

Introducción a la Analítica de datos

Pedro O. Pérez M., PhD.

Herramientas computacionales: el arte de la analítica
Tecnológico de Monterrey

pperezm@tec.mx

10-2021

¿Qué es el análisis de datos?

- ▶ El análisis de datos es el proceso de limpiar, analizar, interpretar y visualizar datos para descubrir información valiosa que impulsa decisiones comerciales más inteligentes y efectivas.
- ▶ Aunque, no sólo es el análisis de datos en sí, sino también la recopilación, la organización, el almacenamiento y las herramientas y técnicas utilizadas para profundizar en los datos, así como las que se utilizan para comunicar los resultados, por ejemplo, las herramientas de visualización de datos.

¿Por qué es importante el análisis de datos?

- ▶ Los datos están en todas partes: en hojas de cálculo, el área de ventas, plataformas de redes sociales, encuestas de satisfacción, tickets de atención al cliente y más. En nuestra era de la información moderna, se crean a velocidades deslumbrantes y, cuando los datos se analizan correctamente, pueden ser el activo muy valioso.
- ▶ El análisis de datos puede ayudar a mejorar aspectos específicos sobre productos y servicios, así como la imagen de marca general, la experiencia del cliente, análisis de experimentos, etc.

Técnicas de análisis de datos

- ▶ Análisis de texto.
- ▶ Análisis descriptivo.
- ▶ Análisis inferencial.
- ▶ Análisis de diagnóstico.
- ▶ Análisis predictivo.
- ▶ Análisis prescriptivo.

Análisis de texto

- ▶ El análisis de texto, también el análisis de texto o la minería de texto, utiliza el aprendizaje automático con procesamiento del lenguaje natural (NLP) para organizar los datos de texto no estructurados de modo que puedan analizarse correctamente en busca de información valiosa. El análisis de texto es una forma de análisis cualitativo que se ocupa de algo más que estadísticas y valores numéricos.
- ▶ El análisis de sentimientos es otro enfoque del análisis de texto, que se utiliza para analizar datos y clasificarlos como Positivo, Negativo o Neutral para obtener un conocimiento profundo sobre cómo se sienten los individuos con respecto a algo.

Análisis descriptivo

- ▶ El análisis de datos descriptivos proporciona el “¿Qué sucedió?” al analizar datos cuantitativos. Es la forma más básica y común de análisis de datos que se ocupa de describir, resumir e identificar patrones a través de cálculos de datos existentes, como media, mediana, moda, porcentaje, frecuencia y rango.

Análisis inferencial

- ▶ El análisis inferencial generaliza o formula hipótesis sobre “¿Qué pasó?” comparando estadísticas de grupos dentro de una población completa: la población de un país, base de clientes existente, pacientes en un estudio médico, etc. Los métodos más comunes para realizar estadísticas inferenciales son las pruebas de hipótesis y las teorías de estimación.

Análisis de diagnóstico

- ▶ El análisis de diagnóstico, también conocido como análisis de causa raíz, tiene como objetivo responder “¿Por qué sucedió 'X'?”. Utiliza conocimientos del análisis estadístico para intentar comprender la causa o la razón detrás de las estadísticas, identificando patrones o desviaciones dentro de los datos para responder por qué.
- ▶ El análisis de diagnóstico puede ayudar a calcular la correlación entre estas posibles causas y los puntos de datos existentes.

Análisis predictivo

- ▶ El análisis predictivo utiliza datos conocidos para postular sobre eventos futuros. Se preocupa por “lo que es probable que suceda”. Utilizado en el análisis de ventas, a menudo combina datos demográficos y datos de compra con otros puntos de datos para predecir las acciones de los clientes.
- ▶ A menudo, hay muchas conjeturas extrapolativas involucradas en el análisis predictivo, pero cuantos más datos tengamos, es probable que la predicción sea más precisa.

Análisis prescriptivo

- ▶ El análisis prescriptivo es la forma más avanzada de análisis, ya que combina todos sus datos y análisis, luego genera una prescripción de modelo: ¿Qué acción tomar? El análisis prescriptivo trabaja para analizar múltiples escenarios, predecir el resultado de cada uno y decidir cuál es el mejor curso de acción en función de los hallazgos.
- ▶ La inteligencia artificial es un ejemplo de análisis prescriptivo que está a la vanguardia del análisis de datos. La IA permite un análisis prescriptivo que puede ingerir y desglosar cantidades masivas de datos y enseñarse a sí mismo de manera efectiva cómo usar la información y tomar sus propias decisiones informadas.

¿Cómo analizar los datos?

- ▶ Si bien puede ser complejo realizar análisis de datos, según el tipo de datos que esté analizando, existen algunas reglas estrictas y rápidas que puede seguir. Incluyen establecer objetivos, recopilar, limpiar y analizar datos, y luego visualizarlos en cuadros de mando llamativos para facilitar la detección de patrones y tendencias.

Pasos para analizar los datos

- ▶ En general, podemos hablar de 7 pasos para analizar los datos:
 1. Define metas: Establecer objetivos claros es clave y ayudará a determinar el tipo de datos que necesitarás recopilar y analizar.
 2. Recopila los datos: Los datos están en todas partes y querrá reunirlos todos en un solo lugar listo para su análisis.
 3. Limpia los datos: Es probable que los datos no estructurados deban limpiarse antes de analizarlos para obtener resultados más precisos. Deshazte del ruido, como caracteres especiales, signos de puntuación, palabras vacías (y, también, ella, ellos), etiquetas HTML, duplicados, etc.
 4. Emplea herramientas de análisis de datos.
 5. Analiza los datos: Ahora que tienes las herramientas de análisis de datos, deberás elegir el tipo de análisis que desea realizar.
 6. Visualiza los datos.
 7. Saca conclusiones.

Algunos ejemplos...

- ▶ Google Tracks Flu Trends
<https://www.youtube.com/watch?v=ytMzI3aphmo>
- ▶ Hans Rosling's 200 Countries, 200 Years, 4 Minutes - The Joy of Stats
<https://www.youtube.com/watch?v=ahp7QhbB8G4&list=PLBE30C2B39FE4BD1C&index=4&t=14s>
- ▶ Coca-Cola: Real-World Data Analytics Example
https://www.youtube.com/watch?v=JIcXC_3Gfow

¿Qué herramienta de análisis usaremos en este curso?

- ▶ Python <https://www.python.org/downloads/>.
- ▶ Anaconda (<https://www.anaconda.com/products/individual-d>) o Miniconda (<https://docs.conda.io/en/latest/miniconda.html>).
- ▶ Jupyter Notebook (<https://www.codecademy.com/articles/setting-up-jupyter-notebook>) o Google Colab.
- ▶ Kaggle (<https://www.kaggle.com/datasets>)

Actividad colaborativa

1. De manera individual,
 - ▶ Instala Anaconda (o Miniconda).
 - ▶ Instala Jupyter Notebook (puedes usar la guía de la filmína anterior).
2. En equipo,
 - ▶ Revisen las bases de datos existentes y seleccionen alguna con la cual les interese trabajar.
 - ▶ Utiliza Excel para revisar la información que contiene la base de datos selecciona. ¿Existe celdas vacías? ¿Caracteres extraños? ¿Signos de puntuación? Tomen esto en cuenta, ya que deberán indicarlos en el reporte a entregar.
 - ▶ En el repositorio de trabajo del equipo, suban la base de datos seleccionada.