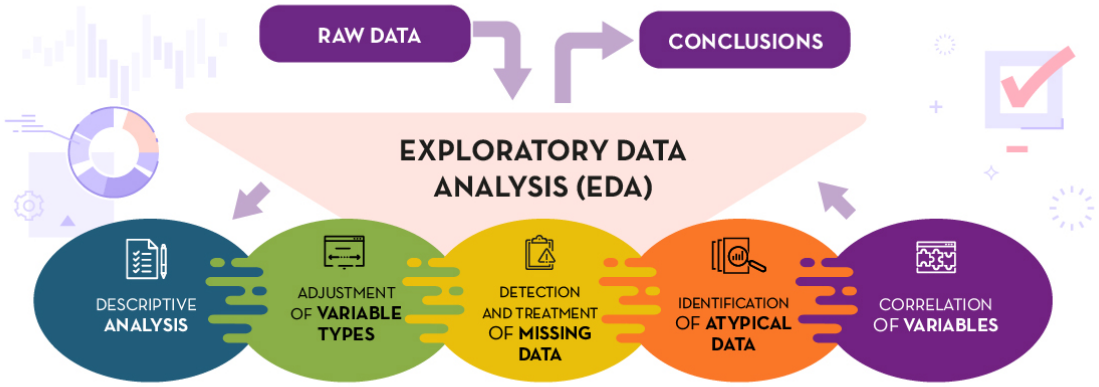
What is Exploratory Data Analysis?

Exploratory Data Analysis (EDA) is understanding the data sets by summarizing their main characteristics and often plotting them visually. This step is very important especially when we arrive at modelling the data in order to apply Machine learning. Plotting in EDA consists of Histograms, Box plots, Scatter plots and many more. It often takes much time to explore the data. Through the process of EDA, we can ask to define the problem statement or definition on our data set which is very important.



Source: https://datos.gob.es/en/documentacion/practical-introductory-guide-exploratory-data-analysis

How to perform Exploratory Data Analysis?

This is one such question that everyone is keen on knowing the answer to. Well, the answer is it depends on the data set that you are working on. There is no one method or common method in order to perform EDA, whereas in this case study you can understand some common methods and plots that would be used in the EDA process.

Important points to remember about EDA

- EDA is applied to investigate the data and summarize the key insights.
- Through EDA you can gain a basic understanding of your data, its distribution and process the data to get insights
- You can either explore data using graphs or through some python functions.
- There will be two types of analysis. Univariate and Bivariate. In the univariate, you will be analyzing a single attribute. But in the bivariate, you will be analyzing an attribute with the target attribute.
- In the non-graphical approach, you will be using functions such as shape, summary, describe, isnull, info, datatypes and more.
- In the graphical approach, you will be using plots such as scatter, box, bar, density and correlation plots.