

DATA INFORMED RECOMMENDATIONS FOR A SHARK REPELLENT COMPANY



# Summary

This report is a collaboration between four Google Certified Data Analysts spanning three continents. We met on LinkedIn after completing the Google Data Analytics course on Coursera and we decided to do a project together to showcase our skills and to show our ability to work together with others even across continents.

For this project we did an analysis of shark attack data to give data-driven marketing advice for the company Sharkbanz, a world leader in shark repellent technology. We will see what marketing strategies Sharkbanz can use to ensure that their marketing is effective.

Also, feel welcome to click on the link for a Slideshow summarizing our case study [Sharkbanz Case Study](https://docs.google.com/presentation/d/1mD-6GKxgRo3ajVTzFETIZQgZKclMVgNHeeK7ph6UEeU/edit?usp=sharing) and on [Tableau.](https://public.tableau.com/views/WorldwideSharkAttacks/SharkAttackCaseStudyGroupProjectbyKristyJohnsenPrinceBhalawatGursharanKaurSainiandCornlHanekom?:language=en-US&:retry=yes&:display_count=n&:origin=viz_share_link)

# Contents

[**Summary**](#_heading=h.gjdgxs) **1**

[**Contents**](#_heading=h.30j0zll) **2**

[**Background information**](#_heading=h.1fob9te) **3**

[**Business task**](#_heading=h.4b1gh5lu75fp) **3**

[**Findings**](#_heading=h.2et92p0) **4**

[**Recommendations**](#_heading=h.tyjcwt) **4**

[**Data**](#_heading=h.3dy6vkm) **5**

[**Data Cleaning and Manipulation**](#_heading=h.5tvncisaj2a9) **5**

[**Analysis**](#_heading=h.1t3h5sf) **6**

[**7.1 shark attacks around the world**](#_heading=h.kowbf1y7y2p3) **6**

[**7.2 Most dangerous locations**](#_heading=h.2s8eyo1) **9**

[**7.3 What time of the year is an attack likely?**](#_heading=h.xnlnjavkx811) **11**

[**7.4 Time of Day**](#_heading=h.c9qqbvtju8y7) **12**

[**7.5 Highest risk activities**](#_heading=h.4unzajr8kjco) **13**

[**7. 6 Attacks by Age and Sex of Victim**](#_heading=h.l4tedbarw5uz) **14**

# Background information

The chance of a shark attack happening is statistically insignificant, about one in 5 million. Shark populations, however, are in dramatic decline. Some estimates are that nearly 100 million sharks are killed each year by human activities, such as longline fishing and pollution. Current research has raised public awareness about how vital sharks are to the health of the oceans and how our actions in the oceans can affect these species.

This new awareness recognizes that it is vitally important to both species that sharks and humans peacefully coexist. Companies around the world are actively looking into shark repellent technology as a way to keep everyone, including the sharks, safe in the water. One such company is Sharkbanz which utilizes magnets to keep sharks away from surfers, swimmers, and fishers. This technology does not harm the sharks, nor does it harm the marine ecosystem.

# Business task

Sharkbanz, a world leader in scientifically proven and tested shark repellent technology, has hired our team to analyze the Global Shark Attack Files to see what insights we can gain to help inform their 2022 marketing strategy. Sharkbanz wants to expand its marketing and needs our insights to make data-driven marketing decisions about potential Sharkbanz customers.

What trends do you see with shark attacks? Is there a group that is more at risk?

# Findings

The average shark attack victim is:

* Male
* Is an average of 27 years old. Looking at standard deviation, ages 14 to 41 as also in the high risk category
* Lives in South Africa, Australia, or the US
* Is surfing, swimming, or fishing

# Recommendations

1. Since most shark attack victims are young men who are surfing, swimming, or fishing, marketing strategies should be targeted towards this group. These marketing strategies should include social media and music streaming services to reach this audience. The parents and guardians of the younger members of this cohort should also be targeted, since they are the ones most likely to purchase the Sharkbanz product.
2. Corporate partnerships should be investigated with national hotels and AirBNB to focus on families traveling to beaches and who may be interested in purchasing or renting Sharkbanz for their vacation. A breakdown of beaches that have a history of attacks can be found here.
3. Local retailers can be partnered with to create beach gear rental packages that include Sharkbanz with the rental of a surfboard, boogie board, fishing gear, or other equipment. This will target families who are visiting the area and looking for package deals to save money.

# Data

The data used in this report was obtained from <https://www.sharkattackfile.net/incidentlog.htm>.

At the time this was written, the dataset had 6701 entries of shark attacks around the world. The dataset is updated regularly when a new shark attack happens, and it is quite impressive that the oldest entry is from 725 BC, where sharks attacked sailors whose ship sank near Italy.

