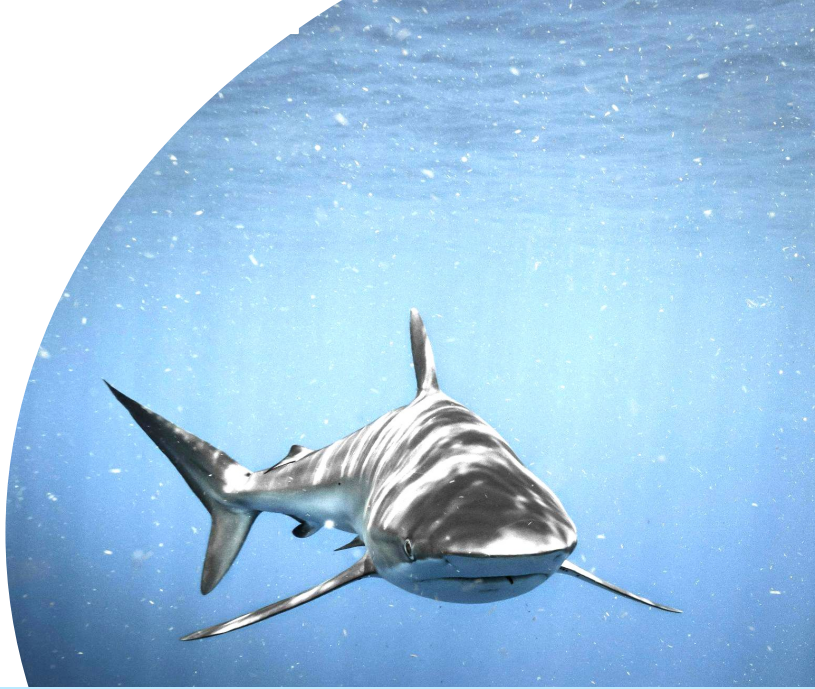
IRON HACK



“3-Headed Sharks”

Decoding Shark Risk:

A Global Intelligence Report for Coastal Industries

Insights from 7,000+ Global Shark Incidents (GSAF)

Project Overview

Why This Project Matters:

7,000+ incidents globally

80% male victims

70% linked to water sports

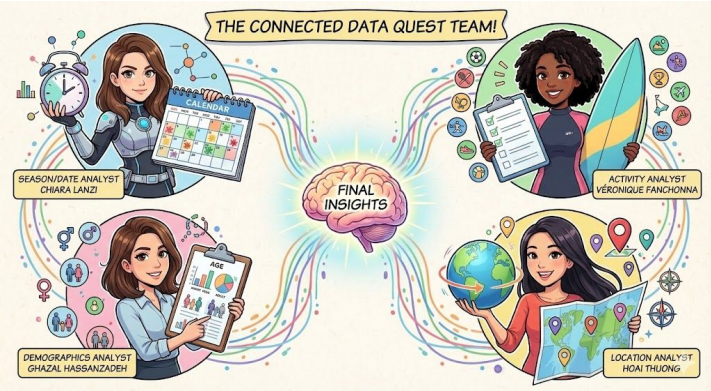
Summer is peak season

Dataset:

Source: **Global Shark Attack File (GSAF)**

Contains **7,000+** global shark incident records

Analyzed variables included season, Year, Activity, Sex, Age, Location and Country



Our Hypothesis

Shark incidents follow predictable patterns based on demographics, location, season, and activity type. These patterns can help coastal industries make smarter safety decisions.

