

# **INTRODUCCIÓN AL ANÁLISIS DE DATOS**

Conocer los tipos de análisis de datos y su propósito, además del ciclo de vida del análisis de datos, para que puedas seleccionar los análisis a efectuar de acuerdo a tu propósito y a los tipos de datos.



# TABLA DE CONTENIDO

> Introducción

**01** Análisis de datos en *big data*

**02** Tipos de análisis

**03** Ciclo de vida del análisis de datos

**04** Caso de estudio

> Cierre

> Referencias



El **análisis de datos** es un término amplio que abarca los procesos, tecnologías, marcos y algoritmos para extraer información significativa de los datos. Los datos sin procesar en sí mismos no tienen significado hasta que se contextualizan y procesan en información útil.



La analítica es este proceso de extraer y crear información a partir de datos sin procesar. La elección de las tecnologías, los algoritmos y los marcos para el análisis depende de los objetivos analíticos de la aplicación.





El análisis de *big data* es la ciencia de examinar o analizar grandes conjuntos de datos con una variedad de tipos de datos (Dorsey, 2017) estructurados, semiestructurados o no estructurados, que pueden ser datos de tiempo real o por lotes.

Este permite tomar **mejores decisiones**, encontrar nuevas oportunidades comerciales, competir contra rivales comerciales, mejorar el rendimiento y la eficiencia y reducir costos mediante el uso de técnicas avanzadas de análisis de datos.

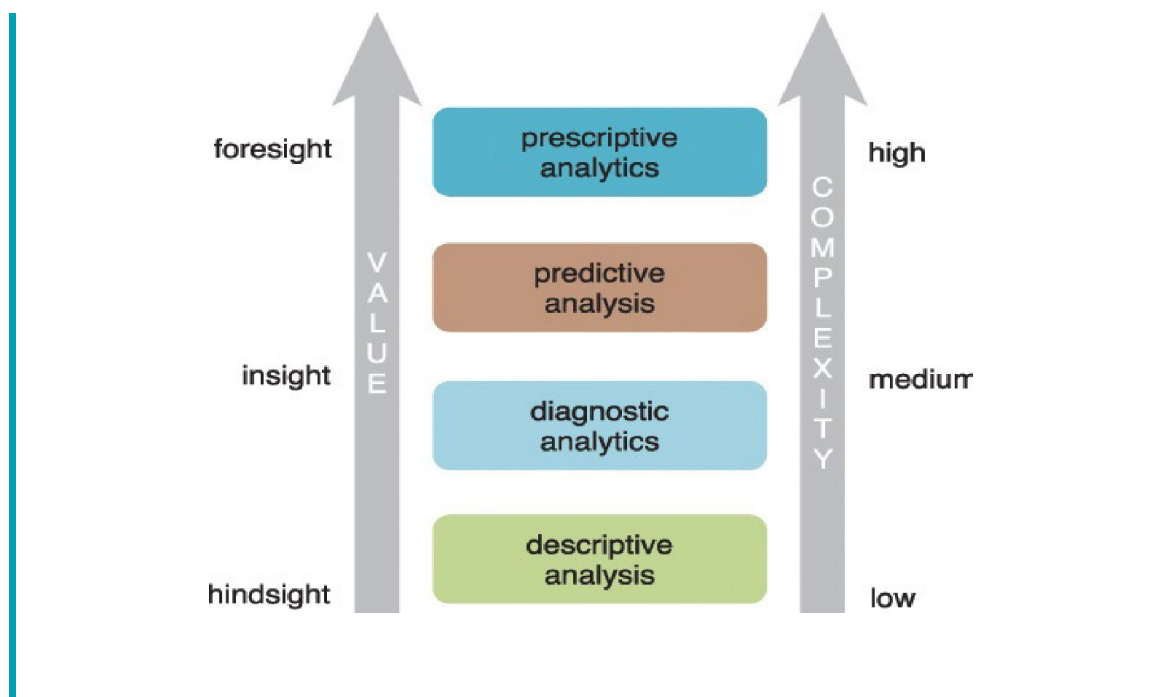
El análisis de datos permite la toma de decisiones basada en datos con respaldo científico para que estas puedan basarse en datos fácticos, y no simplemente en la experiencia pasada o la intuición. Hay cuatro categorías generales de análisis que se distinguen por los resultados que producen:

- Análisis descriptivo
- Análisis de diagnóstico
- Análisis predictivo
- Análisis prescriptivo.

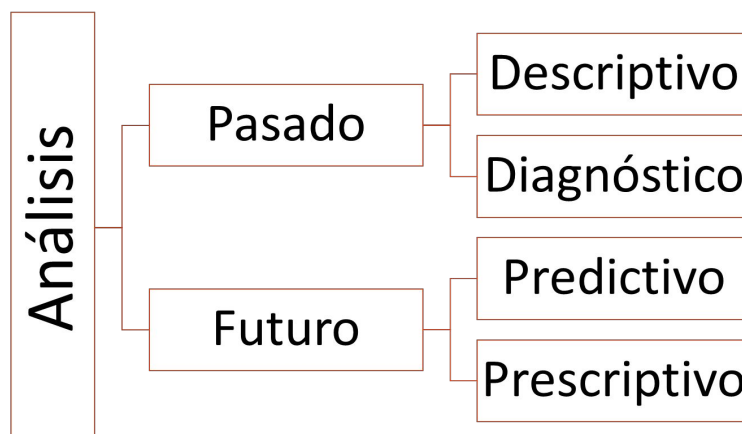
Los diferentes tipos de análisis aprovechan diferentes técnicas y algoritmos de análisis. Esto implica que puede haber diferentes requisitos de datos, almacenamiento y procesamiento para facilitar la entrega de múltiples tipos de resultados analíticos. La siguiente imagen muestra la realidad de que la generación de resultados analíticos de alto valor aumenta la complejidad y el costo del entorno analítico (Erl, Khattak y Buhler, 2016).



El análisis de datos es una disciplina que incluye la gestión del ciclo de vida completo de los datos, que abarca la recopilación, la limpieza, la organización, el almacenamiento, el análisis y el control de los datos. El término incluye el desarrollo de métodos de análisis, técnicas científicas y herramientas automatizadas. En entornos de *big data*, el análisis de datos ha desarrollado métodos que permiten que el análisis de datos ocurra mediante el uso de marcos y tecnologías distribuidas altamente escalables que son capaces de **analizar grandes volúmenes de datos** de diferentes fuentes.



