Analysis of Australian Shark Attacks

BEGINNER DATA VISUALIZATION MACHINE LEARNING PYTHON

This article was published as a part of the Data Science Blogathon.

Introduction

Recently I searched for an interesting dataset to learn something new. After searching for a long time, I got a dataset on Shark Attacks in Australia. This dataset contains about 1,100 + shark bites and attempted shark bites between 1791 and early 2022, gathered by the ***Taronga Conservative Society***.

In this article, We will learn –

- How to use feather files in pandas and why we use them.
- How to plot barplot and pieplot using matplotlib and seaborn.
- Different facts about sharks, as a bonus.

So, fasten your seatbelts, and let's get started.

The Data

This dataset contains 60 columns. That will be overwhelming if I give details about every column. And also, we don't need all columns here. Here I am giving the details about 12 columns. If you want to know about all columns, visit this link.

I also performed some preprocessing before using this data. You can download both the actual and preprocessed data from my repository's data folder.

Data Dictionary

Column Name	*Details*
incident_month	The month in which the attack happens
incident_year	The incident year
victim_injury	How much the victim is injured (fatal, injured, uninjured)
state	Australian State / Territory
location	Closest town and/or beach (descriptive)
shark_common_name	Name of the shark
provoked/unprovoked	Is the shark attacked with or without the human influence.
victim_activity	Activity of victim when the shark attack happens.
victim_gender	Gender of the victim
shark_behaviour	Which part of the victim is bitten
injury_location	The location of the injury (inside body)
victim_age	Age of victim

Data Analysis

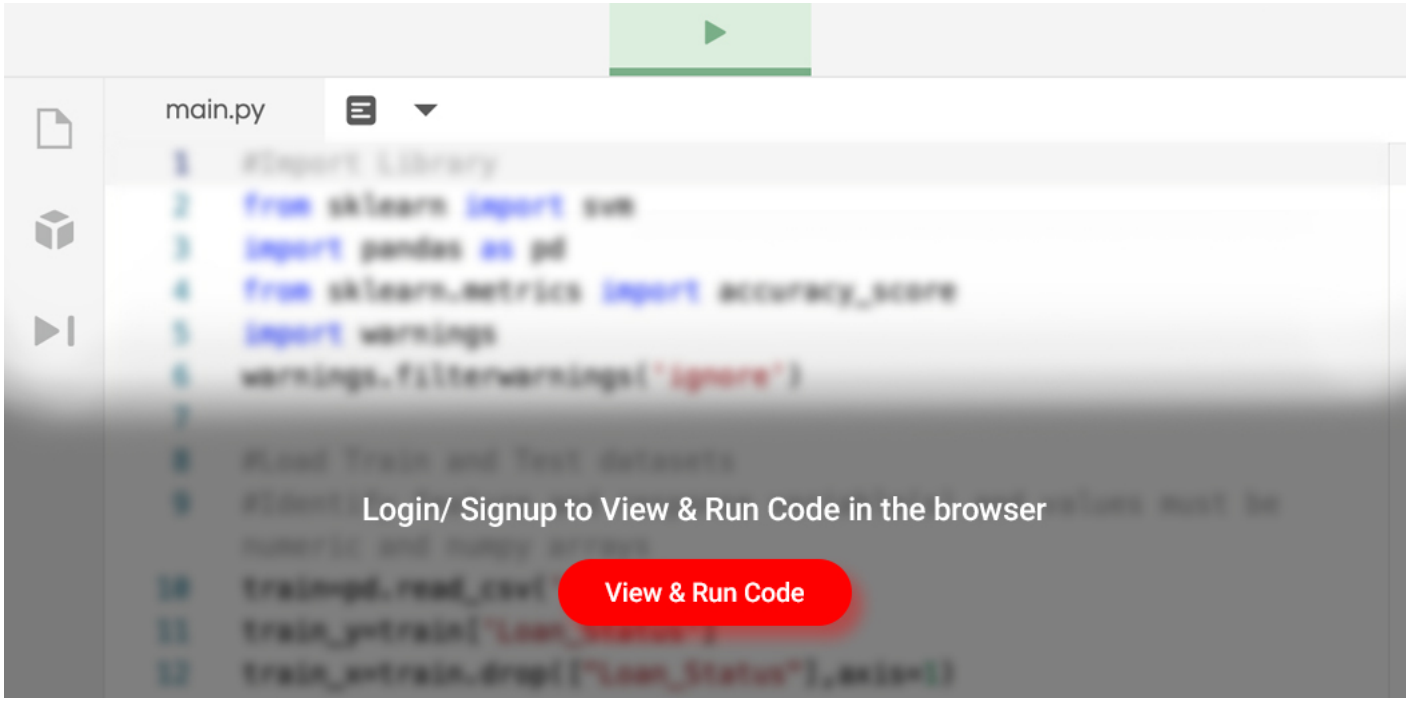
Before analyzing the data, let’s import the necessary libraries.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings

