que a menudo resulta en una elección bastante mala. Si se tienen valores discretos, se configura una columna de estos valores en la hoja de cálculo para el rango de ubicación y se especifica este rango en el campo Rango de ubicación. Se describe cómo manejar datos continuos en breve. Se marca la casilla Salida del gráfico para mostrar un histograma, además de la distribución de frecuencia. También se pueden ordenar los valores como un gráfico de Pareto y mostrar las frecuencias acumulativas marcando las casillas adicionales.

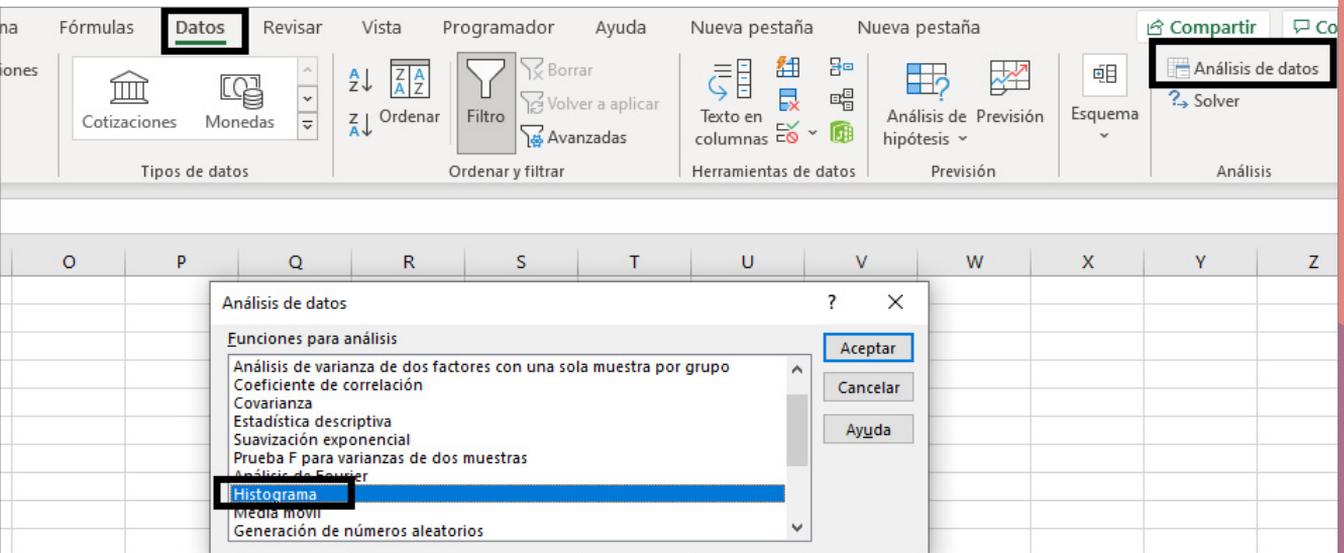


Figura 76. Herramienta de Histograma en función Análisis de datos, en Excel®

Ejemplo 21. Uso de la herramienta Histograma

Crearemos una distribución de frecuencia e histograma si marca la casilla Rótulos (véase la figura 77), es importante que tanto la variable Términos de A/P en la base de datos de órdenes de compra. El rango de entrada y el rango de ubicación tienen etiquetas incluidas. En la figura 78 se muestra el cuadro de diálogo del histograma completado. La primera fila de la figura 79 muestra los resultados de esta herramienta. El rango de entrada incluye el encabezado de columna, así como los datos en la columna H. Se define el rango de ubicación por debajo de los datos en las celdas H99:H103 de la siguiente manera:

Para los datos numéricos que tienen muchos valores discretos diferentes con poca repetición o son continuos, una distribución de frecuencia requiere que se defina especificando el número de grupos, el ancho de cada grupo y los límites superior e inferior de cada grupo.

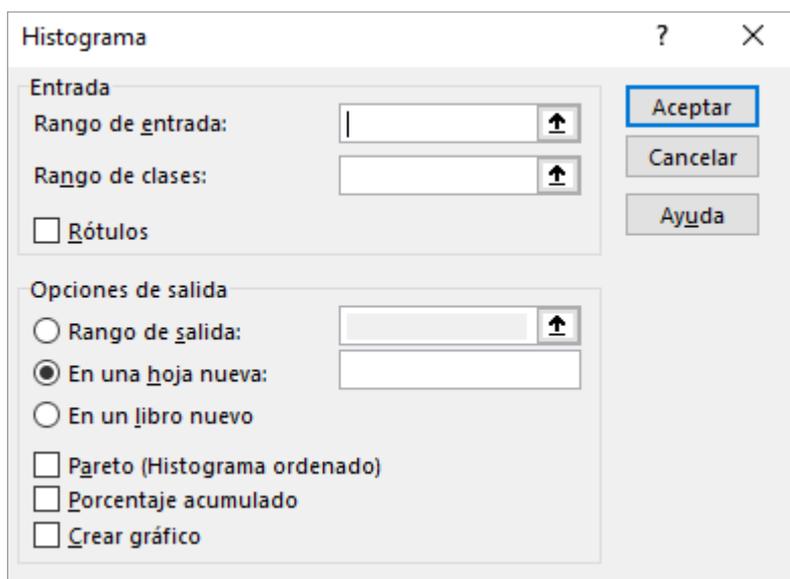


Figura 77. Diálogo de la herramienta histograma

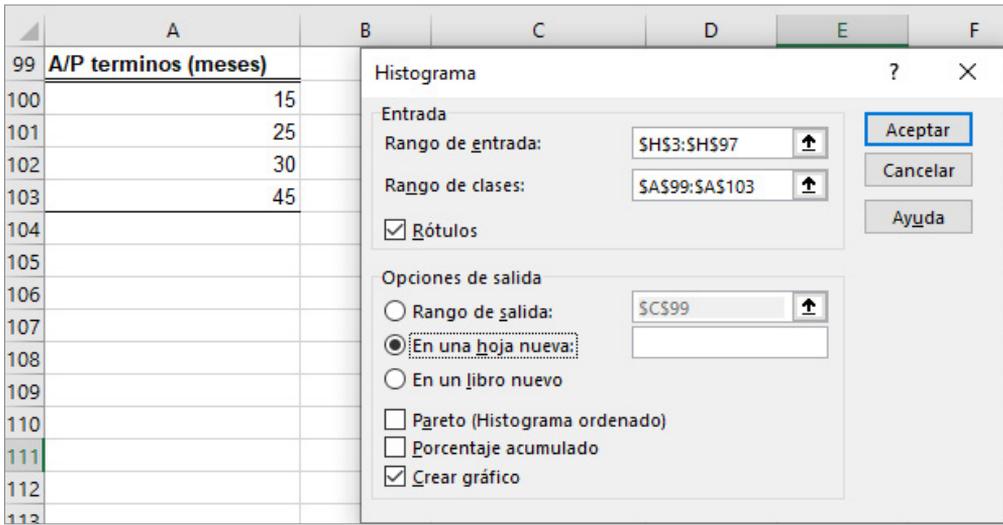


Figura 78. Diálogo del histograma para los datos de términos de A/P

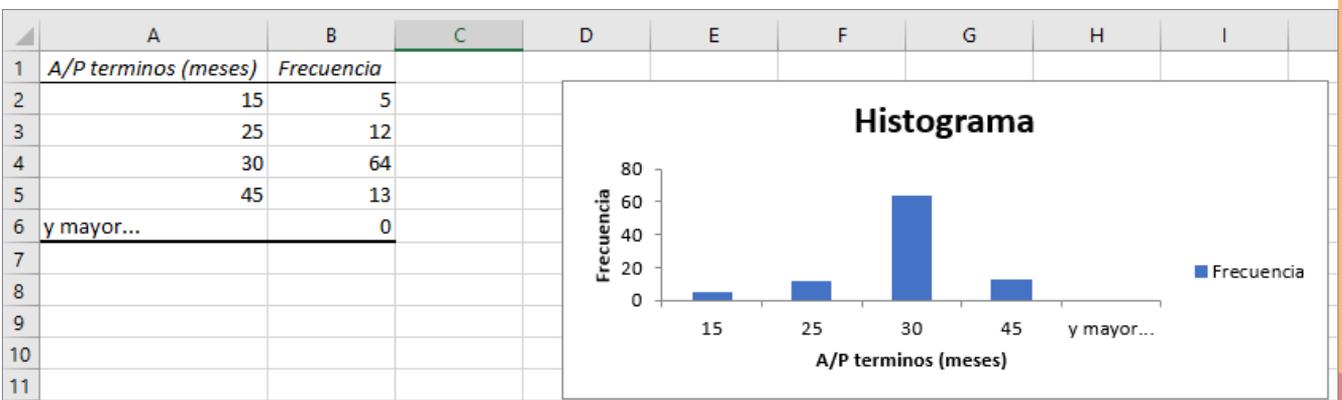


Figura 79. Distribución de frecuencia de Excel® e histograma para términos A/P

Es importante recordar que los grupos pueden no superponerse, por lo que cada valor se cuenta exactamente en un grupo. Se deben definir los grupos después de examinar el intervalo de los datos. Generalmente, se eligen entre 5 a 15 grupos y el rango de cada uno debe ser igual; así, cuantos más datos se tengan, más grupos se deben usar. Se debe tener en cuenta que, con menos grupos, los anchos de grupo serán más amplios. Los anchos de grupo más anchos proporcionan un histograma “ancho”. A veces es necesario experimentar para encontrar el mejor número de grupos para proporcionar una visualización útil de los datos. Se elige el límite inferior del primer grupo

(LL) como un número entero menor que el valor de datos mínimo y el límite superior del último grupo (UL) como un número entero mayor que el valor de datos máximo. Generalmente, tiene sentido elegir números enteros. A continuación, se puede calcular el ancho del grupo como

$$\text{Ancho de grupo} = \frac{\text{Límite superior} - \text{Límite inferior}}{\text{Número de grupos}}$$

Ejemplo 22. Construir una distribución de frecuencia e histograma para el costo por pedido

En este ejemplo, aplicamos la herramienta Histograma de Excel® a los datos de Coste por pedido en la columna G de la base de datos de Órdenes de compra. Los datos van desde un mínimo de \$68.75 hasta un máximo de \$127,500. Se puede encontrar esto mediante las funciones MIN y MAX o simplemente ordenando los datos. Para asegurarse de que todos los grupos de datos se incluirán en algún grupo, tiene sentido establecer el límite inferior del primer grupo en \$0 y el límite superior del último grupo en \$130,000. Por lo tanto, si se seleccionan 5 grupos usando la ecuación del Ancho de grupo, el ancho de cada grupo es $(\$130,000 - 0)/5 = \$26,000$; si elegimos 10 grupos, el ancho es de $(\$130,000 - 0)/10 = \$13,000$. Por tanto, seleccionamos 5 grupos. Al hacerlo, el rango de ubicación se especifica como

Límite de grupo superior

| |
|--------------|
| \$ 0.00 |
| \$ 26,000.00 |
| \$ 52,000.00 |
| \$ 78,000.00 |
| \$104,000.00 |
| \$130,000.00 |

Esto significa que el primer grupo incluye todos los valores menores o iguales a \$0; el segundo grupo incluye todos los valores mayores que \$0

pero menores o iguales a \$26,000, y así, sucesivamente. Tenga en cuenta que los grupos no se superponen porque el límite inferior de un grupo es estrictamente mayor que el límite superior del grupo anterior. Se sugiere usar el encabezado “Límite de grupo superior” para el rango de ubicación para que esto sea claro. En la hoja de cálculo, este rango de ubicación se introduce en las celdas G99:G105. El rango de entrada en el cuadro de diálogo Histograma es G4:G97. La figura 80 muestra los resultados. Estos resultados muestran que la gran mayoría de los pedidos fueron por \$26,000 o menos y caen rápidamente más allá de este valor. Seleccionar un mayor número de grupos, podría ayudar a comprender mejor la naturaleza de los datos. La figura 81 muestra los resultados utilizando 10 grupos. Esto muestra que un porcentaje más alto de pedidos fueron por \$13,000 o menos que estaban entre \$13,000 y \$26,000.

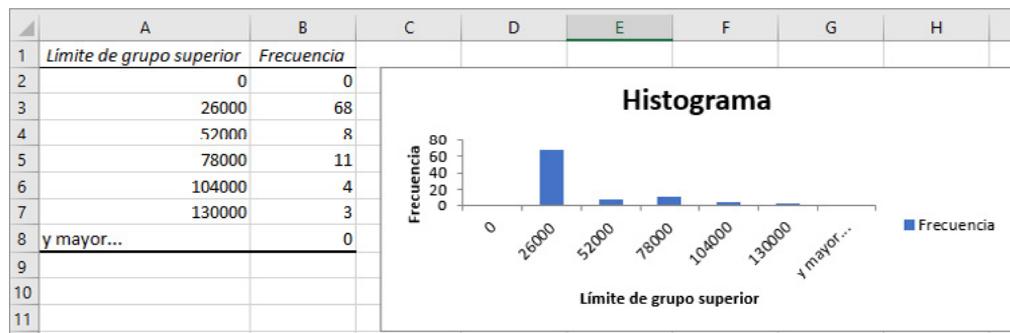


Figura 80. Distribución de frecuencia e histograma para coste por pedido (5 grupos)

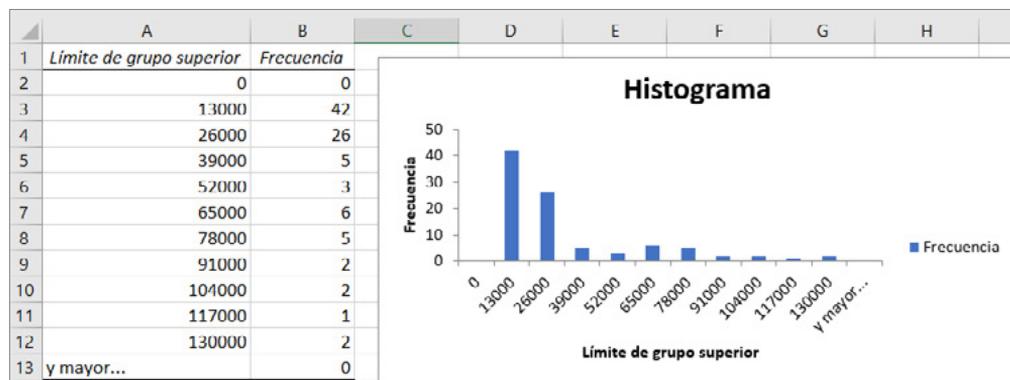


Figura 81. Distribución de frecuencia e histograma para el coste por pedido (10 grupos).

Una limitación de la herramienta Histograma de Excel® es que la distribución de frecuencia y el histograma no están vinculados a los datos; por lo tanto, si se cambia cualquiera de los datos, se debe repetir todo el procedimiento para construir una nueva distribución de frecuencia e histograma.

Distribuciones de frecuencia relativa acumulativa

Para los datos numéricos, también se puede calcular la frecuencia relativa de las observaciones en cada grupo. Al sumar todas las frecuencias relativas en o por debajo de cada límite superior, se obtiene la frecuencia relativa acumulativa. La frecuencia relativa acumulada representa la proporción del número total de observaciones que caen en o por debajo del límite superior de cada grupo. Un resumen tabular de frecuencias relativas acumulativas se denomina distribución de frecuencia relativa acumulativa.

Ejemplo 23. Cálculo de frecuencias relativas acumuladas

La figura 82 se muestra la frecuencia relativa y las distribuciones de frecuencia relativa acumulativa para los datos de Costo por pedido utilizando 10 grupos. Las frecuencias relativas se calculan dividiendo la frecuencia entre el número total de observaciones (94). En la columna D, establecemos la frecuencia relativa acumulativa del primer grupo igual a su frecuencia relativa. Luego, agregamos la frecuencia relativa del siguiente grupo a la frecuencia relativa acumulada.

Por ejemplo, la frecuencia relativa acumulativa en la celda D3 se calcula como “D2+C3” a $0.000+0.447$ a 0.447 ; la frecuencia relativa acumulativa en la celda D4 se calcula como “D3+C4” a $0.447+0.277$ a 0.723 , y así sucesivamente. (Los valores mostrados se redondean a tres decimales.) Dado que las frecuencias relativas deben estar entre 0 y 1 y deben sumar hasta 1, la frecuencia acumulativa para el último grupo debe ser igual a 1.

| | A | B | C | D |
|----|--------------------------|------------|---------------------|-------------------------------|
| 1 | Límite de grupo superior | Frecuencia | Frecuencia relativa | Frecuencia relativa acumulada |
| 2 | \$ - | 0 | 0.0000 | 0.0000 |
| 3 | \$ 13,000.00 | 42 | 0.4468 | 0.4468 |
| 4 | \$ 26,000.00 | 26 | 0.2766 | 0.7234 |
| 5 | \$ 39,000.00 | 5 | 0.0532 | 0.7766 |
| 6 | \$ 52,000.00 | 3 | 0.0319 | 0.8085 |
| 7 | \$ 65,000.00 | 6 | 0.0638 | 0.8723 |
| 8 | \$ 78,000.00 | 5 | 0.0532 | 0.9255 |
| 9 | \$ 91,000.00 | 2 | 0.0213 | 0.9468 |
| 10 | \$ 104,000.00 | 2 | 0.0213 | 0.9681 |
| 11 | \$ 117,000.00 | 1 | 0.0106 | 0.9787 |
| 12 | \$ 130,000.00 | 2 | 0.0213 | 1.0000 |
| 13 | y mayor... | 0 | 0.0000 | 1.0000 |
| 14 | Total | 94 | | |

Figura 82. Distribución de frecuencia relativa acumulativa para datos de costo por pedido

La figura 83 muestra un gráfico para la frecuencia relativa acumulativa, que se denomina Ogiva. Desde este gráfico, se puede estimar fácilmente la proporción de observaciones que están por debajo de un cierto valor. Por ejemplo, se puede ver que un poco más del 70% de los datos caen en \$ 26,000 o menos, alrededor del 90% de los datos caen en o por debajo de \$78,000, y así sucesivamente.



