

SHARK ATTACK QUEST

CORAL REEFS TEAM



Founded by Paula, Charlotte, Marçal,
Elham & Imtiaz



STEP 1

EXPLORATORY DATA

- What parameters make it more likely to be attacked by a shark?



STEP 2 HYPOTHESIS

H1: Geography Influences Shark Attack Likelihood

- H1a: Shark attacks are more likely to occur in Australia and the United States than in other countries.

H2: Demographic Factors Affect Shark Attack Probability

- H2a: Males are more likely to be attacked by sharks than females.
- H2b: Younger individuals (under 35 years old) are more likely to be attacked by sharks than older individuals.

H3: Activity Increases Fatal Shark Attack Risk

- H3a: Surfing is more likely to result in a fatal shark attack compared to swimming.

STEP 3

CLEANING

Findings

- NANS
- empty rows
- unnamed columns
- numerical columns with string
- Numerical Columns with Anomalies

Techniques

- handling null values in different ways (using median, ffill, assigning them to a value category),
- dropping columns, handling duplicates,
- manipulating strings (i.e. filtering, erasing) and categorizing data

STEP 3

DATA CLEANING CHALLENGES



Objective: Clean and categorize shark attack data.

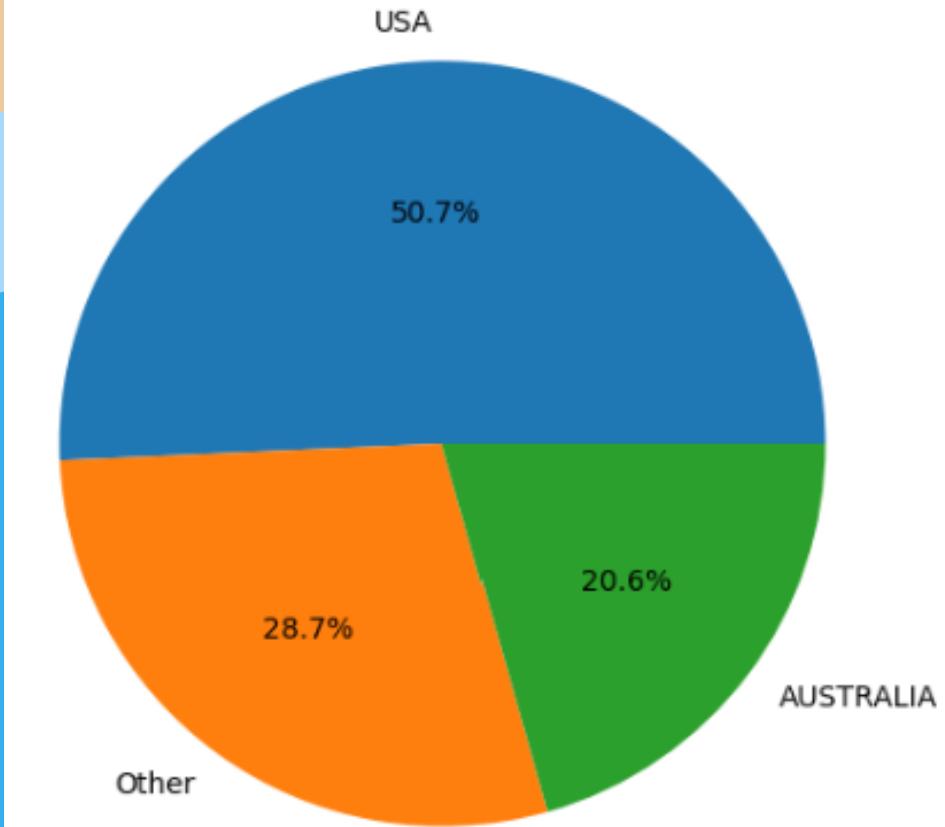
- Key Challenges: Activity Column
- Messy data; created 8 categories (6% "Others").
- Fatality Column: Identified codes ("M" = mortal, "F" = fatal).
- Missing Values: Filled using known mappings.