Fuente: El valor y la complejidad aumentan de la analítica descriptiva a la prescriptiva. Extraída de Erl, Khattak y Buhler (2016)



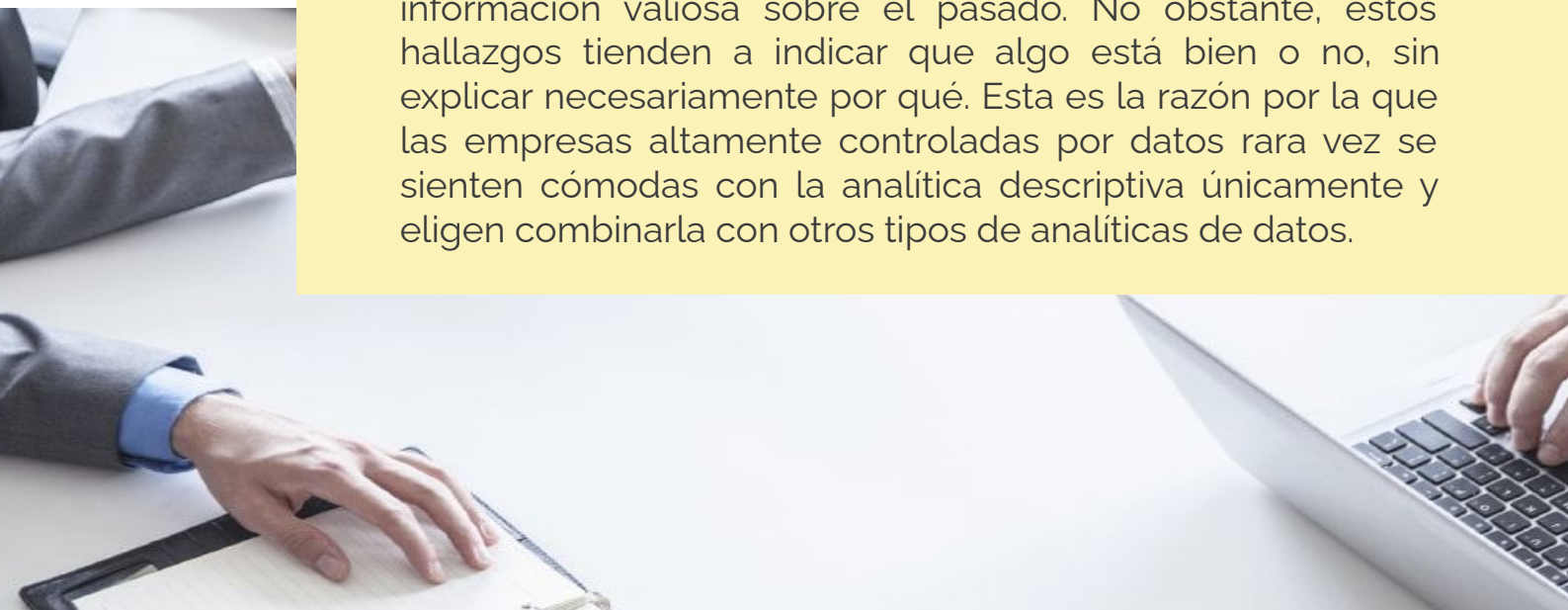
Fuente: Análisis. Elaboración propia (s.f.)



### Análisis descriptivo:

- **Describe, resume y visualiza** cantidades masivas de datos sin procesar (crudos) en una forma en que los usuarios finales pueden interpretar.
- **Describe** los eventos que ocurrieron en cualquier momento del pasado y proporciona una idea de lo que realmente sucedió; se extraen datos anteriores para comprender la razón detrás del fracaso o el éxito.
- **Permite** a los usuarios aprender del desempeño o comportamiento pasado e interpretar cómo podrían influir en los resultados futuros.
- **Encuentra** su aplicación en ventas, marketing, finanzas y más.
- Los podemos **encontrar** en los sistemas de inteligencia de negocio: los indicadores son operaciones aritméticas, media, mediana, máximo, porcentaje, etc.
- Se **calcula** que el 80 % del análisis empresarial implica principalmente descripciones basadas en agregaciones de rendimiento pasado.
- Las dos **técnicas** principales implicadas son la agregación de datos y la minería de datos.

La **analítica descriptiva** esencialmente combina los datos en bruto de diferentes fuentes de datos para ofrecer información valiosa sobre el pasado. No obstante, estos hallazgos tienden a indicar que algo está bien o no, sin explicar necesariamente por qué. Esta es la razón por la que las empresas altamente controladas por datos rara vez se sienten cómodas con la analítica descriptiva únicamente y eligen combinarla con otros tipos de analíticas de datos.





## Análisis diagnóstico:

- **Permite** a los usuarios comprender lo que está sucediendo y por qué sucedió, para que se pueda tomar una acción correctiva si algo salió mal. En términos más simples, los datos históricos se comparan con otros datos para dar una respuesta a la pregunta "¿por qué sucedió?".
- **Beneficia** a los tomadores de decisiones de las organizaciones, al brindarles información procesable. Es un tipo de análisis de causa raíz, de investigación y de detección, que determina los factores que contribuyeron a un determinado resultado.
- Se **utiliza** para analizar redes sociales, datos web o datos de flujo de clics para encontrar un patrón oculto y datos de consumidores. Proporciona información sobre el comportamiento de los clientes rentables y no rentables.



- Se **caracteriza** por técnicas como la exploración, el descubrimiento de datos, la minería de datos y las correlaciones. Las organizaciones generalmente seleccionan este tipo de analítica porque les ofrece una percepción más profunda sobre un problema en particular.

- **Examina** más detalladamente los datos para comprender las causas raíz de los eventos. Es útil para determinar qué factores y eventos contribuyeron al resultado.

- En una **serie temporal** de datos de ventas, el análisis de diagnóstico nos ayudaría a comprender por qué las ventas tienen una disminución o un aumento para un año específico. Sin embargo, este tipo de análisis tiene una capacidad limitada para proporcionar información procesable.





