BE SHARK SMART!

Designing a safety outreach to reduce Florida's shark attacks

Camil, Ceci, Moana, and Victor
Ironhack Data Analytics 2024 Week 2

01 Project Overview

The tourism board of the state of Florida plans to update their public safety outreach to inform tourists about how to lower their risk of shark attack

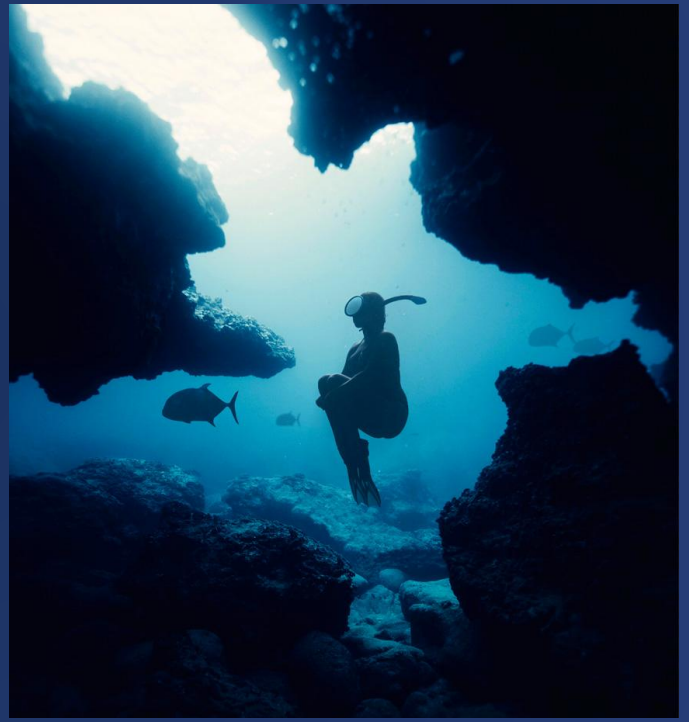
We have been hired to carry out an exploratory data analysis of shark attack reports in Florida. The analysis will be used to create a profile of the most likely shark attack victims, and identify in what locations and under which conditions they are most likely to happen

Our analysis will guide the tourism board in deciding to which demographics they target their campaign and how and where it should be shared for maximum impact

HYPOTHESES

If a significant amount of cases occur during a certain activity (i.e., surfing), safety ads should be posted in certain areas (i.e., popular surfing beaches), and inform tourist service providers (i.e., surf schools and board rental companies) to alert their customers

If the majority of shark attacks are from one species of shark, we can target the information to include that species most active times and particular behaviours so tourists can avoid them



02

Data Wrangling

Challenges

- Many columns contained missing, incorrectly formatted, or incomprehensible data
- The dataset also contained a lot of unnecessary data for our analysis
- Some data/categories which would have been relevant were missing

Solutions

- Implementation of various functions to clean the dataset as thoroughly as possible
- Filter applied to retain only the data relevant to the analysis in the new dataframe (e.g., only date related to Florida)
- Removal of columns with inexplicable data (such as “Unnamed 21/22” , columns with no titles, and data which was impossible to interpret without further explanation)
- Addition of extra categories to conduct our analysis as effectively as possible (i.e., splitting categories with too many unique variables)

