Titanic Crew Member Survival Data

Name: Christopher Anderson
Semester: DSC680-T301 (2213-1) Winter 2020-2021
Github Portfolio URL: https://chrisinoakland.github.io

Project 1 Report

Introduction

The tragedy of the Titanic disaster has captivated me since I was a young boy. When I was twelve, I used to draw the ship using large pieces of paper taped together. After Titanic was discovered on the ocean floor by Dr. Robert Ballard and his crew in late July of 1985, I collected almost everything I could find on it: (newspaper clippings, magazines, books, etc.) And yes, I absolutely loved the movie “Titanic” by James Cameron, not necessarily for the story line of Jack and Rose, but because of the way Cameron painstaking re-created what the ship looked like — it was simply stunning to see something that I had imagined and only seen black and white photos of on full display in full color and vivid and intricate detail! The next great bit for me as a Titanic history buff was taking my family to see artifacts from the ship on display when an exhibit came to Omaha. While I will not delve into the ethics or legality of plucking items from a shipping disaster from the ocean floor, or share my thoughts on who actually owns the items and their rights to salvage them, I will say that seeing those personal items and especially a large chunk of steel from the ship — including port holes — was very moving.

It is with this background that I chose to continue the exploration of Titanic survivor data, by doing analysis on not the passengers that survived the disaster, but the actual crew members employed by White Star Line who survived.

Data Sources

There is a wide and vast amount of information on the *Titanic* available, most of it public and freely available to consume via books, newspapers, magazines, and of course, the Internet. The sinking of this ship on the evening of April 14-15, 1912 was one the largest tragedies of its day, and additionally, it was — at the time — the largest loss of life in maritime history. There

were over 1,500 souls who perished on the sinking of *Titanic*, including men, women, children from all classes and walks of life; however, for this project, I am keying in on those who worked aboard the ship. For information specifically on *Titanic's* crew, I will be using data from the Titanic People Database as created by Encyclopedia Titanica. More information can be found at https://www.encyclopedia-titanica.org. The site has an online and downloadable listing of records of passengers and crew, survivors, etc. I'm a bit of a *Titanic* junkie, so I already had a subscription to this site which allows for more granular downloading of data in various formats. Another source of data is a rather well-known set from Kaggle and it is free to download from https://www.kaggle.com/c/titanic/data.

Data Input

After downloading and looking through the dataset from Encyclopedia Titanica, it was determined that I would not require any additional data from the Kaggle dataset, as I already had everything I looking for. Here's a look at the initial steps of data input and verification.

```
import pandas as pd
import yellowbrick

# Load crew data into a dataframe
list = "titanic_crew_original.csv"
crew = pd.read_csv(list)

# Check the dimension of the table:
print("The dimension of the table is: ", crew.shape)
## The dimension of the table is:  (1126, 16)
```

It looks like there are 1,126 rows (crew members), and 16 columns (variables).

Data Cleanup: Overview

In the data science realm, a lot of the work done is very similar to any other field or line of work — preparation is key. Just as you shouldn't paint a wall, stain a dresser, or lay out a walkway made of brick pavers as the first line of action, you must do the less glamorous but hard work that is required before getting to those "fun" steps. You need to fill any holes in that wall, and perhaps give it a coat of primer. That dresser needs the old stain removed first and/or sanded down, and that walkway is not going to be safe unless you level and pack the ground and fill it with crushed stone to build a solid foundation to set the pavers on. This "grunt" work is the

foundation for what comes next. As it is in data science. Ensure the data is rock solid first, and they you get to enjoy the fruits of your labor with the data exploration and visualization!

The default data file containing all of the crew data for Titanic's maiden voyage as downloaded from https://www.encyclopedia-titanica.org was a tremendous resource for this analysis, and it included a total of sixteen columns of data. After taking an initial look at the data after input, I then began the data analysis, starting with data cleanup. I established the cleanup process with the following four steps:

1. Verify there is no missing data and that the data there is accurate
2. Remove unnecessary data
3. Fix any data formatting issues (if any)
4. Identify and remove any duplicate entries

Data Cleanup: Verify Data

This step was primarily accomplished upon loading the data and looking at all of the column headings and information in the rows. Once I could trust the information downloaded contained the things I wanted, I then created different one-off datasets based upon the category in the `Class/Dept` column and also the `Occupation` column to verify that the total number of crew members that worked in these various departments and parts of the ship were as expected when comparing to other resources that listed this type of information. This is where the verification step proved it's worth. [6] According to multiple references, there were 891 crew members aboard *Titanic* bound for New York, but according to the imported dataset, there are 1,126 rows (crew members).

