A PRELIMENERY REPORT

ON

# “Titanic survivor prediction”

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IIN HE PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE OF

## BACHELOR OF ENGINEERING (COMPUTER ENGINEERING)

## GUIDED BY

## Prof Bhosale.S.S

**SUBMITTED BY**

GAWALI AKANKSHA KUNDLIK (72155648F)



## DEPARTMENT OF COMPUTER ENGINEERING

**HSBPVT’s FACULTY OF ENGINEERING , KASHTI**

SAVITRIBAI PHULE PUNE UNIVERSITY

SUBMISSION: 2023-2024

****

## CERTIFICATE

This is to certify that the mini project report on

**“Titanic survivor prediction”**

SUBMITTED BY,

This is to certify that **Ms. Gawali Akanksha Kundlik** has successfully completed the mini project work entitled “**Titanic survivor prediction**” under my supervision, in the partial fulfillment of Bachelor in Engineering (Computer) of Savitribai Phule Pune University, Pune.

Guide (Bhosale S. S.) Head (Hirnawale S. B.)

Department of Computer Engineering Department of Computer Engineering

Place: Kashti

Date:

# ACKNOWLDGEMENT

The present world of competition there is a race of existence in which those are having will to come forward succeed. Project is like a bridge between theoretical and practical working. First of all, I would like to thank the supreme power the Almighty God who is obviously the one has always guided me to work on the right path of life.

I am indebted to our project guide **Miss. Bhosale S.S**, Department of Computer Science of faculty of engineering, kashti. I feel it’s a pleasure to be indebted to our guide for his valuable support, advice and encouragement and thankful for HOD to cooperate I think him for his superb and constant guidance.

Ms . Gawali Akanksha Kundlik

# INDEX

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Table of Contents** | **Page Number** |
|  | Introduction | **5** |
|  | Abstract | **6** |
|  | Problem Definition | **7** |
|  | System Architecture | **8** |
|  | Description | **9** |
|  | Implementation | **12** |
|  | Conclusion | **16** |
|  | References | **17** |

**INTRODUCTION**

Using the well-known Titanic dataset as a starting point, I would create a machine learning model. This provides a prognosis of the Titanic's likelihood of surviving, taking into consideration a number of variables like economic standing (class), sex, age, etc.

This is taken into account, and many features are compared and found to have relationships in order to estimate whether a passenger would survive on the Titanic. because it is a component of "Titanic: Machine Learning from Disaster." In this exercise, we must determine whether a Titanic passenger would have survived or not.

The RMS Titanic was the largest ship a float at the time it entered service and was the second of three Olympic-class ocean liners operated by the White Star Line. The Titanic was built by the Harland and Wolff shipyard in Belfast. Thomas Andrews, her architect, died in the disaster.

The RMS Titanic was a British passenger liner that sank in the North Atlantic Ocean in the early morning hours of 15 April 1912, after it collided with an iceberg during its maiden voyage from

# PROBLEM DEFINITION

## Titanic Survival Prediction Using Machine Learning:

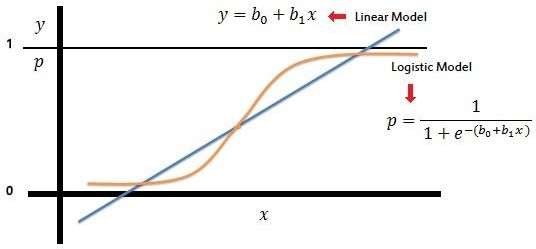
Build a machine learning model that predicts the type of people who survived the titanic shipwreck using passenger data (i.e. name, age, gender, socioeconomic class, etc).

