Titanic Survival Analysis

MIDS INFO W18 Section 2
By: Anjali Shrivastava, Diya Saha, Yui Saeki

Introduction

On April 15, 1912, the RMS Titanic sank after an unfortunate meeting with an iceberg, killing 1502 out of 2224 passengers and crew. The shipwreck shocked the world, and became a call to action for better safety regulations for ships.

Many people died because there were not enough lifeboats for the passengers and the crew. In this report, we will attempt to determine whether gender, class, location and other variables, had an effect on one's chances of survival aboard the Titanic.

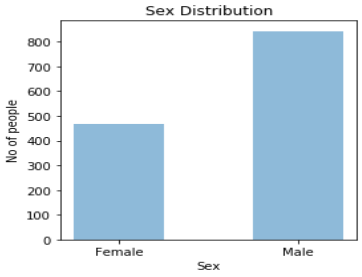
Dataset Structure

The dataset contains 1309 entries (representing 1309 passengers aboard the ship) with 14 columns. These columns are:

- pclass (passenger class)
- survival
- name
- sex
- age
- sibsp (number of siblings and/or spouses aboard)
- parch (number of parents and/or children aboard)
- ticket number
- fare
- cabin
- embarked (port of embarkation)
- boat
- body (body identification number)
- home.dest (home/destination)

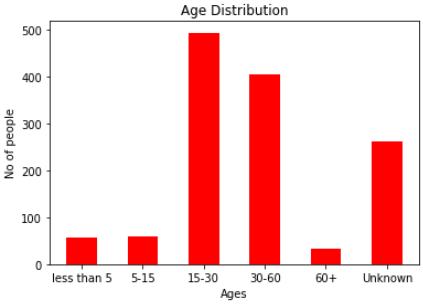
Initial Overview of the Data

In our initial analysis, we found that 500 (38%) of the titanic passengers ultimately survived the disaster, meaning that 809 (62%) of the passengers did not survive. This will be our baseline when comparing whether or not sex or age, for example, had an effect on one's chances of survival.



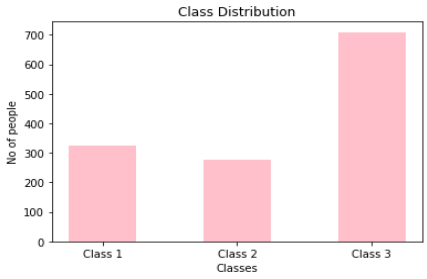
Sex

36% (466) of the Titanic passengers were females and 64% (843) were males.



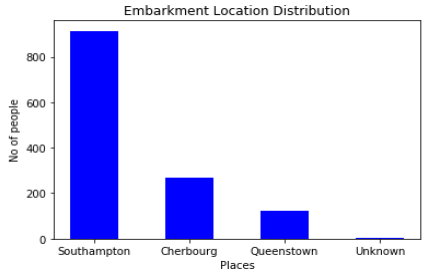
Age

There were 56 (4%) of the passengers were less than five years old, 59 (5%) were between ages of five and fifteen, 494 (38%) were between the ages of fifteen to thirty, 404 (31%) who were between the ages of thirty to sixty and 33 (3%) were above the age of sixty. For 263 (20%) passengers ages were missing in the data.



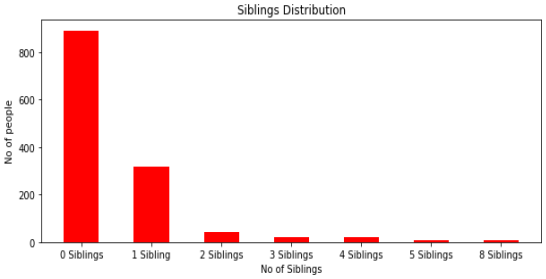
Pclass

323 (25%) of passenger travelled in the first class, 277 (21%) of the passengers travelled in second class and 709 (54%) of the passengers travelled in third class. Most of the passengers were traveling the cheapest class, the third class.



