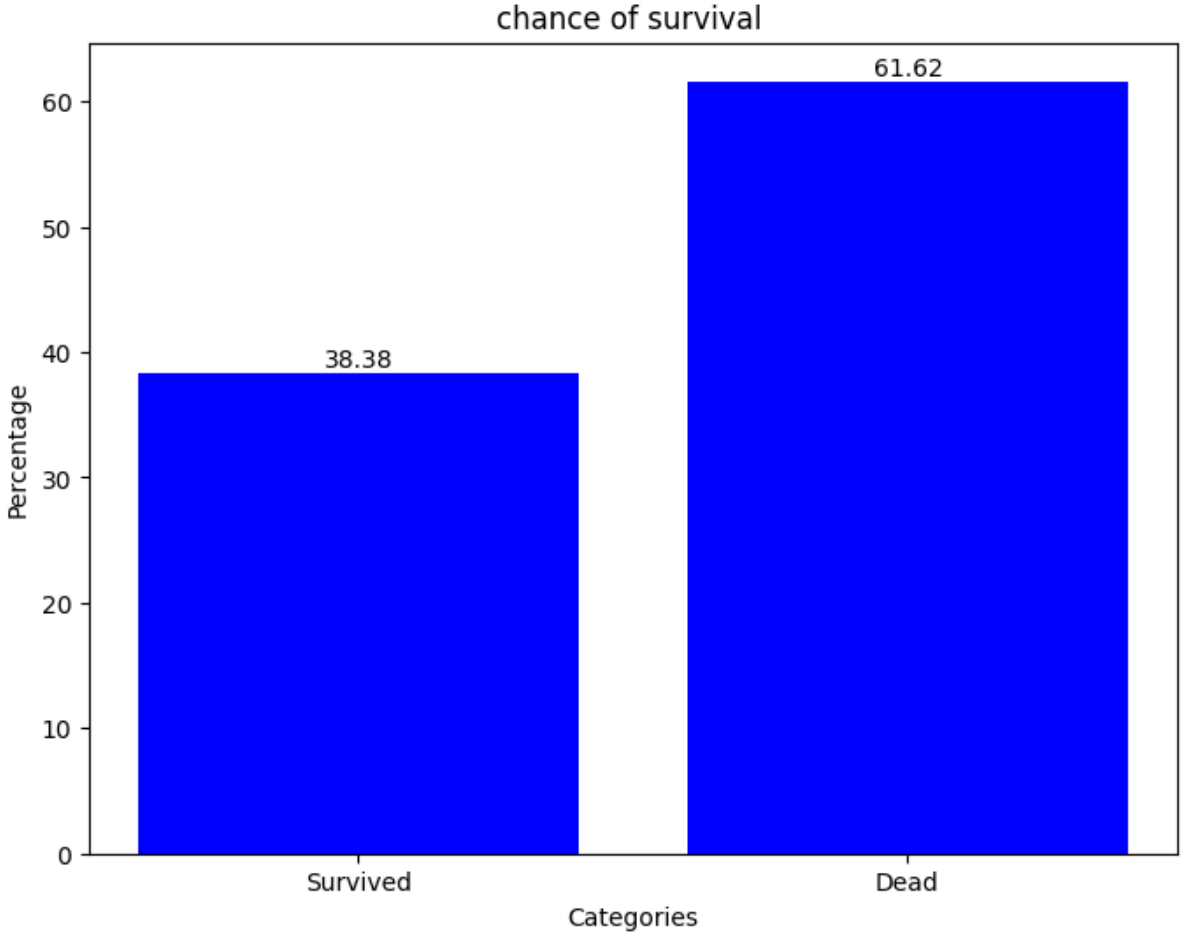
Titanic Data Analysis Overview

Data Description:

The dataset comprises 12 columns containing passenger information such as sex, name, cabin class, number of cabins, etc. It encompasses details of 891 passengers. Notably, the attributes related to age and cabin exhibit numerous missing values (177 for age and 687 for the number of cabins). Consequently, the cabin information is disregarded, and the missing age values are filled with the average age.

The 'Survived' column indicates whether a passenger survived or not, with calculations showing a survival rate of 38%.



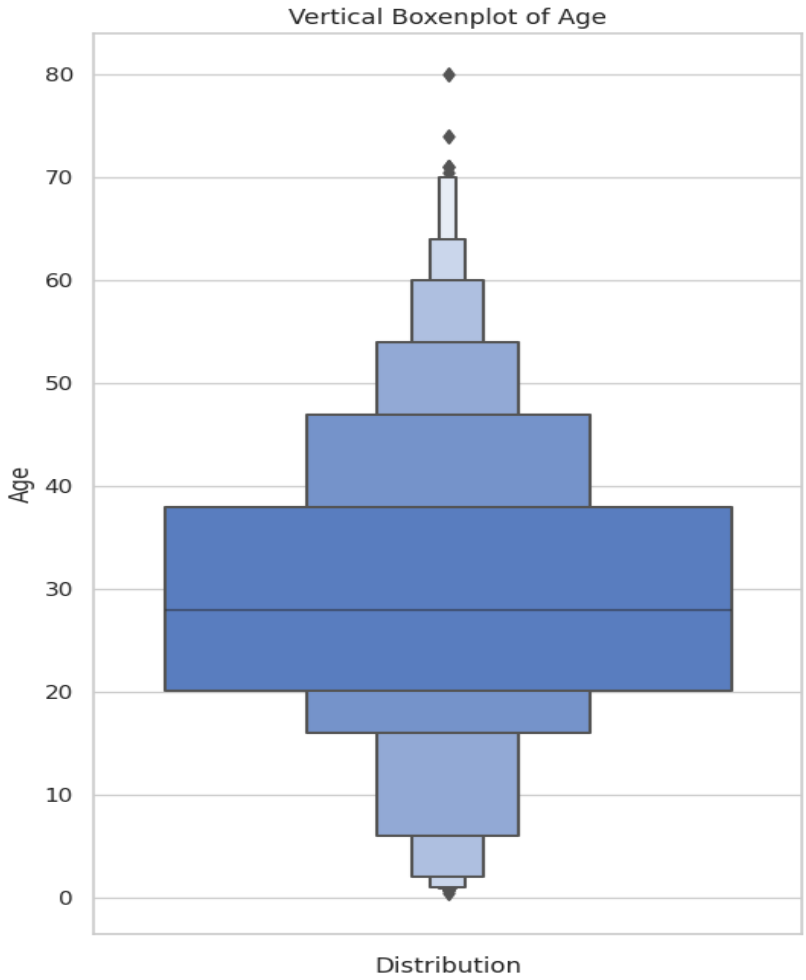
In the following sections, we delve into insights derived from various features.