Why the discrepancy? In attempt to understand the difference, I noticed that some rows had no entry under the `Deceased` column which didn't make sense at first, and then I realized… those crew members were ones that didn't actually make the trans-Atlantic voyage. When *Titanic* departed Southampton, England at noon on April 10th, 1912, bound for New York, it made a few stops along the way, one in Cherbourgh, France that same evening, and then again later that same night in Queenstown, Ireland. The ship actually stayed docked in Queenstown that night and didn't leave again until the next day, April 11th at around 1:30pm. These stops were made for three primary reasons: to take on more passengers, to get and load more supplies onto the ship, and to move mail. Steamships of that era required a lot of coal to burn

to power the engines, turbines, and propellors to move the ship through the water, and also to generate electricity to light the ship. How much is a lot? *Titanic's* massive boilers required 825 tons of coal *per day*.[4] The moving mail portion might come as a surprise to some. What does *Titanic* have to do with mail? The RMS of *RMS Titanic* stands for Royal Mail Steamer, a designation that the ship was used as a traveling post office and mail transportation device for not only passengers and crew who could send postcards, mail, and packages along their journey, but also under contract to the British Royal Mail. [1] As luck would have it, for 236 souls employed by White Star Line, their call of duty had them departing *Titanic* at one of these two stops in Cherbourgh or Queenstown before it left on the tragic, fateful trip to New York City. Actually, that number is more accurate at 228. Wait, what? Why another discrepancy and what about those other eight crew members? Well, it seems that just like in any other job, working for White Star Line had some folks that just weren't cut out for it for the long haul. That's right — eight people that had worked for the company and were assigned to the maiden voyage of *Titanic* ultimately made the smartest decision of their life and quit (in seafaring terms, it is referred to as deserting their post) before the ship left for New York City. In researching this information with more depth, every deserter was a male assigned to the Engineering crew, and seven of the men held the title of "Fireman", and one was a "Trimmer". [4] Both of these roles were some of the most labor intense, and involved the dirty, extremely grueling work of working with coal on a coal-fired, steam-powered ship — either loading, moving, and transporting it, as was a Trimmer's duty, or by shoveling it into the boilers, which was a Firemen's role.

After all of this research, and thereby realizing how the 236 crew member number difference came to be, I proceeded to the next step — getting rid of unnecessary data.

Data Cleanup: Remove Unnecessary Data

As the default downloaded list, the crew members of *Titanic's* maiden voyage contained a total of 1,126 rows, each representing a life aboard the ship working hard to make the trip safe, successful, and enjoyable for the passengers. As discussed in the previous section, this data isn't going to be accurate for the analysis of who survived or not, since 236 crew members actually departed *Titanic* before it took off for New York. While the downloaded dataset featured the expected data with column headers such as `Name`, `Gender`, `Class/Dept`, and `Occupation`, there were also numerous columns that would not be beneficial nor required for this analysis. The columns that I removed from the dataset include `Born, Died, Boat`, `Body`, and URL. Additionally, there are three columns with empty data that won't be used:

`Ticket`, `Fare`, and `Cabin`. Once I had things pared down, this is what the dataset looked like after getting rid of unneeded data:

```
# Remove rows that don't contain survival data:
crew = crew[crew.Survived.notnull()]

# Remove the columns from the df we aren't using:
del crew['Born']
del crew['Died']
del crew['Boat']
del crew['Body']
del crew['URL']
del crew['Ticket']
del crew['Fare']
del crew['Cabin']

# Check the dimension of the table:
print("The dimension of the table is: ", crew.shape)

# Check the results:
## The dimension of the table is:  (891, 8)
print(crew.head(5))
##                          Name   Age  ... Survived       Nationality
## 0     ABBOTT, Mr Ernest Owen  21.0  ...      0.0           English
## 1  ABRAMS, Mr William Thomas  34.0  ...      0.0           English
## 2      ADAMS, Mr Robert John  26.0  ...      0.0           English
## 3    AHIER, Mr Percy Snowden  20.0  ...      0.0  Channel Islander
## 4  AKERMAN, Mr Albert Edward  31.0  ...      0.0           English
##
## [5 rows x 8 columns]
```

By cleaning the data to remove all unneeded information, we get a dataframe with 891 rows and 8 columns (variables), which will aid in trying to find any clues to the research questions; however, before we start asking those research questions, let's get a closer look into what our dataset now contains:

```
# What type of variables are in the table:
print("Describe Data")
## Describe Data
print(crew.describe())
##               Age    Survived
## count  891.000000  891.000000
## mean    31.737374    0.237935
```

```
## std      9.101830    0.426058
## min     14.000000    0.000000
## 25%     25.000000    0.000000
## 50%     31.000000    0.000000
## 75%     38.000000    0.000000
## max     63.000000    1.000000
print("Summarized Data")
## Summarized Data
print(crew.describe(include=['O']))
##                     Name Gender  ... Occupation Nationality
## count                891    891  ...        891         888
## unique               888      2  ...        232          27
## top     COLLINS, Mr John   Male  ...    Fireman     English
## freq                   2    868  ...        161         703
##
## [4 rows x 6 columns]
```

To help understand the meaning of the column headers (the variables), I created the following table:

Name	Definition	Key
Name	Name of crew member	
Age	Age of crew member in years	
Gender	Gender	
Class/Dept	Department of the ship crew member worked in	
Joined	The stop (city) at which the crew member boarded *Titanic*	
Occupation	More detail about the specific role of the crew member	
Survived	Survival status	0 = No; 1 = Yes
Nationality	Country the crew member was from	

Data Cleanup: Fix Data Formatting Issues

After verifying and removing all unnecessary *Titanic* crew member data, the next stage of my cleanup process entailed ensuring that all of the data was formatted correctly. From this standpoint, everything checked out and I did not need to take any corrective action here.

Data Cleanup: Remove Duplicate Entries

After completing the three previous steps, it was time to do the last portion of my cleanup process — looking for and removing any duplicate entries. Let's look at the code to do that:

```
# Find and remove any duplicate entries
crew_duplicates = crew.duplicated()
print('Number of duplicate entries found: {}'.format(crew_duplicates.sum()))
## Number of duplicate entries found: 0
```

It looks like everything checks out, and that we have no duplicates; however, just to be sure, I want to run another test:

```
# Ensuring that the test for duplicates is accurate:
duplicate_test = crew.duplicated('Survived').head()
print('Number of entries with duplicate Survived status in top entries are
{}'.format(duplicate_test.sum()))
## Number of entries with duplicate Survived status in top entries are 4
crew.head()
##                         Name   Age  ... Survived       Nationality
## 0     ABBOTT, Mr Ernest Owen  21.0  ...      0.0           English
## 1  ABRAMS, Mr William Thomas  34.0  ...      0.0           English
## 2      ADAMS, Mr Robert John  26.0  ...      0.0           English
## 3    AHIER, Mr Percy Snowden  20.0  ...      0.0  Channel Islander
## 4  AKERMAN, Mr Albert Edward  31.0  ...      0.0           English
##
```

Okay, I have verified that when I run a duplicates check on the data using the information in the `Survived` column, I show there are duplicates. It never hurts to double-check!

Research Questions

Now that I have completely ran through the data cleanup process and I am comfortable and confident in the data being accurate, error-free, being exactly what I need, and that there are no duplicates, it is time to begin looking into the ways that I can answer the research questions I had in mind when I chose to take on this project.

