**Titanic Survival Prediction Report**

*Task 1: Codsoft project*

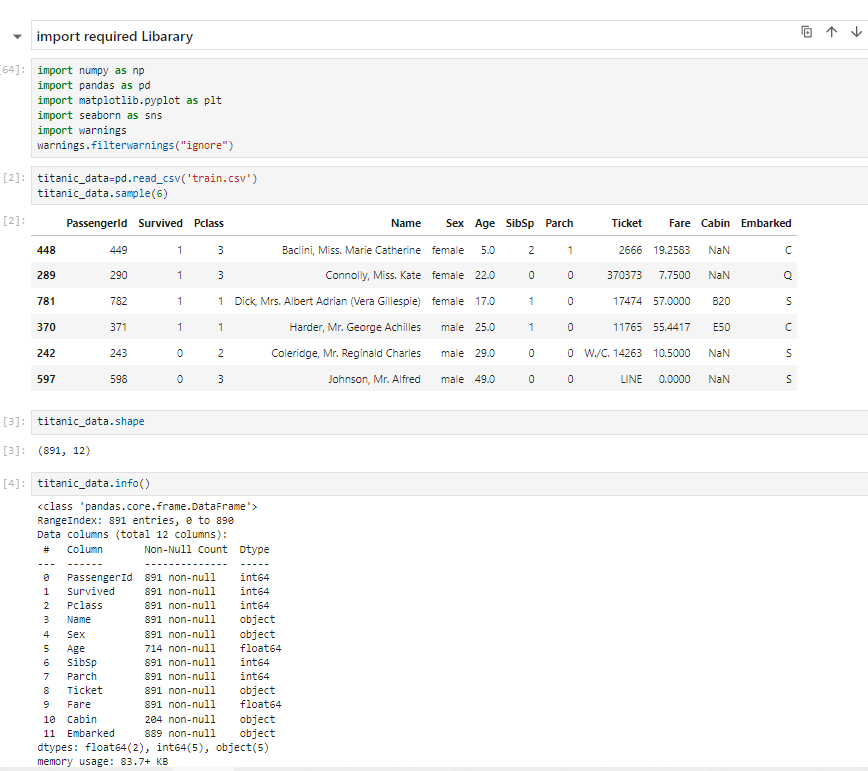
**Problem Statement:** It aims to develop a robust machine learning model that can accurately predict whether a passenger aboard the Titanic survived or not, based on a set of input features. The dataset contains information about various passengers, including their age, gender, class, ticket fare, number of siblings or spouses aboard, number of parents or children aboard, and other relevant attributes.

The tragic sinking of the Titanic in 1912 resulted in a significant loss of lives, and understanding the factors that contributed to survival can provide valuable insights. The challenge is to leverage the available data to build a predictive model that can generalize well to unseen data and accurately classify passengers as survivors or non-survivors.

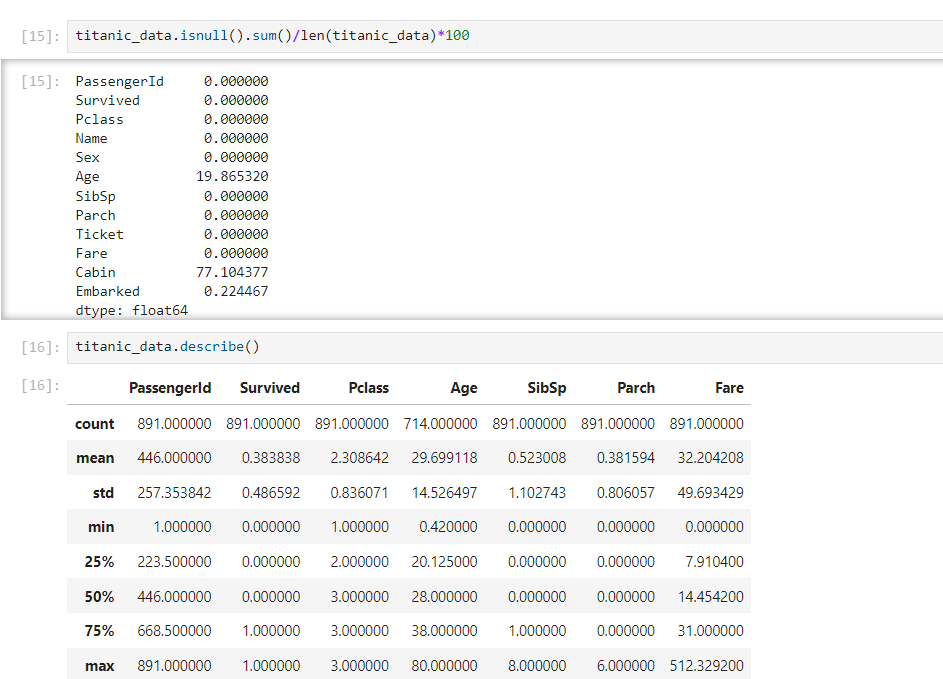


**Process of Analysis:**

1. **Data Acquisition:** I have procured the Titanic Survival Prediction dataset directly from Kaggle. This dataset, in CSV format, encapsulates critical information about passengers. Key variables include PassengerId, Pclass (Ticket class), Fare (Ticket fare), Ticket (Ticket number), Survival (Survival status), Age (Age of the passenger), and Sex (Gender). The acquisition process involved loading the dataset into a Pandas DataFrame using Python. To ensure data integrity, missing values, particularly in the Age column, were addressed through imputation. This dataset is now ready for exploration and subsequent analysis.