Missing values:

In these data there were no missing values except for the age column and for the number of cabins. For the age, we filled missing values with average and deleted whole cabin column and decided to not working with cabin information. Because there were 687 missing values and we only had 891 data.

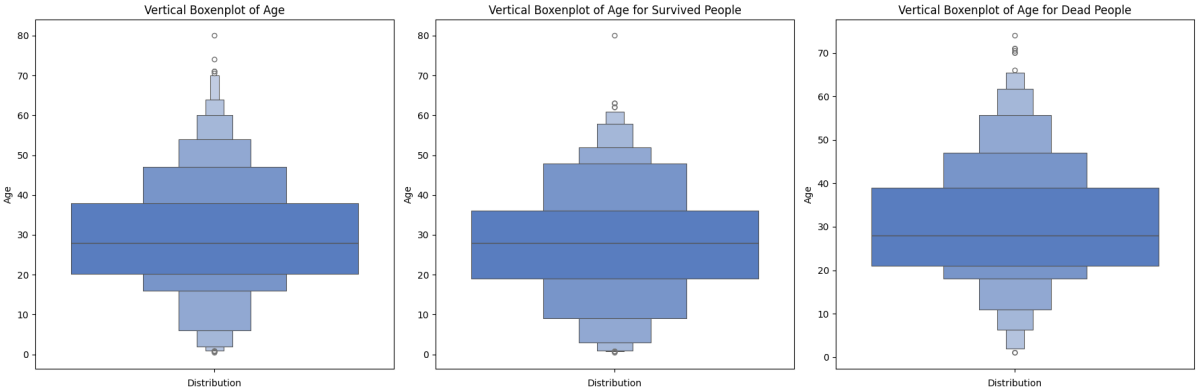
Age:

The average age of passengers was 28 years old, with the oldest being 80 years and the youngest less than one year old.



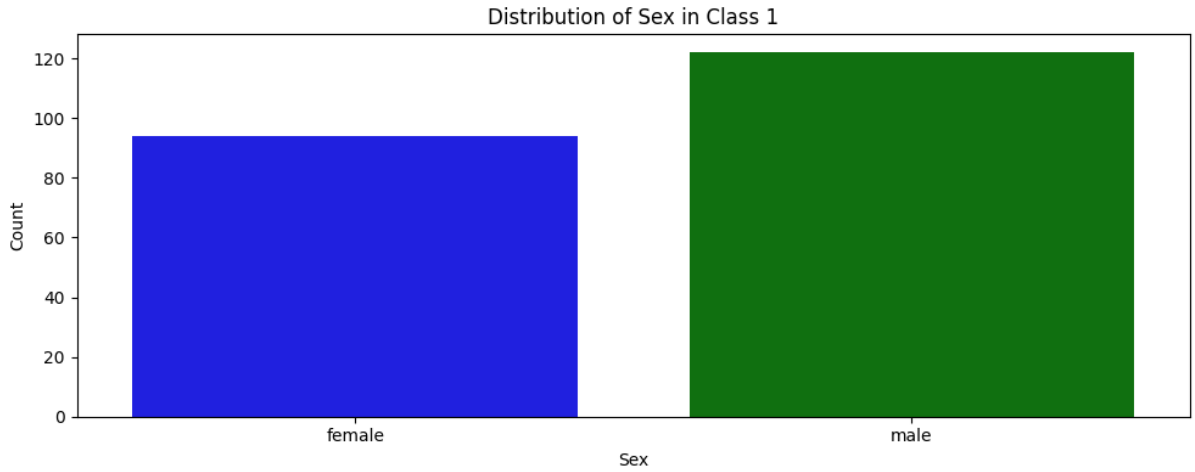
We can see box plot for the age of all passengers, passengers who survived and un-survived passengers, respectively. Which shows there is a slight shift to the younger ages for the survived group but the mean does not different for three

groups.



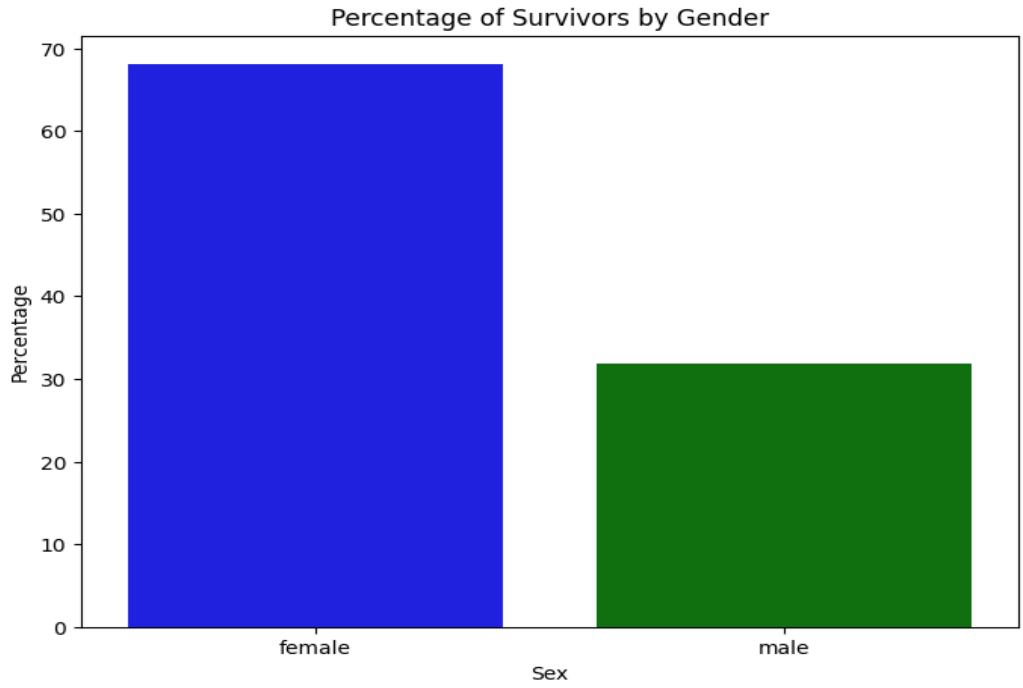
Sex:

Remarkably, the number of men exceeds that of women across all classes, with a more pronounced difference in the third class.



the chance of survival for women where more than men. Here is the

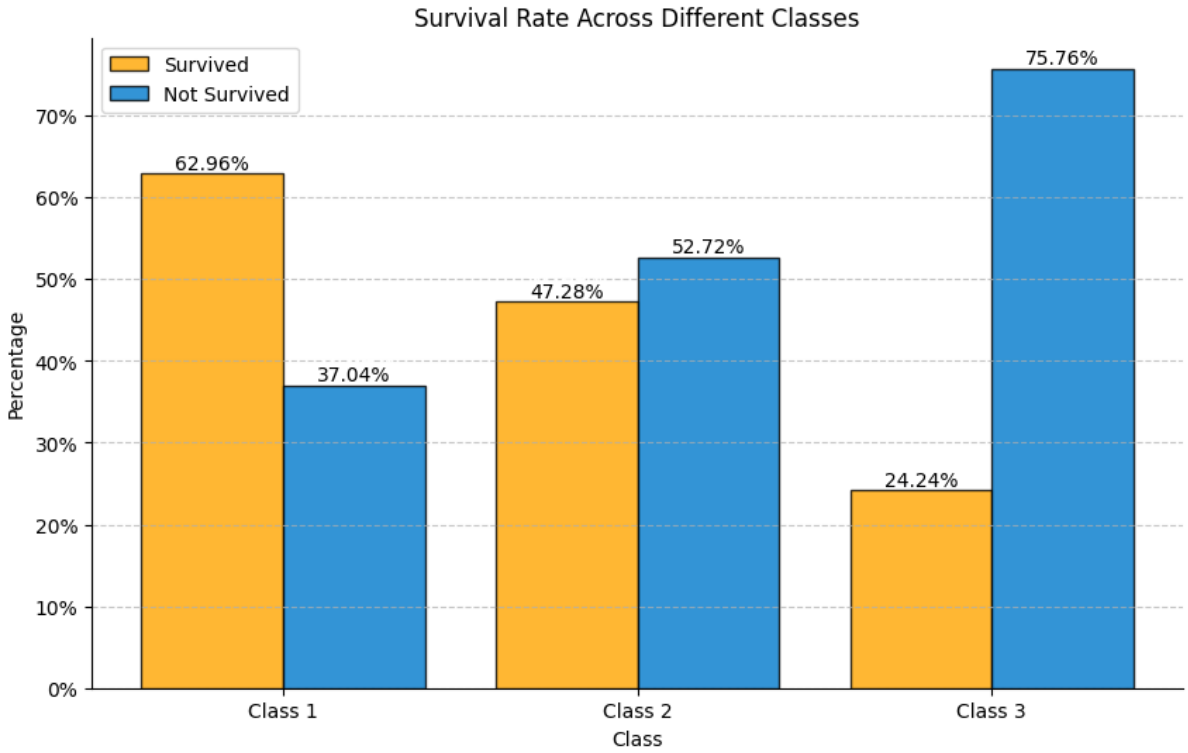
percentage of survivals that were men and the share of them who were



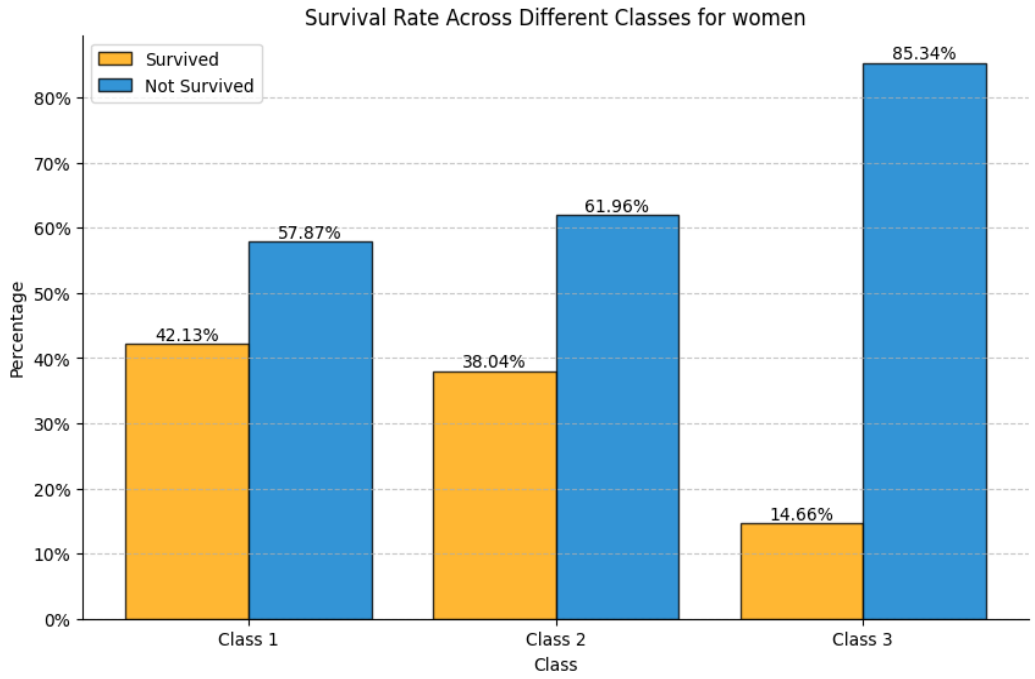
women.

