Technical Report

A graph of a number of values

Description automatically generated

🡨Histogram

Importing The plt.hist function creates a histogram by highlighting columns of the numerical data in the Titanic Dataset that creates a figure representing the category “Age”. I then added titles and labels with the Values and the Frequency on the axis’s .

A pie chart with text and numbers

Description automatically generated

Pie Chart:

Using the Titanic Dataset again, Figure the size of the subplots for the pie chart and sets them all equal. Using the created variable “pclass\_counts” and setting it equal to the dataframe. Highlight the column I want “Pclass”, to then pull out the values in the column. Next I set the proportions of the Pie chart by setting the count of passengers In each class 1 – 3 , I then created the Legend Filler by setting the variable to equal the class for each label. I then defined the function to make autopct labels, by using absolute to equal “pct/100” then multiply the sum by all values and return it back into the % to format on the pie chart. Finally adding the legend using ax.legend, and giving the title of the legend “Passenger Class” and “Passenger Class and the Number of Passengers in Each Class on the Titanic” to add the title for the pie chart. The pie chart shows the classes on the titanic and the passengers in each class on the Titanic.

A graph with blue dots and red line

Description automatically generated🡨 Scatter Plot

The y values are selected based on the rows where x values are present only. By using x.index, I can extract the indices of the non-missing 'Age' values and use them to select the corresponding 'Fare' values to have the same number of datapoints for each column.

The R Value is: 0.0960666917690389

Box and Whisker Plot:

A graph of a box and a diagram

Description automatically generated with medium confidence

The correlation between the numerical values in the columns Survived, Sex, and Age. These columns are interesting because by visualizing age in relation to sex and survival The box plot provides a visual representation of the median, quartiles, and outliers within each group, I can assume that some different factors influenced the chances of survival on the titanic which might’ve lead to a more skewed result.

Contingency Table:

Pclass 1 2 3

Survived

0 80 97 372

1 136 87 119

The Chi Square Test:

The question I was wondering about and tried to solve during this report was, Is if the Class the passenger is located in is relative to if the survived or not .The P-Value is 4.549251711298793e-23 meaning that, survival is not correlated to the class the passenger. The Actual Survived is(136, 87, 119) and deiced being (80, 97, 372). The expected survival being (82, 70, 188) and deiced being (133, 113, 302) with the differences between the two categories being (54, 17, 69) for the survived and for the deiced (53, 17, 70). Meaning that the two categories’ values are independent of one-another.