Quest: Shark Attacks

Ana Cañizares, Samyak Anand, Moses Oloke

22.04.2025

Project Overview

Steps:

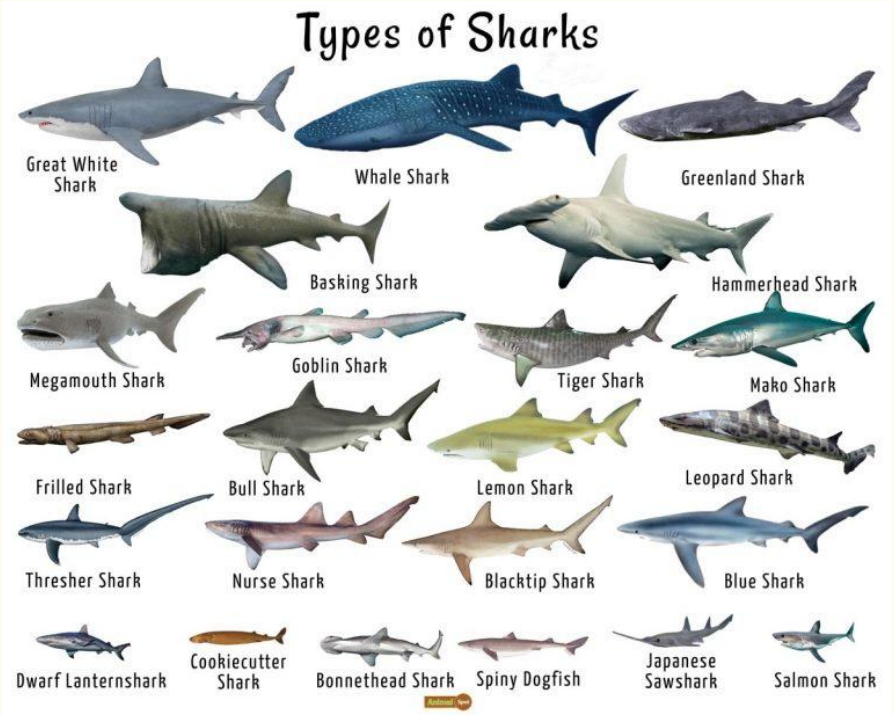
- Examining the Shark Data Set
- Formulating the Hypothesis
- Implementation of Cleaning Techniques
- Exploration of Data to Validate Initial Hypothesis

Hypotheses

- Most shark attacks involve male victims.
- Most shark attacks are unprovoked.
- Shark attacks have increased over the last 50 years.

Data wrangling and Cleaning

1. Strip and trailing whitespace from column names
2. Drop duplicate rows to ensure unique entries
3. Convert all the column names to lowercase for consistency
4. Print confirmation message after cleaning
5. Return the cleaned dataframe