P-class:

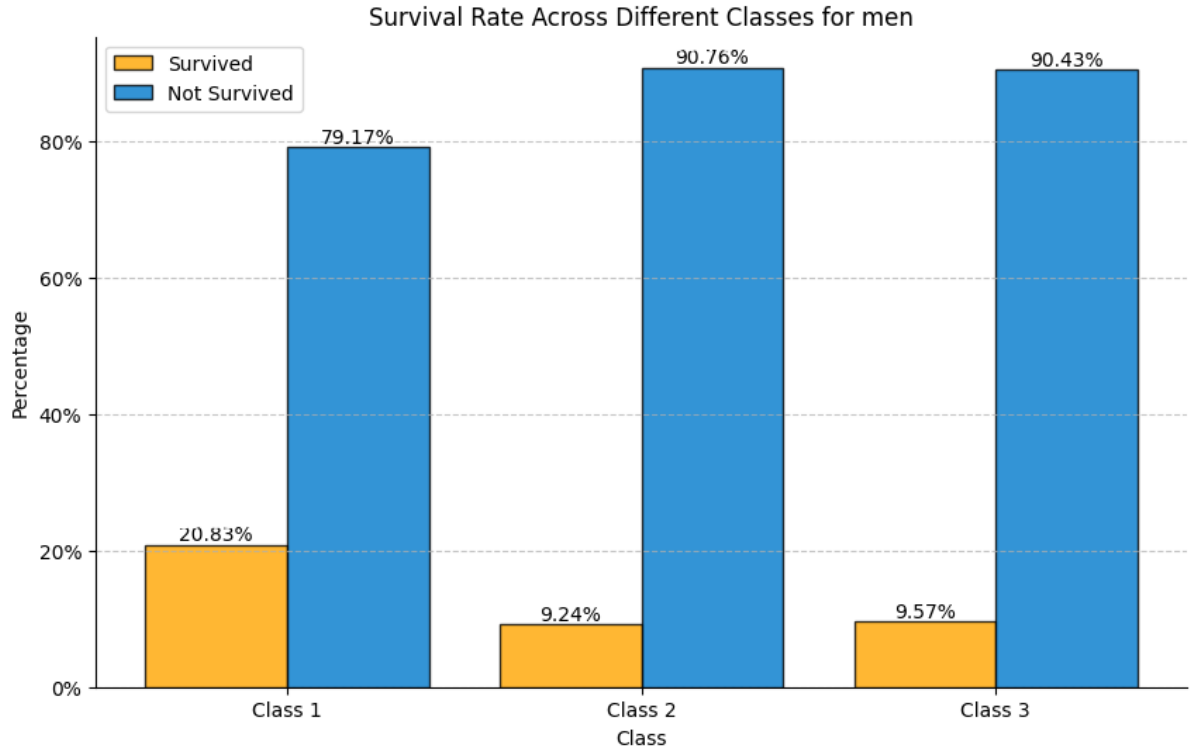
Analysis reveals survival rates of 62% for class 1, 47% for class 2, and 24% for class 3.



Specifically, survival rates among women were 42%, 38%, and 15% for class 1, class 2, and class 3, respectively.

