Titanic Data Set

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history.  On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.

Watch the Video of the Ship sinking here-



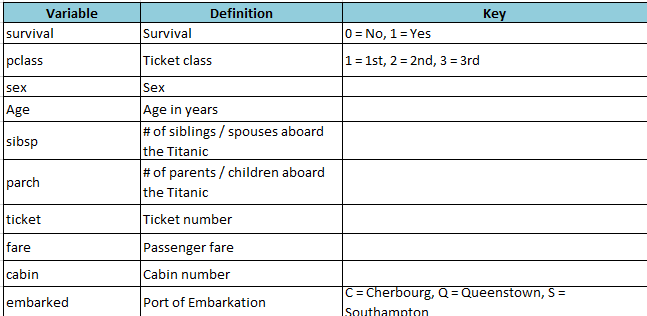
<https://www.youtube.com/watch?v=9xoqXVjBEF8>

# Data Description

The data has been split into two groups:

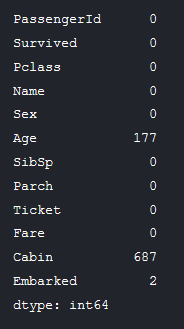
* training set (train.csv)

# Data Dictionary

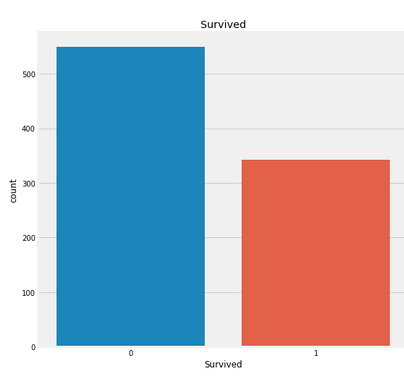


The Dataset is attached in the mail. Please read them into Python and do the following-

1. Read the training data in a dataframe & *Answer the following.*
2. Find out the number of rows and columns. Find the descriptive stats of each column. Find out the data type of each column in the dataframe
3. Do a head or tail of train dataset to look at a few rows.
4. Find out the number of missing values in train for each column. We have to fill these missing values one column at a time. The answer should look like this.



1. How many people Survived the accident. Draw a Plot to find out. The plot should look like this.



## Types of Features

Let us understand the different types of features.

### Nominal Features:

1. **Categorical Features in the dataset: Sex, Embarked, Cabin, Ticket**

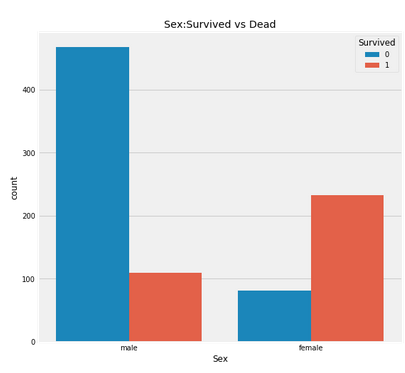
### Ordinal Features

1. **Ordinal Features in the dataset: PClass**

### Continuous Feature:

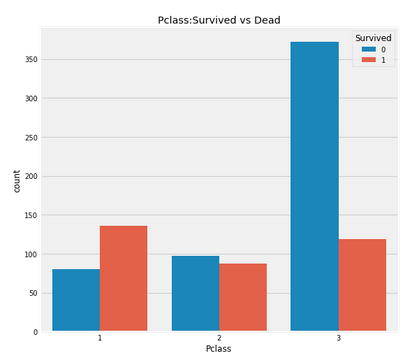
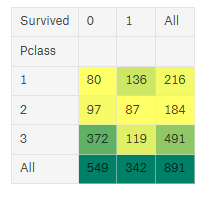
**Continuous Features in the dataset: Age, Fare etc**

1. Who has Higher chance of Survival Male or Female. Try to answer this question looking at the data. Try to get a plot like this to answer this question.

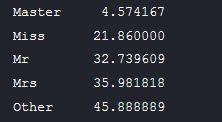


1. Do rich passengers have higher chance of Survival? Which variable will explain this.

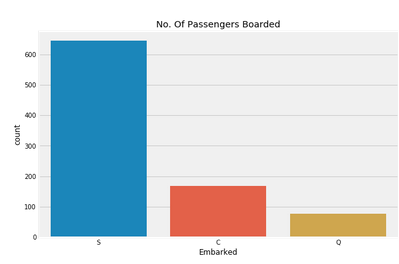
Hint- Get a plot and a table like this-



1. Find the name of the Oldest & Youngest Passenger on the Ship? What is the Average age of all passengers?
2. The Name column has all Title (example Master, Miss etc), First Name and Last Name together. Let’s separate them. Create 3 new columns One each for Title, FirstName and Last Name. Example the firstName is Braund, Mr. Owen Harris where Braund is the Last Name, Mr is the Title and Owen Harris is the first name. You have to separate them.
3. Find the Average Age with respect to each Title. The answer should look like this- Consider all non-Master, Miss, Mr & Mrs title in Others



1. Find out how many passengers embarked at which Port with a plot like this?



# Filling Missing Values

Lets Fill the Missing values in the columns-

1. Fill the Missing Ages with respect to their Title. For example If the Age is Missing for a person with Title Master fill his Age with 4.57 as seen above. Similarly for the other titles. **This is one of the ways to fill missing values for a continuous column. Please note this.**
2. Fill the Missing Values for the Embarked column. How should you fill this? Hint- Take the Mode. **This is one of the ways to fill missing values for a categorical column. Please note this.**
3. Write this as a csv file. We will use this output csv later in ML.

Please do this assignment. It took me 2 hours to prepare.

Else google to search for Codes for the questions. ALL the BEST!! FEEL FREE TO DISCUSS WITHIN YOURSELF.