DSC530 Final: Shark Attack EDA

Isabella Sturm

2021-03-06

When I was looking for a dataset to use for this project, I had a really hard time finding one that I liked. I really wanted a topic that I was interested in but in the end, it would have been better to find something that was better for the techniques we learned in this course.

I chose a dataset on shark attacks, and there were many problems with the dataset. There were a lot of inconsistencies in formatting of the data and missing values. I struggled with getting the data in the correct format or at least removing rows that did not fit what I needed – for instance there were empty strings or strings containing spaces in the ages column and I couldn’t get rid of them to only look at the numeric ages. Because of the issues I faced and the dataset I chose, the EDA was not very successful, and I could not really pull any useful insights from it.

There are some variables I think could have helped in the analysis. When I first decided on shark attack data, I thought it would be really cool and useful to have data about the weather on the day of the attack. Also, while the time of day would also have been useful, I didn’t use it because the data format is so poor and inconsistent.

Overall, the main challenges I faced was getting the data in a format I could use. As stated above, one of the columns that I really wanted to use but had trouble with was Age. The data was very inconsistently formatted and my attempts to fix the data did not work as I had hoped and planned. Another issue I faces was that the data did not have many numeric columns and trying to convert them to numeric is not really ideal – continuous data would have been much better to work with than the data I chose which had a lot of text data, some of which could be translated into discrete numeric categories. But with the techniques we learned in this class, combined with the poorly formatted and difficult to work with data, I was unable to perform the analysis I wanted and pull any useful insights.