Python Technical Report

Introduction: This technical report uses cell-based Python to analyze the titanit_train.csv file. The purpose of this exploration is to see how the age, port of embarkment, fare price, affects one another alongside the survival rat

- To get an overview of the age of the people aboard the titanic, we first programmed a histogram. What we can conclude from this histogram is that most of the people aboard the titanic were between the ages of 20 and 40 alongside a noticeable minority of people between the ages of 0 and 20.

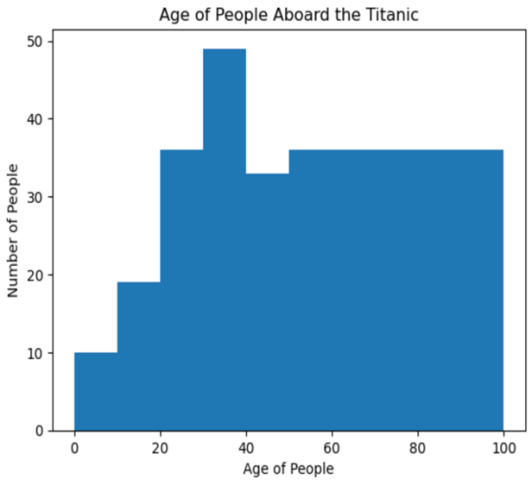

Figure 1: Age of People Aboard

- To get an overview of the of number of people who boarded from each gate, we created a pie chart to visualize this. We can conclude from this that the most people came from Southampton, but more analysis is needed to make that information valuable.

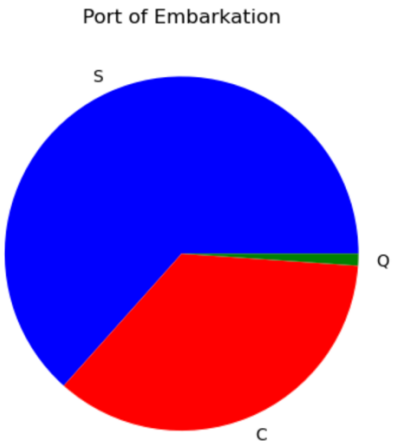

Figure 2: Port of Embarkation

- To find any further correlation between fare and passengers, we must plot it on a scatter plot. From this graph we can conclude that there is no correlation between the age of the passengers and the price of the fare.

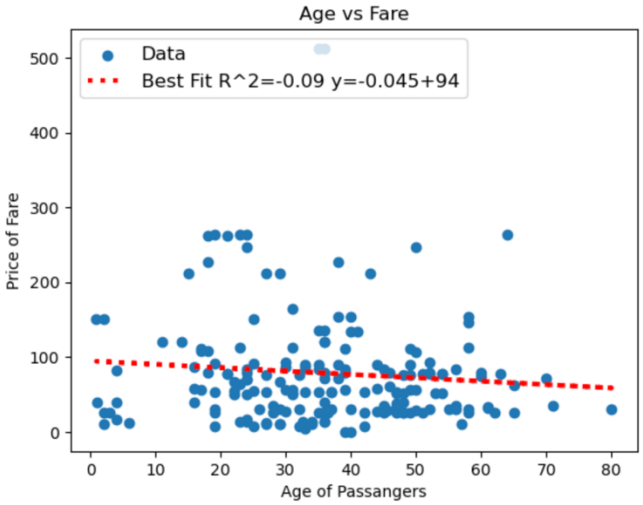

Figure 3: Age vs Fare

- To further prove that the price of the fare is irrelevant, we can include a boxplot of the data. From this we can assume there are factors outside of the data we are viewing affecting the price of the fare.

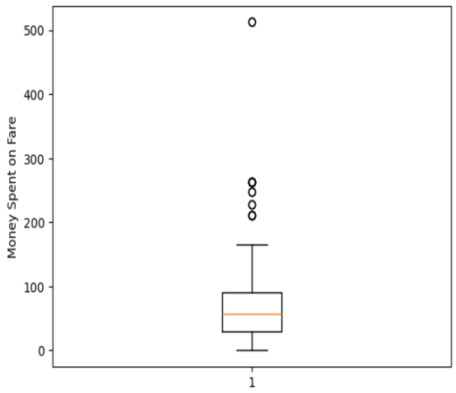

Figure 4: Boxplot of Money Spent on Fare

- Now that we know that the fare is irrelevant, we can create a pivot table comparing age, port of embarkation, and the survival. From this data we can conclude that there are not correlated since the P-Value is greater than 0.05.

```
#Contigency table comparing Age, Port, and Survived by the Average

pt = pd.pivot_table(titanic_train, values="Age", index=["Embarked"],
                    columns=["Survived"], aggfunc="mean")
print(pt)
```

```
Survived          0          1
Embarked
C         43.235294  35.416667
Q         44.000000  33.000000
S         40.523810  31.275946
```

```
c, p, dof, expected = scipy.stats.chi2_contingency(pt)
print("The P-Value is:", p)

#Chi-Squared Test
#No major correlation since P>0.05
```

```
The P-Value is: 0.9618941488583608
```

Figure 5: Pivot Table

Conclusion: The age, fare, and port passengers embark are not correlated to the survival rate of the passengers with the tests we are able to perform using Matplotlib, SciPy, and the Pandas library within Python.