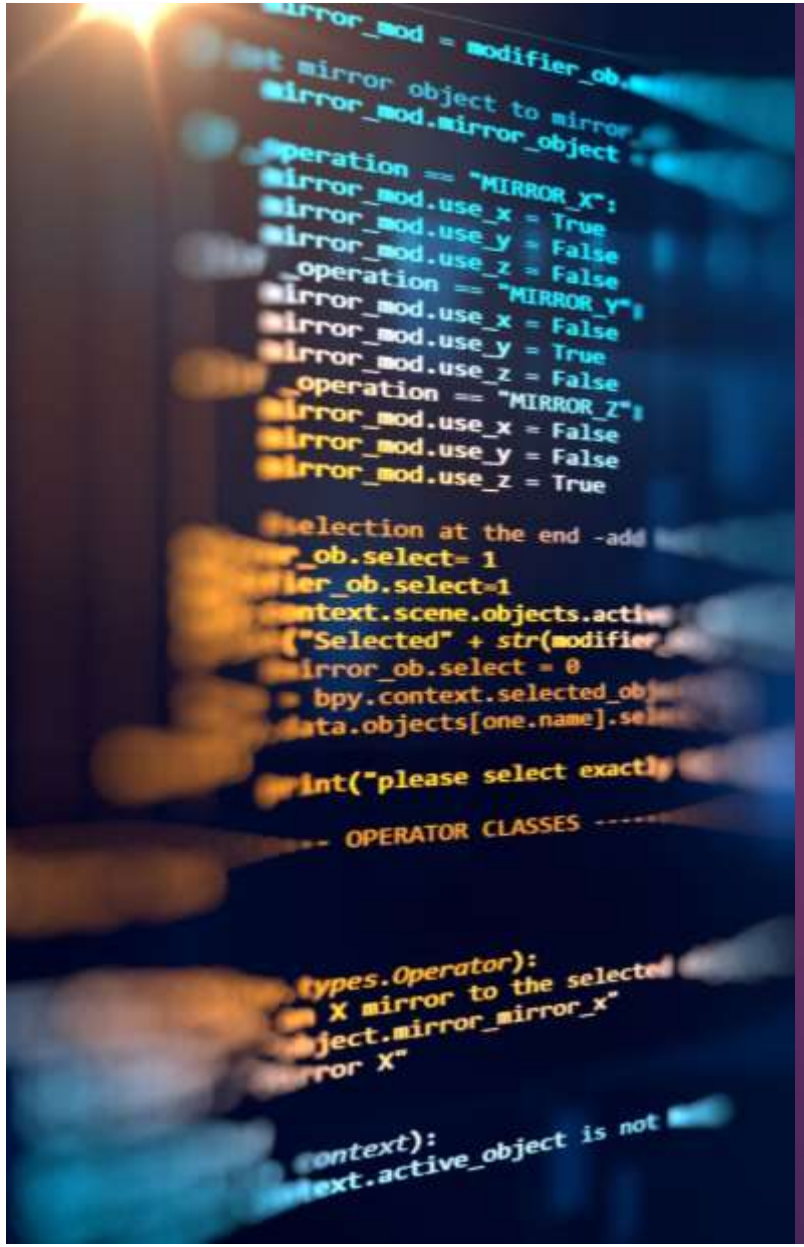
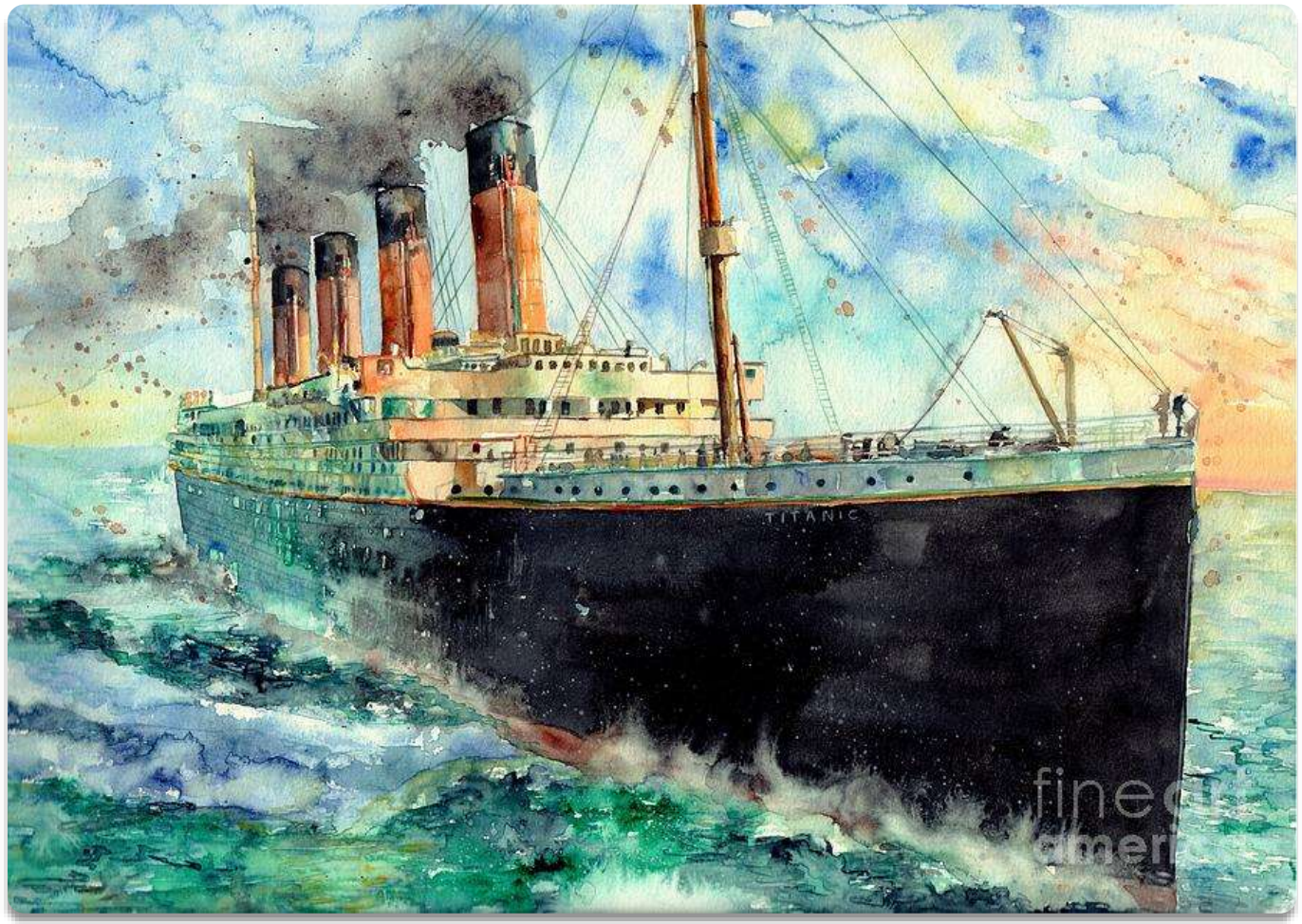
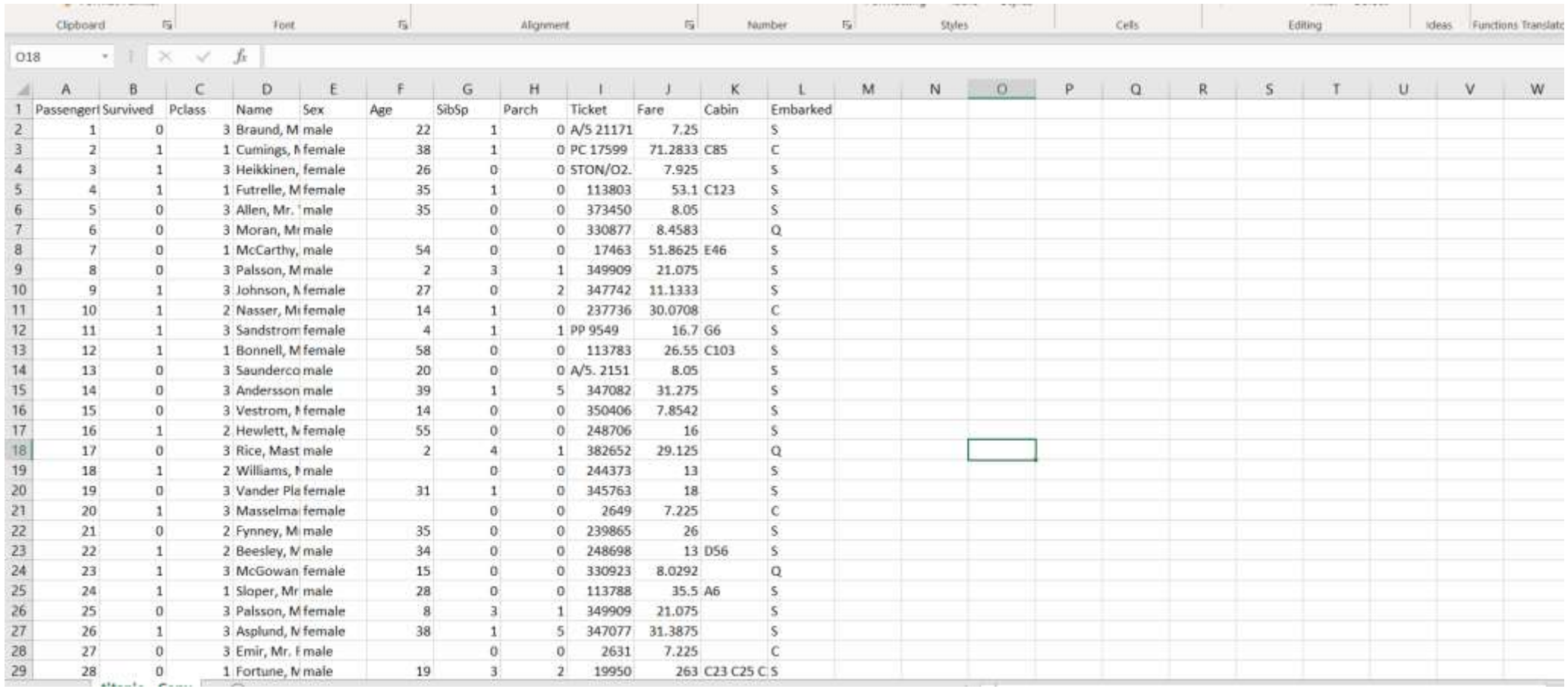


