



Business Study:

Global Market Prioritisation Analysis for new *Sharktooth-Proof*®
product range of surfboard manufacturer “Surfhack”

Code: [Google Colab](#)

Data Analysts & Researchers:

Lora Chuaner Ding, Govarthini G S,
Matt Asquith, José Martins, Daniel Sigge

Project Scope

Background:

A lot of shark attacks happen during surfing. The slogan of Surfhack Ltd. has long been “Surf Fearlessly”, therefore the team went one step further and developed a new product range of Surfboards that secure surfers when attacked by sharks. Surfhack can leverage a network of global surf shops in North America, Europe & Asia.

Problem Statement:

Surfhack Ltd. can activate financial resources to introduce its new range in five distinct markets, but lacks comprehensive insights into which markets offer the optimal product-market fit. Launching a product in a singular market entails significant costs attributed to the activation of international delivery chains and the translation of diverse marketing assets.

Hypothesis:

Countries with the most surfing incidents caused by sharks can significantly increase sales of the new product-range.

Analysis:

Objective: Identify most lucrative markets for launch of product range

Resources: 5 DAs / Researchers, Open Dataset Global Shark Attacks

Key Features: Methodology, Code & Presentation

Prioritisation:

P0: Data Cleaning, P1: Data Vis, P2: Color coded data analysis

Timeline: 5 business days

Constraints:

- Data noise / DA Experience: Poor dataset and DAs in training result in several days spent on cleaning & wrangling available data
- Accuracy: Findings might have low quality due to missing datapoints
- Complexity: Certain DA jobs like analysing color coded data could take too long for timeframe

DA Logbook

Day 1:

- Basic data examination with Google Sheets
- Defined Project Scope & agreed on daily standup
- Removed all columns with irrelevant data with drop-function
- Removed rows with no data, 6944-6968 with drop-function

Day 2:

- Data Cleaning, 3 Columns per DA
- Removing null, removing duplicates, manipulating strings, etc.

Day 3:

- Further data cleaning
- Stackranking & decided on way forward around EDA / Data Vis
- Presentation Preparation

Day 4:

- Finalised data cleaning process
- Finalised presentation

Day 5:

- Presentation day

Roadmap						
Type	Team	Mon	Tue	Wed	Thu	Fri
Team Setup	All					
Data Collection	Dev					
Roadmap Definition	Res					
Data Cleaning	Dev	General	Dates	Aggr. / Filtering		
Data Analysis	Dev					
Insights Deviation / Writing Logbook	All	EOD	EOD	EOD	EOD	
Data Vis	Dev					
Finalisation of MVP	Dev					
Report creation, Presentation	All					

Initial DA Roadmap

Data wrangling and analysis

What did we do / why

- **Column cleaning-** dropped columns with no data or irrelevant data;
- **Row cleaning** - Rows with no data were deleted;
- **Filling NaN Values** - For numerical data, replaced NaN values with mean of the column. For categorical values, we replace NaN with the mode;
- **Replacing inconsistent entries** - To ensure consistency in the dataset, some entries were replaced by standard variables;
- **Creating new columns** - To simplify the analysis.

What techniques did we use

- **Column cleaning-** Used `drop()`;
- **Row cleaning** - Used `drop()` ;
- **Filling NaN Values** - Used `fillna()`;
- **Replacing inconsistent entries** - Used `str.upper()` to capitalize the entries then `replace()` to replace with standard variables. Also achieved this by combining `str.contains()` with `str.replace()` using custom dictionaries;
- **Searching substrings** - Used `str.contains()` method to search substrings;
- **Employed regex expressions** to separate character and numerical data.



Addressing Constraints & Challenges

Poor Data Quality / Noise:

Constraint: Data inconsistency and missing values posed a big obstacle. Required extensive time to clean and understand the importance of the data and its intricacies between columns.

Solution:

To further define launch markets and/or analyse local needs beyond the quantitative findings we can conduct qual. analysis in the pre-defined priority markets.