warnings.filterwarnings('ignore')
```

view raw

import_shark.py hosted with ❤ by **GitHub**

Now let's import the data using pandas.



From the above code, you will notice that the file is in feather format. The feather format is useful for

- When you want to read data faster.
- When you don't need any modification on your data rather than compression.

The machine only reads Feather formatted files. Sometimes we need to load a large amount of data faster. Here csv format is not useful as the loading time of csv formatted data increases when the file size is large. Converting a csv file to a feather file helps us reduce the file size and the loading time of the data file is reduced drastically.

One thing you have to remember, If you have extensive data (like 5 to 6 GB) then try Apache Parquet.

If we run the above code, you will see the output below.

	incident_month	incident_year	victim_injury	state	lo
0	January	1791	fatal	New South Wales	
1	January	1807	injured	New South Wales	
2	January	1825	injured	New South Wales	
3	January	1832	injured	New South Wales	bla
4	June	1832	injured	New South Wales	
...	...	...	...	...	
1191	July	2003	uninjured	Northern Territory	
1192	April	2005	injured	Northern Territory	
1193	September	2018	injured	Northern Territory	
1194	November	2018	injured	Northern Territory	
1195	February	2019	injured	Northern Territory	

This data contains 1196 rows and 12 columns. After the data is successfully loaded, now let's see the statistical summary of the data. First, we see the statistical summary of the numerical columns.

```
shark_data.describe()
```

The output will look something like the one below.

	incident_year	victim_age
count	1196.000000	699.000000
mean	1966.895485	28.164521
std	48.260971	13.794858
min	1791.000000	0.000000
25%	1931.000000	17.000000
50%	1983.000000	25.000000
75%	2010.000000	36.000000
max	2022.000000	84.000000

From the above result, we can easily see that —

- The average age of the victim is 28 years, and the maximum age of a victim is 84 years. I think maximum shark attacks happened to the oldest fishermen. We will verify this fact later.

- The minimum age of a victim is 0, a null value or an error.

For the ***incident_year*** column, the statistical summary doesn't make sense. We have to see this column's statistical summary by converting the column to categorical.