DS 540(STASTICAL PROGRAMMING) FINAL PROJECT

PREDICTING SURVIVAL OF TITANIC PASSENGERS (EDA AND MACHINE LEARNING)



TITANIC SHIP



Clipboard | Font | Alignment | Number | Styles | Cells | Editing | Ideas | Functions Translate

O18

	A	B	C	D	E	F	G	H	I	J	K	L
1	Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings, M	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen,	female	26	0	0	STON/O2.	7.925		S
5	4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mr	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy,	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, M	male	2	3	1	349909	21.075		S
10	9	1	3	Johnson, M	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, Mr	female	14	1	0	237736	30.0708		C
12	11	1	3	Sandstrom	female	4	1	1	PP 9549	16.7	G6	S
13	12	1	1	Bonnell, M	female	58	0	0	113783	26.55	C103	S
14	13	0	3	Saunderco	male	20	0	0	A/5. 2151	8.05		S
15	14	0	3	Andersson	male	39	1	5	347082	31.275		S
16	15	0	3	Vestrom, M	female	14	0	0	350406	7.8542		S
17	16	1	2	Hewlett, M	female	55	0	0	248706	16		S
18	17	0	3	Rice, Mast	male	2	4	1	382652	29.125		Q
19	18	1	2	Williams, M	male		0	0	244373	13		S
20	19	0	3	Vander Pla	female	31	1	0	345763	18		S
21	20	1	3	Masselma	female		0	0	2649	7.225		C
22	21	0	2	Fynney, M	male	35	0	0	239865	26		S
23	22	1	2	Beesley, M	male	34	0	0	248698	13	D56	S
24	23	1	3	McGowan	female	15	0	0	330923	8.0292		Q
25	24	1	1	Sloper, Mr	male	28	0	0	113788	35.5	A6	S
26	25	0	3	Palsson, M	female	8	3	1	349909	21.075		S
27	26	1	3	Asplund, M	female	38	1	5	347077	31.3875		S
28	27	0	3	Emir, Mr. F	male		0	0	2631	7.225		C
29	28	0	1	Fortune, M	male	19	3	2	19950	263	C23 C25 C	S

Details of Data set

- Comprises of 891 rows and 8 columns
- Consists of Categorical and Numerical Data
- Consists of 2 missing values

AIM OF PROJECT

To find the basic statistics of passengers travelled in the ship.

To predict the survival rate of passengers in the ship.

```
import pandas as pd  # Import pandas
import numpy as np   # Import numpy
import seaborn as sns # Import seaborn
import plotly.express as px  # Import plotly express for Interactive Chart
import matplotlib.pyplot as plt  # Import matplotlib
from sklearn.metrics import r2_score # r2score
from sklearn.linear_model import LinearRegression # linear regression model
from matplotlib.animation import FuncAnimation #Import Animation Function
from sklearn.model_selection import train_test_split # train test split
from sklearn.tree import DecisionTreeRegressor  # Import Decision Tree Regr
from sklearn.ensemble import RandomForestRegressor # random forest model
from math import sqrt  # For squareroot operation
from sklearn.cluster import KMeans # kmeans clustering
from plotly.offline import iplot, init_notebook_mode # Standard plotly impo
init_notebook_mode()
init_notebook_mode(connected=True)
import cufflinks as cf
cf.go_offline()

%matplotlib inline
```

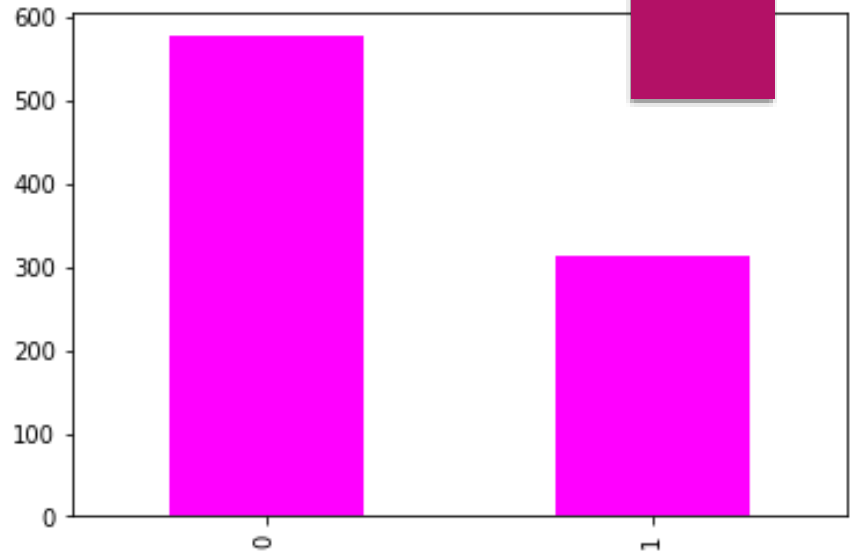
LIBRARIES USED

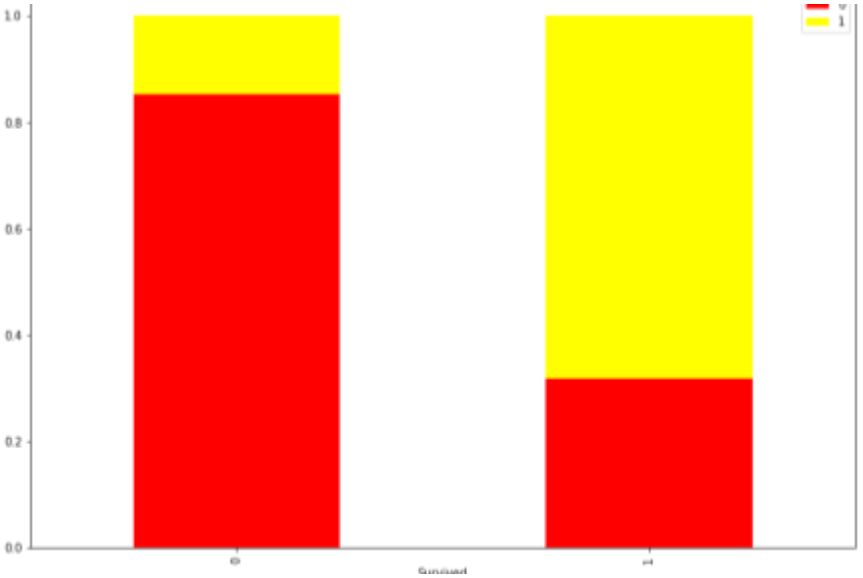
GRAPHS USED FOR FINDINGS

Violin Plot	Bar Plot	Histogram	Count Plot	Heat Map
Cat Plot	Swarm Plot	Interactive 3D Plot	Dist plot	Scatter plot
	Pie Chart	KMeans Clustering	Logistic Regression	

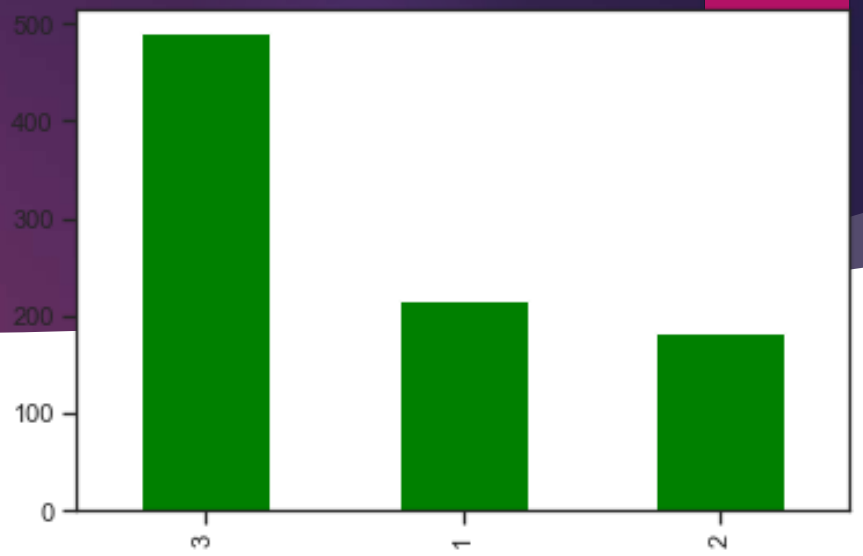
Find out total no of male/female passengers

- The first fig above shows that males are more (577) than females (314) travelling in the ship
- The sec fig above shows that females(1) survival ratio is high than males (0).Yellow color is survived and red is represented by not survived.



