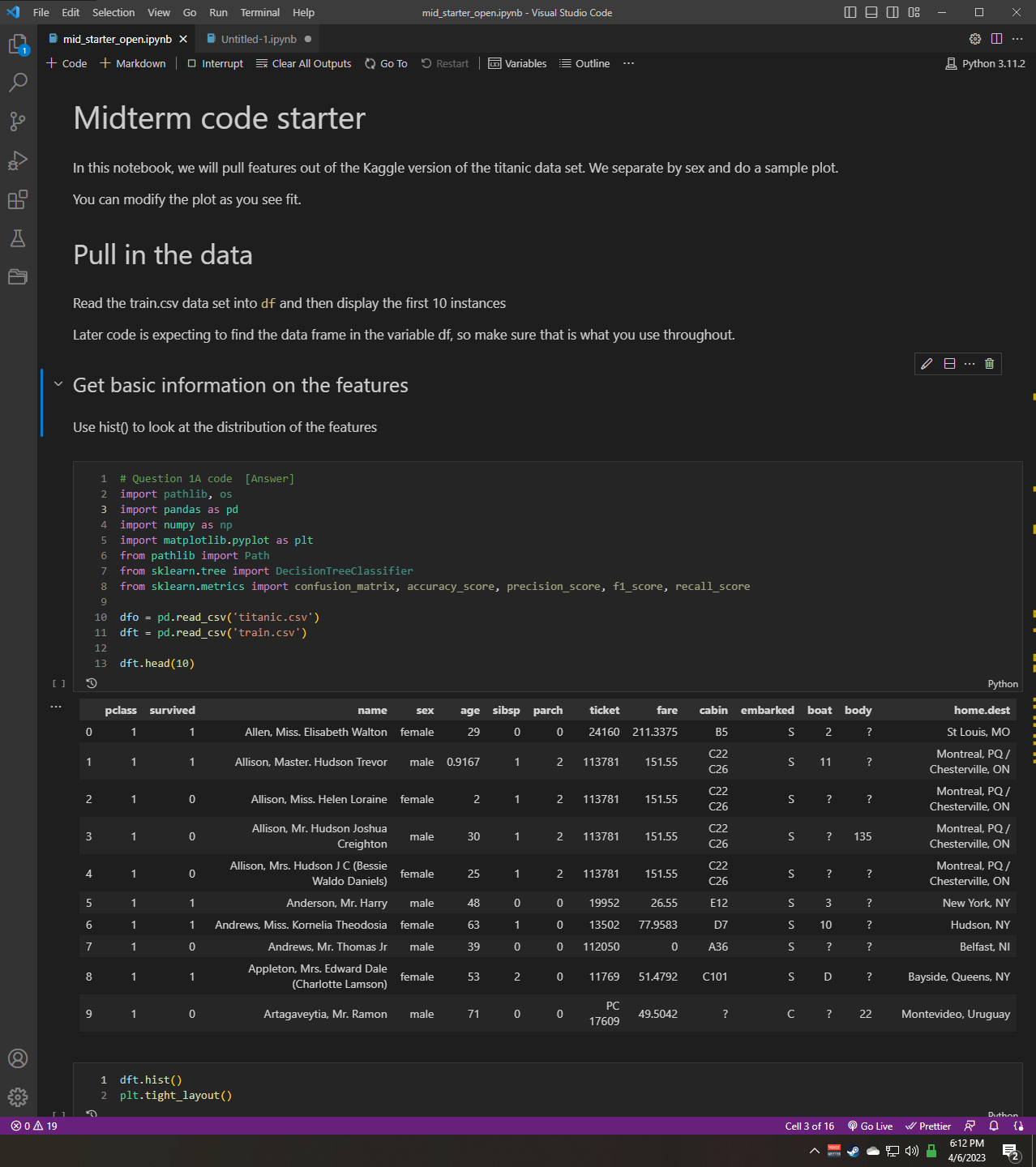
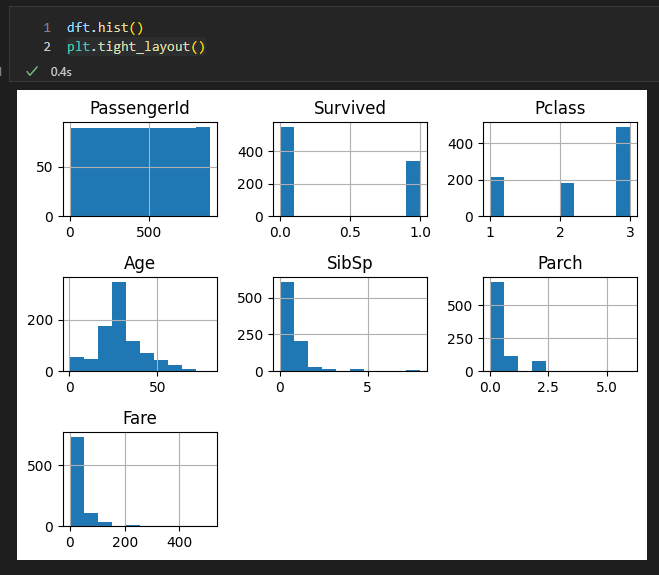
Submission 1-A:

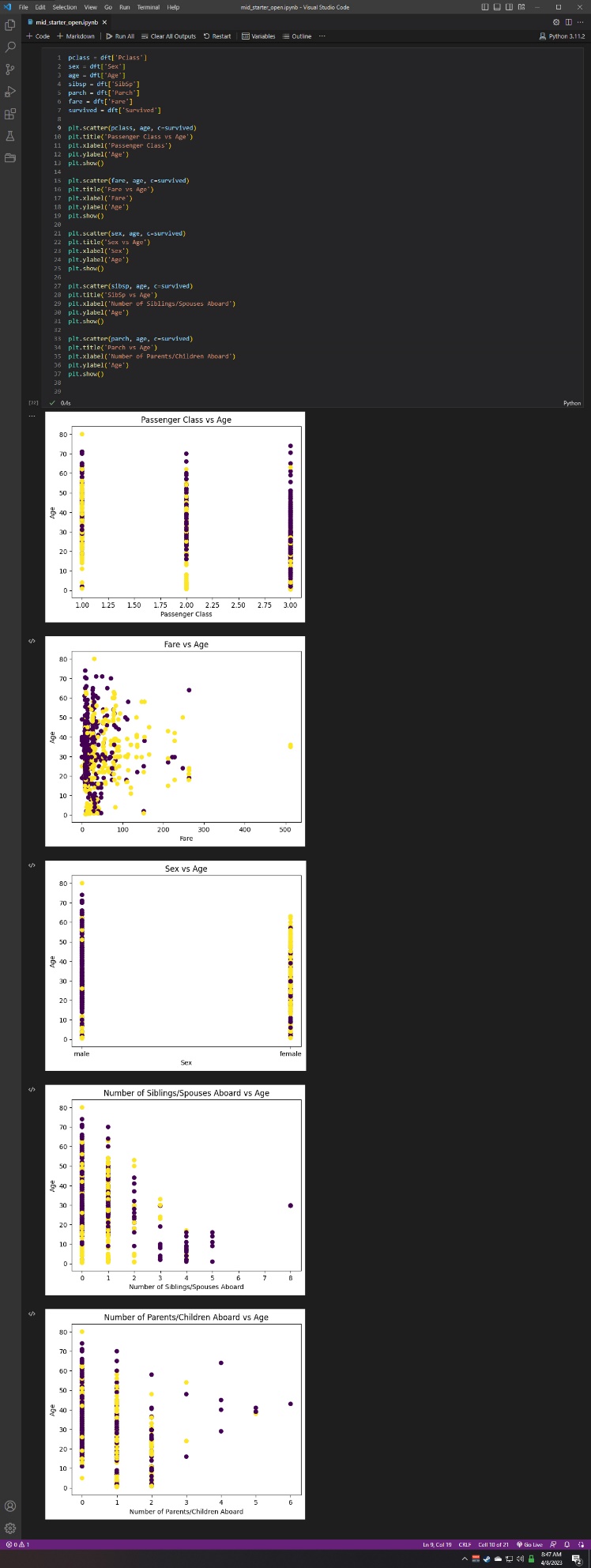


Submission 1-B:

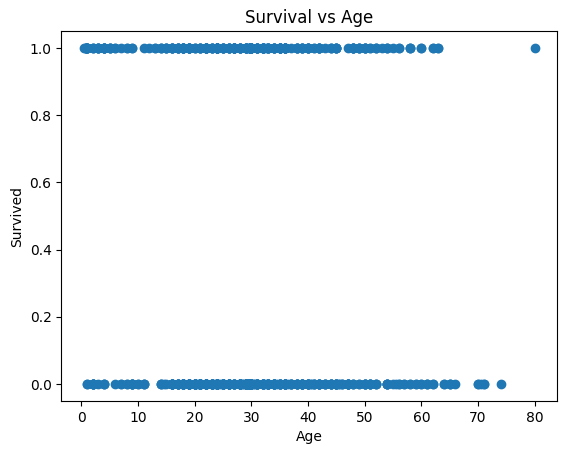
No real anomalies in the charts outside of the age group. The age group is missing values

Submission 1-C:

Home destination and recovered bodies do not have an impact on predicting the causality rate. The life boat amounts would have an impact, but perhaps was left out as there wouldn’t be much of an effect on ratios. I’d argue that cabin could be left out as it’s missing so much data that can’t be filled through means.