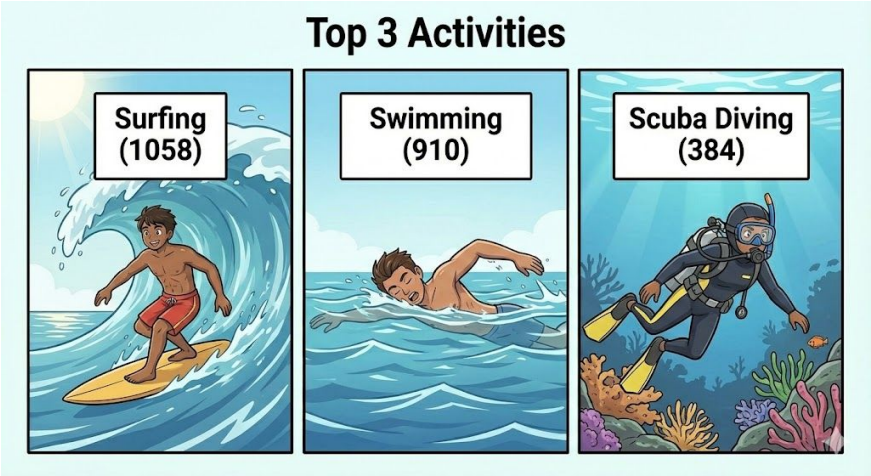
Project Overview

Project Goal

Our goal was to transform raw shark incident data into clear and useful dataset. As a consultancy company specializing in nature and human–wildlife interaction, we want to replace fear and myth with facts. Our work supports:

- **Animal protection associations** — communicating accurate, science-based information
- **Tourism industries** — improving guest experience and safe behavior
- **Healthcare providers** — planning staff around seasonal risk peaks
- **Government agencies** — informing and protecting coastal communities
- **Our mission** is to use data to strengthen safety, awareness, and coexistence between humans and marine wildlife.

Data Wrangling & Cleaning



Significant Data Cleaning Challenges

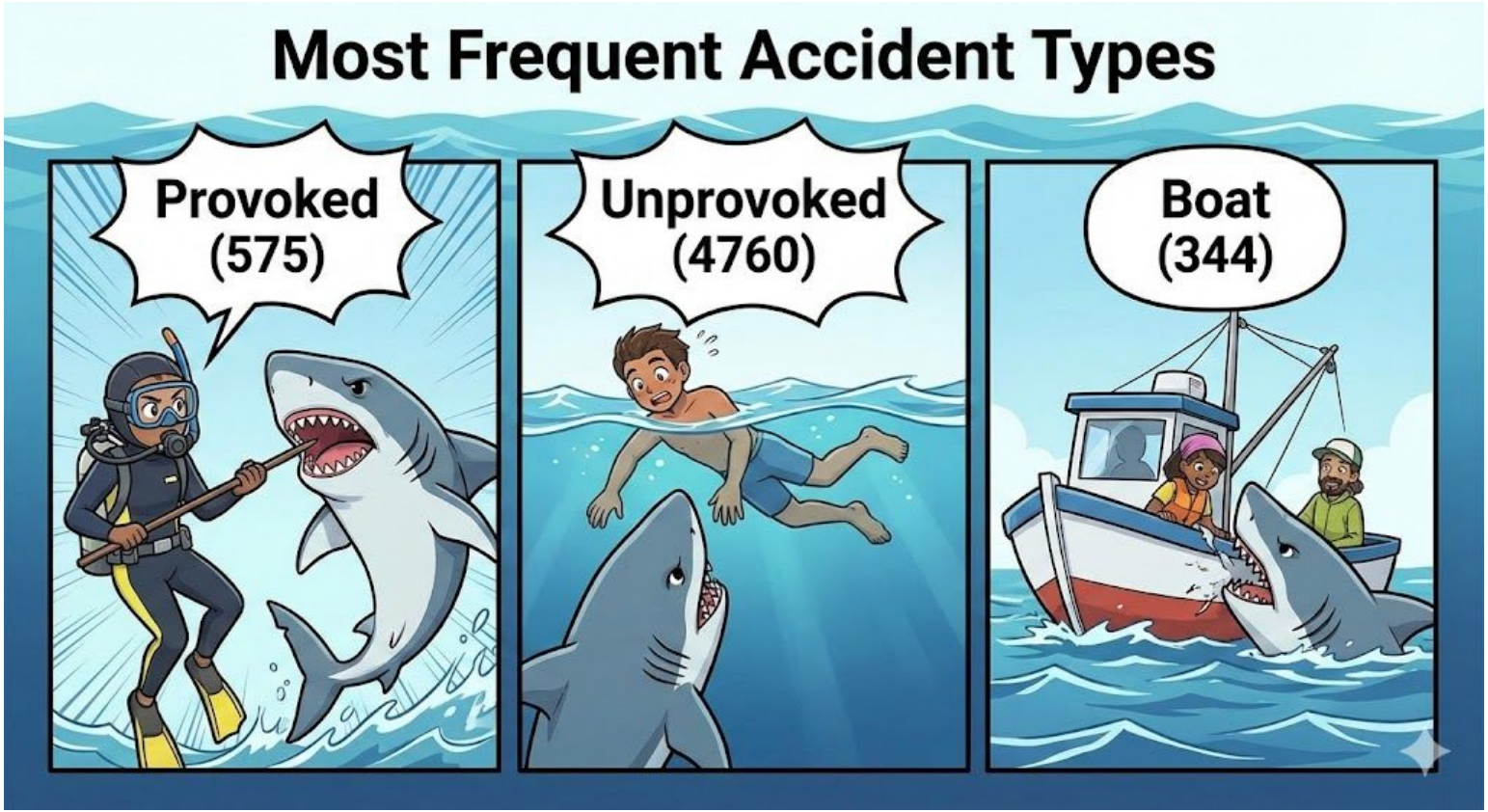
- Missing data
- Inconsistent formats
- Duplicates
- Wrong values, Outliers
- Scattered text formats with delimiters
- Swapped columns' data

