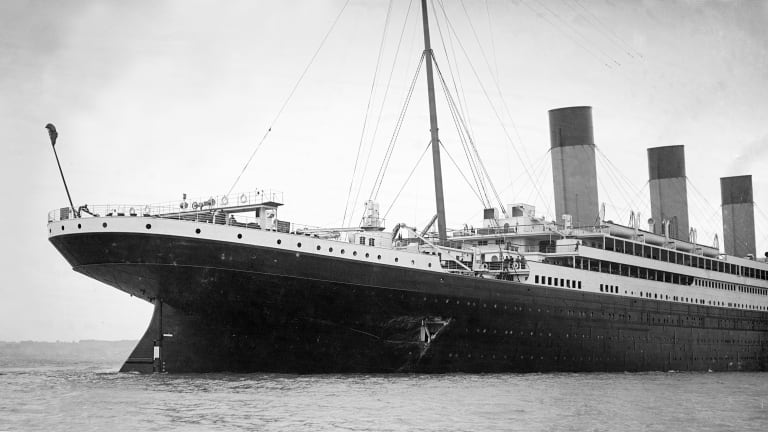
**Titanic Survival Project**

* HRITIK JUYAL



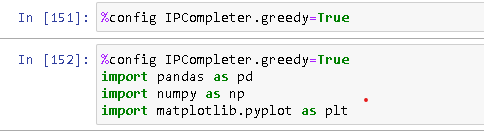
RMS Titanic was a British Passenger liner that sank in the North Atlantic Ocean in the early morning hours of 15th April 1912, after striking an iceberg during her maiden voyage from Southampton to New York City. Of the estimated 2224 passengers and crew aboard, more than 1500 died, making the sinking of the ship one of modern history’s deadliest peacetime commercial marine disasters. Even after having advanced safety features, the sinking caused such huge loss of lives due to the fact that the ship only carried 20 operable lifeboats on its voyage, which at full capacity could only save 1400 people.

This project aims at building a **Machine Learning Model** that can **predict** whether a **person** would or would not have **survived the sinking** of the RMS Titanic, depending on their age, gender, Ticket cost, Place of embarkment, among various other parameters.

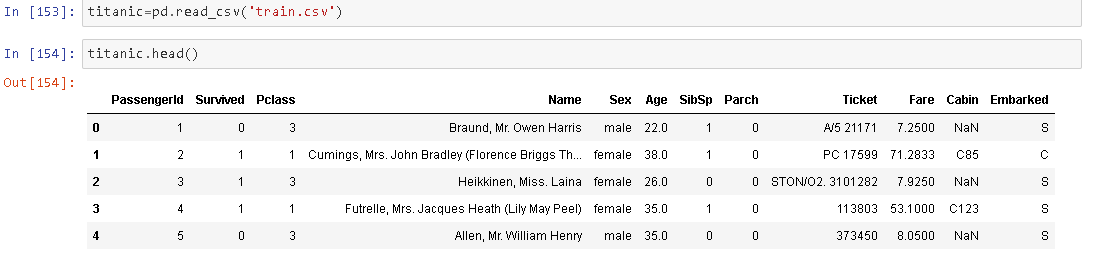
The Dataset used here has the details of 890 passengers onboard the Titanic who either survived or unfortunately did not, along with their details. We will first study the data and convert it into a usable form through data wrangling, by removing the details we don’t need or that would not affect the learning of the model, then we will impute missing values, scale all of them and then encode the character type values to finally use it to visualise the data and use it after selecting and applying an apt model.

**We have a data set of different information about passengers onboard the Titanic, and we see if we can use that information to predict whether those people survived or not**