The sinking of *Titanic* was certainly a tragedy and I specifically want to know about the outcome of *Titanic's* crew. Here are my research questions:

1. What was the survival rate for *Titanic* crew members?
2. What was the survival rate of the crew members compared to the passengers on the ship?
3. How many of the crew members were men and how many were women and did that impact their survival status?
4. In which parts of the ship did the crew members work and did that have any impact on their survival?
5. At which port did the crew members board the ship and did that have any impact on their survival status?

I will be using Python and R to do the data analysis an attempt to find answers to these questions.

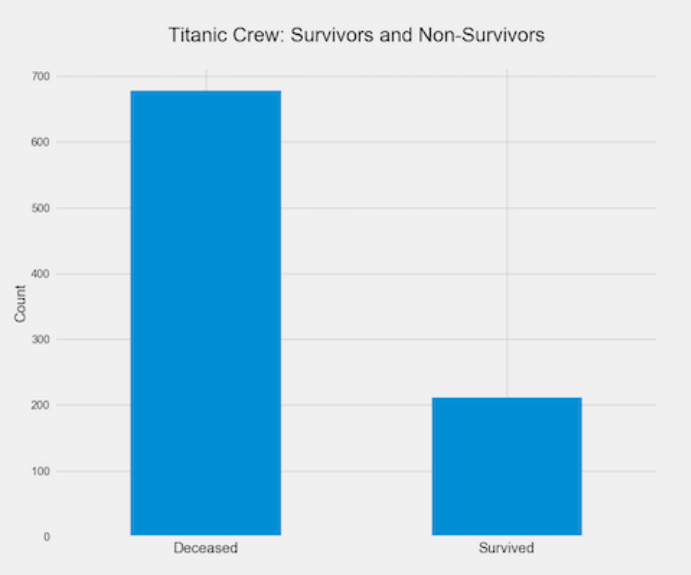
Data Exploration and Analysis and Data Visualization

It is now time to start digging into the data questions I had about the crew members of *Titanic*, and I will use this section by framing out the questions I had and begin the exploration and analysis of the data to answer those questions using R and Python. In particular I will generate visualizations that will look to provide a much more detailed view into the crew data. One of the first questions I had regarding the crew members of *Titanic* is what was their survival rate?

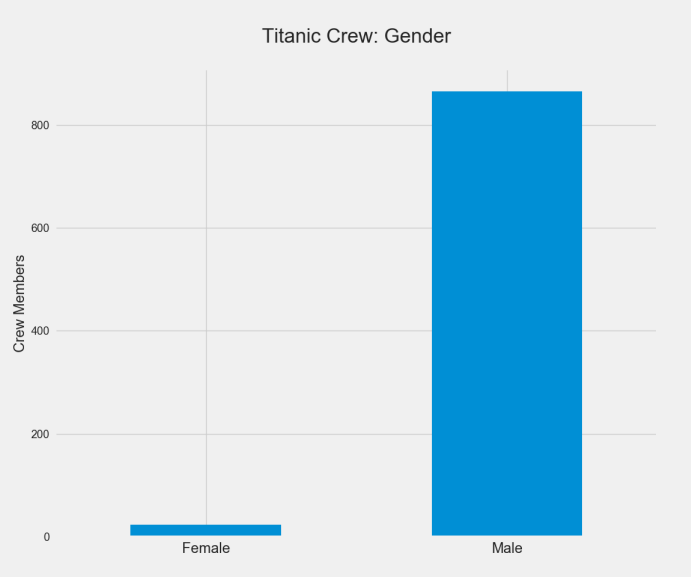
The story is well known that during the sinking of *Titanic*, more than 1,500 people — 1,517 to be exact — lost their lives. [3] The really unfortunate part of this tragedy is that it shouldn't have happened — for numerous reasons — most especially that the ship didn't have enough lifeboats for all on board. Titanic was equipped to carry 64 lifeboats; however, the ship only actually set sail with 20 — and four of those were collapsible, a smaller, more easily storable type of boat. The total number of souls that the lifeboats could carry was 1,178 passengers and crew, but that number was still more than required by the 1883 Merchant Shipping Act. Still, just over 700 made it aboard lifeboats. In 1912, the tradition for loading lifeboats during an emergency was women and children first, and unfortunately, this tradition often caused time delays in filling the lifeboats as the women and children were singled out for priority in lifeboat placement, which often led to lifeboats being launched only half full, if that, amidst the chaos of a sinking ship. This was certainly the case with *Titanic* on that cold, April night with freezing temperatures on the Atlantic Ocean. [11]