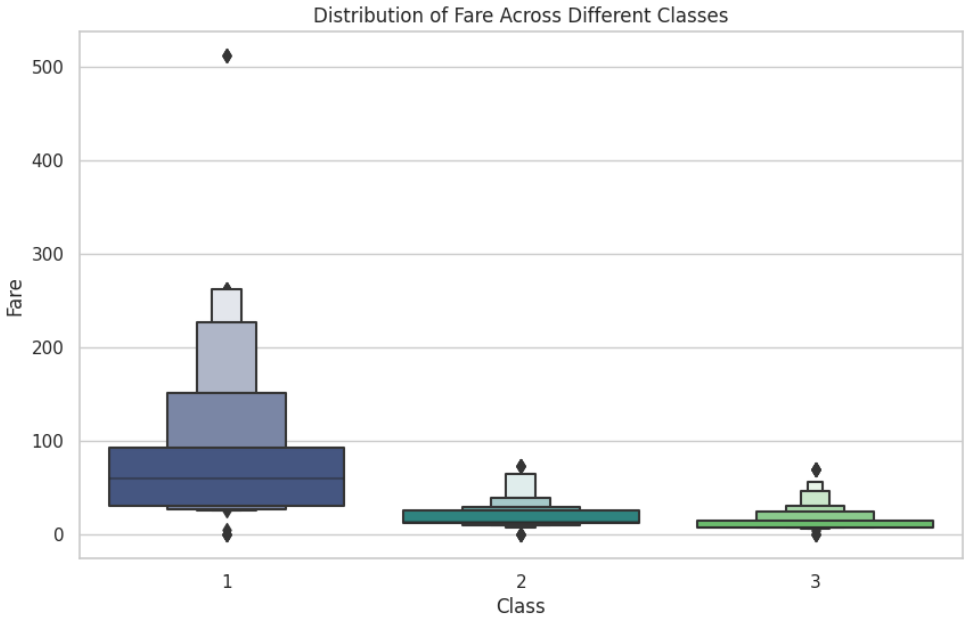


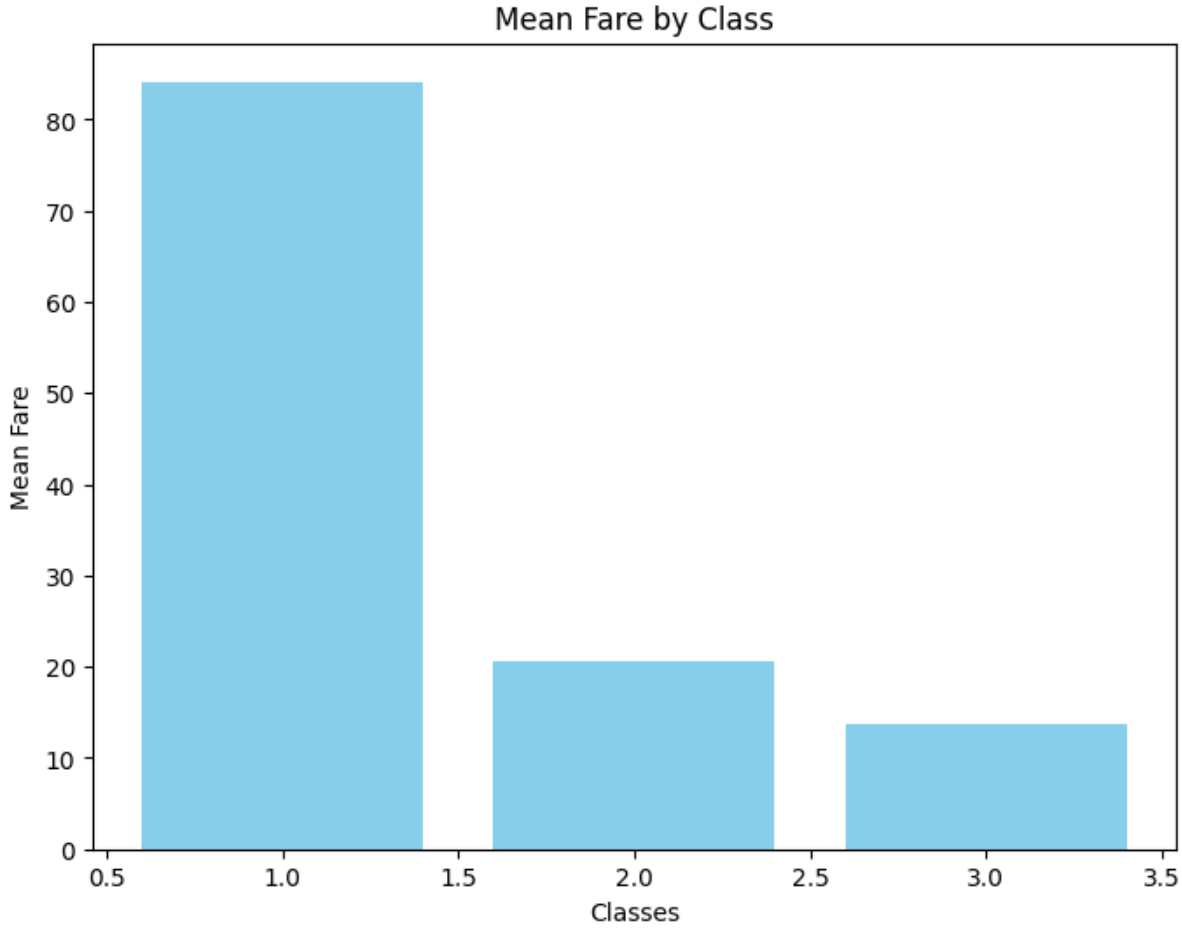
Correspondingly, survival rates for men stood at 21%, 9%, and 9% for class 1, class 2, and class 3, respectively.



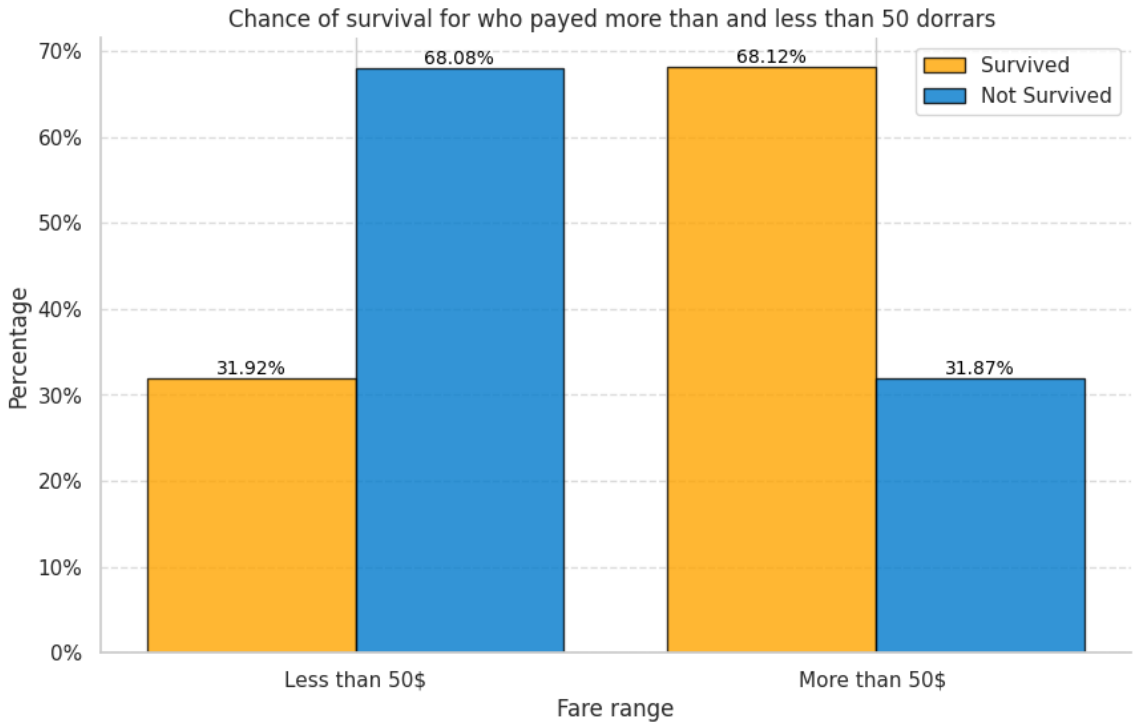
Fare:

The fare represents the amount spent by each passenger on their ticket. On average, the ticket price for class 1 was around 83 dollars, for class 2 approximately 20 dollars, and for class 3 about 10 dollars.

