***Titanic Survival Project***

* **Ashish Sharma**



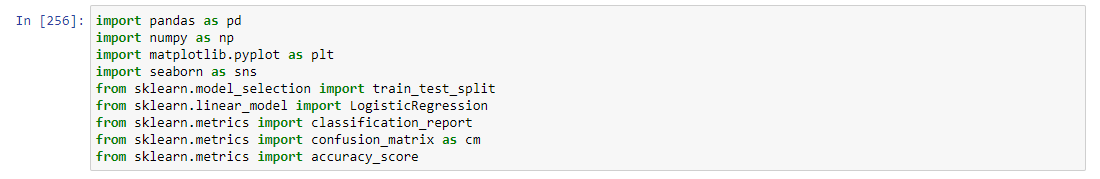
On 10 April 1912, when the largest ocean liner in service at that time, RMS Titanic, started its maiden voyage from Southampton, it was a new beginning in the history. With at least 2,224 people on board, including a number of prominent personalities, the luxury ship that nicknamed as ‘Millionaire’s Special’ was en route to New York City on that day.

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.

Over the course of the next few hours, people witnessed the biggest catastrophe to hit the world, and at 2.20 AM on the morning of the 15th, the biggest ship that the world had seen settled to the bottom of the North Atlantic.

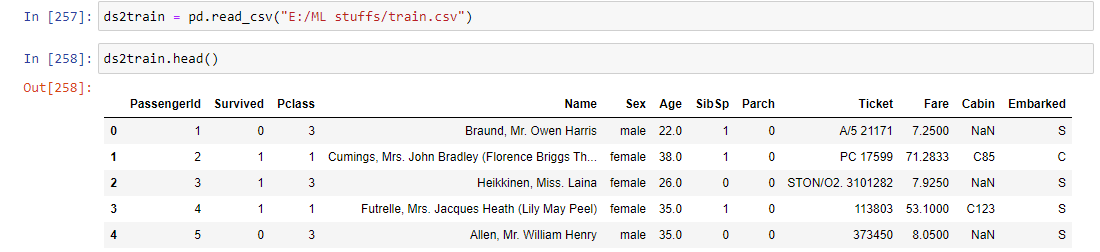
The sea, the sinking of the ship and the cold weather were enough to take the lives of 1517 passengers and crew on board. Just 706 of them survived the ordeal to tell the horrific and sad tale to the world.

* Importing the Libraries



All the libraries and their modules seen here are used in the project.

* **Loading Dataset**



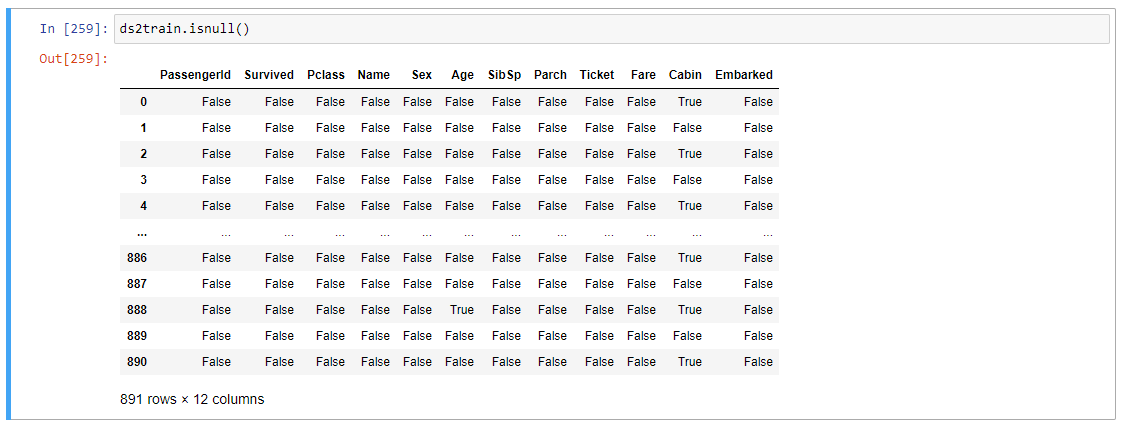
891 rows × 12 columns

Here, as mentioned, the ds2train dataset has 12 columns. It has 891 rows, the dimensions showed in the image are for the 5-row head of the dataframe. The headings of the columns are also mentioned.

