Report on titanic

Introduction:

On April 15, 1912, during her maiden voyage, the widely considered "unsinkable" RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren't enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

So the objective of this task is to conduct a comprehensive analysis on the dataset and provide a report with respect to factor that contributes to a passenger surviving or not. The data has a total of 1309 rows and 13 columns. The following variables where included: survived, name, (Pclass): 1st, 2nd and 3rd class,

Sibsp: number of siblings/spouses aboard, Sex, age, parch, ticket, fare, cabin, embarked. Below is the first 5 rows and columns.

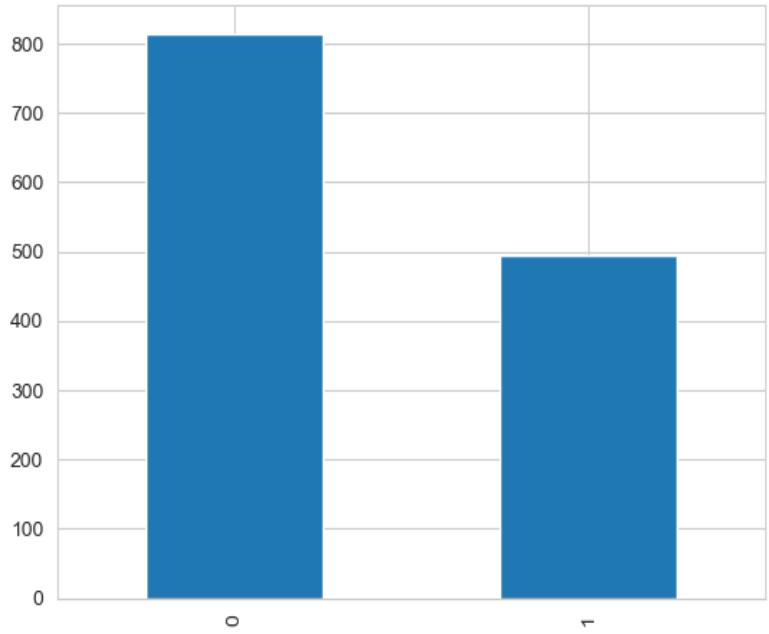
		Passenger Id	Survived	Pclass	Name	SEX	Age	SibSP	Parch	Ticket	Fare	Cabin	Embarked
0	0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	s

1	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	c
2	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	s
3	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	s
4	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	s

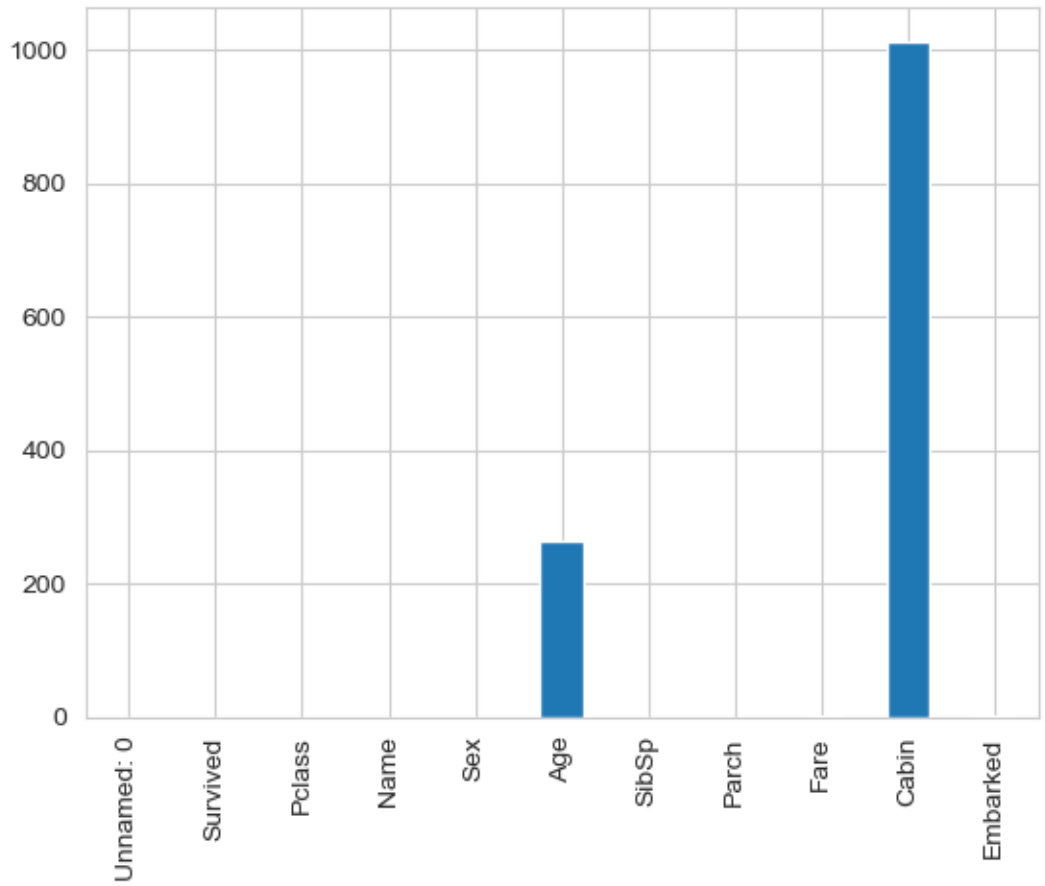
Below are the first 13th rows of the dataset (the rows are picked from 0 -4