## Análisis predictivo:

- **Analiza** los datos para determinar posibles resultados futuros. Es importante tener en cuenta que no puede predecir si se producirá un evento en el futuro; simplemente pronostica cuáles son las probabilidades de la ocurrencia del mismo.
- **Utiliza** muchas técnicas estadísticas, como el aprendizaje automático, el modelado, la inteligencia artificial y la extracción de datos para hacer predicciones. Explora patrones de datos históricos para determinar riesgos y oportunidades.
- La esencia del análisis predictivo es **diseñar** modelos, de tal manera que se entienda que los datos existentes extrapolan la ocurrencia futura o, simplemente, predigan los datos futuros.
- Se **aplica** en el cuidado de la salud, la gestión de relaciones con los clientes, la venta cruzada, la detección de fraudes y la gestión de riesgos.
- Una de las **aplicaciones** comunes de análisis predictivo se encuentra en el análisis de opiniones, donde se recopilan y analizan todas las opiniones publicadas en las redes sociales (datos de texto existentes) para predecir el sentimiento de la persona sobre un tema en particular como positivo, negativo o neutral (predicción futura).





Por ejemplo, se utiliza para analizar el interés del cliente, los patrones de gasto y otros comportamientos a través de los cuales la organización puede realizar ventas cruzadas de sus productos o vender más productos a los clientes actuales.



Fuente: Análisis predictivo de Netflix. Elaboración propia (s.f.)

## Análisis prescriptivo:

- Va más allá del simple análisis de datos y la predicción de resultados futuros al **brindar** sugerencias para extraer los beneficios y aprovechar las predicciones.
- **Combina** modelos matemáticos, algoritmos de aprendizaje automático y datos históricos.
- **Anticipa** lo que sucederá en el futuro, cuándo sucederá y por qué sucederá. Se implementa utilizando dos enfoques principales: la simulación y la optimización.
- **Encuentra** sus aplicaciones en la fijación de precios, la planificación de la producción, el marketing, la planificación financiera y la optimización de la cadena de suministro.



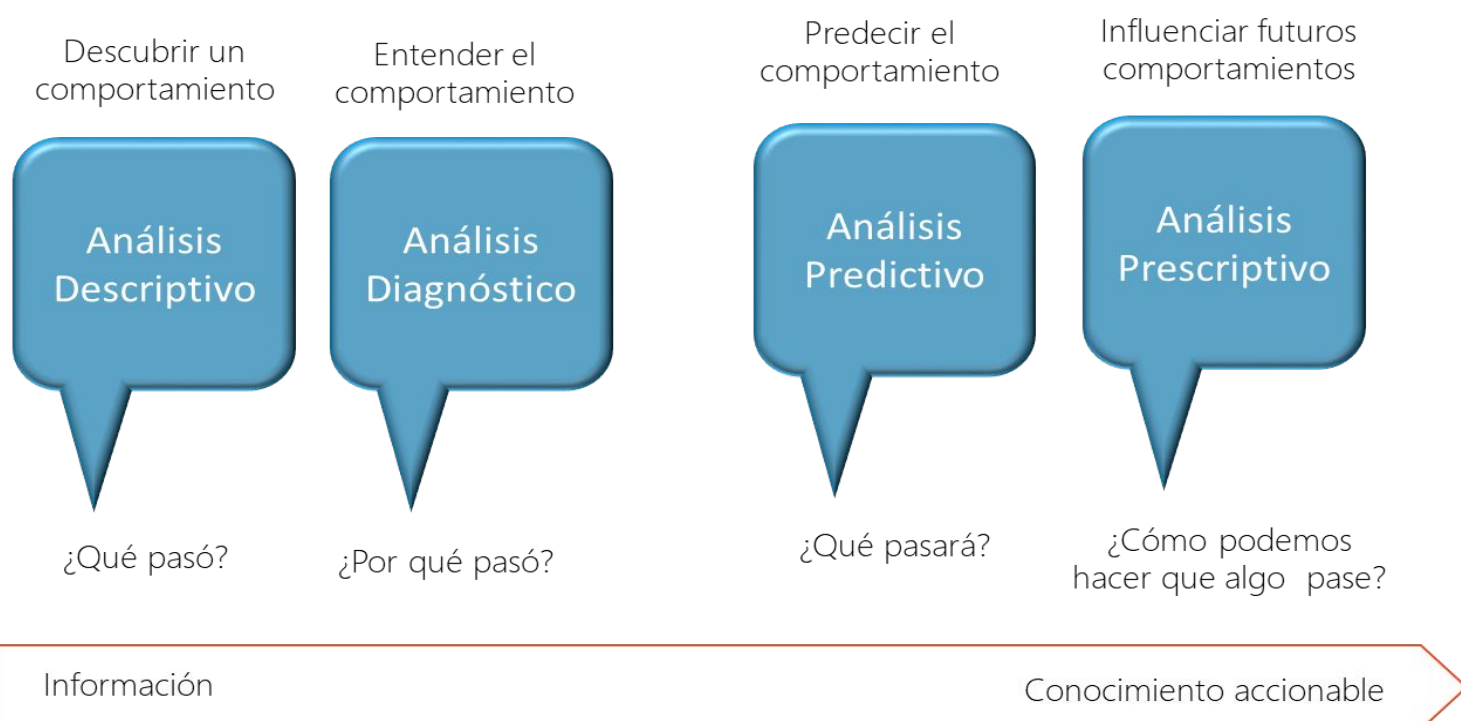
## Análisis prescriptivo vs. análisis predictivo:

- El **análisis predictivo y el análisis prescriptivo** brindan una optimización proactiva de la mejor acción para el futuro, en función del análisis de una variedad de escenarios pasados.

La diferencia real radica en el hecho de que:

- El análisis predictivo ayuda a los usuarios a modelar eventos futuros
- Mientras que el análisis prescriptivo guía a los usuarios sobre cómo afectarán las diferentes acciones al negocio y sugerirles la opción óptima.

A continuación, se muestra un ejemplo de la aplicación de los tipos de análisis en el comportamiento de los clientes:

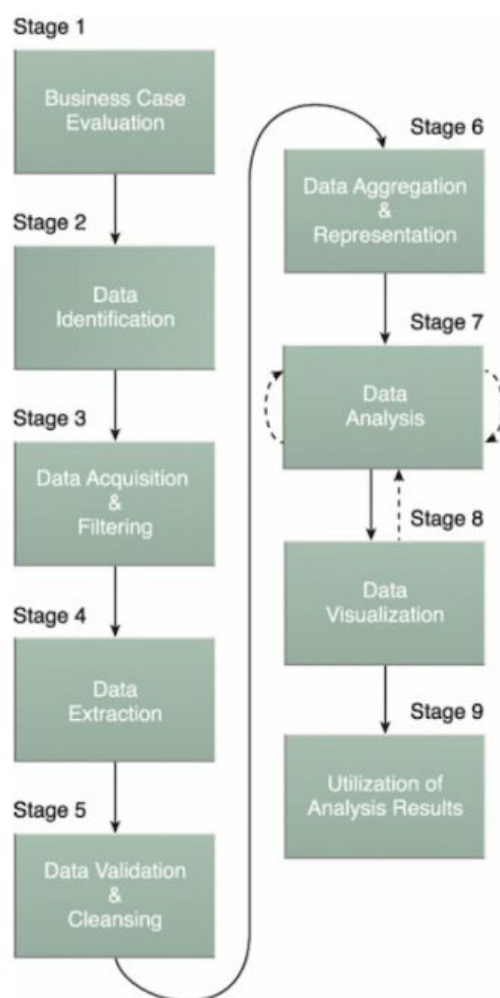


Fuente: Comparación de los tipos de análisis y su propósito en el caso del comportamiento de los clientes. Elaboración propia (s.f.)



