A DATA ANALYSIS ON

AUSTRALIA SHARK ATTACKS

Presented By: Ng Juat Teng

8 April 2025

Capstone Project Report for General Assembly's Data Analytics Bootcamp DAB04

CONTENTS

INTRODUCTION

Australia is the only continent in the world that is fully surrounded by water and it has a booming tourist industry relating to beach and water activities. Australia is also the second-highest number of shark attacks in the world, with the first being USA. There seems to be an increasing trend in shark bite incidents over the years.

How can we make it safer for water activity lovers and for the sharks as well?

PERSONA

June is an avid scuba diver who is planning a shark diving adventure in Australia. However, her family and friends are concerned about her safety due to recent reports of shark attacks globally.

June has read that it is safer to scuba dive as compared to swimming or surfing at a beach where sharks are known to be found, and would like find out if that is true.

Derek is an avid surfer from Germany and would like to conquer the waves in Australia one day. He noticed a lot of online articles regarding shark bites on surfers in Australia and is wondering how to achieve his dream while being safe.

He would like to know which beach is safer to go to, and if there is a particular time of the year to avoid when planning for his trip.

OBJECTIVE (GOALS)

- To reduce shark attack rates in Australia by raising awareness and educating the general public on when to avoid visiting beaches where sharks are known to frequent
- Target Audience: Domestic and international beach-goers, surfers and scuba divers

APPROACH

Step 1: Download dataset

Step 2: Initial data cleaning using Microsoft Excel

Step 3: Import data to Python for further wrangling and Exploratory Data Analysis

Step 4: Visualize trends and insights with Tableau

Step 5: Continue on Python for predictive modelling

Step 6: Consolidate findings into a report and presentation slides

THE DATASET

DATA SOURCE

The main dataset used for this analysis is the **Australian Shark-Incident Database**, which is a log that quantifies temporal and spatial patterns of shark-human interactions in Australia and is considered the principal source of shark-bite data in Australia. This database is a joint partnership with Taronga Conservation Society Australia, along with the Flinders University, and the New South Wales Department of Primary Industries. Maintained as an uninterrupted record by committed team members since 1984, the file currently comprises of more than 1,200 individual investigations from 1791 to today, making it the most comprehensive database of its kind available.

Source: https://zenodo.org/records/11334212

DATA CLEANING/WRANGLING PROCESS

After the database was downloaded, Microsoft Excel was used to clean and prepare the dataset with the steps as follow.

Raw Data File: "Australian Shark-Incident Database Public Version 7 (raw).xlsx"

Column Name	Remarks
Dropped 42 columns	• Original dataset had 59 columns x 1233 rows, reduced to 17 columns. • Dropped columns with high amount of blanks (>50%), overly descriptive/non-categorical texts, or information that is irrelevant to current analysis.
<Shark.common.name> <Provoked/unprovoked> <Victim.activity> <Victim.gender>	• Filled in blanks with 'unknown'
<Victim.injury>	• Corrected 1 count of "injury" as "injured". • Removed the only count of "unknown" which happened in 1959. (UIN #493).
<Shark.common.name>	• Consolidated "broadnose sevengill shark" and "seven gill shark" into "sevengill shark". • Incorporated "bronze whaler shark" into "whaler shark" (different common name for the same species).
<Victim.activity>	• Corrected spelling from "snorkeling" to "snorkelling". • Classified "surfing" and "paddleboarding" under a broader term "boarding". • Classified "other: standing in water" under "swimming" based on data dictionary provided.

<Injury.severity>	• Most of the 255 nos. records with "fatality" under <Victim.injury> column were originally classified under "major lacerations" in this column. These were changed to "fatal" instead. • Records with "uninjured" under <Victim.injury> column had blanks in this column. These were updated with "none" instead. • "Surface wound" is re-classified under "abrasion". • 2 counts of "lacerations" are re-classified under "minor lacerations". • 52 counts of "blanks" are filled in with "injuries unknown", as these records were listed as "injured" under <Victim.injury> column.
<Victim.age>	• 510 blanks were filled in with "999".
<Time.of.incident> <Hour.of.incident>	• Created a new column <Hour.of.incident>. • 710 blanks are imputed with "unknown". • Grouped timings by the hour i.e. "1515" and "1530" will be classified as "1500".

<Depth.of.incident>
<Depth.category>
	• Blanks in <Depth.of.incident> were filled in with "999".
	• Created new column <Depth.category> to categorize <Depth.of.incident>, based on the following:
	    ○ 0 to 1.5m: Water Surface
	    ○ 1.5m to 5m: Shallow Water
	    ○ 5m to 18m: Open Water Diver Depth
	    ○ 18m to 30m: Recreational Diver Depth
	    ○ 31m and above: Technical Diver Depth
	    ○ 999: Unknown

DATA DICTIONARY

The result from the cleaning process was exported to a new csv with the following final columns.

Cleaned Data File: "aus_shark_v7.csv".

Field	Type	Description
UIN	integer	Unique identifier for each case
Incident.month	integer	Month of year
Incident.year	Integer	Year in full
Victim.injury		Outcome of victim's health. Categories: • **Fatal** bites include bites resulting in death to the victim. • **Injured** bites include bites resulting in physical injury to the victim (e.g., bruising, abrasion, punctures, lacerations). • **Uninjured** bites include interactions resulting in no injury to the victim (e.g., shark bit the victim's equipment; surfboard, fishing rod, kayak).
State	text	Australian State/Territory; abbreviated. Categories: WA, SA, VIC, NSW, QLD, NT, TAS.
Latitude	numeric	Latitude of incident
Longitude	numeric	Longitude of incident
Site.category	text	Type of site where incident occurred. Categories: coastal, estuary/harbour, island open ocean (includes offshore shallow reefs), ocean/pelagic, river, others.
Shark.common.name	text	If identified to species, species common name; if identified to family, family group common name; common name should be

