# **SCRIPTS TO CREATE EXPLORATORY AND RESULTS VISUALIZATION:**

The purpose of exploratory data analysis (EDA) is to gain insights and understanding about the underlying structure and relationships within the dataset.

## **Key purposes of conducting exploratory analysis:**

1. **Data Understanding**: helps in understanding the data at hand before proceeding with more advanced analyses.

2. **Identifying Patterns and Trends**: helps in identifying patterns and distributions within the data which includes understanding the central tendencies, variability, and shape of the data distributions.

3. **Detecting Anomalies and Outliers**: enables the detection of anomalies, outliers, and missing values within the dataset. These may indicate errors in data collection or preprocessing, or they may represent interesting insights worth further investigation.

4. **Feature Selection and Engineering**: It helps in determining which features are most important or influential in explaining the target variable, as well as identifying potential interactions or transformations that could improve model performance.

5. **Assumption Checking**: allows for the examination of assumptions underlying statistical models. For example, it helps in assessing whether the data meet the assumptions of linear regression, such as linearity, normality etc.

6. **Visualization and Communication**: involves the creation of visualizations and summary statistics to effectively communicate findings to stakeholders.

7. **Hypothesis Generation:** often leads to the formulation of hypotheses or research questions that can be tested through further analysis. It helps in guiding the direction of subsequent analyses and experiments.

Overall, the primary purpose of exploratory analysis is to serve as a foundation for making informed decisions and building predictive models in data science projects.

**[**The following explorations and visualizations are implemented and stored in separate files. Refer the “**visualizations**” folder in “**src**” parent directory**]**

# **UNIVARIATE ANALYSIS:**

Involves the examination of a single variable at a time such as the distribution, central tendency and the variability of individual variable without considering their relationships with the other variables.

**sweetviz report:**

The sweetviz library is applied to perform a quick data review, data comparison, feature analysis and target analysis.

*# importing sweetviz library*

import sweetviz as sv

*# load the data*

import pandas as pd

data = pd.read\_csv("your\_dataset.csv")

*# Generate the report*

report = sv.analyze(data)

*# Save the report to an HTML file*

report.show\_html("report.html")

**Univariate plots:**

* Histogram
* KDE plot
* Count plot
* Box plot
* Violin plot

# **BIVARIATE ANALYSIS:**

Bivariate analysis is a statistical method used to analyse the relationship between two variables. Unlike univariate analysis, which focuses on a single variable, bivariate analysis examines how two variables are related to each other.

**Bivariate plots:**

* Line plot
* Scatter plot
* Rel plot
* Box plot
* Violin plot
* Bar plot
* Strip plot
* Swarm plot
* Cat plot

# **MULTIVARIATE ANALYSIS:**

Multivariate analysis is a statistical technique used to analyse and understand the relationships between multiple variables simultaneously.

**Multi-variate plots:**

* Pair plot
* Heatmap etc.,