As next step:

Focus group interviews with surfers and surf shop owners help lock in the markets and also define further product innovation of the *Sharktooth-Proof*® range.

Learnings: Our experience taught us the precious lesson of distinguish valuable and noisy data. We discovered the importance of setting priorities and developed new skills in effectively cleaning data.





Executive Summary

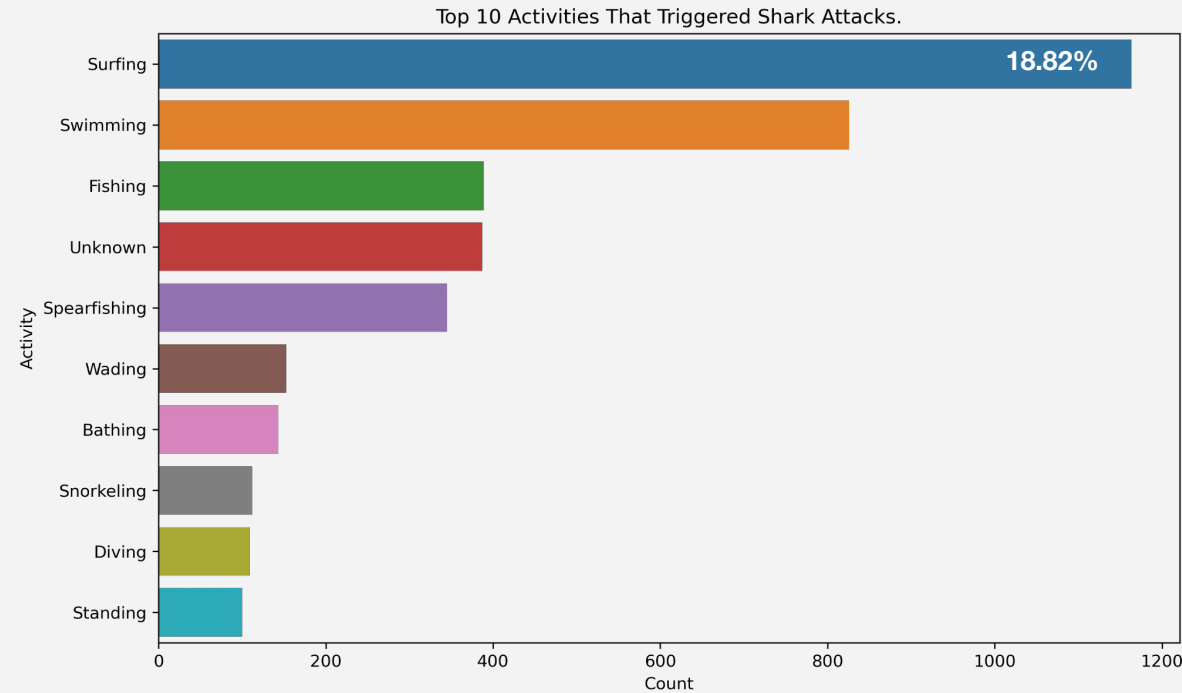
Market Research

1. Most **lucrative markets**: USA (36.8%), Bahamas (1.9%), Mexico (1%): North America congregates by far the most individual lucrative markets with many shark attacks **(Slide 8)**
2. Cost-benefit-analysis to **open businesses** in Australia & South Africa strongly advised: Both Australia (21.4%) & South Africa (8.6%) alone summarise more shark attacks than Europe & Asia combined (8.1%). **(Slide 8)**
3. The **species** most likely to be involved in shark attacks: White sharks. **(Slide 9)**
4. The country with the **highest fatality rate** from shark attacks: USA **(Slide 10)**

Activity and Product Research

5. The **activity with the highest frequency** of triggering shark attacks: Surfing (18.82%) **(Slide 7)**
6. The **colors** most likely to provoke shark attacks: blue and green **(Slide 11)**

Top 10 Activities That Triggered Shark Attacks



- Total number of activities: **6180**
- Total number of activity types: **1478**