		Australian common names (Last and Stevens 2009); all lowercase, e.g., requiem shark
Provoked/unprovoked	text	Categories: • **Unprovoked** is defined as an encounter between a human and a shark where a shark is in its natural habitat and has made a determined attempt to bite a human where that person is not engaged in provocative activities. • **Provoked** is defined as an encounter between a human and a shark where the person attracts or initiates physical contact with a shark (accidentally or on purpose) or was fishing for, stabbing, feeding, netting, or handling a shark, or where the shark was attracted to the victim by activities such as fishing, spearfishing (where a fish has already been speared), commercial diving (e.g., collecting abalone, pearl shells, or other marine animals where catch has already been collected), and cleaning of captured fish. • **Unknown**
Victim.activity	text	Activity at time of incident. Categories: • **Boarding** includes surfboarding, bodyboarding, kiteboarding, sailboarding, wakeboarding, stand-up paddle boarding. • **Diving** includes scuba, hookah, hard-hat diving. • **Motorised boating** • **Fishing** includes cleaning fish. • **Snorkelling** includes freediving. • Spearfishing • **Swimming** includes body surfing, clinging to object, falling into water, floating, or wading. • **Unmotorised boating** includes canoeing, kayaking, sailing. • **Unknow**n • **Others**
Injury.location	text	Areas on victim that are injured by shark. Categories: arm, hand, lower arm, upper arm, shoulder, neck, head, torso, leg, foot, calf, thigh, pelvic region, other; if multiple categories relevant separate by comma.
Injury.severity	text	Severity of victim's injuries sustained from shark. Categories: • **Abrasion** (removal of skin, i.e., grazes/scratches) • **Fatal** (leading to death) • **Major lacerations** (i.e., punctured/torn skin, muscle, tendon, or bone; loss of function seen; surgical procedures required) • **Minor lacerations** (i.e., punctured/torn skin and soft tissue; loss of function is not seen) • **Teeth marks** • **Punctures** • **None** • **Injury unknown**
Victim.gender	text	Categories: female, male, unknown.
Victim.age	integer	Age of victim in years
Hour.of.incident	text	Time of day incident occurred to the nearest hour in 24-hour time, no colon (i.e. 1830 is counted as 1800). Where time is unknown, 'unknown' is used.

Depth.of.incident.m	numeric	Estimated depth in metres at which the shark bite took place (not total water depth). Where depth is unknown, value of "999" is used.
Depth.category	text	Depth of incident categorized into the following based on conventional diving standards: • **0m to 1.5m**: Water Surface • **1.5m to 5m**: Shallow Water • **5m to 18m**: Based on open water recreational diver depth limit • **18m to 40m**: Based on advanced recreational diver depth limit • **Above 40m**: Beyond recreational diver limits, falls under technical diver depth range • **Unknown**: Depth is not given

EXPLORATORY DATA ANALYSIS

Univariate EDA was conducted on Python. Data from before year 1870 falls outside the Interquartile Range (IQR) and were dropped as outliers. Some columns had values that made more sense when regrouped, e.g. hour of day group into 4-hour blocks, and victim age grouped into 10-year blocks. The final file was then exported as a CSV file for visualization in Tableau.

File for Visualization: "aus_shark_v7_export.csv"

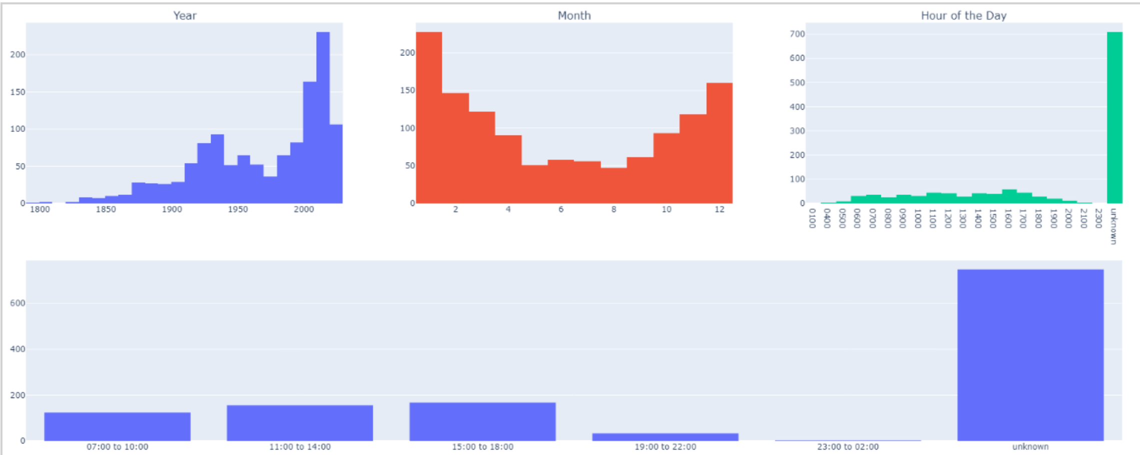

Figure 1: EDA of year, month and hour (top) and 'hour of day' grouped into 4-hour blocks (bottom).

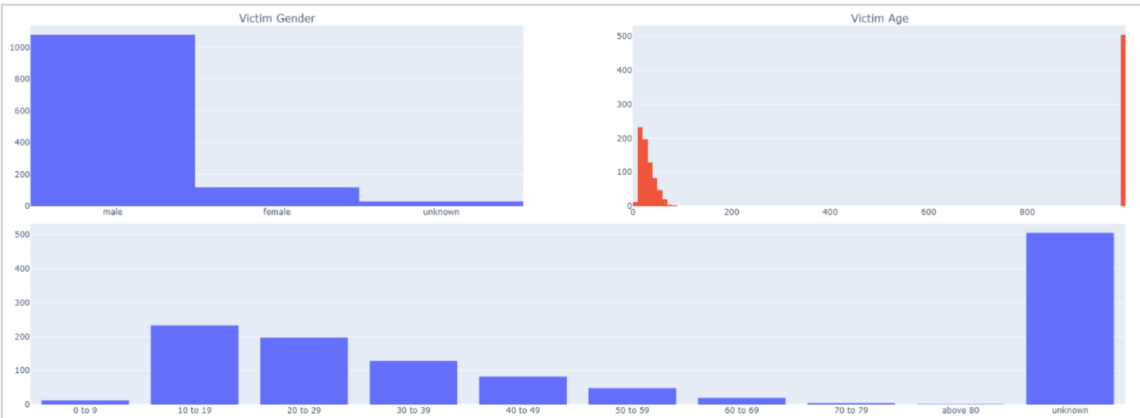

Figure 2: EDA of victim profile (top) and victim age grouped into 10-year blocks (bottom).

KEY TRENDS

In the Tableau dashboards created, we have summarised the key trends of past data for easy reference. The visualizations exclude data from before 1900 as that is more than 125 years ago (may be outdated), and the information available was not totally completed.