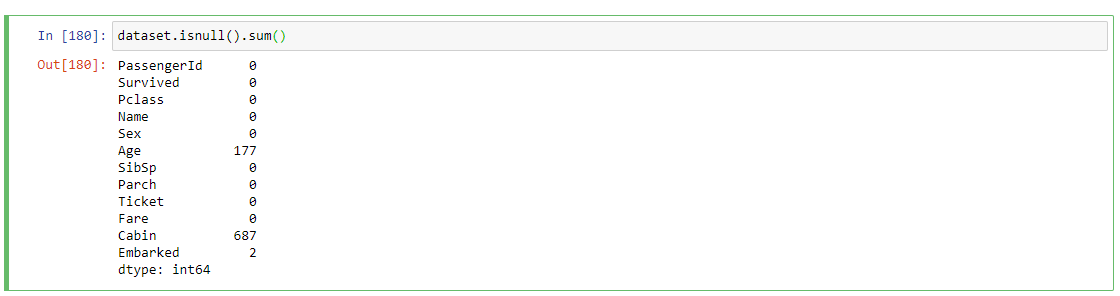
* **Data Wrangling**

Data wrangling is a process to clean the Nan value's data and unnecessary columns from the dataset.

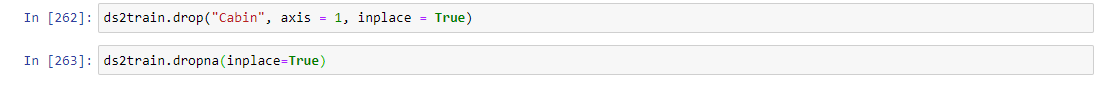
In the table shown below, False shows the data is not null and True means data is null



In the next figure, it is showing the no. of null values in each column



In next figure we will get to see the cabin column has lots of null values , so that it might be true to remove the cabin column. And after that in next cell we will remove all the Null values

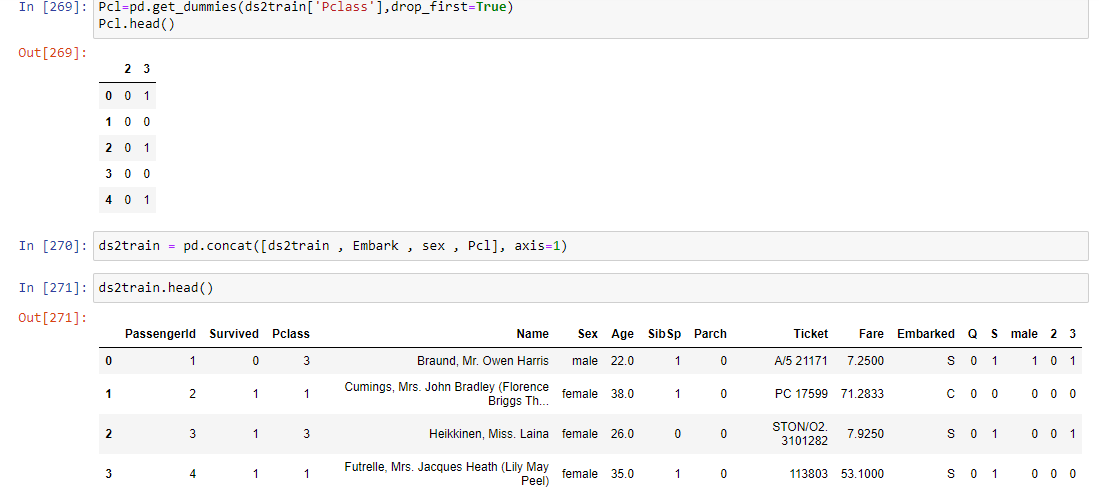


* **Imputing Missing values**

Now we will check for the Null values (if any).