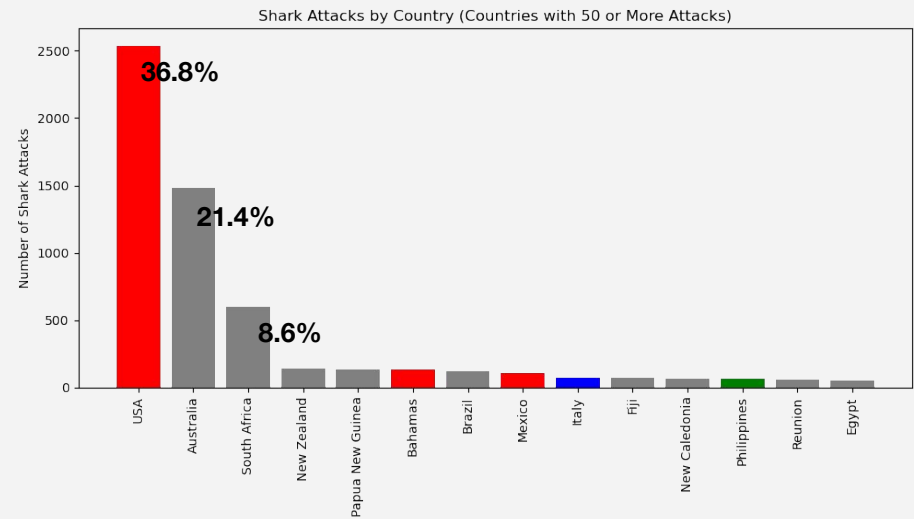
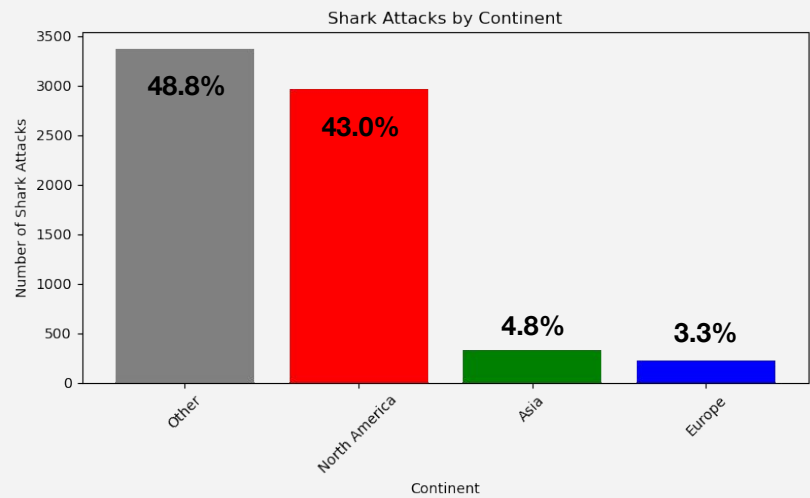
Initial Hypothesis: **TRUE**

Finding countries with the most **surfing incidents** caused by sharks can significantly increase sales of the new product-range.

Conclusion:

In over 1000 types of ocean activities that can trigger shark attacks, **18.82%** are attributed to **surfing**, validating our initial hypothesis.