How We Resolved Them

- **Normalization**: astype, str.lower, str.strip
- **Fixing wrong values:** replace,
  Regex extraction: str.extract, to_numeric, re_match
- **Grouping into categories**: np.select, map,apply
- **Handling missing values**: fillna, dropna
- **Removing duplicates**: drop_duplicates
- **Reset_index**
- **For loop, If statements**
- **idxmax**
- **set up assumptions to filter out obsolete data**

Exploratory Data Analysis

	Yearly Trends	• In 2015, 2% of total accidents took place. due to increasing tourism trends. The number of shark attacks grew steadily YoY. Approx 56% of cases happened from 1900 until today .
	Location	• New Smyrna Beach (Florida) shows the highest number of incidents in a single year (3% Tot share). To note: 8% of location is unknown. • USA accounts for 37% shark attacks followed by Australia with 22% and South Africa 9%.
	Activity Category	• Surfing has the most incidents, especially in 2012 account for 73% of tot Activities. • High-risk activities involve deeper water and strong surface movement.
	Demographics	• About 80% of victims are male. • the share of of young demo group being affected takes 16% of Tot and teen 15%.


Most Frequent Accident Types
Provoked
(575)
Unprovoked
(4760)
Boat
(344)

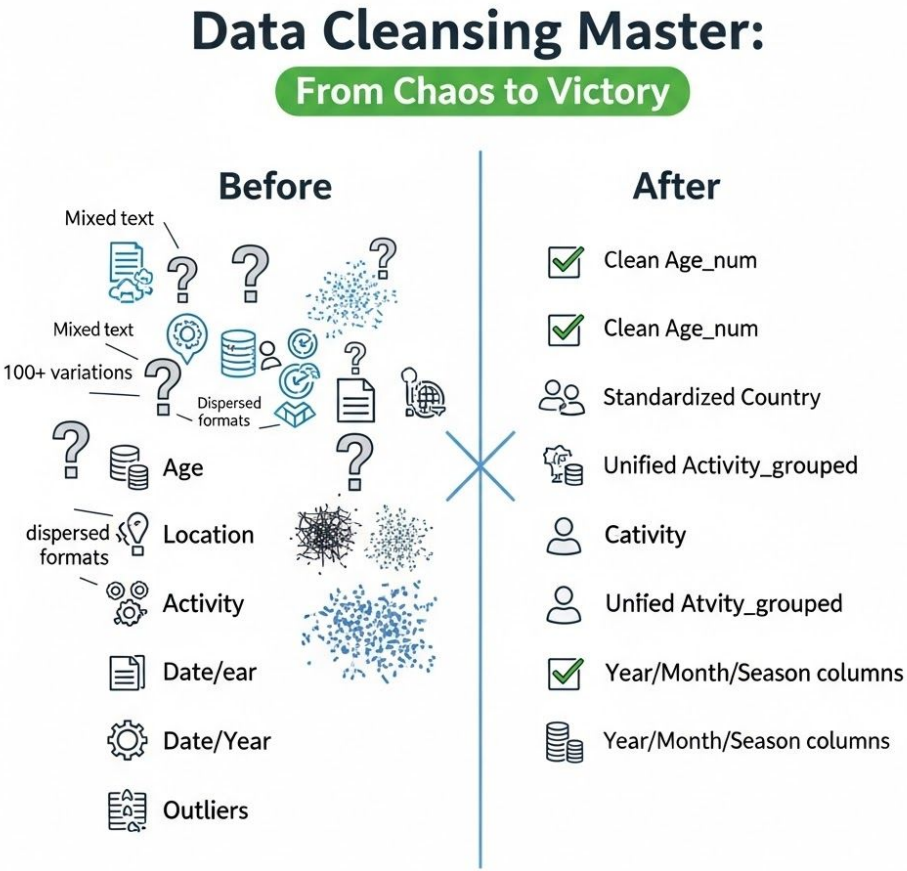
Major Obstacle

Key Domain	Challenge Faced ("The Chaos")	Applied Solution ("The Victory")
Data Consistency	**Age:** Heterogeneous values (mixed text, question marks, ranges, words).	Creation of an **Age_num** column using regex and definition of **Consistent Age Groups** (minor, teen, adult...). over 40% of the values in this column were missing
Standardization	**Location and Date (year, month):** Cities, regions, and countries mixed in the same cell.	Implementation of standardized formats for **Country & Location** for reliable geographical analysis likewise for Month and Year.