Figura 83. Ogiva por costo por pedido

Los datos a menudo se expresan como percentiles y cuartiles. Sin duda, usted está familiarizado con los percentiles de las pruebas estandarizadas utilizadas para los exámenes de ingreso a la universidad o a la escuela de posgrado (SAT, ACT, GMAT, GRE, etc.). Los percentiles especifican el porcentaje de otros que obtuvieron la puntuación de un individuo en particular o por debajo de ellos. En términos generales, el percentil k -ésimo es un valor en o por debajo del cual se encuentran al menos el k por ciento de las observaciones. Sin embargo, la forma en que se calculan los percentiles no está estandarizada. La forma más común de calcular el percentil k -ésimo es ordenar los valores de datos de menor a mayor y calcular el rango del percentil k -ésimo usando la fórmula siguiente

$$\text{percentil } k\text{-ésimo} = \frac{nk}{100} + 0.05$$

donde n es el número de observaciones. Redondee esto al entero más cercano y tome el valor correspondiente a este rango como el percentil k -ésimo.

Ejemplo 24. Cálculo de percentiles

En los datos de las órdenes de compra, se tienen $n = 94$ observaciones. El rango del percentil 90vo. ($k = 90$) para el costo de los datos ordenados es $94 * (90)/100 + 0.5 = 85.1$, o redondeado, 85. El valor ordenado 85 es \$74,375 y corresponde al 90vo. percentil. Esto significa que el 90% de los costos en orden es menor o igual a \$74,375, y el 10% de los costos son más altos. El software estadístico utiliza diferentes métodos que a menudo implican interpolar entre rangos en lugar de redondeo, produciendo así resultados diferentes. La función de Excel® PERCENTIL(matriz,k) calcula el percentil k de datos en el intervalo especificado en el campo de matriz, donde k está en el intervalo 0 a 1.



Ejemplo 25. Cálculo de percentiles en Excel®

Para encontrar el percentil 90 para los datos de costo por pedido en los datos de órdenes de compra, se utiliza la de función de Excel®=PERCENTIL(G4:G97,0.9). Esto calcula el 90vo. percentil de \$73,737.50, que es diferente de usar la fórmula vista en el ejemplo 24.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|-----------------------|-----------|--------------|---------------|-----------|----------|----------------|------------------------|----------|----------|------|
| 1 | Órdenes de compra | | | | | | 73737.5 | =PERCENTIL(G4:G97,0.9) | | | |
| 2 | | | | | | | | | | | |
| 3 | Proveedor | No. Orden | No. Producto | Descripción | Costo del | Cantidad | Costo por orde | A/P termi | Fecha de | Fecha de | cibo |
| 4 | Sujotadores Hulkey | Aug11001 | 1122 | Fijaciones de | \$ 4.25 | 19,500 | \$ 82,875.00 | 30 | 08/05/11 | 08/13/11 | |
| 5 | Láminas Alum | Aug11002 | 1243 | Fijaciones de | \$ 4.25 | 10,000 | \$ 42,500.00 | 30 | 08/08/11 | 08/14/11 | |
| 6 | Aeroespacial Fast-Tie | Aug11003 | 5462 | Cable blindad | \$ 1.05 | 23,000 | \$ 24,150.00 | 30 | 08/10/11 | 08/15/11 | |
| 7 | Aeroespacial Fast-Tie | Aug11004 | 5462 | Cable blindad | \$ 1.05 | 21,500 | \$ 22,575.00 | 30 | 08/15/11 | 08/22/11 | |
| 8 | Steelpin Co. | Aug11005 | 5319 | Cable blindad | \$ 1.10 | 17,500 | \$ 19,250.00 | 30 | 08/20/11 | 08/31/11 | |

Figura 84. Cálculo de percentiles usando Excel®

Excel® también tiene una herramienta para ordenar los datos de alto a bajo y calcular percentiles asociados a cada valor. Se selecciona en el menú Análisis de Datos de la ventana Datos la función para análisis “Jerarquía y Percentil” y se especifica el rango de los datos en el cuadro de diálogo. Asegúrese de marcar la casilla Etiquetas en primera fila si el rango incluye un encabezado en la hoja de cálculo.

The image shows the Excel interface with the 'Datos' ribbon selected. The 'Análisis de datos' button is highlighted. Below it, the 'Análisis de datos' dialog box is open, showing a list of analysis functions. 'Jerarquía y percentil' is selected in the list. The background shows a spreadsheet with a formula bar containing '=PERCENTIL(G4:G97,0.9)' and a data table similar to the one in Figure 84.

Figura 85. Opción para evaluar Percentiles en el menú, Análisis de datos

Ejemplo 26. Excel® Rank and Percentile Tool

Una parte de los resultados de la herramienta Rango y Porcentaje que calculamos en el ejemplo 25, donde el percentil 90vo. es \$73,737.50 para los datos de Costo por pedido, se compara con los datos de la figura 86 utilizando la herramienta de Jerarquía y percentil de Excel®, y genera un cálculo del percentil 90.3vo. que es igual a \$74,375.

Figura 86. Ventana de evaluación Jerarquía y percentil

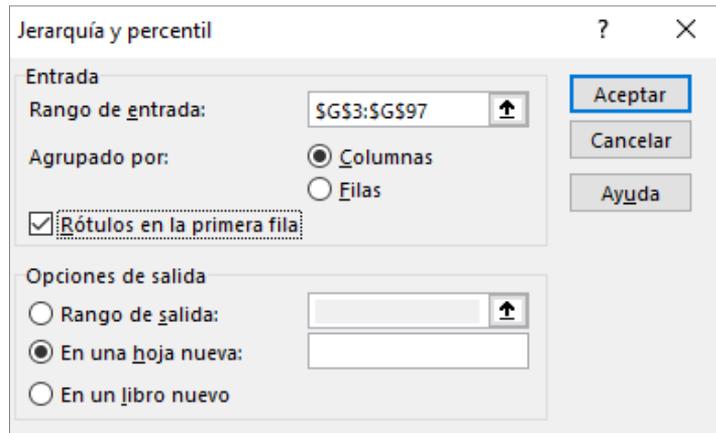


Figura 87. Porción de resultados de herramientas de rango y percentil

| | A | B | C | D |
|----|-----------------|------------------------|------------------|-------------------|
| 1 | <i>Posición</i> | <i>Costo por orden</i> | <i>Jerarquía</i> | <i>Porcentaje</i> |
| 2 | 74 | \$ 127,500.00 | 1 | 100.00% |
| 3 | 62 | \$ 121,000.00 | 2 | 98.90% |
| 4 | 71 | \$ 110,000.00 | 3 | 97.80% |
| 5 | 16 | \$ 103,530.00 | 4 | 96.70% |
| 6 | 73 | \$ 96,750.00 | 5 | 95.60% |
| 7 | 1 | \$ 82,875.00 | 6 | 94.60% |
| 8 | 67 | \$ 81,937.50 | 7 | 93.50% |
| 9 | 82 | \$ 77,400.00 | 8 | 92.40% |
| 10 | 54 | \$ 76,500.00 | 9 | 91.30% |
| 11 | 80 | \$ 74,375.00 | 10 | 90.30% |
| 12 | 68 | \$ 72,250.00 | 11 | 89.20% |

Los cuartiles dividen los datos en cuatro partes. El percentil 25 se llama el primer cuartil, Q1; el percentil 50 se llama el segundo cuartil, Q2; el percentil 75 se llama el tercer cuartil, Q3; y el percentil 100 es el cuarto cuartil, Q4. Una cuarta parte de los datos caen por debajo del primer cuartil, la mitad

están por debajo del segundo cuartil, y tres cuartas partes están por debajo del tercer cuartil. Podemos calcular cuartiles usando la función de Excel® CUARTIL(matriz, cuartil), donde matriz especifica el rango de los datos, y cuartil es un número entero entre 1 y 4, que designa el cuartil deseado. Se pueden extender estas ideas a otras divisiones de los datos. Por ejemplo, los deciles dividen los datos en 10 conjuntos: el percentil 10, el percentil 20, etc. Todos estos tipos de medidas se denominan perfiles de datos o percentiles.



Ejemplo 27. Cálculo de Cuartiles en Excel®

Para los datos de Órdenes de compra se puede usar la función de Excel® CUARTIL(G4:G97,k), donde k oscila entre 1 y 4, para calcular los cuartiles. Los resultados son los siguientes:

| | A | B | C | D |
|-----|-------------------|---------------|-------------------------|---|
| 99 | k Cuartil | Valor | | |
| 100 | 1 Primer cuartil | \$ 6,757.81 | =CUARTIL(G4:G97, A100) | |
| 101 | 2 Segundo cuartil | \$ 15,562.50 | =CUARTIL(G5:G98, A101) | |
| 102 | 3 Tercer cuartil | \$ 26,468.75 | =CUARTIL(G6:G99, A102) | |
| 103 | 4 Cuarto cuartil | \$ 127,500.00 | =CUARTIL(G7:G100, A103) | |

Figura 88. Cálculo de cuartiles 1 al 4

Se puede concluir que el 25% de los costos del pedido caen o por debajo de \$6,757.81; 50% caen por debajo de \$15,656.25; El 75% cae por debajo de \$27,593.75, y 100% cae por debajo del valor máximo de \$127,500.

Tabulaciones cruzadas

Una de las herramientas estadísticas más básicas utilizadas para resumir los datos categóricos y examinar la relación entre dos variables categóricas es la tabulación cruzada. Una tabulación cruzada es un método tabular que muestra el número de observaciones en un conjunto de datos para diferentes subcategorías de dos variables categóricas. Una tabla de tabulación

cruzada a menudo se denomina tabla de contingencia. Las subcategorías de las variables deben ser mutuamente excluyentes y exhaustivas, lo que significa que cada observación puede clasificarse en una sola subcategoría y, en conjunto, en todas las subcategorías deben constituir el conjunto de datos completo. Las tabulaciones cruzadas se utilizan comúnmente en la investigación de marketing para proporcionar información sobre las características de los diferentes segmentos del mercado utilizando variables categóricas como el género, el nivel educativo, el estado civil, etc.

Ejemplo 28. Construir una tabulación cruzada

Examinemos la base de datos Transacciones de ventas, una parte de la cual se muestra en la figura 89. Se supone que deseamos identificar el número de libros y DVDs ordenados por región. Una tabulación cruzada tendrá filas correspondientes a las diferentes regiones y columnas correspondientes a los productos. Dentro de la tabla enumeramos el recuento del número en cada par de categorías. En la figura 90 se muestra una tabulación cruzada de estos datos. Visualizar los datos como un gráfico es una buena manera de comunicar los resultados.

Expresar los resultados como porcentajes de una fila o columna facilita la interpretación de las diferencias entre regiones o productos, especialmente a medida que los totales de cada categoría difieren. En la figura 91 se muestra el porcentaje de ventas de libros y DVD dentro de cada región; esto se calcula dividiendo los recuentos por los totales de fila y multiplicando por 100 (en Excel®, simplemente se divide el recuento por el total y se formatea el resultado como un porcentaje haciendo clic en el botón % en el grupo Número dentro de la pestaña Inicio en la cinta de opciones). Por ejemplo, vemos que, aunque se venden más libros y DVD en la región Oeste que en el norte, los porcentajes relativos de cada producto son similares, especialmente en comparación con las regiones Este y Sur.

| | A | B | C | D | E | F | G | H | I |
|----|----------------------------------|--------|---------|--------------------|-------------------|----------|----------|--------------|---|
| 1 | Transacciones de venta: Julio 14 | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | ID cliente | Region | Pago | Código Transacción | Recurso | Cantidad | Producto | Hora del día | |
| 4 | 10001 | Este | Paypal | 93816545 | Web | \$20.19 | DVD | 22:19 | |
| 5 | 10002 | Oeste | Credito | 74083490 | Web | \$17.85 | DVD | 13:27 | |
| 6 | 10003 | Norte | Credito | 64942368 | Web | \$23.98 | DVD | 14:27 | |
| 7 | 10004 | Oeste | Paypal | 70560957 | Corre electronico | \$23.51 | Libro | 15:38 | |
| 8 | 10005 | Sur | Credito | 35208817 | Web | \$15.33 | Libro | 15:21 | |
| 9 | 10006 | Oeste | Paypal | 20978903 | Corre electronico | \$17.30 | DVD | 13:11 | |
| 10 | 10007 | Este | Credito | 80103311 | Web | \$177.72 | Libro | 21:59 | |
| 11 | 10008 | Oeste | Credito | 14132683 | Web | \$21.76 | Libro | 04:04 | |
| 12 | 10009 | Oeste | Paypal | 40128225 | Web | \$15.92 | DVD | 19:35 | |
| 13 | 10010 | Sur | Paypal | 49073721 | Web | \$23.39 | DVD | 13:26 | |

Figura 89. Porción de la base de datos de transacciones de ventas

| | A | B | C | D |
|---|---|-----|-------|---------------|
| 3 | Cuenta de Producto Etiquetas de columna | | | |
| 4 | Etiquetas de fila | DVD | Libro | Total general |
| 5 | Este | 42 | 56 | 98 |
| 6 | Norte | 42 | 43 | 85 |
| 7 | Oeste | 90 | 100 | 190 |
| 8 | Sur | 37 | 62 | 99 |
| 9 | Total general | 211 | 261 | 472 |

Campos de tabla dinámica

Seleccionar campos para agregar al informe:

Buscar

ID cliente

Region

Pago

Código Transacción

Recurso

Arrastrar campos entre las áreas siguientes:

Filtros

Columnas: Producto

Filas: Region

Valores: Cuenta de Producto

Figura 90. Tabulación cruzada de datos de transacciones de ventas

| | A | B | C | D |
|---|---|--------|--------|---------------|
| 3 | Cuenta de Producto Etiquetas de columna | | | |
| 4 | Etiquetas de fila | DVD | Libro | Total general |
| 5 | Este | 8.90% | 11.86% | 20.76% |
| 6 | Norte | 8.90% | 9.11% | 18.01% |
| 7 | Oeste | 19.07% | 21.19% | 40.25% |
| 8 | Sur | 7.84% | 13.14% | 20.97% |
| 9 | Total general | 44.70% | 55.30% | 100.00% |

Campos de tabla dinámica

Seleccionar campos para agregar al informe:

Buscar

ID cliente

Region

Pago

Código Transacción

Recurso

Arrastrar campos entre las áreas siguientes:

Filtros

Columnas: Producto

Filas: Region

Valores: Cuenta de Producto

Figura 91. Porcentaje de ventas de productos dentro de cada región

4.8 Modelar los datos

Para tomar una decisión, se debe ser capaz de especificar las alternativas de decisión que representan las opciones que se pueden tomar y los criterios para evaluar las alternativas. Especificar alternativas de decisión podría ser muy simple; por ejemplo, es posible que se deba elegir una de las tres opciones del plan de salud corporativo. Otras situaciones pueden ser más complejas; por ejemplo, al localizar un nuevo centro de distribución, es posible que no sea posible enumerar solo un pequeño número de alternativas. El conjunto de ubicaciones potenciales podría estar en cualquier parte de los Estados Unidos o incluso dentro de una gran región geográfica como Asia. Los criterios de decisión pueden ser maximizar los beneficios netos con descuento, la satisfacción del cliente o los beneficios sociales o minimizar los costos, el impacto ambiental o alguna medida de pérdida. Muchos problemas de decisión se pueden formalizar utilizando un modelo, el cual es una abstracción o representación de un sistema, idea u objeto real. Los modelos capturan las características más importantes de un problema y las presentan en una forma fácil de interpretar. Un modelo puede ser tan simple como una descripción escrita o verbal de algún fenómeno, una representación visual como un gráfico o un diagrama de flujo, o una representación matemática u hoja de cálculo.

Ejemplo 29. Tres formas de modelo

Las ventas de un nuevo producto, como un iPad de primera generación, un teléfono Android o una televisión 3D, a menudo siguen un patrón común. Se puede representar esto de una de tres maneras siguientes:

1. Una simple descripción verbal de las ventas podría ser: La tasa de ventas comienza pequeña a medida que los primeros adoptantes comienzan a evaluar un nuevo producto y luego comienza a crecer a un ritmo creciente con el tiempo a medida que se extienden los comentarios

positivos de los clientes. Eventualmente, el mercado comienza a saturarse y la tasa de ventas comienza a disminuir.

- Un boceto de las ventas como una curva en forma de S a lo largo del tiempo, como se muestra en la figura 92, es un modelo visual que transmite este fenómeno.