Data description: from the description of dataset, survived = 1(true) and Not Survived = 0 (false).

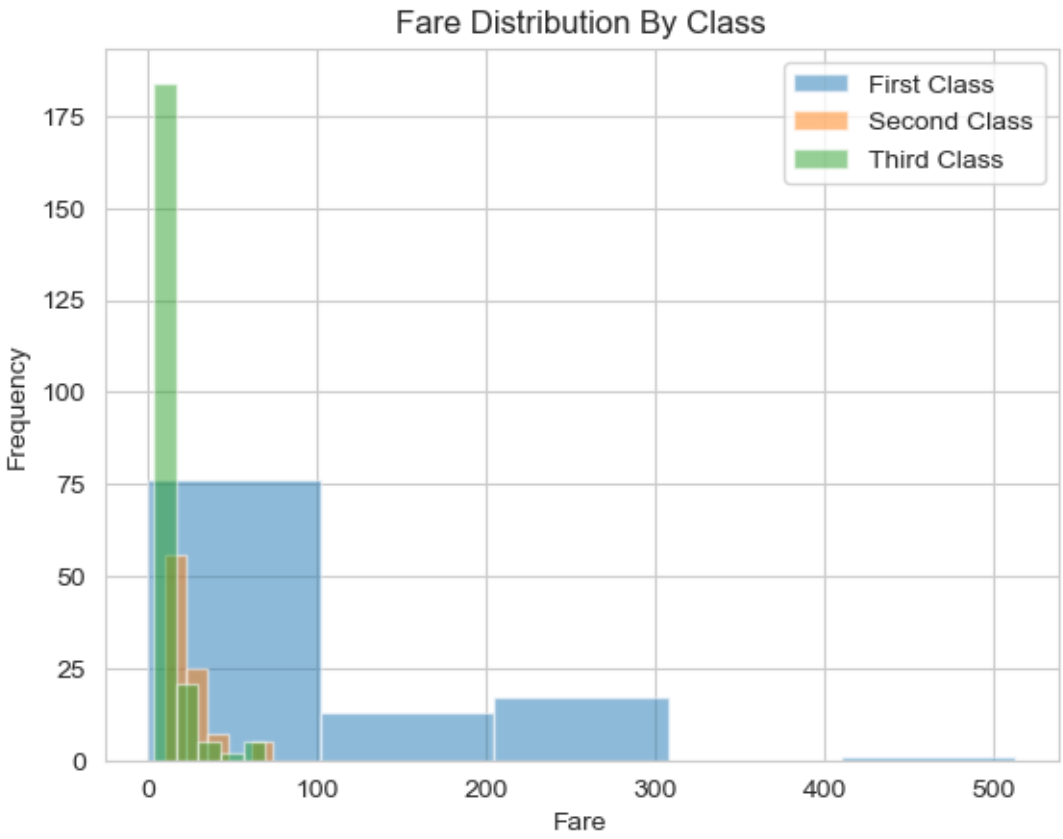
The chart below indicates the total number of people that survive and the total number that died in the titanic (survived and not survived). The number of people that survived is less than the number that did not survived.

