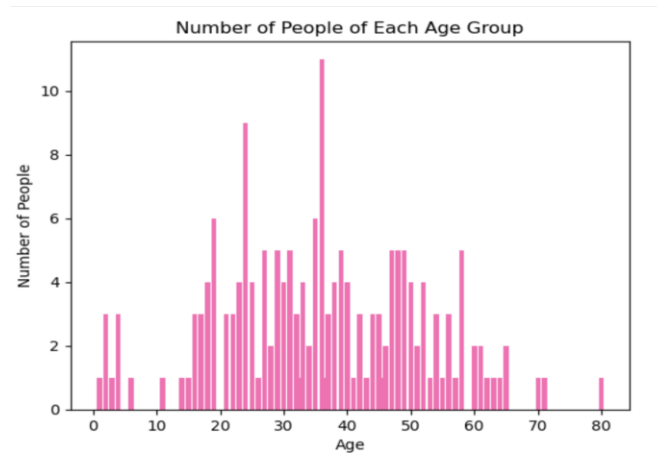


Ava Powers

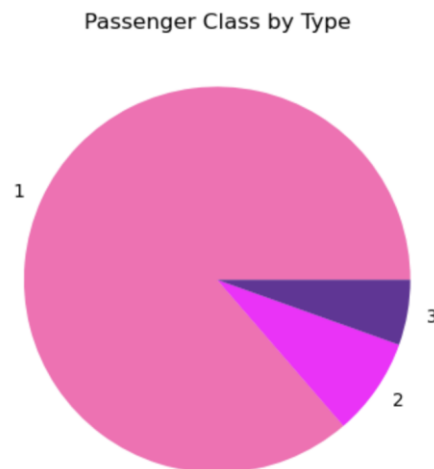
### A Data Analysis on the Titanic

The first task in my data analysis was to create a histogram of a numerical column of data. My numerical column of choice was the “age” column. For the sake of neatness and readability of my histogram, I determined how many people fell into each age group on the titanic. This result was then plotted in the form of a histogram.



The graph shows that the majority of people on the titanic fell around the age of 35, with the second most amount people being around the age of 25.

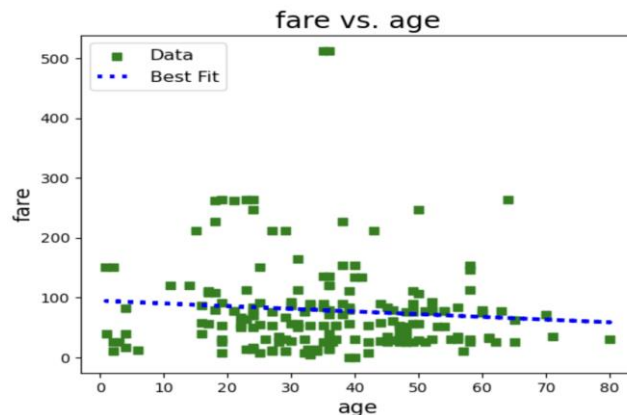
The second task in my data analysis was to create a pie chart of a categorical column of data from the titanic data set. My categorical column of choice was the “Pclass” or passenger class column. For the sake of neatness and readability of my pie chart, I first had to separate my column of data into specific types of passenger classes. I then plotted the result of this separation into a pie chart.



The pie chart shows that the majority of passengers fell into the first category of class.

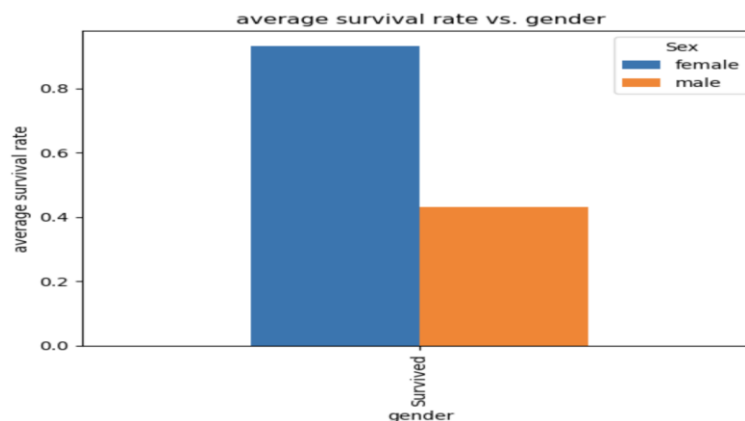
The third task in my data analysis was to create a scatter plot of numerical columns of data with a linear regression. My columns of choice were “Fare” and “Age”. I first calculated the linear regression of the

two columns. Then, I created a scatter plot including the line of best fit previously calculated from linear regression.



The graph shows that there is not a significant increase or decrease between fare and age, meaning that the fare does not necessarily change depending on age.

The fourth task in my data analysis was to show a correlation between survival rate and gender. I first created a pivot table to show the average survival rate between the genders. This was then plotted as a histogram. The graph shows that more women than men survived.



The fifth task in my data analysis was to create a contingency table and perform a chi-squared test on two categories in the data set. My columns of choice were "Pclass" and "Survived". After performing the chi squared test and creating a contingency table, I concluded that the p-value of .29 indicates that there is no statistical significance between the columns.

My final task in my data analysis was to compare two numerical columns of data. The only two numerical columns of data were "Fare" and "Age". After running a correlation calculation, it was determined that the correlation between Fare and Age was  $-0.09$ , showing very little correlation between fare price and age.