El análisis de *big data* **difiere del análisis de datos tradicional** principalmente debido al volumen, la velocidad y las características de variedad de los datos que se procesan. Para abordar los distintos requisitos para realizar análisis en *big data* se necesita una metodología paso a paso para organizar las actividades y tareas involucradas en la adquisición, procesamiento, análisis y reutilización de datos.

El ciclo de vida del análisis de *big data* se puede dividir en las siguientes nueve etapas, como se muestra en la imagen:



Fuente: Las nueve etapas del ciclo de vida del análisis de *big data*. Extraída de Erl, Khattak y Buhler (2016)



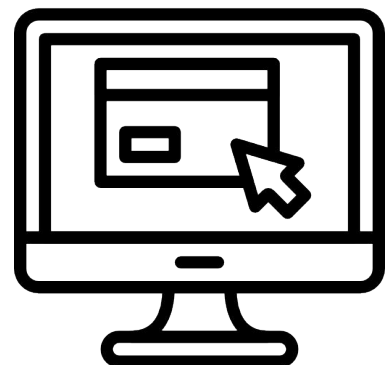


1. Evaluación de caso de negocio
2. Identificación de datos
3. Adquisición y filtrado de datos
4. Extracción de datos
5. Validación y limpieza de datos
6. Agregación y representación de datos
7. Análisis de datos
8. Visualización de datos
9. Utilización de los resultados del análisis

A continuación, describimos cada uno de estos pasos:

## 01 Evaluación del caso de negocio:

- Cualquier ciclo de vida de analítica de datos debe comenzar con un caso de negocio **correctamente definido** que presente un claro conocimiento de la justificación, los objetivos, KPIs y la motivación de llevar a cabo el análisis en primer lugar.
- Un caso de negocio debe crearse, evaluarse y aprobarse antes de continuar con las tareas de análisis práctico reales.
- La evaluación del caso de negocio permite, a quien toma las decisiones:
  - o Comprender los recursos empresariales requeridos: Hw, sw, capacitación
  - o Los desafíos que se enfrentan
  - o Determinar si realmente es un problema de *big data*





## 02 Identificación de los datos:

- Esta etapa está dedicada a **aislar** los conjuntos de datos necesarios para el proyecto de análisis y sus fuentes.
- Cuando **identifica** una variedad más amplia de orígenes de datos, puede aumentar la probabilidad de descubrir correlaciones y patrones ocultos.
- Pueden ser externas e internas a la empresa.

## 03 Adquisición de los datos:

- Los datos externos pueden necesitar **integración** con una API como Twitter. En estos casos, algunos de los datos adquiridos pueden no ser relevantes y ser desechados en etapas tempranas.

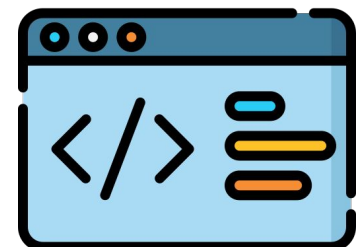
- Los datos corruptos pueden incluir:



- Valores sin sentido
- Registros con valores faltantes
- Tipos de datos no válidos.

- Acá también se agregan metadatos de manera automática, tales como:

- Estructura y tamaño del conjunto de datos
- Tiempo de creación
- Información de origen o recopilación



- Ayudan a mantener la procedencia de los datos a lo largo del ciclo de vida para la precisión y calidad del dato.



## 04

### Extracción de datos:

Algunos datos identificados, como entrada para el análisis, pueden tener un formato incompatible para la solución *big data*. Se busca convertirlos en un formato soportado para que pueda ser almacenado y usado en el análisis.

- El grado de transformación depende de los tipos de analíticas y de la capacidad de almacenamiento y procesamiento de la solución *big data*, por ejemplo: xml, json, txt, csv.

## 05

### Validación y limpieza:

- Los datos no válidos pueden sesgar y falsear fácilmente los resultados del análisis, especialmente los externos.
- Los datos redundantes de varios conjuntos de datos se pueden aprovechar para completar datos faltantes, por ejemplo.



## 06

### Agregación y representación de los datos:

- Los datos pueden distribuirse en muchos conjuntos de datos, lo que requiere que los conjuntos de datos se vinculen mediante campos comunes, por ejemplo, ID o fecha.
- De cualquier manera, es necesario un método de conciliación de datos o se deben determinar los datos que representan el valor correcto.
- La agrupación de datos, así como la etapa de representación, tienen como objetivo integrar diferentes conjuntos de datos para tener una vista unificada.
- Los enormes tamaños de datos pueden hacer que la agrupación sea una operación que requiera tiempo y esfuerzos intensos y requerir una lógica compleja que se ejecute automáticamente sin la intervención humana.



- Los requisitos del análisis de datos deben analizarse durante esta etapa para ayudar a fomentar la reutilización de los datos. Se necesite o no la agrupación de datos, es importante darse cuenta de que los mismos datos se pueden mantener en muchas formas diferentes.
- Esto puede requerir la creación de un repositorio central y estándar para el análisis, como una base de datos NoSQL.