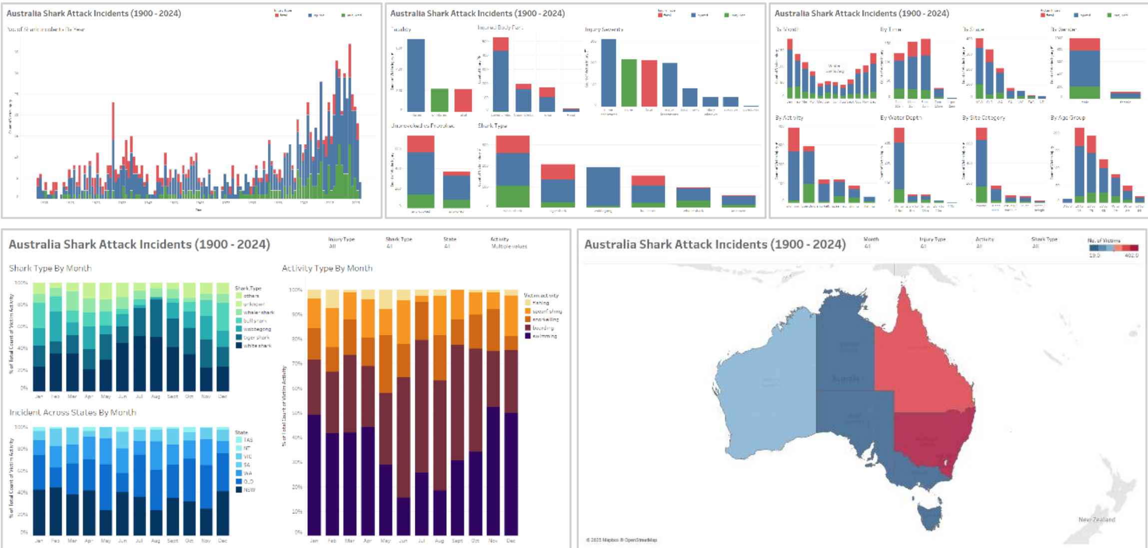

Figure 3: Tableau Dashboards 1 to 3 (top row from left to right) and Dashboards 4 to 5 (bottom row from left to right).

Dashboard 1 to 3 summarised the key trends of individual features while highlighting the segments between the uninjured, injured and fatal cases. Users can toggle on the injury type at the top to highlight the areas that they would like to see.

Dashboard 4 had selected features plotted by month and showed a percentage of each factor within the month. It has selectors for shark type, state and activities on the top for users to toggle between features i.e. how a specific shark type could show up more in certain states and does it have a higher correlation with a certain type of activity.

Dashboard 5 is a map of how shark incidents were distributed across the states of Australia. There are also selectors at the top to view the different distribution by month, injury type, activity and shark type. This will be useful for analysing shark behaviour patterns i.e. migration routes, breeding season, etc.

INJURY TRENDS

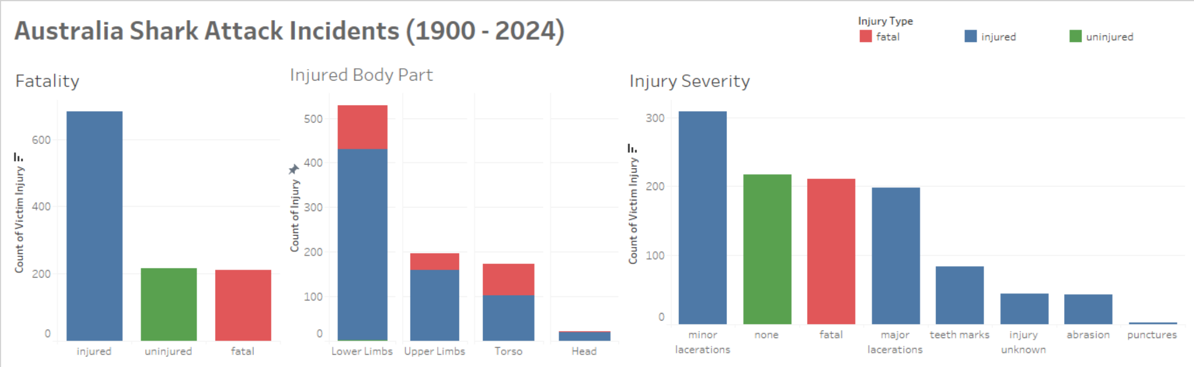

Figure 4: Shark attack incidents by fatality, injured body part and severity

- Majority of injury type is 'injured' with fatality at relatively low rate of 19% of all incidents, which is nearly the same as that for 'uninjured'.
- Most injuries involved the lower limbs of the body. However, the fatality rate is higher when the torso is involved – 40% of torso-related injuries results in fatality. Interestingly, the fatality rate for head-related injuries is very low at just 4.5%.
- Most of the injuries are minor – minor lacerations and all other injuries listed here except for major lacerations make up 43% of the injury severity group.

SHARK TYPE & BEHAVIOUR

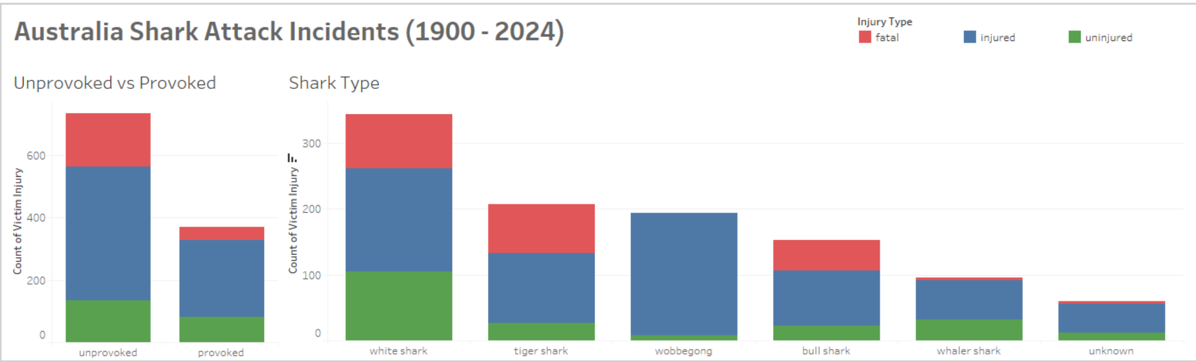

Figure 5: Shark attack incidents by provoked/unprovoked and shark type

- 66.5% of all attacks were unprovoked, of which 15% resulted in deaths. These attacks could be due to accidentally stepping on sharks in shallow beaches or unexpected encounters during in-water activities.
- White shark accounts for the most encountered shark type. This result is expected as great whites are known for their aggressive nature, along with tiger sharks and bull sharks.
- 3rd highest shark on the list is wobbegong shark, which is an interesting result as they are generally not considered aggressive. However, they can bite if provoked or threatened, especially if they are accidentally stepped on. They are also commonly found in shallow, inshore waters along the Australian coastline and are well camouflaged with the sandy beaches, making it easy for humans to accidentally step on them.