Most troublesome columns

Time

Most entries in 00h00 format

Solution: converted to time of day

Species

No consistency

Solution: split into categories of size and those with a listed species

Location

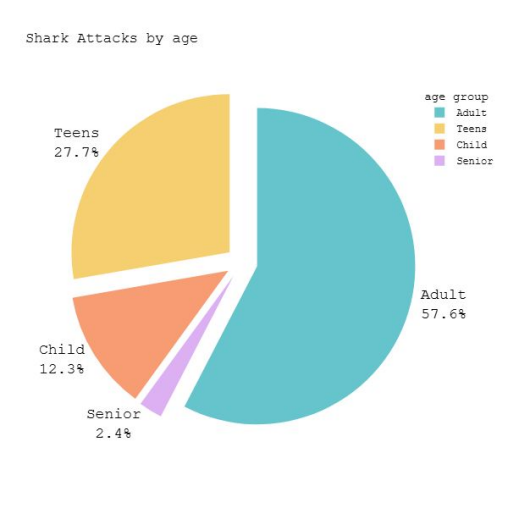
Too many unique values

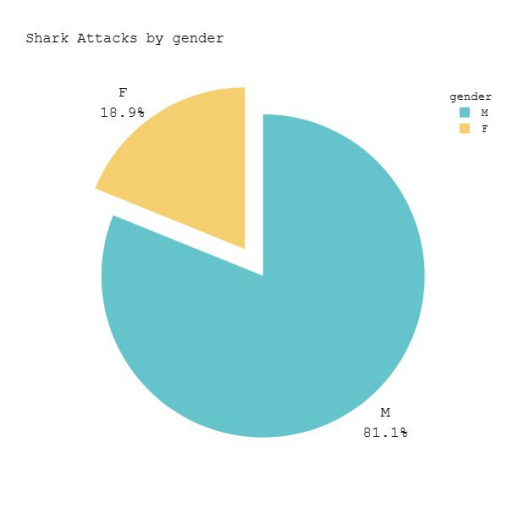
Solution: split by location and county

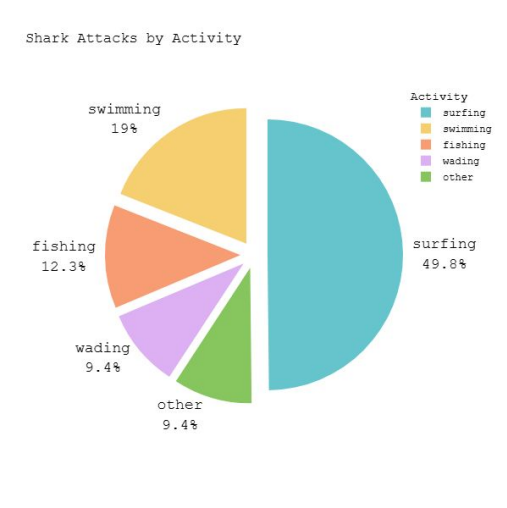
03

Exploratory Data Analysis

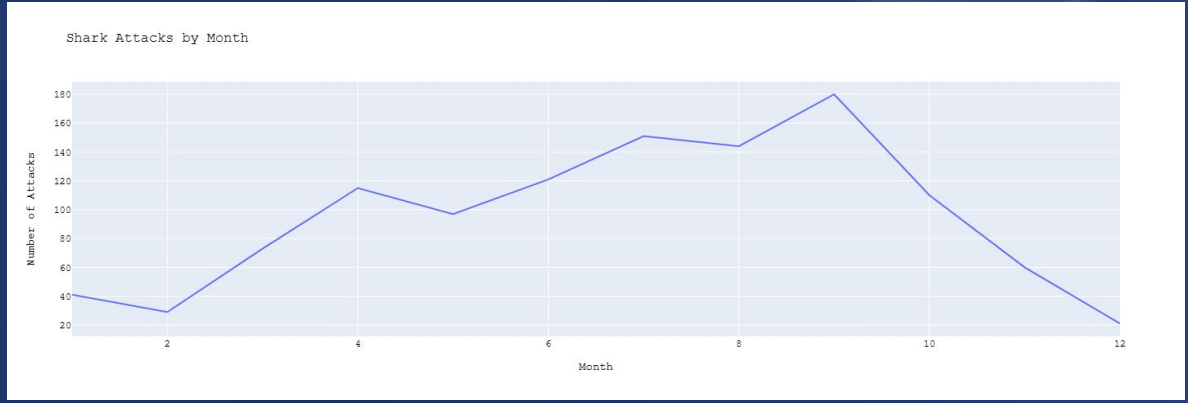
Target demographic by summary statistics

