

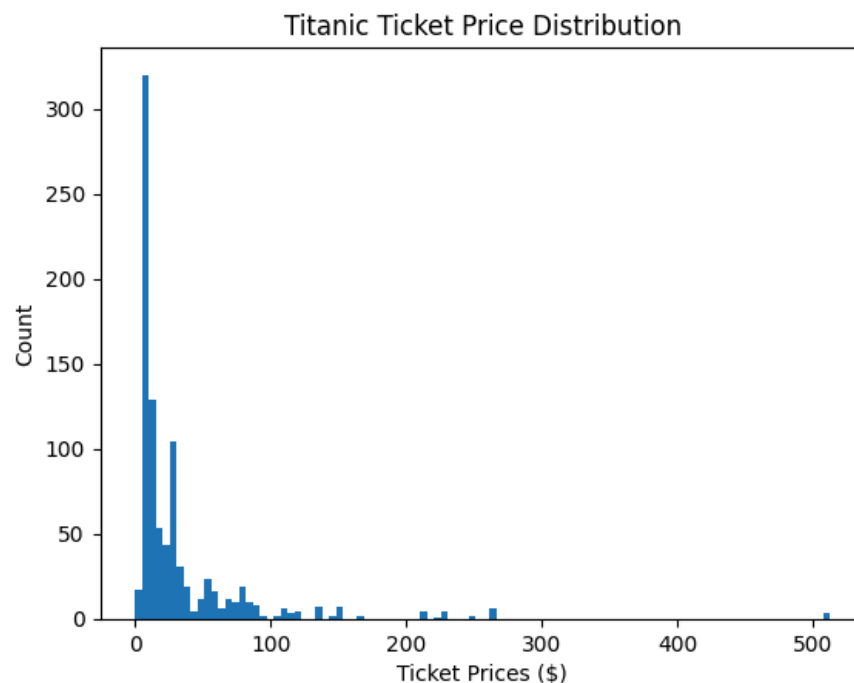
Dennis Houlihan

Homework 5 and 6

### Analysis of Titanic Dataset

The sinking of the Titanic in 1912 remains one of history's most infamous tragedies. However, the detailed records of its passengers offer a unique opportunity for data analysis. This report examines a dataset of 891 individuals aboard the ship to determine what were the key factors that influenced survival rates. More specifically, the role of age, biological sex, ticket price, ticket class, and cabin designation on survival rates will be explored.

The Titanic dataset contains 891 passengers with varying demographics, class distinctions, and survival outcomes. The average ticket price for passengers in this dataset was \$32.20 with a median ticket price of \$14.45. Figure 1 shows the distribution of ticket prices for passengers in the dataset. Most tickets bought were below \$20 but there are smaller bumps around \$30, \$55, and \$80. These could represent different tiers of tickets corresponding to cabin types or positions.



*Figure 1: The distribution of ticket prices for passengers on the Titanic.*

There are 6 types of cabins, labelled with prefixes 'A', 'B', 'C', 'D', 'E', and 'T'. The survival rates were 47%, 74%, 59%, 76%, 75%, and 0%, respectively. There were also 206 survivors and 481 non-survivors (~29% survival rate) without a cabin designation,

so this needs to be accounted for when doing the analysis. Figure 2 shows how many survivors and non-survivors there were for each cabin along with the survival rates. As you can see, despite most passengers not surviving, most of the passengers with a cabin designation did survive. Given this fact, it is difficult to extrapolate the numbers to the total dataset. However, we can still compare survival rates between the cabins. Cabins with prefixes B, D, and E have a significantly higher survival rate than the others, with cabin A and C having much lower survival rates. Cabin T only has one total passenger, so it is difficult to extrapolate any conclusions from it.