TIME AND LOCATION

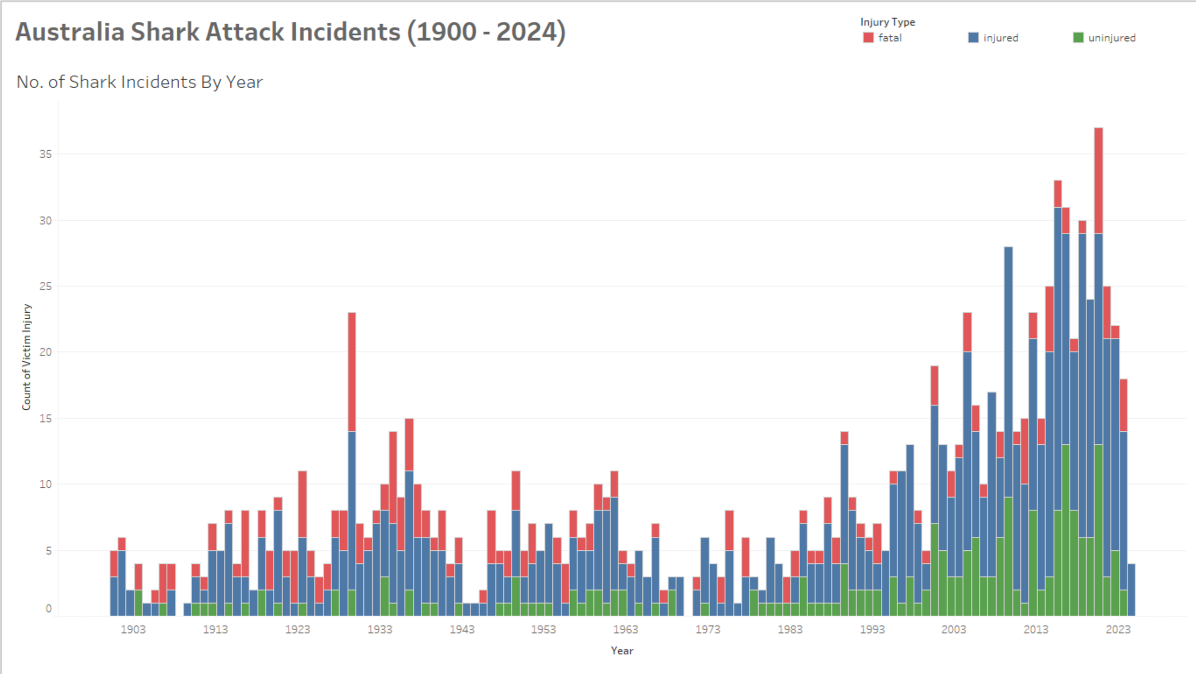

Figure 6: Shark attack incidents over the years

- From the chart above, there is an increasing number of shark incidents in recent years as compared to the olden days. However, we should be cautious to derive that sharks are becoming more aggressive. The presence of modern-day internet means that logging incidents are easier than in the past (which means logs are more accurate), but it also means that the general public will have a higher awareness of such incidents through news channels and online media.

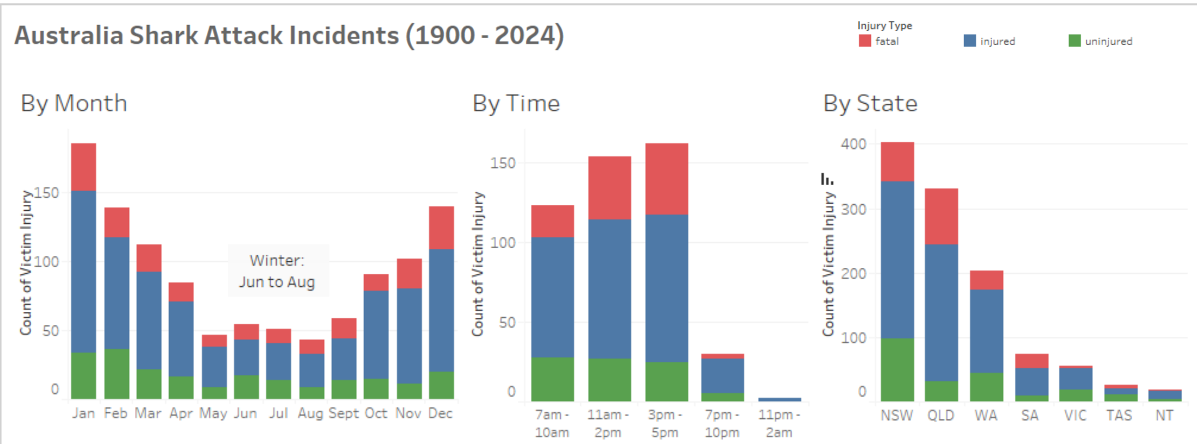

Figure 7: Shark attack incidents by month, time of the day and state.

- Shark incident trend has a very strong correlation with the seasons in Australia. December to February is the summer season, which generally have a lot more beach-goers and people partaking in in-water activities as compared to winter season in June to August.

- Most incidents also happened during daylight hours of 7am to 5pm, which is when most people go to the beach or head out to sea.
- New South Wales and Queensland are the top two states with the most shark incidents, with Queensland having a higher fatality rate than New South Wales. These two states have a large coastal population with high density of living.