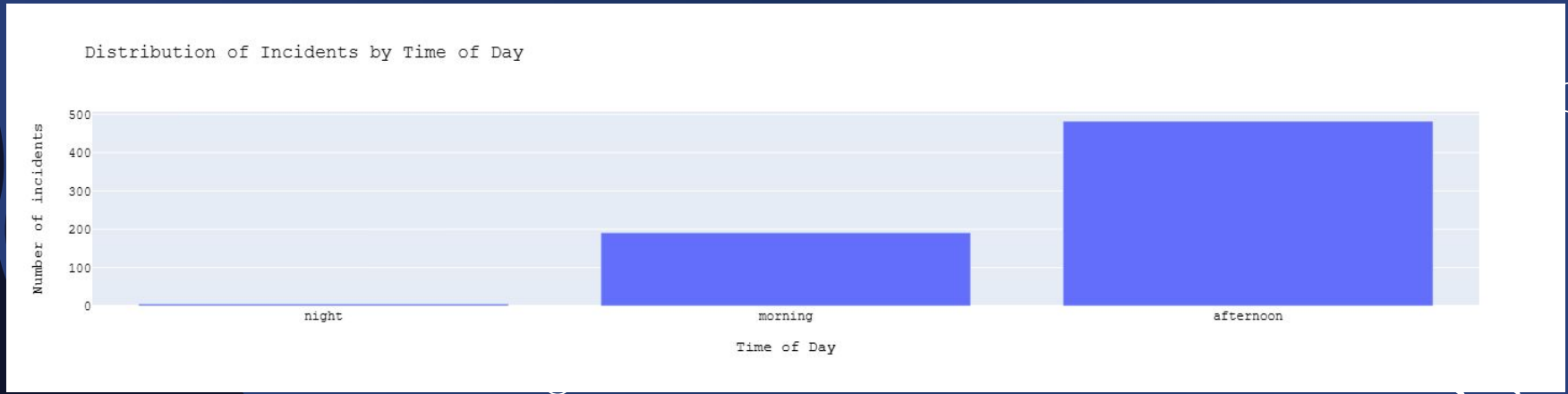




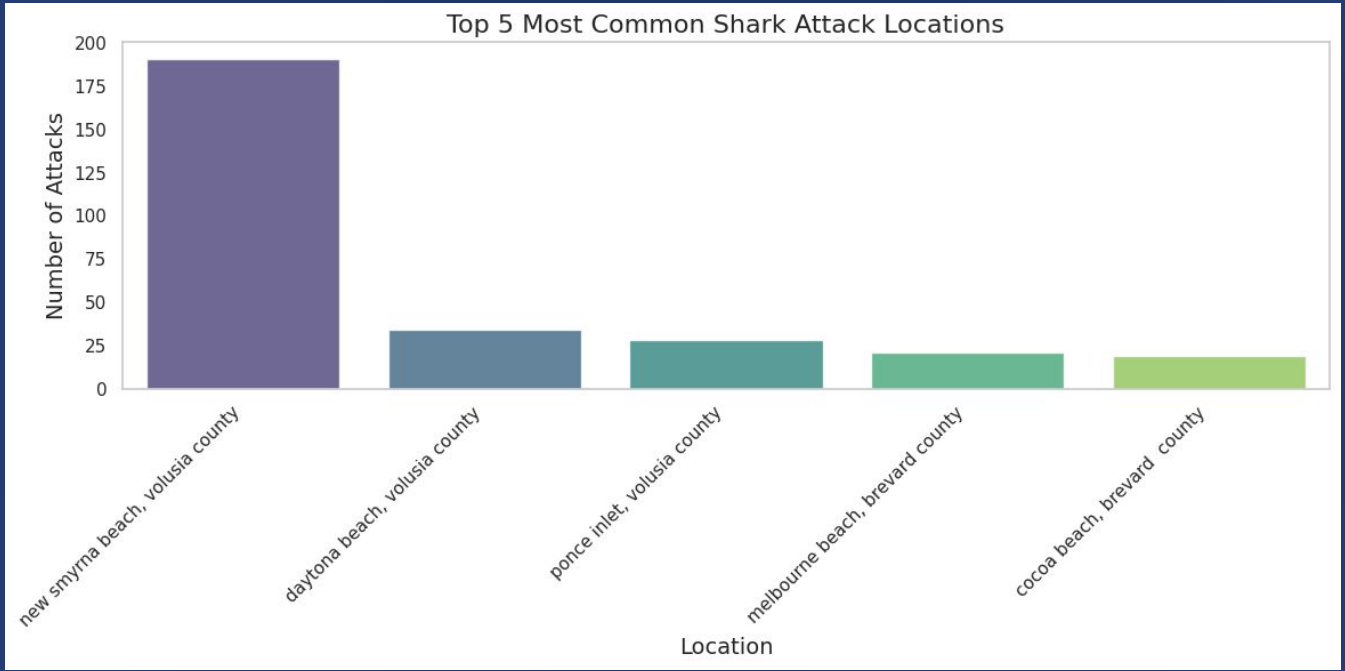


Risk profile by time of year/day





Target location by county



Conclusions from relationships between the surfer dataset

type	
provoked	11
unknown	8
unprovoked	435

Among surfer shark attack victims the vast majority were unprovoked. This shows that the activity of surfing itself is dangerous, rather than the surfer's behavior leading to an attack. Information should thus focus on time and condition most likely to increase the risk of attack, rather than telling surfers how to act in the water

time	
afternoon	231
morning	77
night	2
unknown	144

The vast majority of surfers were attacked in the afternoon. It is unclear whether this is because the sharks are most active at this time, or whether it is the most popular time for surfing, but regardless, an effective outreach campaign should inform surfers to be extra alert during the afternoon, or avoid surfing in the afternoon if possible

04 Major Obstacle

- Cleaning such a large and inconsistent dataset was the biggest challenge we faced
- Decision to limit our analysis to Florida helped to keep cleaning a manageable project
- Keeping the business case in mind was always the key to deciding how to standardise and clean the data

05 Conclusion

In every variable analysed, a single value was significantly higher than all others, meaning that based on our data, the safety campaign can be very easily targeted to a single demographic profile and a handful of individual locations in two counties for maximum effect

Thank You!

Shark Attacks in Florida // Camil, Ceci, Moana and Victor // Ironhack DA FT 2024 Week 2 Project