**Coursera Capstone**

**IBM Applied Data Science Certification**

***Opening a New Steakhouse in Sydney, Australia***

By: Sanya Srivastava

February 2020

A large body of water with a city in the background

Description automatically generated

**Introduction**

Steak is a popular meat that is consumed in Sydney, Australia. However, in recent years, the popularity of beef is declining. As of 2016, the average consumption of beef in Australia is 24.68 kilos, becoming the 3rd most consumed meat after pork and chicken. The consumption of meat has declined in Australia due to the lack of Steakhouses in the Sydney area. Opening a steakhouse can increase the consumption of beef in the future.

**Business Problem / Target Audience**

The objective of this capstone project is to find an area in Sydney, Australia to open a new Steakhouse. Using data science skills, we answer the question: In the city of Sydney, Australia, if a steakhouse restaurant owner wants to open a restaurant, where would you recommend opening it?

This project is targeting families in the Sydney, Australia. The owner is attempting to open a steakhouse that is family friendly. This steakhouse can compete with Sydney’s popular restaurant, Kinsley’s CBD Steakhouse.

**Data**

**To solve the problem, we will need the following data:**

* List of neighborhoods in Sydney, Australia.
* Longitude and latitude coordinates of the neighborhoods.
* Venue data

**Source of data and methods to extract them**

The Wikipedia page (<https://en.wikipedia.org/wiki/Category:Suburbs_of_Sydney>) contains a list of neighborhoods in Sydney, Australia with 200 neighborhoods. I used the beautifulsoup packages to scrape the data and Python packages for insight on the data. Additionally, I used Foursquare API to get venue data of the neighborhoods.

Foursquare API allows me to see the venue data, especially for the Steakhouses in Australia. Web scraping (Wikipedia), Foursquare API, Machine Learning, and more are being used for finding where to place the next best Steakhouse.

**Methodology**

The first thing I did was get the neighborhoods in Sydney; Australia on the Wikipedia page provided on the data section of this report. After that, I used the beautifulsoup package to get the list of neighborhoods. Once the neighborhoods were accessed, I used Foursquare API and the geocoder package on python to attain the longitude and latitude. However, we still do not have enough data understanding. Thus, I used the Folium package to data visualize the neighborhood on the map.

After, I used Foursquare API to find the top 100 venues within a 3000-mile radius. Since I had a Foursquare Developer Account, I used the Foursquare Client ID and secret key. Foursquare returned the venue name, category, latitude and longitude. Once I was able to see how many venues there are, I filtered out “Steakhouse” for the neighborhoods. I was able to get all the information for Steakhouses in Australia.

Finally, I used k-means clustering on the data I have acquired. I used this machine learning algorithm because it is one of the algorithms that is the easiest to use in a mapping situation. I used 3 clusters based on the “Steakhouse” amount in Sydney, Australia. Based on how many steakhouses there are in different neighborhoods, I could see what cluster needs a new or new steakhouse.

**Results**

The results for the k-means clustering algorithm on the “Steakhouses” in Sydney, Australia shows three clusters based on this specific venue frequency:

* Cluster 0: Low number of steakhouses.
* Cluster 1: Low number of steakhouses
* Cluster 2: High number of steakhouses

Below us a folium map that shows the cluster of steakhouses in neighborhoods.

A close up of a map

Description automatically generated

**Discussion**

Based on my results, cluster 2 has the most steakhouses and could have intense competition. Therefore, it is not beneficial for the steakhouse to be built in this cluster, since most likely it will be unable to do well. However, clusters 1 and cluster 0 have low amount of competition for Steakhouses, therefore it will be beneficial for a new steakhouse to be located in cluster 1 and cluster 0. This can help a steakhouse restaurant thrive in these clusters since high concentration of steakhouses will cause this restaurant to fail.

**Conclusion**

This project focuses on building a new steakhouse restaurant in Sydney, Australia to expand the beef market in Australia. From downloading the data required, extracting and preparing the data and using a machine learning algorithm known as k-means clustering, using three clusters to find where to place the next steakhouse. I found that clusters 1 and 0 should have the next steakhouse to profit and be successful in Sydney, Australia.