ACTIVITY

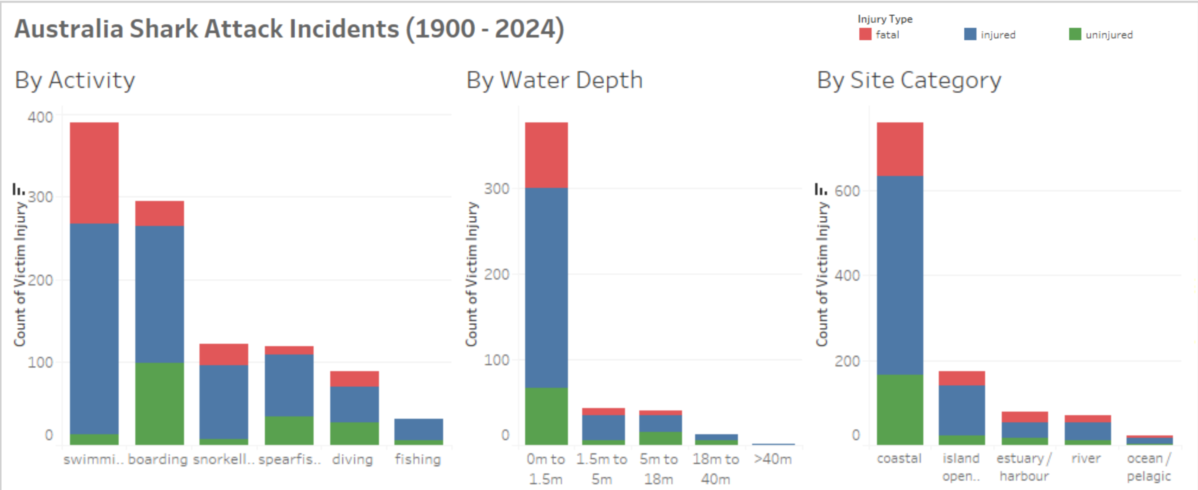

Figure 8: Shark incidents by activity type, water depth and site category

- Top activity when met with a shark incident is swimming at 37%, followed by boarding at 28% (including surfboarding, bodyboarding, kiteboarding, sailboarding, wakeboarding, stand-up paddle boarding). The fatality rate for swimming at 12% is notably higher than boarding at 3%, likely because of having a board to defend against the shark in an incident. This could also be the reason that boarding has the highest rate of uninjured incidents at nearly 10% of overall numbers.
- The rates for incidents by water depth corroborate with the findings by activity, with nearly 80% of incidents happening near the water surface (i.e. swimming, boarding, etc).
- The chart by site category also points in the same direction, with almost 70% of incidents happening in coastal regions (i.e. beaches).

VICTIM PROFILE

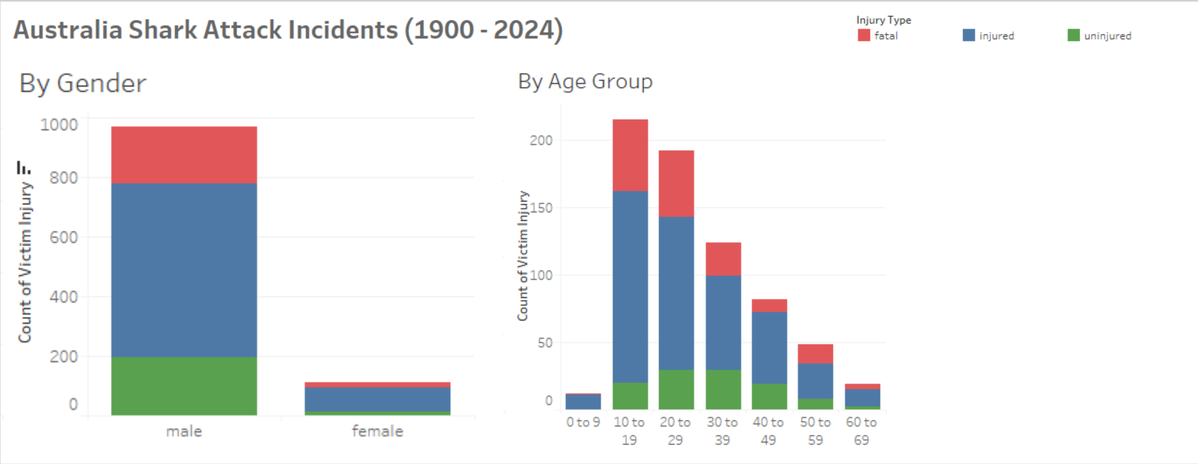

Figure 9: Victim profiles of shark attack incidents

- Nearly 90% of shark attack victims are male. In the previous section (*Figure 8*), swimming and boarding are the top activities. While swimming is generally gender neutral, studies have indicated that surfing is a male-dominated sport especially amongst adult surfers, with a gender split of 72:28 (men:women).
- 31% of victims are in their teenage years and 28% are in their 20s. This could be attributed to more younger people partaking in water activities as compared to the older group.

INSIGHTS

WHITE SHARK

White shark has been selected as a focus for insight analysis due to their famed and aggressive reputation. In Australia, they are typically found along the southern and western coasts with two distinct populations:

- **Eastern Population:** ranges from Victoria, NSW and up to Queensland. This population is also known to undertake long migrations, including to New Zealand, New Caledonia and Papua New Guinea.
- **South-Western Population:** ranges from western Victoria across south of Australia and up to Western Australia coast. to Western Australia and a southern-western one that ranges from Victoria in the New South Wales (NSW) region. The NSW population is also known to undertake long migrations.

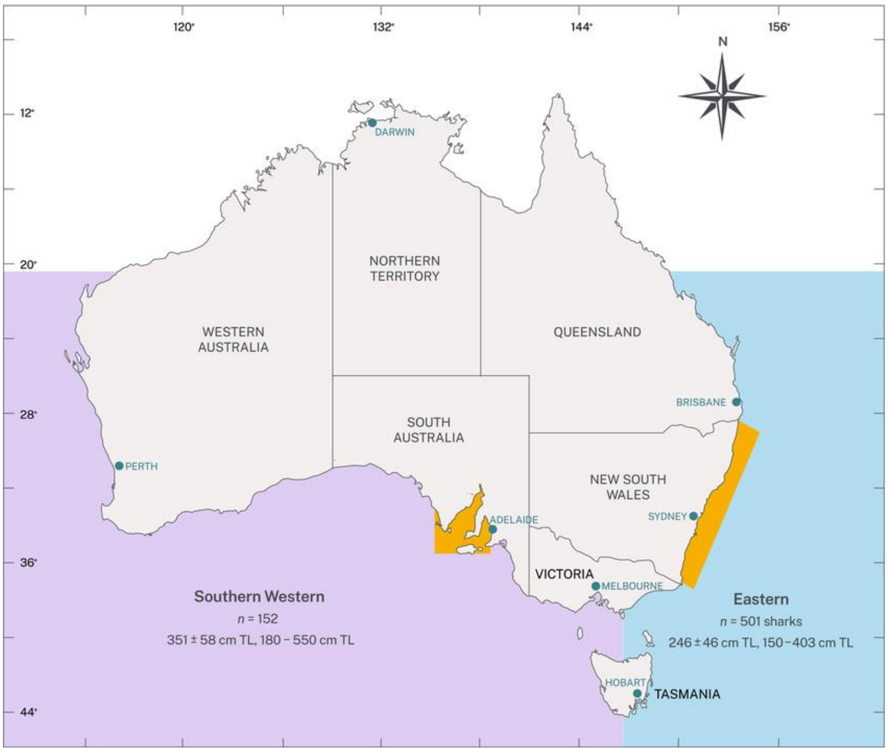

Figure 10: Distribution of two distinct white shark populations in Australia. (Zach S. R. Clark, 2024)