```
shark_data_copy = shark_data.copy() shark_data_copy['incident_year'] = shark_data_copy['incident_year'].astype('object') shark_data_copy.describe(include='O')
```

	incident_month	incident_year	victim_injury	state	loca
count	1196	1196	1196	1196	
unique	12	176	4	7	
top	January	2020	injured	New South Wales	thu
freq	223	37	722	438	

provoked/unprovoked	victim_activity	victim_gender	shark_behaviour	inj
1192	1171	1179	980	
2	10	2	271	
unprovoked	swimming	male	bit victim on leg	
794	452	1062	241	

From the above result, we can observe that —

- The maximum incident happened in ***January***.
- Shark incidents were mostly recorded in the year ***2020***.
- ***722 people out of 1196*** were injured in shark attacks.
- Most of the shark attacks were reported from the ***New South Wales*** state of Australia.
- Most attacks are made by ***White Shark***. If you search on the internet, you will see that ***White Shark is responsible for by far the largest number of recorded shark bite incidents on humans***. Below is the picture which proves the statement.

Relationship with humans

Shark bite incidents

Main article: Shark attack

Of all shark species, the great white shark is responsible for by far the largest number of recorded shark bite incidents bite incidents on humans as of 2012.[19]

More than any documented bite incident, Peter Benchley's best-selling novel *Jaws* and the subsequent 1975 film adap great white shark with the image of being a "man-eater" in the public mind.[155] While great white sharks have killed hu incidents, they typically do not target them: for example, in the Mediterranean Sea there have been 31 confirmed bite centuries, most of which were non-fatal. Many of the incidents seemed to be "test-bites". Great white sharks also test- and they might grab a human or a surfboard to identify what it is.

The great white shark is one of only four kinds of shark that have been involved in a significant number of fatal unprovoked attacks on humans.

Contrary to popular belief, great white sharks do not mistake humans for seal visibility or other situations which impair the shark's senses. The species appe finds the taste unfamiliar. Further research shows that they can tell in one bite upon. Humans, for the most part, are too bony for their liking. They much pref

Studies published in 2021 by Ryan *et al.* in the Journal of the Royal Society In case for many shark bite incidents perpetrated by great white sharks. Using c models and mounted cameras moving at the same speed and angle as a crui from below, the experiment suggests that the sharks are likely colorblind and whether the silhouette above them is a pinniped or a swimming human, poten

Humans are not appropriate prey because the shark's digestion is too slow to and fat. Accordingly, in most recorded shark bite incidents, great whites broke usually caused by blood loss from the initial bite rather than from critical organ 2011 there have been a total of 139 unprovoked great white shark bite incider

However, some researchers have hypothesized that the reason the proportion

- Most shark attacks are ***unprovoked*** and happen while ***swimming***.

- Most victims' gender is ***male*** and they got injuries mostly in their ***legs***.

Are Shark Attacks increasing over the years in Australia?

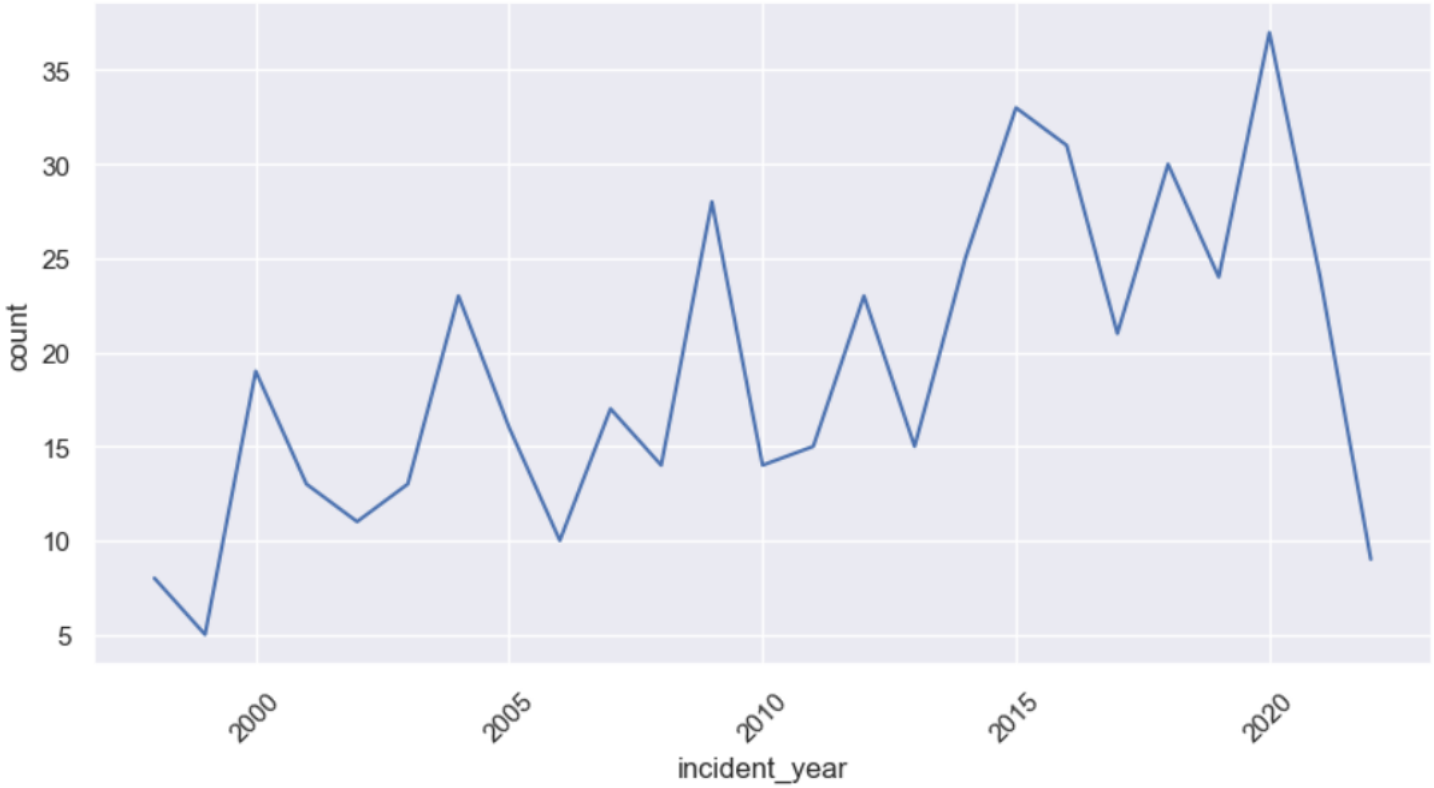
Here we are going to see whether shark attacks are increasing over the years or not. We don't need all the years from the table to know this. So, I selected the data count from 1998 to 2022 and plot a graph.