It's now time to tell the data story of *Titanic's* crew members. In total, there were 1,517 souls lost during this tragedy. From the total of 891 crew members, 679 did not survive, a survival rate of around 24%. From the ship's passenger perspective, there were a total of 1,352 passengers aboard and 500 survived, a survival rate of 40%. The data tells us that the *Titanic* crew members had less of a chance of survival than the passengers did.

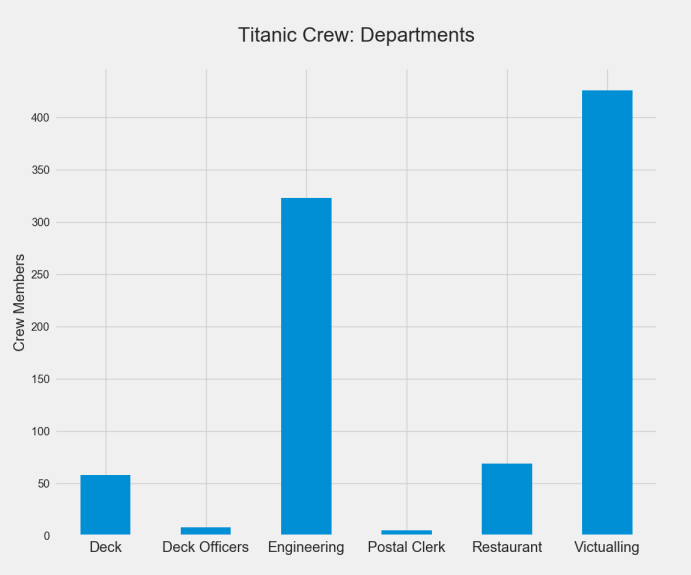
Let's take a closer look at our *Titanic* crew survival data using bar charts:



To answer the research question of how many of the crew members were men and how many were women, let's again turn to a bar chart:



Looking at the data, the number of male crew members aboard the Titanic's maiden voyage far outnumbered the women. [3] In fact, of the 891 crew members, only 23 were women. Why such a large discrepancy? For that we should look more into the different departments that worked aboard the ship. Here's that data:



This information shows that most of the crew members of *Titanic* were in the victualling department; this is the department that provides all the services for the occupants of the ship: food, housekeeping, laundry, room service, etc. The next highest amount of crew members aboard *Titanic* were its engineers. The engineers were responsible for keeping the engines, generators, and other mechanical equipment on the ship running. They were also the highest paid members of the crew (not counting the officers) and had the education and technical expertise to operate, maintain, and repair the engineering plant.
Why is there such a disparity in the number of men vs women that worked aboard *Titanic*? Mostly likely, this was due to the social norms of the early 1900s, and the fact that the majority of the employees worked deep in the depths of the ship producing its power using coal.