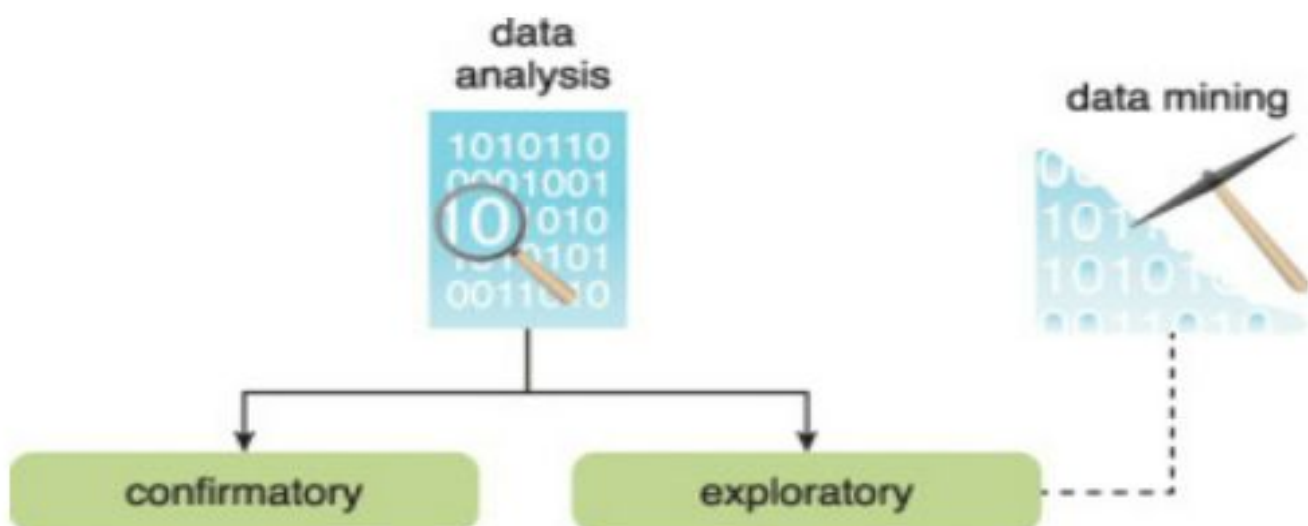




## 07

### Análisis de datos:

- La etapa de análisis de datos normalmente involucra uno o más tipos de análisis.
- Por su naturaleza es iterativa: el análisis de caso se repite hasta que se descubra la correlación o patrón correcto.
- Puede ser tan fácil como una consulta a un conjunto de datos para calcular una agrupación por comparación, o tan desafiante como reunir técnicas de análisis estadístico complejo y minería de datos.
- Se puede clasificar como análisis exploratorio o confirmatorio.
- El confirmatorio es un enfoque deductivo en el que se sugiere la causa del fenómeno investigado (hipótesis), la cual se comprueba o refuta y se ofrecen respuestas definitivas a las preguntas.



**Fuente:** Análisis de datos. Extraída de Erl, Khattak y Buhler (2016)



- La utilización de los resultados del análisis tiene por objeto determinar cómo y dónde se pueden aprovechar aún más los datos procesados del análisis.
- Durante esta etapa tenemos áreas comunes que generalmente se exploran, que incluyen lo siguiente:
  - o **Entrada en sistemas empresariales:** por ejemplo, en una tienda en línea, usar los resultados de análisis relacionados con el cliente que impacta en las recomendaciones de productos.
  - o **Orientación del proceso de negocio:** por ejemplo, consolidación de las rutas de transporte como parte del proceso de la cadena de suministro.
  - o **Alertas:** por ejemplo, puede crear alertas para informar a los usuarios sobre un evento, que necesita que realicen una acción correctiva a través de mensajes de texto SMS o correo electrónico.

## 08 Visualización de los datos:

- La capacidad de analizar una cantidad masiva de datos para obtener información importante tiene muy poco valor si los analistas son los únicos que pueden interpretar los resultados.
- La etapa de visualización de datos tiene como objetivo utilizar herramientas y técnicas de visualización de datos para representar los resultados del análisis gráficamente para que los usuarios de negocio los interpreten de manera efectiva.
- Estos deben poder comprenderlos para tener la capacidad de proponer mejoras y a su vez descubrir respuestas a preguntas que los usuarios aún no han formulado.



## 09

### Utilización de los resultados de los análisis:

- La utilización de los resultados del análisis tiene por objeto determinar cómo y dónde se pueden aprovechar aún más los datos procesados del análisis.
- Durante esta etapa, tenemos áreas comunes, que generalmente se exploran, que incluyen lo siguiente:
  - o **Entrada en sistemas empresariales:** por ejemplo, en una tienda en línea, usar los resultados de análisis relacionados con el cliente que impacta en las recomendaciones de productos.
  - o **Orientación del proceso de negocio:** por ejemplo, consolidación de las rutas de transporte como parte del proceso de la cadena de suministro.
  - o **Alertas:** por ejemplo, puede crear alertas para informar a los usuarios sobre un evento, que necesita que realicen una acción correctiva a través de mensajes de texto SMS o correo electrónico.

