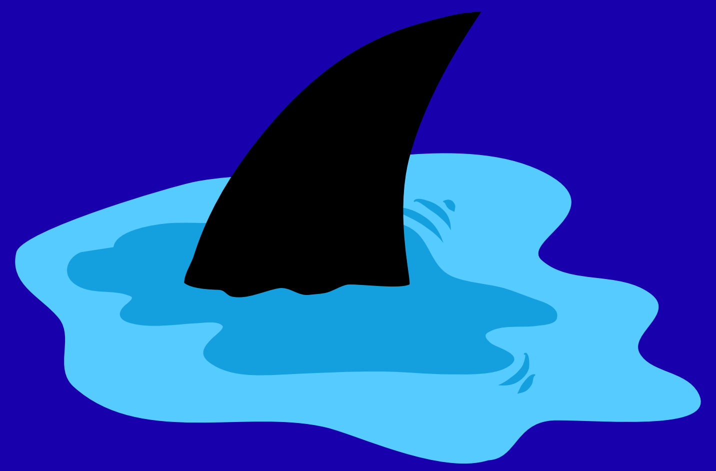
IRON HACK

SHARKSAFE:

ENHANCING DIVER SAFETY AND EXPERIENCES THROUGH PREDICTIVE MAPPING OF SHARK ENCOUNTERS

Presented by
Bruno Sousa, Gonçalo Trindade
Rachel Vianna, Sarah Jane Nede



PROJECT OVERVIEW

Business Proposition: A company specializing in mapping services aims to enhance diver safety and experience by providing detailed insights into shark encounters. We wish to empower divers to make informed decisions, dramatically enhancing both safety and the quality of the diving experience.

Problem Statement: This project seeks to predict the likelihood and nature of shark encounters at specific locations and times, educating divers on where and when to dive for desired sightings while minimizing safety risks.

Data Set: The Dataset utilized for the SharkSafe project is sourced from the Global Shark Attack File Incident Log, a comprehensive compilation of shark encounter and attack incidents worldwide. It includes key fields such as Date and Time for temporal analysis, Year for trend tracking, and Type to categorize encounters. Geographic indicators (Country, State, Location) pinpoint where incidents occur.

HYPOTHESIS 1

Primary Hypothesis: Shark attacks are more likely to occur around midday due to increased human and shark activity in the water, thus informing crucial safety advisories for divers to minimize unforeseen risks.

- Objective: Determine if shark attacks are more common around midday.

- Approach:
  - Quickly explore the dataset for a timestamp or time-related column.
  - Create a pie chart or bar plot showing attack frequencies by time of day.
  - Analyze peak periods and assess if there's empirical support for increased attacks around midday.

HYPOTHESIS 2

Secondary Hypothesis: Shark visibility is highest in January due to seasonal migration patterns and optimal water clarity, which informs safe and rewarding diving experiences for enthusiasts seeking specific encounters.

- Objective: Evaluate shark visibility patterns in January.

- Approach:
  - Filter or plot data points specifically for January around the world.
  - Examine patterns in species observations and encounters during this timeframe.
  - Use geolocation data to visualize sightings spread over the relevant regions.

DATA WRANGLING AND CLEANING

Key Data Cleaning Challenges:

1. Data Optimization: To enhance clarity and relevance, we removed extraneous columns. Our focus was on identifying the temporal and locational aspects of the attacks. Consequently, we retained only the essential columns: dates, times, locations, species involved, and fatality information.

2. Column Identification for Cleaning: In order to substantiate our hypotheses, it was necessary to pinpoint specific data attributes for refinement. We concentrated on the seasonal and daily timing of the incidents, as well as the geographical locations where attacks occurred. The columns subjected to cleaning include:
    - Date
    - Time
    - Countries

DATA WRANGLING AND CLEANING

Cleaning Details for Key Columns:

- Time:
  - Removed extraneous characters to ensure data consistency.
  - Standardized all time formats for uniformity.
  - Categorized times into defined groups for analytical purposes.
- Country:
  - Standardized country names to ensure consistency, using lowercase with an initial capital letter.
  - Eliminated blank spaces and extraneous characters to maintain data integrity.
  - Ensured consistent spelling for matching countries and filtered data to include only those in the Southern Hemisphere.
- Date:
  - Removed all NaN values to ensure data completeness.
  - Uniformly formatted all entries to maintain consistency.