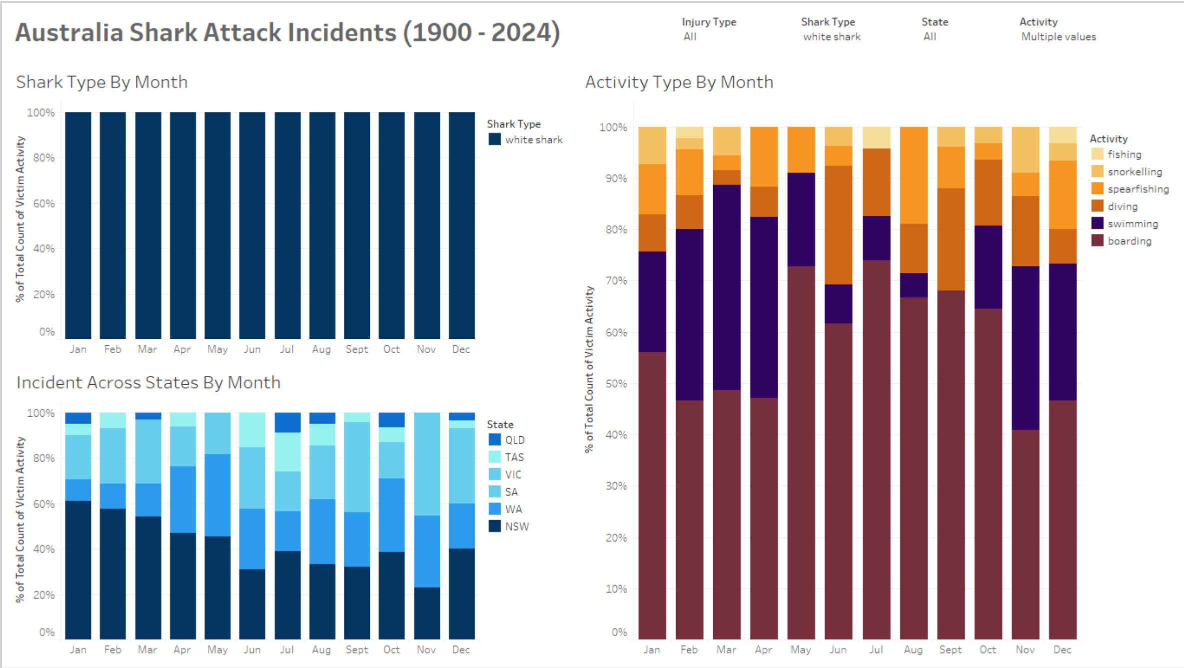

Figure 11: Incidents across states and activity type for white shark.

From **Dashboard 4**, we can observe that the main activity type with white shark encounters is primarily boarding, which is consistently high across all months. This is followed by swimming, which have high incidents in warmer months (November all the way to May), low incidents in winter months (June to August), and none in September. It seems that colder weather can deter swimmers but not surfing enthusiasts and possibly divers too as we also see higher incidents amongst divers in winter. This could also be a sign that white sharks may hang out in the shallow waters more during summer and stay in deeper waters during winter.

We can also see that white sharks are mostly encountered in New South Wales (NSW) which has the famous Bondi Beach, but with a generally decreasing trend from January to November. The lowest point for NSW is in November and in the same month, there are more white shark encounters in Western / Southern Australia instead, which could indicate a link with the shark's seasonal migration routes. We can visualise this trend with **Dashboard 5** by using the month selector, as demonstrated in Figure 12 below.

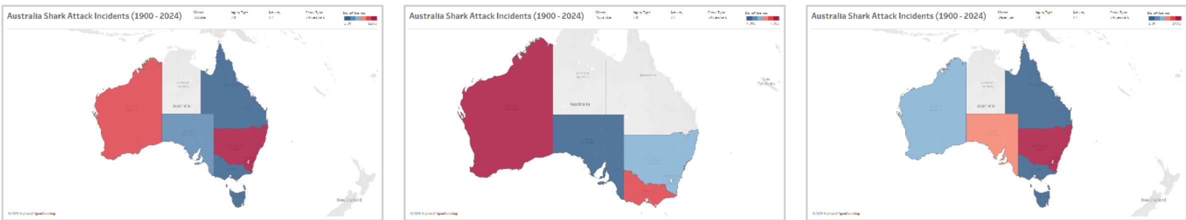

Figure 12: White shark incidents across different states in Oct (left), Nov (middle), and Dec (right).

WOBBEGONG SHARK

Wobbegong sharks are chosen for analysis as they are generally non-aggressive, yet they are the third highest encountered shark on our list. They are also generally not as well-known as other sharks like white shark or bull sharks.

Wobbegong sharks are generally found along the southern coastline of Australia, from southern Queensland to south-western Western Australia. There are a few types of wobbegongs, but the spotted wobbegong and gulf wobbegongs are endemic to Australia. They inhabit in shallow, inshore waters (less than 100m deep) in areas with rock and weed, coral reef lagoons, and continental shelves.

Wobbegong's mating season is believed to occur in the summer months (similarly for white sharks too), which also coincides with higher human activity in waters. During this period, male sharks may display more territorial behaviour leading to increased aggressiveness if their territory is threatened.

