**Titanic Ship Fares Analysis**

A graph of passengers on a ship

Description automatically generatedThe tragedy of the titanic is one of the most known and heartbreaking events of the past. However, when taking a deeper look into the data, there is an opportunity to see connections between different variables, like ticket fare, survived, embarkation location, etc. Specifically, this report investigates the connections found with the ticket fares of the passengers. These connections deal with age, class, and location of embarkation.

To begin, I started by making a histogram of the ages of the passengers abord. As seen below, most the passengers on the titanic fall in the age range of 18-55 years.

I then used a scatter plot to see if there was a linear connection between the age of a passenger and their ticket fare. The correlation score between age and fare is approximately -0.092424.

A graph with blue dots and a red line

Description automatically generated

Strong correlation scores are close to 1 or -1, so this score is extremely weak. I then found the P-Value between age and fare to be approximately 0.98909. Because the P-Value is greater than 0.05, these data distributions are not statistically significant from one another.

I then started to look for a connection between fare and embarkation location. To do this, I used a contingency table to view the data and calculate a p-value.

A screenshot of a computer

Description automatically generatedThe contingency table I created shows the relationship between the ticket fare and location of embarkation. I then calculated the P-Value which came out to approximately 0.0015, as listed below. Because the P-Value is less than 0.05, this data is statistically significant.



A pie chart with a blue and orange circle

Description automatically generatedI then used a pie chart to see if the distribution of passengers was even between all three locations of embarkation.

This pie chart shows that the representation of the three locations is not even. Southampton makes up over half of the data for location. With Queenstown having barely any representation.

I used box plots to look at ticket fare distribution at the three ports of embarkation.

A diagram of a ticket fare

Description automatically generated

A white rectangular object with a line in the middle

Description automatically generatedA diagram of a ticket

Description automatically generated

The boxplots make it easy to compare the fare distribution. Cherbourg and Southampton had a median ticket fare of about 50 dollars (orange line). However, Cherbourg has an outlier of an over 500 dollars. The only tickets purchased by class 1 in Queenstown were about 90 dollars.

In conclusion, this data is not as correlated as thought and more entries from Queenstown are needed before making any conclusions for embarkation location.