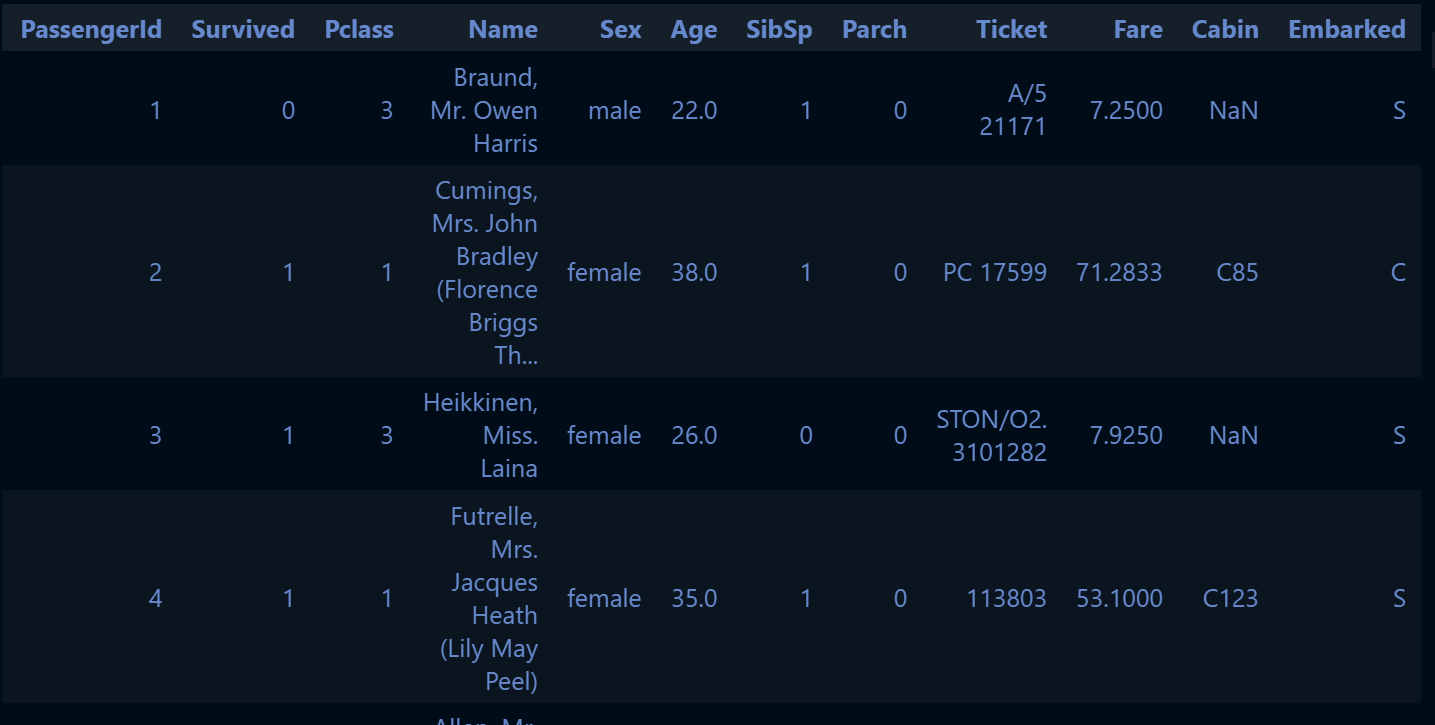
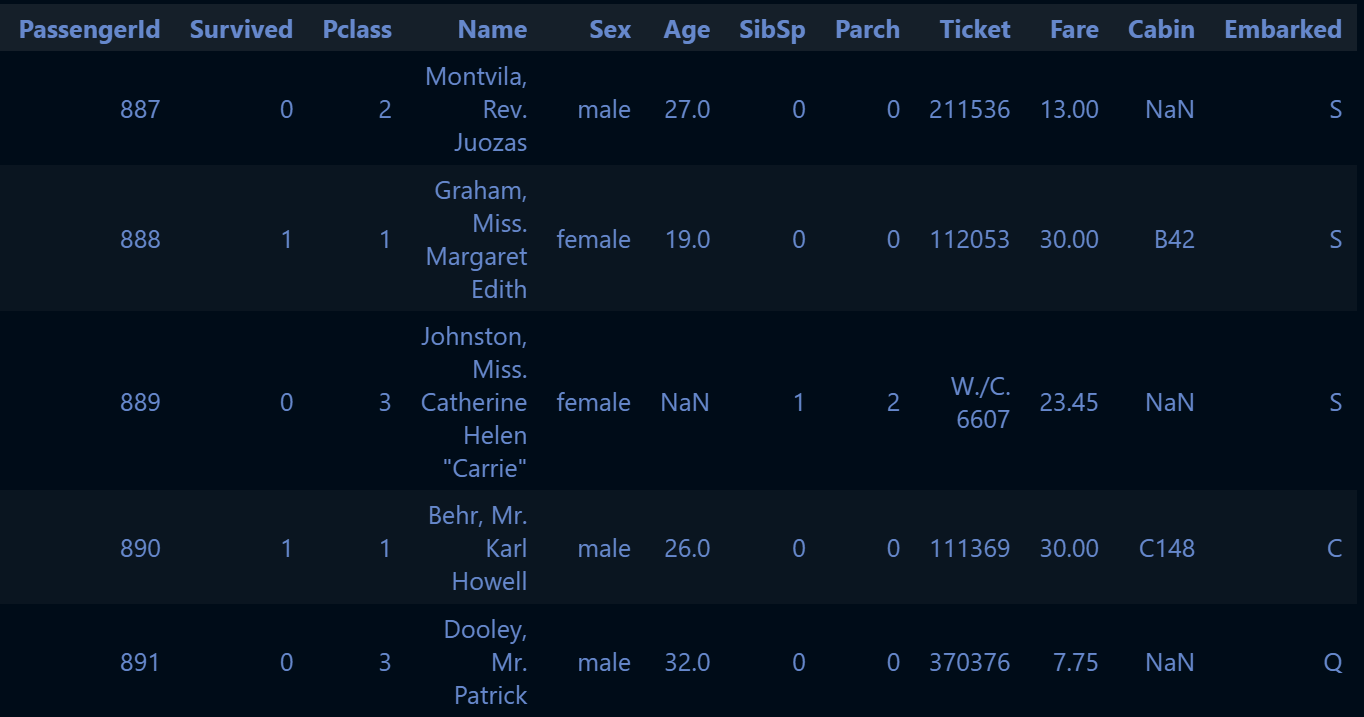
**TITANIC SURVIVAL PREDICTION**