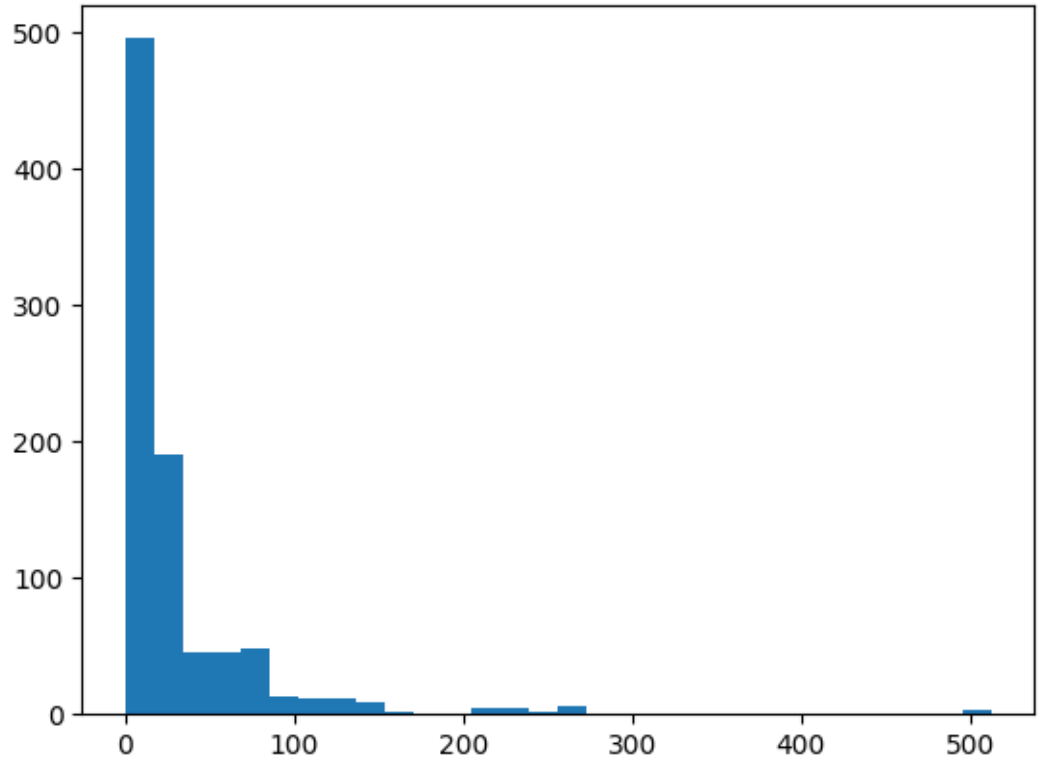

Distribution of Fare Across Different Classes
Fare
500
400
300
200
100
0
1
2
3
Class


Mean Fare by Class
Mean Fare
80
70
60
50
40
30
20
10
0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Classes

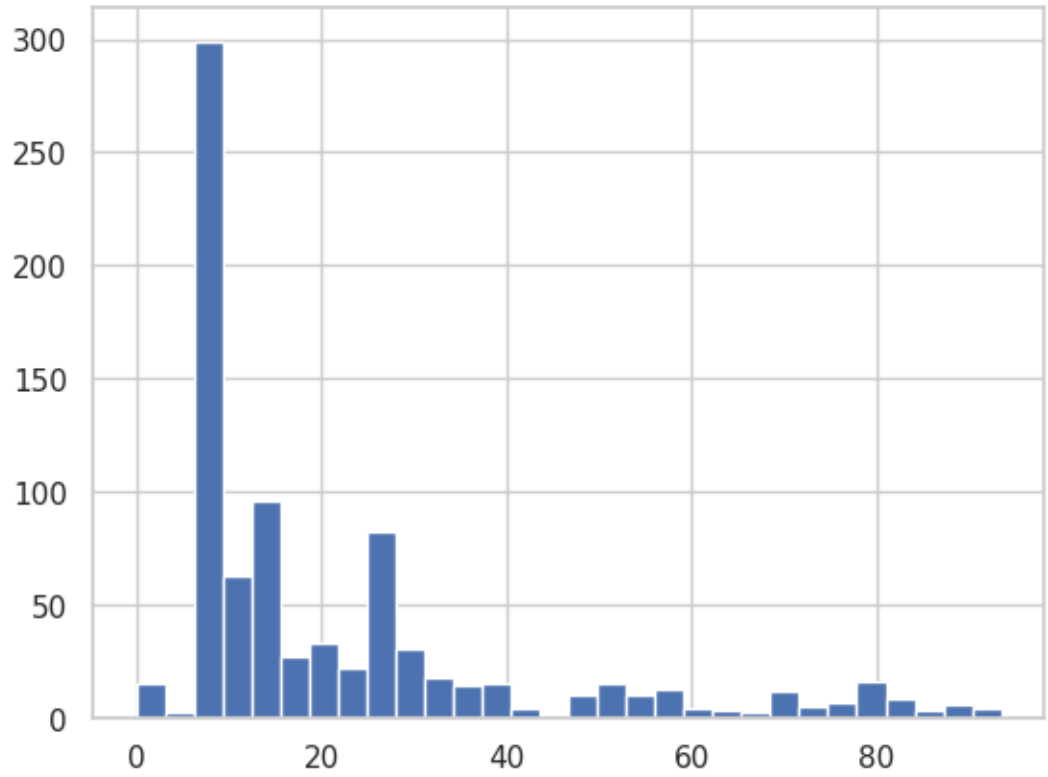
The survival rate for passengers who paid more than 50 dollars was 68%, while it was 31% for those who paid less than 50 dollars.