# SYSTEM ARCHITECTURE

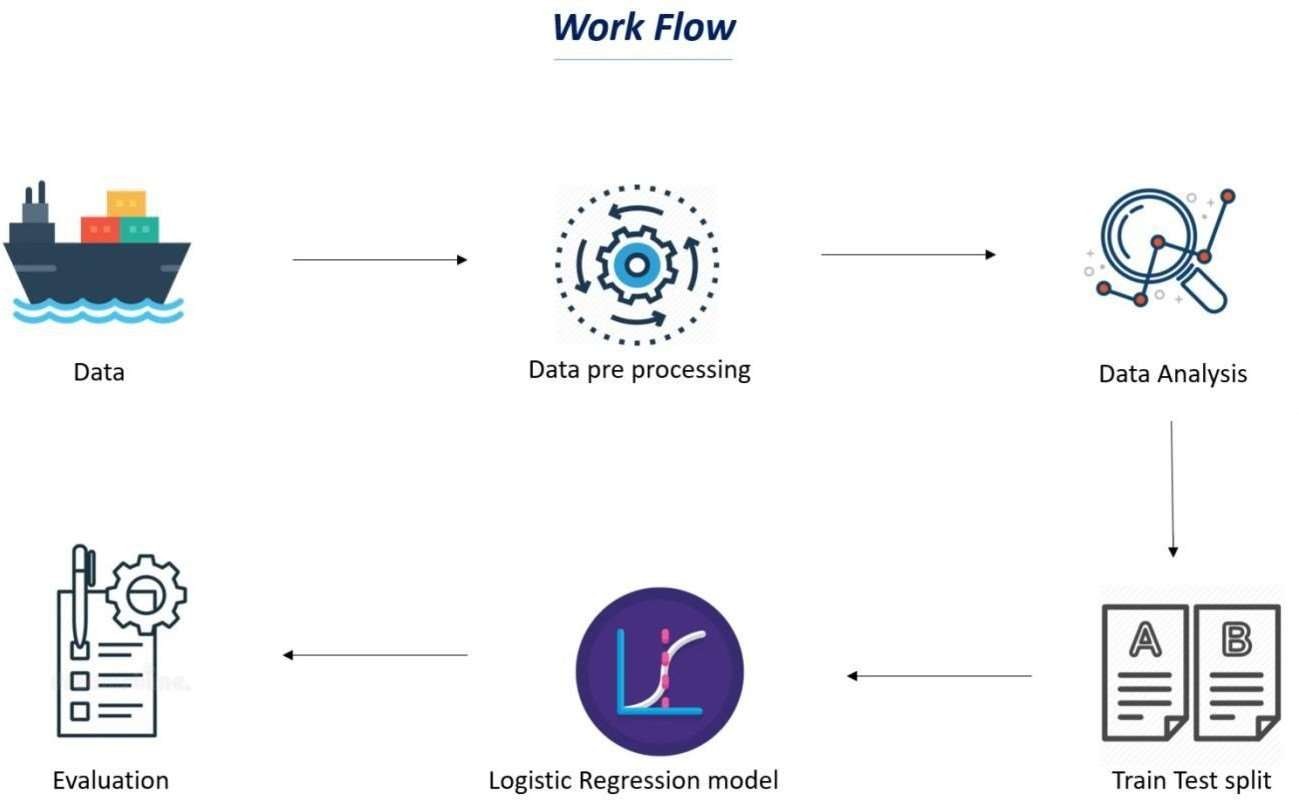
I will be understanding, how to analyze and predict, whether a person, who had boarded the RMS Titanic has a chance of survival or not, using Machine Learning’s Logistic Regression model.

## Brief description about Logistic Regression:

A simple yet crisp description of Logistic Description would be, “it is a supervised learning classification algorithm used to predict the probability of a target variable. The nature of target or dependent variable is dichotomous, which means there would be only two possible classes.” The graph of logistic regression is as shown below:



For better understanding, let’s split the task into smaller parts and depict them in a work flow as shown below :



As I now know what I have to do, to accomplish this task, I shall begin with the very first and the most important thing needed in machine learning, a **Dataset.**

**What is a dataset:**

A data set, as the name suggests, is a collection of data. In Machine Learning projects, we need a training data set. It is the actual data set used to train the model for performing various actions.

# DESCRIPTION

**Data Set Column Descriptions:**

* **pclass:** Passenger Class (1 = 1st; 2 = 2nd; 3 = 3rd)
* **survived:** Survival (0 = No; 1 = Yes)
* **name:** Name
* **sex:** Sex
* **age:** Age
* **sibsp:** Number of siblings/spouses aboard
* **parch:** Number of parents/children aboard
* **fare:** Passenger fare (British pound)
* **embarked:** Port of embarkation (C = Cherbourg; Q = Queenstown; S = Southampton)
* **adult\_male:** A male 18 or older (0 = No, 1=Yes)
* **deck:** Deck of the ship
* **who:** man (18+), woman (18+), child (<18)
* **alive:** Yes, no
* **embarked\_town:** Port of embarkation ( Cherbourg, Queenstown, Southampton)
* **class:** Passenger class (1st; 2nd; 3rd)
* **alone:** 1= alone, 0= not alone ( you have at least 1 sibling, spouse, parent or child on board) **age:**

Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

## Sibsp:

The dataset defines family relations in this way:

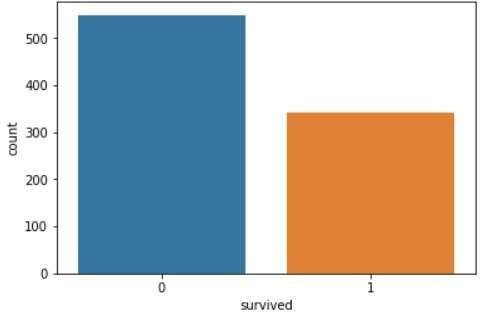
* Sibling= brother, sister, stepbrother, stepsister
* Spouse= husband, wife (mistresses and fiancés were ignored)

## Parch:

The dataset defines family relations in this way:

* Parent= mother, father
* Child= daughter, son, stepdaughter, stepson

Some children traveled only with a nanny, therefore parch=0 for them.