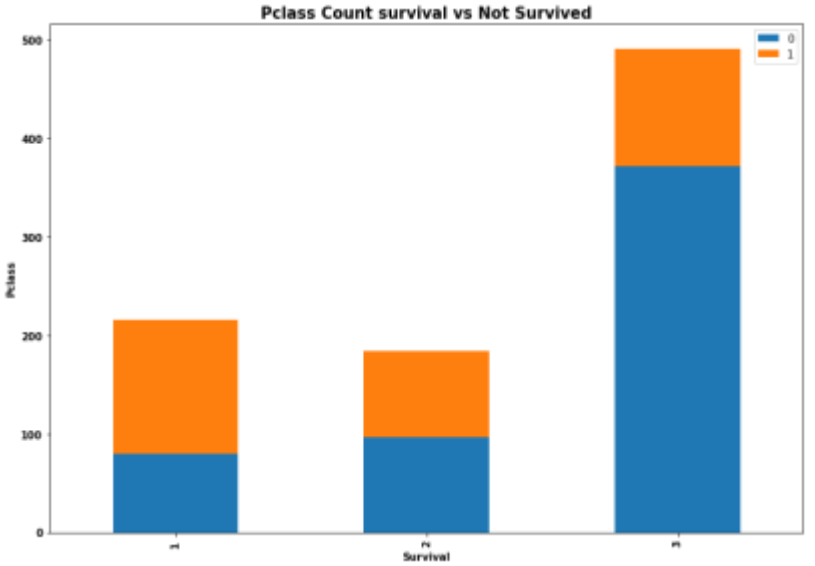


Find total number of passengers in each passenger class

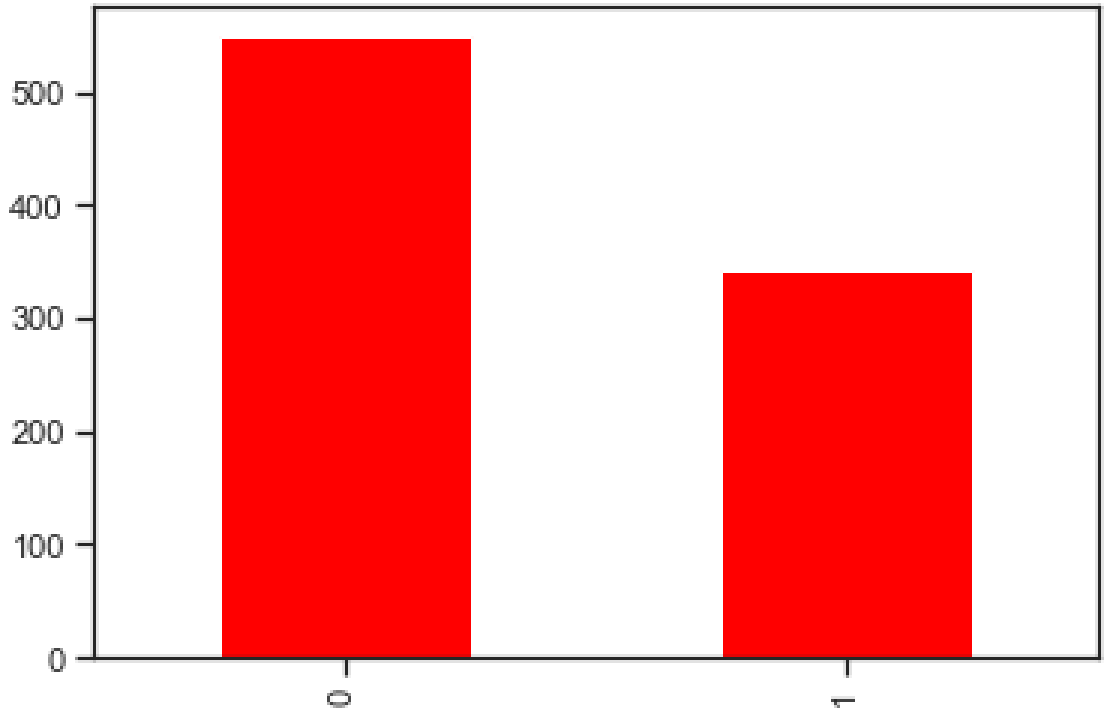


The first fig above shows that more number pf passengers (491) are travelling in 3rd class,216 are travelling in 1st calss and least number(184) passengers are travelling in 2nd class

The second fig above shows that survival rate is high in Pclass 1 and survival rate is almost equal in Pclass2 and Pclass1 Orange color is survived and blue color represented by not survived

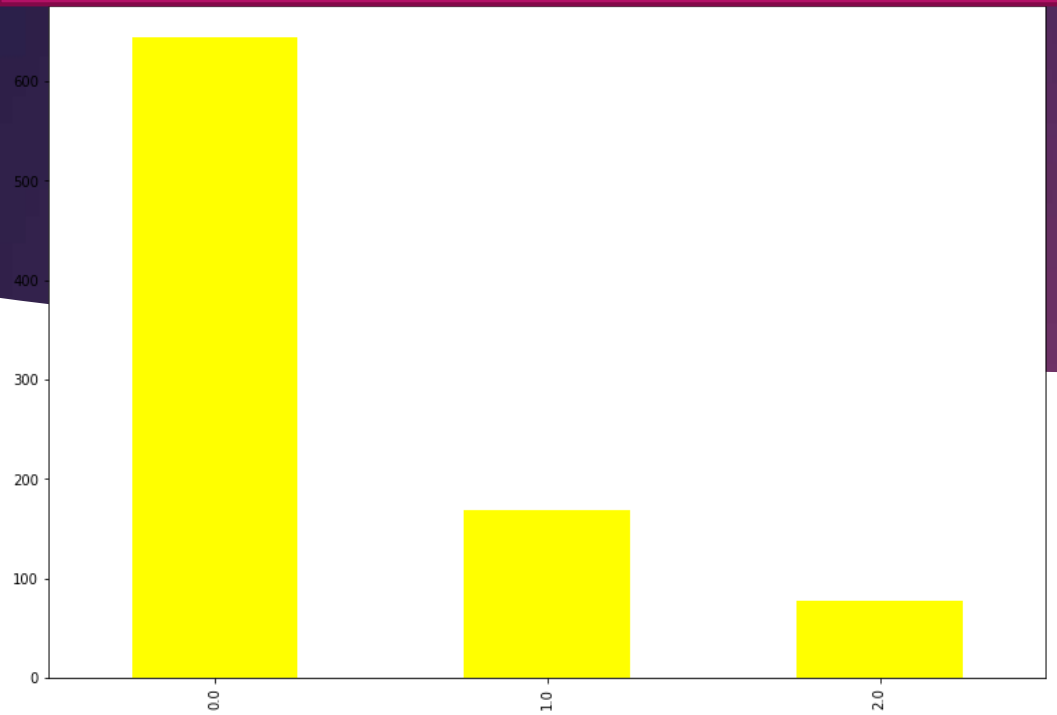


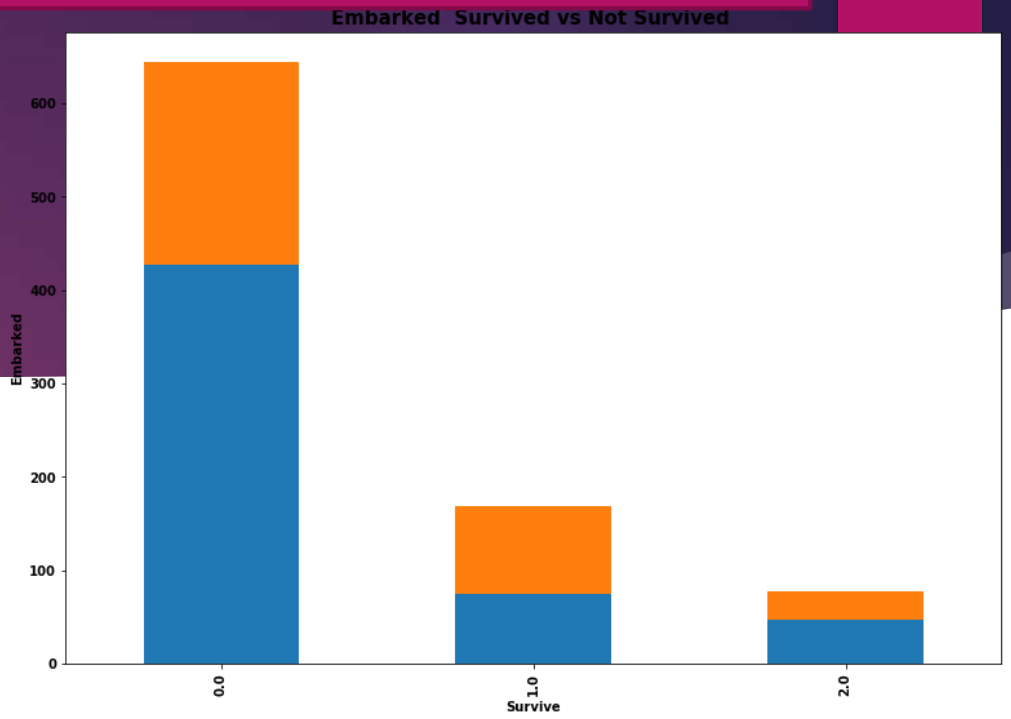
Find out total number of survived/not survived passengers