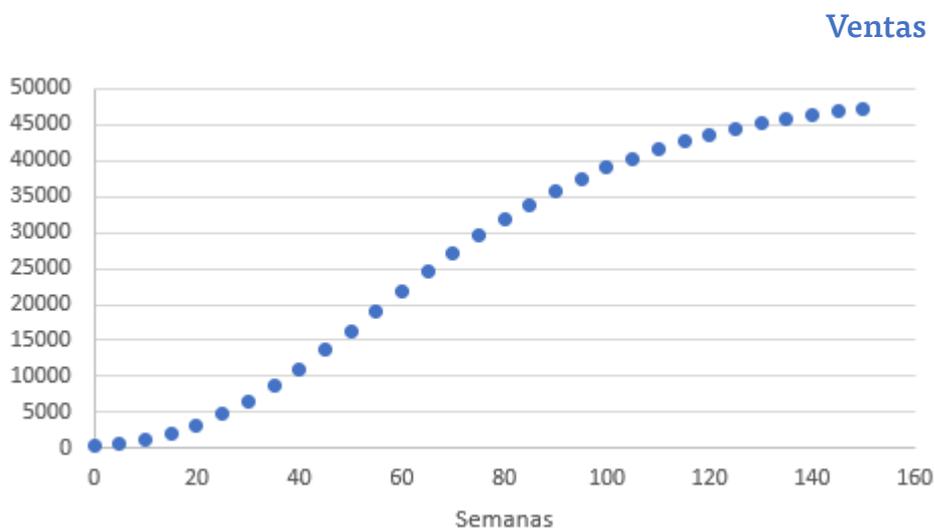


Figura 92. Ventas de nuevos productos a lo largo del tiempo

- Por último, los analistas podrían identificar un modelo matemático que caracterice esta curva. Varias funciones matemáticas diferentes hacen esto; uno se llama curva de Gompertz y tiene la fórmula: $S = ae^{-be^{-ct}}$, donde S son las ventas, t es el tiempo, e es la base de los logaritmos naturales, y a , b , y c son constantes. Por supuesto, no se espera que lo sepa; eso es lo que hacen los profesionales de análisis. Este modelo matemático proporciona la capacidad de predecir las ventas cuantitativamente y analizar las decisiones potenciales sobre su comportamiento.

Los modelos pueden ser descriptivos, predictivos o prescriptivos y, por lo tanto, se utilizan en una amplia variedad de aplicaciones de análisis empresarial. En el ejemplo anterior, la descripción verbal y gráfica del modelo son puramente descriptivas; simplemente explican el fenómeno. Mientras que el modelo matemático también describe el fenómeno, se puede utilizar para predecir las ventas en un momento futuro. Los modelos generalmente se

desarrollan a partir de la teoría u observación y establecen relaciones entre las acciones que los responsables de la toma de decisiones podrían tomar y los resultados que se podrían esperar, lo que permite a los tomadores de decisiones predecir lo que podría suceder en función del modelo.

Los modelos complementan la intuición de los tomadores de decisiones y a menudo proporcionan ideas que la intuición no puede. Un modelo descriptivo simple es una representación visual denominada diagrama de influencia porque describe cómo varios elementos del modelo influyen o se relacionan con otros. Un diagrama de influencia es un enfoque útil para conceptualizar la estructura de un modelo y puede ayudar a crear un modelo matemático u hoja de cálculo. Los elementos del modelo se representan mediante símbolos circulares denominados nodos. Las flechas denominadas ramas conectan los nodos y muestran qué elementos influyen en otros. Los diagramas de influencia son muy útiles en las primeras etapas de la construcción del modelo cuando es necesario entender y caracterizar las relaciones clave. En el ejemplo 29 se muestra cómo construir diagramas de influencia simples, y en el ejemplo 30, cómo construir un modelo matemático basándose en el diagrama de influencia.

Ejemplo 30. Un diagrama de influencia para el costo total

A partir de los principios básicos del negocio, sabemos que el costo total de producir un volumen fijo de un producto se compone de costos fijos y costos variables. Por lo tanto, un diagrama de influencia simple que muestra estas relaciones se observa en la figura 93. Podemos desarrollar un modelo más detallado observando que el costo variable depende del costo variable unitario, así como de la cantidad producida. El modelo expandido se muestra en la figura 94. En esta todos los nodos que no tienen ramas que apuntan a ellos son entradas para el modelo, se puede ver, además, que el costo variable unitario y los costos fijos son entradas de datos en el modelo. La cantidad producida, sin embargo, es una variable de decisión porque puede ser controlada por el administrador de la operación. El costo

total es la salida que estaríamos interesados en calcular (tenga en cuenta que no tiene ramas que apunten fuera de ella). El nodo de costo variable vincula algunas de las entradas con la salida y se puede considerar como un “bloque de construcción” del modelo para el costo total.

Figura 93. Un diagrama de influencia que relaciona el costo total de sus componentes clave

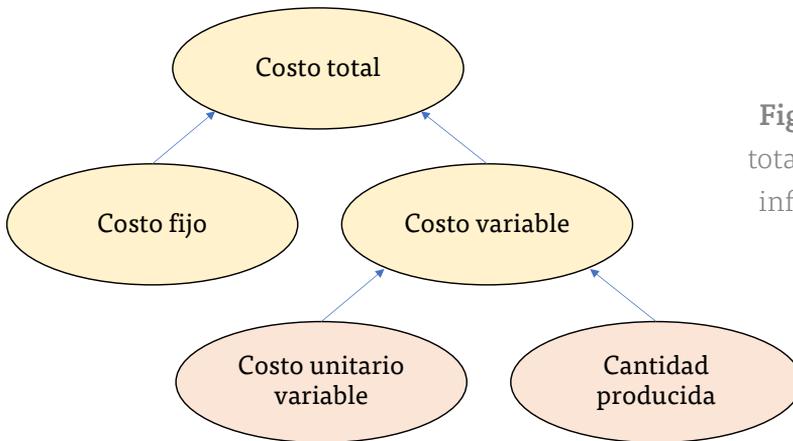
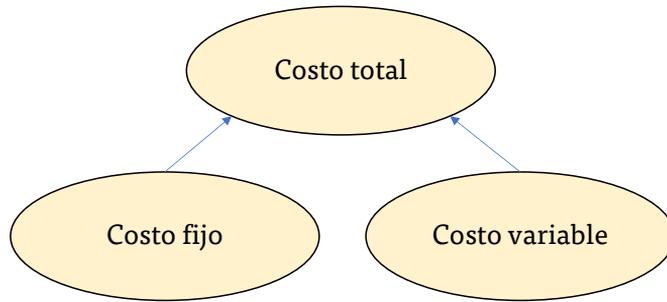


Figura 94. Un costo total del diagrama de influencia ampliado

Ejemplo 31. Construir un modelo matemático a partir de un diagrama de influencia

Se desarrolla un modelo matemático a partir del diagrama de influencia en las figuras 93 y 94. En primer lugar, debemos especificar la naturaleza precisa de las relaciones entre las distintas cantidades. Por ejemplo, se puede afirmar fácilmente que

$$\text{Costo total} = \text{Costo fijo} + \text{Costo variable} \tag{1.1}$$

La lógica también sugiere que el costo variable es el costo unitario variable multiplicado por la cantidad producida. Así

$$\text{Costo variable} = \text{Costo variable unitario} * \text{Cantidad producida} \quad (1.2)$$

Al sustituir esto en la ecuación 1.1, se completa del análisis de costo total como resultado de la suma del costo fijo más el producto de costo variable unitario por cantidad producida.

$$\begin{aligned} \text{Costo total} &= \text{Costo fijo} + \text{Costo variable} \\ \text{Coste total} &= \text{Costo fijo} + \text{Costo variable unitario} * \text{Cantidad producida} \quad (1.3) \end{aligned}$$

Al utilizar estas relaciones, se puede desarrollar una representación matemática definiendo símbolos para cada una de estas cantidades:

CT - costo total

F - costo fijo

V - costo variable unitario

Q - cantidad producida

Esto da como resultado el modelo

$$CT = F + V * Q \quad (1.4)$$

Un modelo de decisión es una representación lógica o matemática de un problema o situación empresarial que se puede utilizar para comprender, analizar o facilitar la toma de una decisión. La mayoría de los modelos de decisión tienen tres tipos de entrada:

1. Datos, que se supone que son constantes a efectos del modelo. Algunos ejemplos serían los costos, las capacidades de la máquina y las distancias interurbanas.
2. Variables incontrolables, que son cantidades que pueden cambiar, pero no pueden ser controladas directamente por el responsable de la

toma de decisiones. Algunos ejemplos serían la demanda de los clientes, las tasas de inflación y los retornos de inversión. A menudo, estas variables son inciertas.

3. Variables de decisión, que son controlables y pueden ser seleccionadas a discreción del responsable de la toma de decisiones. Algunos ejemplos serían las cantidades de producción, los niveles de personal y las asignaciones de inversión.

Los modelos de decisión caracterizan las relaciones entre los datos, las variables incontrolables y las variables de decisión, así como los resultados de interés para el responsable de la toma de decisiones. Los modelos de decisión se pueden representar de varias maneras, normalmente con funciones matemáticas y hojas de cálculo. Las hojas de cálculo son vehículos ideales para implementar modelos de decisión debido a su versatilidad en la gestión de datos, la evaluación de diferentes escenarios y la presentación de resultados de una manera significativa.

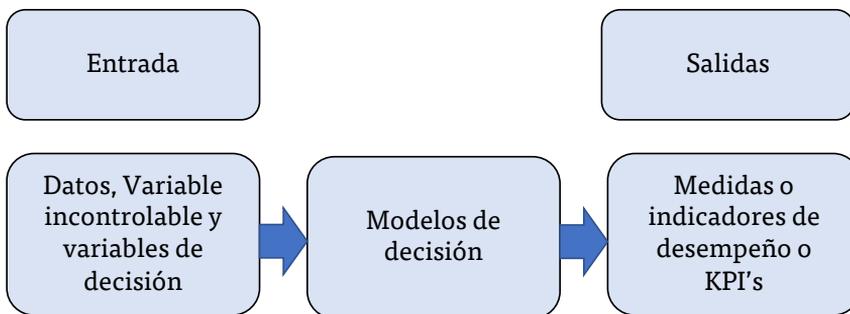


Figura 95. Naturaleza de los modelos de decisión

Ejemplo 32. Un modelo de decisión de equilibrio

Supongamos que un fabricante puede producir una pieza por \$125/unidad con un costo fijo de \$50,000. La alternativa es externalizar la producción a un proveedor a un costo unitario de \$175. El costo total de fabricación se expresa mediante la ecuación (1.5):

$$TC(\text{fabricación}) = \$50,000 + \$125 * Q \tag{1.5}$$

y el costo total de la subcontratación se puede escribir como

$$TC(\text{subcontratación}) = \$175 * Q$$

Los modelos matemáticos son fáciles de manipular; por ejemplo, es fácil encontrar el volumen de equilibrio estableciendo $TC(\text{fabricación}) = TC(\text{subcontratación})$ y resolviendo para Q:

$$\begin{aligned} \$50,000 + \$125 * Q &= \$175 * Q \\ \$50,000 &= 50 * Q \\ Q &= 1,000 \end{aligned}$$

Por lo tanto, si el volumen de producción previsto es superior a 1000, es más económico fabricar la pieza; si es menos de 1000, entonces debe ser subcontratado. Esto se muestra gráficamente en la figura 96.

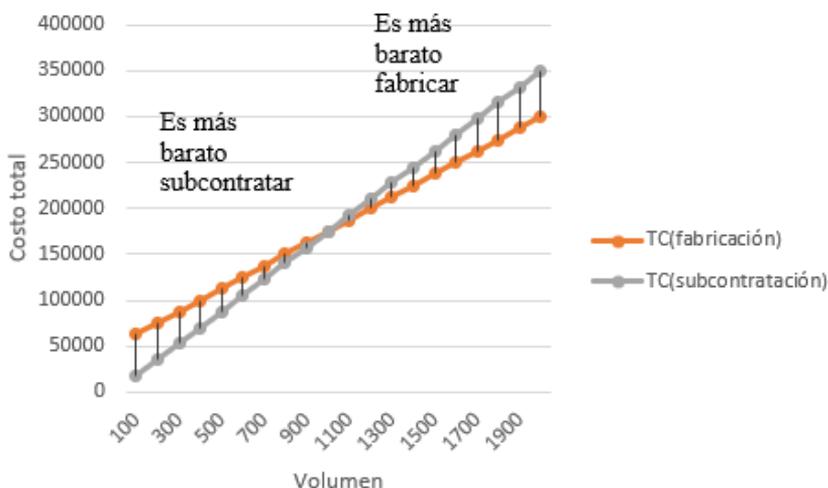


Figura 96. Ilustración gráfica del análisis de equilibrio

También podemos desarrollar una fórmula general para el punto de equilibrio, dejando que C sea el costo unitario de la subcontratación de la pieza y estableciendo $TC(\text{fabricación}) = TC(\text{subcontratación})$. Utilizando las fórmulas:

$$F + VQ = CQ$$

$$Q = \frac{F}{V - C}$$

Ejemplo 33. Un modelo de decisión de promoción de ventas

En la industria de comestibles, los gerentes normalmente necesitan saber la mejor manera de usar los precios, los cupones y las estrategias publicitarias para influir en las ventas. Los tenderos a menudo estudian la relación del volumen de ventas con estas estrategias mediante la realización de experimentos controlados para identificar la relación entre ellos y los volúmenes de ventas³. Es decir, implementan diferentes combinaciones de precios, cupones y publicidad, observan las ventas que resultan y usan análisis para desarrollar un modelo predictivo de ventas en función de estas estrategias de decisión.

Por ejemplo, supongamos que un tendero que opera tres tiendas en una pequeña ciudad varió el precio, los cupones (sí, se representa con un valor de 1; no, se representa con un valor de 0) y los gastos publicitarios en un periódico local durante un período de 16 semanas y observó las siguientes ventas:

| Semana | Precio (\$) | Cupón (0,1) | Publicidad (\$) | Tienda 1 Ventas (Unidades) | Tienda 2 Ventas (Unidades) | Tienda 3 Ventas (Unidades) |
|--------|-------------|-------------|-----------------|----------------------------|----------------------------|----------------------------|
| 1 | 6.99 | 0 | 0 | 501 | 510 | 481 |
| 2 | 6.99 | 0 | 150 | 772 | 748 | 775 |
| 3 | 6.99 | 1 | 0 | 554 | 528 | 506 |
| 4 | 6.99 | 1 | 150 | 838 | 785 | 834 |
| 5 | 6.49 | 0 | 0 | 521 | 519 | 500 |
| 6 | 6.49 | 0 | 150 | 723 | 790 | 723 |

³ (Calantone, Cornelia, Litvack, & Benedetto, 1989).

| Semana | Precio (\$) | Cupón (0,1) | Publicidad (\$) | Tienda 1 Ventas (Unidades) | Tienda 2 Ventas (Unidades) | Tienda 3 Ventas (Unidades) |
|--------|-------------|-------------|-----------------|----------------------------|----------------------------|----------------------------|
| 7 | 6.49 | 1 | 0 | 510 | 556 | 520 |
| 8 | 6.49 | 1 | 150 | 818 | 773 | 800 |
| 9 | 7.59 | 0 | 0 | 479 | 491 | 486 |
| 10 | 7.59 | 0 | 150 | 825 | 822 | 757 |
| 11 | 7.59 | 1 | 0 | 533 | 513 | 540 |
| 12 | 7.59 | 1 | 150 | 839 | 791 | 832 |
| 13 | 5.49 | 0 | 0 | 484 | 480 | 508 |
| 14 | 5.49 | 0 | 150 | 686 | 683 | 708 |
| 15 | 5.49 | 1 | 0 | 543 | 531 | 530 |
| 16 | 5.49 | 1 | 150 | 767 | 743 | 779 |

Figura 96. Gastos publicitarios en 16 semanas

Para comprender mejor las relaciones entre el precio, los cupones y la publicidad, el tendero podría haber desarrollado el siguiente modelo utilizando herramientas de análisis de negocios:

$$\text{Ventas} = 500 - 0.05 * \text{precio} + 30 * \text{cupones} + 0.08 * \text{publicidad} + 0.25 * \text{precio} * \text{publicidad}$$

En este modelo, las variables de decisión son precio, cupones y publicidad. Los valores 500, -0.05, 30, 0.08 y 0.25 son efectos de los datos de entrada en el modelo, que se estiman a partir de los datos obtenidos del experimento. Reflejan el impacto en las ventas al cambiar las variables de decisión. Por ejemplo, un aumento en el precio de \$1 resulta en una disminución de 0.05 unidades en las ventas semanales; el uso de cupones resulta en un aumento de 30 unidades en las ventas semanales. En este ejemplo, no hay variables de entrada incontrolables. La salida del modelo son las unidades de venta del producto. Por ejemplo, si el precio es de \$6.99, no se ofrecen cupones y no se hace publicidad (el experimento correspondiente a la semana 1), el modelo estima las ventas como

$$\text{Ventas} = 500 - 0.05 * 6.99 + 30 * 0 + 0.08 * 0 + 0.25 * 6.99 * 0 = 500 \text{ unidades}$$

Se observa en la figura que las ventas reales en la semana 1 variaron entre 481 y 510 en las tres tiendas. Por lo tanto, este modelo predice una buena estimación para las ventas; sin embargo, no nos dice nada sobre la posible variabilidad o error de predicción. Sin embargo, el gerente puede utilizar este modelo para evaluar diferentes estrategias de precios, promoción y publicidad, y ayudar a elegir la mejor estrategia para maximizar las ventas o la rentabilidad.

Suposición del modelo

Todos los modelos se basan en suposiciones que reflejan la visión del modelador del “mundo real”. Algunas suposiciones se hacen para simplificar el modelo y hacerlo más manejable; es decir, que sea capaz de ser fácilmente analizado o resuelto. Se podrían hacer otras suposiciones para caracterizar mejor los datos históricos u observaciones pasadas. La tarea del modelador es seleccionar o crear un modelo adecuado que represente mejor el comportamiento de la situación real. Por ejemplo, la teoría económica nos dice que la demanda de un producto está negativamente relacionada con su precio. Por lo tanto, a medida que los precios aumentan, la demanda cae y viceversa (un fenómeno que puede reconocer como elasticidad de precios, la relación entre el cambio porcentual en la demanda y el cambio porcentual en el precio). Diferentes modelos matemáticos pueden describir este fenómeno.

