

Capstone Project - The Battle of Neighbourhoods Report

Contents

| | |
|--|---|
| 1.Introduction: Business Problem | 2 |
| 2.Data Acquisition and Cleaning | 2 |
| 2.1 Data Cleaning | 2 |
| 3.Methodology..... | 3 |
| 4.Results | 3 |
| 5.Discussion | 5 |
| 6.Recommendation | 5 |
| 7.Conclusion..... | 5 |

1.Introduction: Business Problem

This project aims to find a suitable location for an Indian restaurant in Toronto area. This project will be targeting entrepreneurs who are interested in establishing an Indian restaurant. Indian food is popular among Asians; therefore, entrepreneurs may want to service this group of people by opening a restaurant in Asian communities. The focus will be areas in Toronto that are mainly inhabited by Asians. This report will be using the data science methods and tools along with machine learning, this project aims to answer the following question: Which areas are most suitable for a new Indian Restaurant?

2.Data Acquisition and Cleaning

Based on our identified problem, factors that will influence our decision are: -The most common areas in Asian communities within Toronto.

The following data will be extracted and generated:

1. List of neighbourhoods in Toronto, Canada.
2. Latitude and Longitude of these neighbourhoods.
3. Venue data related to Indian restaurants. This will help find the areas that suitable to open an Indian Restaurant.

Data extraction methods:

1. Scrapping of Toronto neighbourhoods via Wikipedia
2. Getting Latitude and Longitude data of these neighbourhoods via Geocoder package
3. Using Foursquare API to get venue data related to these neighbourhoods.

2.1 Data Cleaning

The Toronto dataset was merged with the latitude and longitude data which resulted in three clusters of Indian Restaurants.

3. Methodology

Firstly, the list of neighbourhoods in Toronto, Canada is required. This is possible by extracting the list of neighbourhoods from Wikipedia:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. I did the web scraping by utilizing pandas HTML table scraping method as it is easier and more convenient to pull tabular data directly from a web page into the data frame.

However, it is only a list of neighbourhood names and postal codes. I need to get corresponding coordinates to utilize Foursquare to pull the list of venues near these neighbourhoods. To get the coordinates, I tried using Geocoder Package but it was not working so I used the CSV file provided by IBM team to match the coordinates of Toronto neighbourhoods. After gathering these coordinates, I visualized the map of Toronto using Folium package to verify whether these are correct coordinates.

Thereafter, I used Foursquare API to pull the list of top 100 venues within 500 meters radius. I have created a Foursquare developer account in order to obtain account ID and API key to pull the data. Foursquare gives me the ability to extract the names, categories, latitude, and longitude of the venues. With this data, I was able to check how many unique categories that I can get from these venues.

Then, I analysed each neighbourhood by grouping the rows by neighbourhood and taking the mean on the frequency of occurrence of each venue category. This is to prepare clustering to be done later.

Here, I made a justification to specifically look for "Indian restaurants".

Lastly, I performed the clustering method by using k-means clustering. K-means clustering 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 it is highly suited for this project as well. I have clustered the neighbourhoods in Toronto into 3 clusters based on their frequency of occurrence for "Indian food". Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

4. Results

After running the K-means clustering we can access each cluster created to see which neighbourhoods were assigned to each of the three clusters. Looking into the neighbourhoods in cluster 0, there are no Indian Restaurants in this cluster (see fig 1).

| Neighborhood | Indian Restaurant | Cluster Labels | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|--------------|-------------------|----------------|-----------------------|------------------------|-------|----------------|-----------------|----------------|
|--------------|-------------------|----------------|-----------------------|------------------------|-------|----------------|-----------------|----------------|