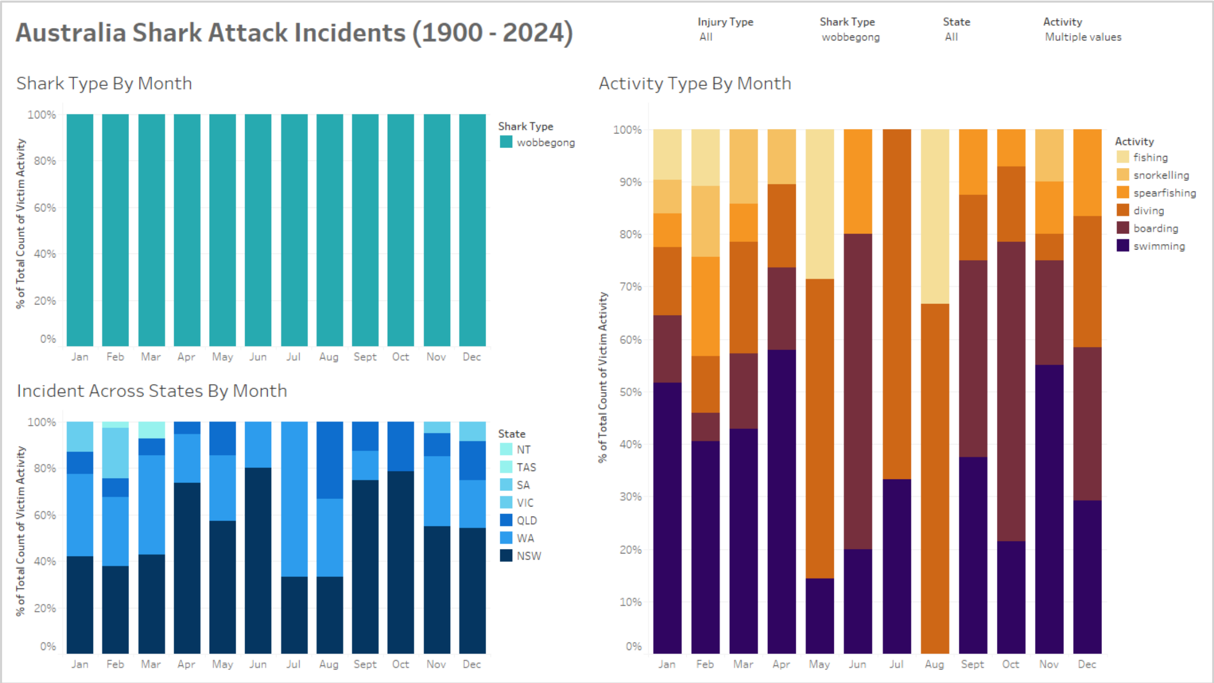

Figure 13: Incidents across states and activity type for wobbegong shark.

In the activity trends, we can clearly see that the pattern differs greatly for wobbegong sharks as compared to white shark. Boarding does not dominate the trend, and instead we see a mix of swimming and diving being the top activity depending on the month. In November to April (the warmer half of the year), swimming is the top activity, whereas in May to Aug, diving is the dominating activity. In August (the coldest month), there is only fishing and diving related incidents as water temperatures in New South Wales can hit as low as 13°C to 16°C, which is possible for divers diving in drysuits, but most swimmers/surfers will generally avoid.

The key area for wobbegong shark encounters is New South Wales, which is similar to white shark even though the geographical trend is not as distinctive. However, we can see that wobbegong sharks are mostly found in 2 – 3 states most of the time (New South Wales, Western Australia and Queensland), as compared to the 4 – 5 states for white shark.

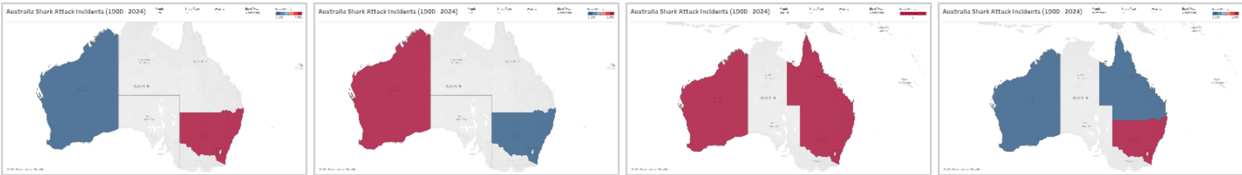

Figure 14: Wobbegong shark incidents across different states in June, July, August and September (from left to right).

PREDICTIVE MODEL

The original intention of this analysis was to use time series ARIMA/SARIMA model to predict safer months for users. However, due to the limited size of the dataset (1,200 samples) and lack of consistent data by month and even by quarters in recent years, we are unable to use time series modelling. The lack of data mainly falls within the winter months from May to July where there is less human activity in water.

First, we test for baseline accuracy score using Decision Tree Classifier with only 1split (simplest possible tree). It uses the best single feature to make predictions to see if a single feature can outperform randomness.

Decision Tree Classifier Accuracy: 0.62

MODEL CHOICES

The dataset was put through 3 different models, both before and after scaling or tuning to determine which one has the best scores. The models tested are Logistic Regression (LR), Random Forest Classifier (RF) and XGBoost.

```
Classification Report for Logistic Regression (without scaler):
              precision    recall  f1-score   support

           0       0.59      0.43      0.49        47
           1       0.70      0.82      0.76       152
           2       0.42      0.30      0.35        47

    accuracy                           0.65       246
   macro avg       0.57      0.52      0.53       246
weighted avg       0.62      0.65      0.63       246

Classification Report for Logistic Regression (with scaler):
              precision    recall  f1-score   support

           0       0.56      0.40      0.47        47
           1       0.70      0.80      0.74       152
           2       0.39      0.32      0.35        47

    accuracy                           0.63       246
   macro avg       0.55      0.51      0.52       246
weighted avg       0.61      0.63      0.62       246
```

```
Classification Report for Random Forest (before tuning):
              precision    recall  f1-score   support

           0       0.47      0.36      0.41        47
           1       0.73      0.75      0.74       152
           2       0.49      0.55      0.52        47

    accuracy                           0.64       246
   macro avg       0.56      0.55      0.56       246
weighted avg       0.63      0.64      0.63       246

Classification Report for Random Forest (after tuning):
              precision    recall  f1-score   support

           0       0.64      0.38      0.48        47
           1       0.75      0.77      0.76       152
           2       0.48      0.64      0.55        47

    accuracy                           0.67       246
   macro avg       0.63      0.60      0.60       246
weighted avg       0.68      0.67      0.67       246
```

```
Classification Report for XGBoost (before tuning):
              precision    recall  f1-score   support

           0       0.62      0.38      0.47        47
           1       0.78      0.59      0.67       152
           2       0.37      0.79      0.50        47

    accuracy                           0.59       246
   macro avg       0.59      0.59      0.55       246
weighted avg       0.67      0.59      0.60       246

Classification Report for XGBoost (after tuning):
              precision    recall  f1-score   support

           0       0.59      0.40      0.48        47
           1       0.79      0.62      0.69       152
           2       0.37      0.74      0.49        47

    accuracy                           0.60       246
   macro avg       0.58      0.59      0.56       246
weighted avg       0.67      0.60      0.61       246
```

Figure 15: Classification reports for Logistic Regression, Random Forest Classifier and XGBoost (from left to right) before and after scaling/tuning.