Dataset reference -> (<https://www.kaggle.com/datasets/yasserh/titanic-dataset>)

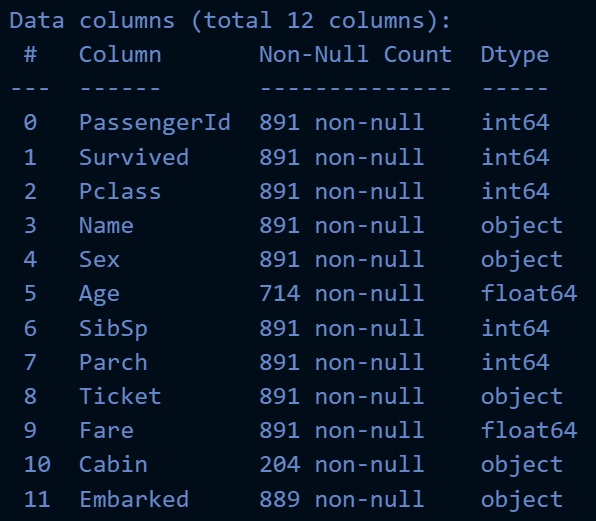
We used the titanic dataset to analyse the hidden information about the survival of the passengers.

* The dataset has 12 Columns Namely, ['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp','Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked']
* The first five and last Five rows of the data using pandas Head() and Tail() Functions.

