Titanic Dataset

## Variable Description

### **Data Dictionary**

VariableDefinitionKey survival Survival 0 = No, 1 = Yes pclass Ticket class 1 = 1st, 2 = 2nd, 3 = 3rd sex Sex Age Age in years sibsp # of siblings / spouses aboard the Titanic parch # of parents / children aboard the Titanic ticket Ticket number fare Passenger fare cabin Cabin number embarked Port of Embarkation C = Cherbourg, Q = Queenstown, S = Southampton

### **Variable Notes**

pclass: A proxy for socio-economic status (SES)  
1st = Upper  
2nd = Middle  
3rd = Lower  
  
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 travelled only with a nanny, therefore parch=0 for them.

## Steps of Data modelling:

1. Data Cleaning
2. EDA
3. Data Wrangling
4. Data PreProcessing
5. Building Models
6. Data Visualisation

## MY Kernel

1. Import important Libraries:Pandas,Numpy,matplotlib
2. Open File as “gender”,”train\_data”,”test\_data
3. Find Categories of people in Ship by making Dictionary of their titles and columns of title names
4. Group data by PClass,Sex,Title,and fill NA values of Age by Median