Submission 2-A:

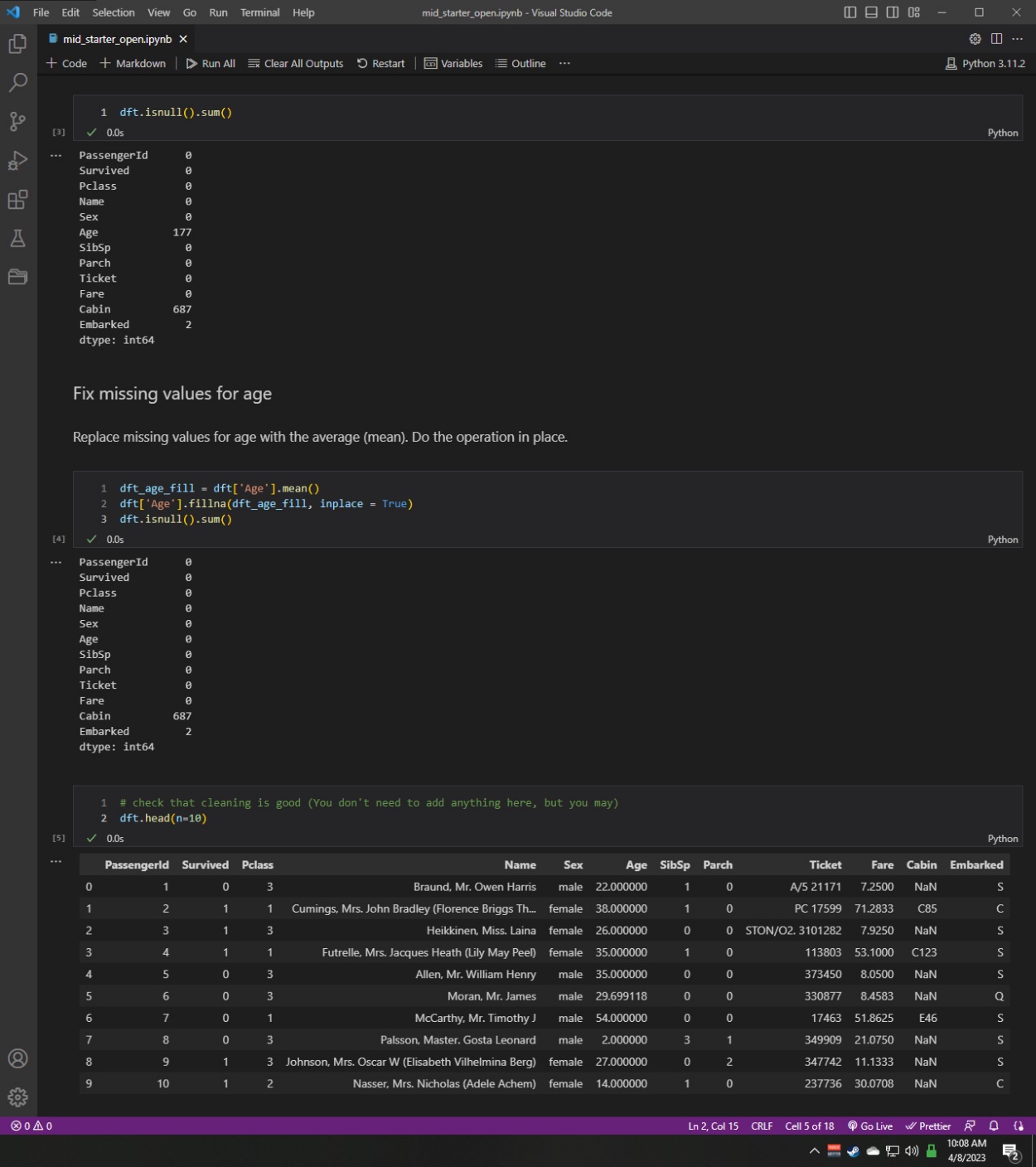
I would like to know positioning. Where they were as the boat was sinking (port, aft, bow, or stern) whenever the Titanic finally split, I’m curious if survival improved depending on the location. Of course, the water would have killed off any survivors that were not pulled under or who weren’t able to find some sort of floatation device to prevent from freezing. I’m simply curious if the location had any part to play in the fatalities. I would assume those located anywhere close to where the ship split would have had higher fatalities. I do not believe that combining any features would produce this data, partially due to a lack of data on where people were when they died. We could make assumptions however on the area they may have perished based on *pclass*, *ticket,* and *cabin*.