Text Normalization	**Activity:** Over 100 variations, symbols, typos.	Unification into a single **Activity_grouped** column with only 6 main activity types after intensive string cleaning and standardization of one single string..
Format Management	**Date & Year:** Dispersed formats (6094 unique date values, approx. 260 unique year values).	Extraction of **Year, Month, Year_Group** columns using regex and derivation of a **clean Season** column from the date.
Outlier Management	Presence of irrelevant or aberrant years (e.g., < 1500).	Implementation of a rule to reset or exclude values outside the study scope (**Out of scope Year < 1500 reset_index**).

Major Obstacle

The Human Qualities Behind the datas

- **Resilience to Ambiguity:** Ability to transform raw data chaos. Familiarity with the data-frame is key.
- **Analytical Rigor & Quality Control**: Implemented standardization and outlier management rules, ensuring data integrity and reliability for the final analysis.
- **Solution Focus & Business Impact:** Constant focus on business value by making data immediately actionable (demographics, season)
- **Synthesis & Clarity** : Aptitude for translating complex technical work into a simple narrative, facilitating stakeholder decision-making.



Conclusion

Our core hypothesis: shark attacks do not happen evenly over time and place, instead, there are specific periods, locations, and populations that face higher risk.

- ## Major findings

Month: Major shark attacks (approx > 70%) occurred during the warm summer months when more people go to the beach.

Year Group: The number of shark attacks grew steadily YoY. Approx 56% of cases happened from 1900 until today as ocean activities increased.

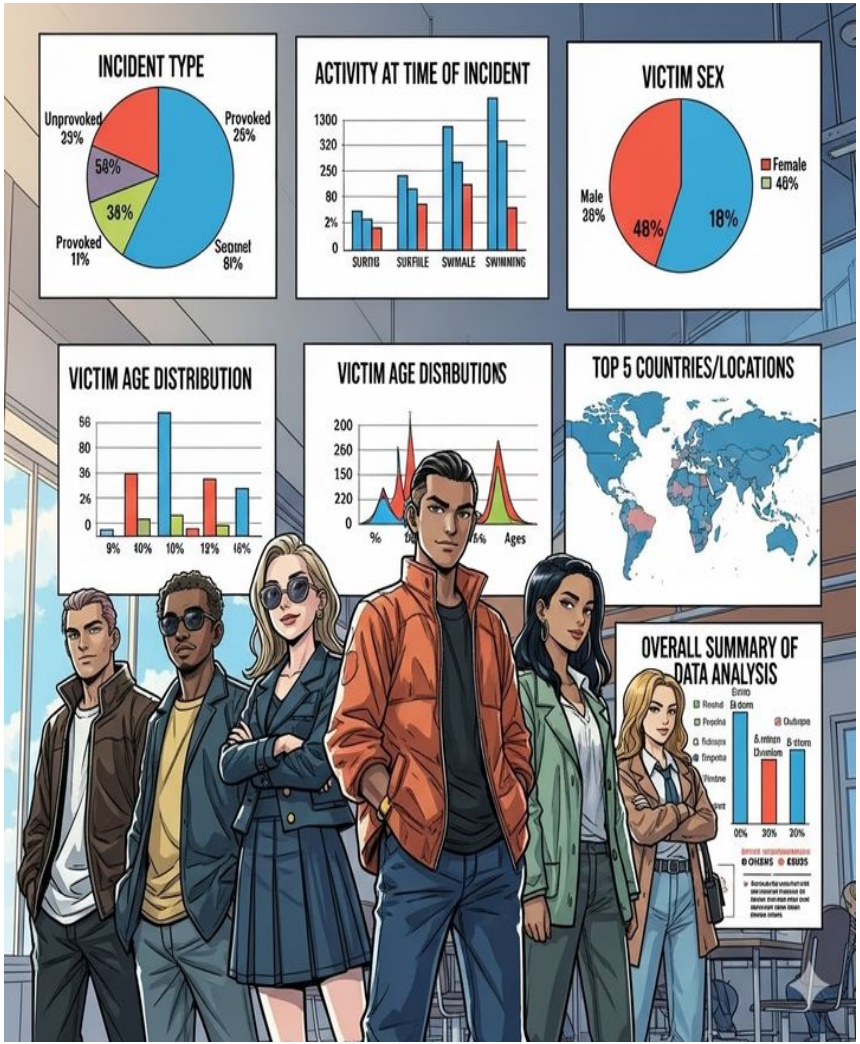
Country: Top 3 countries are USA, Australia, and South Africa given the presence of beaches and many water activities.

