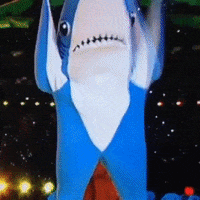
# Shark Attack

### **Day 1**



**Problem Statement:**

Shark attacks have long captivated public interest and spurred scientific inquiry. Despite widespread fascination, a comprehensive understanding of the species responsible for these incidents remains elusive. Our dataset comprises documented shark attacks, providing a valuable resource for exploring patterns and trends in shark-human interactions.

**Hypothesis:**

We hypothesize that the majority of documented shark attacks are not perpetrated by Great White sharks. Additionally, we propose that the majority of shark attacks are not fatal. This hypothesis challenges the prevailing perception of Great Whites as the primary perpetrators of shark attacks and suggests the presence of other species as significant contributors to human-shark encounters.

**Objective:** The primary objective of this data analysis is to investigate the veracity of our hypothesis by examining the distribution of shark species involved in documented shark attacks. By analyzing the dataset, we aim to discern whether Great Whites indeed dominate the roster of shark aggressors or if other species exhibit comparable or greater involvement in such incidents.

**Approach:**

**Data Collection:** Gather and curate a comprehensive dataset of documented shark attacks, encompassing variables such as date, location, victim demographics, severity of attack, and, crucially, the species of the attacking shark.

**Data Preprocessing:** Cleanse and preprocess the dataset to rectify inconsistencies, missing values, and outliers, ensuring its suitability for rigorous analysis.

**Exploratory Data Analysis (EDA):** Conduct exploratory analysis to gain insights into the overall distribution of shark attacks, identifying geographic hotspots, temporal trends, and victim profiles.

**Statistical Analysis:** Apply statistical methods to assess the significance of the observed distribution of shark species involved in attacks. Utilize hypothesis testing to evaluate whether the proportion of Great White involvement differs significantly from alternative species.

**Conclusion:** Draw conclusions based on the empirical evidence derived from the analysis. Evaluate the support for or against the hypothesis and discuss implications for our understanding of shark behavior and risk mitigation strategies.

## Roadmap

### **Day 2**

**Dev Team Key Tasks & Topics:**

* Clean & format everything but dates.
* Apply at least five data cleaning techniques, such as handling null values, dropping columns, removing duplicated data, manipulating strings, formatting the data, etc., to prepare the dataset for analysis.

**Research Team Key Tasks & Topics:**

* Read the self-guided material on Regular Expressions (Regex) in Python if you haven’t done so before. Explore how Regex can be utilized to extract or manipulate strings in a dataset.
* Research on the patterns which might be relevant for this particular dataset (like patterns in location names, activity descriptions, etc.).
* Read the self-guided material on Dates in Python. Explore how manipulate dates in a dataset.
* Share your findings with the Dev team. Apply Regex in the data cleaning process. If you need the date for your analysis, apply the appropriate techniques to the dataset.

### **Day 3**

**Dev Team Key Tasks & Topics:**

* Finish cleaning and formatting the data.
* Structure the data to make it suitable for analysis. This can involve creating new fields, grouping data, or creating pivot tables.
* Use aggregation and filtering techniques to further refine the dataset and to start drawing conclusions about your hypotheses.

**Research Team Key Tasks & Topics:**

* Read the self-guided material on Exploratory Data Analysis (EDA), focusing on the use of statistical measures, and various types of visualizations (histograms, box plots, scatter plots, etc.) to start understanding the trends and patterns in the data and support your findings.
* Understand how these techniques can help validate hypotheses and provide insights into the dataset.
* Share findings with the Dev team, decide and incorporate basic EDA methods and visualization tools (such as graphs) into the project.

### **Day 4**

On the last day of the quest, you will divide the remaining tasks however you prefer as a team.

**Dev Team & Research Team Key Tasks & Topics:**

* Finalize all cleaning, transformation, and analysis tasks.
* Finish doing EDA and use the insights gained from it to validate or disprove the initial hypotheses.
* Finish refining your code according to best practices.
* Prepare a visually appealing presentation with minimal text to effectively communicate the insights and conclusions to stakeholders. Make sure to build a compelling narrative that highlights the significance of your analysis