EXPLORATORY DATA ANALYSIS

What we did:

- Cleaned and standardized messy data (dates, countries, categories)
- Handled missing and non-conformative values
- Created derived columns (Year, Month, Month_name, Time categories)
- Used groupby, counts, percentages, and averages to explore patterns
- Visualized trends with line charts and bar charts
- Compared distributions across: Years (trend over time), Months (seasonality), Time of day (risk periods)
- Tested assumptions by checking correlations between time, month, and number of attacks and validating or rejecting common beliefs using data

Why?:
To understand patterns, trends, and anomalies before validating (or invalidating) assumptions

RESEARCH INSIGHTS

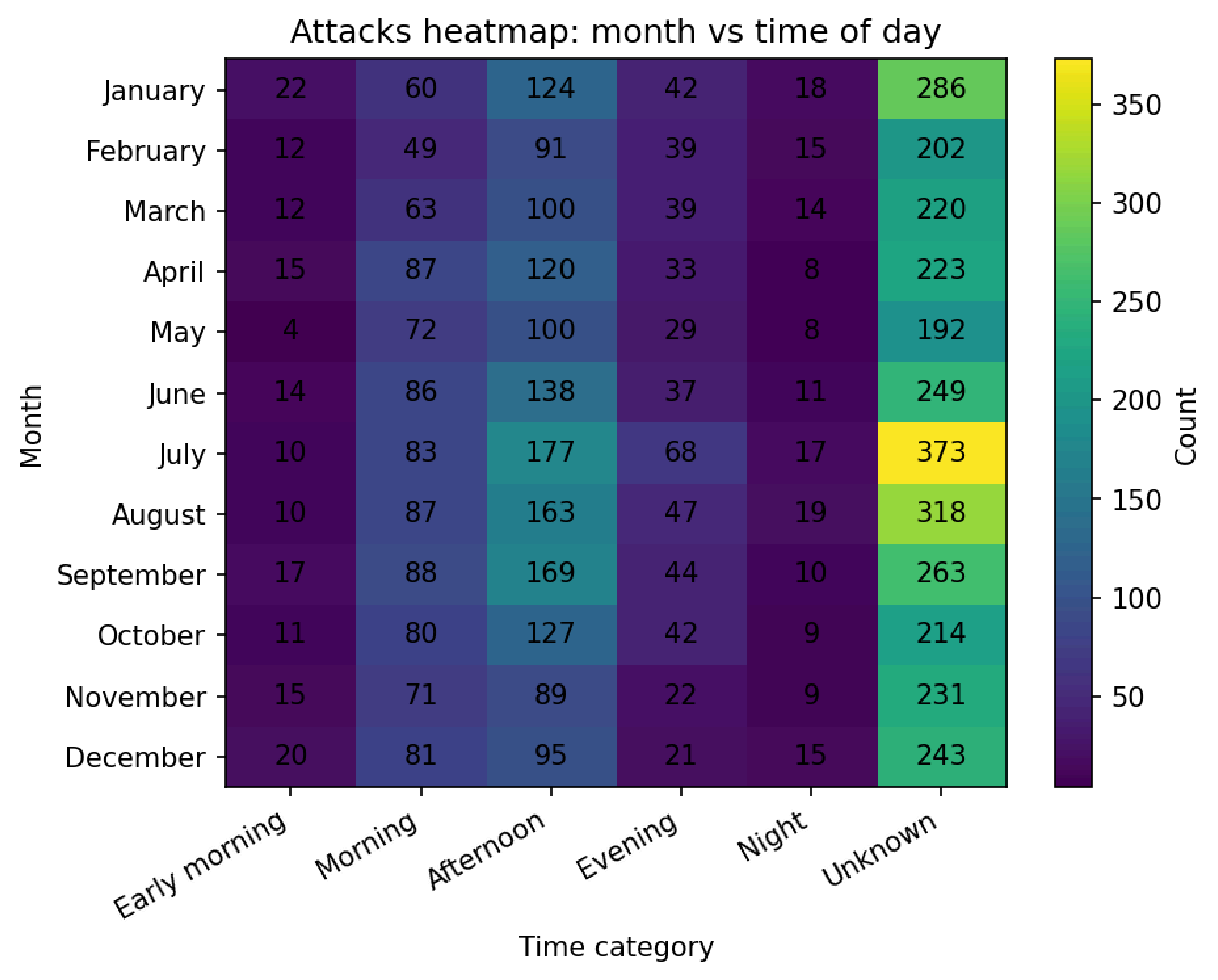
- Shark attacks have increased over the last 40 years, linked to human population growth and increased ocean activity
- July is actually the most dangerous month on average globally
- Contrary to popular belief, the most dangerous time is not dusk or dawn, but midday/afternoon
- The distribution of attacks shows clear seasonal patterns
- The data helped validate and invalidate assumptions instead of relying on stereotypes