STEP 4 DATA ANALYSIS

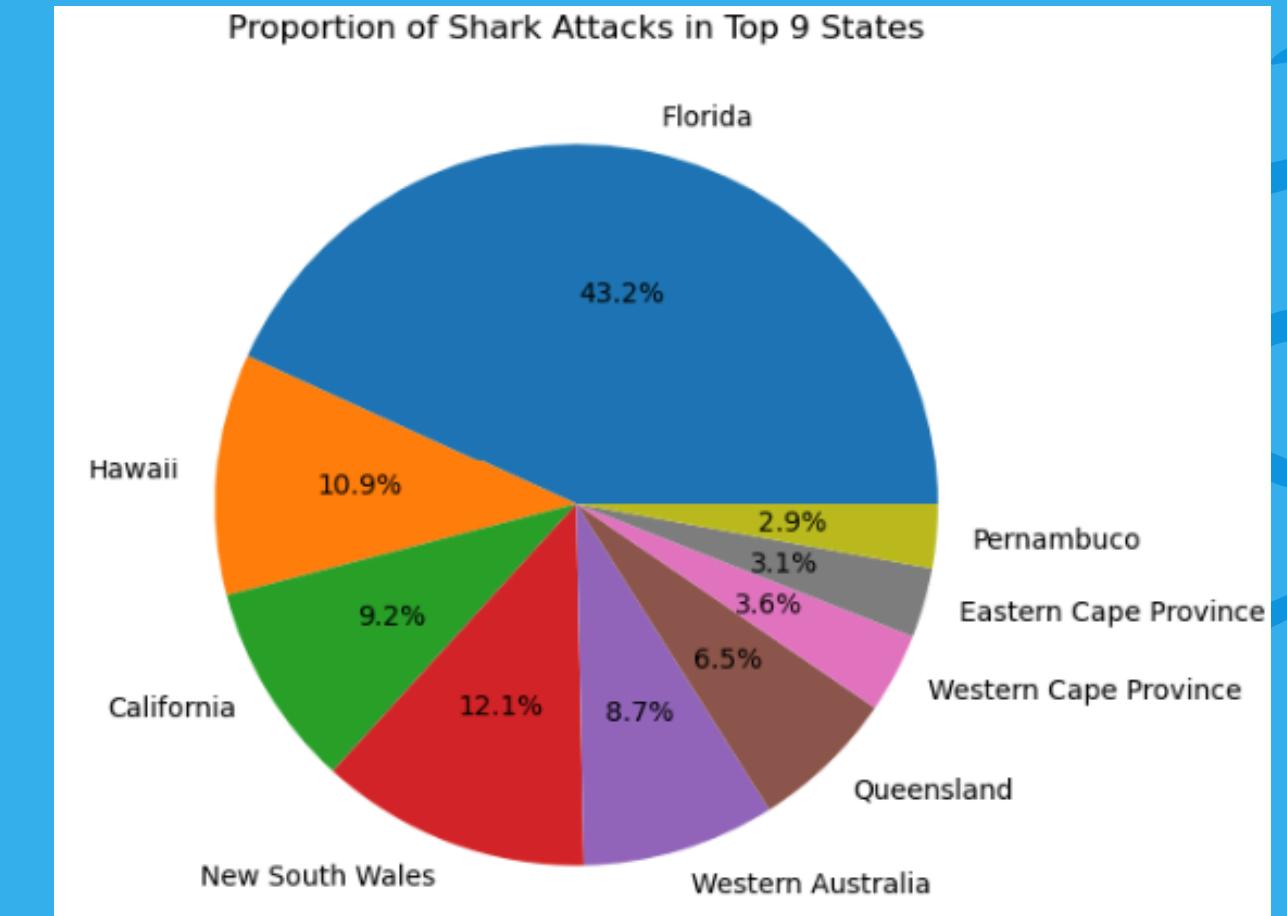
Methods

- Grouping countries into 'USA', 'AUSTRALIA', and 'Other' for all other countries
- Finding the top three states per country and concat them
- pie plot

