# **Pasos en el análisis exploratorio de datos**

La tarea de analizar datos utilizando gráficos y estadísticas simples se denomina **EDA**. Esto se denomina Exploratorio porque no sabemos nada sobre los datos, necesitamos comprenderlos.

## Pasos para explorar EDA

**Importar conjunto de datos**

1. Intente comprender la información presente en el conjunto de datos. Obteniendo la forma del conjunto de datos, nombres de columnas, información sobre variables dependientes e independientes.
2. Analiza las métricas básicas. Mín, Máx, Mediana, Media, estándar, percentiles: esto se puede hacer usando la función de descripción.
3. Análisis univariante no gráfico: obtenga el recuento de valores únicos. Con esto obtenemos si los datos están equilibrados o desequilibrados.
4. Rellenar valores nulos.
5. Análisis gráfico univariante. Los histogramas brindan información sobre la distribución de datos y los valores atípicos presentes. Los diagramas de caja brindan información estadística como primer cuartil, segundo cuartil, tercer cuartil, máximo y cualquier valor atípico presente.
6. Análisis bivariante. Gráficos de pares (el análisis de gráficos de pares es fácil de entender cuando la dimensionalidad de los datos es pequeña, si no es así, necesitamos analizar los gráficos de pares nc2, lo cual es un desafío).
7. Tratar los valores perdidos (missing values), valores atípicos.
8. Análisis de correlación.
9. Reducción de dimensionalidad.

.shape returns the number of rows by the number of columns for my dataset. My output was (525839, 22), meaning the dataset has 525839 rows and 22 columns.

.head() returns the first 5 rows of my dataset. This is useful if you want to see some example values for each variable.

.columns returns the name of all of your columns in the dataset.