```
year_count = shark_data['incident_year'].value_counts() \
                                        .rename_axis('incident_year') \
                                        .reset_index(name='count')
kamor_increase = year_count[year_count['incident_year'] >= 1998].sort_values(by='incident_year')


# line plot
plt.figure(figsize=(10,5))
sns.set()
sns.lineplot(x='incident_year', y='count', data=kamor_increase)
plt.xticks(rotation=45)
plt.show()
```

view raw

attack_plused.py hosted with ❤ by GitHub

After running the above code, the plot will look like the one below.



From the above result, we can see an increasing trend which tells that shark attacks are increasing. Below is a screenshot of an Australian newspaper where this fact is stated.

New dataset shows shark bites in Australia are increasing and researchers want to know why

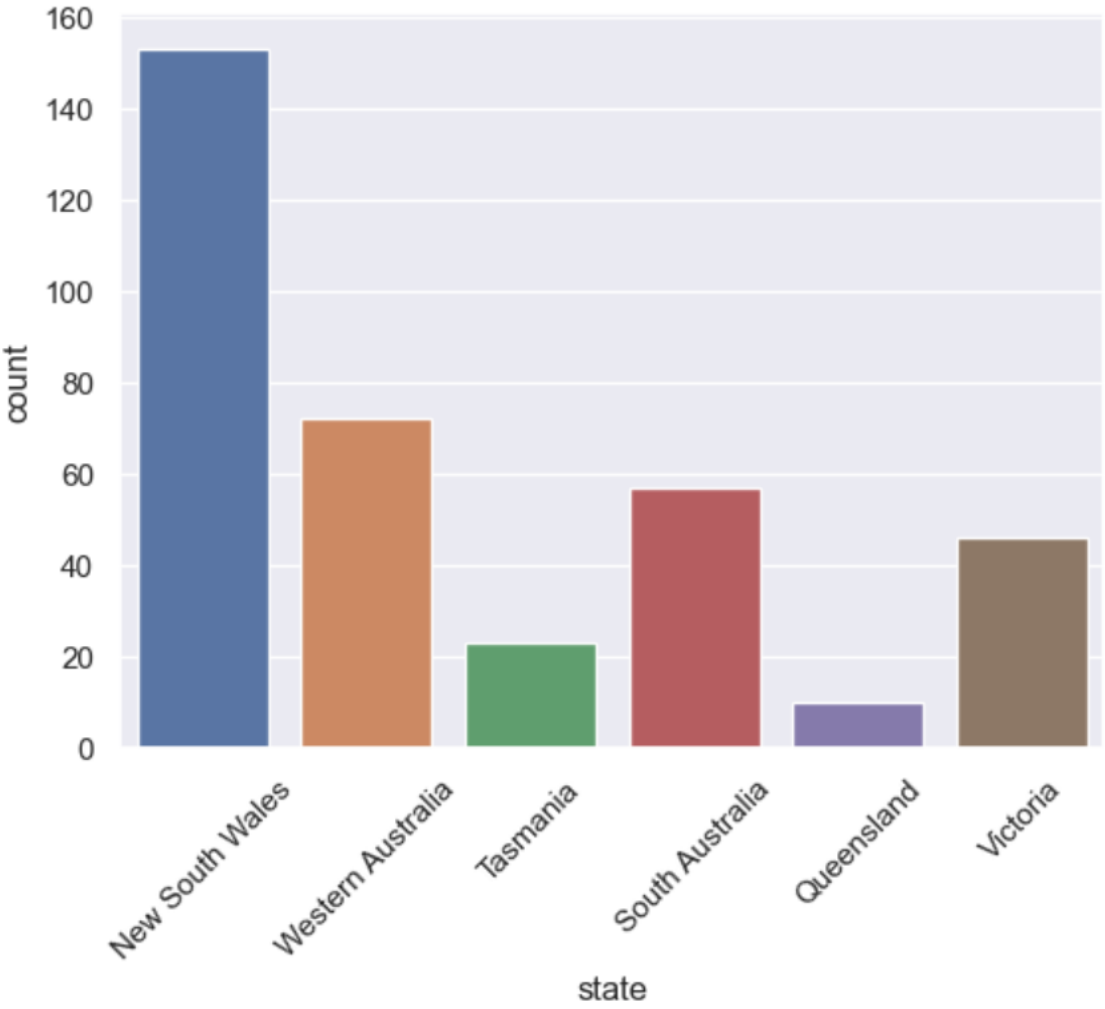
Analysts caution changes in the manner of reports need to be considered when examining the data

- Follow our Australia news live blog for the latest updates
- Get our free news app, morning email briefing and daily news podcast

White Shark Attacks

Previously, we saw from the statistical inference that white sharks attack most victims. Let's see where the most attacks happened.