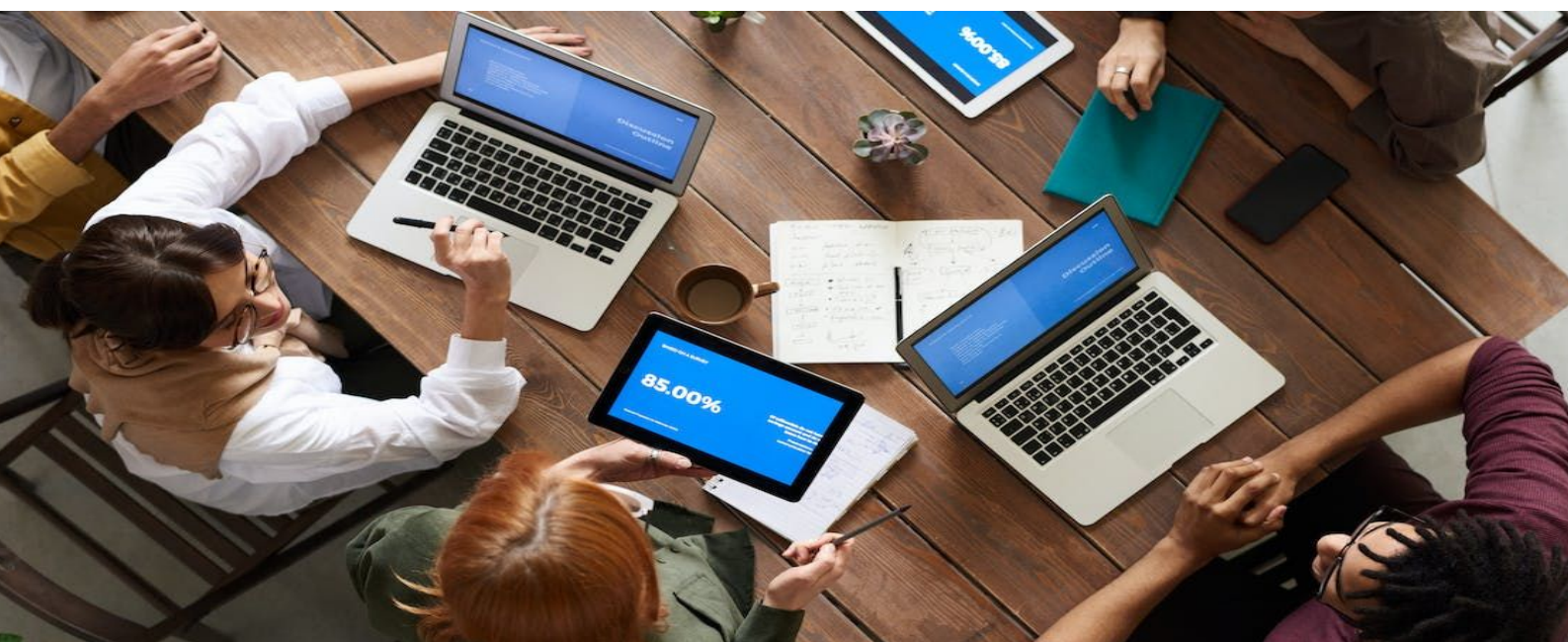




ETI es una compañía de seguros líder que ofrece una gama de planes de seguros en los sectores de la salud, la construcción, la marina y la aviación a su base de clientes de 25 millones dispersos en todo el mundo. La empresa consta de una plantilla de alrededor de 5.000 empleados y genera ingresos anuales de más de 350.000.000 USD.

## 01 | Historia:

ETI comenzó su vida como un proveedor exclusivo de seguros de salud hace 50 años. Como resultado de múltiples adquisiciones durante los últimos 30 años, ETI ha ampliado sus servicios para incluir planes de seguros de propiedad y accidentes en los sectores de la construcción, la marina y la aviación. Cada uno de sus cuatro sectores está compuesto por un equipo central de agentes, actuarios, suscriptores y ajustadores de siniestros especializados y experimentados.

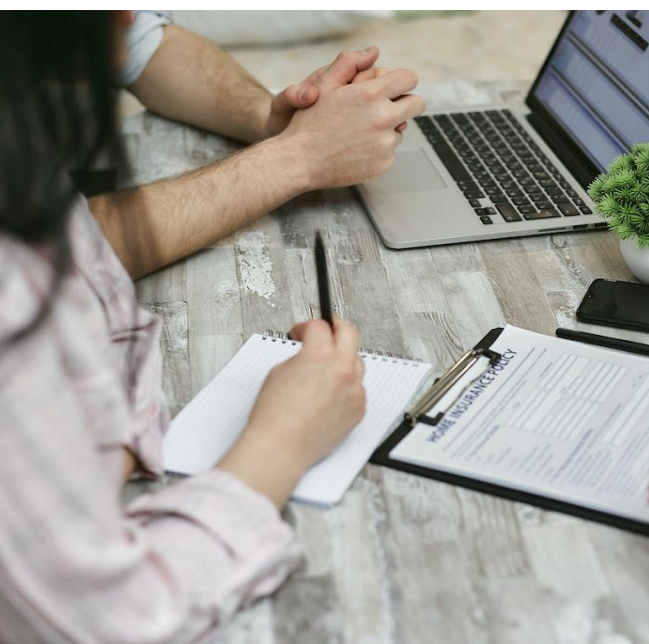


Algunos de los departamentos clave dentro de ETI incluyen los departamentos de suscripción, liquidación de reclamaciones, atención al cliente, legal, marketing, recursos humanos, cuentas y TI. Tanto los clientes potenciales como los existentes generalmente se comunican con el departamento de atención al cliente de ETI por teléfono, aunque el contacto por correo electrónico y redes sociales ha aumentado exponencialmente en los últimos años.





Los agentes generan los ingresos de la empresa mediante la venta de pólizas, mientras que los actuarios son responsables de la evaluación de riesgos, la creación de nuevos planes de seguros y la revisión de los planes existentes. Los actuarios también realizan análisis hipotéticos y utilizan tableros y cuadros de mando para la evaluación de escenarios. Los suscriptores evalúan las nuevas solicitudes de seguros y deciden el monto de la prima. Los ajustadores de reclamos se ocupan de investigar los reclamos hechos contra una póliza y llegan a un monto de liquidación para el titular de la póliza.



ETI se **esfuerza** por distinguirse al ofrecer pólizas competitivas y un servicio al cliente de primer nivel que no finaliza una vez que se vende una póliza. Su gerencia cree que hacerlo ayuda a lograr mayores niveles de adquisición y retención de clientes. ETI depende en gran medida de sus actuarios para crear planes de seguro que reflejen las necesidades de sus clientes.

## 02 | Problema:

En las últimas décadas, la **rentabilidad** de la empresa ha ido en declive. Se formó un comité compuesto por altos directivos para investigar y hacer recomendaciones. Los hallazgos del comité revelaron que la principal razón detrás del deterioro de la posición financiera de la empresa es el aumento en el número de reclamos fraudulentos y los pagos asociados que se realizan en su contra. Estos hallazgos mostraron que el fraude cometido se ha vuelto complejo y difícil de detectar porque los estafadores se han vuelto más sofisticados y organizados. Además de incurrir en pérdidas monetarias directas, los costes relacionados con la tramitación de reclamaciones fraudulentas dan lugar a pérdidas indirectas.



Otro factor que contribuye es un aumento significativo en la ocurrencia de catástrofes como inundaciones, tormentas y epidemias, que también han aumentado el número de reclamos genuinos de alto nivel. Otras razones de la **disminución** de los ingresos incluyen la deserción de clientes debido a la lentitud en el procesamiento de reclamaciones y productos de seguros que ya no satisfacen las necesidades de los clientes. La última debilidad ha quedado expuesta por la aparición de competidores expertos en tecnología que emplean el uso de la telemática para proporcionar pólizas personalizadas.

El comité señaló que la frecuencia con la que se modifican las reglamentaciones existentes y se introducen nuevas reglamentaciones ha aumentado recientemente. Lamentablemente, la empresa ha tardado en responder y no ha podido garantizar el cumplimiento total y continuo. Debido a estas deficiencias, ETI ha tenido que pagar cuantiosas multas.



