Teen Shark Inc.

Protecting Teens from…

doo-doo, doo-doo, doo-doo

Meet the Team: Aisyah Amatul Ghina, Paola Garay, Sasha Crowe, & Bana Kourouma

Project Overview

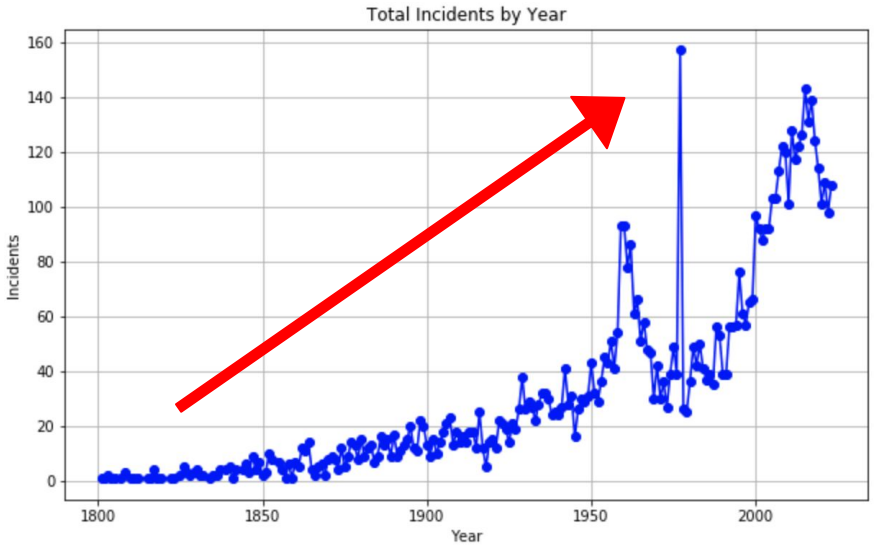
What Teen Shark Inc. Does

Main Objective

> Develop shark-repellent devices & protective gear for teens

Our Approach

1. Preliminary look at data
   a. -> GLOBAL SHARK ATTACK FILE
2. Conducted supporting researching into
   a. Shark attack prevention
   b. MVP development
3. Decided on project scope
4. Adapted as needed



Project Overview

Data Cleaning Process & Techniques

O.G. Dataframe

Look at the shape of the df -> 6969 rows & 23 columns

Cleaning Process

- Review the values in each column & formatting data type
  - eg. Age had strings instead of numeric data
- Keeping only relevant variables (13 variables) for further data cleaning process
- Dropping rows containing all NaN values & removing duplicates
- Understanding the meaning of unique values & make logical groups
  - replacing values where it made sense, eg. Cali & California

Technique to Highlight

```
# Activity
shark_attacks["activity"] = shark_attacks["activity"].apply(lambda x: x.lower() if isinstance(x, str) else x)

shark_attacks["activity"] = shark_attacks["activity"].replace(r'.*fishing.*', 'fishing', regex=True)
shark_attacks["activity"] = shark_attacks["activity"].replace(r'.*diving.*', 'diving', regex=True)
shark_attacks["activity"] = shark_attacks["activity"].replace(r'.*surfing.*', 'surfing', regex=True)
shark_attacks["activity"] = shark_attacks["activity"].replace(r'.*surf.*', 'surfing', regex=True)
```

Data Wrangling and Cleaning | Challenges Faced

Typos

Null Values

Duplicates

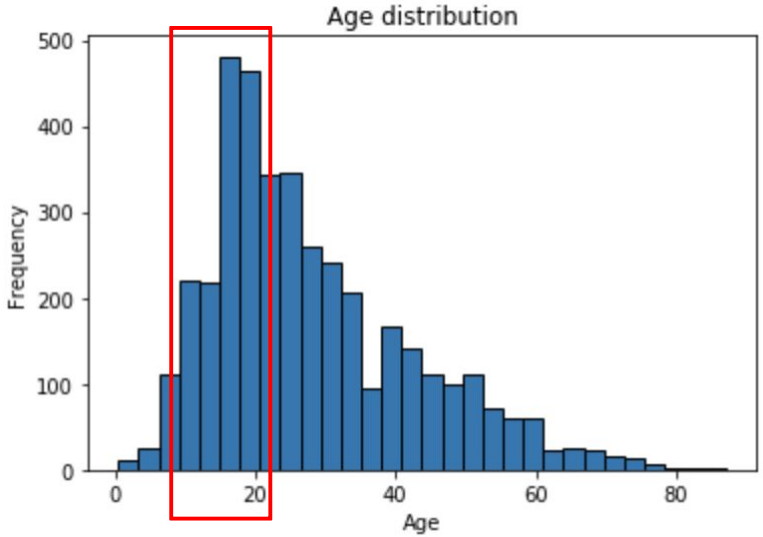
Inconsistent values

Data Types

Exploratory Data Analysis | Q1+2 of 5

Let's go hunt

Questions	Variables Used
How many teens are attacked by sharks?	Age & Sex
What does our demographic look like?	