Location: Florida accounts for 2494 shark attacks. South Africa for 1448

Activity: 73% of shark attacks happened during surfing followed by swimming

Sex: male injuries account for 80% driven by activities such as surfing and diving.

Age Group: Young people (20-30y) accounted for 16 % of accidents followed by teens (15-20y) for 15%.

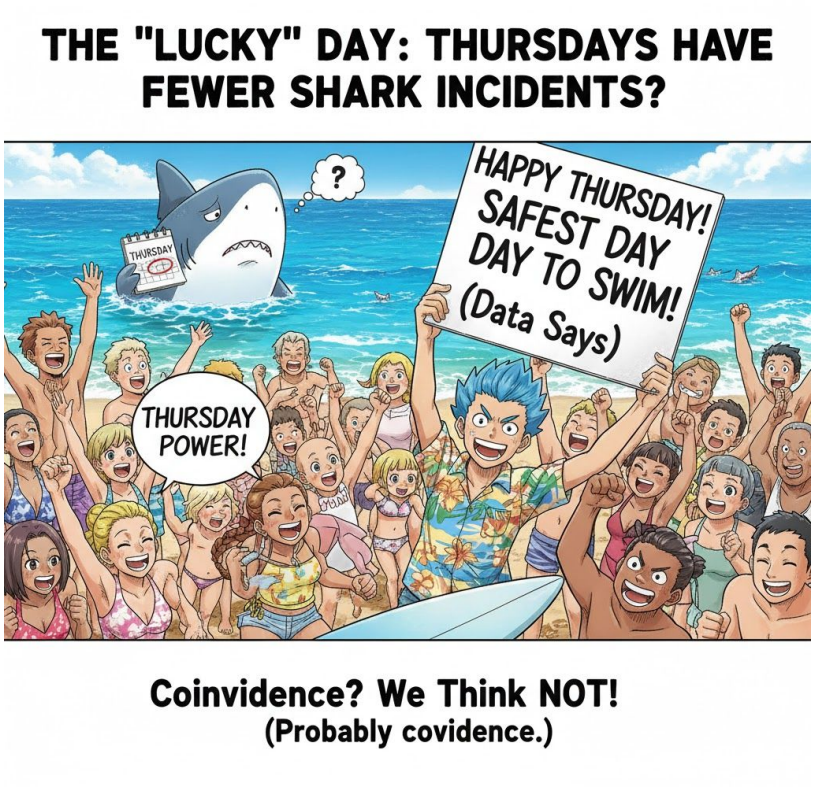


Recommendation

- Based on these patterns, we recommend focusing safety and public-awareness efforts on **hotspot coastal areas** and **peak seasons** where water activity is frequent.

- We advise local authorities and beach managers to ensure **clear warnings, education campaigns**, and s**afety guidelines** (e.g., avoid swimming at dawn/dusk, avoid murky water, avoid lone swimming) for visitors.

- Encourage surfers and swimmers to be particularly cautious and **stay informed on shark-presence reports** when visiting high-risk zones



THANK YOU

for your attention!

Presented by:

Veronique Fanchonna

Ghazal Hassanzadeh

Chiara Lanzi

Hoai Thuong

Ironhack Data Analytics Bootcamp — Nov 2025— Mini Project 1