Proportion of Shark Attacks by countries



Proportion of Shark Attacks in Top 9 States

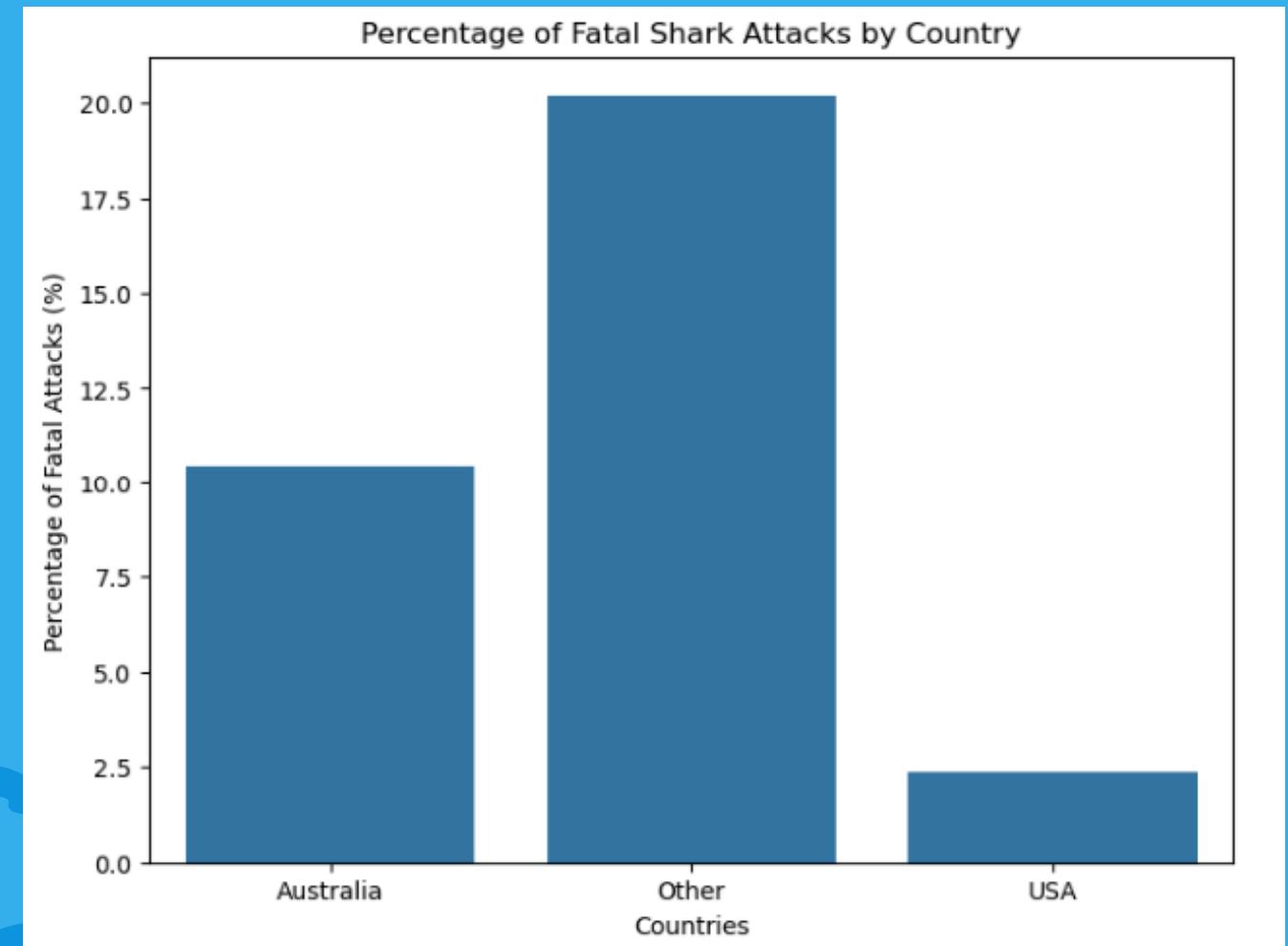


STEP 4 DATA ANALYSIS

Methods and Insights

- 80% male
- median age 28
- Other countries accounted for 20% fatal shark attacks, USA and Australia for 12.5%