```
plt.figure(figsize=(10,      6))      sns.countplot(x='state',      data=white_shark_case,      palette='RdYlBu')
plt.xticks(rotation=45) plt.show()
```



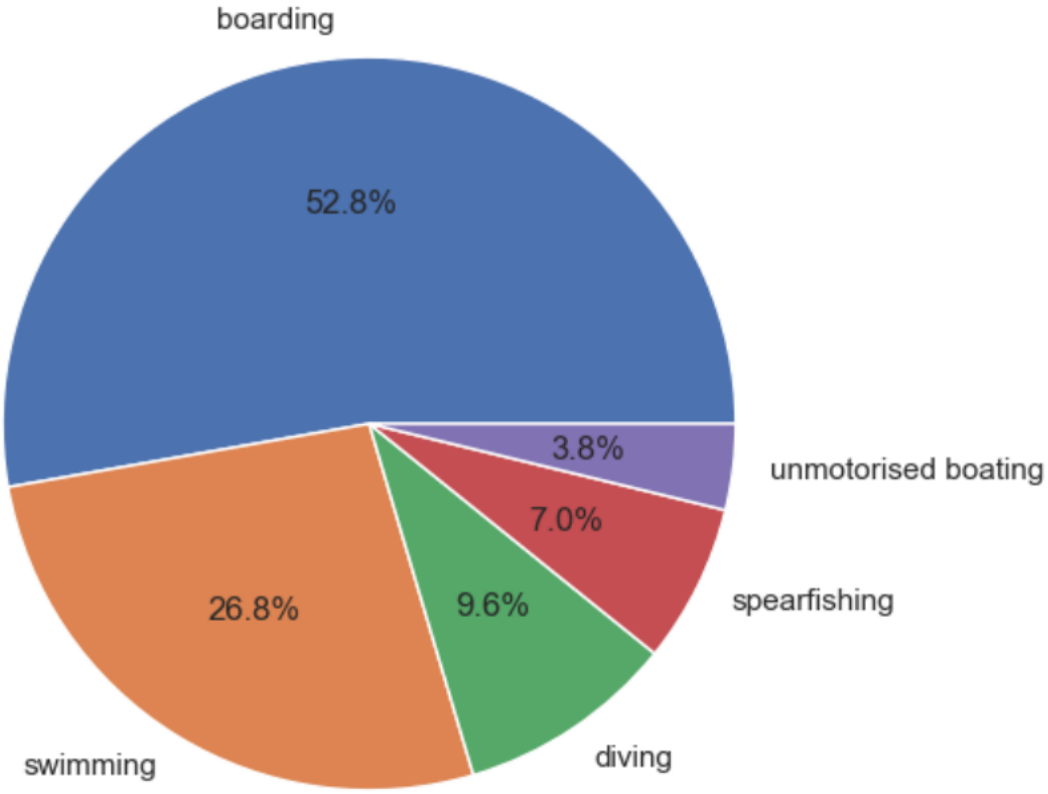
From the above bar plot, we can see that most attack happens in New South Wales. This type of shark is mostly found in ***New South Wales, Australia***.

Now we see the victim's activity when the shark attack happened.

```
plt.figure(figsize=(10,6))

df1 = pd.DataFrame(white_shark_case['victim_activity'].value_counts().reset_index())
df1 = df1[df1['victim_activity'] > 10]