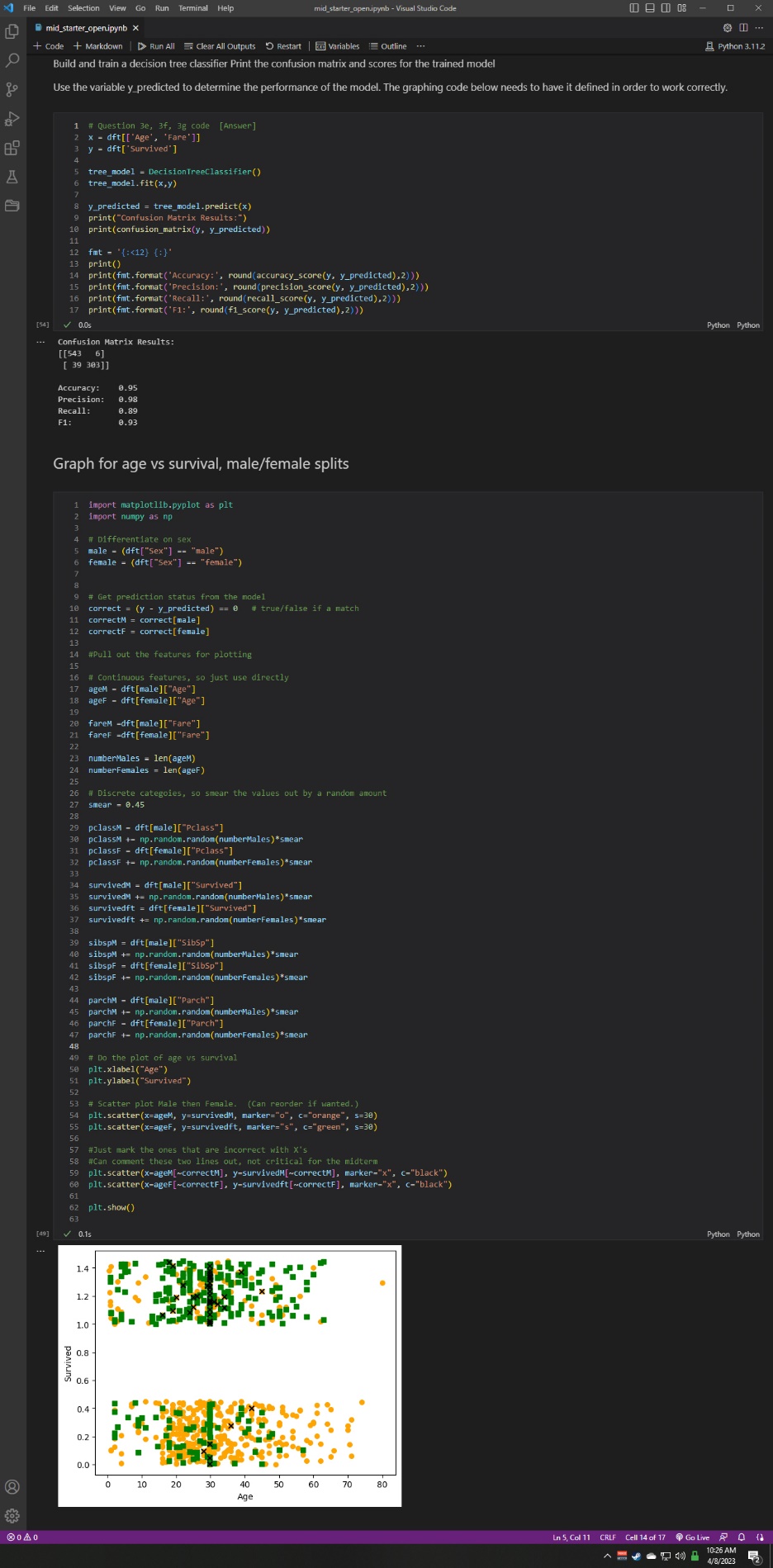
Submission 2-B:

Submission 2-C:

I chose to use age as the primary comparison as it was my assumption that children would get priority towards lifeboats. Based on the results it would appear that the two most important factors for survival was not really age, but being female and in first class. While younger ages did affect survival rates, it wasn’t as clear as male vs female or the class of the passengers. Even though the data is skewed due to the missing values, I believe it represents this point adequately.

Submission 3-B:



Submission 3-E, F, G:

The model performed very well with a high accuracy and F1 score. This tells us that accuracy and recall are relatively high, which is what we want to see. Again, while the data is not perfect due to missing values, by filling the values we have obtained an excellent model which we can accurately state that there was a difference between the survival rates of gender and age when compared to various factors such as class, ticket prices, and family count.