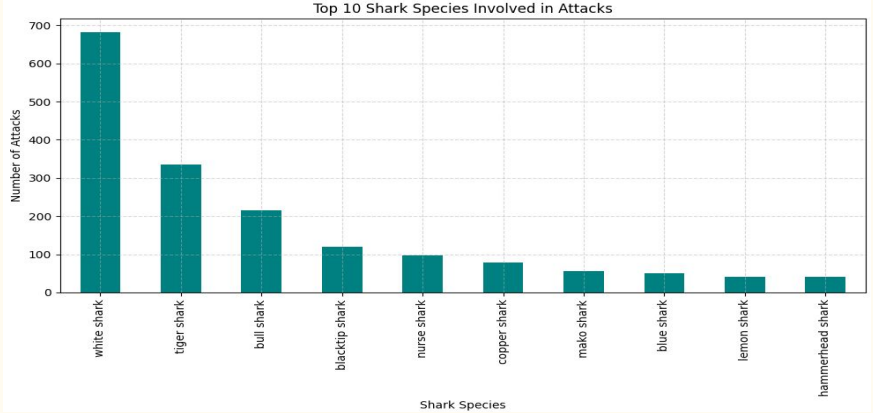
Exploratory Data Analysis

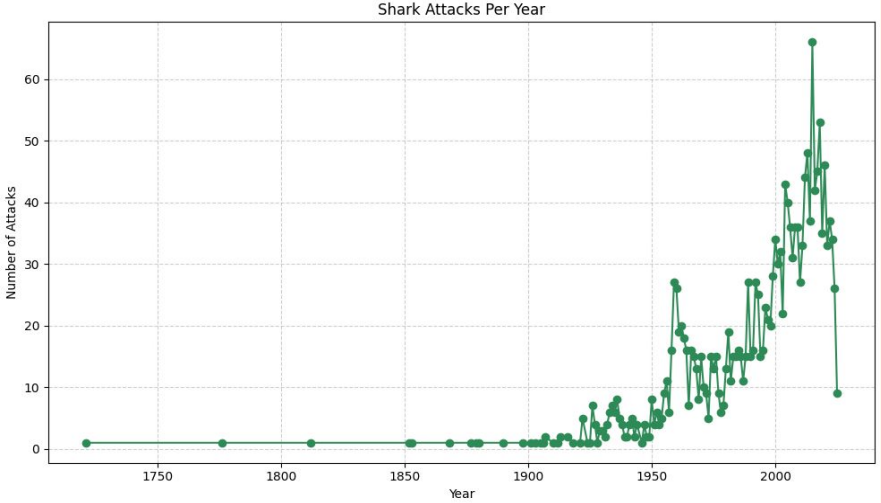
- Understanding the structure of the data
- Data cleaning check
- Summarizing the statistics
- matplotlib
- Seaborn
- Visual exploration.

Exploratory Data Analysis

Insights and interesting patterns found during the analysis:

- The above bar chart illustrating the frequency of shark species involved in attacks.
- The line plot below, showing the number of attacks over time by year.

