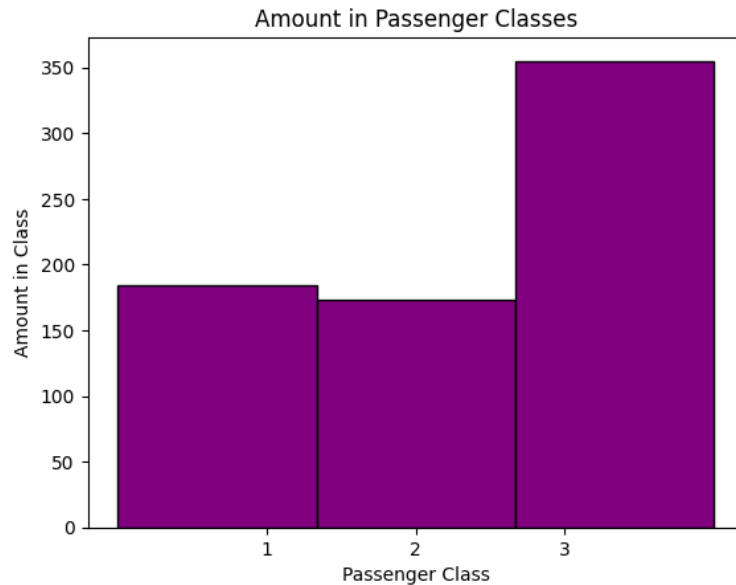# Analysis of Titanic Data Set

Introduction:

This report dives into analysis of the titanic data set to analyze and interpret the numerical and categorical data. To illustrate the multitude off data within this set, the use of histograms, charts, plots, and contingency tables were used.

Passenger Classes:

To show the variability of how many people were in each passenger class, the use of a histogram was used.

Amount in Passenger Classes

Based on this histogram, the interpretation is that the largest amount of people was in third class, with roughly 350 people. On the other hand, first and second class were very close together, however, second class ended up being the lowest.
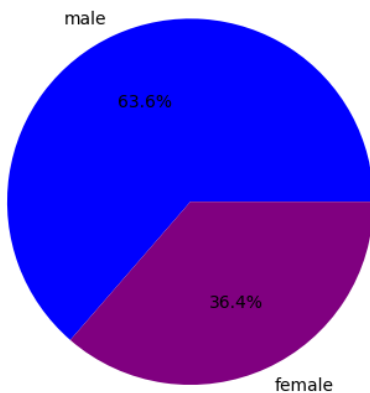
Males Versus Females:

When examining the difference between the male and females that were on the titanic, a two pie charts were used. The first chart was used to measure a male to female ratio.

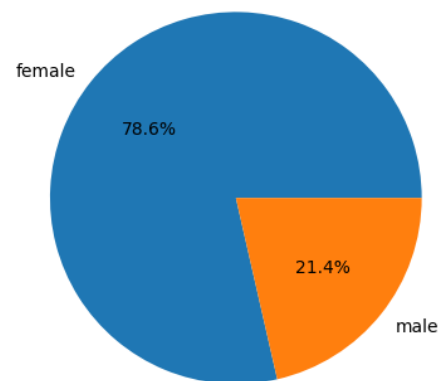Based on the Male vs Female graph, 63.6% of passenger were male, while only 36.4% of passenger were female.

The second pie chart that was used, measured the percentage of survivors that were male or female. Unlike the male to female ratio, females were higher, having 78.6% of the survivors. Due to this, males resulted in only having 21.4% of survivors.
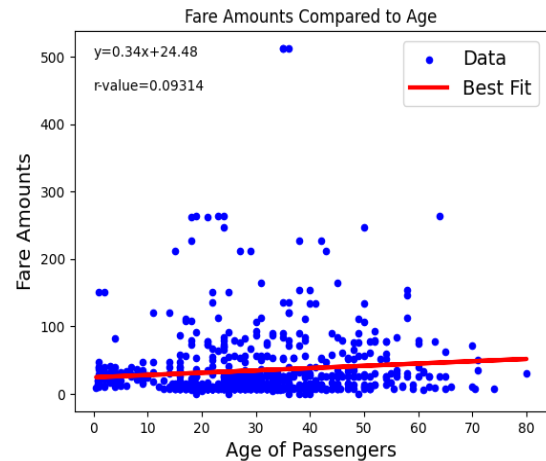

Male vs. Female
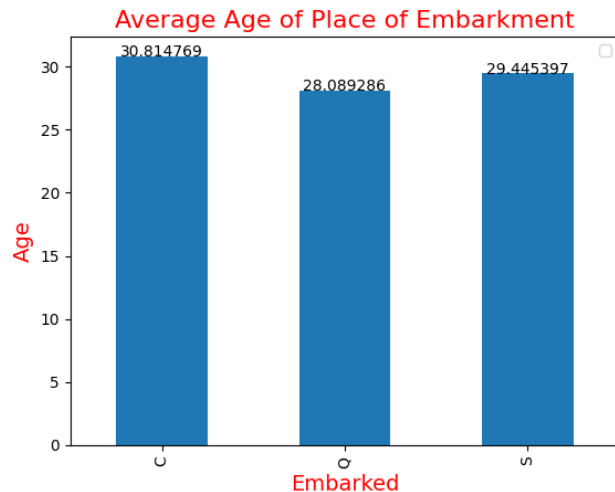

Survival Rates of Males and Females

Passenger Age Vs. Fare Amounts:

Next, the comparison of the ages of passengers and the fares that they paid was to be analyzed using a scatter plot. Here, it can be viewed that most of the fare prices are below $100, with some outliers above $500. It can also be interpreted that most people were around the age of 30 because that is where most of the points are.

Age to Embarkment:

To go along with passenger age. The average age was also taken from each place of embarkment. As it's seen, the average age was between 28-30 years old from the three places of embarkment. The highest average came from Cherbourg, while the lowest average come from Queenstown.
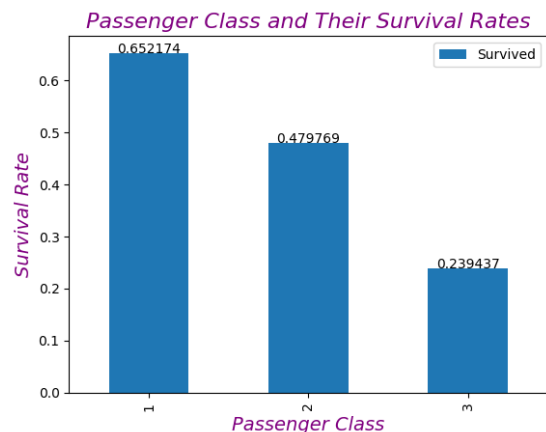
Contingency Table of Sex Vs. Embarkment:

Next, the comparison between sex and place of embarkment was made using a contingency table. The p-value from a chi-square test was also measured as 0.013327. Showing that there was a significant correlation as it was below 0.05.

	Cherbourg	Queenstown	Southampton
Female	61	12	186
Male	69	16	368

Passenger Class and Their Survival Rates:

Furthermore, the passenger classes and their survival rates were measured using a bar graph. Here, we can see that 0.65 of class 1 survived while only 0.24 of class 3 survived.

Conclusion:

In conclusion, the titanic data set has many categorical and numerical types of data that can be analyzed. With this report, we analyzed age, sex, embarkment, and survival rates.