The histogram for the ticket price is like this:

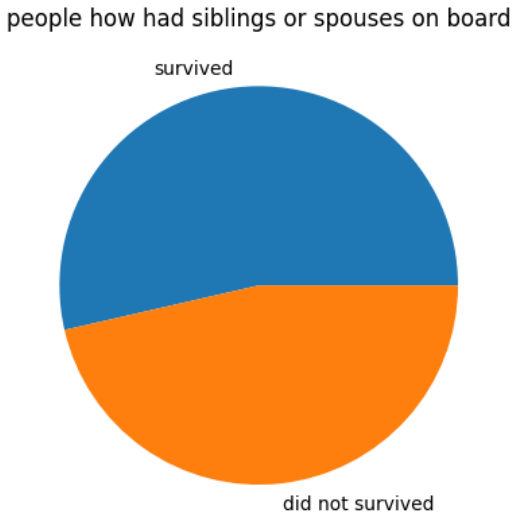


And the histogram for ticket prices under 100 is:

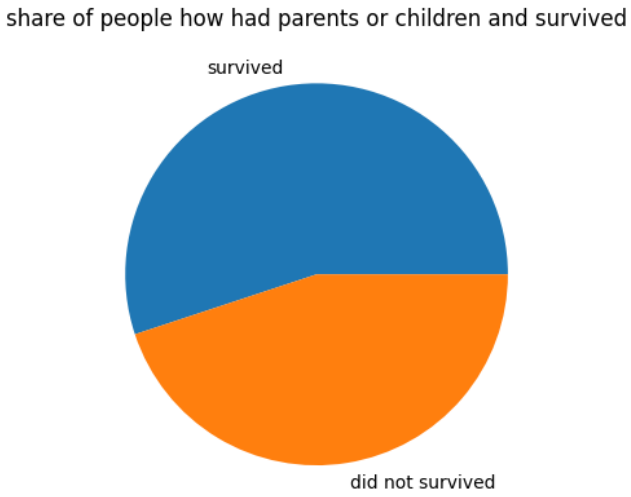


SibSp and Prach:

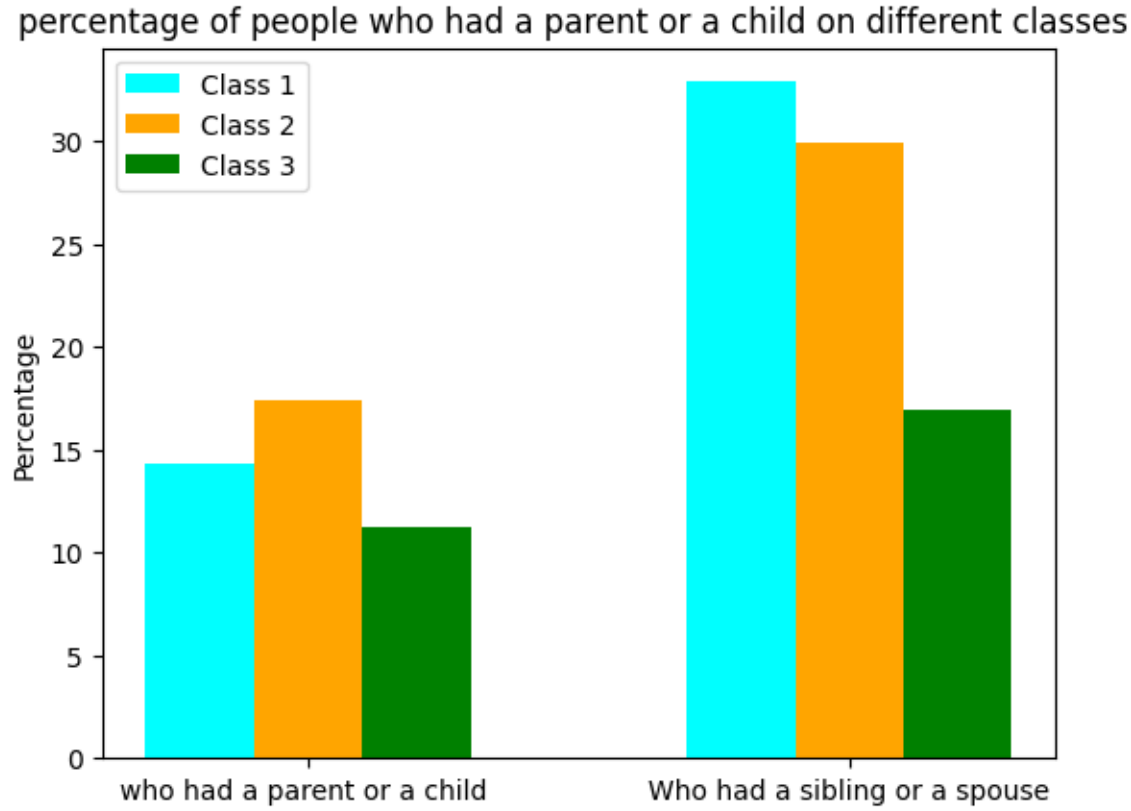
The share of people who had a sibling or a spouse on board and survived is 53%.



The share of people who had a parent or a child on board and survived is 55%.



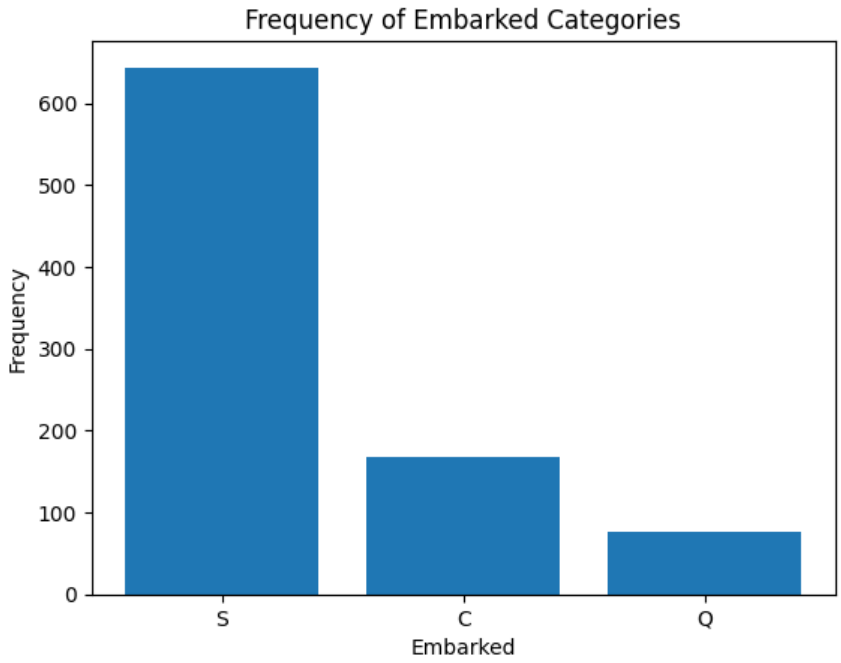
The percentage of people who had a sibling or a spouse or a child or a parent on the bard and survived is 88%.
Here is the diagram that show the percentage of people how had a sibling or a spouse and a parent or child on board in different classes.