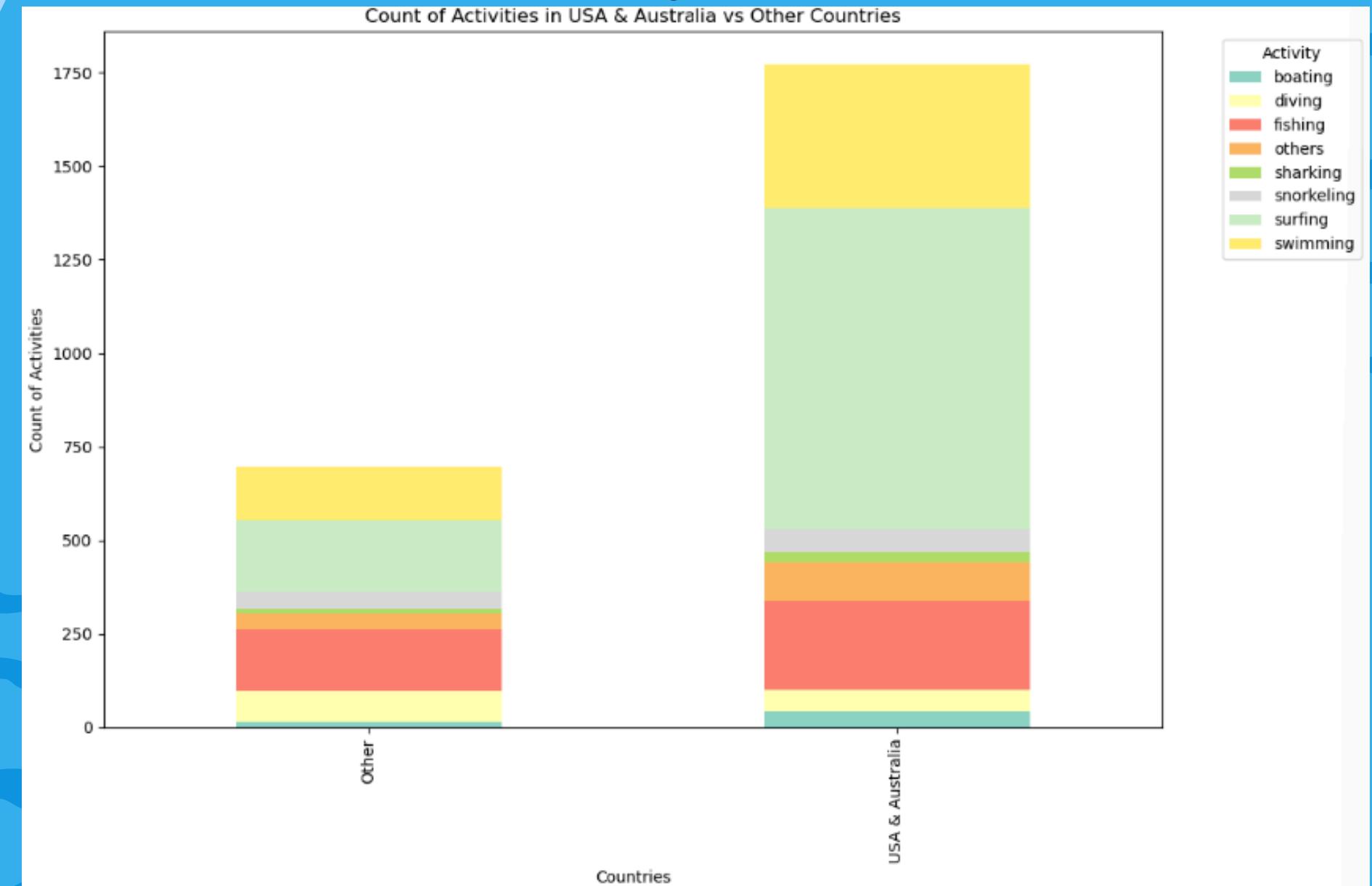
Percentage of Fatal Shark Attacks:
USA, Australia and Other Countries



STEP 4 DATA ANALYSIS



Count of Different Activities in USA, Australia, and Other Countries



Methods and Insights

- Used groupby for country and activity
- Surfing is the activity most prominent
- Second is swimming in USA and AUSTRALIA and fishing on other countries

STEP 4

MAJOR OBSTACLE



- Time-Consuming Fill Process:
 - Extensive effort spent searching through cells.
- Retention of “Age” Column:
 - Numerous NaN's but essential for analysis.
- Choosing the right method:
 - Best approach to handle NaN values, `#fillna()`, drop...
- Balancing the impact of cleaning methods on data quality while ensuring usability.





CONCLUSION & INSIGHTS



- Tip: It's safer to stay at the bar, skip swimming or surfing.
- Where to travel: (% fatality)
 - USA 2,3%
 - Australia 10%
 - South Africa 16%





THANK YOU