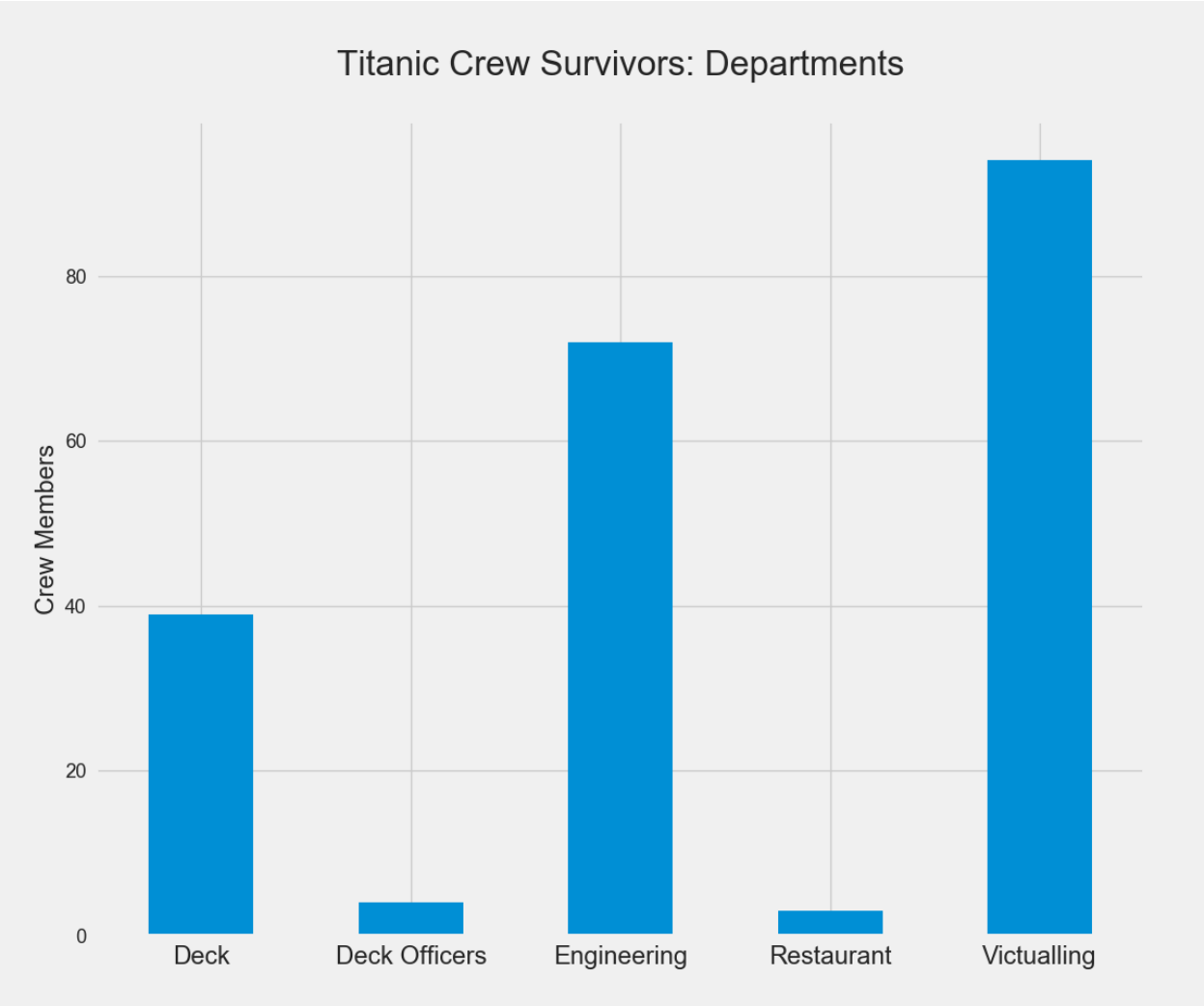
Survivor Data

Now let's look at some of the same data showing just survivors of the sinking. First we'll create the new dataframe of just the survivors:

```
# Create a new dataframe of just survivors:
survivors = crew.loc[crew['Survived'] == 1]
```

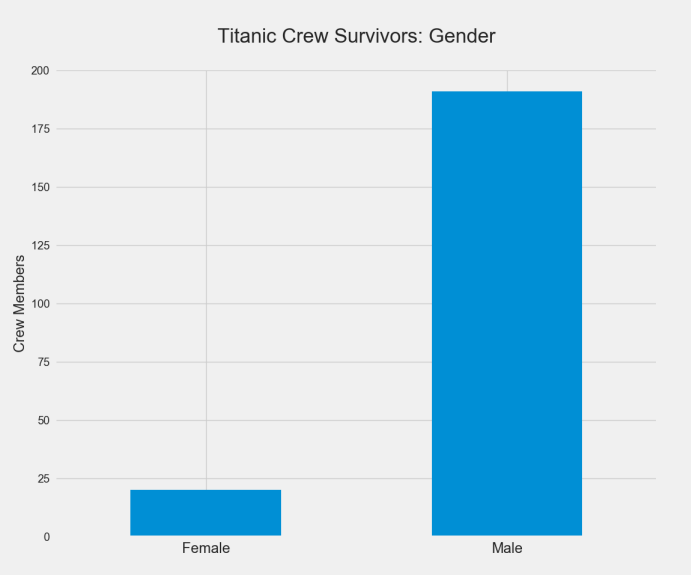
```
# Check the dimension of the table:
print("The dimension of the table is: ", survivors.shape)

# Get a look at the first five rows of the survivor df:
print(survivors.head(5))
```