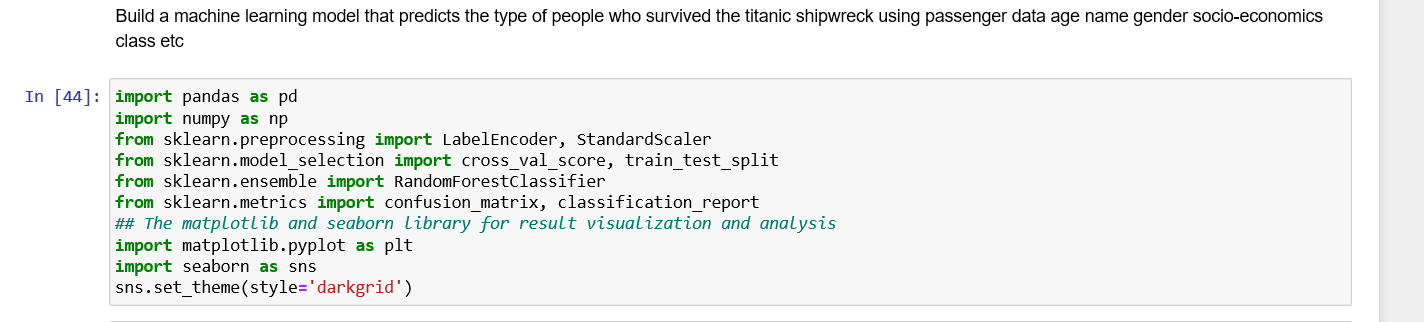


Visualize the count of survivors for the columns who*,* sex*,* pclass*,* sibsp*,* parch*,* and embarked*.*

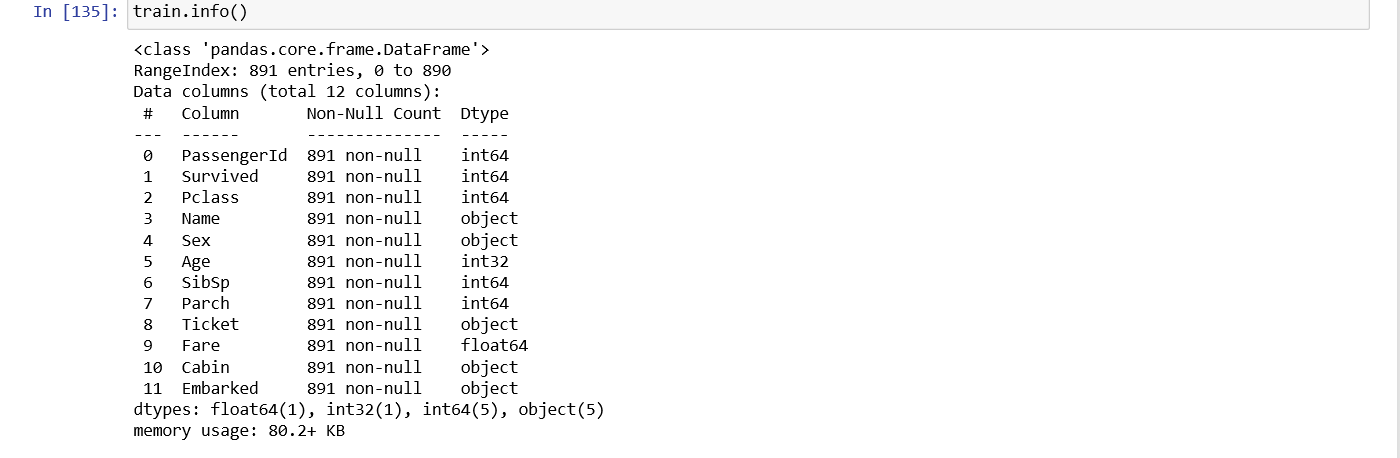
* From the charts below, we can see that a man (a male 18 or older) is not likely to survive from the chart who.
* Females are most likely to survive from the chart sex.
* Third class is most likely to not survive by chart pclass.
* If you have 0 siblings or spouses on board, you are not likely to survive according to chart sibsp.
* If you have 0 parents or children on board, you are not likely to survive according to the parch chart.
* If you embarked from Southampton (S), you are not likely to survive according to the embarked chart.