# Data Cleaning and Manipulation

Cleaning data is obligatory and an essential step for Data analysis. For this case study, we have used Python for all the cleaning, wrangling, and manipulating the data and to find out meaningful insights from it. Initially, we started to have a look around the messy data [shark attack](https://www.sharkattackfile.net/incidentlog.htm). Imported it, used our data wrangling skills to clean it up, prepared it to be analyzed, and then exported it as a clean CSV data file.

Steps involved in the data cleaning:

1. We started with importing the necessary packages that were needed for cleaning. Then imported the dataframe using Pandas with the correct encoding. The initial messy data contains 24 columns and 25827 rows. We decided to work on the following variables: *Fatal(Y/N), Age, Sex, Time, Activity, Type, Country, Year, Date, Species, Area, and Month.*
2. All the columns mentioned above contained null values, blanks, duplicated entries, and some missing and unwanted information not needed for analysis. Some columns, such as Case Number, Investigator or source, href formula, etc., were dropped by us as they were not required for finding insights.
3. Here is the link for the notebook where all the code is present that was required to clean the original data [Shark Attack Data Cleaning](https://colab.research.google.com/drive/1VIsypMcPyXHGFA2qnFjQswAe0cHEP08Q#scrollTo=vzeGNhfd3Uy3)
4. Most attacks, around 77.26%, are not fatal attacks, and 9.44% are Fatal attacks, while the remaining were unspecified.
5. Sharks mostly attack men (87.67% male vs.9.33% female)
6. Most attacks were unprovoked (75.61% unprovoked, 15.24% provoked, and 9.15% were unspecified).
7. The cleaned CSV data file containing the results of data cleaning work can be found here [SharkFinal](https://docs.google.com/spreadsheets/d/1ytCxojjigYY7uCxVjRAgMfmGrsOn0ePP5CxGJy8YhK4/edit?usp=sharing).

# Analysis

# 7.1 shark attacks around the world

First, we have the following visualization that was made in Tableau. The visualization shows each country and the number of shark attacks recorded for that country from 1900-2021. We can see from this visualisation that the US, Australia, and South Africa are the countries with the most reported shark attacks.

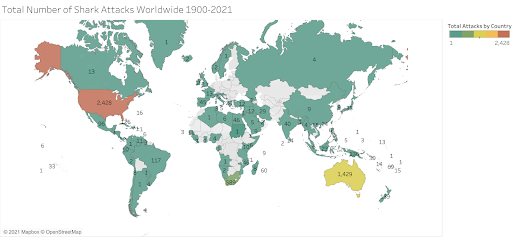
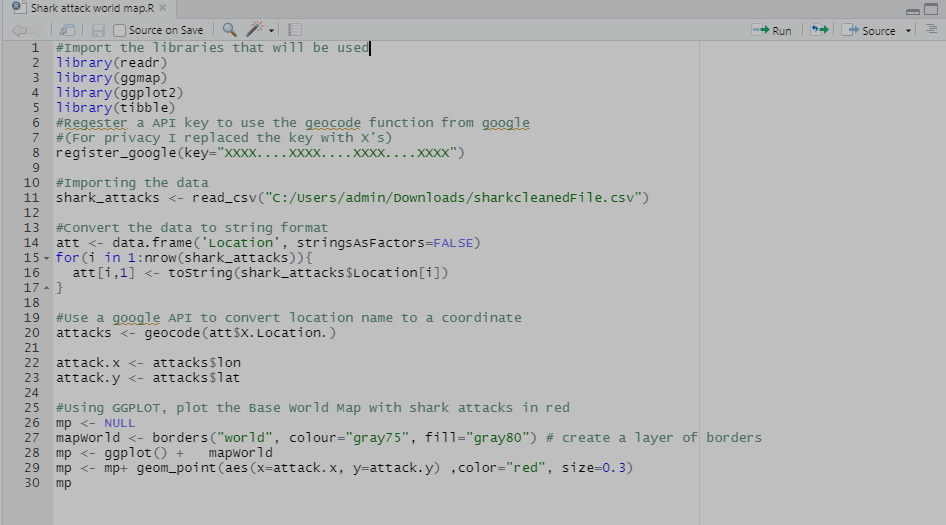


Figure 1.1 - Shark attacks per country

The following R code will plot the exact location of a shark attack on the world map. To do this, the data was imported from <https://www.sharkattackfile.net/incidentlog.htm> on 4 November 2021. The data was imported to excel and cleaned. Then columns *Country, Area,* and *Location* were combined into one column and all the rows missing an entry were removed. All the entries containing nonspecific locations, *like 46 miles off Mayport, Duval County*, and *Between Pescadero Point & Bean Hollow Beach, San Mateo County,* were removed since these are not locations that can be pinpointed with coordinates. Unfortunately, the program could not process entries that contained special characters, and they had to be removed. This may create a slight bias in the data, but since these are in the minority, and we are interested in the locations where shark attacks are frequent, this will not influence our analysis. In the end, 915 of the most recent data points were used in the visualization.