plt.pie(x='victim_activity', labels='index', data=df1, autopct='%.1f%%')
plt.show()
```

view raw

victim_activity.py hosted with ❤ by **GitHub**



From the above pie chart, it is visible that most of the victims enjoy ***boarding*** when the shark attack happens. That means this shark frequently comes to the surface of the water.

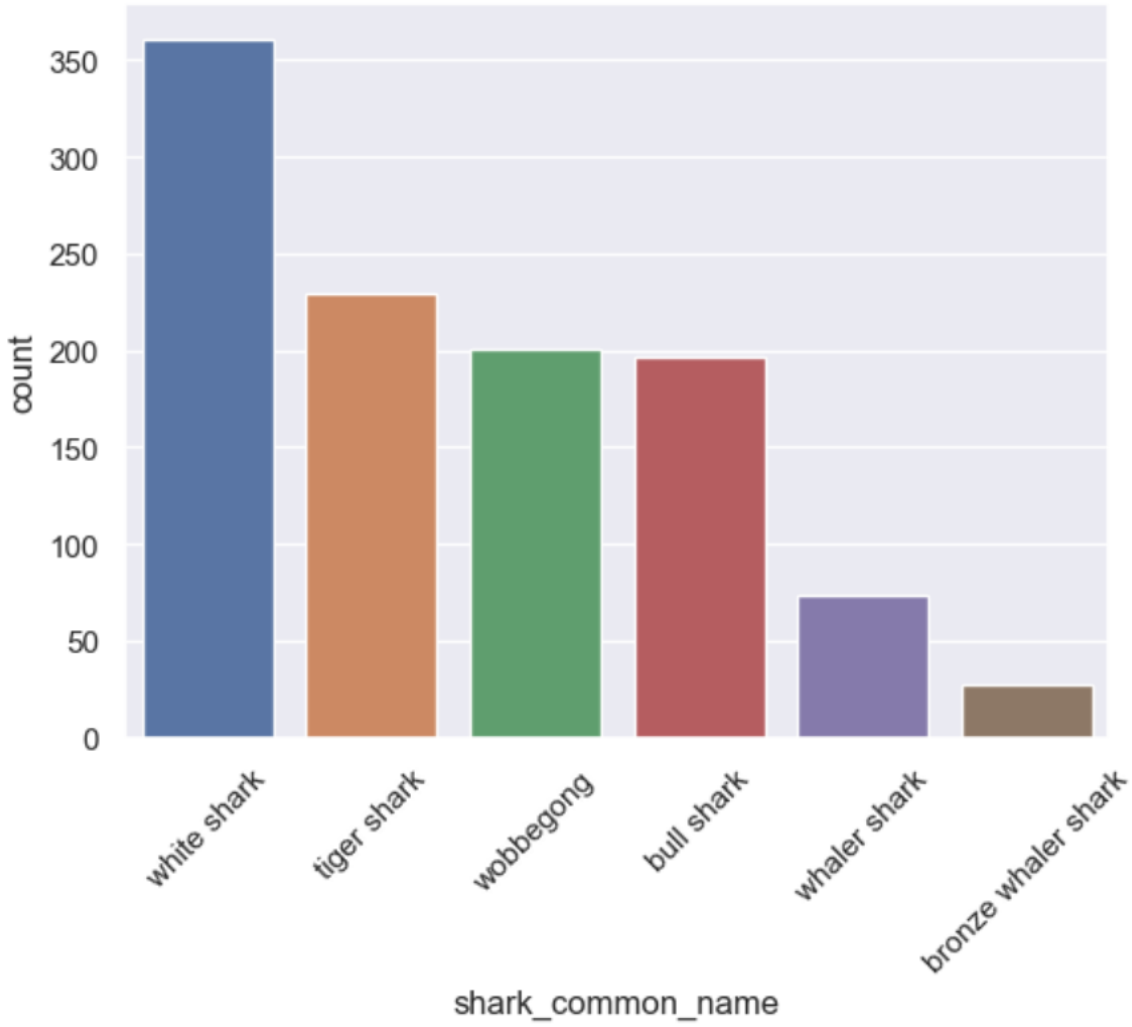
Who has more Shark Attack Cases after White Shark?

Now let's see who has the second and third place based on shark attack cases.

```
shark_incident_count = shark_data['shark_common_name'].value_counts() \
                                    .rename_axis('shark_common_name') \
                                    .reset_index(name='count')
second_shark = shark_incident_count[shark_incident_count['count'] >= 10]


sns.barplot(x='shark_common_name', y='count', data=second_shark)
plt.xticks(rotation=45)
plt.show()
```

view raw

gistfile1.txt hosted with ❤ by **GitHub**

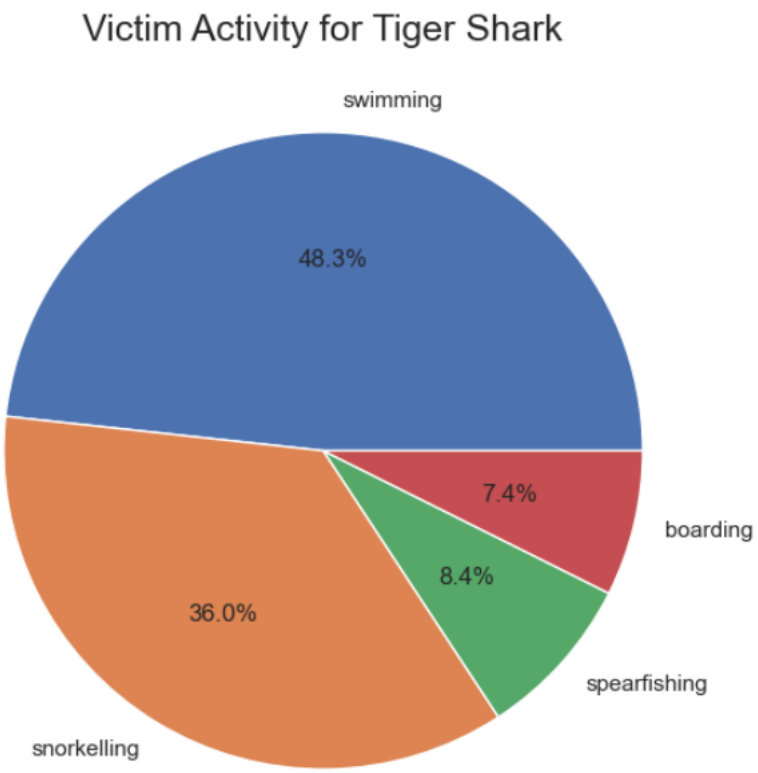


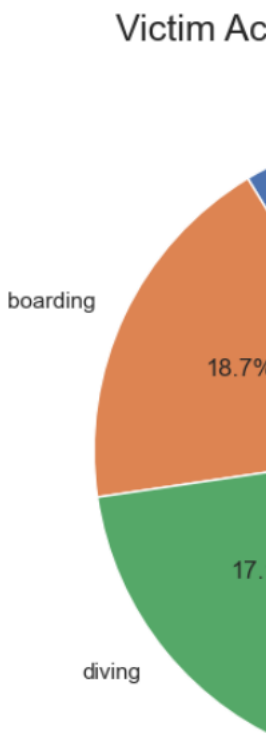
From the above bar chart, ***tiger shark*** and ***wobbegong*** have second and third place respectively. Now we are going to see what is the victim activity when these two sharks attacked them.