For our analysis, we maintained the following class weightage as a constant amongst all model versions due to the sample size and the critical nature of Class 2 (Fatal). The class weightage is derived from ratio of each injury type within the dataset.

Class Types	Class Name	Class Weight
Class 0	Uninjured	1
Class 1	Injured	1
Class 2	Fatal	3

Model	Parameters Before Tuning/Scaling	Parameters After Tuning/Scaling
LR	Max_iter = 1000	Max_iter = 1000 Used StandardScaler()
RF	Class_weight = {0:1, 1:1, 2:3} N_estimators = 100	Class_weight = {0:1, 1:1, 2:3} N_estimators = **200** Max_depth = **None** Min_samples_split = **5** Max_features = **'sqrt'**
XGBoost	Class_weight = {0:1, 1:1, 2:3} N_estimators = 100 Max_depth = 3 Learning_rate = 0.1	Class_weight = {0:1, 1:1, 2:3} N_estimators = 100 Max_depth = 3 Learning_rate = 0.1 Subsample = **0.8** Colsample_bytree = **0.8**

KEY METRICS

While accuracy score is a common approach for comparison, we also want to have higher importance for other metrics due to the critical nature of predicting fatalities (Class 2).

1. **Precision** (TP / (TP + FP))
   - Interpretation: *"How many predicted 'injured' attacks were correct?"*
   - High precision = Few false alarms (e.g., 0.75 for "injured" means 75% of predicted "injured" cases were correct).
2. **Recall (Sensitivity)** (TP / (TP + FN))
   - Interpretation: *"How many actual 'injured' attacks were caught?"*
   - High recall = Few missed attacks (e.g., 0.83 for "injured" means 83% of actual "injured" cases were detected).
3. **F1-Score** (2 * (Precision * Recall) / (Precision + Recall))
   - Balances precision and recall.
   - Useful for imbalanced data (e.g., 0.79 for "injured").

In our analysis, we will be focusing on achieving a high recall rate for Class 2 (fatality) as missing attacks is costly (high consequence), but at the same time the precision rate should be decent to reduce false alarms.

KEY METRICS COMPARISON

We picked the better version of each model for comparison of key metrics, based on the version that has a higher recall score for Class 2.

Model	Recall (Class 2)	Precision (Class 2)	F1-Score (Class 2)	F1-Score (Class 1)	Accuracy	Key Advantage
RF (Tuned)	0.64	**0.48**	**0.55**	**0.76**	**0.67**	Best overall accuracy
XGBoost (Tuned)	**0.74**	0.37	0.49	0.69	**0.60**	Best fatality detection
LR (Scaled)	0.32	0.39	0.35	0.74	0.63	Worst fatality detection
Baseline	-	-	-	-	0.62	

Based on the results, Random Forest Classifier (Tuned) emerged as the best choice for our shark attack prediction model as it optimally trades off fatality detection (64% recall) with reasonable false alarms (48% precision), while maintaining high overall accuracy (67%).

1. **Best Overall Accuracy (67%)**
    - Higher than untuned RF (64%) and tuned XGBoost (60%, which falls below baseline benchmark).
2. **Balanced Fatality Detection**
    - **64% recall** - catches most fatalities.
    - **48% precision** - good precision (48%) considering the high recall and fewer false alarms than XGBoost's 37%.
3. **Robust Performance**
    - Best F1-score for both fatalities (55%) and injuries (76%).

In most cases, we would expect XGBoost to perform better than Random Forest, especially if it is properly tuned. However, RF performs better on small datasets by averaging predictions across many shallow trees and reduces variances whereas XGBoost requires more data to optimize its gradient boosting approach. RF also has a natural advantage of being less prone to overfitting the minority class (fatalities in our case). XGBoost on the other hand tends to overfit despite having class weights. If our dataset is much larger (more than 10,000 data points), like in the case of a Global Shark Attack Database, XGBoost might work better.

LIMITATIONS & RECOMMENDATIONS

KEY DATA LIMITATIONS

1. **Limited Dataset**
   - Small dataset of ~1,200 samples, limits the usage of more powerful predictive models.
   - Imbalanced classes: Most critical Class 2 (Fatalities) have only around a third of the sample size compared to Class 1 (Injured). Even when class weightage is used for predictive modelling, there is still some trade-off between recall and precision.

2. **Missing Factors and Data**
   - Environmental factors like water temperature, tide conditions and weather data
   - Human activity metrics like tourist density, beach attendance and activity-specific participation
   - Shark behaviour data like migratory patterns, breeding seasons and species-specific presence

RECOMMENDATIONS

Our best model Random Forest Classifier having a 48% precision means that more than 50% of predicted fatalities are false alarms, hence we recommend the following:

1. Deploy the model with a confidence threshold, e.g. only warn lifeguards if fatality probability is more than 60%.
2. Work with marine life experts to finetune the model based on shark behaviour patterns.
3. Improve model accuracy by factoring environmental and human activity metrics.

Other than building a chatbot with the predictive model at the backend, other future improvements could be:

1. Establish standardised reporting protocols for community or hospitals.
2. Develop community reporting apps to capture near-miss incidents and even sightings.
3. Develop a mobile app for lifeguards with location-based risk alerts based on live sightings.

CONCLUSION

SUMMARY OF FINDINGS

1. Attacks peak during **warmer months**.
2. **New South Wales** and **Queensland** are high-risk zones.
3. Most attacks are **unprovoked**.
4. Activities at water surface like **swimming and surfing had higher risks** vs. diving at depth.

SUMMARY FOR PREDICTIVE MODEL

Random Forest Classifier had the highest performance of all models tested:

1. **Highest accuracy** among all models (67%).
2. **Best fatality detection** among all models with 64% recall for fatalities (identifying most high-risk cases) and 48% precision (fewer false alarms).
3. **Most robust performance** among all models with the best F1-score for both fatalities (55%) and injuries (76%).

FINAL THOUGHTS

While no model can perfectly predict shark attacks, this analysis provides actionable insights to **reduce risk and improve emergency response**. The tuned Random Forest offers the best balance between **detecting fatalities** and **minimizing false alarms**, making it suitable for pilot deployment in high-risk regions.

WORKS CITED

Zach S. R. Clark, P. A. (2024). *Genomic Assessment of Australian White Sharks (Carcharodon carcharias) Challenges Previous Evidence of Population Subdivision*. Retrieved from ResearchGate: https://www.researchgate.net/publication/389051262_Genomic_Assessment_of_Australian_White_Sharks_Carcharodon_carcharias_Challenges_Previous_Evidence_of_Population_Subdivision