The fig above shows that survived ratio is less (342) when compared to not survived (549)

Find total number of passengers by point of Embarkation



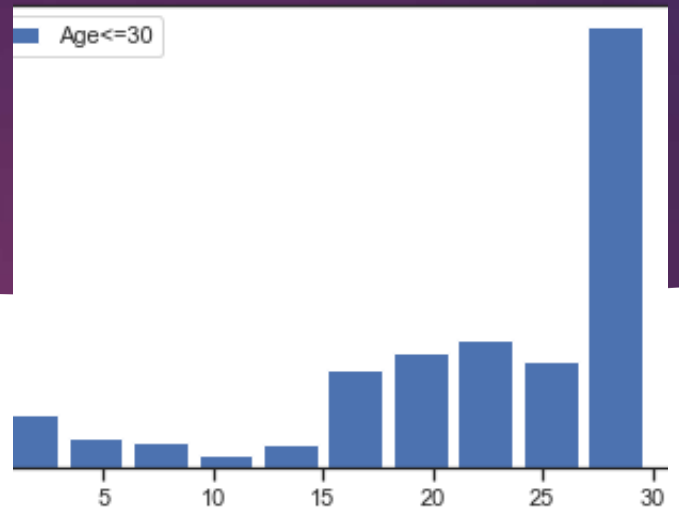


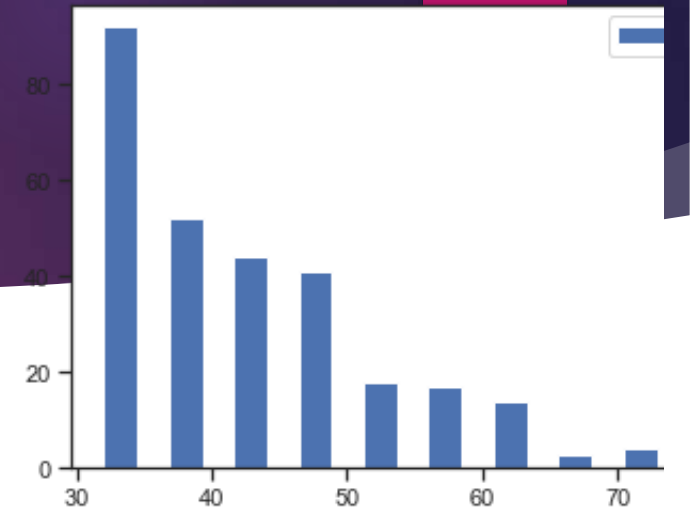
The first fig above shows that More number of passengers are from southhampston(644),the count of passengers from cherbourg are 168 and least number of passengers are from Queenstown(77)
The sec fig above shows that survived ratio is more in passengers from southhampston,next is cherbourg and least is queenstown.

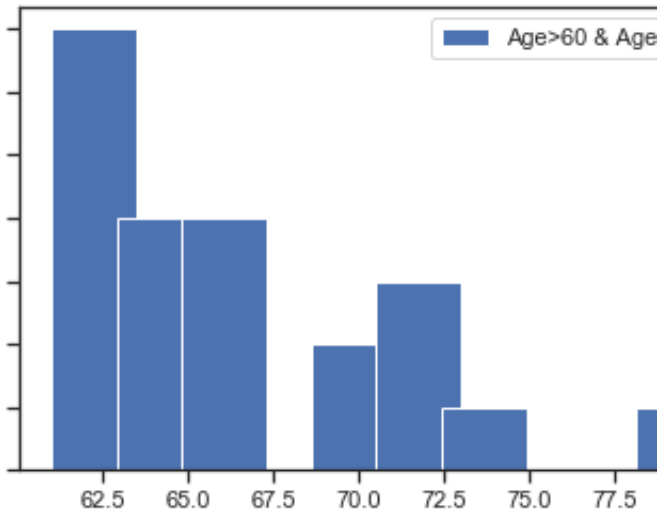
Histogram of Number of passengers in each age group

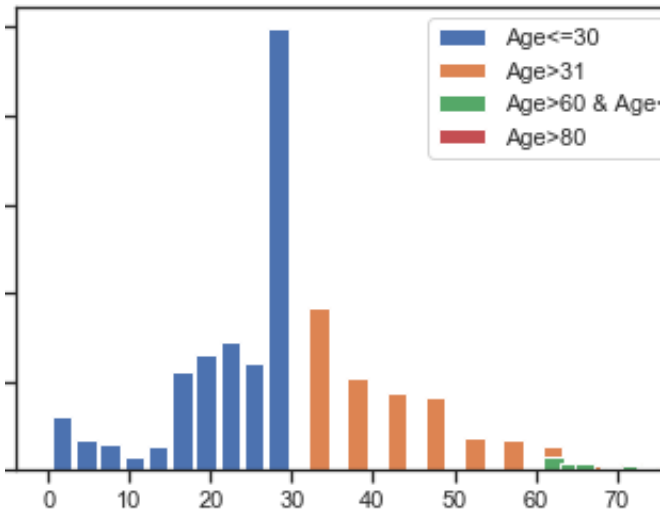
The above figs shows that more number of passengers are in the age group 25-30 and least number of passenger in the age gropup from 70 and above

Infants below 1 yr are the youngest passengers and passengers above 70 are the oldest passengers travelling in the ship



