|  |
| --- |
| Business Intelligence Project Report – Surfing Business |
| Christopher Goodfellow |
|  |

5/30/2018

Table of Contents

**1.0 Executive Summary3**

**2.0 Problems Being Addressed4**

**3.0 Description of Data4**

**4.0 Analysis Conducted5**

**5.0 Results5**

**6.0 Recommendations7**

# Executive Summary

This report answers a number of business questions, in regards to the surfing experience business client. The first question involves understanding the optimal periods of the year to run the service. The second pertains finding relevant real-time data to allow the business to make an informed decision as to which beach, north or south, they would prefer to take their customers to, on any particular day. The final question is in regards to finding a way for the business to be able to guarantee their customers a certain amount of time at the beach itself, considering they leave the Brisbane CBD at 5:00am and are dropped back at 12:30pm. Using Python and various data analysis techniques, this report has answered these questions and provided visualisations. Findings and recommendations include the fact that, based off of wave height, the first 6 months and the last 2 months of the year are the best periods to offer this service, with the Sunshine Coast consistently offering higher waves. Therefore, the business should increase its marketing and operations during these periods, to maximise customer satisfaction. Using twitter sentiment analysis, it was found that the Gold Coast had a slight edge over the Sunshine Coast in terms of positive sentiment, with the recommendation being to compare the two results from both of the beaches and take into account the ratio of positive/negative sentiment for each beach and how the results may change over time. Utilising the Google Maps API, the current time it would take to reach either beach could be retrieved, with the recommendation being that the business checks this data each day before departure, to be able to tell if there has been a collision or other event that has slowed down traffic enormously, to either location. Additionally, it will make it possible for the business to be able to determine how long they can guarantee their customers will be able to surf, by knowing when to leave to make it back to the Brisbane CBD by 12:30pm. Finally, by employing the use of the last technique, web-scraping, the current wave and weather data of both beaches can be extracted. Should the business incorporate this data into the decision making process each morning before they leave, this should be enough to allow them to make an informed decision as to which beach would be more suitable to travel to on that day.

# 2.0 Problems Being Addressed

There are a number of problems that will be addressed in this report. This first is pinpointing and recommending the best period of the year to run the surfing experience service. The second is to try and harness real-time data that would help the business in the daily decision of which beach, north or south, would be more desirable on any particular day for their customers. The final issue this report will address is providing the surfing business with the means of working out how long of a surfing experience they can guarantee customers if they are leaving at 5:00am and dropping them back to the Brisbane CBD by 12:30pm.

# 3.0 Description of Data

Due to the nature of the business questions that were posed, most of the data that was used came from live sources, such as websites and API’s. The only historical data used in this analysis was wave height data of both the Sunshine Coast (Mooloolaba) and the Gold Coast. This data was selected in an effort to try and reveal which parts of the year would be best suited to run the surfing experience business in. The data examined was organised into separate csv files and came from the years 2015-2017 and consisted of thousands of rows of wave heights (ft) collected over this time period. This data was helpful as it was able to be manipulated to reveal the average highs and lows of wave height throughout certain time periods.

The twitter API was also used to extract tweets pertaining to certain hash tags such as ‘#sunshinecoast’ and ‘#surfersparadise’. The reasons for this data selection included the fact that the business could potentially be shown the latest tweets regarding the beaches, notifying them of any potential factors that could impact their business whilst also allowing for sentiment analysis of both beaches. This would be helpful as the business would be able to keep track of how public perception of both beaches are changing and whether this should be taken into account when deciding on a daily basis which beach to travel to.

Another API, this time Google Maps, was also employed in this analysis. It was selected as the data could be trusted to be up-to-date and accurate. Its use case in this analysis was the ability to extract the predicted time it would take to reach either locations, at any given time. This information is crucial to the business, who would value the knowledge of estimated traffic conditions in either direction before they left at 5:00am.

Finally, data was web-scraped off of a popular website which tracks, in real-time, the wave and weather conditions of the world’s most popular beaches. This data was selected as it is imperative that the business be aware of both the weather and wave conditions, of both beaches, before they pick their customers up in the morning.

There were no ethical considerations to take into account in this analysis, as all data used was wilfully made available to the public.

# 4.0 Analysis Conducted

In regards to the historical data concerning wave height (ft), recorded at both beaches, the data focused on examining one wave height value per day, during the time customers would actually be surfing, 8am. Monthly averages were then found for each beach, for each year (2015-2017), which were then integrated, leaving the monthly average wave height for both beaches, over the course of 3 years.

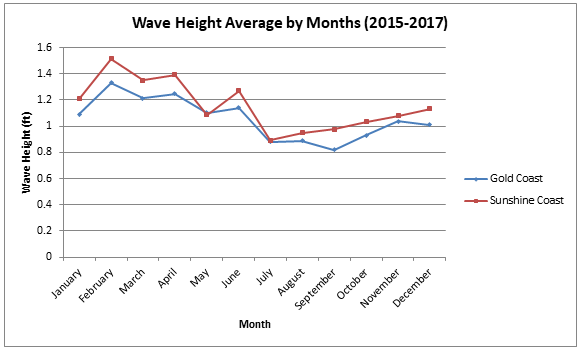
Using the Twitter API, tweets were extracted and the most recent 50 tweets containing the relevant hash tags were analysed for current sentiment (positive, neutral and negative).