Ejemplo 34. Un modelo de predicción lineal de la demanda

Un modelo simple para predecir la demanda en función del precio es

$$D = a - bP \quad (1.6)$$

donde D es la tasa de demanda, P es el precio unitario, a es una constante que estima la demanda cuando el precio es cero, y b es la pendiente de la

función de demanda. Este modelo es más aplicable cuando queremos predecir el efecto de pequeños cambios en torno al precio actual. Por ejemplo, se supone que cuando el precio es de \$100, la demanda es de 19,000 unidades y esa demanda cae en 10 por cada dólar de aumento de precio. Usando álgebra simple, podemos determinar que $a = 20,000$ y $b = 10$. Por lo tanto, si el precio es \$80, la demanda prevista es

$$D = 20000 - (10) * (80) = 19,200 \text{ unidades}$$

Si el precio aumenta a \$90, el modelo predice la demanda como

$$D = 20000 - (10) * (90) = 19,100 \text{ unidades}$$

Si el precio es de \$100, la demanda sería

$$D = 20000 - (10) * (100) = 19,000 \text{ unidades}$$

Y así sucesivamente, el gráfico de la demanda como función del precio se muestra en la figura 97, donde el precio varía entre \$80 y \$120. Vemos pequeños cambios alrededor del precio actual. Por ejemplo, hay una disminución constante en la demanda por cada aumento de \$10 en el precio, una característica de un modelo lineal.

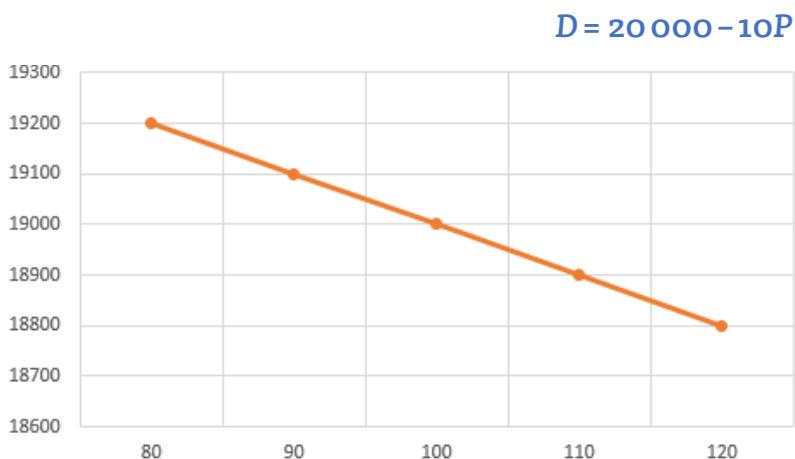


Figura 97. Gráfico de la demanda lineal Modelo $D = a - bP$

Ejemplo 35. Un modelo de predicción de la demanda no lineal

Un modelo alternativo supone que la elasticidad de los precios es constante. En este caso, el modelo apropiado es

$$D = cP^{-d} \quad (1.7)$$

donde, c es la demanda cuando el precio es 0 y $d > 0$ es la elasticidad del precio. Para ser coherentes con el ejemplo 1.9, suponemos que cuando el precio es cero, la demanda es de 20 000. Por lo tanto, $c = 20\,000$. También asumiremos, como en el ejemplo 34, que cuando el precio sea de 100 dólares, $D = 19\,000$. Usando estos valores en la ecuación 1.7, se puede determinar el valor para d (se puede hacer matemáticamente usando algoritmo log), esto es, $d = -0.0111382$. Por lo tanto, si el precio es 80, la demanda sería

$$D = 20\,000 (80)^{-0.0111382} = 19\,047$$

si el precio es 90, la demanda sería

$$D = 20\,000 (90)^{-0.0111382} = 19\,022$$

Si el precio es 100, la demanda es

$$D = 20\,000 (100)^{-0.0111382} = 19\,000.$$

En la figura 98 se muestra un gráfico de la demanda en función del precio. La demanda pronosticada cae de una manera ligera no lineal a medida que aumenta el precio. Por ejemplo, la demanda disminuye en 25 unidades cuando el precio aumenta de \$80 a \$90, pero solo en 22 unidades cuando el precio aumenta de \$90 a \$100. Si el precio aumenta a \$100, se vería una disminución menor en la demanda. Por lo tanto, se observa una relación no lineal en contraste con la del ejemplo 34.

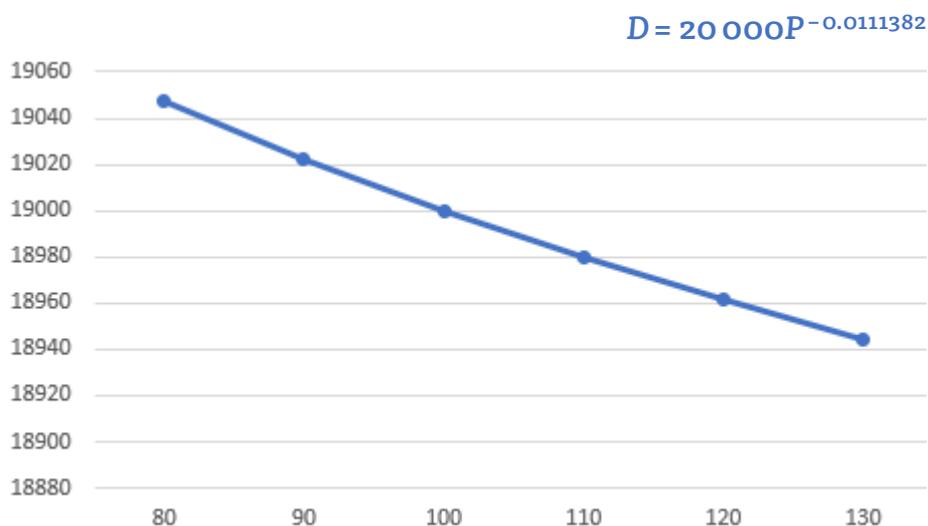


Figura 98. Gráfico del Modelo de demanda no lineal $D = cP^{-d}$

Ambos modelos de los ejemplos 34 y 35 hacen predicciones diferentes de la demanda para diferentes precios (excepto \$90). ¿Qué modelo es el mejor? La respuesta puede ser ninguna de las dos. En primer lugar, el desarrollo de modelos realistas requiere muchos cambios de punto de precio dentro de un experimento cuidadosamente diseñado. En segundo lugar, también debe incluir datos sobre la competencia y la renta disponible de los clientes, que son difíciles de determinar. Sin embargo, es posible desarrollar modelos de elasticidad de precios con rangos de precios limitados y segmentos de clientes estrechos. Un buen punto de partida sería crear una base de datos histórica con información detallada sobre todas las acciones de precios anteriores.

Incertidumbre y riesgo

Como se sabe, el futuro siempre es incierto. Por lo tanto, muchos modelos predictivos incorporan incertidumbre y ayudan a los tomadores de decisiones a analizar los riesgos asociados con sus decisiones. La incertidumbre es el conocimiento imperfecto de lo que sucederá; el riesgo se asocia con las consecuencias y la probabilidad de lo que podría suceder. Por ejemplo, el cambio en el precio de las acciones de Apple en el día siguiente

de la negociación es incierto. Sin embargo, si usted es dueño de acciones de Apple, entonces se enfrenta al riesgo de perder dinero si el precio de las acciones cae. Si usted no posee ninguna acción, el precio sigue siendo incierto, aunque no tendría ningún riesgo. El riesgo se evalúa por la magnitud de las consecuencias y la probabilidad de que ocurran. Por ejemplo, una caída del 10% en el precio de las acciones incurriría en un riesgo mayor si se posee \$1 millón que si solo se tiene \$1,000. Del mismo modo, si las posibilidades de una caída del 10% fueran 1 en 5, el riesgo sería mayor que si las posibilidades fueran solo 1 de cada 100.

La importancia del riesgo en los negocios ha sido reconocida durante mucho tiempo. El renombrado escritor de gestión, Peter Drucker, observó en 1974:

“Tratar de eliminar el riesgo en las empresas es inútil. El riesgo es la inherente a la asignación de recursos actuales para cumplir con demandas futuras. De hecho, el progreso económico puede definirse como la capacidad de asumir mayores riesgos. El intento de eliminar los riesgos, incluso el intento de minimizarlos sólo puede hacerlos irracionales e insoportables. Sólo puede resultar en el mayor riesgo de todos: rigidez”⁴.

La consideración del riesgo es un elemento vital de la toma de decisiones. Por ejemplo, probablemente no se elegiría una inversión simplemente sobre la base del rendimiento que podría esperar porque, por lo general, los rendimientos más altos están asociados con un mayor riesgo. Por lo tanto, se tiene que hacer un equilibrio entre los beneficios de mayores recompensas y los riesgos de pérdidas potenciales. Los modelos analíticos pueden ayudar a evaluar esto. Este tema se abordará en capítulos posteriores.

⁴ Drucker, 1974.

Modelos de decisión prescriptiva

Un modelo de decisión prescriptivo ayuda a los responsables de la toma de decisiones a identificar la mejor solución a un problema de decisión. La optimización es el proceso de encontrar un conjunto de valores para variables de decisión que minimicen o maximicen cierta cantidad de interés (beneficio, ingresos, costo, tiempo, etc.) llamados la función objetiva. Cualquier conjunto de variables de decisión que optimice la función objetivo se denomina solución óptima. En un mundo altamente competitivo donde un punto porcentual puede significar una diferencia de cientos de miles de dólares o más, conocer la mejor solución puede significar la diferencia entre el éxito y el fracaso.

Ejemplo 36. Un modelo prescriptivo para los precios

Para ilustrar un ejemplo de un modelo prescriptivo, se supone que una empresa desea determinar los mejores precios para uno de sus productos para maximizar los ingresos durante el próximo año. Un estudio de mercado ha recopilado datos que estiman las ventas anuales esperadas para diferentes niveles de precios. Los analistas determinaron que las ventas se pueden expresar mediante el siguiente modelo:

$$\text{ventas} = -2.9485 * \text{precio} + 3240.9$$

Debido a que los ingresos son iguales a las ventas de precio, un modelo para los ingresos totales es

$$\text{Ingresos totales} = \text{precio} * \text{ventas}$$

$$\text{Ingresos totales} = \text{precio} * (-2.9485 * \text{precio} + 3240.9)$$

$$\text{Ingresos totales} = -2.9485 * \text{precio}^2 + 3240.9 * \text{precio}$$

La firma desea identificar el precio que maximiza los ingresos totales. Una manera de hacerlo sería probar diferentes precios y buscar el que produce

los ingresos totales más altos. Esto sería bastante tedioso para hacer a mano o incluso con una calculadora.

Aunque el modelo de precios no lo hizo, la mayoría de los modelos de optimización tienen restricciones: limitaciones, requisitos u otras restricciones que se imponen a cualquier solución, como “no excedan el presupuesto permitido” o “garantizan que se cumple toda la demanda”. Por ejemplo, un gerente de la empresa de productos de consumo probablemente desearía asegurarse de que se logra un nivel específico de servicio al cliente con el rediseño del sistema de distribución. La presencia de restricciones hace que el modelado y la resolución de problemas de optimización sean más difíciles.

Para algunos modelos prescriptivos, las soluciones analíticas (expresiones matemáticas de forma cerrada o fórmulas simples) se pueden obtener utilizando técnicas como el cálculo u otros tipos de análisis matemáticos. En la mayoría de los casos, sin embargo, se necesita algún tipo de procedimiento basado en el equipo para encontrar una solución óptima. Un algoritmo es un procedimiento sistemático que encuentra una solución a un problema. Los investigadores han desarrollado algoritmos eficaces para resolver muchos tipos de problemas de optimización. Por ejemplo, Microsoft Excel® tiene un complemento integrado llamado Solver que le permite encontrar soluciones óptimas a los problemas de optimización formulados como modelos de hoja de cálculo. Se utiliza Solver en capítulos posteriores. Sin embargo, no nos preocuparemos por la mecánica detallada de estos algoritmos; nos centraremos en el uso de los algoritmos para resolver y analizar los modelos que desarrollamos.

Si es posible, nos gustaría asegurarnos de que un algoritmo como el que utiliza Solver encuentre la mejor solución. Sin embargo, algunos modelos son tan complejos que es imposible resolverlos de manera óptima en una cantidad razonable de tiempo de computadora debido a la cantidad extremadamente grande de cálculos que pueden ser necesarios o porque son tan complejos que no se puede encontrar la mejor solución garantizada.

En estos casos, los analistas utilizan algoritmos de búsqueda, procedimientos de solución que generalmente encuentran buenas soluciones sin garantías de encontrar la mejor. Existen potentes algoritmos de búsqueda para obtener buenas soluciones a problemas de optimización extremadamente difíciles.

Los modelos de decisión prescriptivos pueden ser deterministas o estocásticos. Un modelo determinista es aquel en el que toda la información de entrada del modelo se conoce o se supone que se conoce con certeza. Un modelo estocástico es aquel en el que parte de la información de entrada del modelo es incierta. Por ejemplo, supongamos que la demanda del cliente es un elemento importante de algún modelo. Se puede suponer que la demanda se conoce con certeza; por ejemplo, 5,000 unidades al mes. En este caso estaríamos tratando con un modelo determinista. Por otro lado, supongamos que tenemos pruebas que indican que la demanda es incierta, con un valor medio de 5 000 unidades al mes, pero que suele variar entre 3 200 y 6 800 unidades. Si se hace esta suposición, se estaría tratando con un modelo estocástico.

4.9 Comunicar los datos

Consultas de datos: tablas, ordenación y filtrado

Los administradores realizan numerosas consultas sobre los datos. Por ejemplo, en la base de datos Órdenes de compra, es posible que estén interesados en buscar todos los pedidos de un proveedor determinado, todos los pedidos de un artículo determinado o los pedidos de seguimiento por datos de pedido. Para abordar estas consultas, se necesita ordenar los datos de alguna manera. En otros casos, los administradores podrían estar interesados en extraer un conjunto de registros que tengan ciertas características. Esto se denomina filtrado de los datos. Por ejemplo, en la base de datos Pedidos de compra, un administrador podría estar interesado en extraer todos los registros correspondientes a un artículo determinado.

Excel® proporciona una forma conveniente de dar formato a las bases de datos para facilitar el análisis, se denominan tablas.

Ejemplo 37. Creación de una tabla de Excel®

Usaremos el archivo de Datos de riesgo crediticio para ilustrar una tabla de Excel®. En primer lugar, seleccione el rango de los datos, incluidos los encabezados (un acceso directo útil es seleccionar la primera celda en la esquina superior izquierda, luego haga clic en Ctrl+Mayús+flecha_abajo y luego Ctrl+Mayús+flecha_derecha). A continuación, haga clic en Tabla en el grupo Tablas de la ficha Insertar y asegúrese de que la casilla Mi tabla tenga encabezados y esté marcada. (También puede seleccionar una celda dentro de la tabla y luego hacer clic en Tabla desde el menú Insertar. Excel® elegirá el rango de tablas para que lo verifique.) El rango de tablas ahora se formateará y continuará automáticamente cuando se introduzcan nuevos datos. La figura 99 muestra una parte del resultado. Tenga en cuenta que las filas están sombreadas y que cada encabezado de columna tiene una flecha desplegable para filtrar los datos (lo discutiremos en breve). Si hace clic dentro de una tabla, la pestaña Diseño de herramientas de tabla aparecerá en la cinta de opciones, lo que le permitirá hacer una variedad de cosas, como cambiar el esquema de color, eliminar duplicados, cambiar el formato, etc.

| | E | F | G | H | I |
|-----|----------------|--|------------|------|--------|
| rus | \$770,339 | -SUMA(Tabla1[Ahorro]) | | | |
| pia | 292 | =CONTAR.SI(Tabla1[Vivienda], "Propia") | | | |
| | Meses empleado | Gene | Estado ci | Edad | Vivien |
| 13 | 12 | M | Soltero | 23 | Propi |
| 25 | 0 | M | Divorciado | 32 | Propi |
| 19 | 119 | M | Soltero | 38 | Propi |
| 13 | 14 | M | Soltero | 36 | Propi |
| 40 | 45 | M | Soltero | 31 | Rentad |
| 11 | 13 | M | Casado | 25 | Propi |
| 13 | 16 | M | Casado | 26 | Propi |

Figura 99. Porción de datos de riesgo de crédito formateados como una tabla de Excel®

Una tabla de Excel® le permite utilizar referencias de tabla para realizar cálculos básicos, como se muestra en el ejemplo siguiente.



Ejemplo 38. Cálculos basados en tablas

Supongamos que en la tabla Datos de riesgo de crédito se desea calcular la cantidad total de ahorros en la columna C. Se podría, por supuesto, simplemente utilizar la función =SUMA(C4:C428). Sin embargo, con una opción de “Tabla”, se podría usar la fórmula =SUMA(Tabla1[Ahorro])

El nombre de la tabla, Tabla1, se puede encontrar (y cambiar) en el grupo Propiedades de la pestaña Diseño de herramientas de tabla. Una de las ventajas de hacer esto es que, si agregamos nuevos registros a la tabla, el cálculo se actualizará automáticamente, y no tenemos que cambiar el rango en la fórmula u obtener un resultado incorrecto si nos olvidamos. Como otro ejemplo, se podría encontrar el número de propietarios con vivienda propia usando la función =CONTAR.SI(Tabla1[Vivienda],”Propia”)

| fx | | =SUMA(Tabla1[Ahorro]) | | | | | | |
|-------|---------|--|---------------|---------------------------------------|------------|------|--------|--|
| B | C | D | E | F | G | H | I | |
| | | Suma total de ahorros | \$770,339 | =SUMA(Tabla1[Ahorro]) | | | | |
| | | Número de propietarios con vivienda propia | 292 | =CONTAR.SI(Tabla1[Vivienda],”Propia”) | | | | |
| Cheq | Ahor | Meses de cliente | Meses emplead | Gene | Estado ci | Edad | Vivien | |
| \$0 | \$739 | 13 | 12 | M | Soltero | 23 | Propia | |
| \$0 | \$1,230 | 25 | 0 | M | Divorciado | 32 | Propia | |
| \$0 | \$389 | 19 | 119 | M | Soltero | 38 | Propia | |
| \$638 | \$347 | 13 | 14 | M | Soltero | 36 | Propia | |

Figura 100. Uso de función Tabla en Excel®

Si se agregan registros adicionales al final de la tabla, se incluirán y calcularán automáticamente, y si crea un gráfico basado en los datos, el gráfico se actualizará automáticamente si se agregan nuevos registros.

