Christopher Anderson
Professor Williams
December 13, 2020

DSC 680 Applied Data Science

Project 1 Check-In/Milestone 2

Any surprises from your domain from 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 is an analysis of the crew members employed by the White Star Line
that were aboard the fateful voyage. So far, there isn't anything too surprising with the
data; however, it was a bit jarring to see how far many more men were members of the
crew than women. But after looking over the data and seeing it from the perspective of
the different departments, it began to make sense as there were many more men that
were parts of the crew to shovel coal into the boilers to provide power and electricity to
the steam-powered ship.

Is the dataset what you thought it was?

Going into it, I was pretty confident that that dataset from the Titanic People Database
as created by Encyclopedia Titanica — https://www.encyclopedia-titanica.org — was
going to be what I expected, and it turned out that it was. As I'm a bit of a Titanic
diehard, I already had a subscription to this site which allows for more granular
downloading of data in csv format and I was able to view online and download a listing

of records of all crew data including survival data. I also retrieved data from Kaggle: https://www.kaggle.com/c/titanic/data as a way to do cross-referencing and build out a more complete dataset if/as needed.

Have you had to adjust your approach or research questions?

We are a few weeks in, and so far I have not yet had to adjust my approach or research questions. With the data that I have, I can see how many total crew members there were on the ship, and how many of them survived. I have a dataset showing how many crew members were men and how many were women, and I can also see the different departments of the ship they worked in — for example, engineering, hospitality, etc. Another interesting data point is the port (city) in which they boarded *Titanic* and began their working journey on the trip to New York.

Is your method working?

As I make my way through the data and look to get answers to my questions, the methods for analyzing *Titanic* crew survivor data appear to be working. I have gotten a good look at some of the information using bar charts, a nice and simple way to see the answers to my questions. Another part of the process has been converting categorical data for the port city information to numerical data and that has gone smoothly as well. Time will tell if the ideas that I have here will completely pan out but I believe that they will based upon what I am seeing so far.

What challenges are you having?

Regarding challenges, I am still working through the best way to present the data. I am comfortable displaying this information using a combination of bar charts and other methods, but am wondering if there is something else that would work as effectively or better. I am contemplating the idea of creating something more visually appealing such as a nice infographic and using multiple graphics to help better tell the data story of the crew and where they worked on the ship, such as their physical location and deck level

on the ship. As I work more through the data, I am sure that the best answer will present itself.