Figure 1: Cluster 0

Most of the Indian restaurants are in Cluster 1 which is around Riverdale, Church and Wellesley, Cabbagetown, Union Station, Davisville. See fig (2) below:

| | Neighborhood | Indian Restaurant | Cluster Labels | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|----|---|-------------------|----------------|-----------------------|------------------------|------------------------|----------------|-----------------|-------------------|
| 36 | The Danforth West / Riverdale | 0.023256 | 1 | 43.679557 | -79.352188 | Sher-E-Punjab | 43.677308 | -79.353066 | Indian Restaurant |
| 14 | Harbourfront East / Union Station / Toronto Is... | 0.010000 | 1 | 43.640816 | -79.381752 | Indian Roti House | 43.639060 | -79.385422 | Indian Restaurant |
| 8 | Davisville | 0.027778 | 1 | 43.704324 | -79.388790 | Marigold Indian Bistro | 43.702881 | -79.388008 | Indian Restaurant |
| 30 | St. James Town / Cabbagetown | 0.021739 | 1 | 43.667967 | -79.367675 | Butter Chicken Factory | 43.667072 | -79.369184 | Indian Restaurant |
| 4 | Central Bay Street | 0.015873 | 1 | 43.657952 | -79.387383 | Colaba Junction | 43.660940 | -79.385635 | Indian Restaurant |
| 6 | Church and Wellesley | 0.012821 | 1 | 43.665860 | -79.383160 | Kothur Indian Cuisine | 43.667872 | -79.385659 | Indian Restaurant |

Figure 2: Cluster 1

The lowest in Cluster 2 area which is in North Toronto West. Also, there are good opportunities to open near Midtown, Yorkville.

Looking at nearby venues it seems cluster 2 might be a good location as there are not a lot of Indian restaurants in these areas (see Fig 3 below). Therefore, this project recommends the entrepreneur to open an authentic Indian restaurant in these locations.

| | Neighborhood | Indian Restaurant | Cluster Labels | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|----|---------------------------------------|-------------------|----------------|-----------------------|------------------------|-----------------------|----------------|-----------------|-------------------|
| 34 | The Annex / North Midtown / Yorkville | 0.045455 | 2 | 43.67271 | -79.405678 | Roti Cuisine of India | 43.674618 | -79.408249 | Indian Restaurant |

Figure 3: Cluster 2

Visualising the clustered neighbourhoods on a map using the folium library (see fig 4).



Figure 4: Clustered neighbourhoods in Toronto

Each cluster is colour coded for the ease of presentation; we can see that majority of the Indian restaurant neighbourhoods falls in the red cluster which is the first. The green cluster has one Indian restaurant in the neighbourhood, the remainder of the Indian restaurants are in the blue cluster (cluster 0).

- Cluster 0: Neighbourhoods with no Indian restaurants.
- Cluster 1: Neighbourhoods with a greater number of Indian restaurants
- Cluster 2: Neighbourhoods with the smaller number of Indian

restaurants.

5. Discussion

The aim of this project is to assist entrepreneurs who want to open an Indian restaurant in Toronto. They can open the restaurant in neighbourhoods with a smaller number of existing Indian restaurants. Most of the Indian restaurants are in Cluster 1 which is around Riverdale, Church and Wellesley, Cabbagetown, Union Station, Davisville. The lowest in Cluster 2 area which is in North Toronto West. Also, there are good opportunities to open near Midtown, Yorkville. Looking at nearby venues it seems cluster 2 might be a good location as there are not a lot of Indian restaurants in these areas. Therefore, this project recommends the entrepreneur to open an authentic Indian restaurant in these locations.

6. Recommendation

The future research of this project can include other factors such as potential income made in the suitable Indian restaurant areas in Midtown, Yorkville.

7. Conclusion

This project helps entrepreneurs get a better understanding of the Toronto neighbourhoods with respect to the most common area with Indian restaurants. It's imperative for the entrepreneurs to utilise this data when deciding where to place their restaurants in order to avoid opening a restaurant in areas that already saturated. Using the data gathered, Midtown Yorkville is the most suitable neighbourhood for an Indian restaurant.