Clasificación de datos en Excel®

Excel® proporciona muchas maneras de ordenar las listas por filas o columna o en orden ascendente o descendente y mediante esquemas de ordenación personalizados. Los botones de ordenación en Excel® se pueden encontrar debajo de la pestaña Inicio en el grupo Edición (consulte la figura 101). Se selecciona una sola celda en la columna en la que se desea ordenar y se hace clic en el botón “AZ↑” para ordenar de menor a mayor o el botón “ZA↓” para ordenar de mayor a menor. También se puede hacer clic en el botón Ordenar para especificar los criterios para las capacidades de clasificación más avanzadas

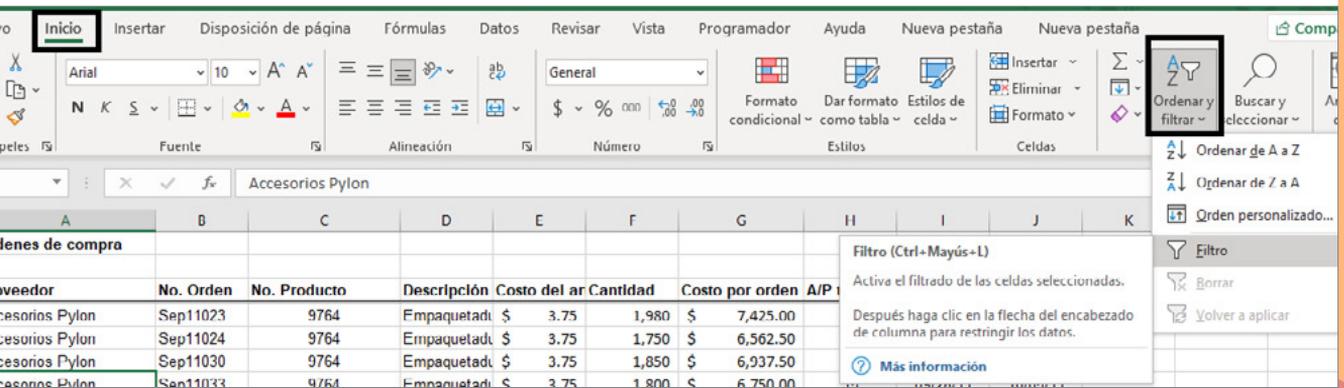


Figura 101. Pestaña de datos para función “Ordenar y filtrar”

Ejemplo 39. Clasificación de datos en la base de datos de órdenes de compra

Para la base de datos de Órdenes de compra, supongamos que deseamos ordenar los datos por proveedor. Haga clic en cualquier celda de la columna A de los datos (pero no en la celda de encabezado A3) y luego en el botón “ZA↓” en la pestaña Datos. Esto le permite identificar fácilmente los registros que corresponden a todos los pedidos de un proveedor en particular.

| | A | B | C | D | E | F | G | H | I |
|----|--------------------------|------------------|---------------------|--------------------|---------------------|-----------------|------------------------|--------------------|--------------------|
| 1 | Ordenes de compra | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | Proveedor | No. Orden | No. Producto | Descripción | Costo del ar | Cantidad | Costo por orden | A/P termino | Fecha de or |
| 4 | Accesorios Pylon | Sep11023 | 9764 | Empaquetadu | \$ 3.75 | 1,980 | \$ 7,425.00 | 15 | 09/20/11 |
| 5 | Accesorios Pylon | Sep11024 | 9764 | Empaquetadu | \$ 3.75 | 1,750 | \$ 6,562.50 | 15 | 09/20/11 |
| 6 | Accesorios Pylon | Sep11030 | 9764 | Empaquetadu | \$ 3.75 | 1,850 | \$ 6,937.50 | 15 | 09/25/11 |
| 7 | Accesorios Pylon | Sep11033 | 9764 | Empaquetadu | \$ 3.75 | 1,800 | \$ 6,750.00 | 15 | 09/28/11 |
| 8 | Accesorios Pylon | Oct11004 | 6433 | Empaque orir | \$ 2.95 | 1,500 | \$ 4,425.00 | 15 | 10/01/11 |
| 9 | Aeroespacial Fast-Tie | Aug11003 | 5462 | Cable blindad | \$ 1.05 | 23,000 | \$ 24,150.00 | 30 | 08/10/11 |
| 10 | Aeroespacial Fast-Tie | Aug11004 | 5462 | Cable blindad | \$ 1.05 | 21,500 | \$ 22,575.00 | 30 | 08/15/11 |
| 11 | Aeroespacial Fast-Tie | Aug11006 | 5462 | Cable blindad | \$ 1.05 | 22,500 | \$ 23,625.00 | 30 | 08/20/11 |
| 12 | Aeroespacial Fast-Tie | Aug11009 | 6321 | Empaque orir | \$ 2.45 | 1,300 | \$ 3,185.00 | 30 | 08/25/11 |
| 13 | Aeroespacial Fast-Tie | Aug11010 | 5462 | Cable blindad | \$ 1.05 | 22,500 | \$ 23,625.00 | 30 | 08/25/11 |

Figura 102. Parte de la base de datos de pedidos de compra ordenada por nombre de proveedor

Análisis de Pareto

El análisis de Pareto es un término que lleva el nombre de un economista italiano, Vilfredo Pareto, quien, en 1906, observó que una gran proporción de la riqueza en Italia era propiedad de una proporción relativamente pequeña del pueblo. El principio Pareto se ve a menudo en muchas situaciones de negocios. Por ejemplo, un gran porcentaje de las ventas suele provenir de un pequeño porcentaje de clientes, un gran porcentaje de defectos de calidad proviene de solo un par de fuentes, o un gran porcentaje del valor de inventario corresponde a un pequeño porcentaje de artículos. Como resultado, el principio de Pareto también se llama a menudo la “regla 80-20”, refiriéndose a la situación genérica en la que el 80% de algunas salidas provienen del 20% de algunas entradas. Un análisis de Pareto se basa en la clasificación de datos y el cálculo del porcentaje acumulado de la característica de interés.



Ejemplo 40. Aplicando el principio Pareto

El archivo de Excel® Inventario de bicicletas enumera el inventario de modelos de bicicletas en una tienda de artículos deportivos (consulte las

columnas A a F en la figura 103). Para realizar un análisis de Pareto, primero calculamos el valor de inventario de cada producto multiplicando la cantidad disponible por el costo de compra; este es el importe invertido en los artículos que están actualmente en stock. A continuación, ordenamos los datos en orden decreciente del valor de inventario y calculamos el porcentaje del valor total del inventario para cada producto y el porcentaje acumulado. En las columnas G a I en la figura 103, se observa que alrededor del 75% del valor de inventario está representado por menos del 40% (9 de 24) de los artículos. Si estos inventarios de alto valor no se venden bien, es posible que el gerente de la tienda desee mantener menos en stock.

| | A | B | C | D | E | F | G | H | I | J | K |
|----|-----------|---------------------------------|----------------------------|------------------------|------------------------|-----------------------|----------------------------|-------------------------------|-----------------------------|-------------------|------------------|
| 1 | | Inventario de bicicletas | | | | | | | | | |
| 2 | | | | | | | | | | | |
| 3 | ID | Categoría de producto | Nombre del producto | Costo de compra | Precio de venta | Proveedor | Cantidad disponible | Valor de los artículos | Valor del inventario | Porcentaje | Acumulado |
| 4 | 20 | Road | Runroad 5000 | \$450.95 | \$599.99 | Run-Up Bikes | 5 | 0.041666667 | \$2,254.75 | 0.111880107 | 0.111880107 |
| 5 | 18 | Road | Runroad 1000 | \$250.95 | \$350.99 | Run-Up Bikes | 8 | 0.083333333 | \$2,007.60 | 0.099616588 | 0.211496695 |
| 6 | 11 | Road | Elegant 210 | \$281.52 | \$394.13 | Bicyclist's Choice | 7 | 0.125 | \$1,970.64 | 0.097782643 | 0.309279338 |
| 7 | 19 | Road | Runroad 4000 | \$390.95 | \$495.99 | Run-Up Bikes | 5 | 0.166666667 | \$1,954.75 | 0.096994185 | 0.406273523 |
| 8 | 9 | Mtn. | Eagle 3 | \$350.52 | \$490.73 | Bike-One | 5 | 0.208333333 | \$1,752.60 | 0.086963555 | 0.493237078 |
| 9 | 5 | Road | Classic 109 | \$207.49 | \$290.49 | Bicyclist's Choice | 7 | 0.25 | \$1,452.43 | 0.072069198 | 0.565306275 |
| 10 | 10 | Hybrid | Eagle 7 | \$150.89 | \$211.46 | Bike-One | 9 | 0.291666667 | \$1,358.01 | 0.067384102 | 0.632690377 |
| 11 | 23 | Hybrid | Tea for Two | \$429.02 | \$609.00 | Simpson's Bike Supply | 3 | 0.333333333 | \$1,287.06 | 0.063863581 | 0.696553959 |
| 12 | 2 | Mtn. | Bluff Breaker | \$375.00 | \$495.00 | The Bike Path | 3 | 0.375 | \$1,125.00 | 0.055822207 | 0.752376165 |
| 13 | 8 | Mtn. | Eagle 2 | \$401.11 | \$561.54 | Bike-One | 2 | 0.416666667 | \$802.22 | 0.039805947 | 0.792182112 |
| 14 | 4 | Leisure | Breeze LE | \$109.95 | \$149.95 | The Bike Path | 5 | 0.458333333 | \$549.75 | 0.027278452 | 0.819460564 |
| 15 | 17 | Children | Runkidder 100 | \$50.95 | \$75.99 | Run-Up Bikes | 10 | 0.5 | \$509.50 | 0.025281257 | 0.844741821 |
| 16 | 13 | Mtn. | Jetty Breaker | \$455.95 | \$649.95 | The Bike Path | 1 | 0.541666667 | \$455.95 | 0.02262412 | 0.867365941 |
| 17 | 16 | Leisure | Runcool 3000 | \$85.95 | \$135.99 | Run-Up Bikes | 5 | 0.583333333 | \$429.75 | 0.021324083 | 0.888690024 |
| 18 | 6 | Children | Coolest 100 | \$69.99 | \$97.98 | Bicyclist's Choice | 6 | 0.625 | \$419.94 | 0.020837313 | 0.909527337 |
| 19 | 7 | Mtn. | Eagle 1 | \$410.01 | \$574.01 | Bike-One | 1 | 0.666666667 | \$410.01 | 0.020344589 | 0.929871926 |
| 20 | 12 | Children | Green Rider | \$95.47 | \$133.66 | Simpson's Bike Supply | 4 | 0.708333333 | \$381.88 | 0.018948786 | 0.948820712 |
| 21 | 3 | Leisure | Breeze | \$89.95 | \$130.95 | The Bike Path | 4 | 0.75 | \$359.80 | 0.017853182 | 0.966673895 |
| 22 | 1 | Leisure | Blue Moon | \$75.29 | \$105.41 | Simpson's Bike Supply | 4 | 0.791666667 | \$301.16 | 0.014943481 | 0.981617375 |
| 23 | 22 | Leisure | Supreme 350 | \$50.00 | \$70.00 | Bicyclist's Choice | 3 | 0.833333333 | \$150.00 | 0.007442961 | 0.989060336 |
| 24 | 14 | Children | Red Rider | \$15.00 | \$25.50 | Simpson's Bike Supply | 8 | 0.875 | \$120.00 | 0.005954369 | 0.995014705 |
| 25 | 21 | Leisure | Starlight | \$100.47 | \$140.66 | Simpson's Bike Supply | 1 | 0.916666667 | \$100.47 | 0.004985295 | 1 |
| 26 | 15 | Hybrid | Runblend 2000 | \$180.95 | \$255.99 | Run-Up Bikes | 0 | 0.958333333 | \$0.00 | 0 | 1 |
| 27 | 24 | Road | Twist & Shout | \$490.50 | \$635.70 | Simpson's Bike Supply | 0 | 1 | \$0.00 | 0 | 1 |
| 28 | | | | | | | | Total | \$20,153.27 | 1 | |

Figura 103. Análisis de Pareto de la bicicleta

Exploración de datos mediante tablas dinámicas

Excel® proporciona una poderosa herramienta para destilar un conjunto de datos complejo en información significativa: tablas dinámicas. Las tablas dinámicas permiten crear resúmenes personalizados y gráficos de información clave en los datos, las que se pueden usar para crear rápidamente tabulaciones cruzadas y para profundizar en un gran conjunto de datos de muchas maneras.

Para aplicar tablas dinámicas, se necesita un conjunto de datos con etiquetas de columna en la primera fila, similar a los archivos de datos que hemos estado utilizando. Se selecciona cualquier celda del conjunto de datos y se elige Tabla dinámica en el grupo Tablas de la pestaña Insertar y se siguen los pasos del asistente. Excel® primero pide que se seleccione una tabla o rango de datos; si se hace clic en cualquier celda dentro de la matriz de datos antes de insertar una tabla dinámica, Excel® se establecerá de forma predeterminada en el rango completo de los datos. Se puede colocar la tabla dinámica en una nueva hoja de cálculo o en un rango en blanco de la hoja de cálculo existente. Excel®, a continuación, crea una tabla dinámica en blanco, como se muestra en la figura 104.

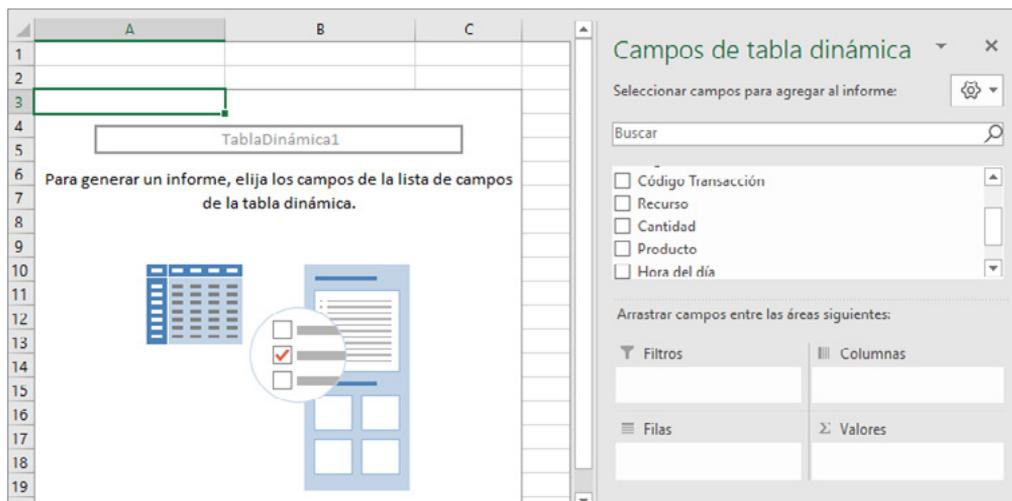


Figura 104. Tabla dinámica en blanco

La lista de campos de la tabla dinámica que se observa en el lado derecho de la figura 104 es una lista de los campos que corresponden a los encabezados en el archivo de datos. Se selecciona la que se desea incluir, ya sea como etiquetas de fila, etiquetas de columna, valores o lo que se denomina un filtro de informe. En primer lugar, se debe decidir qué tipos de tablas se van a crear, es decir, qué campos se desean para las filas, columnas y valores de datos.

Ejemplo 41. Creación de una tabla dinámica

Vamos a crear una tabulación cruzada de las ventas regionales por producto, como se hizo en la sección anterior. Se debe seleccionar la tabla donde se encuentran los datos. Posteriormente se selecciona la pestaña del menú de Excel, “Insertar”. En la pestaña de herramienta “Tabla” se selecciona el icono “Pivot Table” o “Crear tabla dinámica”. Posteriormente, aparece una ventana donde se seleccionan los datos que se desean analizar, de manera predefinida aparece la selección realizada. También se debe elegir dónde se desea colocar el informe de la tabla dinámica, de manera predefinida aparece “Nueva hoja de cálculo”. Finalmente, se puede elegir si se quiere analizar varias tablas. Ya ingresada la información deseada se selecciona “Aceptar” y aparece una hoja nueva: del lado derecho se tiene un menú desplegable titulado “Campos de tabla dinámica”; del lado izquierdo aparece un cuadro con título “TablaDinámica1” y la leyenda “Para genera un informe, elija los campos de la lista de campos de la tabla dinámica.

En los campos de tabla dinámica se selecciona la etiqueta “Region” con el botón derecho del ratón (mouse), sin dejar de seleccionar el botón, trasladamos el texto al campo de la tabla dinámica “Filas” y soltamos el botón del ratón. Se debe observar la etiqueta “Region” dentro del cuadro de campos “Filas” y se actualiza en la hoja de Excel la información de la tabla dinámica en las etiquetas de filas disponibles (para este ejemplo son: Este, Norte, Oeste y Sur).

