

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Restaurant in Auckland, New Zealand



By: Hardik Modha
July 2020

Introduction

The population migration, tourism and busy lifestyle in the financial hub of New Zealand, i.e. Auckland, lead to a surge in a food serving business. The restaurants provide ease to singles, working parents, older people and tourists who have some inconvenience with cooking. It shows that this food industry is of bright future in terms of monetary return. As a result, there are many restaurants in the Auckland suburbs, and more will come. The reality is that many such restaurants fail in a few years of their business due to poor planning. The prime factor of this failure is the location of the business. As this industry is highly competitive, serious consideration should be given when deciding the place of business.

Problem Statement

The main objective of this project is to analyze the location of existing restaurants and to select the best place for a new restaurant in Auckland. By using the data science methodology and machine learning techniques, this project aims at providing a solution to the following question:

In Auckland, if an entrepreneur or investor is looking to establish a new restaurant, where would you recommend it?

Target Audience

New Zealand is the popular tourist destination in the world whose tourism industry and hospitality sector are one of the significant contributors to the GDP. Also, New Zealand is a prominent venue for many international events, and concerts, which are generally held in Auckland. This makes Auckland the first choice for the investor for any business starts up. This project is prepared by mainly targeting the local entrepreneur who is a novice in this industry and also to the financial investors in making an investment decision for opening a new restaurant.

Data

To solve this problem, the following data will be needed:

- List of suburbs of Auckland, New Zealand. These suburbs are termed as “Neighborhoods” in this project.
- Latitude and Longitude of those neighbourhoods. This coordinates will help in plotting the map and getting the venue data.
- Venue data for the restaurants. This data will be used to perform clusters in these neighbourhoods.

Data Collection and Extraction Method

The Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Auckland) contains a list of neighbourhoods in Auckland, which will be extracted by using the method of web scraping techniques with the help of Python 'Beautifulsoup' packages. After that, we will get the geographical coordinates of the neighbourhoods using Python 'Geocoder' package, which will give us the latitude and Longitude of the neighbourhoods.

Once the coordinate is collected, we will use Foursquare API to get the venue data for those neighbourhoods. The reason for using Foursquare API is that it has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data; we are particularly interested in the restaurant category in order to help us to solve the business problem put forward.

This project will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

In the next section, we will present the Methodology section, where we will discuss the steps taken in this project for the data analysis that we did and the machine learning technique that was used.

Methodology

Firstly, we need to get the list of neighbourhoods in the city of Auckland from the Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Auckland). We will do web scraping by using Python 'requests' and 'beautifulsoup' packages to extract the list of neighbourhoods data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and Longitude to be able to use Foursquare API. To do so, we will use the 'Geocoder' package that will allow us to convert an address into geographical coordinates in the form of latitude and Longitude. After gathering the data, we will populate the data into a pandas Data Frame and then visualize the neighbourhoods in a map using 'Folium' package. This allow us to perform a sanity check to make sure that the geographical coordinate's data returned by 'Geocoder' are correctly plotted in the city of Auckland.

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical Coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format, and we will extract the venue name, venue category, venue latitude and Longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyze each neighbourhood by grouping the rows by neighbourhood and taking the mean of the

frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analyzing the “Restaurant” data, we will filter the “Restaurant” as venue category for the neighbourhoods.

Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering the algorithm identifies k number of centroids, and then allocates every data point to the nearest Cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this Project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for “Restaurant”. The results will allow us to identify which neighbourhoods have a higher concentration of restaurants while which neighbourhoods have a fewer number of restaurants. Based on the occurrence of restaurants in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new restaurants.

Finding

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Restaurants”:

- Cluster 0: Neighborhoods with Low number of restaurants
- Cluster 1: Neighborhoods with moderate to a high number of restaurants
- Cluster 2: Neighborhoods with a high concentration of restaurants

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour.



Discussion

As observations noted from the map in the finding section, most of the restaurants are concentrated in the central area of Auckland city, with the highest number in cluster 2 and moderate number in cluster 1. On the other hand, cluster 0 has a very low number to no restaurants in the neighbourhoods. This represents a great opportunity and high potential areas to open new restaurants as it has very little to no competition from existing restaurants. Meanwhile, restaurants in cluster 2 are likely suffering from intense competition due to oversupply and high concentration of restaurants.

From another perspective, the results also show that the oversupply of restaurants mostly happened in the central area of the city, with the suburb area still have very few restaurants. Therefore, this project recommends property developers to capitalize on these findings to open new restaurants in neighbourhoods in cluster 0 with little to no competition. Investors with unique selling propositions to stand out from the competition can also open new restaurants in neighbourhoods in cluster 1 with moderate to high competition. Lastly, Investors are advised to avoid neighbourhoods in cluster 2 which already have a high concentration of restaurants and suffering from intense competition.