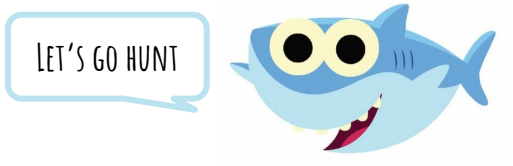
- Graph shows the age distribution of shark attack victims; Young adults & teens appear with the highest frequency
- Taking a closer look at teens; Teen boys take up 85.8% of the pie

sex	
M	85.8%
F	14.2%

MVP Takeaway

Make protective gear with a focus on teen boys

Exploratory Data Analysis | Q3 of 5



Question	Variables Used
What kind of shark injury is most common for teens?	Injury & Age

- We made a frequency analysis of attack type - by age group
- Teens are subject to unprovoked attacks more than any other kind of attack

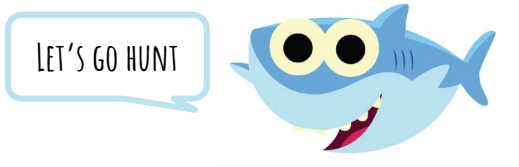
type age_group	Provoked	Questionable	Sea Disaster	Unprovoked	incidents Watercraft
Adult	217	159	31	1840	45
Child	15	31	3	315	1
Senior	331	334	201	1934	309
Teens	74	54	7	1018	7

MVP Takeaway

Make protective gear that
- ➔ is comfortable
- ➔ easy to wear &
- ➔ fun to wear

so that the product can preemptively protect against attacks

Exploratory Data Analysis | Q4 of 5



Question	Variables Used
What are teens doing when being attacked?	Activity & Age

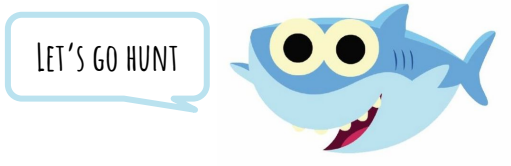
- We made a frequency analysis of activities - by age group
- Teens are attacked with a higher frequency when they are surfing or swimming

activity	diving	fishing	surfing	swimming	incidents wading
age_group					
Adult	240	461	637	390	59
Child	2	16	40	106	40
Senior	280	705	307	445	43
Teens	48	105	392	264	46

MVP Takeaway

Make protective gear that is perfectly suited to water sports such a swimming or surfing

Exploratory Data Analysis | Q5 of 5



Question	Variables Used
What country (+state) would be ideal for starting up Teen Shark Inc.	Country & State

- Country with the highest attack frequency was the USA
  - Florida most freq. state

MVP Takeaway
Find Teen Shark Inc.
in Florida, USA

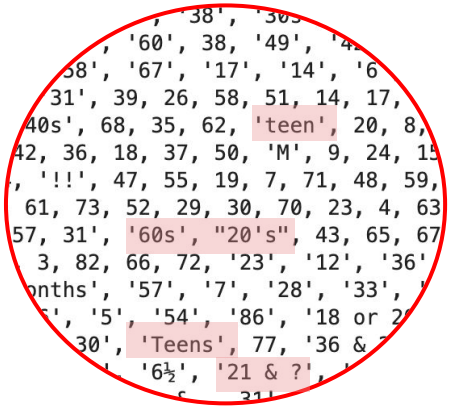
```
country
USA                  36.8%
AUSTRALIA            21.5%
SOUTH AFRICA          8.7%
NEW ZEALAND           2.1%
BAHAMAS               2.0%
PAPUA NEW GUINEA      2.0%
BRAZIL                1.8%
MEXICO                1.5%
ITALY                 1.0%
FIJI                  1.0%
Name: country, dtype: object
```

```
state
Florida                  0.182297
New South Wales          0.080625
Queensland               0.053699
Hawaii                   0.052770
California               0.049985
Western Australia        0.035747
KwaZulu-Natal            0.033736
Western Cape Province    0.030486
South Carolina           0.026617
Eastern Cape Province    0.025998
Name: state, dtype: float64
```

Major Obstacle

Biggest Obstacle

- 60-80% of the total time was spent on cleaning the data before we could make any meaningful analysis
  - Messy data - ie. missing values or inaccurate values



Major Learning

- **"Teamwork makes the dream work"**
  - We started with a plan & organized ourselves
  - We adapted quickly to problems & took initiative to step-in & help each other

Conclusion & Insights

- **We were quick to recognize that shark attacks are trending upwards**
  - **identified a potential business opportunity**

- **Room to refine the business**
  - **More injury details**
  - **Potential for global expansion (ie. AU)**
  - **Are kids an underserved market?**

- **Biggest surprise: sometimes the messy data can be quite funny**