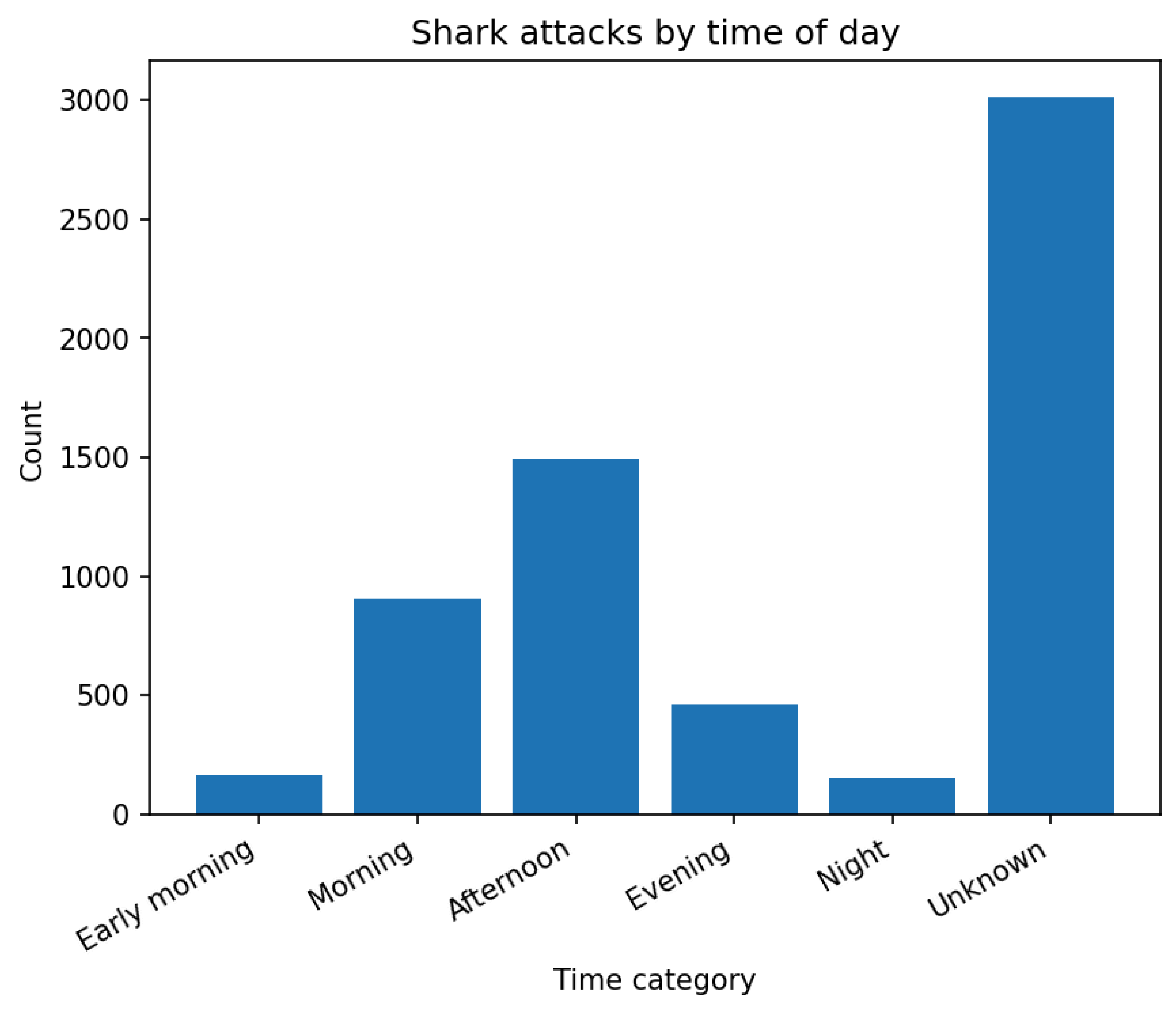
Conclusion:

- EDA revealed strong temporal patterns and showed that human activity and environmental changes play a major role in observed trends
- Safety in diving and shark experiences varies wildly by time of the year

EXPLORATORY DATA ANALYSIS

- Our findings show that most shark incidents happen during full daylight (over 21%). External studies also back this finding in different ways: Source.





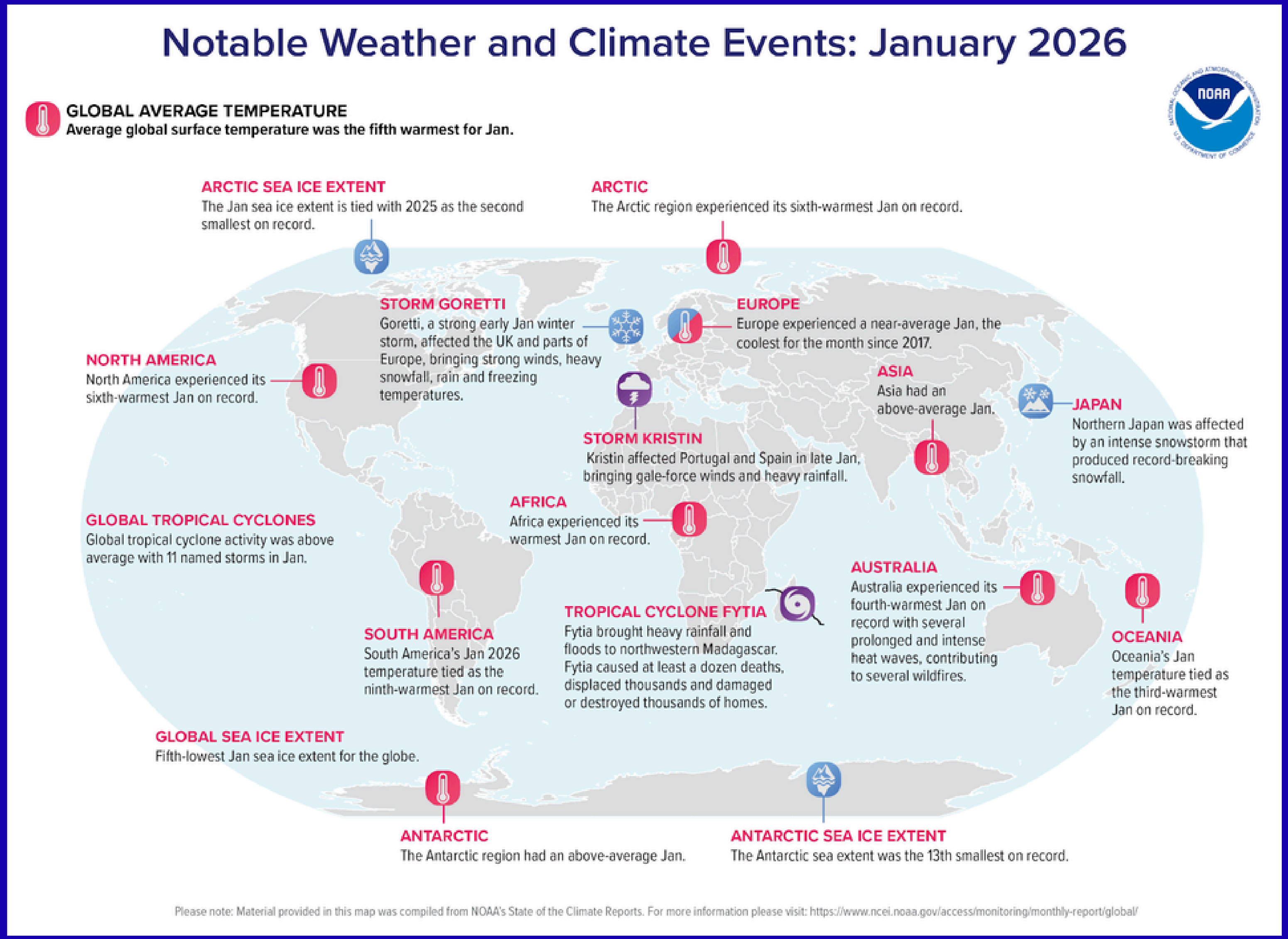
MAJOR OBSTACLE

The data:

- The original dataset was very messy and inconsistent
- Multiple date formats and incomplete dates made time analysis difficult
- Many columns contained missing values (Country, Species, Time, Fatal Y/N, etc.)
- The same information appeared in different formats (e.g. country names, month names, text casing)
- A lot of non-standard values like "Unknown", "?", or empty cells were used instead of real missing values
- Several columns had mixed data types (numbers stored as text, partial dates, etc.)
- Some fields (e.g. Species, Activity, Injury) were free-text instead of structured categories
- The dataset also contained redundant and overlapping columns (multiple date and case number fields)

ENVIRONMENTAL INSIGHTS

- **Impact of Global Warming**: Climate change is increasing the frequency and intensity of extreme weather events, such as heavy summer storms. These storms lead to increased rainfall and subsequent flooding, altering coastal ecosystems by changing salinity, temperature, and visibility. The influx of nutrients and sediment creates murky water conditions that favor species like Bull Sharks, which thrive in low-visibility environments.



CONCLUSION AND INSIGHTS

Validating Diving Safety at Given Times and Seasons:

- **Primary Hypothesis Validation: Time of Day for Shark Attacks**
  - **Finding**: Data analysis confirms that shark attacks are more prevalent around midday, aligning with peak human activity times such as midday when swimming and surfing peak.
  - **Insight**: This correlation guides recommendations for divers to avoid peak attack hours, enhancing safety through better timing awareness.

- **Secondary Hypothesis Validation**: **Shark Visibility in January**
  - **Finding**: In January, visibility conditions are actually decreased due to increasingly extreme weather events. This can lead to a worse diving experience because of both a hindered ability to actually see the sharks AND an overall higher safety risk.
  - **Insight**: Our analysis suggests that diving experiences during January (and July, August and September) will lead to more shark encounters, BUT that this is NOT conducive to tourism, since the chances of an attack will also be much higher as we encroatch onto what is left of an altered natural habit that can only be managed by some of the most dangerous species (like the highly adaptable Bull shark).

IRON HACK

THANK YOU

SharkSafe:
Enhancing Diver Safety and Experiences Through Predictive Mapping of Shark Encounters

Presented by
Bruno Sousa, Gonçalo Trindade
Rachel Vianna, Sarah Jane Nede