Limitation and Suggestion for Future Researcher

In this project, we only consider one factor, i.e. frequency of occurrence of restaurants; there are other factors such as population and income of residents that could influence the location decision of a new restaurant. However, to the best knowledge of this researcher, such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new restaurant. Besides, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of a paid account to bypass these limitations and obtain more results.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders, i.e. Local entrepreneurs and investors regarding the best locations to open a new restaurant. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 are the most preferred locations to open a new restaurant. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new restaurant.

Appendix

Clusters

Neighborhoods in Cluster – 0

| | | | |
|--------------------|-------------------|--------------------|---------------------|
| Airport Oaks | Sandringham, | Murrays Bay | Te Atatū South |
| Ponsonby | Highbrook | Ōtara | Te Papapa |
| Remuera | Herne Bay | Shelly Park | Birkenhead |
| Manukau | Henderson North | Hauraki | Birkdale |
| Mairangi Bay | Orakei | St Johns | Belmont |
| Lynfield | Manurewa East | Grey Lynn | Beach Haven |
| Long Bay | Schnapper Rock | Te Atatū | Bayswater |
| Rosebank, Auckland | Marlborough | Clover Park | Balmoral |
| Laingholm | Matakatia | Clevedon | Avondale |
| Konini | Onehunga | Te Atatū Peninsula | Auckland waterfront |
| Rosehill | One Tree Hill | Cheltenham | Totara Heights |
| Rothsay Bay | Northcross | Chatswood | Ardmore |
| Royal Heights | Oratia | Chapel Downs | Totara Vale |
| Kauri Park | Northcote Central | Castor Bay | Unsworth Heights |
| Hunua | North Harbour | Campbells Bay | Viaduct Harbour |
| Howick | Orere Point | Bucklands Beach | Albany |
| Homai | Oteha | Point Chevalier | McLaren Park |

| | | | |
|------------------------|-----------------|-------------------|---------------|
| Hobsonville | Owairaka | Pinehill | Mechanics Bay |
| Hingaia | New Lynn | Parnell | Penrose |
| Hillsborough, Auckland | Pahurehure | Mission Bay | Parau |
| Hillpark, Auckland | Māngere East | Papatoetoe | Morningside |
| Saint Marys Bay | Māngere Bridge | Pakuranga Heights | Mount Eden |
| Highland Park | Pakuranga | Crown Hill | Dannemora |
| Conifer Grove | Duders Point | Stanley Bay | Greenlane |
| Stanmore Bay | Green Bay | Grafton | Sunnyhills |
| Glenfield | Glenfield North | Glendowie | Sunnynook |
| Glen Innes | Glen Eden | Sunnyvale | Devonport |
| Oranga | Ellerslie | East Coast Bays | Flat Bush |
| Favona | East Tamaki | Takapuna | Tamaki |
| Fairview Heights | Epsom | Swanson | Eden Valley |
| Eden Terrace | | | |

Neighborhoods in Cluster – 1

| | | | |
|--------------|------------|-----------------|------------|
| Forrest Hill | Farm Cove | Kelston | Kohimarama |
| Cockle Bay | Lincoln | Manukau Heights | Manurewa |
| Bayview | Browns Bay | Newmarket | Māngere |
| Massey | Ōtāhuhu | Freemans Bay | Henderson |

| | | | |
|----------|---------------|----------------|----------------|
| Rānui | Saint Heliers | Point England | Papakura North |
| Sandspit | Three Kings | Torbay Heights | Titirangi |

Neighborhoods in Cluster – 2

| | | | |
|-----------------------|---------------|-------------------|-----------------|
| Blockhouse Bay | Auckland CBD | New Windsor | Eastern Beach |
| Goodwood Heights | Stillwater | Newton | Greenmount |
| Glendene, New Zealand | Brookby | Torbay | The Gardens |
| Ōpaheke | Stonefields | Arch Hill | Army Bay |
| Wai o Taiki Bay | Vauxhall | Algies Bay | Clendon Park |
| Takanini | Alfriston | Tamaki City | Narrow Neck |
| Mount Wellington | Mount Roskill | Mount Albert | Millwater |
| Paremoremo | Milford | Middlemore | Mellons Bay |
| Meadowbank | One Tree Hill | Red Beach | Red Hill |
| Rosedale | Kingsland | Keri Hill | Royal Oak |
| Huia | Hillcrest | Highbury | Hatfields Beach |
| Northcote | Southdown | Greenwoods Corner | Greenhithe |
| Half Moon Bay | | | |