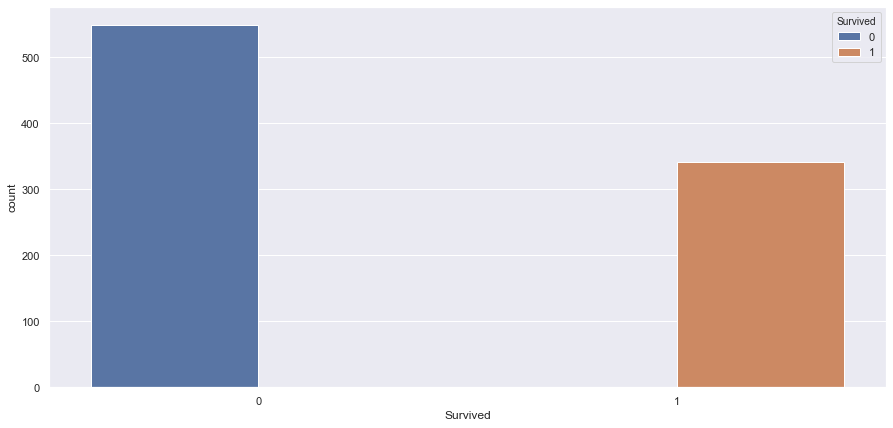
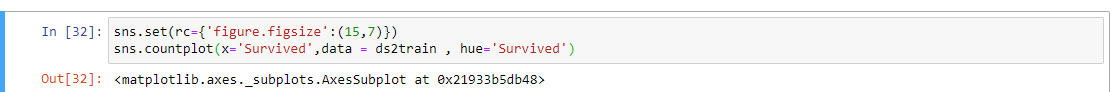


As we see we have many categorical data in our dataset , but this data is not acceptible in the analysis , so we have to convert those data in the numerical data.And we can do it using pandas

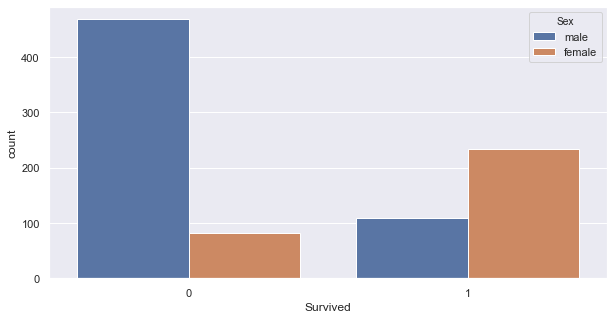
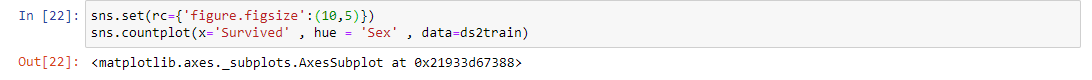


* **Visualisation**

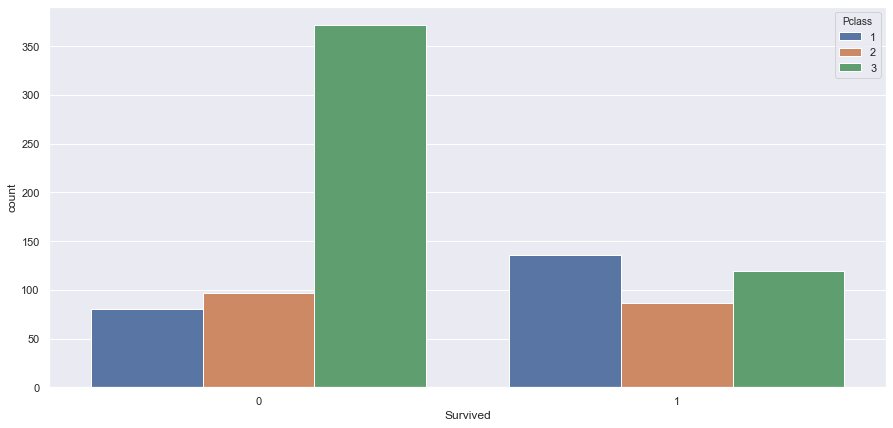
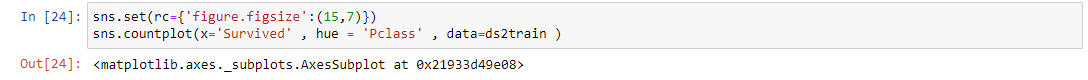
Inputs and Outputs :-



