



# AVOID SHARKS? USE SHARKAVOIDER!

BY

JONATHAN, JOHANNA, BRU + TUNG

# PROJECT OVERVIEW

## Business Problem

How can insurance companies accurately assess the risk of shark attacks to offer appropriately priced insurance premiums for individuals and ocean activity businesses, such as surf and dive schools?

## Hypothesis

By using exploratory data analysis, we can accurately classify individuals and businesses into risk categories (low, medium, high) based on a points system to price insurance premiums more appropriately.

# PROJECT OVERVIEW

## The original dataset description

- *Over 6000 rows spanning from 2026 to year zero*
  - *many unnamed columns*

## Structure and process of cleaning & analysis

- *limited the range of rows (years: 2024 - 2014)*
- *assigned attacks to seasons for better assessment*
  - *split the cleaning to two columns per person*

## Unique data cleaning technique & methods

- *Using regex to extract months from different date formats*

# DATA WRANGLING + CLEANING

## Significant challenges when cleaning

- *date strings vs datetime objects hard to extract*
- *finding a way to extract the top 10 most recurring words*

## *How to overcome*

- *parsing via regex or using datetime.month*
  - *import Counter function*

# EXPLORATORY DATA ANALYSIS

## EDA methods used

- *text cleaning, data type correction, handling missing values*
  - *summary statistics*
  - *visualizations*
- *relationships between variables*

## Insights and interesting patterns

- *the data shows a significant skew towards the USA, males, summer months, and ages 10-20*
  - *how messy real datasets can be*
  - *how time consuming data cleaning is*

# MAJOR OBSTACLE

## Biggest obstacle or mistake

- *GitHub*
- *working on the same notebook*
- *Deciding which correlation in the data to observe to calculate a simple yet reasonable insurance risk  
(all variables correlate with each other --> only correlated fatality with other variables)*

## What did we learn from our mistakes?

- *restart kernel and clear outputs of all cells before committing*
  - *how to Resolve conflicts*
  - *work on separate notebooks*
    - *keep it simple!*

# CONCLUSION AND INSIGHTS

## Initial hypothesis supported or refuted?

- *Hypothesis was supported by creating a functional risk calculator based on our dataset (seasons, country, activity, sex, age, fatalities)*

## Surprising insights or findings?

- *few valid rows compared to the whole data (1200 vs 6600)*
- *shocking amount of people injured by taking selfies with sharks*

## Potential implications of findings if we used data going

- *Darwin at its best*
- *very different results if the whole set had been used*

A blue and white shark is shown from a top-down perspective, swimming in a body of water with a mottled orange-brown background. The shark's mouth is open, revealing a row of sharp, black teeth. The text "SHARKVOIDER" is written in a bold, black, sans-serif font across the shark's head.

**SHARKVOIDER**

**SAYS THANK YOU**

**BY**

**JONATHAN, JOHANNA, BRU + TUNG**