

Are shark attacks rising ?



By:
Natalie Lewis & Federico Sarmiento

1.Project Overview -How does human coastal development and fishing activity influence the behavior of sharks, and is there a correlation between increased human presence in shark habitats and the frequency of shark attacks?

Hypothesis:

We hypothesize that as human invasion and fishing activities intensify in shark habitats, there will be observable changes in shark behavior, potentially leading to an increased frequency of shark attacks. This hypothesis is based on the assumption that alterations in the sharks' natural environment and food sources due to human activities may trigger defensive or predatory behaviors, contributing to a higher incidence of interactions between sharks and humans.

Data:

We utilized the Global Shark Attack File, a publicly accessible database that compiles shark attacks worldwide since 1500.

Data cleaning technique or methods employed

1. **Pre-upload:** 'usecols' to upload only the data necessary for analysis.
2. **General cleaning:** methods such as '.lower' or '.replace' for replacing values in columns, and using 'dropna' to eliminate empty rows.
3. **Funneling process:** created filters based on year (>1800) and country (shark habitats: USA, Australia, South Africa)
4. **Standardization:** the creation of functions and dictionaries to automate the replacement of values in different columns.
5. **Analysis:** use pivot tables, groupby functions, plots, and mathematical methods to derive insights and explore the data comprehensively.

2.Data Wrangling and Cleaning:

Challenges in data cleaning:

Viewing all unique values: Large datasets with unique values within columns, make it challenging to get an overview of the data and identify ways to standardized and/or clean the data.

Automating Cleaning: Creating efficient automated processes for data cleaning tasks.

Excessive Uniqueness: Dealing with columns containing too many unique values, complicating standardization or normalization.

We solve the problem by:



Ask your team



ChatGPT



The teacher



Bird's eye view

Solution:

- Realizing you can access a text will all the unique values. (using Excel or Tableau)
- Creating a function
- Prioritize achieving a comprehensive dataset representation, then remove unique values occurring less than twice for better data quality.

3.Exploratory Data Analysis: -pivot tables, groupby functions, plots, and mathematical methods to derive insights and explore the data comprehensively.

Shark attack by country

	Count	Percentage
country		
USA	2259	40.98
AUSTRALIA	1254	22.75
SOUTH_AFRICA	521	9.45
NEW_ZEALAND	131	2.38
BRAZIL	101	1.83

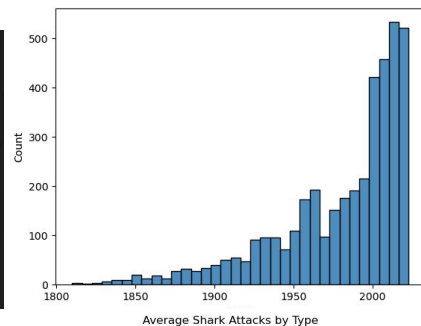
Shark attack by activity

	Count	Percentage
activity		
water activity	3281	59.52
fishing	1318	23.91
underwater activity	597	10.83
work activity	236	4.28
accident	80	1.45

Shark attack by type

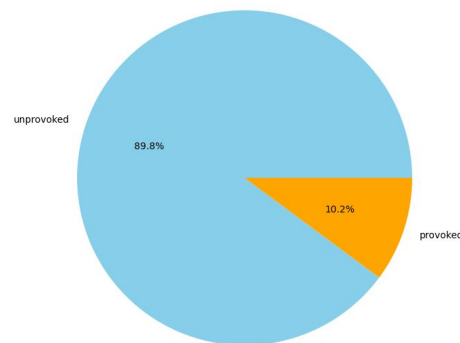
	Count	Percentage
type		
unprovoked	4638	84.14
provoked	524	9.51
invalid	350	6.35

Shark attack since 1800



Principal findings:

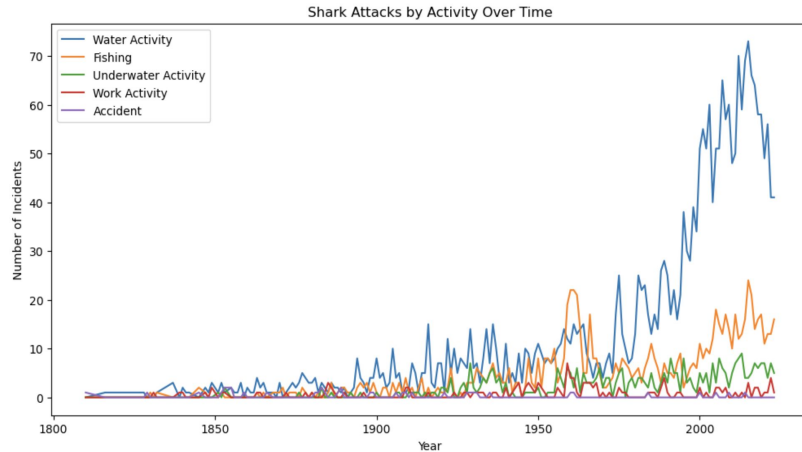
- 73% of shark attacks happened in the USA, Australia, and South Africa, recognized as shark habitats.
- 84% of these attacks are unprovoked incidents, occurring without initial interaction or disturbance.
- Approximately 70% of the attacks took place while individuals were engaged in leisure-related activities.
- Since 1800, there has been a noticeable upward trend in shark attacks.



5. Conclusion and Insights - Shark Habitats:



Principal shark Habitats



AUSTRALIA

State	Attacks
New South Wales	447
Queensland	284
Western Australia	196
South Australia	103
Victoria	84

USA

State	Attacks
Florida	1053
Hawaii	300
California	299
South Carolina	132
North Carolina	105

SOUTH AFRICA

State	Attacks
KwaZulu	184
Western Cape	174
Eastern Cape	156

	activity	accident	fishing	underwater activity	water activity	work activity
country	type					
AUSTRALIA	invalid	1	20	7	52	5
	provoked	1	80	12	18	14
	unprovoked	21	196	165	638	24
SOUTH_AFRICA	invalid	0	11	5	21	0
	provoked	0	32	8	6	3
	unprovoked	0	99	25	283	28
USA	invalid	1	29	22	83	11
	provoked	0	137	17	31	20
	unprovoked	14	293	123	1444	34

Conclusion:

Shark attacks predominantly occur in natural shark habitats, often unprovoked and involving individuals participating in leisure-related activities. This trend appears to be increasing over time.

It might suggest that heightened human presence and fishing activities in these habitats could contribute to the rising frequency of shark attacks.

5. Major Obstacle:

Biggest mistake or obstacles:

- **Visualizing the entire dataset:** Gaining a comprehensive overview of the dataset to understand its nature and contents.
- **Collaborative work using GitHub:** Proficiency in utilizing GitHub for effective collaboration and version control in team projects.
- **Accidentally deleting a code cell:** Mistakenly deleting a code cell due to a misclick, resulting in the loss of valuable code.
- **Changing variable names or cases:** Adjusting variable names or cases, potentially causing confusion or inconsistency in the codebase.

Our learning:

- Utilize tools like VS Code, Excel, or similar programs to visualize a sample of the dataset.
- Establish dataset-specific terminologies such as dataset names, functions, columns, and variables.
- Maintaining a backup copy ensures data safety.

Are shark attacks
rising ?



Thank you for the meal !