



Práctica 2: Exploratory Data Analysis

Presentado por:
Jorge Rafael Martínez Buenrostro

Profesor: René Mac Kinney Romero

México, CDMX, a 20 de octubre de 2025

El análisis exploratorio de datos (EDA) es un paso crucial en el proceso de análisis de datos, ya que permite comprender mejor los conjuntos de datos mediante diversas técnicas. El objetivo principal del EDA es extraer variables significativas, identificar valores atípicos o errores y explorar las relaciones entre las variables, maximizando así los conocimientos y minimizando los posibles errores en futuros análisis.

El EDA cumple dos funciones principales: limpiar el conjunto de datos y mejorar la comprensión de las variables y sus interrelaciones. Este proceso se puede dividir en tres componentes esenciales:

- **Comprensión de la variables:** esto implica comandos como `.shape`, `.head()`, `.columns`, `.nunique` y `.describe()` para obtener información sobre la estructura, los valores y las estadísticas básicas del conjunto de datos. Es fundamental comprender lo que representan los datos antes de sumergirse en el análisis.
- **Limpieza del conjunto de datos:** este paso incluye eliminar variables redundantes, seleccionar columnas pertinentes (eliminando aquellas con un número elevado de valores nulos), eliminar valores atípicos basándose en rangos lógicos y descartar filas con datos faltantes. También puede implicar reclasificar variables discretas para mejorar la claridad.
- **Análisis de las relaciones entre variables:** Este componente utiliza visualizaciones, como matrices de correlación (mapas de calor) para discernir relaciones generales, diagramas de dispersión para investigar relaciones específicas entre dos variables, histogramas para examinar la distribución de variables individuales y diagramas de caja para identificar valores atípicos y obtener información sobre la distribución de los datos.

Estas representaciones visuales facilitan la comunicación rápida de información compleja y son esenciales para seleccionar modelos adecuados y perfeccionar el proceso general de análisis de datos.