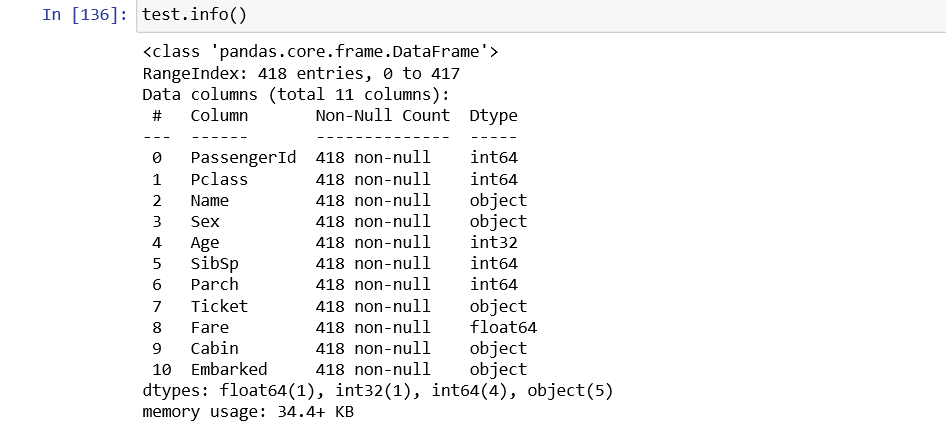
# IMPLEMENTATION

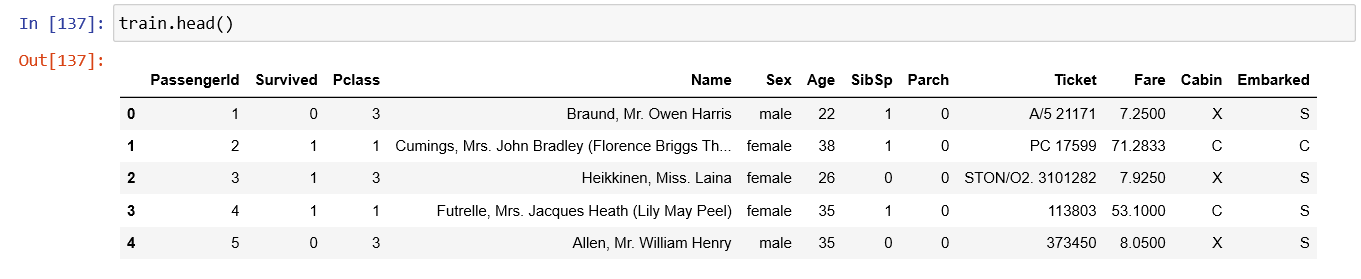
**Start Programming:** Now import the packages /libraries to make it easier to write the program.