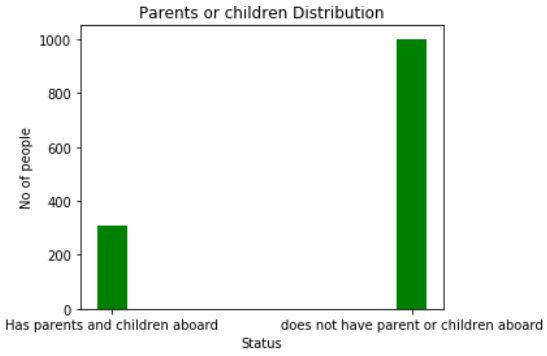
Embarked

Most of the passengers, 914 (70%) boarded the Titanic from Southampton, 270 (21%) embarked from Cherbourg and 123(9%) boarded from Queenstown. The embarkment location for two passengers were missing.



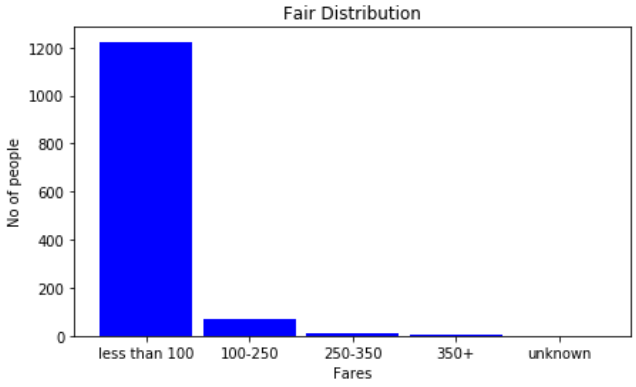
Sibsp

Most of the passengers, 891(68%) of the passengers were travelling alone. 319 (24%) of the passengers had only one co-passenger aboard and the rest 99 (8%) passengers were travelling with 2 or more siblings/spouse.



Parch

307 (23%) passengers were travelling with parents or children and 1002 (77%) of them were travelling without them.



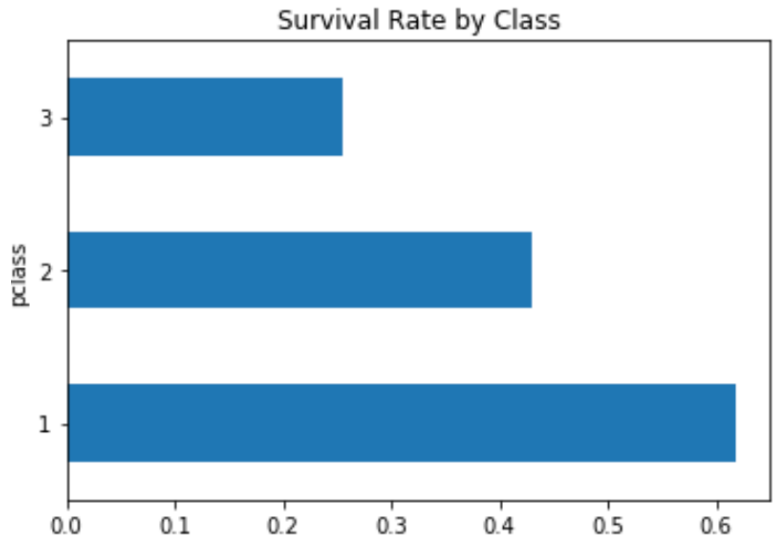
Fare

1224 (94%) of the passenger had bought their ticket under $100, 67 (5%) had paid some amount between $100 and $250, 13 (1%) has paid an amount in between $250 and $350. $ passengers over $350 for their tickets.