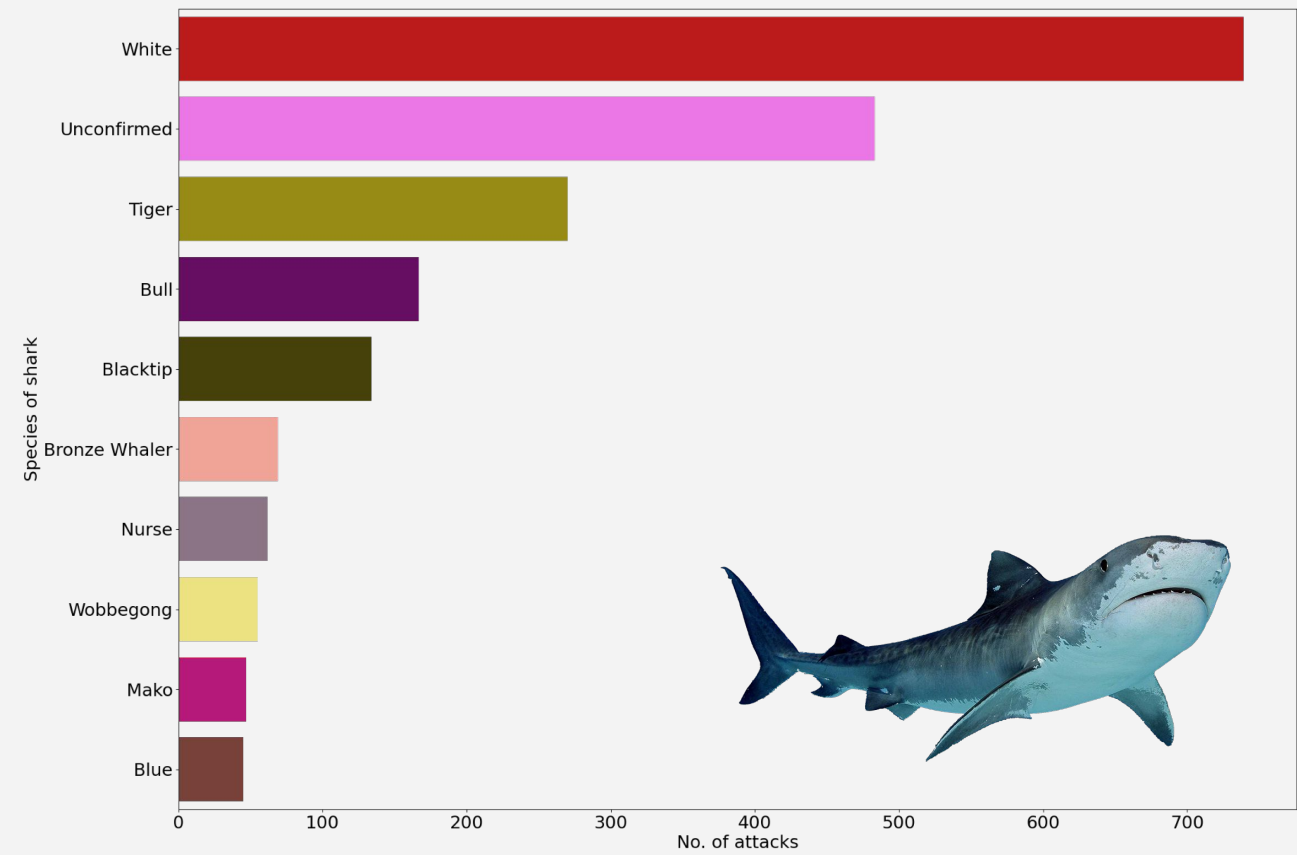
NA congregates by far the most lucrative markets, other advantageous places aren't (yet) addressable



Convincing Fact: Both Australia (21.4%) & South Africa (8.6%) account for more shark attacks than Europe & Asia combined (8.1%).

Of a total of 6.9k global shark attacks, **NA congregates many markets within countries with most shark attacks** (#1: USA - 2538, #6: Bahamas - 136, #8: Mexico - 106) and faces 43% of all global attacks. Above analysis show that other continents from SurfHACK's store & e-commerce network are trailing behind: Asia (4.8%) & Europe (3.3%). Due to this marginal number of attacks and **huge potential in off-markets Australia (#2: 1481) and South Africa (#3: 598)**, it's strongly advised to evaluate the costs to leverage new partnership networks spanning over these countries as next step.