Per the chart in our initial look at all of the crew member data, our new `survivors` dataframe verifies that only 212 out of the 891 crew members of *Titanic* survived the sinking. Now that the data is isolated to show just survivors, let's look at some of the same bar charts as before showing survivor-related data only:



This chart looks a bit different that the listing of all crew members. You can see that all five of the postal clerks aboard the ship did not survive, and that while a good majority of the deck crew did, many of the members of the engineering and victualling departments did not.

One final chart, this showing the crew survivors by gender:



Of the 23 crew members aboard *Titanic* that were women, 20 of them survived the disaster, right around 87%, whereas of the 868 men, 192 of them survived, a rate of 22%.

Conclusion

There were many people that died when RMS *Titanic* sank on its maiden voyage in April 1912, including those who were crew members, employed by the White Star Line to service the passengers on the ship. After exploring data on the crew members to look for any trends in the data from those that survived the disaster, some items stood out.

Recall there were five research questions that I wanted to answer when looking through this data:

1. What was the survival rate for *Titanic* crew members?
2. What was the survival rate of the crew members compared to the passengers on the ship?
3. How many of the crew members were men and how many were women and did that impact their survival status?
4. In which parts of the ship did the crew members work and did that have any impact on their survival?

5. At which port did the crew members board the ship and did that have any impact on their survival status?

While there was not much correlation to the port (city) in which the crew members boarded *Titanic*, I was able to glean information on the other questions, notably that crew members' survival rate (24%) was less than that of the passengers (40%). Additionally, there were only 23 women crew members out of the 891 total, and 20 of them survived, an 87% survival rate. The last research question regarding crew department and survival rate shows that those with less seniority and status as far as department goes were most likely to perish in the disaster. [7] A fair amount of the ship's deck crew survived, most likely because they were put in charge of manning the lifeboats, and also assigned to them as rowers as they were lowered to the ocean during the calamity. Unfortunately, the majority of the workers who were in lower parts of the ship — especially right above the hull (bottom) where the boilers were located [10] did not survive. All five of the postal clerks aboard the ship perished, most likely due to the mailroom being one of the first rooms impacted by the iceberg strike and first to become flooded with freezing sea water. [8]

References and Annotations:

1. Harrell, Frank E. (2002, December 27). "Titanic Data" from http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/titanic.html.
2. Zuckerman, Arthur. (2020, May 24). "57 Titanic Statistics: Deaths, Passengers & Survivors" from https://comparecamp.com/titanic-statistics/.
3. Encyclopedia Titanica. (2020). "Titanic Crew Survivors" from https://www.encyclopedia-titanica.org/titanic-crew-survivors/.
4. Ultimate Titanic. (2012, April). "Titanic: By the Numbers" from https://www.ultimatetitanic.com/facts-statistics/.
5. Encyclopedia Britannica. (2020). "Titanic" from https://www.britannica.com/topic/Titanic.
6. Titanic Database. (2012, April). "List of crew members on board RMS Titanic" from https://titanicdatabase.fandom.com/wiki/Crew_of_the_RMS_Titanic.
7. History on the Net. (2020). "The Titanic — Crew" from https://www.historyonthenet.com/the-titanic-crew.
8. Lehnhardt, Karin. (2019, July 3). "45 Unsinkable Titanic Facts" from https://www.factretriever.com/titanic-facts.

9. RMS Titanic - Ship of Dreams. (2014). “Designing & Building Titanic” from http://www.titanicandco.com/construction.html.
10. History. (2020, October 30). “Titanic by the Numbers: From Construction to Disaster to Discovery” from https://www.history.com/news/titanic-facts-construction-passengers-sinking-discovery.