

Product Description

Willemijn Barens - Data Processing 2022

In this project I aim to answer the following question:

'What trends can we identify in shark attacks that have occurred in the passed 200 years?'

In order to answer this question I used two different datasets in csv format. One of these contained information about all the shark attacks that have been reported up until June 2018. The second dataset contained geographical information on all big cities and localities in the world, including their longitude and latitude.

Firstly I created two functions to read these datasets, clean them, and join them appropriately. These functions can be found in `read_csv_file.py`. To get the most accurate geographical details per shark attacks, I joined the datasets on 'Area' and 'admin_name'. However, some shark attacks did not include area data. Therefore I matched those ones on country instead.

After the data was put together in one pandas DataFrame, I wrote a program to visualise the shark attacks on a world map, `world_map_sharks.py`. This map was created with the use of Folium and saves the map as an HTML file. On this map the places where shark attacks have occurred are represented as blue circles. The more shark attacks there have been in one area, the bigger the circle. A few places that stand out with the biggest circles are Australia, South Africa, and the USA. This does not necessarily mean that these are the most dangerous places to go into the ocean. Since these are all fairly developed countries it is possible that non-fatal encounters get reported more often than in less developed countries. To get a better insight in this outcome one could take the number of fatal attacks per area into account. When going over a circle with the cursor a tooltip pops up with the name of the location and the number of fatal shark attacks. For example, the circle for Florida is much bigger than the circle for the Madang Province in Papua New Guinea, even though the number of fatal attacks is lower in Florida.

Other interesting data to take into account are the activities that people were doing before getting attacked. When one clicks on one of the circles on the map a popup will be displayed with the top 5 most common activities per location. Many places where a higher number of both fatal and non-fatal attacks was reported will have 'surfing' as the number 1. Other common activities, unsurprisingly, include swimming, fishing, and more specifically, spearfishing.

For the second visualisation I used the Bokeh library to output two line plots with the shark attack data per year, this code can be found in `visualise_sharks.py`. For the first plot (p1) I created a second DataFrame which included the amount of fatal shark attacks per year. This is displayed in the plot as the red line. The blue line displays the amount of both fatal and non-fatal shark attacks per year. The light coloured lines show the linear regression of the aforementioned lines. From the plot it becomes clear that the total amount of shark attacks has gone up massively. The amount of fatal attacks however has been quite steady. An explanation for this finding is that they only started reporting non-fatal attacks in recent years. Another element that could play a role is that the global population has grown substantially and therefore, the number of people in the ocean has grown as well. If there are simply more people in the water it is not surprising that it is more likely for some of those to encounter sharks.

In the second plot (p2) the percentage of fatal shark attacks is shown in relation to the total amount of attacks. Here the percentage slowly decreases over the years. This is consistent with the findings from the first plot, where the total amount of attacks increases over time.

In conclusion, it is questionable whether the amount of shark attacks has actually increased or if they just get reported more often. It does seem that swimming and surfing are the most common activities for people to be doing while getting attacked. Luckily, most attacks are non-fatal.

Appendix

Shark attack dataset. (2021, 26 juli). Kaggle. <https://www.kaggle.com/datasets/felipeesc/shark-attack-dataset>

World Cities Database | Simplemaps.com. (z.d.). <https://simplemaps.com/data/world-cities>