```
fig, (ax1, ax2) = plt.subplots(ncols=2, figsize=(20, 7))

df2 = pd.DataFrame(tiger_shark_case['victim_activity'].value_counts().reset_index())
df2 = df2[df2['victim_activity'] > 10]

df3 = pd.DataFrame(wobbegong_shark_case['victim_activity'].value_counts().reset_index())
df3 = df3[df3['victim_activity'] > 10]

ax1.pie(x='victim_activity', labels='index', autopct='%.1f%%', data=df2)
ax2.pie(x='victim_activity', labels='index', autopct='%.1f%%', data=df3)

ax1.set_title("Victim Activity for Tiger Shark", fontsize=18)
ax2.set_title("Victim Activity for Wobbegong Shark", fontsize=18)
fig.show()
```

view raw

victim_of_tiger_and_wobbengong.py hosted with ❤ by **GitHub**





From the above charts, we notice that tiger shark attacks mostly happen when the victim is ***swimming*** or ***snorkeling***. If you don't know what is snorkeling, here is the definition from Wikipedia.

> **Snorkeling** (British and Commonwealth English spelling: **snorkelling**) is the practice of swimming on or through a body water while equipped with a diving mask, a shaped breathing tube called a snorkel, and usually swimfins. In cooler water wetsuit may also be worn. Use of this equipment allows the snorkeler to observe underwater attractions for extended per with relatively little effort and to breathe while face-down at the surface.

Wobbegong shark attack happens when the victim is ***swimming, boarding, or diving***. Those sharks are interesting. If you see this type of shark at a glance, it seems like a carpet. Below is a picture of wobbegong.

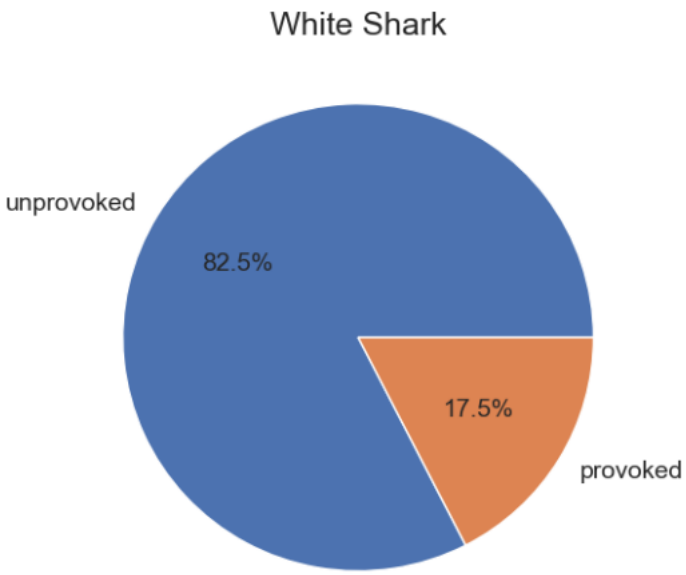
If you want to know more about wobbegong, you can see this video.

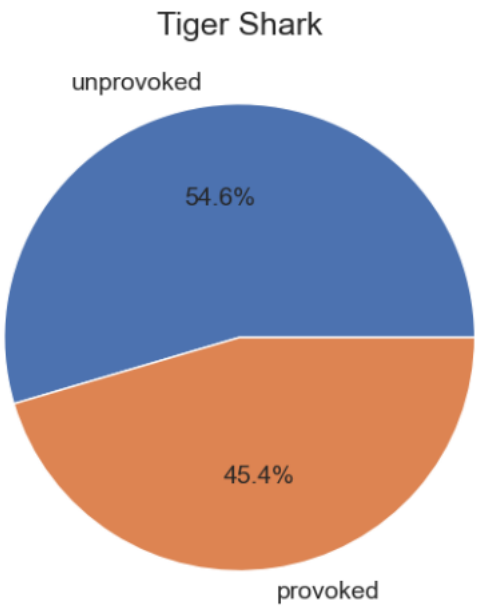
Let's see the provoked/ unprovoked ratio for the white shark, tiger shark, and wobbegong shark, respectively.