Utilising the Google Maps API, the total distance in kilometres from the Brisbane CDB to both beaches was retrieved, as well as the current estimated time it would take to travel there by car.

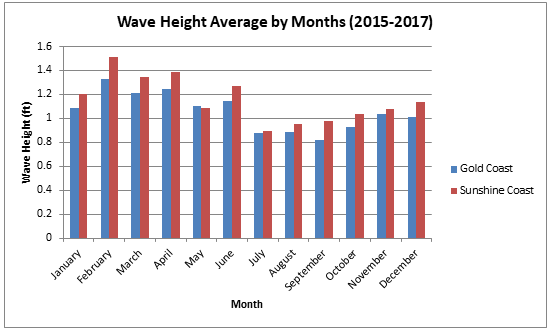
Finally, using web-scraping techniques, the current wave and weather conditions were retrieved from a surfing website ‘Magic Seaweed’. This was achieved by parsing through the website’s HTML code and extracting the relevant information.

# 5.0 Results

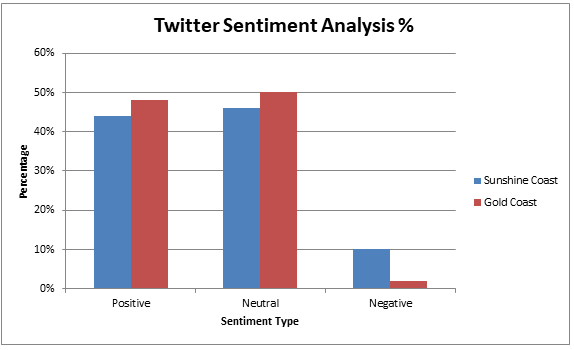
The first set of visualisations pertains to the historical wave height data of both beaches. As shown in the line graph below, it is clear that the first six months of the year have had, on average, bigger wave heights during the last three years (2015-2017). It is also clear that the Sunshine Coast consistently ranked higher than the Gold Coast. Noticeably, the final two months of the year for both beaches seem to improve in terms of average wave height.



A bar chart further aids in visualising these findings. The month of May is the only period in which the Gold Coast had a bigger wave height average.



Below is a bar chart of the percentage of positive, neutral and negative tweets attributed to the Sunshine Coast (blue) and the Gold Coast (red). As of the 27th May 2018, it is obvious that the Gold Coast holds an edge over the Sunshine Coast in sentiment, with negative sentiment 5 times more prevalent (10% compared to 2%). Due to the changing nature of sentiment, it would be ideal to track the changes of these results over time, to identify if an area is improving or worsening in the eyes of the public.



The results from the next analysis were the distance and travel duration it would take to reach both of these locations from the Brisbane CBD. The example shown below was taken at 6pm 27th May 2018. As this analysis technique relies on live data, the duration result will fluctuate throughout the day. This will be useful for the business, as they will have the ability to check the current time it would take to reach either location before they pick up their clients at 5am. Most importantly, it may also alert the business to the fact there may have been a collision or another incident that has resulted in heavy traffic congestion, in either direction.

|  |  |  |
| --- | --- | --- |
| *6pm 27th May 2018* | **Distance** | **Duration** |
| **Sunshine Coast** | 104 km | 1 hour 20 mins |
| **Gold Coast** | 79.1 km | 1 hour 5 mins |

The final results include the live wave and weather conditions on both the Sunshine and Gold Coast. The particular example shown below was extracted at 5pm 27th May 2018, however much like the Google Maps API data, the business must check this for themselves each morning to know the current conditions of the beaches, to help them make an informed decision of where to take their customers that day.

*The current wave conditions at the Gold Coast are:* ***2-3ft.*** *The current weather conditions at the Gold Coast are:* ***Clear.***

*The current wave conditions at the Sunshine Coast are:* ***3-5ft.*** *The current weather conditions at the Sunshine Coast are:* ***Clear.***

# 6.0 Recommendations

There are a number of recommendations that should be considered on account of this analysis. Firstly, in regards to the problem of determining when the best time of year it would be to run the service, it is clear that the first 6 months, and to a degree the final 2 months of the year, yield the largest waves, with the Sunshine Coast consistently producing higher waves than the Gold Coast. Therefore, the business should consider focusing its marketing and operations around this time period, to maximise customer satisfaction. In regards to the twitter sentiment analysis technique, my recommendation is to compare the two results from both of the beaches hash tags and take into account the ratio of positive/negative sentiment for each beach and how the results may change over time. Should drastic changes occur in the sentiment of either beach, this factor could be taken into account when deciding which beach should be preferred for the customers. Utilising the live travel time provided by the Google Maps API, my recommendation would be that the business check this data each day before departure, to be able to tell if there has been a collision or other event that has slowed down traffic enormously, to either location. Additionally, it will make it possible for the business to be able to determine how long they can guarantee their customers will be able to surf, by knowing when to leave to make it back to the Brisbane CBD by 12:30pm. It will also give the business another variable to consider when they make their decision incorporating the weather and wave height situation of each beach. That particular data is also available for the business to check each morning through the use of web-scraping live data. Using these tools and techniques, the business will not only have an idea of the best time of year to run the business, but also will have the necessary information to make an informed decision each morning as to what beach they should take their customers, on any particular day.