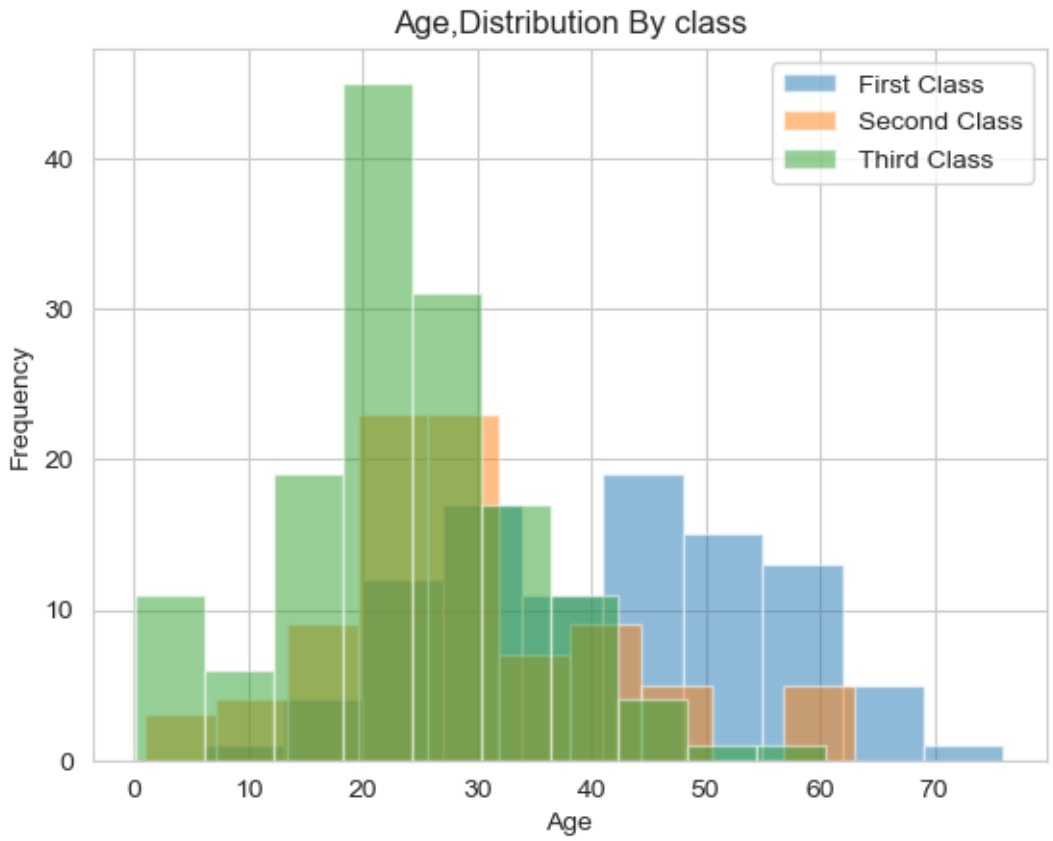


Data.null: is the missing value in the titanic data, empty column in the chart below indicating the missing values in the data. There was also missing value in Age, cabin, columns.