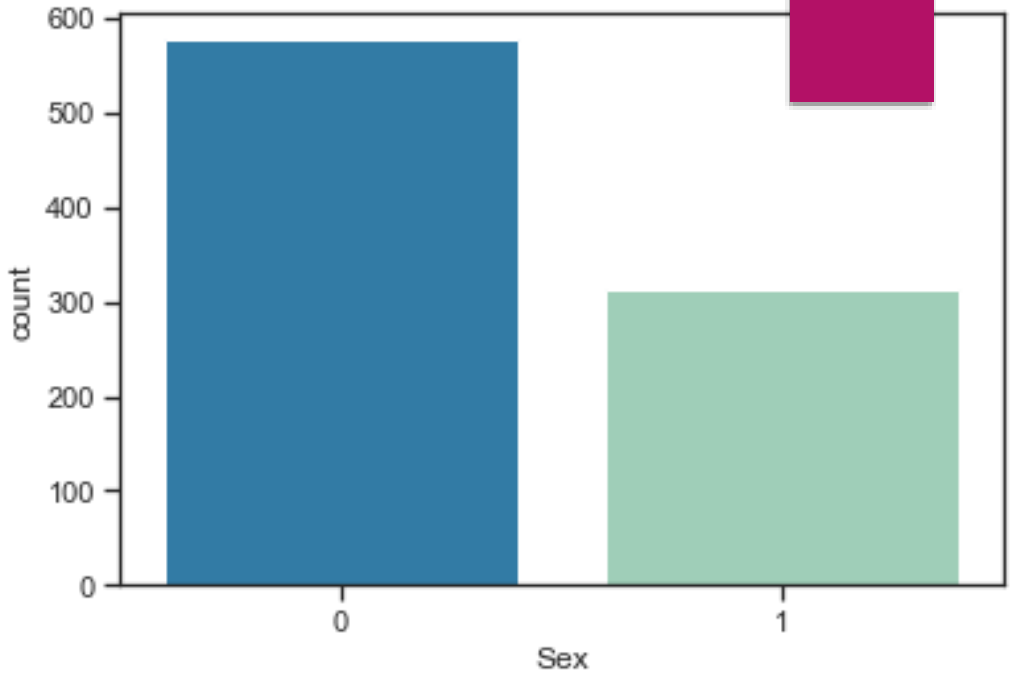


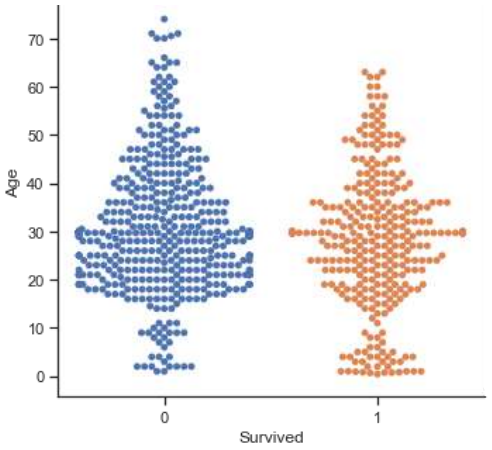


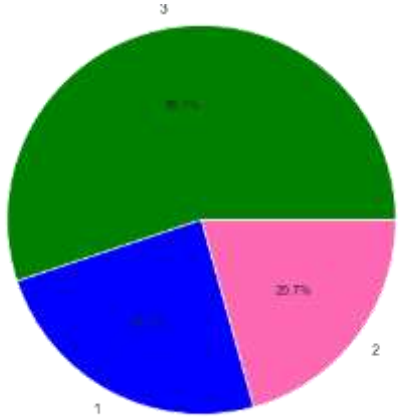


Histogram,Pie Chart and swarm plot

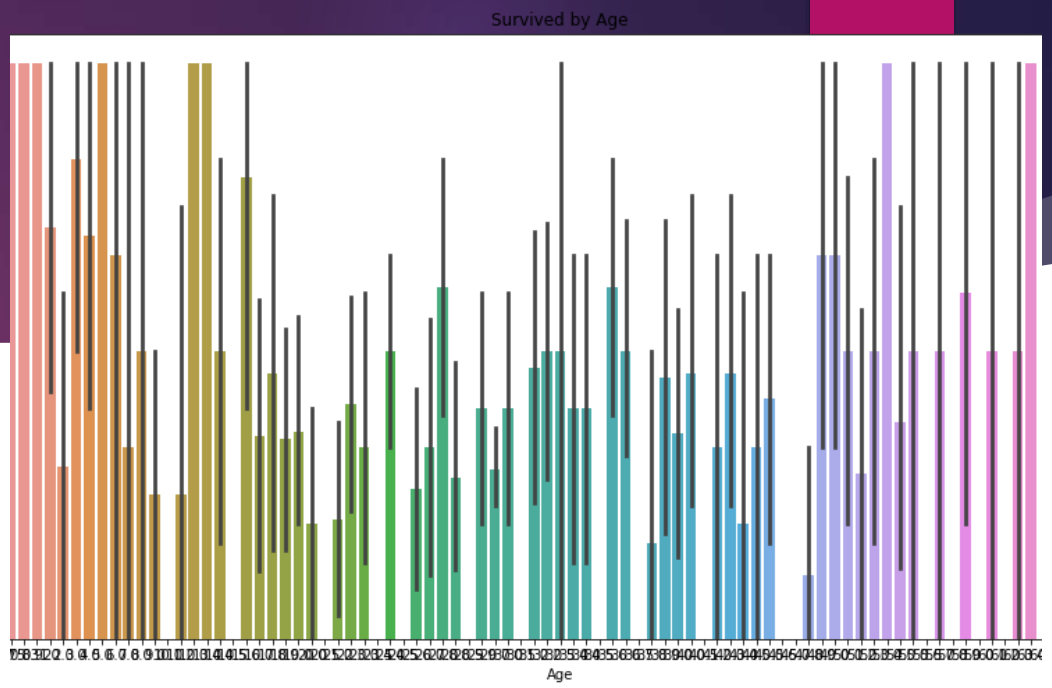
- The first fig above shows that males are more than females
- The sec pie chart shows that 55% are travelling in 3rd class,24% are travelling in 1st class and 20% are travelling in 2nd class
- The third swarm plot above shows that in both male and females survived passengers are more in the age group between 20-30



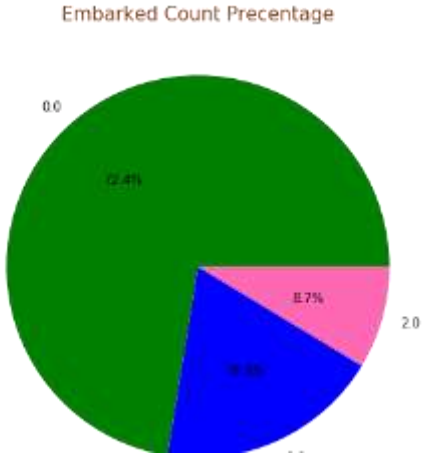


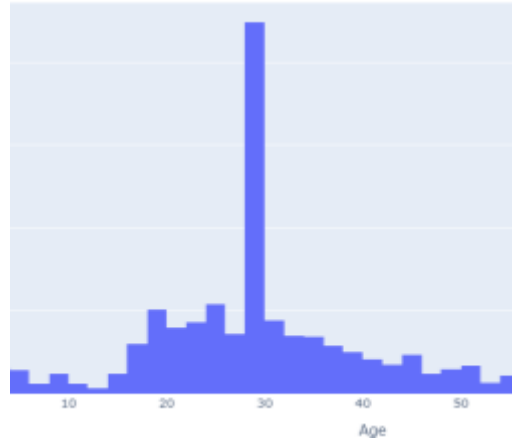


Interactive Histogram Pie chart and bar chart,countplot

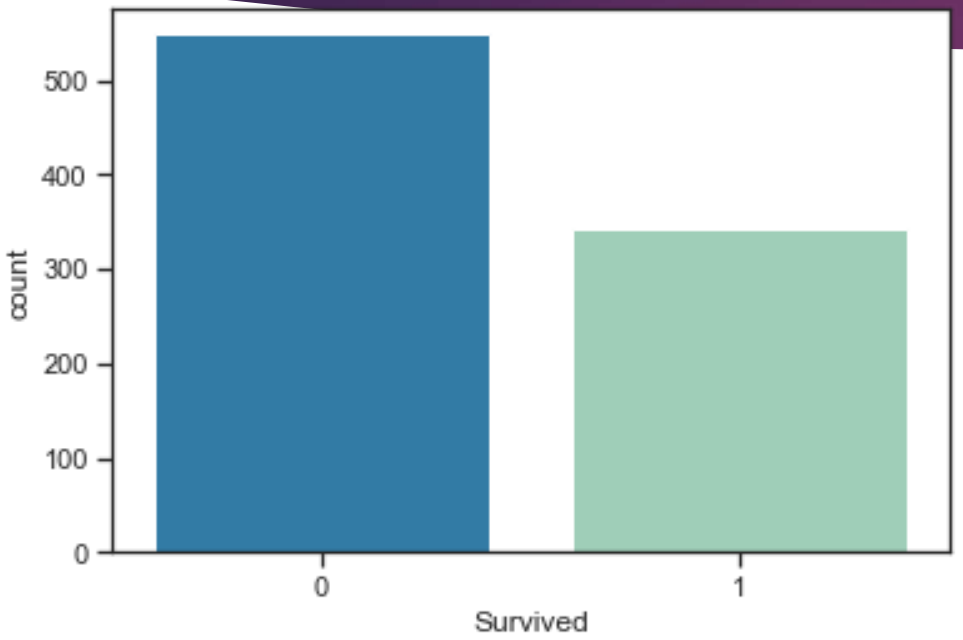


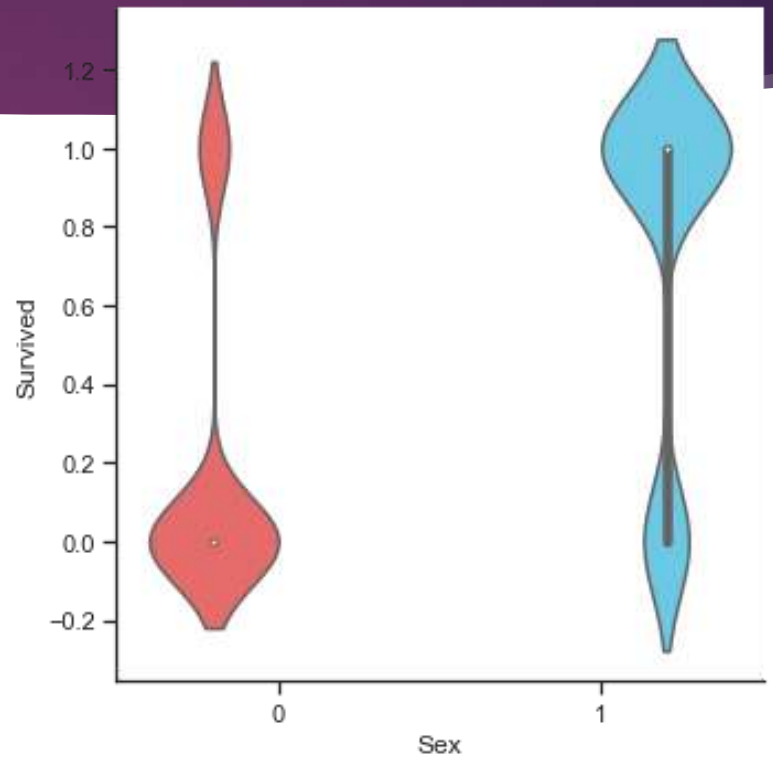
- The first fig above shows that count of passengers between 28-29 is 224,1 passenger is travelling in age of 80,5 passengers travelling in the age of 70-71 14 are travelling in age group of 0-2 yrs old.
- The sec pie chart shows that 72% passengers point of embarkation is southhampston,cherbourg is 18% and 8.7% are from queenstown
- The third fig above shows survival rate in each age group