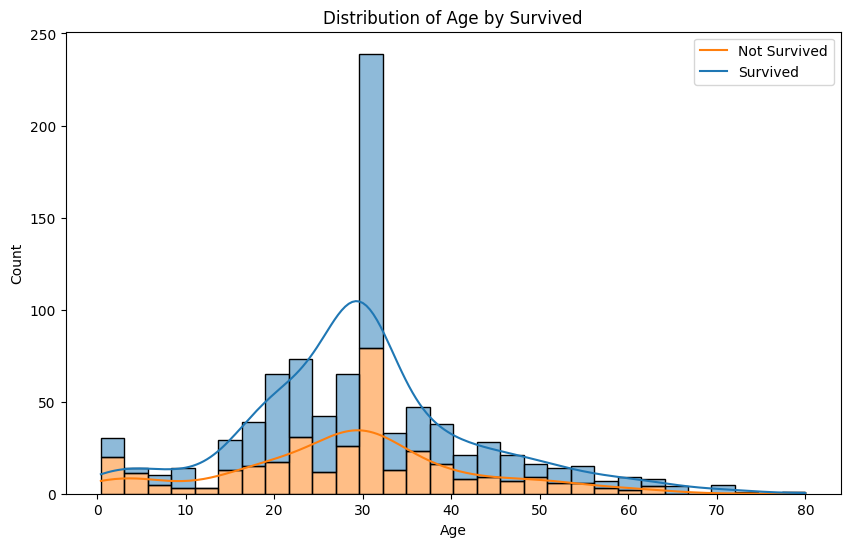


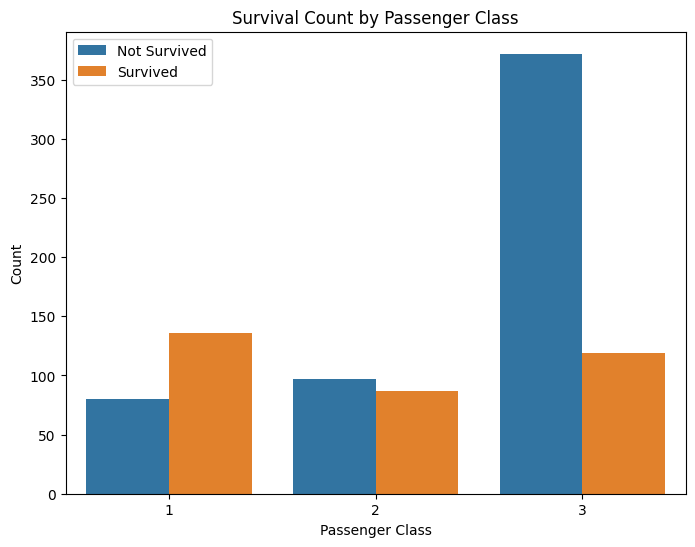


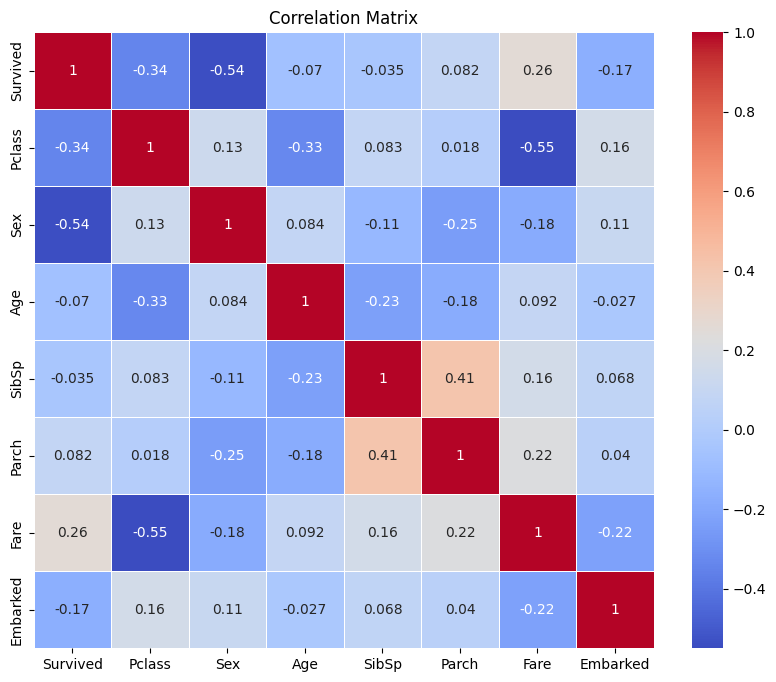
* Shape of the dataset is (891, 12) i.e. there are 891 different tuples representing unique passengers
* Info about the columns is:



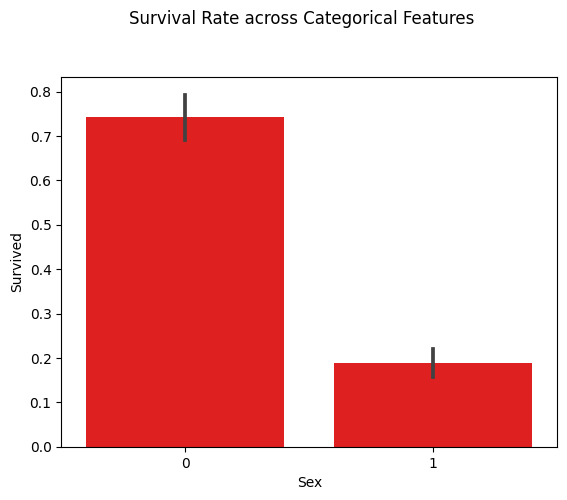
* We handled the null values using mean simple imputer from scikit-learn library and dropped the irrelevant columns like [ 'PassengerId','Name','Ticket','Cabin' ]
* Now the new data has only 8 columns namely, ['Survived', 'Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']



* The above Age Vs count histogram distribution shows the count of survived and not survived passengers of different age groups. We can see that the proportion of survival of passenger is more in the age group between 15 to 40.
* The below bar plot shows the survival and not survival count of passengers of different classes. We can see that the passengers of all classes have same survival count but the number of not survivors in the 3rd class is quite high.



* The above correlation matrix shows the correlation of different columns with each other.



* The above plot shows the survival rate of different genders; we can see that the 0 class has very high rate of survival unlike the 1 class.

**Author:**

**Name: Asutosh Kataruka**

**ID: B122029**

**7th Semester CSE Dept.**

**IIIT Bhubaneswar**