Broad geographical range for top attacker



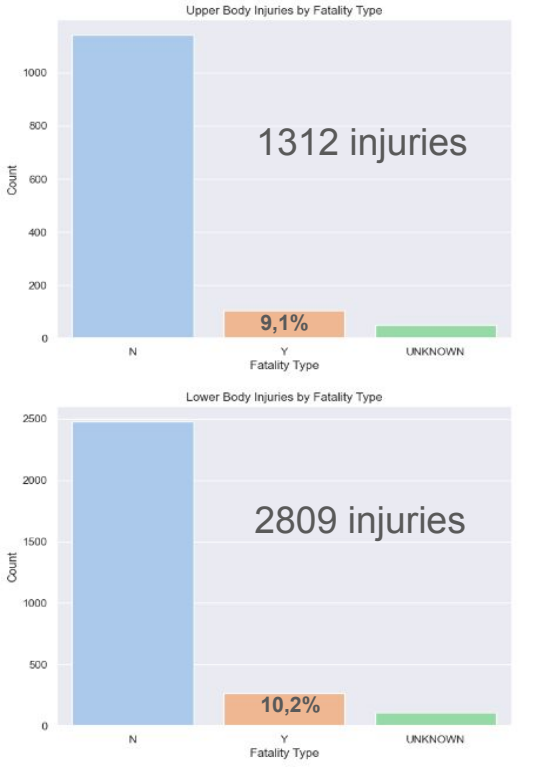
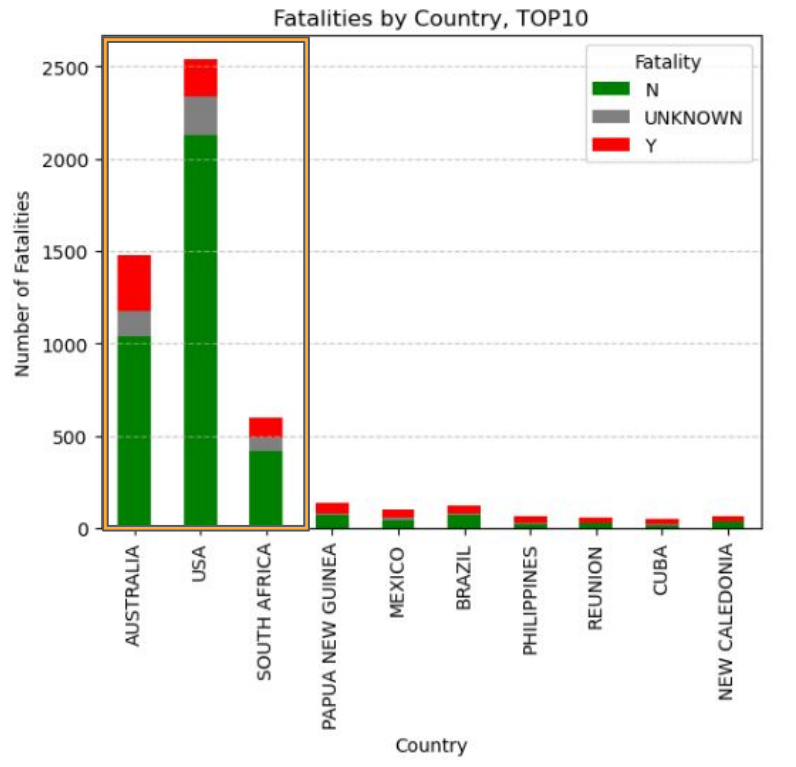
Geographical coverage:

- **White sharks** were by far the most common attackers.
- Many attacks were **unconfirmed** to involve sharks at all.
- Whites occupy **a broad swathe of geographic territories**, such as off the coasts of the northeastern and western United States, Chile, northern Japan, southern Australia, New Zealand, southern Africa, and the Mediterranean.

Business conclusion:

- **Species distribution** should not have undue impact on market choice.