### 03 | Objetivos:

La mayoría del equipo de TI de ETI está convencido de que *big data* es la panacea que resolverá todos sus problemas actuales. Sin embargo, los miembros de TI capacitados señalan que adoptar *big data* **no es lo mismo** que simplemente adoptar una plataforma tecnológica.



Más bien, primero se debe considerar una variedad de factores para garantizar la adopción exitosa de *big data*. Por lo tanto, para garantizar que se comprenda completamente el impacto de los factores relacionados con el negocio, el equipo de TI se reúne con los gerentes comerciales para crear un informe de viabilidad. Involucrar al personal comercial en esta etapa inicial ayudará a crear un entorno que reduzca la brecha entre las expectativas percibidas de la administración y lo que TI realmente puede ofrecer.

Con base en los hallazgos del comité, los directores de ETI establecen los siguientes objetivos estratégicos:

- **Disminuir** las pérdidas, (a) mejorando la evaluación de riesgos y maximizando la mitigación de riesgos, lo que se aplica tanto a la creación de planes de seguro como cuando se evalúan nuevas solicitudes al momento de emitir una póliza, (b) implementando un sistema proactivo de gestión de catástrofes que disminuye el número de reclamos potenciales resultantes de una calamidad y (c) detección de reclamos fraudulentos.



- **Reducir** la deserción de clientes y mejorar la retención de clientes con (a) resolución rápida de reclamos y (b) políticas personalizadas y competitivas basadas en circunstancias individuales en lugar de la generalización demográfica solamente.
- **Lograr y mantener** el pleno cumplimiento normativo en todo momento mediante el empleo de técnicas mejoradas de gestión de riesgos que puedan predecir mejor los riesgos, ya que la mayoría de las normativas requieren un conocimiento preciso de los riesgos para garantizar el cumplimiento.



04

#### Estrategia:

Después de consultar con su equipo de TI, el comité **recomendó** la adopción de una estrategia basada en datos con análisis mejorados para aplicar en múltiples funciones comerciales, de tal manera que los diferentes procesos comerciales tengan en cuenta los datos internos y externos relevantes. De esta manera, las decisiones pueden basarse en la evidencia y no solo en la experiencia y la intuición. En particular, se enfatiza el aumento de grandes cantidades de datos estructurados con grandes cantidades de datos no estructurados para respaldar la realización de análisis de datos profundos pero oportunos.





El comité preguntó al equipo de TI si existen obstáculos que puedan impedir la implementación de la estrategia antes mencionada. Se recordó al equipo de TI las limitaciones financieras dentro de las cuales debe operar. En respuesta a esto, el equipo preparó un informe de viabilidad que destaca los siguientes obstáculos:



- **Adquirir, almacenar y procesar datos no estructurados de fuentes de datos internas y externas:** actualmente solo se almacenan y procesan datos estructurados, porque la tecnología existente no admite el almacenamiento y procesamiento de datos no estructurados.
- **Procesamiento de grandes cantidades de datos de manera oportuna:** aunque el EDW se usa para generar informes basados en datos históricos, la cantidad de datos procesados no se puede clasificar como grande y los informes tardan mucho tiempo en generarse.
- 
- **Procesamiento de múltiples tipos de datos y combinación de datos estructurados con datos no estructurados:** se producen múltiples tipos de datos no estructurados, como documentos de texto y registros del centro de llamadas, que actualmente no pueden procesarse debido a su naturaleza no estructurada. En segundo lugar, los datos estructurados se utilizan de forma aislada para todo tipo de análisis.

El equipo de TI emitió una recomendación para que ETI adopte *big data* como el medio principal para superar estos impedimentos para lograr los objetivos establecidos.



## 05

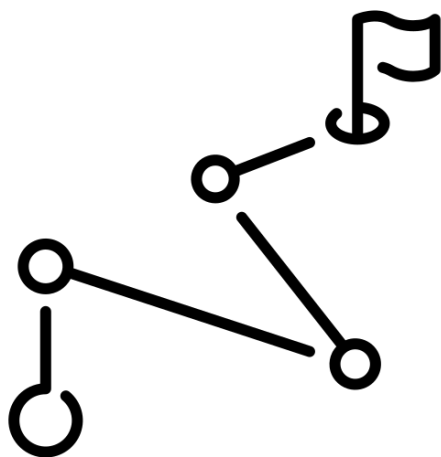
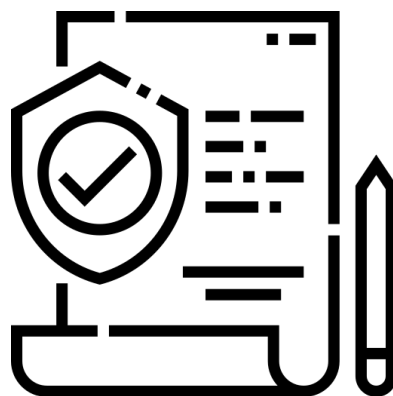
### Análisis:

La implementación del análisis de *big data* para la detección de reclamaciones fraudulentas en los seguros de construcciones corresponde directamente a una reducción en la pérdida monetaria y, por lo tanto, conlleva un respaldo empresarial completo.

El alcance del análisis de *big data* se limita a la identificación de fraudes en el sector de la construcción, para que el análisis sea más sencillo.

En términos de análisis, ETI hace uso de análisis descriptivos y de diagnóstico.

El **análisis descriptivo** incluye consultar el sistema de administración de pólizas para determinar la cantidad de pólizas vendidas cada día, consultar el sistema de administración de reclamos para averiguar cuántos reclamos se envían diariamente y consultar el sistema de facturación para averiguar cuántos clientes están atrasados en sus pagos de primas.



Los **análisis de diagnóstico** se llevan a cabo como parte de varias actividades de BI, como realizar consultas para responder preguntas como por qué no se cumplió el objetivo de ventas del mes pasado. Esto incluye realizar operaciones de desglose para desglosar las ventas por tipo y ubicación, para poder determinar qué ubicaciones tuvieron un rendimiento inferior para tipos específicos de pólizas.



ETI actualmente **no utiliza** análisis predictivos ni prescriptivos. Sin embargo, la adopción de *big data* le permitirá realizar este tipo de análisis, ya que ahora puede hacer uso de datos no estructurados, que cuando se combinan con datos estructurados proporcionan un rico recurso en apoyo de estos tipos de análisis. ETI ha decidido implementar estos dos tipos de análisis de manera gradual, primero implementando análisis predictivos y luego desarrollando lentamente sus capacidades para implementar análisis prescriptivos.