Posteriormente, se busca la etiqueta “Producto” y se selecciona con el botón derecho del ratón (mouse), sin dejar de seleccionar el botón, trasladamos

el texto al campo de la tabla dinámica “Columnas” y soltamos el botón del ratón. Se debe observar la etiqueta “Producto” dentro del cuadro de campos “Columnas” y se actualiza en la hoja de Excel la información de la tabla dinámica en las etiquetas de columna (para este ejemplo son: DVD y Libro).

Finalmente, se ingresa la información a ser representada dentro de la tabla: se busca la etiqueta “ID cliente” y se selecciona con el botón derecho del ratón (mouse), sin dejar de seleccionar el botón, trasladamos el texto al campo de la tabla dinámica “Valores” y soltamos el botón del ratón. Se debe observar en la hoja de Excel la suma de productos que corresponden a Región y Producto. De manera predeterminada se realiza el cálculo de la suma. Si seleccionamos el campo Región de la lista de campos de la tabla dinámica en la figura 105 al área Etiquetas de fila, el campo Producto en el área Etiquetas de columna y cualquiera de los otros campos, como ID del cliente, en el área Valores, creará la tabla dinámica que se muestra en la figura.

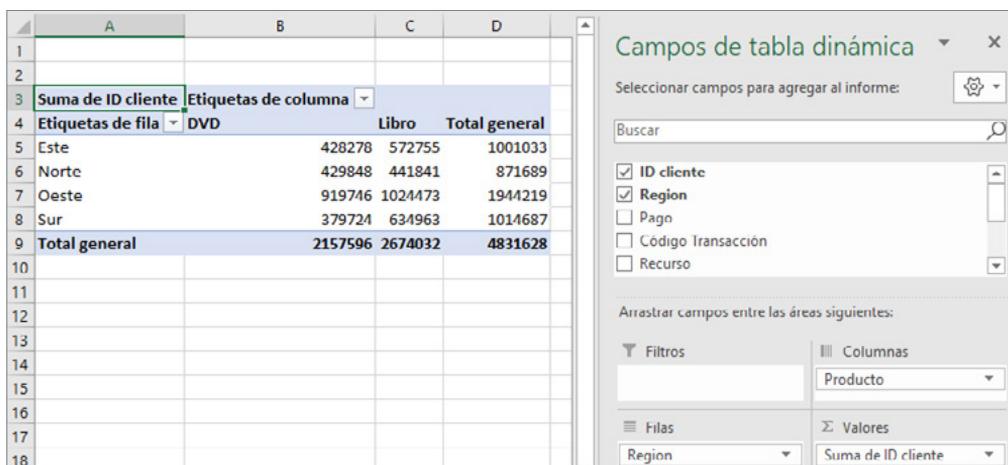
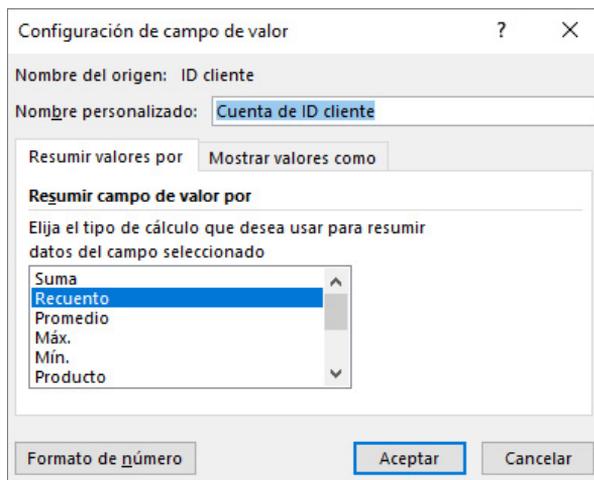


Figura 105. Tabla dinámica predeterminada para ventas regionales por producto

Sin embargo, la suma de los valores de ID de cliente (el valor predeterminado) no tiene sentido; simplemente queremos un recuento del número de registros en cada categoría. Haga clic en la pestaña Analizar y, a continuación, en el grupo Campo activo y elija Configuración de campo. Podrá cambiar el método de integración en la tabla dinámica del cuadro de diálogo Configuración de campo de valor que se muestra en la figura 106.

Figura 106. Cuadro de diálogo Configuración de campo de valor de campo de valor



Las opciones de Configuración de campo de valor de la figura 106 incluyen opciones como Promedio, Máximo, Mínimo y otras medidas estadísticas que se analizarán en el capítulo siguiente. También permiten formatear los datos correctamente (por ejemplo, moneda o para mostrar un número fijo de decimales) haciendo clic en el botón Formato de número.

La selección de Recuento de resultados en la tabla dinámica es la tabulación cruzada, como se muestra en la figura 107.

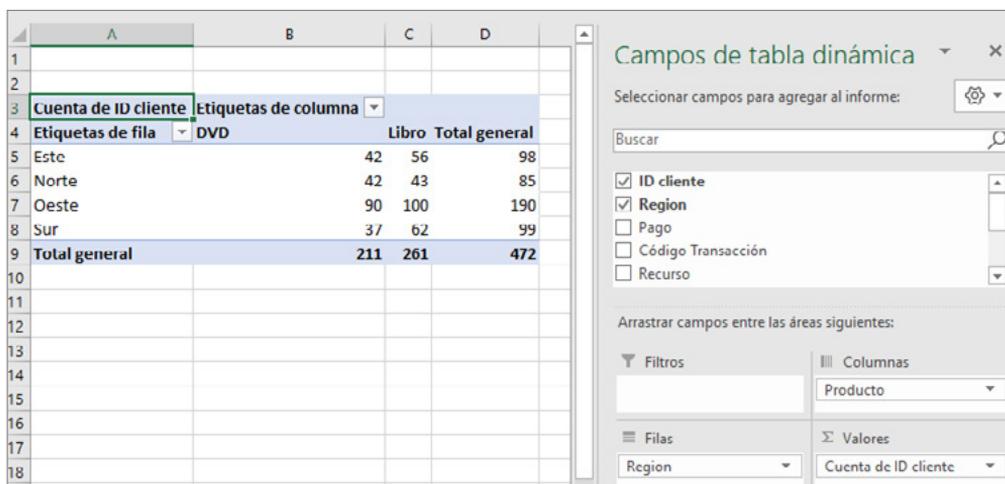


Figura 107. Tabla dinámica para el recuento de ventas regionales por producto

La ventaja del uso de las tablas dinámicas es que, si se desea cambiar el análisis, simplemente se puede desmarcar las casillas de la lista de campos

de tabla dinámica o arrastrar los nombres de campo a diferentes áreas, también se pueden agregar fácilmente varias variables en los campos para crear diferentes vistas de los datos. Por ejemplo, si arrastra el campo Origen al área Etiquetas de fila, creará la tabla dinámica que se muestra en la figura 108. Esto muestra un recuento del número de ventas por región y producto que también se desglosa por cómo se realizaron los pedidos, ya sea por correo electrónico o en la Web.

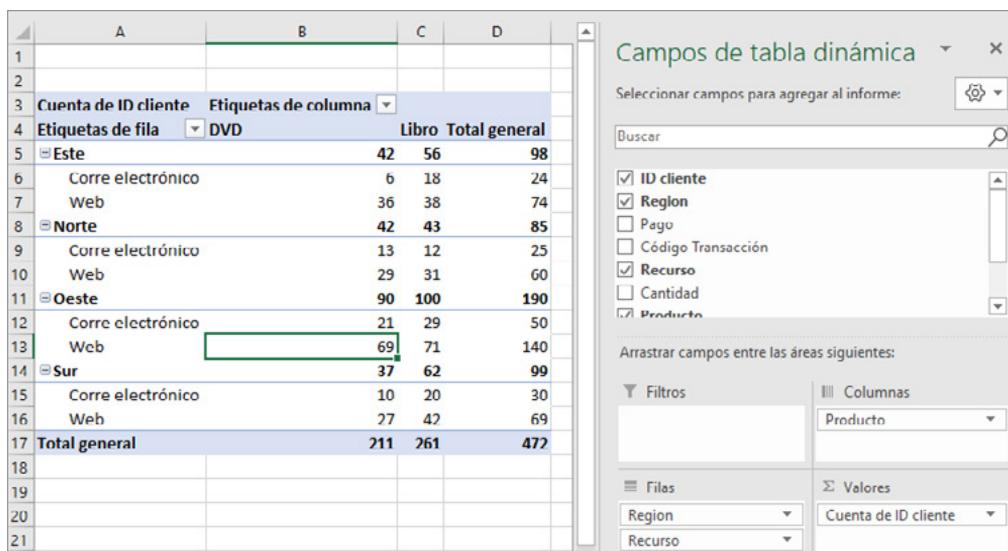


Figura 108. Tabla dinámica para ventas por región, producto y origen de pedidos

Arrastrar un campo al área Filtro de informe de la lista Campo de tabla dinámica permite agregar una tercera dimensión al análisis. El ejemplo 42 ilustra esto. Se pueden crear otras tablas dinámicas sin repetir todos los pasos del asistente. Simplemente copie y pegue la primera tabla. La mejor manera de aprender acerca de las tablas dinámicas es simplemente experimentar con ellas.



Ejemplo 42. Uso del filtro de informe de tabla dinámica

Volviendo a la tabla dinámica de tabulación cruzada de ventas regionales por producto, arrastre el campo Pago al área Filtro de informe. Esto coloca

el pago en la fila 1 de la tabla dinámica y permite desglosar la tabulación cruzada por tipo de pago, como se muestra en la figura 109.

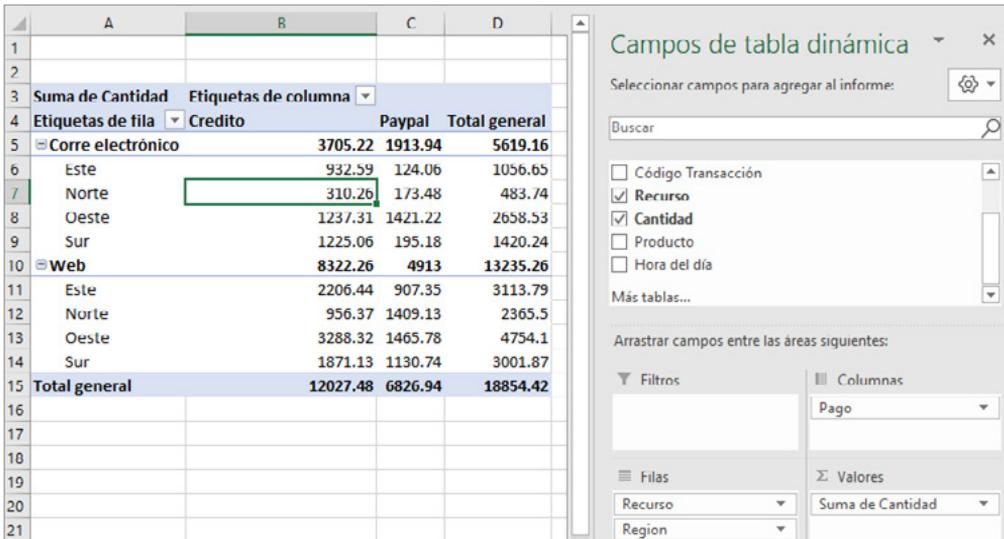


Figura 109. Tabulación cruzada en tabla dinámica

Haga clic en la flecha desplegable en la fila 1, y puede elegir mostrar una tabulación cruzada para uno de los diferentes tipos de pago, Crédito o Paypal. La figura 110 muestra los resultados de los pagos con tarjeta de crédito que representaron 299 del número total de transacciones.

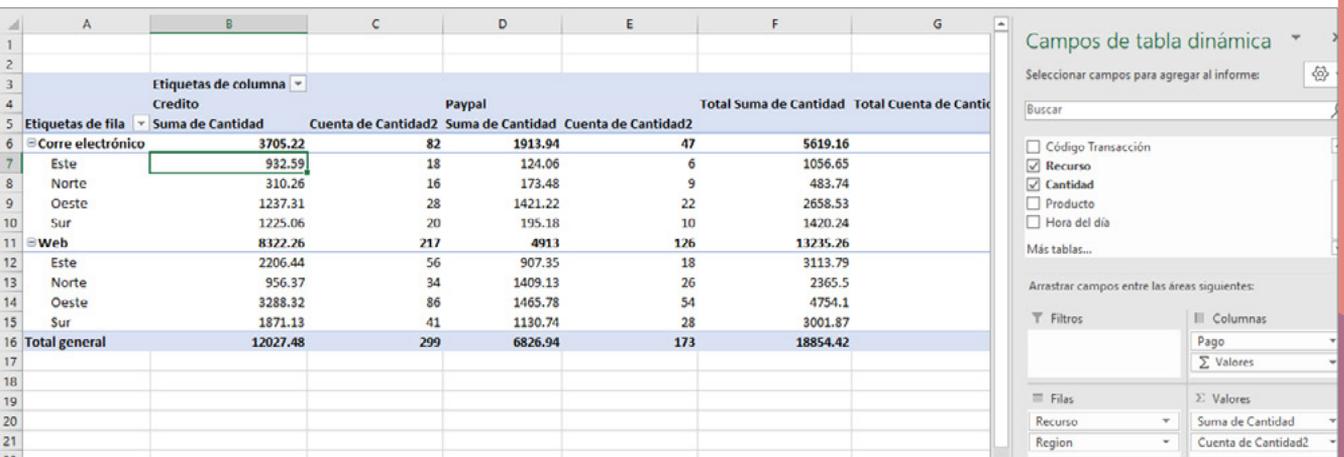


Figura 110. Tabulación cruzada con información de cálculos diferentes en valores

Gráficos dinámicos

Microsoft Excel® proporciona una forma sencilla de crear gráficos dinámicos con un solo clic para visualizar datos en tablas dinámicas. Para mostrar un gráfico dinámico para una tabla dinámica, se selecciona primero la tabla dinámica. En la pestaña Analizar, se hace clic en Gráfico dinámico. Excel® mostrará un cuadro de diálogo Insertar gráfico que permite elegir el tipo de gráfico que desea mostrar.

Ejemplo 43. Gráfico dinámico para datos de ventas

Para la tabla dinámica que se muestra en la figura 111, elegimos mostrar un gráfico de columnas desde el cuadro de diálogo Insertar gráfico. La figura 11 muestra el gráfico generado por Excel®. Al hacer clic en los botones desplegable, se pueden cambiar fácilmente los datos que se muestran filtrando los datos. Además, haciendo clic en el gráfico y seleccionando la pestaña Diseño de herramientas de gráfico dinámico, se pueden cambiar las filas y columnas para mostrar una vista alternativa del gráfico o cambiar el tipo de gráfico por completo.

| Etiquetas de fila | Etiquetas de columna | Libro | Total general | |
|----------------------|----------------------|-------|---------------|-----|
| Este | | 42 | 56 | 98 |
| Corre electrónico | | 6 | 18 | 24 |
| Web | | 36 | 38 | 74 |
| Norte | | 42 | 43 | 85 |
| Corre electrónico | | 13 | 12 | 25 |
| Web | | 29 | 31 | 60 |
| Oeste | | 90 | 100 | 190 |
| Corre electrónico | | 21 | 29 | 50 |
| Web | | 69 | 71 | 140 |
| Sur | | 37 | 62 | 99 |
| Corre electrónico | | 10 | 20 | 30 |
| Web | | 27 | 42 | 69 |
| Total general | | 211 | 261 | 472 |

Figura 111. Tabla dinámica para ventas por región, producto y fuente de pedidos

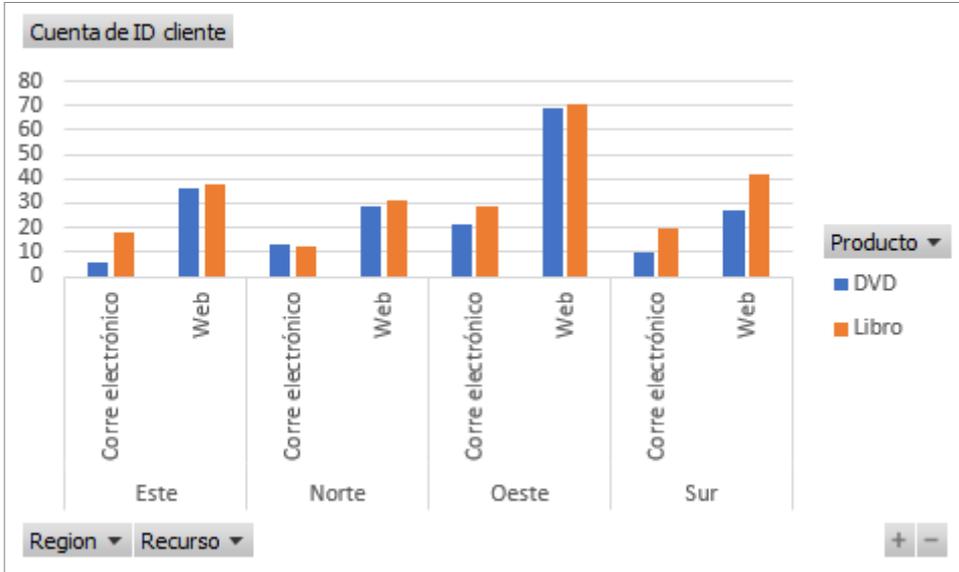


Figura 112. Gráfico dinámico para ventas por región, producto y fuente de pedidos

Segmentaciones y paneles de tabla

Excel® 2010 introdujo segmentaciones de datos: una herramienta para explorar en profundidad para “cortar” una tabla dinámica y mostrar un subconjunto de datos. Para crear una segmentación de datos para cualquiera de las columnas de la base de datos, se hace clic en la tabla dinámica y se elige Insertar segmentación en la pestaña Analizar de la cinta herramientas de tabla dinámica.