Resulting in the following visualization

Map

Description automatically generated

Figure 1.2- Shark attacks around the world

In this visualization, you see a world map with red dots representing shark attacks on the location where they occurred. One can see that North America, South Africa, and Australia are the countries with the most shark attacks. This supports the other visualization we have with the number of attacks per country. We can also note that if Sharkbanz wants the most significant return on investment, it should advertise in regions with the most shark attacks. So, Sharkbanz should advertise on the east and west coast of North America, on the east coast of South Africa, and the east and west coast of Australia.

Furthermore we notice that this map roughly correlates with ocean temperatures as seen in the following visualization obtained from the NRF website at <http://app01.saeon.ac.za/sadcofunstuff/OceanTemperature.htm> on 16/10/2021.

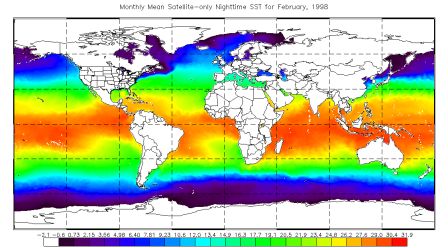


Figure 1.3 - Ocean temperatures

If we look at South Africa in Figure 1.1 and Figure 1.2, we see that there are many more shark attacks on the east coast than on the west coast. Significantly, on the east coast, ocean temperatures are much higher than on the west coast. So, there is some correlation between shark attacks and ocean temperatures. This can be of great value in predicting shark attacks in the long term. Suppose there is an increase in ocean temperatures at some location due to climate change. In that case, there should be a greater emphasis on investigating shark attacks in that region due to a corresponding increase in human presence in the water. However, it should be noted that ocean temperatures are not the only factor that affects the number of shark attacks. There are many more factors that play a role. For instance, some cultures are more likely to swim in the ocean than others. Another factor can be environmental. Some oceans are too polluted for people to want to use for recreational purposes.

# 7.2 Most dangerous locations

Next, let's examine where the most shark attacks occur. This can be useful when Sharkbanz wants to advertise at a few specific locations where many shark attacks are likely to happen. To do so, I imported the cleaned dataset sharkfinal.csv into excel. Next, I combined the *Country, Area,* and *Location* columns and used a pivot table to group the exact locations. After this, I ordered the locations with the most attacks first. Here is a list of the most dangerous places in ascending order.

| Location | Number of shark attacks |
| --- | --- |
| USA Florida New Smyrna Beach, Volusia County | 185 |
| USA Florida Daytona Beach, Volusia County | 30 |
| USA Florida Ponce Inlet, Volusia County | 26 |
| USA Florida Melbourne Beach, Brevard County | 20 |
| USA South Carolina Myrtle Beach, Horry County | 18 |
| USA Florida Cocoa Beach, Brevard County | 17 |
| SOUTH AFRICA KwaZulu-Natal Durban | 16 |
| USA South Carolina Isle of Palms, Charleston County | 14 |
| USA Florida Jacksonville Beach, Duval County | 13 |
| AUSTRALIA Torres Strait | 12 |
| USA Florida Cocoa Beach, Brevard County | 12 |
| USA Florida Ponce Inlet, New Smyrna Beach, Volusia County | 12 |
| SOUTH AFRICA Western Cape Province Mossel Bay | 11 |
| USA Florida Juno Beach, Palm Beach County | 11 |
| USA Florida Ormond Beach, Volusia County | 11 |
| USA Florida Palm Beach, Palm Beach County | 11 |
| USA Hawaii | 11 |
| AUSTRALIA New South Wales Sydney | 10 |
| AUSTRALIA New South Wales Sydney Harbor | 10 |
| SOUTH AFRICA Eastern Cape Province Nahoon | 10 |
| USA Florida Singer Island, Riviera Beach, Palm Beach County | 10 |
| USA Florida Vero Beach, Indian River County | 10 |
| AUSTRALIA Queensland Brisbane River | 9 |
| SOUTH AFRICA KwaZulu-Natal North Beach, Durban | 9 |
| USA Florida Florida Keys | 9 |
| USA Florida Florida Keys, Monroe County | 9 |
| AUSTRALIA New South Wales | 8 |
| AUSTRALIA Queensland Moreton Bay | 8 |
| AUSTRALIA Torres Strait Near Thursday Island | 8 |
| AUSTRALIA Torres Strait Thursday Island | 8 |
| SOUTH AFRICA Eastern Cape Province Nahoon, East London | 8 |
| SOUTH AFRICA KwaZulu-Natal Amanzimtoti | 8 |
| SOUTH AFRICA KwaZulu-Natal Country Club Beach, Durban | 8 |
| SOUTH AFRICA Western Cape Province False Bay | 8 |
| USA Florida Sanibel Island, Lee County | 8 |
| USA South Carolina Folly Beach, Charleston County | 8 |