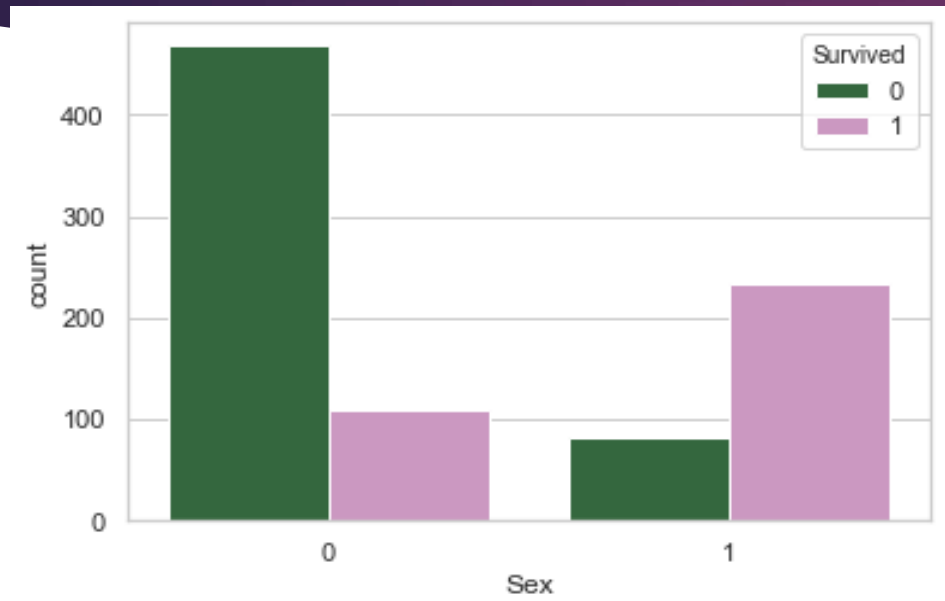
Histogram & Violin plot

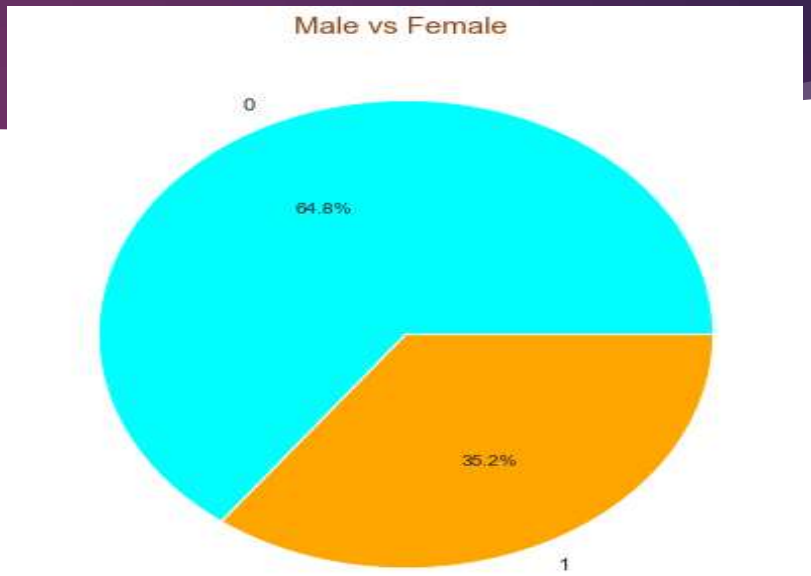




The first fig above shows that not survived (549) are higher than the survived (342)
The sec violin plot above shows that survided ratio is high in females than males

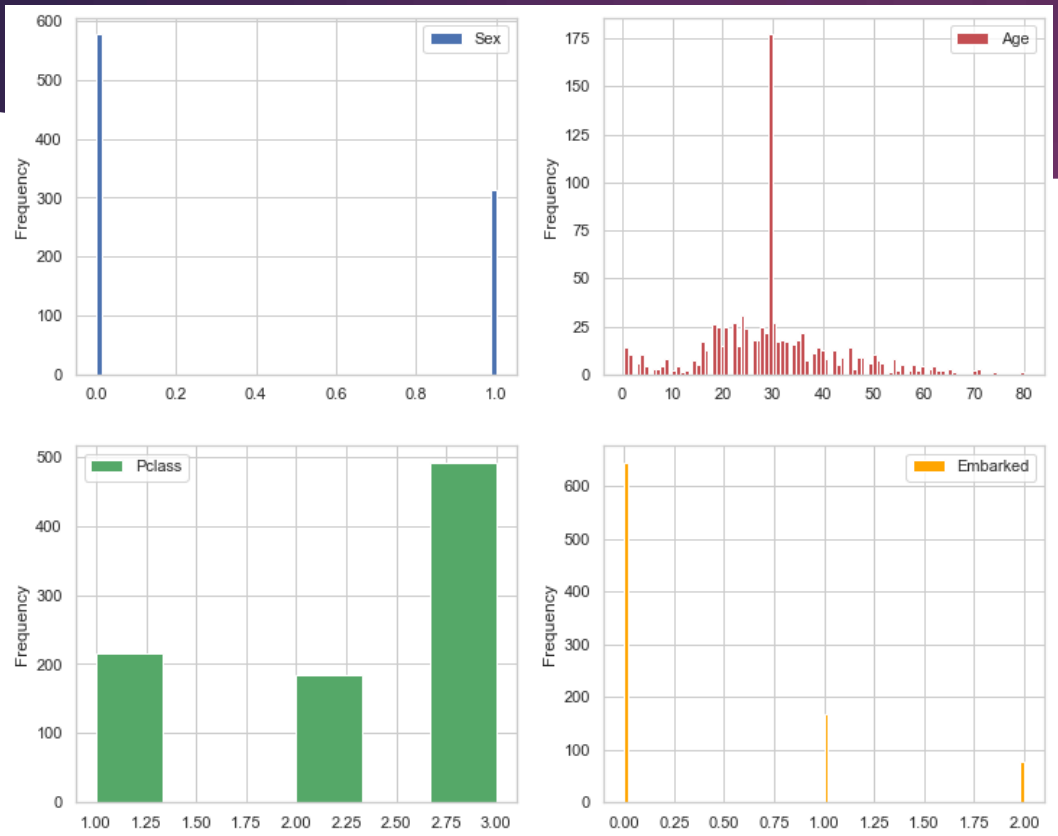
Count plot & Pie chart





The first fig shows that survived ratio is high in females and not survived ratio is high in males
The sec fig shows that 64.8% are males and 36.2% are females travelling in the ship

Histogram of count of passengers in each variable

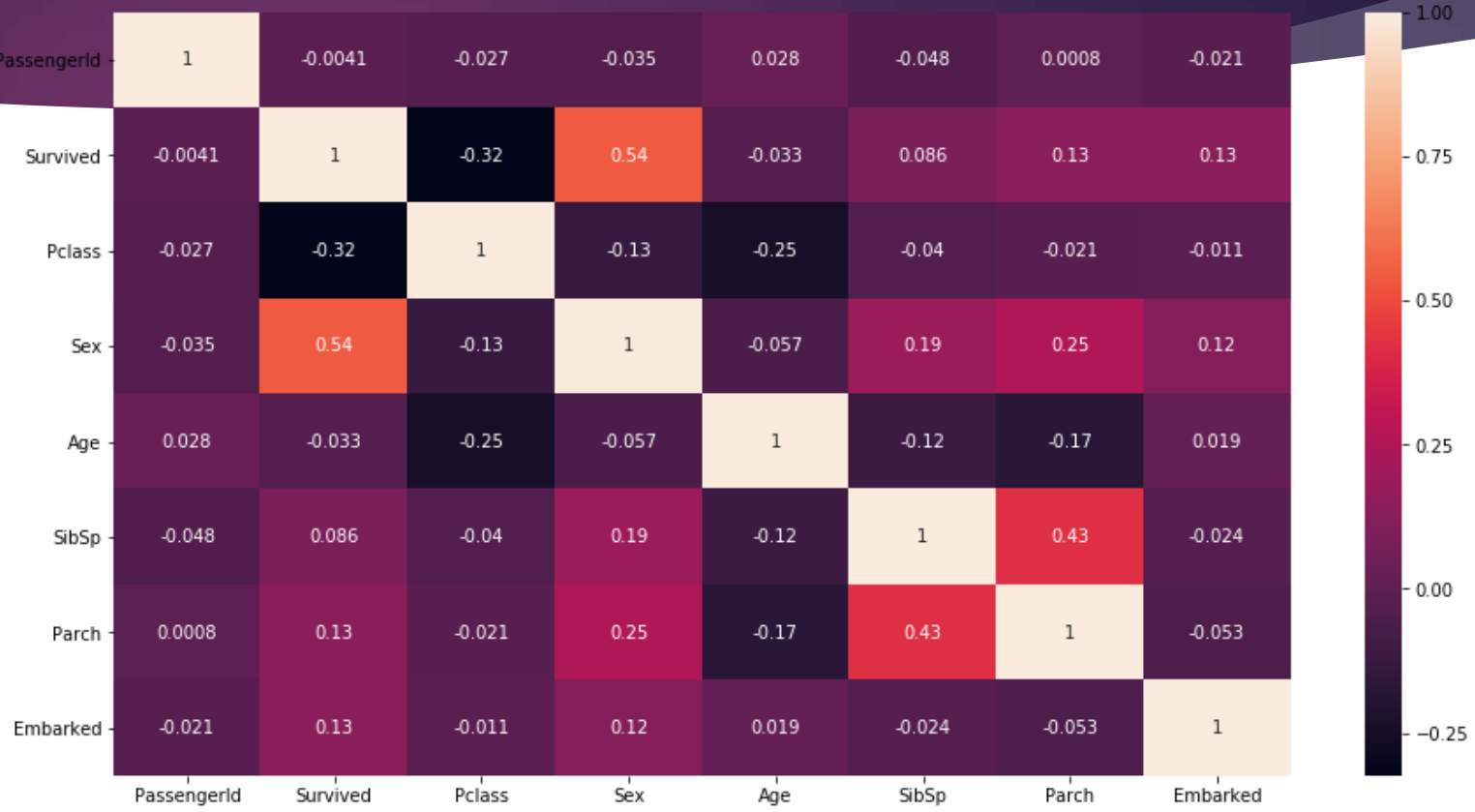


- The above histograms shows the count in each category.
- Males are more than females
- Passengers are more in between age group 25-25
- Passengers count is more in Pclass 3
- Passengers count is more from Southhampston