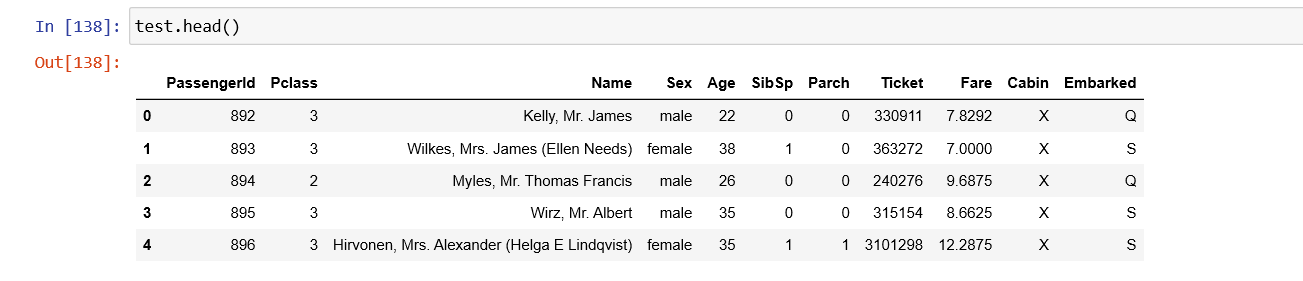


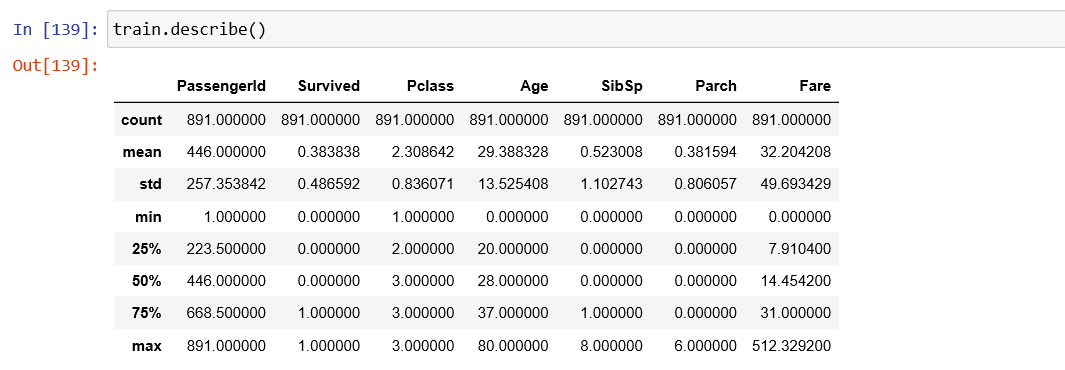


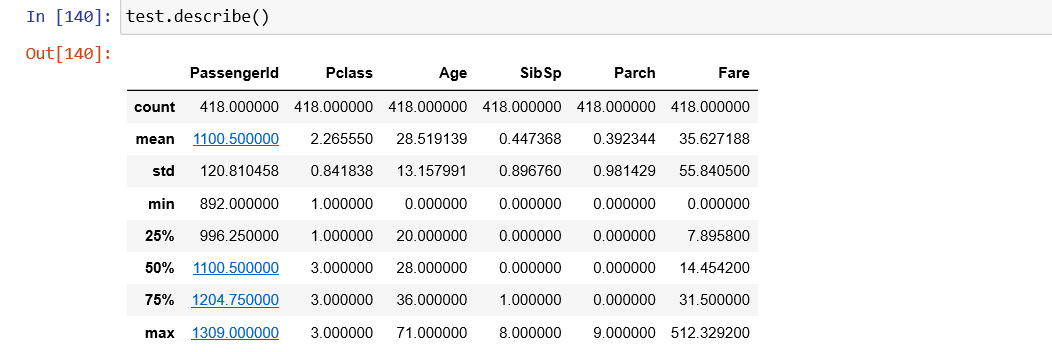


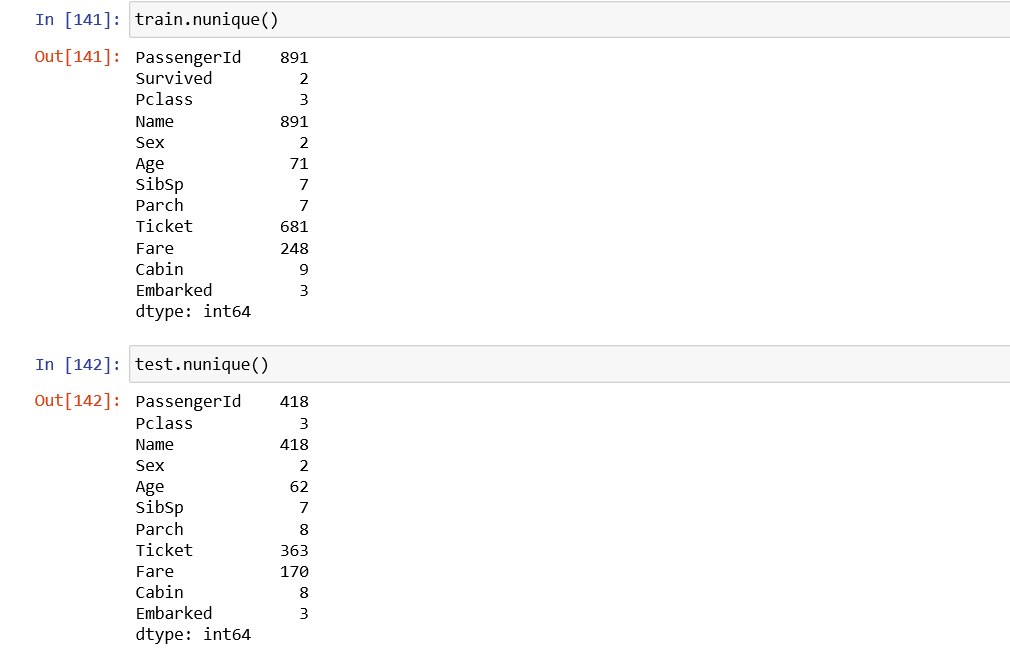


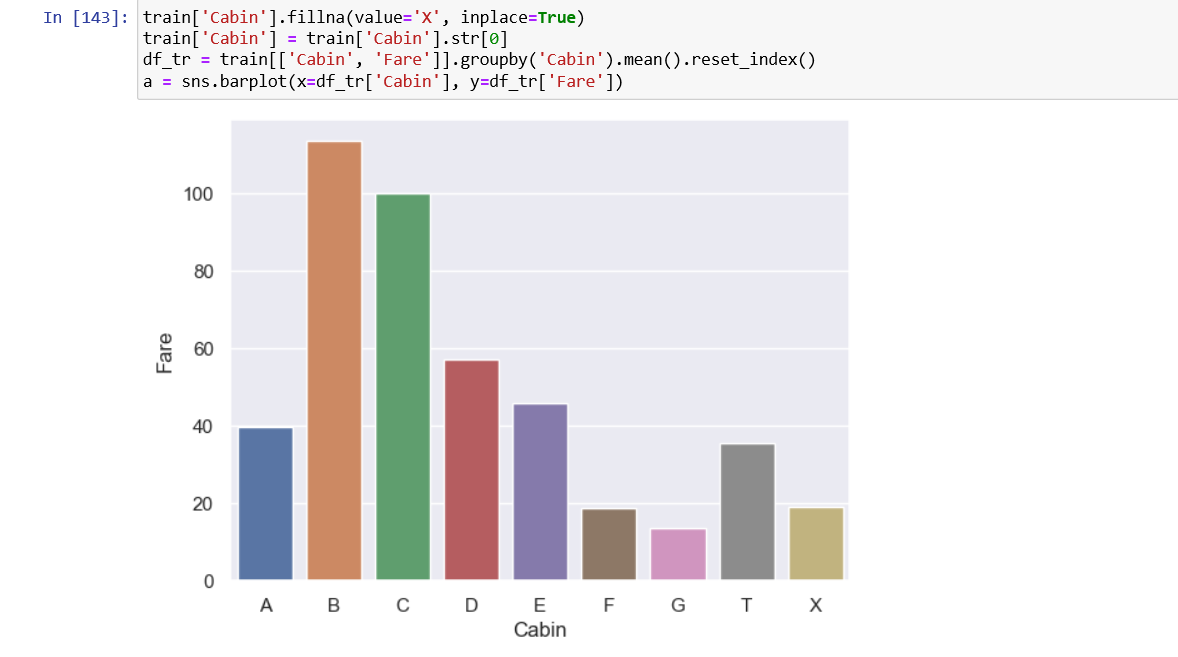


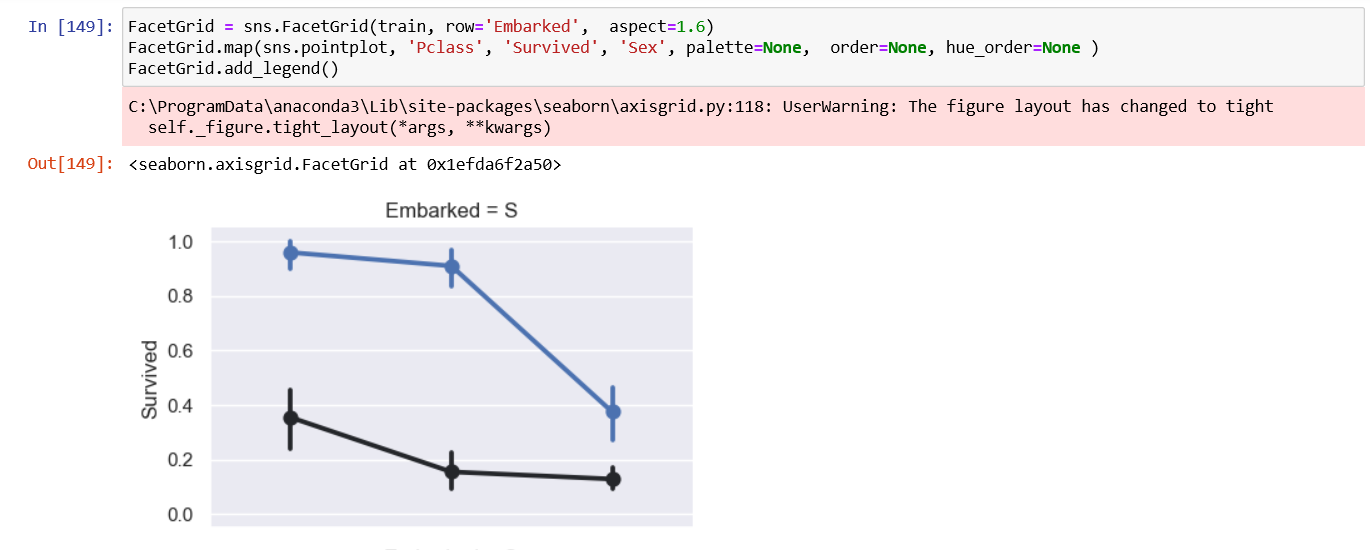


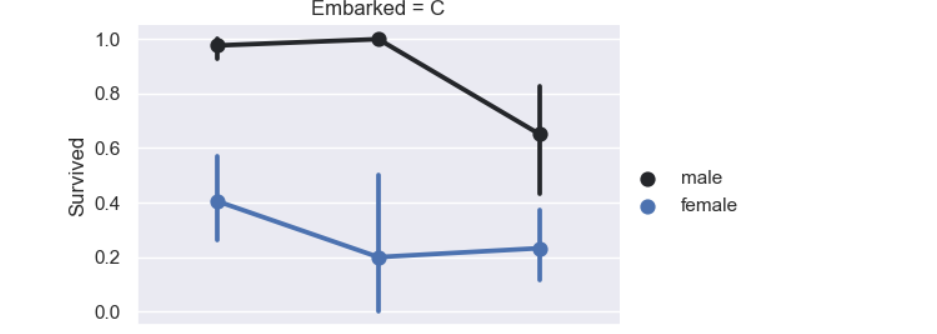


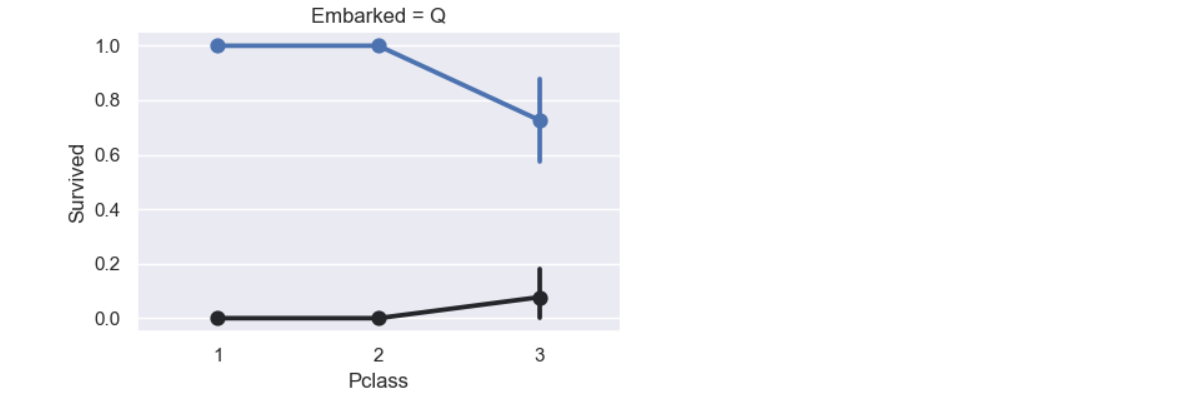