En esta etapa, ETI **planea** hacer uso de análisis predictivos para respaldar el logro de sus objetivos. Por ejemplo, el análisis predictivo permitirá la detección de reclamos fraudulentos al predecir qué reclamo es fraudulento y, en caso de deserción del cliente, al predecir qué clientes probablemente desertarán. En el futuro, a través de análisis prescriptivos, se prevé que ETI pueda mejorar aún más la realización de sus objetivos. Por ejemplo, el análisis prescriptivo puede prescribir el monto correcto de la prima, teniendo en cuenta todos los factores de riesgo, o puede prescribir el mejor curso de acción para mitigar las reclamaciones cuando se enfrentan a catástrofes, como inundaciones o tormentas.

Aunque el fraude de seguros puede ser organizado y oportunista al mismo tiempo, el fraude oportunista en forma de exageración y mentiras cubre la mayoría de los casos.



## 06 | Identificación de tipos de datos:

Los miembros del equipo de TI realizan un ejercicio de categorización de los diversos conjuntos de datos que se han identificado hasta ahora y elaboran la siguiente lista:



- **Datos estructurados:** datos de pólizas, datos de siniestros, datos de perfiles de clientes y datos de cotizaciones.
- **Datos no estructurados:** datos de redes sociales, documentos de solicitud de seguros, notas de agentes del centro de atención telefónica, notas de liquidadores de siniestros y fotografías de incidentes.
- **Datos semiestructurados:** registros de salud, datos de perfiles de clientes, informes meteorológicos, datos de censos, registros de servidores web y correos electrónicos.

## 07 | Caso de negocio:

La implementación del análisis de *big data* para la detección de reclamaciones fraudulentas, corresponde directamente a una reducción en la pérdida monetaria y, por lo tanto, conlleva un respaldo empresarial completo.

El alcance del análisis de *big data* se limita a la identificación de fraudes en el sector de la construcción, para que el análisis sea más sencillo.





Para medir el éxito de la solución de *big data* con respecto a la detección de fraude, uno de los indicadores clave de rendimiento (KPI) es la disminución de las reclamaciones fraudulentas en un 15 %.

Cuando se toma en cuenta el presupuesto, el equipo decide centrar su mayor gasto en la adquisición de una nueva infraestructura que sea adecuada para crear un entorno para la solución *big data*.

Principalmente **aprovecharán** las tecnologías de código abierto para implementar el procesamiento por lotes y, por lo tanto, no consideran que se necesite una gran inversión inicial para las herramientas.

No obstante, cuando ponen en consideración el ciclo de vida de análisis de *big data* más grande, se dan cuenta de que deben presupuestar para adquirir herramientas adicionales de limpieza y calidad de datos y tecnologías más novedosas de visualización de datos.

Una vez que se tienen en cuenta estos gastos, un análisis de costo-beneficio muestra que la inversión de la solución de *big data* puede muy bien recuperarse varias veces si se pueden lograr los KPI para la detección de fraudes.

## 08 | Análisis de los datos:

El equipo de TI **involucra** a los analistas de datos en esta etapa, ya que no tiene el conjunto de habilidades adecuado para analizar datos en apoyo de la detección de reclamos fraudulentos. Para poder detectar transacciones fraudulentas, primero se debe analizar la naturaleza de los reclamos fraudulentos para encontrar qué características diferencian un reclamo fraudulento de un reclamo legítimo. Para ello, se adopta el enfoque de análisis exploratorio de datos.



Como parte de este análisis, se aplica una variedad de técnicas de análisis. Esta etapa se repite varias veces, ya que los resultados generados después de la primera pasada no son lo suficientemente concluyentes para comprender qué hace que un reclamo fraudulento sea diferente de un reclamo legítimo. Como parte de este ejercicio, los atributos que son menos indicativos de un reclamo fraudulento se eliminan, mientras que los atributos que tienen una relación directa se mantienen o agregan.



## 09 | Visualización de datos:

El equipo ha descubierto algunos hallazgos interesantes y ahora necesita transmitir los resultados a los actuarios, aseguradores y ajustadores de siniestros.

Se utilizan métodos que incluyen gráficos de barras y líneas y diagramas de dispersión. Los diagramas de dispersión se utilizan para analizar grupos de reclamaciones fraudulentas y legítimas a la luz de diferentes factores, como la edad del cliente, la antigüedad de la póliza, la cantidad de reclamaciones realizadas y el valor de la reclamación.



## 10 | Utilización de los resultados del análisis:

Con base en los resultados del análisis de datos, los usuarios de suscripción y liquidación de reclamos ahora han desarrollado una comprensión de la naturaleza de los reclamos fraudulentos. Sin embargo, para obtener beneficios tangibles de este ejercicio de análisis de datos, se genera un modelo basado en una técnica de aprendizaje automático, que luego se incorpora al sistema de procesamiento de reclamos existente para marcar los reclamos fraudulentos.





El **análisis de datos** es un área ampliamente estudiada desde hace muchos años. Sin embargo, las características de los grandes volúmenes de datos como son el volumen, la variedad y la velocidad, generan retos importantes que deben ser enfrentados.

Por ello, en este contenido se definen una serie de pasos en el ciclo de vida de análisis *en big data*, que pretenden guiar cualquier proyecto de esta naturaleza desde sus inicios. También se identifican los distintos tipos de análisis y sus aplicaciones para, finalmente, mostrar un caso de estudio.





Dorsey, R. (2017). *Data Analytics: Become A Master In Data Analytics*. CreateSpace Independent Publishing Platform.

Erl, T., Khattak, W. y Buhler, P. (2016). *Big Data Fundamentals, Concepts, Drivers & Techniques*. Prentice Hall.

### **Referencias de las imágenes**

Erl, T., Khattak, W. y Buhler, P. (2016). Análisis de datos [Imagen]. Disponible en: *Big Data Fundamentals, Concepts, Drivers & Techniques*. Prentice Hall.

Erl, T., Khattak, W. y Buhler, P. (2016). El valor y la complejidad aumentan la analítica descriptiva a la prescriptiva [Imagen]. Disponible en: *Big Data Fundamentals, Concepts, Drivers & Techniques*. Prentice Hall.

Erl, T., Khattak, W. y Buhler, P. (2016). Las nueve etapas del ciclo de vida del análisis de *big data* [Imagen]. Disponible en: *Big Data Fundamentals, Concepts, Drivers & Techniques*. Prentice Hall.



**Has culminado la revisión del  
tema**