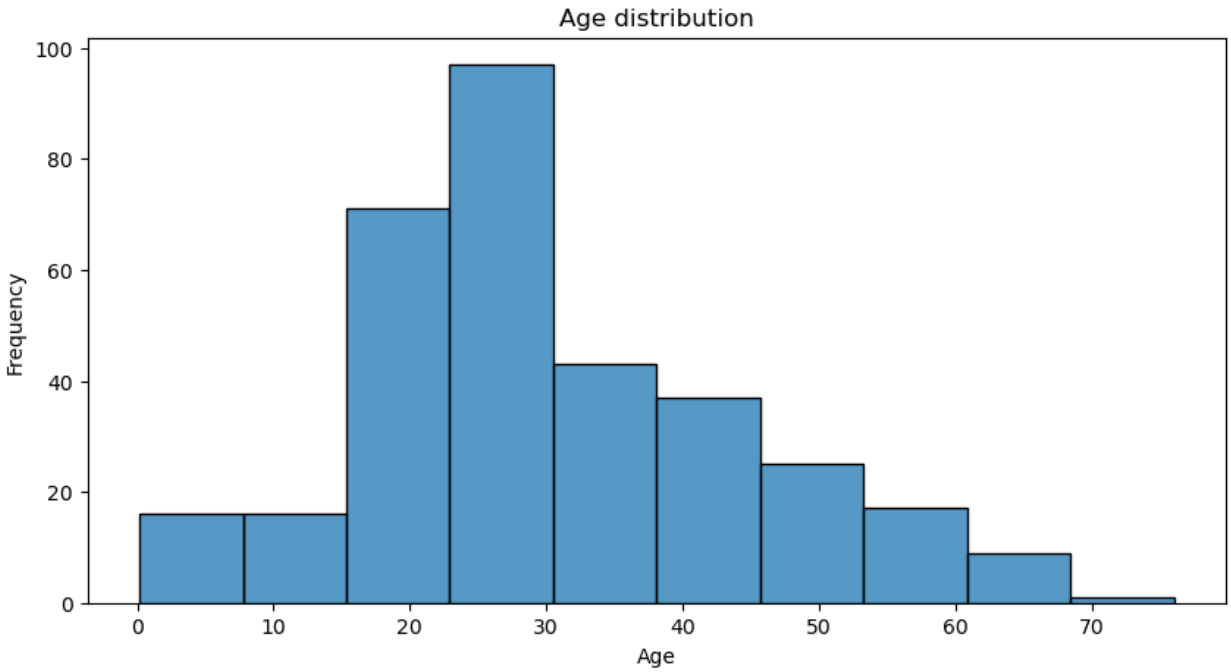


The plot below indicates that lower fare price decreases the ratio of survival. First class passengers had the highest rate of survival, then second class passengers and the least passengers in 3rd class. Passengers in 3rd class has the largest population, followed by second class and first class has the lowest Passengers with the highest fare.


Fare Distribution By Class
First Class
Second Class
Third Class
Frequency
175
150
125
100
75
50
25
0
0
100
200
300
400
500
Fare



From the above chart, the age distribution of first class ranges from 6-78. While age distributions of second class and third class range from 2-63 and 0-60 respectively. This is to imply that people in first class are of older age.



The plot above describe that the age of females was less than the number males. The age of oldest passenger age was 80.0 and 68.0 was less.

From the above chart, there is a strong positive correlation between variables (Family, Fare, Parch, Sibsp, Pclass, Survived and Age)

Family and Parch has a positive correlation of 0.85, family and sibsp also has a positive relationship of 0.79, while Age and Survived has a strong negative correlation of -1.3e-05, Pclass and Fare also has a strong negative correlation of -0.58.

SUMMARY

In conclusion, the dataset provides insight on the survival and non-survival in titanic disaster. The following factors determine the survival of the passengers; Age, gender, and socio-economic status. The data has a total of 1309 rows and 13 columns. Data information include:

survival: 1=true, 0= false. Total number of passengers that survived was 494 and not survived was 815. Percentage of people that survive was 37.79%. Percentage of people that did not survive was 37.79%. Greater number of male passengers died. Female passengers have the highest number of people that survive and male passengers have the highest dead rate. The survival rate between different class recorded that: 1st class was 323, 2nd class was 277 then 3rd class has the highest of 709. In the titanic dataset greater number of passengers died.