Best Neighbourhood for an Indian Restaurant

H. Dewhurst

18 January 2021

Introduction

Business Problem:

In this hypothetical scenario a successful restaurateur is looking to expand their business into a new city. Their Indian restaurant has been the talk of Wellington City for 25 years and now the eldest daughter is keen to open and run a restaurant of her own. She would like to expand the family business into Auckland, New Zealand's largest city.

Auckland is home to 1.6 million people covering an area 1,000 km2. The daughter is confident she can establish a restaurant and build a reputation in fine dining that will rival her parents in Wellington. However, not knowing Auckland city well, she is keen to identify the best neighbourhood to open her first fine dining Indian restaurant.

Based on the family's experiences with their restaurant in Wellington, seeking out an area that already attracts customers who enjoy Indian cuisine would be a good indicator for the location of the new restaurant.

Approach:

For the purposes of this project this problem will be investigated using the Foursquare API to identify which neighbourhoods meet the preferences outlined by the client.

Data Collection

Part 1 - Data Preparation:

In order to utilize the Foursquare API the following information was needed:

- A list of Auckland neighbourhoods
- The post codes of each neighbourhood
- Latitude and longitude co-ordinates for each post code.

This data was found at the following website:

https://www.geonames.org/postalcode-search.html?q=auckland&country=NZ .

The data was scrapped and then cleaned through a series of steps into a data frame of 4 features, Neighbourhood, Postcode, Latitude, and Longitude.

The Folium library was then used as a visual aid while examining the data, rendering maps of Auckland's neighbourhoods. Later in the analysis folium was be used to show the different venue clusters that were identified using machine learning methods.

Once the neighbourhood data was arranged into a workable data frame, The Foursquare API software identified:

- The different types of venues (e.g., restaurants, parks) in each neighbourhood.
- The popularity of each venue in each of these neighbourhoods.
- The focus of our search will be on the Food category, specifically on Indian cuisine.

To do this the API to build a list of venues of up to 100 candidates, within a 500-metre radius of each of Auckland's neighbourhood centroids. The number of candidates was limited to 100 as that is the maximum number allowed as a guest user of this website.

The venue search include only data from the "Food" and "Indian restaurant" categories (Table: 1). The category hierarchy ID codes were taken from: https://developer.foursquare.com/docs/build-with-foursquare/categories/

Food ID: “4d4b7105d754a06374d81259”	
Indian restaurant categories ID: “4bf58dd8d48988d10f941735”	
Subset Category	Category ID:
Andhra Restaurant	54135bf5e4b08f3d2429dfe5
Awadhi Restaurant	54135bf5e4b08f3d2429dff3
Bengali Restaurant	54135bf5e4b08f3d2429dff5
Chaat Place	54135bf5e4b08f3d2429dfe2
Chettinad Restaurant	54135bf5e4b08f3d2429dff2
Dhaba	54135bf5e4b08f3d2429dfe1
Dosa Place	54135bf5e4b08f3d2429dfe3
Goan Restaurant	54135bf5e4b08f3d2429dfe8
Gujarati Restaurant	54135bf5e4b08f3d2429dfe9
Hyderabadi Restaurant	54135bf5e4b08f3d2429dfe6
Indian Chinese Restaurant	54135bf5e4b08f3d2429dfdf
Indian Sweet Shop	54135bf5e4b08f3d2429dfe4
Irani Cafe	54135bf5e4b08f3d2429dfe7
Jain Restaurant	54135bf5e4b08f3d2429dfea
Karnataka Restaurant	54135bf5e4b08f3d2429dfeb
Kerala Restaurant	54135bf5e4b08f3d2429dfed
Maharashtrian Restaurant	54135bf5e4b08f3d2429dfee
Mughlai Restaurant	54135bf5e4b08f3d2429dff4
Multicuisine Indian Restaurant	54135bf5e4b08f3d2429dfe0
North Indian Restaurant	54135bf5e4b08f3d2429dfdd
Northeast Indian Restaurant	54135bf5e4b08f3d2429dff6
Parsi Restaurant	54135bf5e4b08f3d2429dfef
Punjabi Restaurant	54135bf5e4b08f3d2429dff0
Rajasthani Restaurant	54135bf5e4b08f3d2429dff1
South Indian Restaurant	54135bf5e4b08f3d2429dfde
Udupi Restaurant	54135bf5e4b08f3d2429dfec

Table: 1. Category ID’s for “Food”, “Indian Restaurant” and “Indian Restaurant subcategories” taken from the Foursquare API library.

At the end of this process there was a data frame of 7 features x 848 rows, comprising of 63 unique categories for the 202 neighbourhoods.

Part 2 - Data Preparation:

The second part of the data preparation addresses the two remaining points the client wishes to consider.

- Attracting return patrons.
- A centrally located neighbourhood.