Ejemplo 44. Uso de segmentaciones

En la tabla dinámica, para el recuento de ventas regionales por producto, vamos a insertar una segmentación de datos para el origen de la transacción, como se muestra en la figura 113. En este caso, elegimos Source como segmentación de datos. Esto da como resultado la ventana de segmentación de datos que se muestra en la figura 114. Si hace clic en uno de los botones de origen, Correo electrónico o Web, la tabla dinámica refleja solo

los registros correspondientes a ese origen. En la figura 114, ahora tenemos una tabulación cruzada solo para pedidos por correo electrónico.

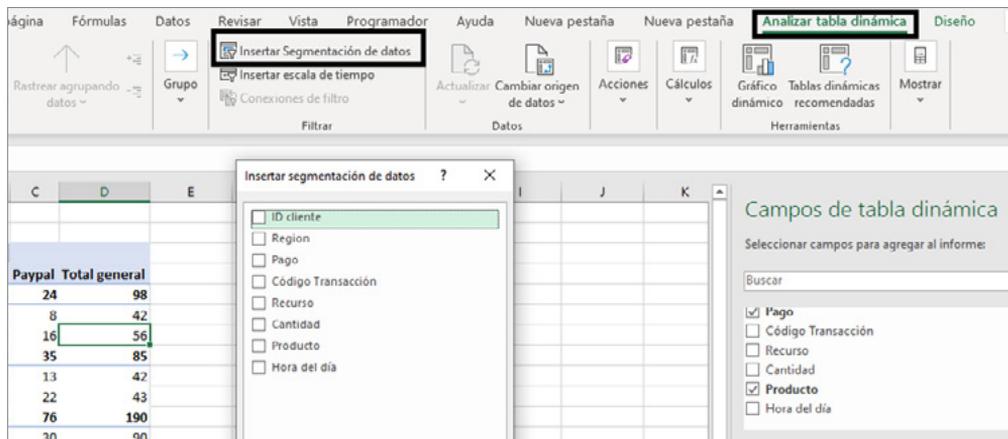


Figura 113. Insertar ventana de segmentaciones

| Cuenta de ID cliente | Etiquetas de columna | Libro | Total general |
|----------------------|----------------------|------------|---------------|
| Este | DVD | 42 | 56 |
| Corre electrónico | | 6 | 18 |
| Web | | 36 | 38 |
| Norte | | 42 | 43 |
| Corre electrónico | | 13 | 12 |
| Web | | 29 | 31 |
| Oeste | | 90 | 100 |
| Corre electrónico | | 21 | 29 |
| Web | | 69 | 71 |
| Sur | | 37 | 62 |
| Corre electrónico | | 10 | 20 |
| Web | | 27 | 42 |
| Total general | | 211 | 261 |

Figura 114. Tabulación cruzada rebanada por correo electrónico

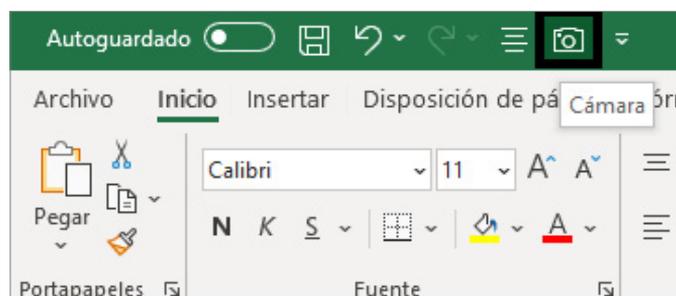
Herramienta Cámara de Excel®

Una característica poco conocida de Excel® es la herramienta de cámara. Esto permite crear imágenes en vivo de varios rangos de diferentes hojas de trabajo que se pueden colocar en una sola página, dimensionarlas y

organizarlas fácilmente. Son simplemente imágenes vinculadas de los rangos originales y la ventaja es que a medida que los datos cambian o actualizan, las tomas de la cámara también se actualizan. Esto es especialmente valioso para imprimir resúmenes cuando necesita extraer datos de varias hojas de trabajo, consolidar tablas dinámicas en una página o para crear paneles cuando las tablas y gráficos están dispersos en varias hojas de trabajo. Para utilizar la cámara, primero se agrega a la barra de herramientas de acceso rápido (el conjunto de botones situados encima de la cinta de opciones). En el menú Archivo, se elige Opciones y, a continuación, Barra de herramientas de acceso rápido. Se elige Comandos y, a continuación, Comandos no en la cinta de opciones. Se selecciona Cámara y se agrega.

A continuación, aparecerá como se muestra en la figura 115. Para usarla, simplemente resalte un rango de celdas. Si desea capturar un gráfico, resalte un rango de celdas que lo rodean, se hace clic en el botón de la herramienta de la cámara y, a continuación, se hace clic en la ubicación donde desea colocar la imagen. Puede cambiar el tamaño de la imagen como cualquier otro objeto de Microsoft Excel®.

Figura 115. Botón de la herramienta de la cámara de Excel®



La herramienta “cámara” es útil para crear paneles visualizaciones de varias tablas dinámicas y gráficos diferentes. Permite tomar fotografías de las imágenes generadas en tablas y gráficos, para posteriormente consolidarlos en una hoja de trabajo. De esta manera, todavía se pueden realizar cambios en las tablas dinámicas que se reflejarán automáticamente en las tomas de la cámara. La figura 116 muestra un panel simple creado con la herramienta de cámara para la base de datos de transacciones de ventas.

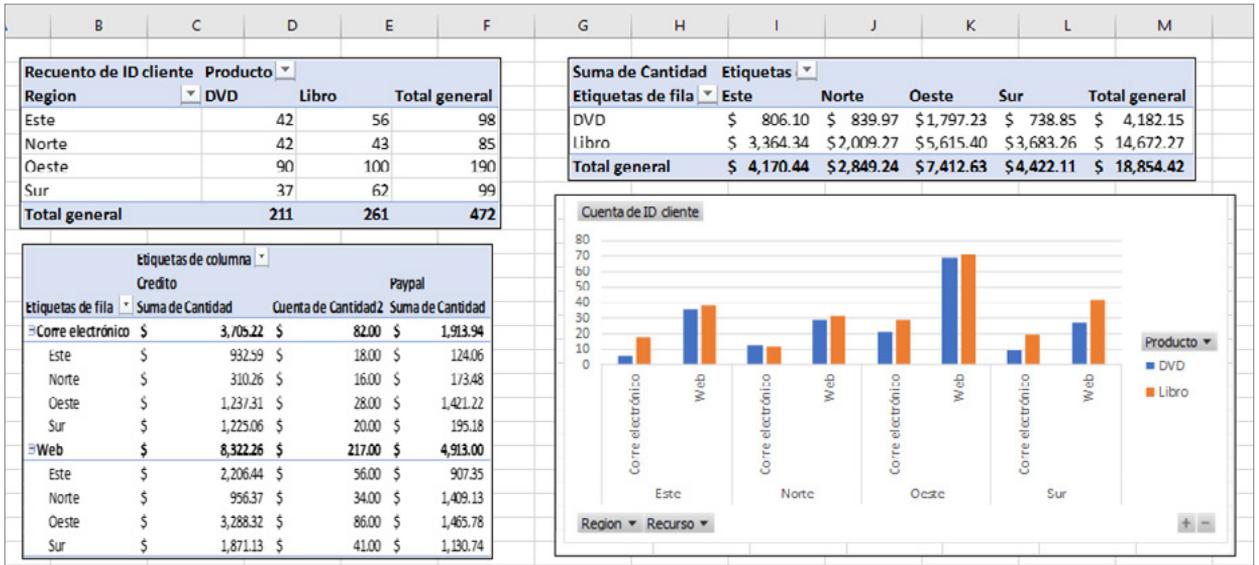


Figura 116. Panel basado en cámara

REFERENCIAS

CALANTONE, R., Cornelia , D., Litvack, D., & Benedetto, A., “Flanqueando en un precio guerra”. *Interfaces*, 1989, pp. 1–12.

DRUCKER, P., *The Manager and the Management Sciences in Management: Tasks, Tasksies, Practices*. Londres: Harper and Row, 1974.

JAMES, R., E., *Business Analytics*. Vivar, Malaysia: Pearson, 2017.

LAUDON, K., & Laudon, J., *Essentials of Management Information Systems*. Superior Silla Río, NJ: Prentice Hall, 2011.

LUSTIG, I., Dietrich, B., Johnson, C. & Dziekan, C., *The Analytics Journey*. *Analytics Magazine* noviembre/diciembre de 2010.

WICKHAM, H. & Grolemond, G., *R for Data Science*, O’Reilly, 2017.

ANALYTICS INFORMS

<https://pubsonline.informs.org/magazine/analytics>

Consultado el 28 de septiembre de 2021.

CAPÍTULO 5

Enfoque directivo y estratégico de la inteligencia analítica de negocios

5.1 Definiciones de dirección y estrategia

La *inteligencia analítica de negocios* debe formar parte e incluso, en algunos casos, ser el soporte de la dirección estratégica de las empresas u organizaciones. Para poder entender lo que esto significa, primero se deben comprender los conceptos de dirección y estrategia para finalmente utilizar alguno de los modelos planteados para realizar la dirección estratégica en las empresas u organizaciones.

Definición y concepto de dirección

Revisaremos tres definiciones y conceptos de lo que significa dirección, las cuales son enunciadas por Peter F. Drucker, Henry Mintzberg y Michael E. Porter.

Para Drucker (2001), la dirección en una empresa es: “hacer que las personas sean capaces de desempeñarse en forma conjunta mediante objetivos y valores conjuntos, una estructura adecuada y la capacitación y desarrollo necesarios para responder a los cambios”. En otras palabras, el director es aquella persona que establece los objetivos de una empresa u organización y mediante el trabajo en equipo, capacitación y establecimiento de un marco de principios los alcanza.

Por otro lado, Mintzberg (2009) menciona que la dirección no es una ciencia ni una profesión, es una práctica aprendida principalmente a través de la experiencia y basada en el contexto en el que se desarrolla; con esta consideración la dirección tiene lugar dentro de un triángulo en donde el arte, el trabajo y el uso de la ciencia se encuentran (véase la figura 117). El arte aporta las ideas y la integración; el trabajo hace las conexiones basándose en experiencias tangibles; y la ciencia proporciona el orden, a través del análisis sistemático del conocimiento. En la figura 118 se muestra un modelo conceptual de la forma en que Mintzberg conceptualiza a la dirección.

Para Porter (1996), el director debe establecer la disciplina para decidir cuáles cambios en la industria y las necesidades de los clientes deben responderse, mientras evita distracciones organizacionales y mantiene el carácter distintivo de la empresa. Otra función importante del director es el establecimiento de límites y mejorar la eficiencia operacional.

Figura 117. La dirección como una composición de arte, trabajo y ciencia.
Fuente: Adaptado de Mintzberg (2009)

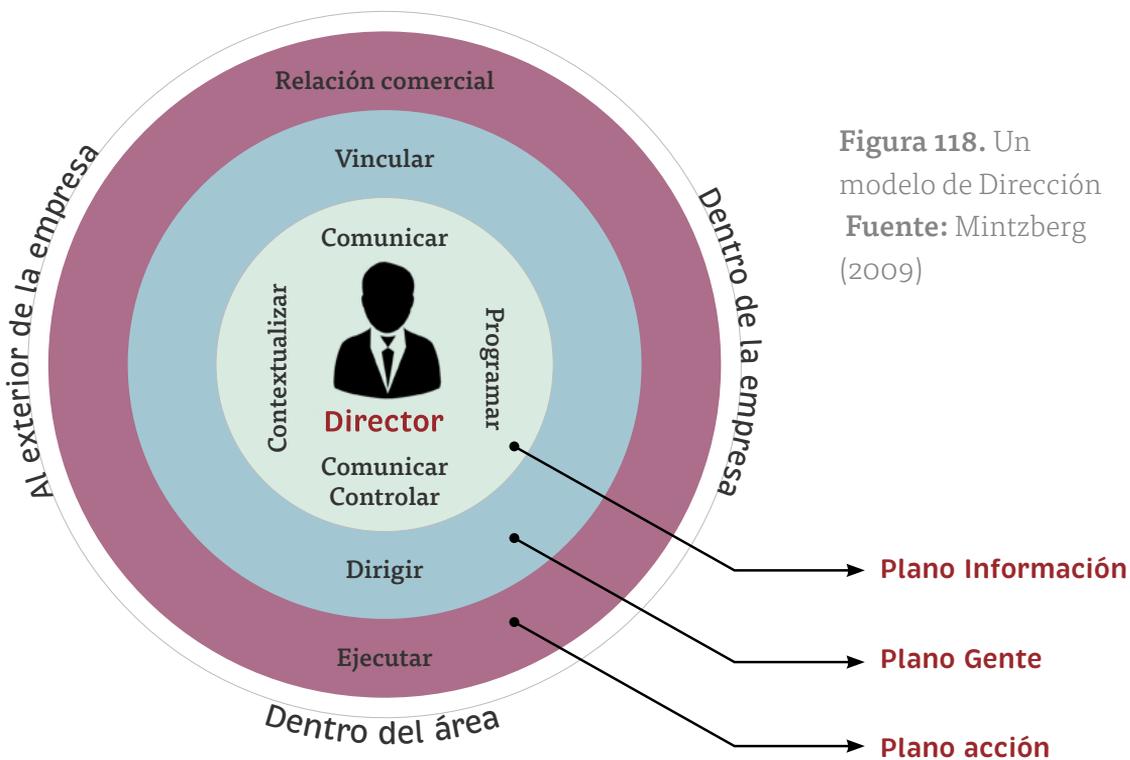
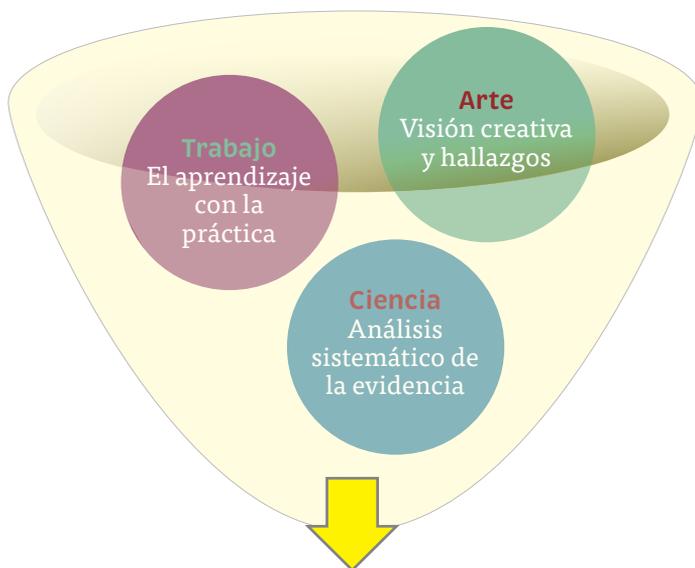


Figura 118. Un modelo de Dirección
Fuente: Mintzberg (2009)

Definición y concepto de estrategia

Los autores que revisaremos para entender el concepto y la definición de estrategia serán: Phillip Kotler, Arthur A. Thompson y Michael E. Porter.

Para Kotler (2010), el objetivo general de cualquier estrategia es garantizar la supervivencia a largo plazo, y el enfoque se encuentra en los mercados relevantes y sus oportunidades y amenazas. Es decir, que la estrategia trata de definir el o los mercados que son relevantes para las empresas u organizaciones desde una perspectiva material, geográfica y temporal para así permitirle desarrollar dicha estrategia. Esta definición abona en dos dimensiones importantes, una parte técnica al introducir el concepto de lo material y, otra, la dimensión geográfica, que se relaciona directamente con la perspectiva operacional y se extrapola a la perspectiva financiera. El mismo Kotler menciona que una estrategia debe tener 5 características:

1. Tiene un aspecto integrado, es decir, que se relaciona con todas o la mayor parte de las áreas de la empresa.
2. Se pretende intencional por parte de los tomadores de decisiones.
3. Está orientada a la actividad, es decir, se crea con un enfoque en la acción directa.
4. Es sistemática, en otras palabras, comprensible a terceros.
5. Persigue el logro a largo plazo de los objetivos.

Por su parte, Thompson (2012) menciona que una estrategia es un plan de acción de una empresa para competir con éxito y obtener utilidades, este plan brinda una ventaja competitiva sustentable (este término se refiere a que se puede mantener en el tiempo) y para ello debe evolucionar con el tiempo. Asimismo, Thompson menciona que la forma en que una estrategia se evalúa y se considera ganadora, se logra mediante el sometimiento de esta a tres pruebas:

1. Prueba del ajuste: una estrategia se considera ganadora si se ajusta bien a las condiciones externas (industria) respecto a sus competidores y al mismo tiempo a los recursos y capacidades de la empresa.

2. Prueba de la ventaja competitiva: las estrategias exitosas permiten que una empresa consiga una ventaja competitiva perdurable.
3. Prueba del desempeño: Dos clases de indicadores de desempeño son los que mejor revelan el alcance de la estrategia de una compañía: los primeros serían la rentabilidad y la fortaleza financiera y, los segundos, su fuerza competitiva y su posición en el mercado.