Embarked Cities:

Passengers boarded from three cities: Southampton, Queenstown, and Cherbourg.

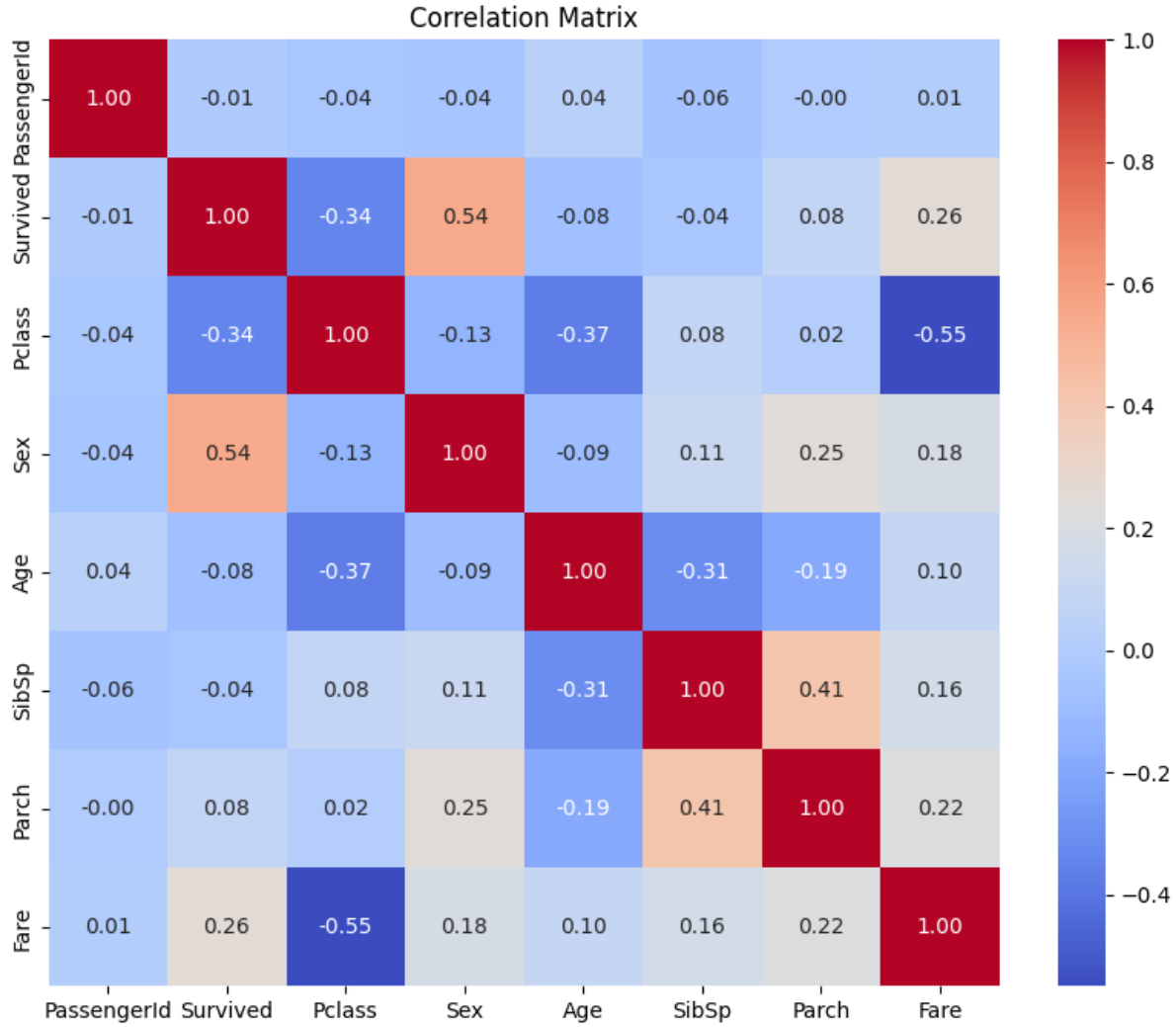
Over 600 of the 891 passengers boarded from Southampton, around 150 from Cherbourg, and about 50 from Queenstown.



The majority boarding from Southampton and Queenstown were in the third class, while most passengers from Cherbourg were in class 1. More than 55% and 95% of class 3 ticket holders boarded from Southampton and Queenstown, respectively, while around 50% of those boarding from Cherbourg were in class 1.

Correlation Between:

After extracting correlations between different features, we know that the most correlated features are fare a and class, which is -0.55. The second most correlated features are sex and survival, 0.54, which means that women had the most opportunity to survive. After that we have correlation between the SibSP (shows if passengers have siblings or spouses on board) and the Parch (shows if a passenger has a parent or children on board). The correlation is positive, which means that there is a good chance for a person to have a parent or children onboard to have a spouse or sibling on board



Conclusion:

At this section we will have a brief discussion about the results. The number of men where more than women in this ship and the chance of survival for women were more than men (By about 40%). The results shows that the rate of

survival was more among passengers who had first class tickets and the more expensive their tickets were, the more likely they were survived. Here we can see a little correlation between the fact that if the passengers had a relative on the board and the chance of survival (by 5%). The results shows that the age of survived passengers was younger than the ones who died. But the age average does not different for different groups. Furthermore, it seems that the passengers in the upper classes had the better chance to survive.