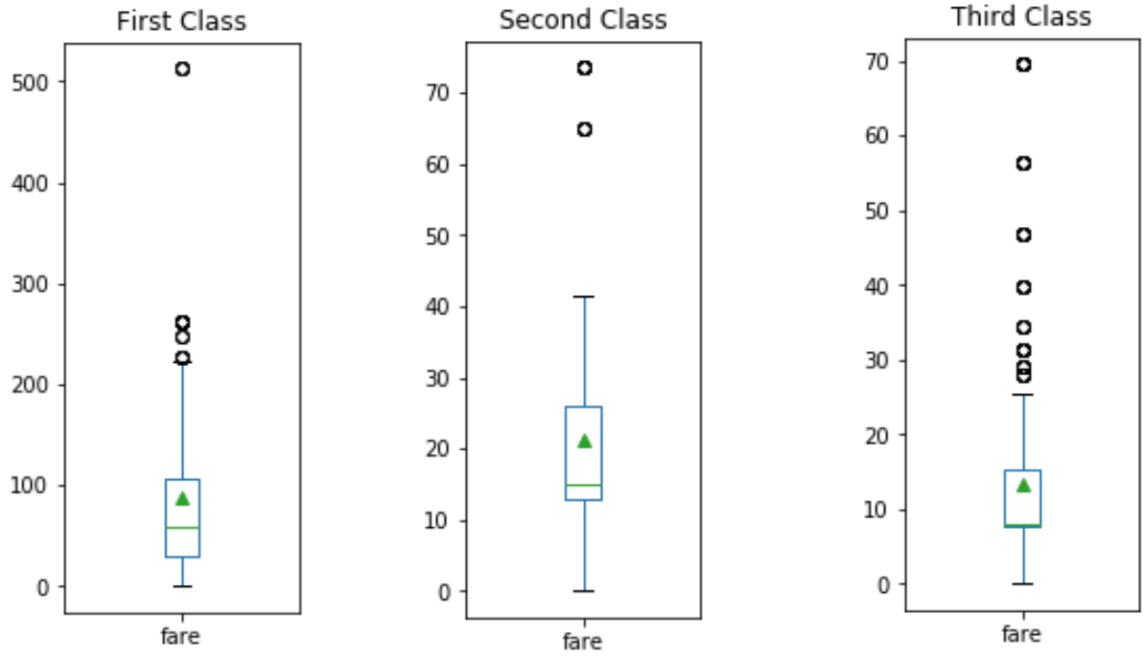
Several of the passengers' ages and one passenger fare was missing from the dataset and was represented by a '?'. These values were replaced with np.NaN before analysis of the dataset.

Analysis

The Relationships Between Class, Fare and Survival



There is a clear correlation between class and survival rate. Passengers in class 1 had the greatest survival rate (61.9%), while those in class 3 had the smallest (25.5%). 43.0% of the passengers in class 2 survived. The data indicates that passengers in higher classes had a greater chance of survival.



Unsurprisingly, the first class tickets were the most expensive, with an average fare of $87.51, and the third class tickets were the least expensive, with an average fare of $13.30. The average fare of a second class ticket was $29.50. Excluding outliers, third class tickets fall into a range of $0-25, second class tickets between a range of $0-45 and first class tickets between a range of $0-220.

The maximum fares for both second and third classes were both below the average fare of the first class ticket, so it's fair to assume that first classes tickets were more expensive, and by extension first class passengers were wealthier, than those in the second and third classes.

There are a number of passengers whose fare amounted to $0, and many of them are in the first and second classes. I suspect that these passengers were employees aboard the Titanic.

A quick sanity check affirms my speculation — I took a random sample of passengers with $0 fares, and researched them — all of them, with the exception of Mr. Jonkheer John George Reuchlin, worked for the ship in some capacity. Reuchlin's ticket was made complementary because of his position at another cruise line.

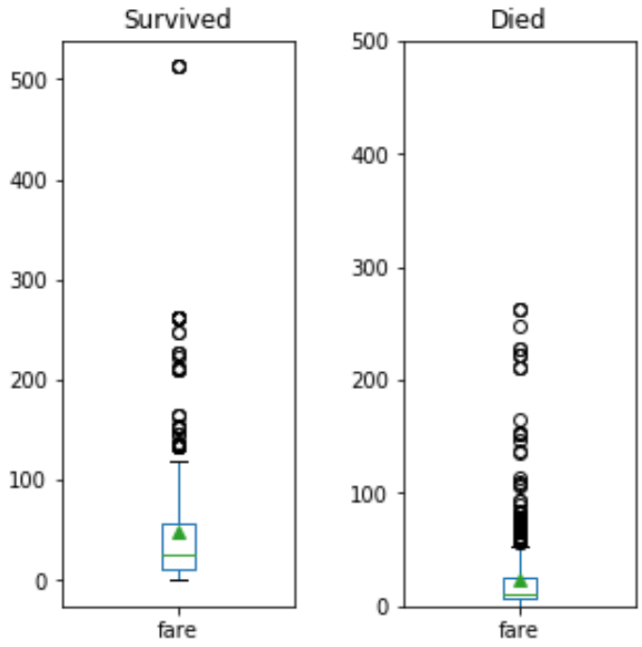
Below is a brief summary of the job descriptions for each individual selected in my random sample:

- Mr. Richard Fry: personal valet to Mr. Joseph Bruce Ismay, the founder of the White Star Line, the shipping company behind the Titanic
- Mr. Anthony Wood 'Archie' Frost: foreman engineer who oversaw the voyage.
- Mr. Lionel Leonard: seaman for American Line
- Mr. Ennis Hastings Watson: apprentice electrician for Titanic
- Mr. Francis 'Frank' Parkes: apprentice plumber for Titanic
- Mr. Robert J Knight: engine fitter for Titanic
- Mr. Alfred Fleming Cunningham: apprentice ship's fitter for Titanic
- Mr. William Campbell: carpenter/'joiner' for Titanic
- Mr. Jonkheer John George Reuchlin: director of the Holland Amerika Line; ticket was made complementary because of his position
- Mr. William Henry Marsh Parr: electrician for Titanic

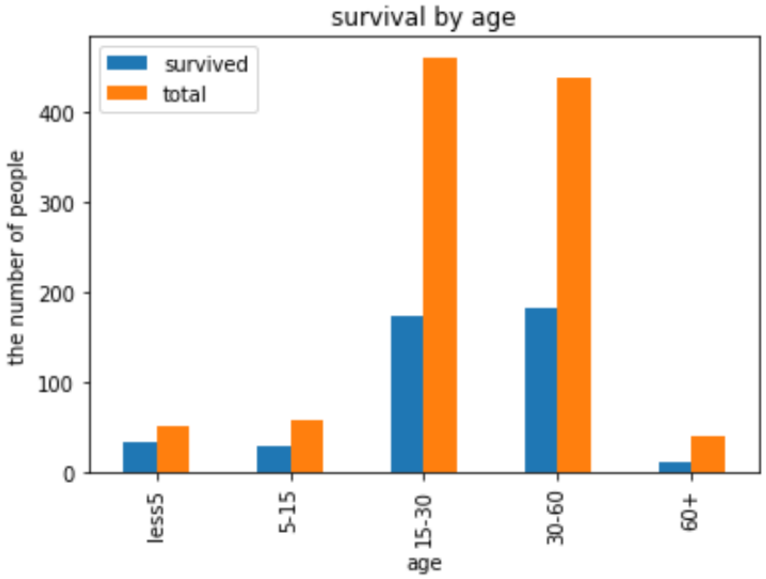
In total, there were 17 passengers with complementary fares, henceforth referred to as "workers" in this report. All of these passengers were male, and 2 of them (11.8%) survived. It is clear that the workers had a much lower survival rate than the rest of the

passengers — the overall survival rate was 38%, and the workers' survival rate is less than a third of that. Perhaps the workers' were preoccupied with saving the passengers, or preventing the ship from sinking, and could not save themselves in time.