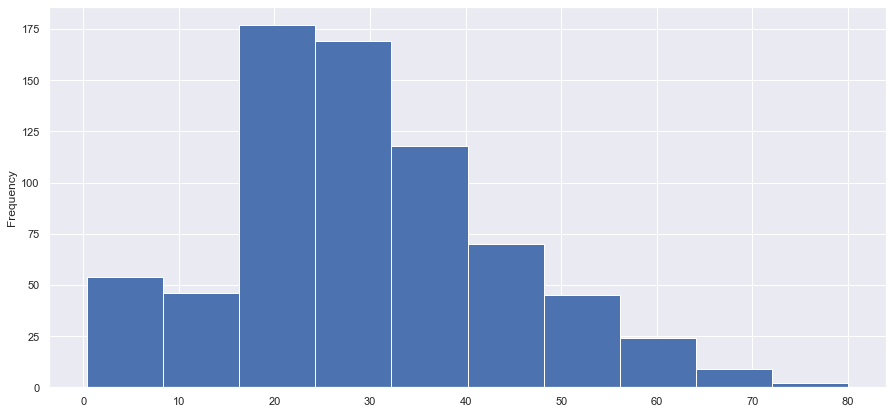
This graph shows the data of people who were survived or not. So blue color shows not survived , and Orange color shows survived.



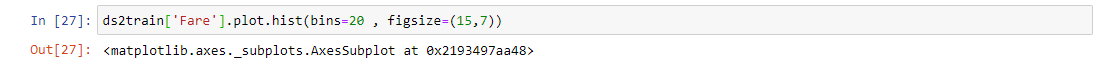
This graph shows the survived or unsurvived people data on the basis of their gender. Majority of male were not survived and majority of female survived.

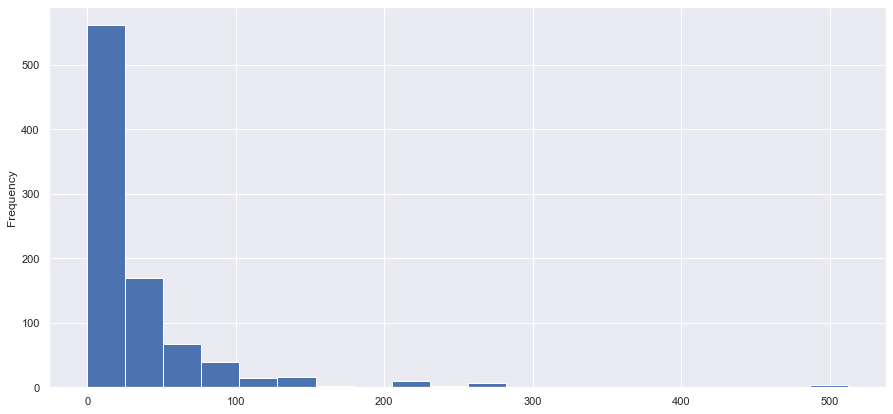


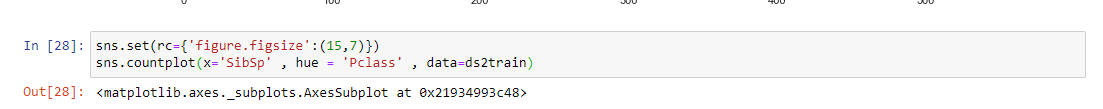
This grapsh shows the class of passengers in which they travelled. According to dataset it shows the majority of unsurvived people were belong from the third class.