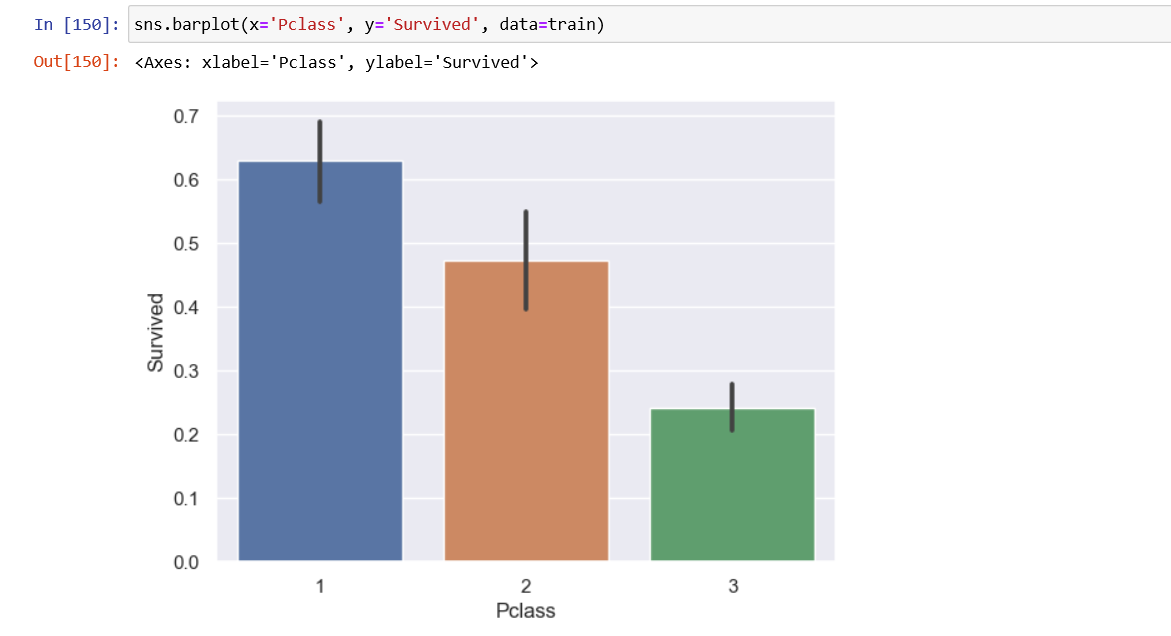




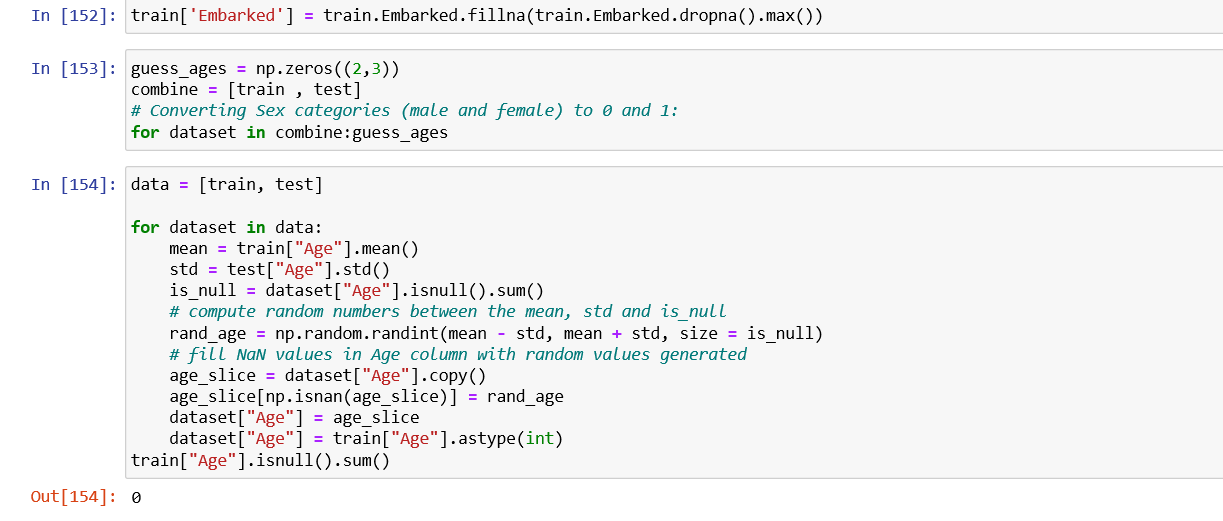




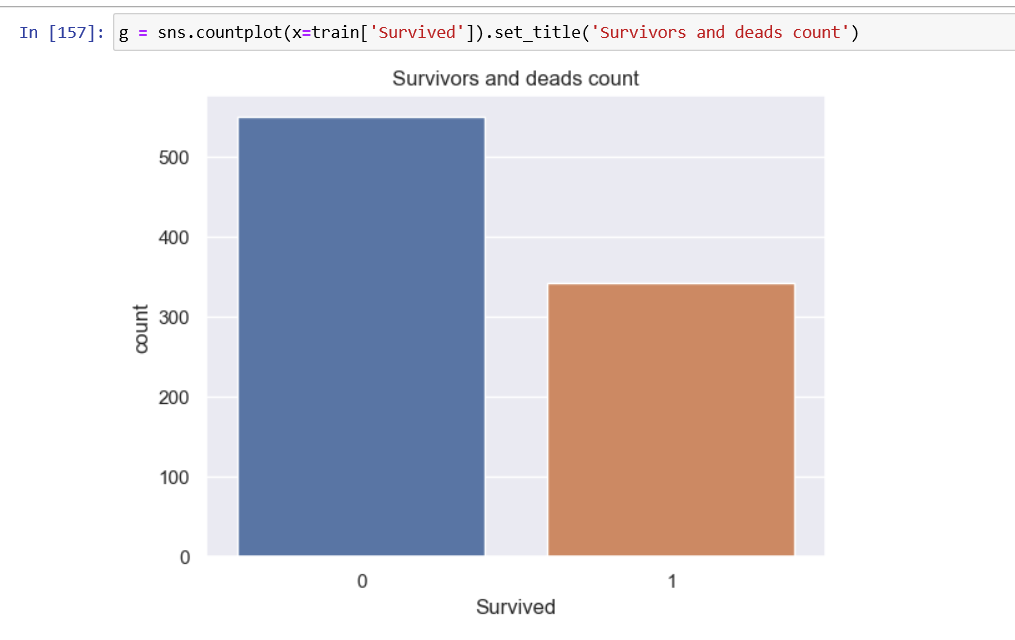


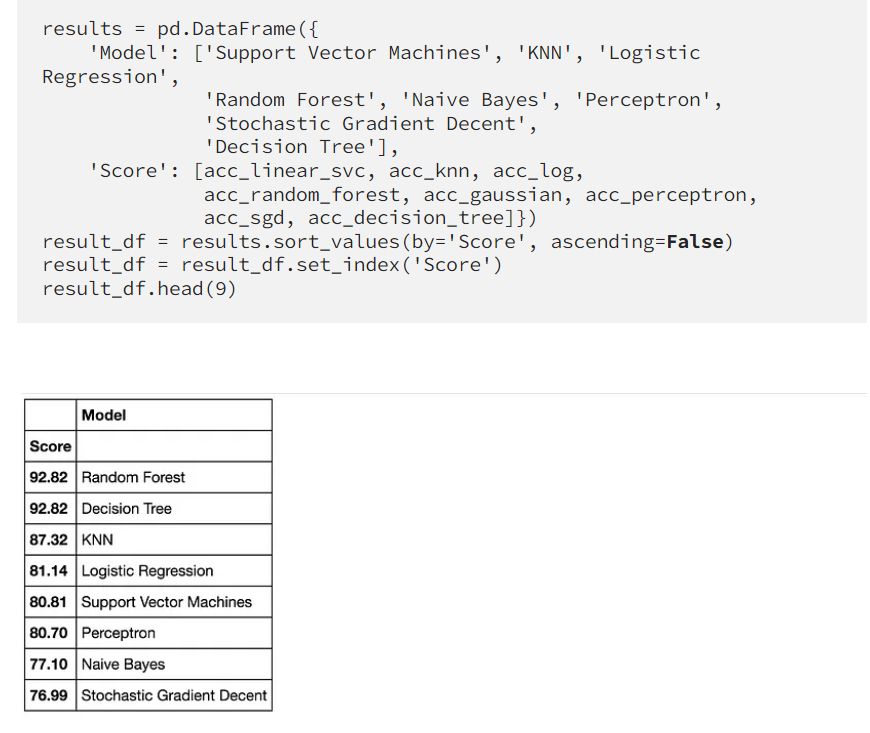


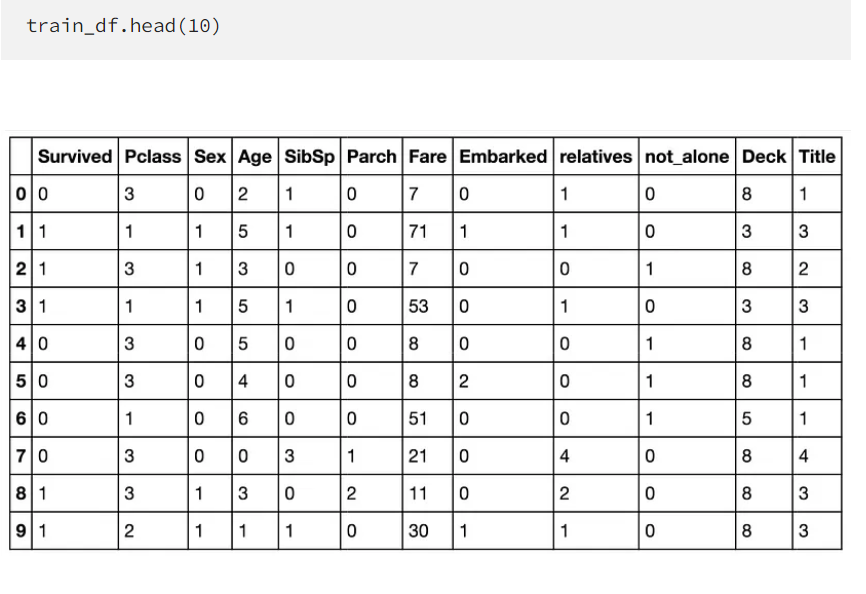












# EXPECTED OUTPUT

# 

# CONCLUSION:

We started with the data exploration where we got a feeling for the dataset, checked about missing data and learned which features are important. During this process we used seaborne and matplotlib to do the visualizations. During the data preprocessing part, we computed missing values, converted features into numeric ones, grouped values into categories and created a few new features.

**REFERENCES:**

1.Michalski R S,et al.Machine Learning:Challenges of the eighties.Machine Learning,1986,99-102.

2.Analyzing Titanic disaster using machine learning algorithms-Computing, Communication and Automation (ICCCA), 2012 International Conference on 21 December 2017, IEEE.

3.Eric Lam, Chongxuan Tang, "Titanic Machine Learning From Disaster", LamTang-Titanic Machine Learning From Disaster, 2012.