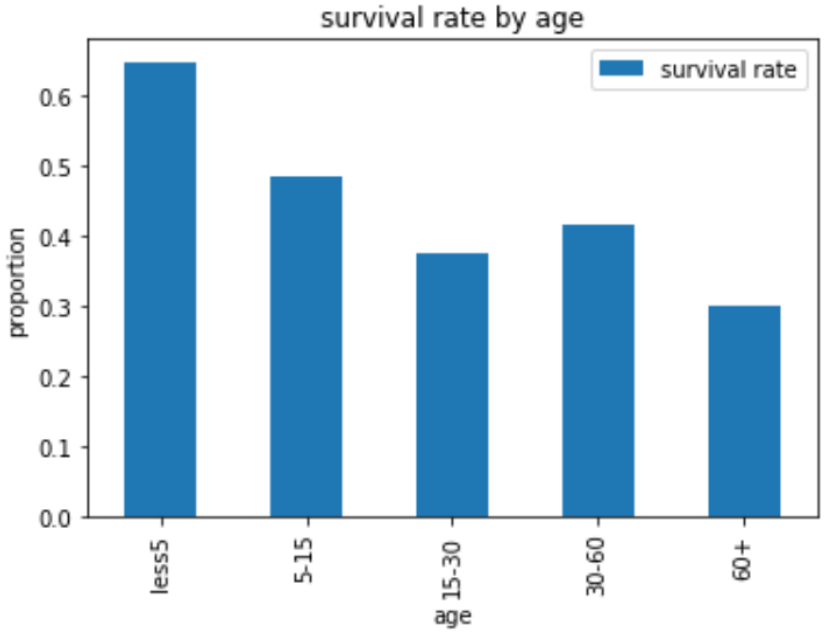
There is a positive, direct correlation between a passenger's fare and their chance of survival, as evidenced by the boxplots to the right. The fares of those who survived were much greater, on average, than those who died. This correlation is likely because those in higher classes had a greater chance of survival, and their tickets were much more expensive on average. Also the workers, who had complimentary tickets, had a much lower chance of survival than other passengers.



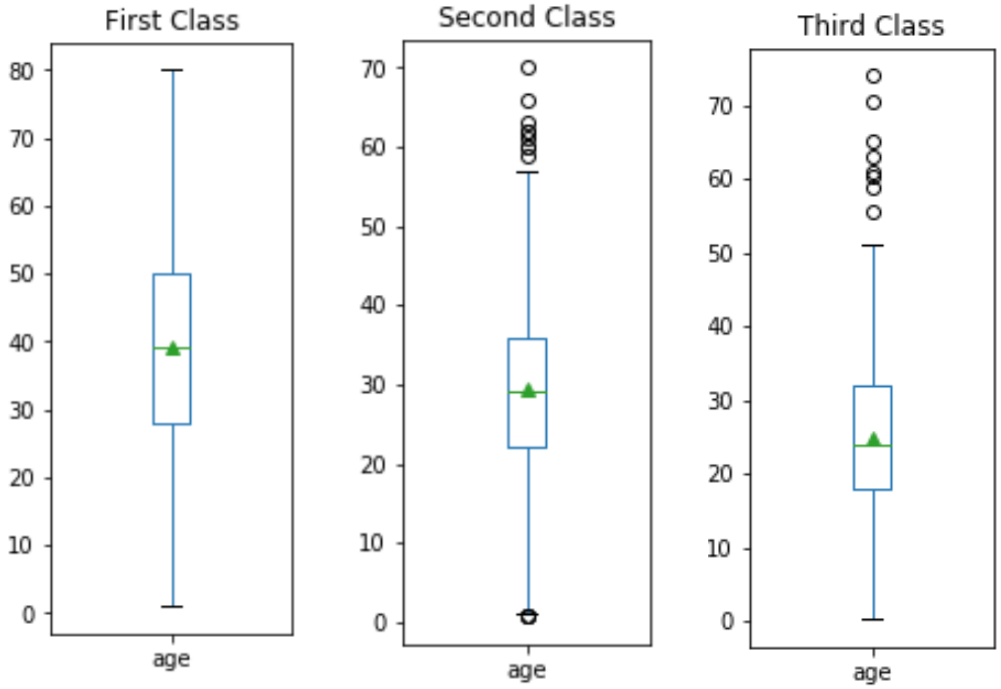
The Relationships Between Age, Gender and Survival



Most of the survivors were aged 15-60 years, according to our analysis. This would make sense, because our initial analysis found that 38% of the passengers were aged 15 to 30 years, and 31% percent of the passengers were aged 30 to 60 years.