Attracting return patrons.

The client wishes to attract returning patrons who are seeking a high-end dining experience. An initial way to explore this question would be to look at household incomes in the potential neighbourhoods, as household incomes (or disposable incomes) may provide an indication of the frequency potential customers might seek to dine out. Higher earners may be more inclined to seek out the more exclusive dining experiences the client wishes to provide in her restaurant.

Household income data was taken from New Zealand's most recent Census (2018). These were downloaded as a.csv file from:

https://www.stats.govt.nz/ .

This data was not easily obtained and could not be "scrapped" from a webpage. It was however available as a downloadable .csv file. Once downloaded it was examined and found to only contain three features, median household income, latitude and longitude. In order to be able to match this data frame with the existing data frame prepared for Foursquare, extra information was required. It was decided that the best way to join the datasets was via "Postcodes". To do this postcodes associated with geolocations were acquired by using the centroid for each postcode's location and reverse geocoding with https://nominatim.org/ .

Missing Neighbourhood names were scrapped from https://www.nzpost.co.nz/personal/sending-within-nz/how-to-address-mail/postcodes/postcode-directory as a pdf and converted to a .csv file. This data frame was merged with the household income data frame using the "Postcode" feature.

The final data frame consisted of 5 features: Income, postcode, latitude, longitude and neighbourhood.

A centrally located neighbourhood.

The remaining point the client wished to consider was how centrally located any potential neighbourhood was within the city. Her preference is to establish her first restaurant within the inner-city neighbourhoods, although she will not dismiss and outlying neighbourhood if the conditions are favourable.

Method of Analysis

Part 1 - Analysis:

The analysis began with determining the density of restaurants in each neighbourhood, followed by grouping them into which venues were the most popular. In this case the top 10 venues for each neighbourhood were presented.

The next step was to employ K-means clustering to organise the top venues into clusters of neighbourhoods based on the most common venues. The clusters were displayed using Folium, and tables generated for each cluster.

Part 2 - Analysis:

The neighbourhoods from the selected cluster were further investigated by analysing the household income data frame and Neighbourhood Sprawl:

The data frame of median household incomes was investigated with basic statistics to determine the mean incomes for each of the selected neighbourhoods found in the "Part 1 - Analysis".

The final part of the analysis was to build a table of distances between the city centre and each of the selected neighbourhoods. Distances (km's) were determined by calculating the distance between 2 geolocations. Auckland central business district (CBD) was chosen as the centroid for Auckland City.

A second "Distance" feature was added to describe distance between the CBD and each neighbourhood of interest. "Distance via road" provide the most direct route driven to provide a more practical measure.

Results and Discussion

Foursquare revealed a total of 62 unique FOOD categories were identified in the 202 Auckland neighbourhoods that make up the greater Auckland City region. This data set was reduced to just the top 10 venues food venues and the frequency in which they appeared in each neighbourhood.

K-means clustering reduced the neighbourhoods into 3 clusters with the 2nd cluster grouping six neighbourhoods: McLaren Park, Birkdale, Paremoremo, Meadowbank, Otara and Snell's Beach together (Fig: 1).

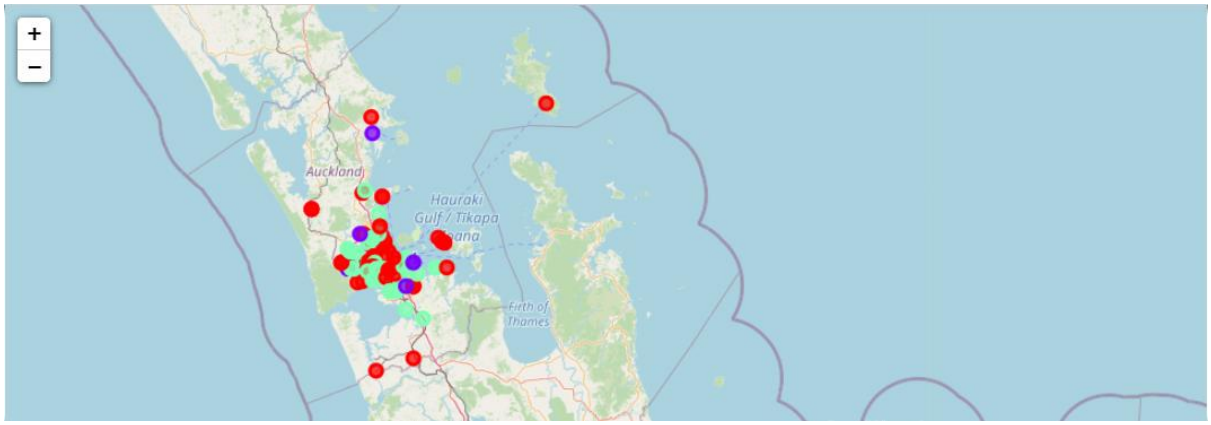

Figure: 1. Clustering of the Auckland neighbourhoods' top venues using the K-means clustering method and displayed using Folium.