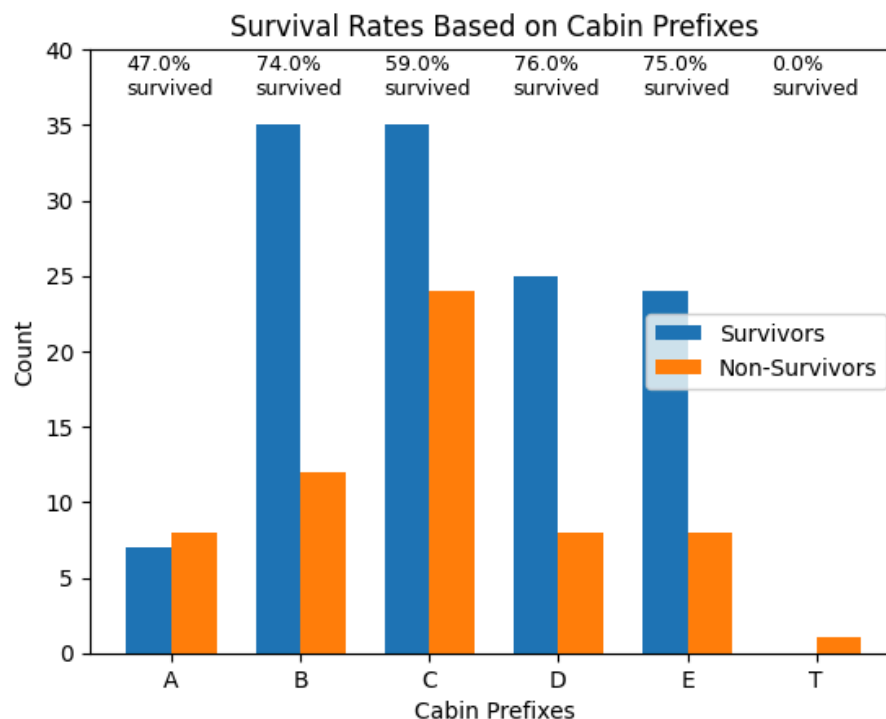


Figure 2: A comparison of survivors and non-survivors for different cabin designations.

Figure 3 goes further to see if there was any correlation between a cabin's average ticket price and the survival rate. Cabins A, D, and E cost less on average while B and C have the highest average ticket prices. There was no clear link between the average ticket prices and survival rate. Cabins D and E had the highest survival rate while also costing the least and cabin B cost the most on average but has a lower survival rate than B, D, and E. However, given that the average ticket price for someone without a cabin designation was \$19.16 and the survival rate was 29%, it is possible that cabins themselves were cost-prohibitive and most people without a cabin designation were in more vulnerable positions on the ship.

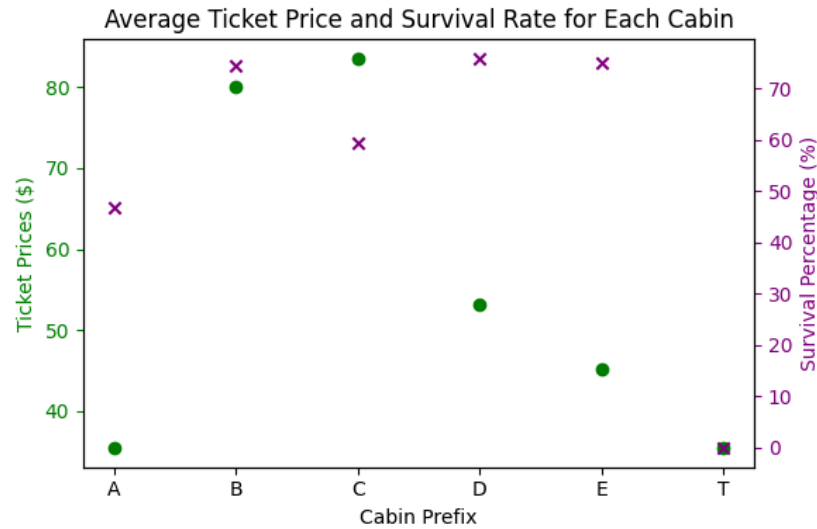


Figure 3: The average ticket price and survival percentage for each cabin designation.

This correlation between survivor rate and ticket prices is further supported by Figure 4. It shows ticket price distributions for both survivors and non-survivors. Most non-survivors had lower-priced tickets, while there were very few non-survivors with tickets above \$100. We can also see the average ticket price for a survivor was \$48.40 and \$22.12 for non-survivors. This could be explained by the fact that more expensive tickets afforded passengers better cabins, access to lifeboats, or access to other resources to escape the flooding water.

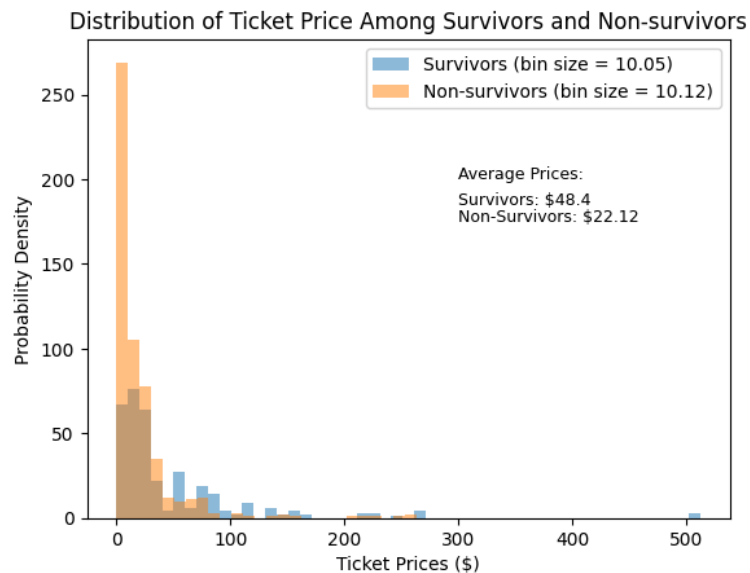


Figure 4: The distribution of ticket prices for both survivors and non-survivors.