****

1. **Data Preprocessing:** In the second phase of Analysis (EDA) is where I dive into the data details. I look closely at things like passenger IDs, ticket classes, fares, and survival status. I use graphs to see patterns and relationships, and I make sure the data is complete. If something seems out of the ordinary, like extreme values, I take note. This phase helps me understand the story the data is telling and guides decisions for the next steps in building the survival prediction model.

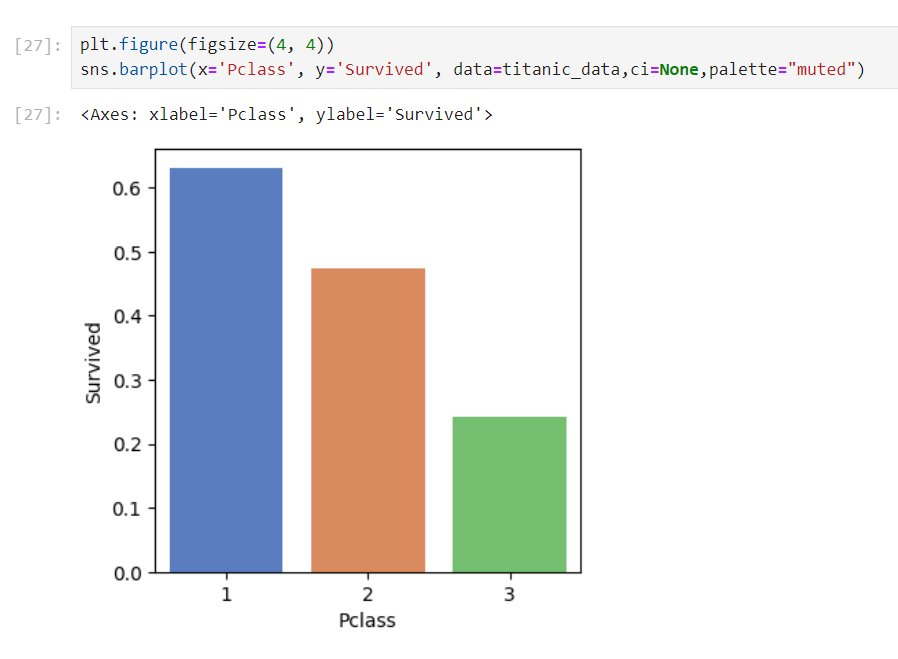




A white rectangular frame with colorful text

Description automatically generated

1. **Data Exploration and Visualization**: In the third phase of my Titanic Survival Prediction project, the spotlight shifts to data visualizations, where I aim to bring the dataset to life through meaningful graphs and charts. This involves creating visual representations to gain deeper insights into the patterns and relationships within the data.

A screenshot of a computer screen

Description automatically generatedA screenshot of a computer screen

Description automatically generated

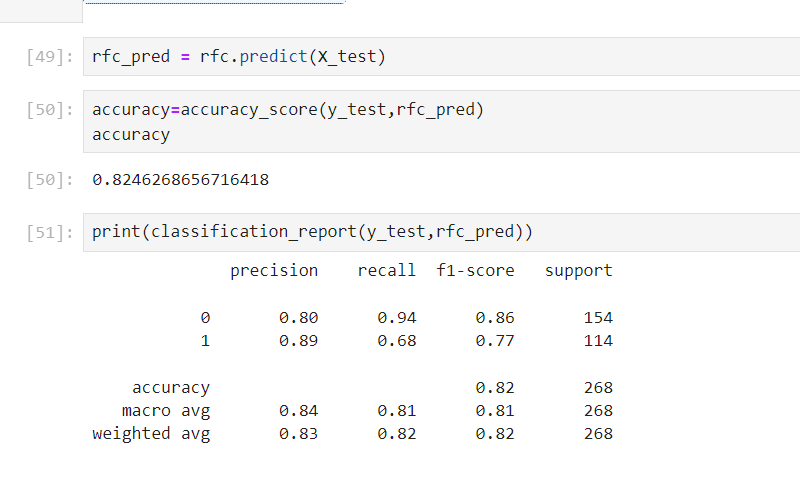
1. **Building Model: - LogisticRegression & RamdomForestClassifier**

****

**A screenshot of a computer program

Description automatically generated**

1. **Model Evaluation:** Model evaluation involves assessing the performance of the trained models using metrics such as accuracy, precision, recall, F1 score, and ROC-AUC.

****

**A screenshot of a computer

Description automatically generated**

1. **Interactive Dashboard:** the focus is on data visualizations; I've gone beyond creating static graphs and charts. I've developed an interactive dashboard to provide a dynamic and user-friendly exploration of the dataset. Utilizing tools like Plotly, I've designed a dashboard that allows users to interactively navigate through various features, zoom in on specific data points, and gain real-time insights.

To enhance the report and make the findings more tangible, I've included a screenshot of the interactive dashboard.