# 7.3 What time of the year is an attack likely?

In this section, I want to take some of the locations where the most shark attacks occur and see which month of the year these attacks are likely to happen. This might be useful when advertising shark repellent technology and the company wants to know when in the year to advertise at a specific location. To do this, I used pivot tables in Excel to get the number of attacks that happened within a particular month in a specific location.

First, let us plot the number of attacks per month for the most popular location, i.e., USA, Florida, New Smyrna Beach, Volusia County. The x-axis has each month of the year, and the value on the y-axis shows the number of attacks that occurred in that month. We can see from this visualization that in the USA, Florida, New Smyrna Beach, Volusia County, attacks are likely to occur from March to November and are most likely to occur in April, July, August, and September.

Figure 3.1 - Attacks per month at USA Florida New Smyrna Beach, Volusia CountyChart

Description automatically generated

In the next visualization we see the number of attacks per month for USA Florida Daytona Beach, Volusia County. We can see that a shark attack is likely to occur between April and

November with the most likely time in July.

Figure 3.2 - Attacks per month at South Africa Eastern Cape Province Port Alfred Chart

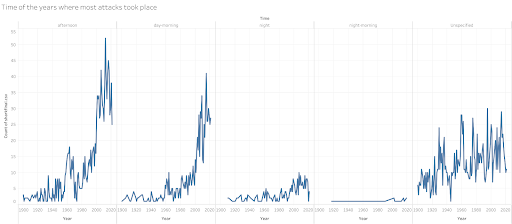
Description automatically generated

Next, I want to show this visualization for a different continent as the distribution is affected by climate. Here follows this graph for SOUTH AFRICA Eastern Cape Province Port Alfred. We can clearly see that the months where a shark attack is most likely to happen are between December and April, with a peak in April. This makes sense since these are the summer months in South Africa.

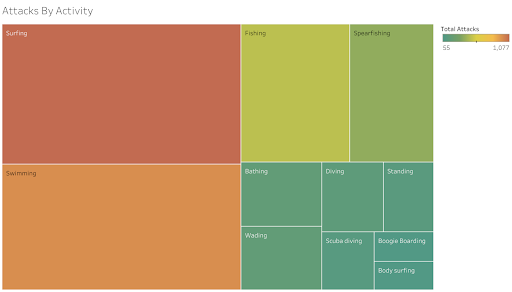
These are some specific examples but if the company wants to see the distribution for a different location all they must do is click on a different location on the following google sheet file [Probability of a shark attck per month at a specific location](https://docs.google.com/spreadsheets/d/1iMG16n99fP-HJ9bOfDyxXg9TdrTcnU8aMio3eWB1Geo/edit?usp=sharing)

# 7.4 Time of Day

The following visualization shows the number of shark attacks per year separated by the time of day the attack took place. We can see that shark attacks increased in the morning and afternoons during the years.

Figure 4.1- Time of day

# 7.5 Highest risk activities

The following [Tableau visualization](https://public.tableau.com/views/WorldwideSharkAttacks/SharkAttackCaseStudyGroupProjectbyKristyJohnsenPrinceBhalawatGursharanKaurSainiandCornlHanekom?:language=en-US&:retry=yes&:display_count=n&:origin=viz_share_link) shows which activities are most at risk of a shark attack. The larger the area in the visualization the more shark attacks were recorded in that category. Figure 5.1 - High risk activities

We can clearly see that the activities with the highest risk are Surfing, Swimming, and Fishing.

# 7. 6 Attacks by Age and Sex of Victim

In the following visualization we see a histogram with the number of attacks for each age with the blue representing males and the red representing females. We can see that the average is 27 years old and people from 14 to 41 years of age are most at risk with a peak at 17 years of age.

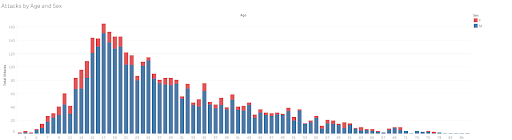


Figure 6.1- Attacks by age and sex