Furthermore, we can look at how ticket class impacted the survival rate of the passengers. There are three distinct classes labelled '1', '2', and '3' as seen in Figure 5. Class 3 showed the biggest discrepancy, as there was only about a 24% survival rate. The other classes had a more even split between survivors and non-survivors. This could indicate that passengers with class 3 tickets were housed in more compromising places within the ship, or perhaps did not have as much access to lifeboats or other resources.

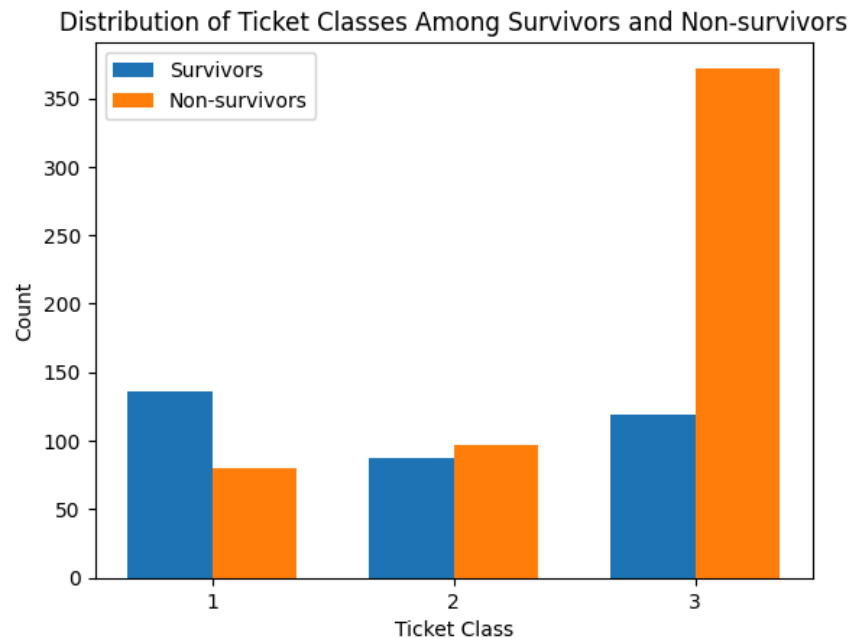
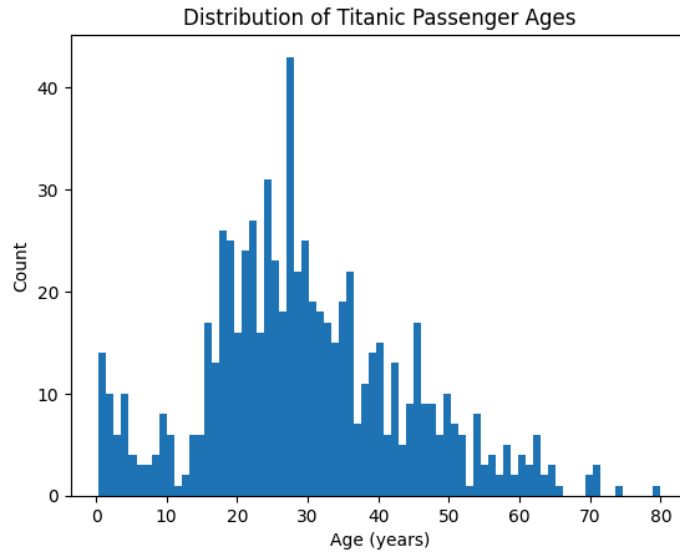


Figure 5: A comparison of the number of survivors and non-survivors for each ticket class.

Exploring the demographics of the dataset could also reveal more factors that influenced survival rate. Of the passengers, 64.8% were male with an average age of 30.7 years while 35.2% were female with an average age of 27.92 years. The average age of survivors was 28.34 years and 30.63 years for non-survivors, which does not show any correlation with age and survival rate. Figure 6 may explain why as it shows the distribution of age for all passengers on the ship. When one looks at the distribution, it makes sense then that the average ages of survivors and non-survivors would not yield a good description since most passengers were between the ages of 20 and 40. But if we dig a bit deeper and plot the distributions of age for survivors and non-survivors, we begin to see a pattern.