Cluster 2 revealed the category "Indian restaurant" as the most popular venue for people to visit (Table: 2). A refined search of these six neighbourhoods showed that the density of Indian restaurants is low in all six neighbourhoods, with each only having one or two existing restaurants in the "Indian cuisine" category that would be considered in direct competition for customers.

Neighbourhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
McLaren Park	Indian Restaurant	Malay Restaurant	Dim Sum Restaurant	German Restaurant	Gastropub	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	Filipino Restaurant
Birkdale	Indian Restaurant	Bakery	Fast Food Restaurant	Yakitori Restaurant	Dim Sum Restaurant	German Restaurant	Gastropub	Fried Chicken Joint	French Restaurant	Food Court
Paremoremo	Indian Restaurant	Yakitori Restaurant	Dim Sum Restaurant	German Restaurant	Gastropub	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	Filipino Restaurant
Snells Beach	Indian Restaurant	Restaurant	Deli / Bodega	German Restaurant	Gastropub	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	Filipino Restaurant
Meadowbank	Indian Restaurant	Café	Yakitori Restaurant	Dim Sum Restaurant	German Restaurant	Gastropub	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop
Otara	Indian Restaurant	Yakitori Restaurant	Dim Sum Restaurant	German Restaurant	Gastropub	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	Filipino Restaurant

Table: 2. Table of Cluster 2, showing the top 10 most common venues. Clusters determined using the K-means clustering method.

An examination of median household incomes showed Birkdale ($41,684 ± 8,319.0) and Meadowbank ($35,716 ± 14,452.0) had the highest mean incomes (Fig: 2, Table: 3). This could suggest that people in these 2 neighbourhoods might be inclined to eat out more often, and perhaps seek out a high-end restaurant. However, while household income might indicate disposable income it does not take into consideration debt. Household debt would also be a factor in potential customers' willingness to eat out at a fine dining restaurant. These questions can only be addressed by further research that is beyond the scope of this project.

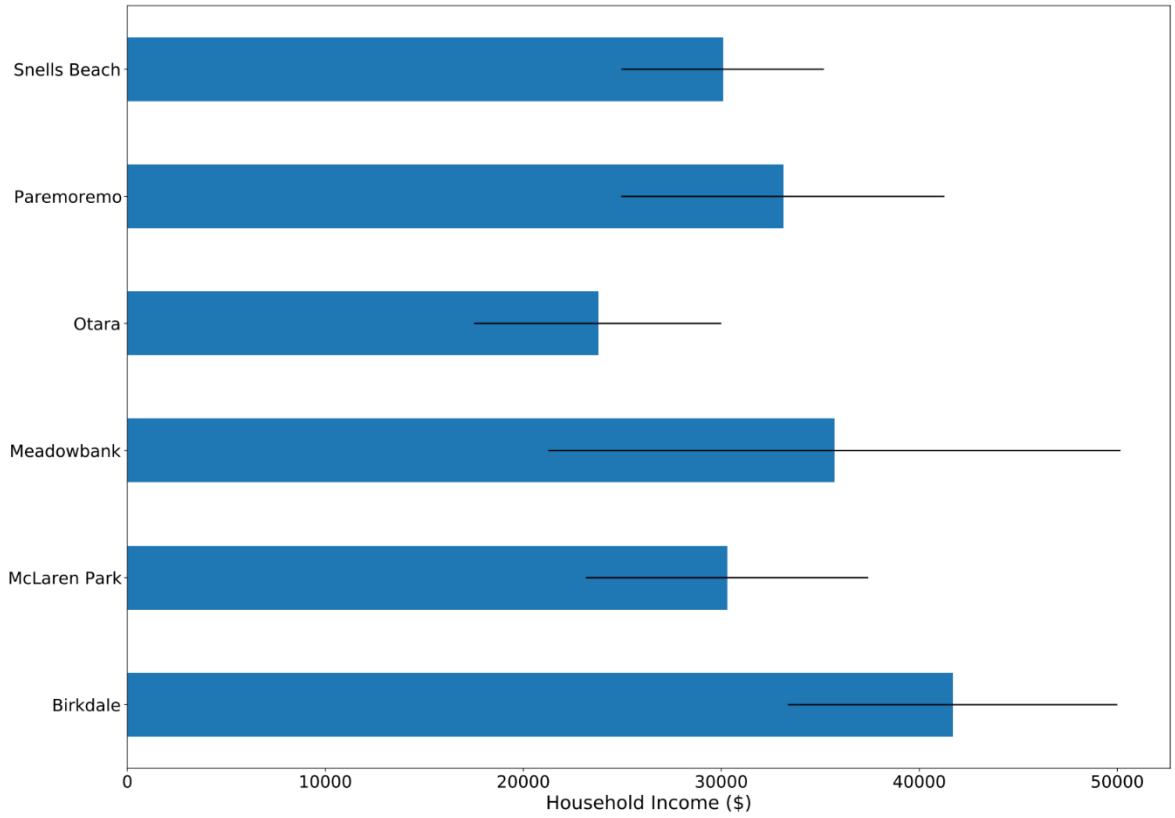

Figure: 2. Means & Standard Deviations of Household Incomes for Neighbourhood's in Cluster 2

