Titanic Crew Member Survival Data

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

Which Domain?

What domain is this data going to come from? Please list 10 references (with a brief annotation) to use to make sense of what you're doing with these data.

The sinking of the Titanic is a heart-breaking tale, and while much of the tragedy focuses on the passengers of the ship, I feel the workers have a story that goes untold. My first project will be an analysis of the crew members employed by the White Star Line that were aboard the fateful voyage.

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

Which Data?

What is the dataset you'll be examining? Please provide a codebook if there is one or a link to the dataset as well as a detailed description.

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 csv format. Another source of data is Kaggle: https://www.kaggle.com/c/titanic/data.

These are some of the variables I will be exploring:

Variable	Definition	Key
Name	Name of passenger	
Age	Age in years	
Gender	Gender	
Class/Dept	Department of the ship worked in	
Joined	The stop (city) at which they boarded the ship	
Occupation	More detail about the specific role of the crew member	
Survived	Survival	0 = No; 1 = Yes
Boat	If survived, which lifeboat they were in when they were rescued	
Body	If lost at sea, a number was used to identify body/burial information	
Nationality	Where crew member was from	

Research Questions? Benefits? Why Analyze these Data?

How are you proposing to analyze this dataset? This is about your approach. Here, you'll be proposing your research questions as well as justifications for why you'd offer these data in this way.

I specifically want to know about the survival rate of Titanic's crew. How many total crew members were there and how many survived? How many were men and how many were women? In which parts of the ship did they work? At which port did they board the ship? I will be using Python and/or R to do the data analysis an attempt to find answers to these questions.

What Method?

What methods will you be using? What will those methods provide in terms of analysis? How is this useful?

The methods I will use for the analysis of the crew survivors will most likely include bar charts, ROC curve, and logistic regression. These will provide clarity on crew members, who survived, and if there are any trends or patterns in the data as to why they survived while others did not.

Potential Issues?

What challenges do you anticipate having? What could cause this project to go off schedule?

I am worried about not having enough information to draw any conclusions that I will be able to state definitively whether there are any actual predictors to which crew members would have survived the sinking.

Concluding Remarks

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. I am exploring data on the crew members to look for any trends in the data from those that survived the disaster.