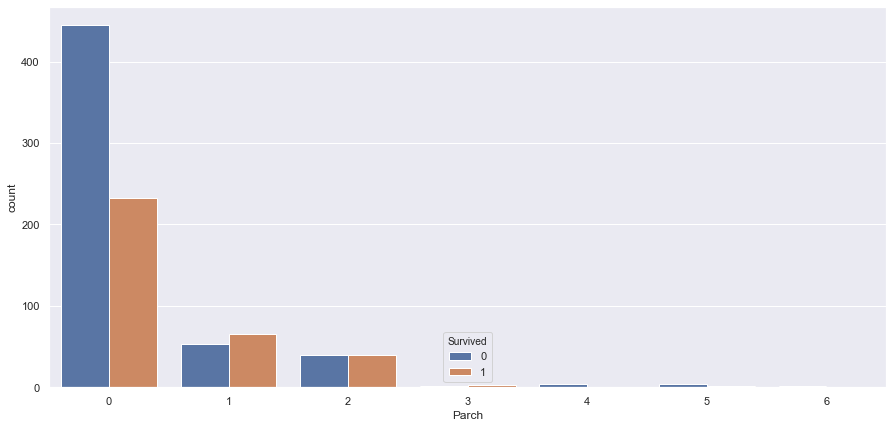
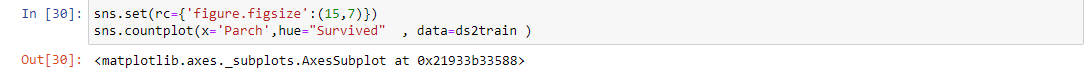
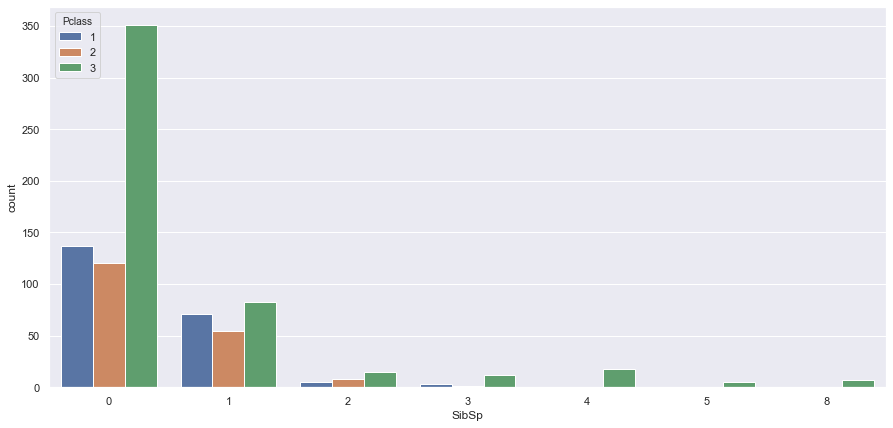


This histogram shows the age level of passengers. According to the datsets histogram shows the mostly passengers were above 20 age.

