



## Conceptos básicos de ciencia de datos

### LECCIÓN 2



En la segunda sesión se analizarán los de análisis y los roles que existen en la ciencia de datos.

En la sesión 1 se habló acerca de cómo son los procesos que requieren los datos para un análisis. En esta sesión nos enfocaremos en entender los tipos de análisis existentes y cómo se emplea cada tipo de análisis para extraer información.

En primer lugar:

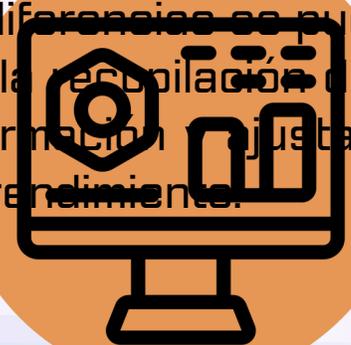
1

Tenemos el análisis descriptivo. Este tipo de análisis se basa en explorar los datos existentes, tal como se ha realizado en las actividades 1 y 2 de la lección 2. El propósito de este análisis es tener una visión general de lo sucedido en el histórico de los datos. El análisis descriptivo debe ayudar a comprender lo que sucede y cómo cambian los fenómenos observados en el tiempo y ante eventos. En el caso del análisis de fútbol se pueden tener muchas variables adicionales que aportarían datos relevantes para entender lo que sucede en los mundiales de fútbol. Por ejemplo, se podrían incluir los datos del porcentaje del presupuesto nacional invertido en deportes, los resultados de los clubes de fútbol de cada país en torneos internacionales, los jugadores de mejor desempeño, entre otros datos que permitan determinar qué ventajas tiene cada equipo sobre los demás y entender por qué sucedieron los resultados observados en los mundiales de fútbol. En síntesis, el análisis descriptivo busca responder ¿Qué ha sucedido?, es decir métricas cuantitativas de los sucesos. Las conclusiones de estos análisis se plasman en informes y gráficos interactivos llamados dashboards. Para que resulte fácil interpretar grandes cantidades de datos sin tener que leer uno a uno cada registro.

# 2

## En segundo lugar:

Está el análisis predictivo. Para realizar este tipo de análisis es necesario recurrir a técnicas estadísticas y de modelado para hacer predicciones con base en los datos. Para esto se analizan patrones en los históricos de datos, se clasifican los datos y se separan de tal forma que se puedan hacer predicciones de eventos futuros respaldadas por los datos de comportamiento de eventos anteriores. El análisis predictivo busca responder la cuestión ¿Qué es lo más probable que ocurra en el futuro? En este tipo de análisis se utilizan modelos estadísticos que son capaces de aprender patrones desde los datos, a esta rama se le conoce como aprendizaje de máquina (machine learning). Este tipo de analítica se emplea en casos de inteligencia de negocios para prever demanda, ventas, precios, entre otros pronósticos de comportamiento y así tomar decisiones informadas basadas en datos. Cabe recordar que el análisis predictivo entrega una probabilidad de ocurrencia de eventos, sin embargo, cualquier evento puede ocurrir ya sea poco probable o altamente probable. Para hacer frente a estas diferencias se pueden realizar ciclos de análisis predictivo junto con la recopilación de los datos reales. Comparar estas fuentes de información y ajustar los modelos de análisis predictivo para mejorar su rendimiento.



En tercer lugar:



Existe la analítica prescriptiva. En este tipo de análisis se quiere vislumbrar qué va a suceder e influir en los acontecimientos para que ocurran de manera beneficiosa a los intereses propios. Para realizar análisis prescriptivo se necesita analizar datos históricos de eventos de interés junto con las predicciones realizadas antes de que tales eventos sucedieran y las decisiones tomadas antes de los sucesos. Con esta información es posible determinar cómo se puede influir en el comportamiento de un fenómeno e intentar que suceda un comportamiento de acuerdo con entradas que se pueden generar. Por ejemplo, el análisis prescriptivo se puede emplear para analizar las estrategias de los políticos y sus impactos en los votantes, con el fin de establecer cuales son las mejores pautas para la campaña y atraer el interés de los votantes. De la misma manera se puede emplear para una campaña de marketing donde se quiere vender un producto y se pueden aprovechar datos de venta de producto similares para tomar las acciones adecuadas según los datos.