Neighbourhood	count	mean	std	min	25%	50%	75%	max
Birkdale	5124.0	41683.606557	8318.946846	19800.0	36400.0	41800.0	47800.0	61200.0
McLaren Park	30178.0	30289.240506	7132.470791	10300.0	24400.0	29900.0	36000.0	50500.0
Meadowbank	5796.0	35715.527950	14451.863287	13500.0	22400.0	35200.0	45500.0	72200.0
Otara	15703.0	23755.951092	6244.625565	11000.0	18900.0	22200.0	28400.0	58800.0
Paremoremo	2142.0	33105.555556	8162.100243	11700.0	26900.0	32700.0	39300.0	57700.0
Snells Beach	1984.0	30071.875000	5111.953432	21300.0	26800.0	29650.0	33275.0	45600.0

Table: 3. Table of basic statistics for household incomes in neighbourhoods from Cluster 2.

Due to the sprawl of neighbourhoods in Cluster 2, the distances of each neighbourhood from the city centre are added for the client's consideration. For the purposes of this report the Auckland CBD (central business district) has been used as the centre point. Of the six neighbourhoods, Birkdale and Meadowbank are the most central to the city at 8.9 km's and 9.2 km's, respectively. Snell's Beach is the most remote at 46.6 km (Table: 4). While these distances are accurate (in a direct line) they are not realistic of how far it is to drive between any of the two given points. Therefore, the distance between the CBD and each neighbourhood via the most direct road route using Google maps has been added in the table below (Table: 5).

Neighbourhood	Income_mean	Income_stddev	Latitude	Longitude	PostCode	Km's
Birkdale	41684.0	8319.0	-36.791	174.693	0626	8.9
McLaren Park	30289.0	7132.0	-36.876	174.636	0612	11.7
Meadowbank	35716.0	14452.0	-36.879	174.859	1072	9.2
Otara	23756.0	6245.0	-36.967	174.885	2023	17.1
Paremoremo	33106.0	8162.0	-36.732	174.699	0632	14.1
Snells Beach	30072.0	5112.0	-36.430	174.722	0920	46.6

Table: 4. Table of data frame information for household incomes. Includes distance information for each neighbourhood from the Auckland CBD. Distance is given in kilometres (km's) and represents the distance via direct line of sight between two given points.

Neighbourhood	Direct Distance (Km's) [1]	Distance via road (Km's) [2]
McLaren Park	11.7	18.9
Birkdale	8.9	12.6
Paremoremo	14.1	25.3
Meadowbank	9.2	10.1
Otara	17.2	18
Snells Beach	46.6	66.6

Table: 5. Table of Neighbourhoods in Cluster 2 and their respective distances in kilometres (km's) from Auckland CBD.

[1] Distance calculated as a direct line between latitude and longitude points.

[2] Distance calculated by most direct driving route, as determined by Googlemaps.com

Looking at the distances in the above table, Birkdale and Meadowbank are the most central neighbourhoods and Snell's Beach the most removed from the city centre.

Conclusion

The purpose of this project was to find neighbourhoods within Auckland City where Indian cuisine is rated amongst the top venues sort out by diners. Using Foursquare data, the search was reduced to a cluster of six areas, Birkdale, Meadowbank, Otara, Snell's Beach and McLaren Park. Foursquare also revealed a relatively light density of Indian cuisine restaurants in all these neighbourhoods, with Meadowbank and Birkdale having the highest number. Both neighbourhoods currently have two existing restaurants that might be considered in direct competition with the client's business model.

Household incomes placed Birkdale and Meadowbank at the higher end of the six neighbourhoods and Otara at the lower end. In terms of distance from the city centre Meadowbank and Birkdale are the closest neighbourhoods and Snell's Beach requiring the greatest distance to travel.

Based on the findings of this investigation I would recommend the client consider either the neighbourhoods of Birkdale or Meadowbank as possible locations for her new restaurant. While both areas already have two Indian Restaurants which would indicate there is some competition, both locations are the most centrally located and are on the upper end of household incomes. Further information beyond the scope of this report would assist the client decide on the best location to open her first restaurant.