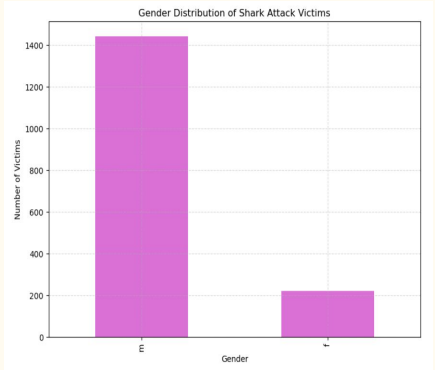


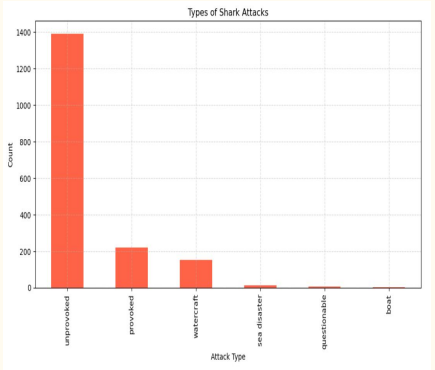


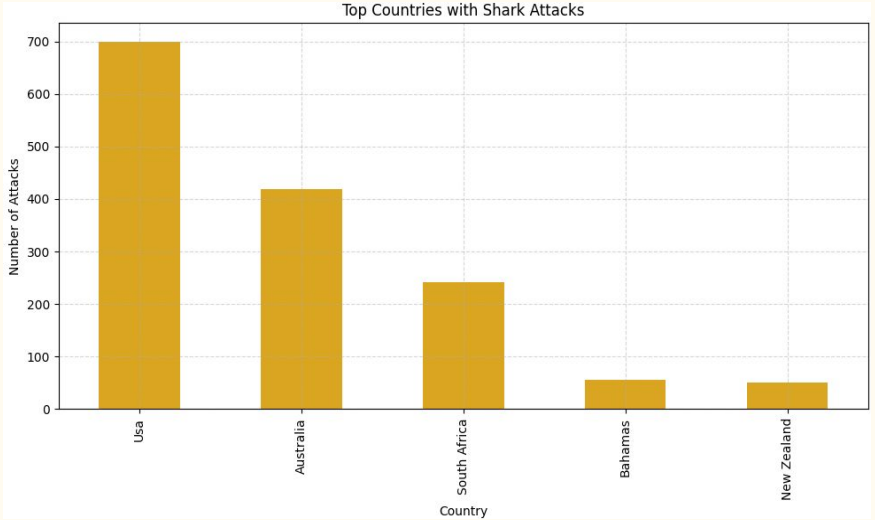
Exploratory Data Analysis

Insights and interesting patterns found during the analysis:

- Above left bar chart representing the gender distribution of victims.
- Above right bar chart showing the categorization of incident types.
- Below bar chart highlighting the top five countries with the highest number of reported attacks.





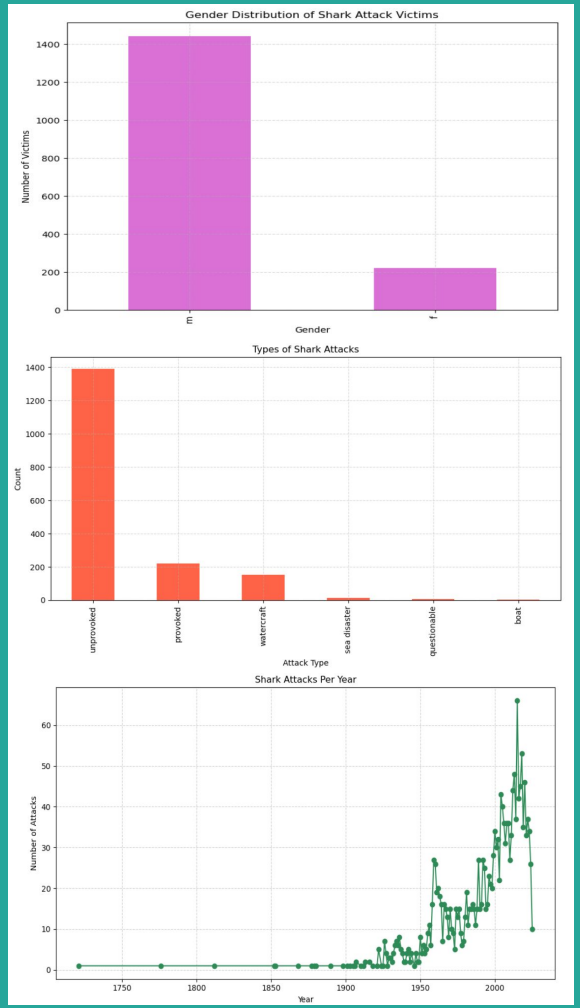


Major Obstacle:

- **String manipulation issues** when standardizing the "Species" column.
- **Initial method only replaced specific words**, leading to incomplete modifications.
- **Inconsistent species names** retained fragments instead of full replacements.
- **Solution**: A custom function **checked entire cell values** before rewriting them.
- **Outcome**: Improved **data consistency** and eliminated partial mismatches.
- **Key lesson: Conditional logic is essential** for effective data cleaning.

Conclusions & Insights:

- *Most shark attacks involve male victims.*
  - **Male victims** (m) represent **80.22 %** of the attacks, with 1,436 cases.
  - The top graph shows that the data supports our hypothesis.
- *Most shark attacks are unprovoked.*
  - **Unprovoked** attacks account for **77.71 %** of all attacks, or 1,391 cases.
  - In the middle graph we see that our hypothesis is confirmed.
- *Shark attacks have increased over the last 50 years.*
  - Shark attack incidents have shown a significant increase over time, highlighting a long-term rise in recorded cases over the decades. **Between 1975 and 2024**, there were 1,353 recorded attacks, making up **75.59 %** of all reported cases since 1721. In contrast, the 1925 to 1974 period saw only 397 attacks.
  - In the bottom graph we see that our hypothesis is backed up by data.



Demo (repository)

Thank you!