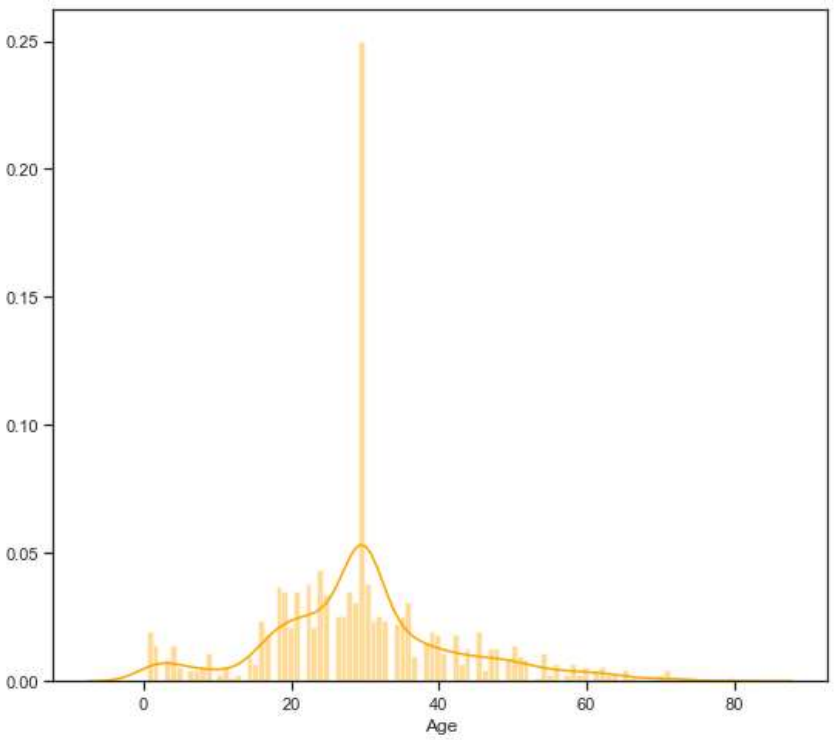
Heat Map

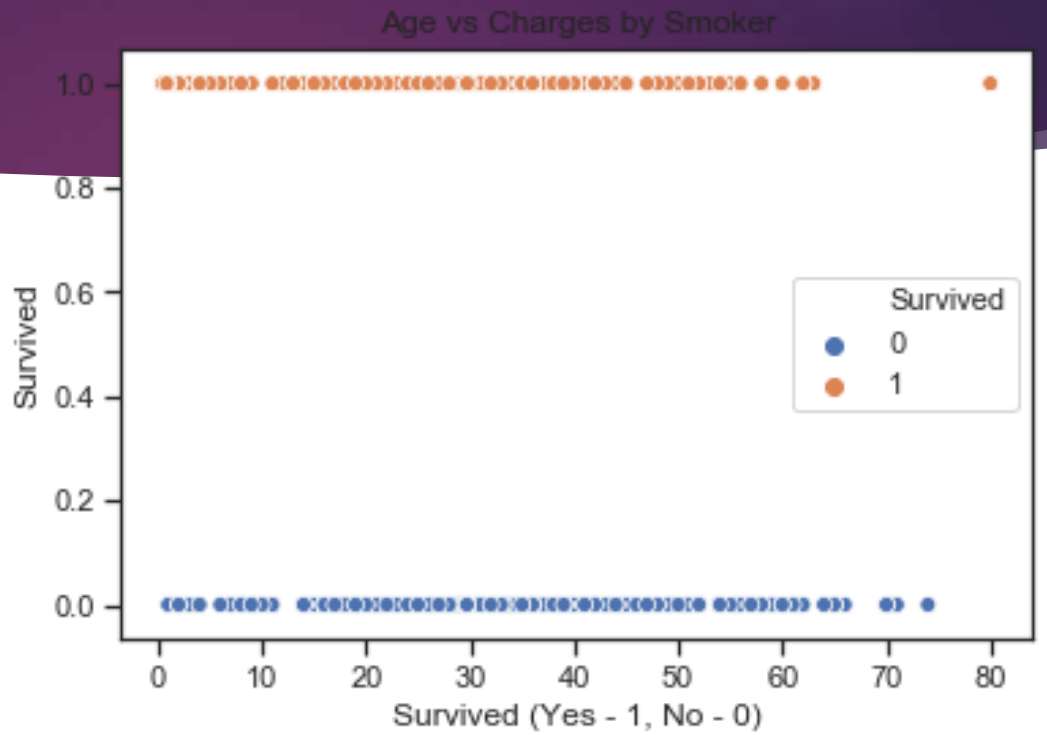
The above heatmap shows correlation beween different variables

There is a high correlation between Sex and Survived and Pclass and Survived,Age and Pclass.