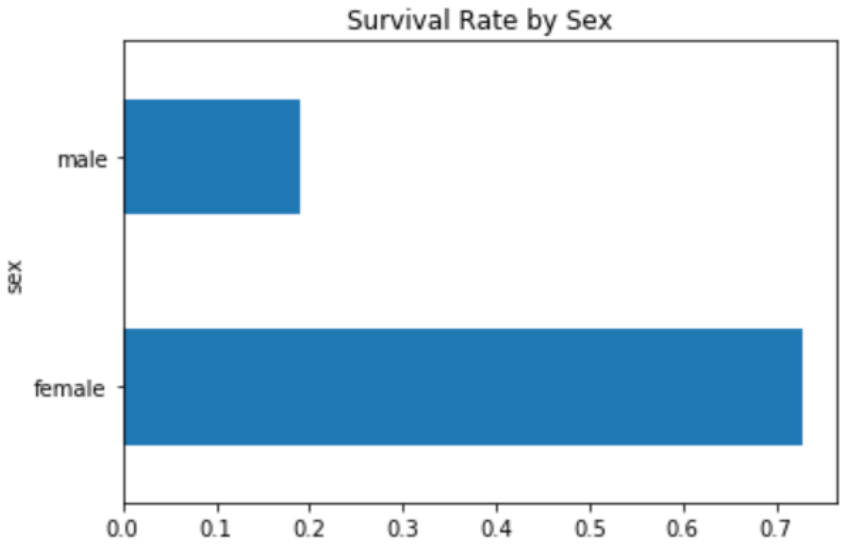
However, passengers under the age of 5 years old had the greatest survival rate (64.7%) and passengers above the age of 60 years had the least (30%). The data indicates that there is a negative correlation between age and survival — that is, the older a passenger was, the less likely the were to survive.



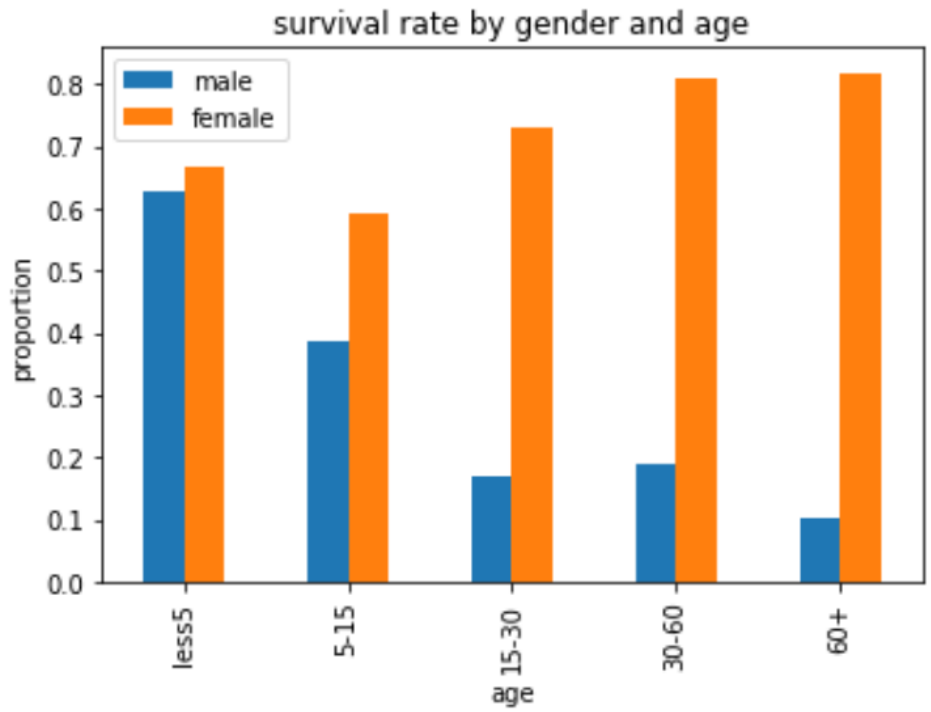
This is interesting because not only did passengers in the first class have a greater chance of surviving, but they were also much older, on average, than passengers in the second and third classes. The mean age of a first class passenger was 39.2 years, and the mean age for the second and third classes were 29.5 and 24.8 years respectively. All classes had relatively similar standard deviations with respect to age (14.54 for first, 13.6 for second and 12.0 for third) so the ages were equally varied.