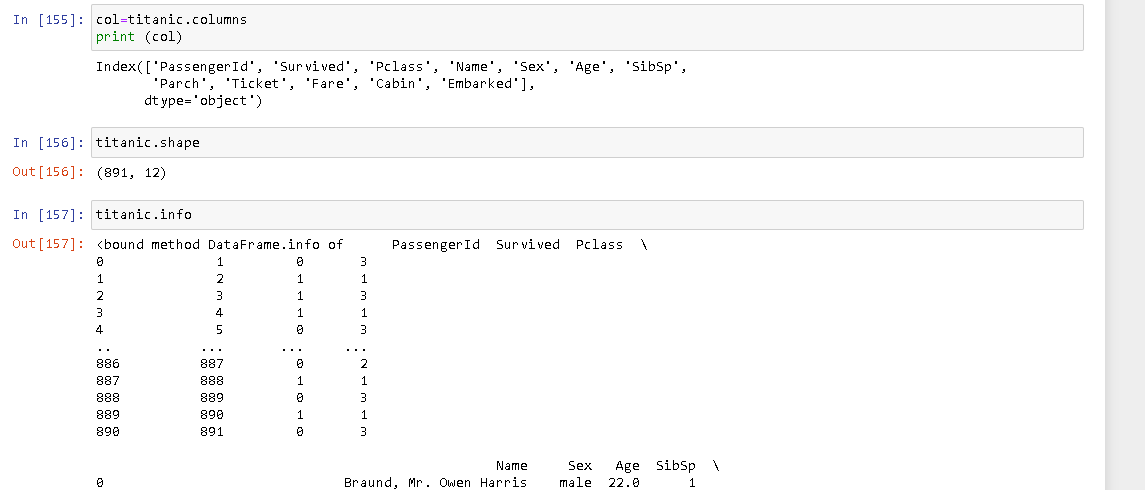
**Import important libraries**



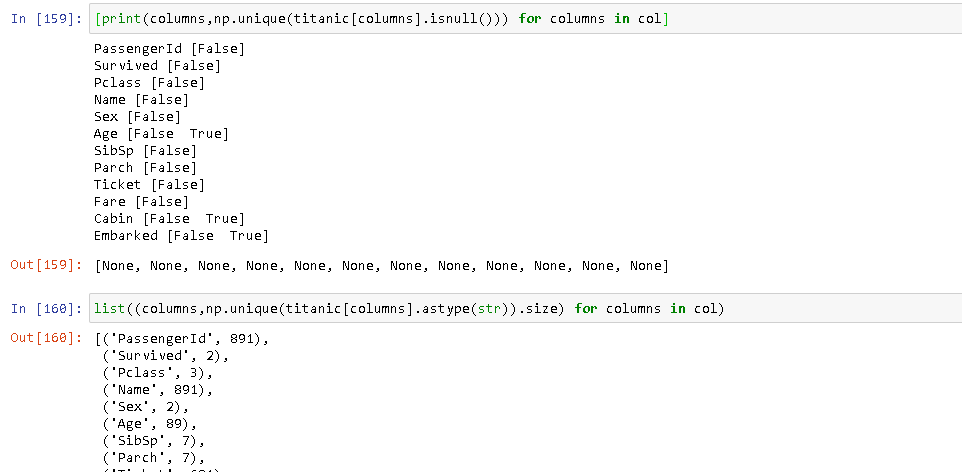
**load titanic data**



**What are all the columns present**



**Which columns have none of nan values**

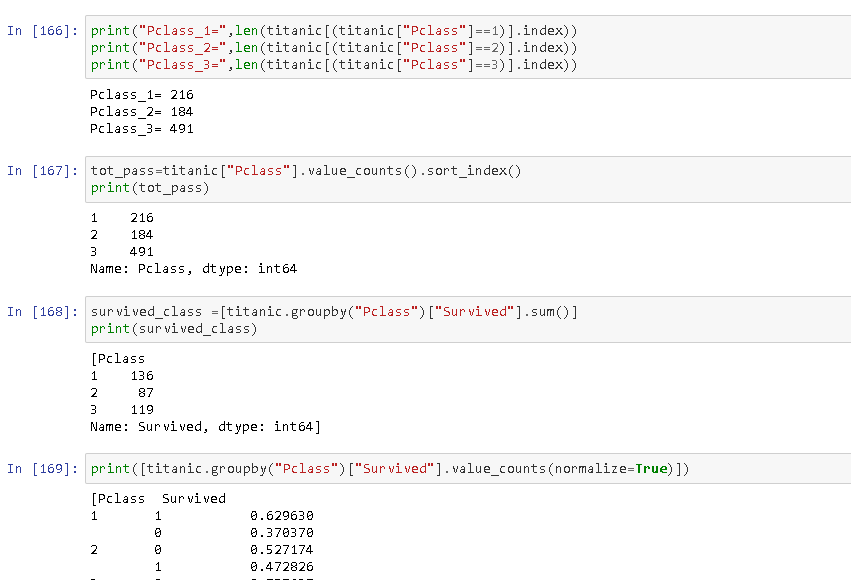


**Survived column contains 0 if the passenger did not survive and 1if they did,**

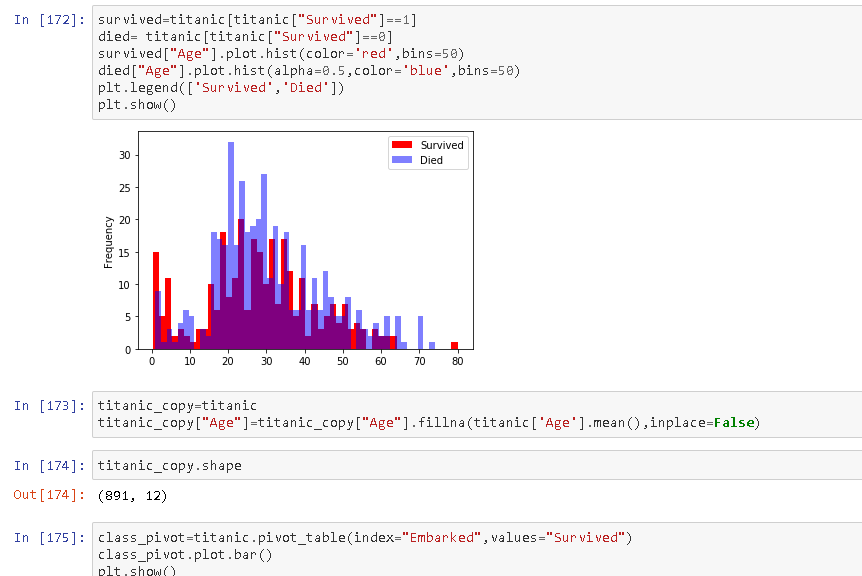
**Survival based on gender of passenger**



**Survival based on passenger class**



**We are interested in relation of other features with whether the passenger survived and we can see that passenger class PClass is an important parameter for determining if the passenger survived fare is also giving high value but fare is in relation with Pclass.this does not give the complete picture as the survived class is not a continuous variable but still we can get the idea**



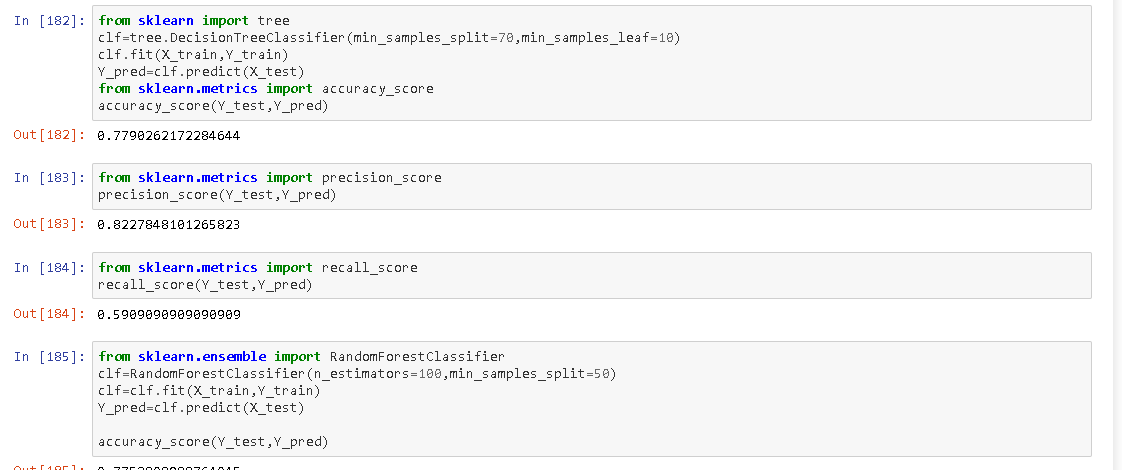
**Dropping rows where embarked is NAN**



**Model Selection**

**The input and output Datasets x and y were split into two parts, i.e. test and train. Train data was used to train each model and test data was used to determine accuracy of the predictions made by each model on the test data.**

**Different models were made and compared, namely LogisticRegression, RandomForestClassifier , Support Vector Machine, etc. Each model was separately trained and tested on the same data and each model’s confusion matrix and accuracy score were compared to determine which classifier worked best for the given dataset.**





**On comparison, we can see that SVM classifier gives the best result with precision (86.66)**