Scatterplot & Dist plot

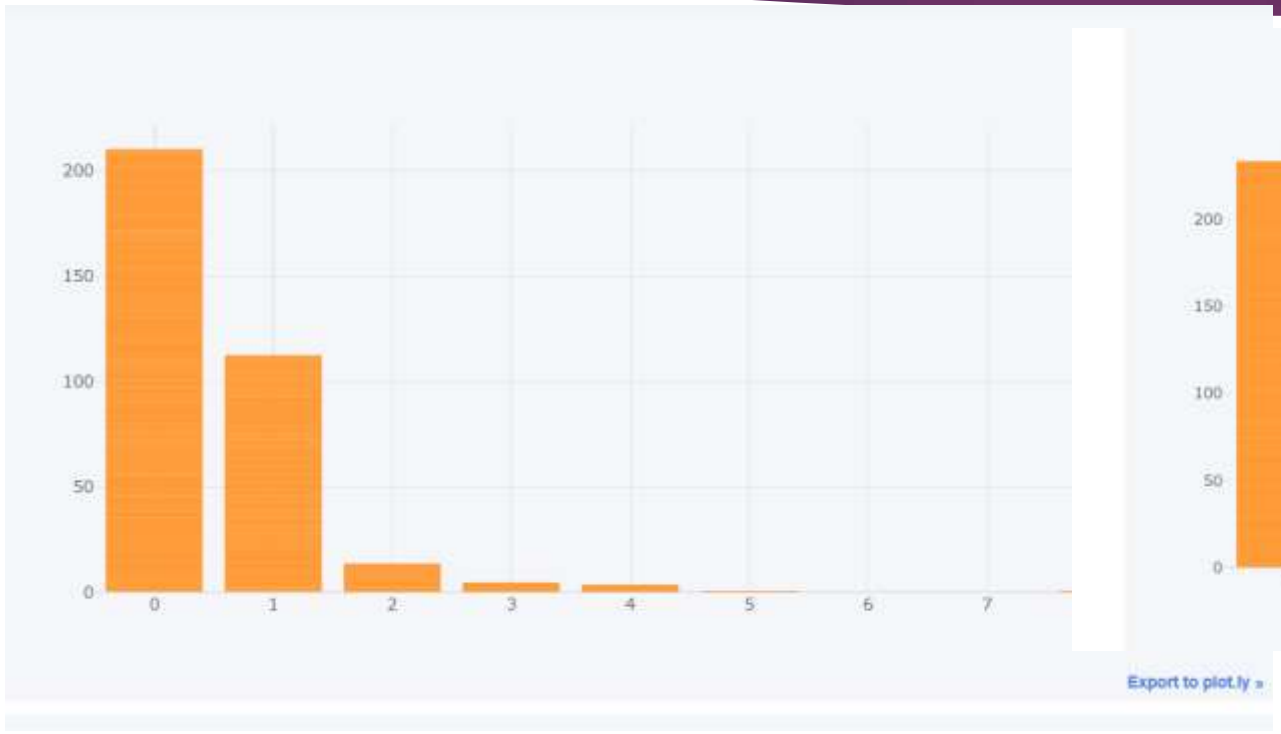


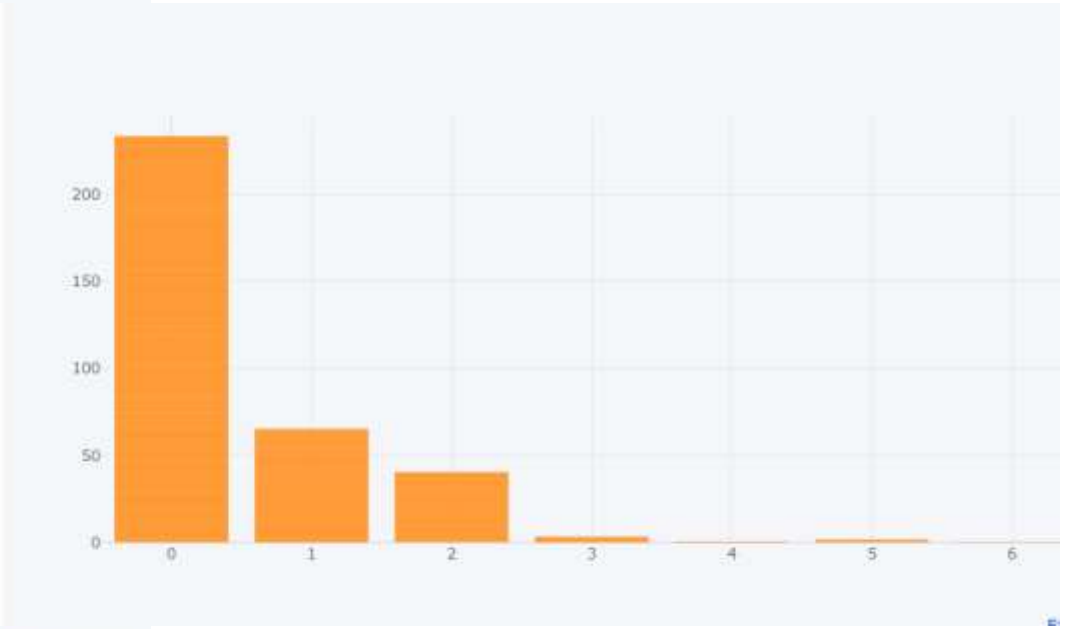


The above dist plot shows that there is a spike of passengers in between age group 25-35

The above scatter plot shows Survived and not survived in different age groups

Interactive barplot

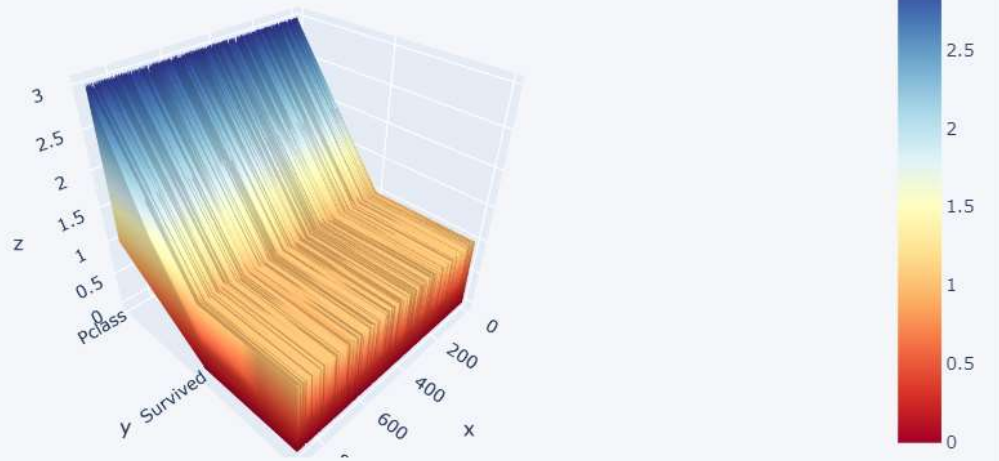




The fist fig shows that passengers travellling alone survival ratio is high compared to passengers travelling with spouse and siblings

The sec fig shows that passengers travellling alone survival ratio is high compared to travelling with

Interactive 3D plot

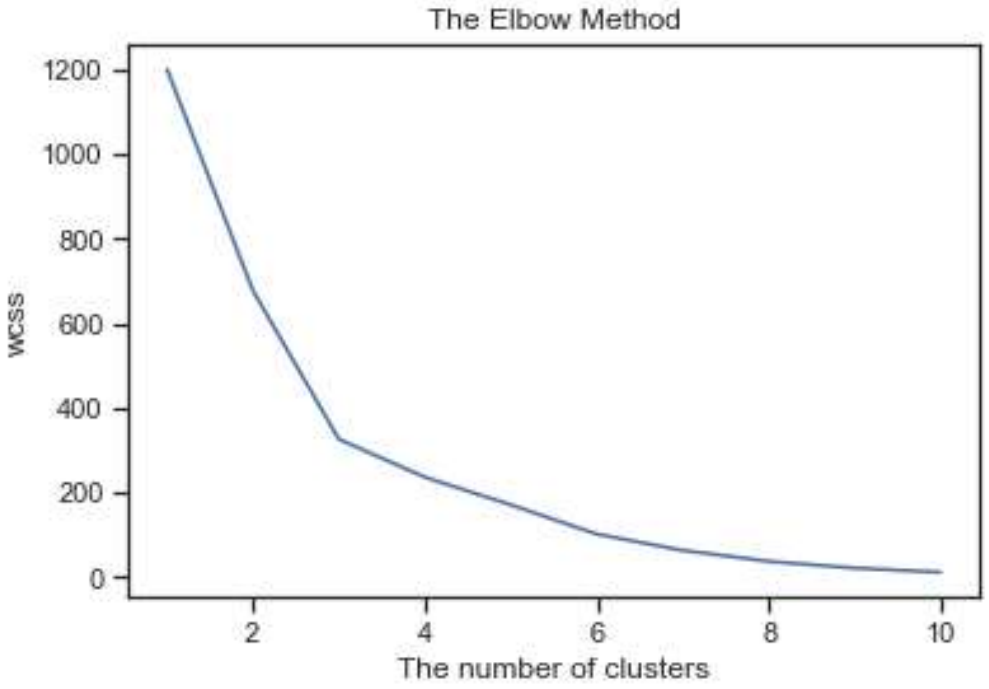


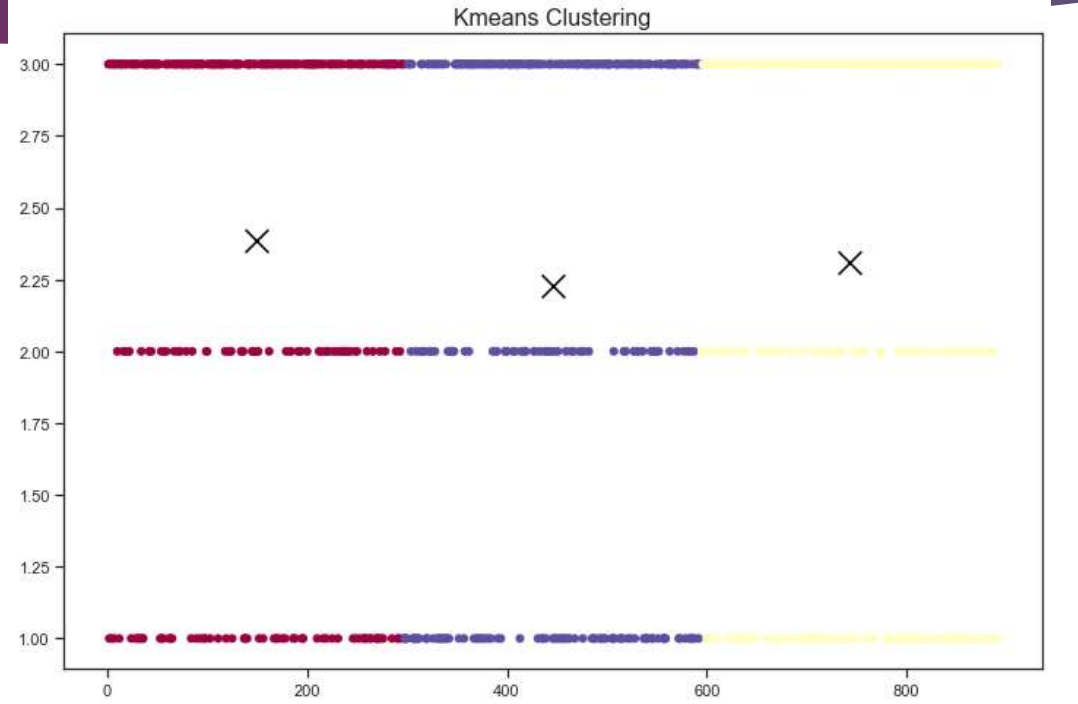
Interactive 3D plot shows Survival rate by Pclass and sex

Machine Learning Models

- R2 Score for Multiple Linear Regression Model is 0.42
- R2 Score for Random Foresrt Linear Regression Model is 0.38
- R2 Score for Decision Tree Regression Model is 0.10
- Hence Multiple Regression Model is the best model

Kmeans Clustering





The elbow method is a technique to choose the most number of clusters. *The graph above shows that there are 3 clusters in the data set.

Conclusion :

- Survival ratio is high in females than males
- More number of passemgers are in the age group 25-25
- More number of passengers are from point of embarkation southhampston
- More number of passengers have surbived in Pclass1
- More number of passengers travelled in Pclass 3
- More number of male passengers are than the female passengers
- Survival rate is more in Southhampston embarkment than cherbourg and queenstown
- Age group who have survived more are in the range of 25-30
- Youngest traveller is below 1 yr and oldest traveller is 80 yr old
- Out of total 891 passengers 342 have survived and 549 have died
- Avg age of passengers is 30.Avg male age is 24 and Avg fem age is 23
- Females survival rate is 74% and males survival rate is 19%
- R2 score for multiple linear regression is 0.42 which is best score to test new data for predicting survival rate.


Thank
You