**Name:** Swaraj Pawar

**Roll no:** 282072

**Batch:** B3

**Assignment 2**

**Problem Statement:** The objective of this project is to perform exploratory data analysis (EDA) on a given dataset, including computing summary statistics, visualizing data distributions, and then building a machine learning classification model. The dataset contains multiple features, and the goal is to understand the characteristics of the data and create a model that can classify instances accurately.

**Software used:**

1. Python 3.x
2. Google colab

**Libraries and packages used:** NumPy, pandas, matplotlib, sklearn

**Theory:**

**Methodology:**

1. Summary statistics: Computing summary statistics helps in understanding the basic properties of each feature in the dataset, such as mean, standard deviation, minimum and maximum values, percentiles, etc.
2. Data visualization: Creating histograms for each feature provides insights into the distribution of data, revealing patterns, skewness, and potential outliers.
3. Data cleaning, Integration, Transformation: These steps involve handling missing values, encoding categorical variables, scaling features, etc., to prepare the data for modeling.
4. Model Building: Building a classification model using machine learning algorithms such as Decision Trees, Random Forests, or Support Vector Machines.

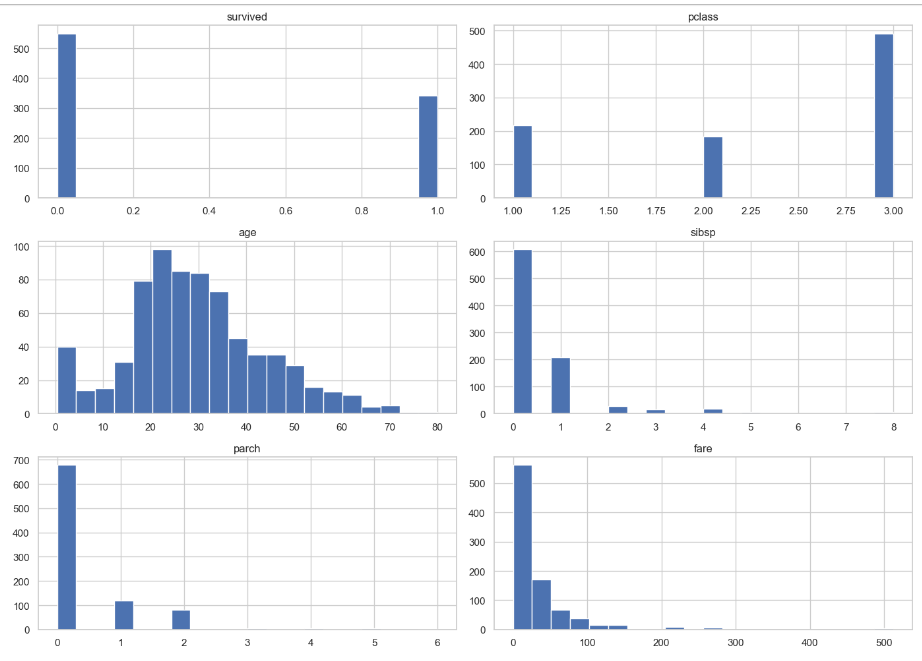
**Working/ Algorithm:**

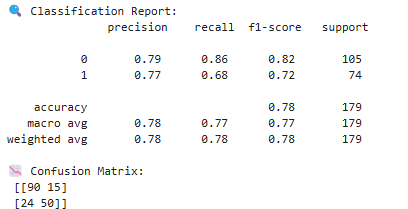
1. Load the dataset using Pandas.
2. Compute summary statistics using the **describe()** function.
3. Visualize data distributions using histograms with Matplotlib and Seaborn.
4. Perform data cleaning, integration, and transformation as necessary.
5. Build a machine learning classification model using Scikit-learn.
6. Evaluate the model's performance using appropriate metrics such as accuracy, precision, recall, etc.

**Diagram:**

****

# ****Data Visualization****





**Conclusion:** In conclusion, this project demonstrates the importance of exploratory data analysis and machine learning modeling in understanding and extracting insights from data. By following a systematic approach, we can gain valuable insights into the data, identify patterns, and build predictive models that can be applied to real-world problems across various domains.