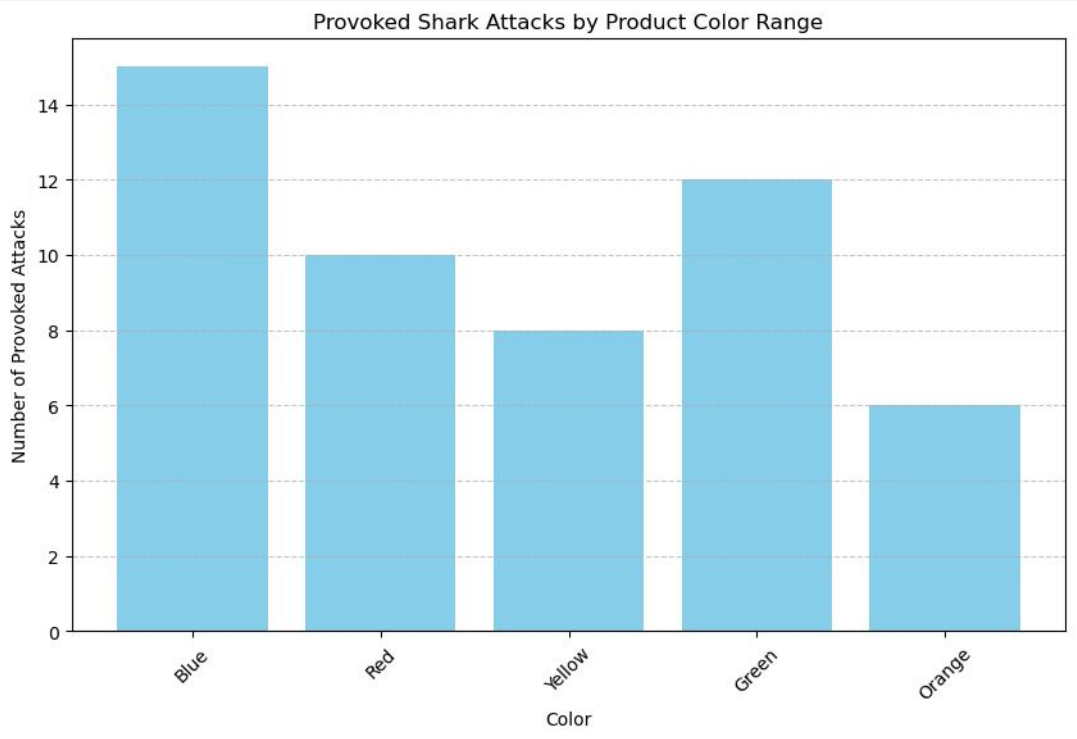
Comparing fatal attacks in different countries and injuries



Conclusions:

- Research indicates that the majority of fatal attacks are concentrated in the top three countries with the highest number of attacks.
- Additionally, findings suggest that both lower and upper body attacks exhibit a similar fatality rate.

Provoked attacks could give hints on product color range selection



Conclusions:

- Examining the occurrence of provoked shark attacks in relation to product colors could offer insights and make an effective decision related to product design, marketing, or safety measures.
- However, the low number of provoked attacks may affect the statistical significance of the data.

Thank you!

SURFHACK
Ride fearlessly



1. **Title Slide** (1 slide): Your project's name and your names.
2. **Project Overview** (1-2 slides):
 - Briefly describe the original dataset and the hypothesis you've formulated.
 - Explain the structure and process of your data cleaning and analysis.
 - Highlight any unique data cleaning techniques or methods you've employed.
3. **Data Wrangling and Cleaning** (1-2 slides):
 - Discuss the significant data cleaning challenges you encountered (missing data, duplicates, formatting issues, etc.).
 - Explain how you resolved these challenges.
4. **Exploratory Data Analysis** (1-2 slides):
 - Discuss the exploratory data analysis methods you used.
 - Share insights and interesting patterns you found.
5. **Major Obstacle** (1-2 slides):
 - Discuss the biggest obstacle or mistake you encountered during this project.
 - Share what you learned from it and how it influenced your project.
6. **Conclusion and Insights** (1-2 slides):
 - Discuss whether your initial hypothesis was supported or refuted.
 - Share any surprising insights or findings.
 - Discuss potential implications of your findings.
7. **Closing Slide** (1 slide): Your project's name, your names, and a "Thank You" message.
8. **Total:** 7-10 slides.