Para Porter (1996), la estrategia se relaciona con un posicionamiento estratégico, y este se consigue al desempeñar actividades diferentes a los rivales o actividades similares, pero de formas diferentes. Así, enuncia tres principios clave para conseguir un posicionamiento estratégico:

1. **La estrategia es la creación de una posición única y valiosa que involucra un conjunto de actividades.** Dichas actividades surgen de tres diferentes motivaciones: primero, se debe hacer la selección de los productos o servicios; segundo, se tiene que hacer la segmentación de los clientes a los que se dirigen esos productos o servicios; y tercero, se debe decidir sobre la forma en que se entregarán a los clientes dichos productos o servicios.
2. **La estrategia requiere hacer negociaciones respecto a las actividades que se deben o no hacer.** Estas negociaciones se hacen con base en tres razones: la primera razón es la congruencia entre las actividades y la imagen que proyecta la compañía; la segunda razón tiene que ver con respecto al uso efectivo de los recursos de la empresa; y la tercera razón es las limitaciones respecto a la coordinación y control interno en la empresa.
3. **La estrategia involucra crear “ajustes” entre las actividades de la empresa.** Esto surge debido a que generalmente las actividades de manera frecuente tienen repercusiones entre sí. Se tienen que hacer tres tipos de ajustes, el primero es simple y consiste en ajustar las actividades para que sean consistentes con la estrategia global de la empresa; el segundo se refiere a la ejecución de actividades para reforzar

a actividades previas; y el tercer ajuste se refiere a la ejecución de actividades encaminadas a optimizar los recursos.

Los “ajustes” en la estrategia son fundamentales para mantener una ventaja competitiva, la probabilidad de que los competidores puedan imitar un “ajuste” en el sistema o estrategia total es altamente improbable. De esta forma, el mantener una ventaja competitiva depende de la cantidad de ajustes necesarios y bien ejecutados en las actividades que conforman la estrategia.

5.2 Metodología de la dirección estratégica

Para David (2010), la dirección estratégica puede ser definida como: “el arte y la ciencia de formular, implementar y evaluar decisiones multifuncionales que permiten a una organización conseguir sus objetivos”. Asimismo, la dirección estratégica considera tres preguntas importantes que debe responder, las cuales son: ¿Dónde nos encontramos ahora?, ¿A dónde queremos ir? y ¿Cómo vamos a llegar allá? Para responder a estas preguntas se tiene un proceso para elaborar la dirección estratégica (véase la figura 119), el cual consiste en:

1. **Formulación de la estrategia:** en esta etapa se incluye el desarrollo de una visión y misión, identificar las oportunidades y amenazas para la organización, determinar las fortalezas y debilidades internas, establecer los objetivos de largo plazo y generar diferentes alternativas de estrategia, de las cuales se deben seleccionar cuáles de ellas deben ser implementadas.

Algunas de las preguntas que deben responderse en esta etapa son:

- ¿En qué nuevos negocios entrar?
- ¿Qué negocios deben abandonarse?
- ¿Cómo distribuir los recursos?

- ¿Se deben expandir las operaciones o diversificarse?
- ¿Se deben entrar a mercados internacionales?
- ¿Fusiones? ¿Adquisiciones? ¿Aventuras conjuntas?

Las decisiones en esta etapa hacen que una organización especifique los productos y/o servicios, mercados, recursos y tecnologías que utilizará en un período de tiempo prologado.

2. Implementación de la estrategia: en esta etapa la empresa requiere establecer objetivos anuales, diseñar políticas, motivar a los empleados y distribuir recursos para que las estrategias formuladas puedan ser ejecutadas.

Algunas de las actividades más importantes y representativas de esta etapa son:

- Crear e implementar una cultura organizacional que soporte la estrategia.
- Crear una efectiva estructura organizacional.
- Dirigir los esfuerzos de mercadotecnia.
- Preparar presupuestos.
- Desarrollar y utilizar sistemas de información.
- Vincular la compensación de los empleados al desempeño organizacional.

A la etapa de implementación de la estrategia se le denomina la “etapa de acción” de la dirección estratégica.

3. Evaluación de la estrategia: es la etapa final de la dirección estratégica, se basa en la necesidad de los directores de saber cuándo las estrategias no están trabajando bien. Todas las estrategias están sujetas a futuras modificaciones por cambios en los factores externos o internos.

Los pasos por seguir en esta etapa son: (1) revisar los factores externos e internos que son las bases de las estrategias actuales, (2) medir el desempeño del sistema con base en los objetivos de desempeño a alcanzar, y (3) tomar acciones correctivas en caso de ser necesario.

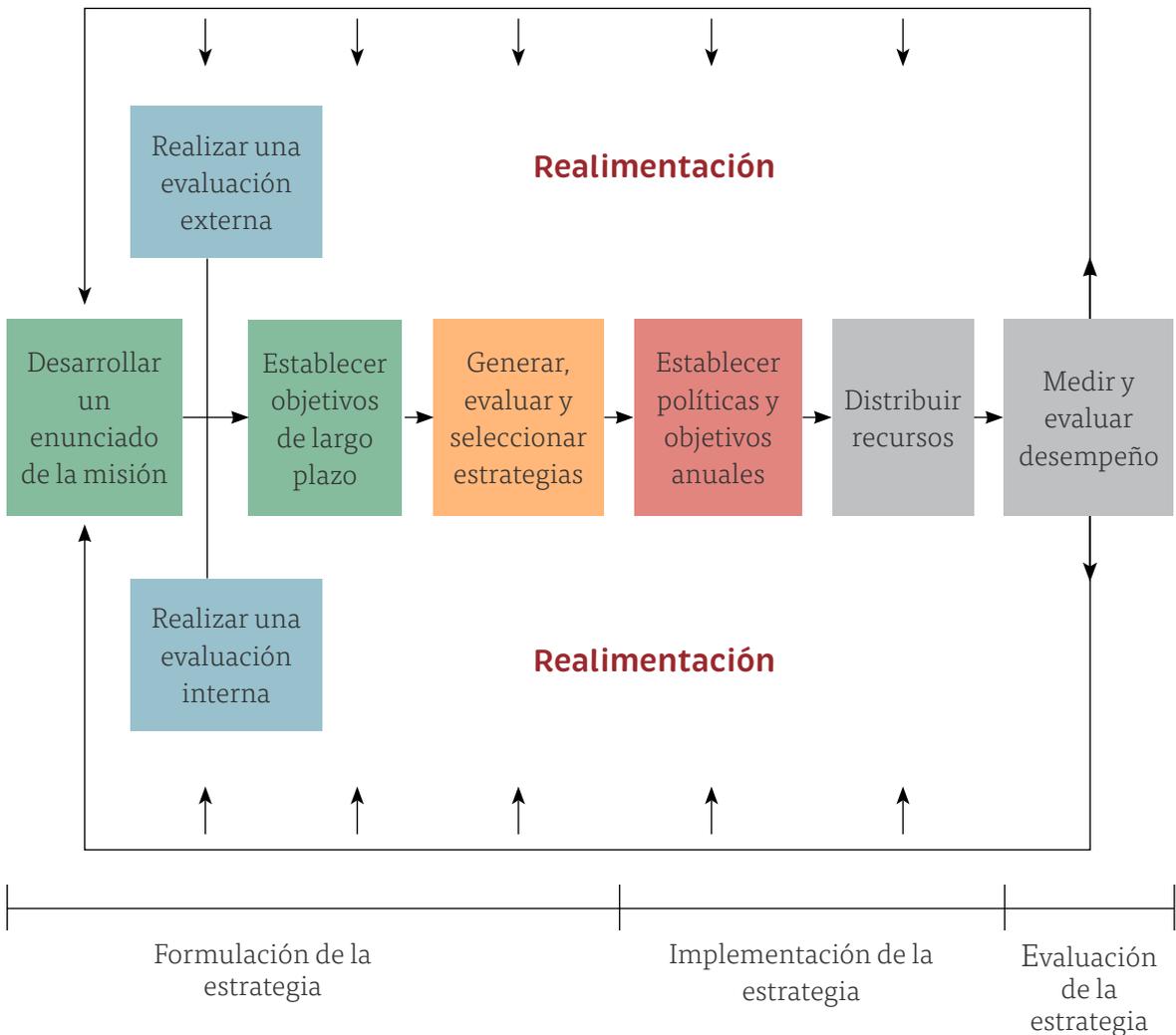


Figura 119. Un modelo integral de dirección estratégica
Fuente: David (2010)

5.3 Inteligencia analítica de negocios y la dirección estratégica

El objetivo principal de la inteligencia analítica de negocios es recopilar diferentes tipos de datos que se obtienen de fuentes variadas que podrían incluir canales diferentes, aparentemente aislados el uno del otro, pero que permiten analizar la organización como un todo y transformar los datos en información con el fin de tomar decisiones. La inteligencia de negocios actúa como un factor estratégico para una empresa u organización, generando una ventaja competitiva al proporcionar información que permite medir el desempeño de la empresa para identificar las áreas de mejora y de oportunidad en la operación de la empresa, tales como entrada a nuevos mercados, promociones y ofertas de productos, eliminación de vacíos de información, control financiero, optimización de costos, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, entre otros. En la figura 120 se ve de manera esquemática simple el proceso de la inteligencia analítica de negocios.



Figura 120. Proceso simplificado de la inteligencia de negocios Fuente: Elaboración propia.

Para tener una estrategia sólida actualmente, se requiere de la digitalización de la empresa porque que es muy útil para competir en un mercado cada vez más globalizado, tomando en cuenta que la transformación digital no solo es tecnológica sino social y económica.

Una empresa debe tener un mejor desempeño en cada aspecto de su organización con respecto a la competencia al generar diferenciadores que aporten valor dentro de la perspectiva de los clientes y que representen

un costo bajo para la empresa y, de esta manera, obtener mayores ganancias. Lo anterior se logra realizando acciones específicas de forma eficiente, estas acciones y la forma en que se implementen es lo que permite desarrollar una estrategia competitiva. Si las acciones que se realizan son las mismas que la competencia, pero mejores, se está hablando de que la empresa posee eficacia operacional, pero si se habla de acciones distintas, tenemos un posicionamiento estratégico.

La estrategia competitiva se define como el conjunto de acciones que se ponen en marcha para alcanzar una ventaja competitiva frente al resto de los competidores, que pueda ser sostenida a lo largo del tiempo y que genere una mayor rentabilidad. Una empresa que pretende tener una estrategia competitiva que se base en la inteligencia analítica de negocios, debe tomar en cuenta gestionarse a nivel conjunto, es decir, poner a disposición de toda la organización los datos y su análisis para que al momento de tomar decisiones se contemplen todas las partes que integran la empresa.

Para lograr implementar una estrategia competitiva que se base en la inteligencia analítica de negocios, se requiere que la empresa haga cambios en la cultura, procesos, conducta y competencias de quienes laboran en la organización; estos cambios deben ser liderados por las personas que ejerzan los puestos directivos, quienes deben tener el compromiso de ser capaces de poder discutir con personas expertas en competitividad con enfoque analítico, y tener la fortaleza para empujar a las demás personas a pensar y actuar de forma analítica.

Es importante considerar que una empresa se mide por sus resultados y, hablando de estrategia competitiva, siempre deben observarse y analizarse los resultados que aspiran conseguir, los cuales deben tener la magnitud y alcance suficiente. Una forma de medir el impacto de los resultados es por medio de la generación de ahorros o incremento de las utilidades, esto sin dejar de tomar en cuenta la cuota de mercado y fidelidad de las y los clientes.

Davenport resume lo anterior en lo que llama “los cuatro pilares de la competitividad basada en la inteligencia analítica” (véase la figura 121), los cuales son:

- Capacidad distintiva
- Inteligencia analítica a nivel de empresa
- Compromiso de la alta dirección
- Ambición a gran escala

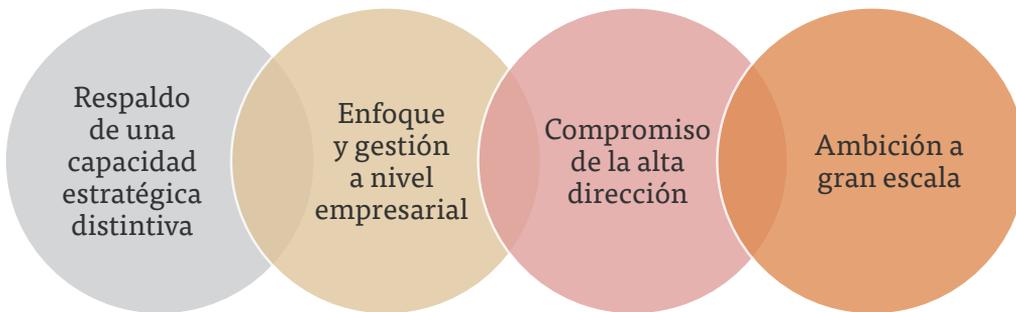


Figura 121. Cuatro pilares de la competitividad basada en la inteligencia analítica

Fuente: Extraído de Davenport (2009).

Ahora bien, es claro que no todas las empresas se desarrollan de la misma manera e implementan una estrategia competitiva similar, y aunque las empresas tengan planteadas estas estrategias basadas en la inteligencia analítica de negocios, al evaluar, se observa lo mucho o poco que aplican la inteligencia analítica; por lo que se han identificado cinco etapas de la estrategia competitiva, como se muestran en la figura 122 y se describen en la tabla 11. En estas etapas se pueden posicionar las empresas, describiendo la ruta que la organización puede seguir desde la base hasta la cima de la pirámide.



Figura 122. Las cinco etapas de la estrategia competitiva basada en la inteligencia analítica **Fuente:** Extraída de Davenport (2009).

Las etapas propuestas por Davenport (2009) son:

- **Etapa 1.** Empresas u organizaciones con inteligencia analítica reducida. Se refiere a aquellas empresas que no tienen datos, por lo que al mismo tiempo no tienen información y no pueden saber la situación de la empresa en un momento dado.
- **Etapa 2.** Empresas con inteligencia analítica localizada. Se refiere a aquellas organizaciones que tienen la capacidad de recolectar datos de manera sistemática en una o varias actividades, aunque la mayor parte de las veces con actividades aisladas.
- **Etapa 3.** Empresas con aspiraciones de competir con base en la inteligencia analítica. Se refiere a aquellas organizaciones que tienen la capacidad de recolectar datos de manera sistemática y coordinada para alguno de los procesos o áreas de la organización.

- **Etapas 4.** Empresas con inteligencia analítica. Se refiere a aquellas organizaciones que tienen la capacidad de recolectar datos de manera sistemática y coordinada para la mayor parte de los procesos de la empresa, y que está usando esos datos para convertirlos en información para la toma de decisiones.
- **Etapas 5.** Empresas cuya estrategia competitiva se basa en la inteligencia analítica. Se refiere a aquellas organizaciones que tienen inteligencia analítica y usan esta para generar conocimiento que les permite elaborar estrategias que dirigen a la empresa u organización a conseguir una ventaja competitiva.

Tabla 11. Las cinco etapas de la estrategia competitiva basada en la inteligencia analítica Fuente: Extraída de Davenport (2009).

| Etapa | Capacidad distintiva | Objetivo | Preguntas por responder |
|---|--------------------------------|---|--|
| Empresas con inteligencia analítica reducida | Actuación a ciegas | Conseguir datos precisos para mejorar la operatividad | ¿Qué ha ocurrido en nuestra actividad empresarial? |
| Empresas con inteligencia analítica localizada | Local y oportunista | Utilizar la inteligencia analítica para mejorar una o más actividades funcionales | ¿Qué podemos hacer para mejorar esta actividad? ¿Cómo podemos conocer mejor nuestra empresa? |
| Empresas con aspiraciones de competir con base en la inteligencia analítica | Esfuerzo en integrar los datos | Utilizar la analítica para mejorar una capacidad distintiva | ¿Qué está ocurriendo ahora? ¿Podemos extrapolar las tendencias actuales? |

| Etapa | Capacidad distintiva | Objetivo | Preguntas por responder |
|--|---|---|---|
| Empresas con inteligencia analítica | Capacidad de utilizar la inteligencia analítica para identificar ventajas | Desarrollar una amplia capacidad analítica. Usar la inteligencia analítica para diferenciarse | ¿Cómo podemos emplear la inteligencia analítica para innovar y diferenciarse? |
| Empresas cuya estrategia competitiva se basa en la inteligencia analítica | Resultados basados en la inteligencia analítica, ventaja sostenible | Usar una estrategia competitiva basada plenamente la inteligencia analítica | ¿Qué viene a continuación? ¿Qué es posible? ¿Cómo seguimos estando a la vanguardia? |

Las empresas cuya ventaja competitiva se basa en la inteligencia analítica de negocios son organizaciones que han seleccionado unas pocas capacidades distintivas sobre las cuales desarrollan sus estrategias, tomando decisiones basadas en datos e información que respalden a esas capacidades seleccionadas. La inteligencia analítica de negocios no constituye una estrategia, pero su utilización para optimizar una capacidad distintiva de la empresa u organización sí constituye una estrategia.

REFERENCIAS

DRUCKER, Peter F. *The Essential Drucker*. USA: Harper Collins Publishers Inc., 2001.

MINTZBERG, Henry. *Management*. USA: Berret-Koehler Publishers Inc., 2009.

PORTER, M. E. "What Is Strategy?" *Harvard Business Review* 74, no. 6 (November–December 1996): 61–78.

KOTLER, Philip. *The Quintessence of Strategic Management*. Alemania: Springer, 2010.

THOMPSON, Arthur A. *Administración Estratégica*. México: McGraw Hill, 2012.

DAVID, Fred R. *Strategic Management Concepts and Cases*. New Jersey: Prentice Hall, 2010.

DAVENPORT, Thomas H. y Harris Jeanne G. *Competing on analytics: inteligencia competitiva para ganar*. México: Alfaomega, 2009.