Finalmente se tiene el análisis diagnóstico en donde interesa saber por qué se ha producido un evento. Las técnicas de análisis de diagnóstico complementan el análisis descriptivo básico y usan los resultados del análisis descriptivo para identificar la causa de estos eventos. Este análisis es muy común ya que constantemente interesa saber qué es lo que sucede en los procesos para que se dé un resultado determinado. De esta forma se puede establecer procesos de mejora, decisiones que aumenten el desempeño, o que reduzcan costos, entre otras que estén basadas en la necesidad de una empresa o un proceso productivo de una organización.

Existen diferentes roles que desempeñan muchas funciones en el proceso de análisis y extracción de conocimiento a partir de datos. Estos roles pueden variar según la estructura y necesidades específicas de una organización, pero algunos de los roles más comunes en ciencia de datos son:

## Roles en la ciencia de datos

### Científico de Datos

**(Data Scientist):** Este es el rol que se enfoca en el campo de la ciencia de datos. Los científicos de datos son responsables de diseñar y construir modelos predictivos, realizar análisis estadísticos y descubrir patrones en los datos para tomar decisiones informadas.

## Ingeniero de Datos

**(Data Engineer):** Los ingenieros de datos se centran en la construcción y mantenimiento de la infraestructura necesaria para procesar grandes volúmenes de datos. Esto incluye la gestión de bases de datos, la integración de sistemas y la creación de tuberías de datos eficientes.

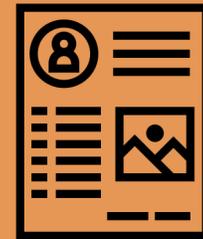
## Analista de Datos



**(Data Analyst):** Los analistas de datos se centran en la interpretación y visualización de datos. Utilizan herramientas estadísticas y de visualización para traducir los datos en información que pueda ser comprendida por no expertos, facilitando así la toma de decisiones.

## Científico de Datos de Negocios

**(Business Data Scientist):** Este rol se enfoca en aplicar técnicas de ciencia de datos para resolver problemas específicos del negocio. Los científicos de datos de negocios trabajan en estrecha colaboración con los equipos de gestión para identificar oportunidades y desafíos comerciales.



## Ingeniero de Machine Learning

**(Machine Learning Engineer):** Estos profesionales se especializan en el desarrollo e implementación de modelos de aprendizaje automático en aplicaciones y sistemas. Su enfoque principal es llevar los modelos de los laboratorios de investigación a la producción.

## Arquitecto de Datos

**(Data Architect):** Los arquitectos de datos diseñan la estructura de los sistemas de almacenamiento de datos, definen cómo los datos se integran y fluyen a través de la organización, y garantizan la coherencia y la seguridad de los datos.



## Especialista en Big Data

**(Big Data Specialist):** Con el auge de grandes volúmenes de datos, los especialistas en Big Data se centran en la gestión y análisis de conjuntos masivos de datos. Trabajan con tecnologías como Hadoop y Spark para procesar y extraer información de datos a gran escala.



## Ingeniero de Optimización

**(Optimization Engineer):** Estos profesionales se enfocan en optimizar algoritmos y procesos para mejorar la eficiencia y rendimiento de los modelos de aprendizaje automático y las aplicaciones relacionadas con la ciencia de datos.

Estos roles pueden variar en sus responsabilidades según la empresa y el proyecto específico, y en muchos casos, los profesionales de la ciencia de datos pueden desempeñar funciones que abarcan más de una de estas categorías. Además, el campo de la ciencia de datos sigue evolucionando, y nuevos roles pueden surgir a medida que las tecnologías y las necesidades empresariales cambian.