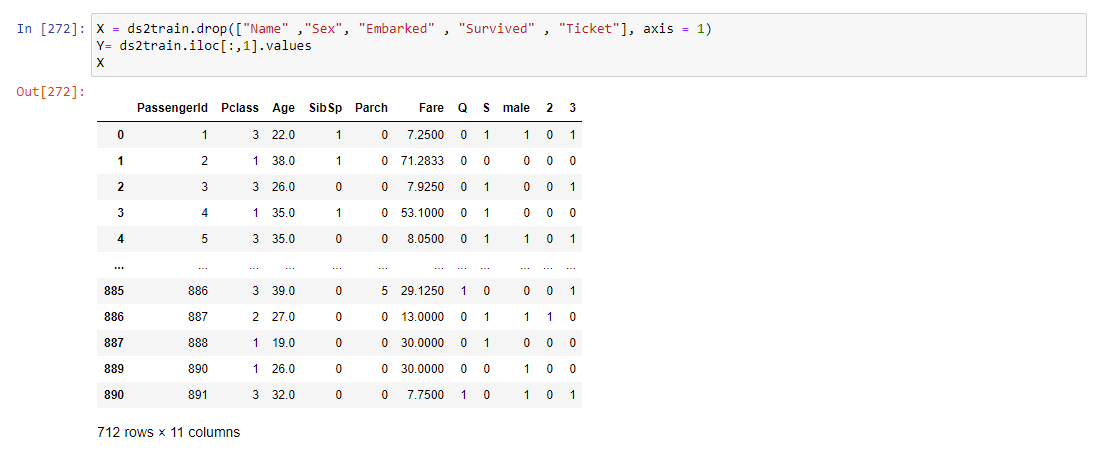




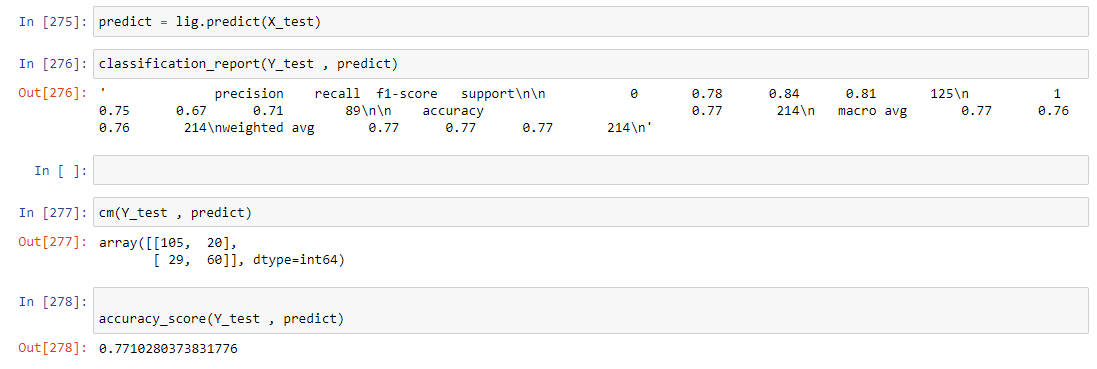
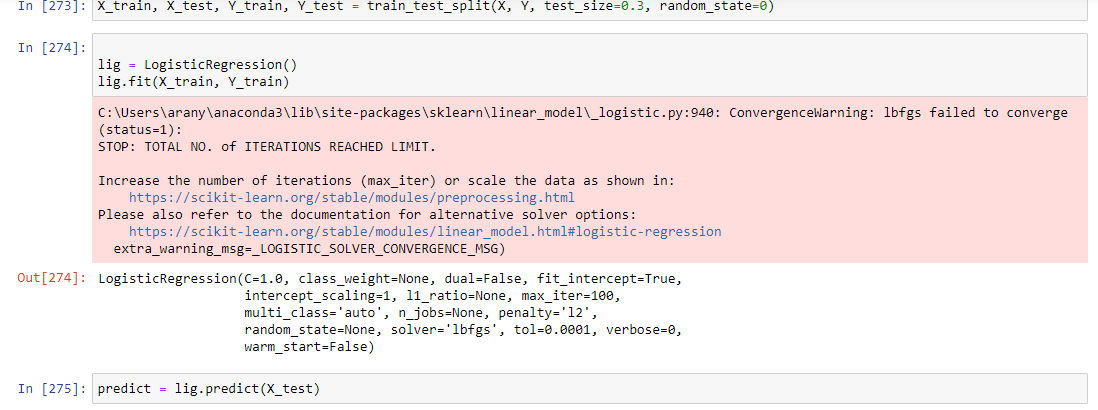




* **Training Model**



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.



On comparison, we can see that the best result which gives us an accuracy of **77.10%.**