*Figure 6: The distribution of age of passengers on the Titanic.*

Figure 7 shows those exact distributions. As you can see, most survivors and non-survivors were between 20 and 40, but we can see some other noticeable features. The older demographics tended to have lower survival rates. This could be because they were less able to maneuver and compete for lifeboat spots, or if they were in the water, their bodies were less likely to be able to withstand the cold. We also see a spike in survival rate with children aged 5 and below. This could be that during evacuations, children were prioritized for lifeboat spots.

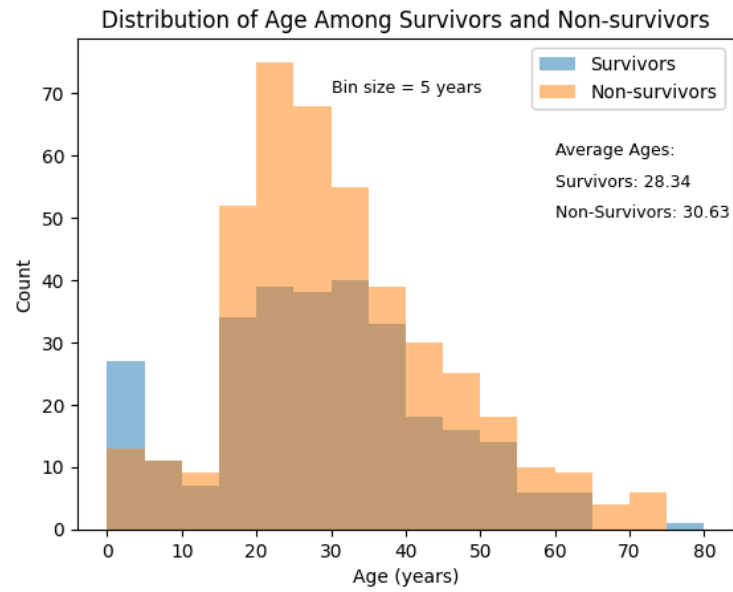


Figure 7: The age distributions for both survivors and non-survivors.

Along with age, the biological sex of the passengers had a big impact on survival rate. Figure 8 shows a histogram that outlines the gender breakup for both survivors and non-survivors. Most survivors were women while most non-survivors were men. This could be because during evacuations, both women and children were prioritized.

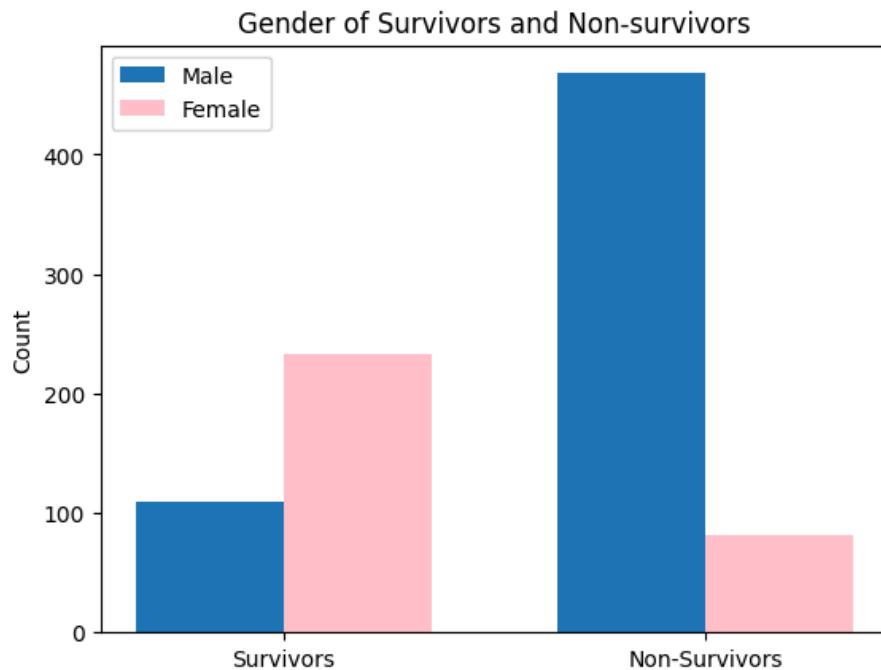


Figure 8: A comparison of the gender for both survivors and non-survivors.

The analysis of the Titanic passenger dataset reveals several key factors that influenced survival during the disaster. Ticket class and cabin designation had a noticeable impact, with higher-class passengers and those assigned cabins, particularly in sections D and E, experiencing higher survival rates. Demographics also played a critical role: women and young children had significantly higher chances of survival, likely due to evacuation priorities, while older passengers and men faced much lower survival rates. Overall, the data underscores how social structure and situational access to resources heavily influenced those who lived and who perished aboard the Titanic.