```
fig, (ax1, ax2, ax3) = plt.subplots(1, 3, figsize=(20, 8))

df1 = pd.DataFrame(white_shark_case["provoked/unprovoked"].value_counts().reset_index())
df2 = pd.DataFrame(tiger_shark_case["provoked/unprovoked"].value_counts().reset_index())
df3 = pd.DataFrame(wobbegong_shark_case["provoked/unprovoked"].value_counts().reset_index())

ax1.pie(data=df1, x='provoked/unprovoked', labels='index', autopct='%.1f%%', textprops={"fontsize": 14})
ax2.pie(data=df2, x='provoked/unprovoked', labels='index', autopct='%.1f%%', textprops={"fontsize": 14})
ax3.pie(data=df3, x='provoked/unprovoked', labels='index', autopct='%.1f%%', textprops={"fontsize": 14})

ax1.set_title("White Shark", fontsize=18)
ax2.set_title("Tiger Shark", fontsize=18)
ax3.set_title("Wobbegong Shark", fontsize=18)

fig.show()
```

view raw

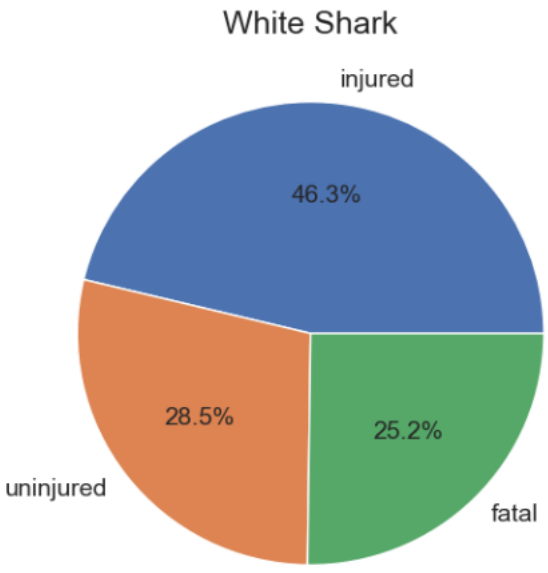
provoked.py hosted with ❤ by **GitHub**

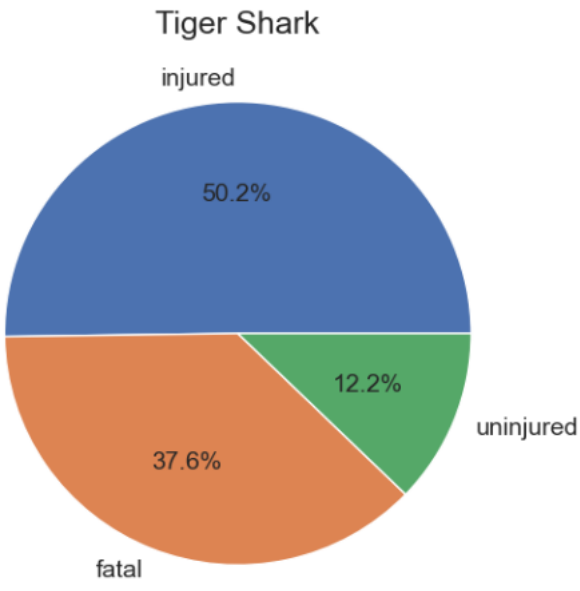




White shark has more unprovoked attacks than others. Wobbegong shark has an almost equal ratio of provoked and unprovoked attacks.

Now let's see how many people are injured, uninjured, or take fatal damage in their attacks. For this, just use the previous code snippet replacing ***provoked/unprovoked*** to ***victim_injury.***





Surprisingly, though the wobbegong shark injured so many people, they are not in the first place based on shark attacks. It may happen because of the highest ratio of fatal injuries dealt with by white sharks.

Conclusion

So, That's all I got. I know that this article is pretty long, but I think it is worth reading. In this article, you learned how to –

- Read the feather file using pandas and when to use it.
- Plot different graphs using seaborn and matplotlib.
- Modify the data according to your needs.
- Customize different charts.

But the analysis doesn't end here. If you got something, let me know in the comments. If there is something wrong from my side, I am always here to listen to you. You can also do some more in the plots – there are so many options for customization. I made some basic customization so that the beginners don't find those plots overwhelming.

You can read my article on the analysis of dark chocolates. Here is the link. You can also check my medium profile.

Article Url - https://www.analyticsvidhya.com/blog/2022/09/analysis-of-australian-shark-attacks/

Subhradeep Rang