The second and third classes, while having lower standard deviations than the first, also had more outliers with respect to age. However, for each of the classes, the mean and median age were nearly the same, indicating that none of the datasets are particularly skewed and the distribution is relatively symmetric.

The data seems to indicate that younger passengers were provided lifeboats first, despite the first class passengers being much older than the rest of the passengers.



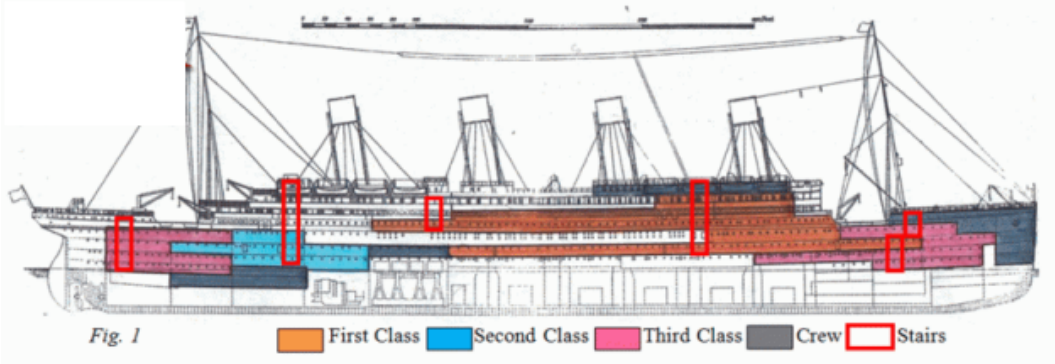
Similarly, the data indicates that women were provided lifeboats first. 72.7% of female passengers survived, which is especially high compared to the 19.1% of male passengers that survived.



The above graph illustrates that women had a much higher survival rate than men for each age group, and the gap widens between the two sexes as the ages increase. The survival rates for men older than 60 years, for example, is a paltry 10.3% compared to to the 81.1% survival rate of women older than 60 years.

It would seem that both male and female passengers under the age of five years had a higher survival rate (63.0% for males and 66.7% for females), and were therefore likely given lifeboats first. This graph affirms our speculation that women and children were given lifeboats first, as even elderly women above 60 years had a high survival rate, and previous graphs indicate that passengers in older age groups had a lower chance of survival.

Cabins (not used)



This picture demonstrates the location of all the classes. The red highlighted area indicates the location of the stair cases. According to the history of the titanic we know that when the ship

broke into two halves, the electricity cut off. The stairway gates were electrically operated. This meant that when the electricity got cut off the gates shut close. This meant that the classes that had access to more staircases might have had a better chance of survival. Unfortunately, we couldn't analyze the cabin distribution vs survival rate. This was because in our data some of passengers had booked more than one cabins. So we cannot really find out which cabin that particular passenger was located during the time of the accident and hence it is hard to analyze the survival rate based on the locations of the cabins.

Conclusions

According to our analysis, there is a clear relationship between class and survival rate. First class passengers had the greatest survival rate, whilst third class passengers had the least. The average price of a first class ticket was much greater than the average fares of a second and third class ticket, so it is fair to conclude that first class passengers were, on average, wealthier than the passengers travelling in the second and third classes.

Our analysis also found that passengers with complimentary tickets had an extremely low survival rate. This is likely because many of these passengers were workers who were preoccupied with saving other passengers and preventing the RMS Titanic from sinking. We found that the passengers who survived paid a higher fare, on average, than those who died, which makes sense when you take into account the relationship between class and survival, and the survival rate of those with complimentary tickets.

We also found that there is a negative correlation between age and survival — that is, the older a passenger is, the less likely they were to survive. This is interesting because first class passengers were much older than other passengers, on average.

Finally, women were much likelier to survive than men, according to our analysis. This is likely because women and children were given lifeboats first, which accounts for the difference in survival rates by age group as well.