

Finding a Suitable Location for Opening a New Restaurant in Neighborhoods of Toronto

Introduction

A restaurant's success and failure depends on a variety of factors. One of those many factors is its location. The location a restaurant is as important as its quality and taste, besides location of a restaurant is a one-off factor. So it's important that we find an optimal location to run a restaurant. As a data scientist, we are required to properly analyse the location of various neighborhoods in the city and find a proper location to open a restaurant.

This problem is mostly faced by all the entrepreneurs thriving to open up a business. In this project we are going to focus on the restaurant location.

Data sets to be used:

For the given problem I may use the following data sets,

- Canada neighborhood data
- Latitude and Longitude data for the corresponding neighbourhood
- Venue Data

Description on dataset

First, the Canada neighborhood data is available in this Wikipedia page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. This dataset contains the Borough and corresponding neighborhood data. We will be using a Web scraping tool called BeautifulSoup to scrape the website and develop a data frame. The latitude and longitude dataset is already provided by the IBM team. This dataset contains all the latitude and longitude of all the neighborhoods. So we can match the neighborhood in both the dataframe and create a new dataframe (Figure 1 shows the sample of dataset). For the venue data, we will be using the foursquare API. I have already created a developer account in the website which gives us access to its venue data. This API will be used to extract all the restaurants in a particular neighborhood.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304
11	M1R	Scarborough	Wexford, Maryvale	43.750071	-79.295849
12	M1S	Scarborough	Agincourt	43.794200	-79.262029
13	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	43.781638	-79.304302
14	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-79.284577
15	M1W	Scarborough	Steeles West, L'Amoreaux West	43.799525	-79.318389
16	M1X	Scarborough	Upper Rouge	43.836125	-79.205636
17	M2H	North York	Hillcrest Village	43.803762	-79.363452
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556
19	M2K	North York	Bayview Village	43.786947	-79.385975

Figure 1 Canada Neighborhood dataset

Methodology

The final location for restaurant is suggested based on the fact that an optimal location of a restaurant is where there is little to medium competition. Based on this assumption, we recommend the location where there is fewer 'Indian Cuisine Restaurants'.

Step1: Data cleaning and Visualization

In this step, I have imported all the required libraries and imported the dataset which was prepared in another notebook. So the code to prepare the dataset is not mentioned here. Further, in this step we will be visualizing the data using folium library.

Step2: Fetching Indian restaurants in every neighbourhood

In this step we will be using Foursquare API to fetch all the Indian cuisine restaurants from each neighborhood. Its returns the value as a JSON file. So I have extracted only important elements from the JSON and is stored in "res_list". From this list a

dataframe is created which contains elements like borough, Neighborhood name, latitude, longitude and restaurants name, latitude, longitude. we will be using the data frame for further clustering and analysis. This data frame is called as “new_df” (shown in figure 2).

After creating the database, we will be visualizing the restaurants in folium map by plotting their latitude and longitude coordinates. This is done to get a better idea on how close the restaurants are located to each other.

	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Name of Restaurant	Restaurant Latitude	Restaurant Longitude
0	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	Wonder Season Chinese Restaurant	43.765354	-79.190536
1	Scarborough	Cedarbrae	43.773136	-79.239476	Terry's Restaurant	43.774702	-79.241175
2	Scarborough	Cedarbrae	43.773136	-79.239476	Terry's Restaurant & Bar	43.774780	-79.241043
3	Scarborough	Cedarbrae	43.773136	-79.239476	terry's restaurant	43.774969	-79.240872
4	Scarborough	Cedarbrae	43.773136	-79.239476	Thai One On	43.774468	-79.241268
...
617	Etobicoke	New Toronto, Mimico South, Humber Bay Shores	43.605647	-79.501321	Lucky Dice Restaurant	43.601392	-79.503056
618	Etobicoke	Alderwood, Long Branch	43.602414	-79.543484	Il Paesano Pizzeria & Restaurant	43.601280	-79.545028
619	Etobicoke	Alderwood, Long Branch	43.602414	-79.543484	ASA Korean Restaurant	43.603950	-79.537700
620	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999	McDonald's	43.630002	-79.518198
621	North York	Humber Summit	43.756303	-79.565963	E Caribbean Delight Restaurant	43.760230	-79.568913

622 rows × 7 columns

Figure 2 Final dataframe for clustering

Step3: Clustering the Data

Clustering the restaurants using K-Means clustering. Clustering is done to help us find which part of Toronto has most no. of restaurants. We can use this insight to find where there will be more demand for a particular restaurant. The clustered data will be plotted by 4 different colours on Toronto map. The below figure shows the no. of restaurants in each cluster.

Cluster Labels							
0	529	529	529	529	529	529	529
1	39	39	39	39	39	39	39
2	26	26	26	26	26	26	26
3	28	28	28	28	28	28	28

Step4: Examine the Cluster

Examining the clusters. From each cluster we can find the most no. of restaurants in a particular neighborhood.

Step5: Grouping

Grouping by neighborhood and borough to find the no. of restaurants in each.

Result

From the analysis we conclude that the neighborhoods listed in cluster0 is not an ideal place to open a new Indian restaurant as the location is already packed with more no. of Indian restaurants. So neighborhood from other three clusters are recommended. To be more accurate and specific the following neighborhood is recommended from each cluster.

1.1 Recommended location from cluster 2 (Blue):

The below neighborhoods are recommended for restaurant,

- Steeles West, L'Amoreaux West
- Cedarbrae

1.2 Recommended location from cluster 3 (Red):

The below neighborhoods are recommended for restaurant,

- Runnymede, Swansea
- Parkdale, Roncesvalles
- High Park, The Junction South

1.3 Recommended location from cluster 4 (Orange):

The below neighborhoods are recommended for restaurant,

- Davisville
- Willowdale, Willowdale East

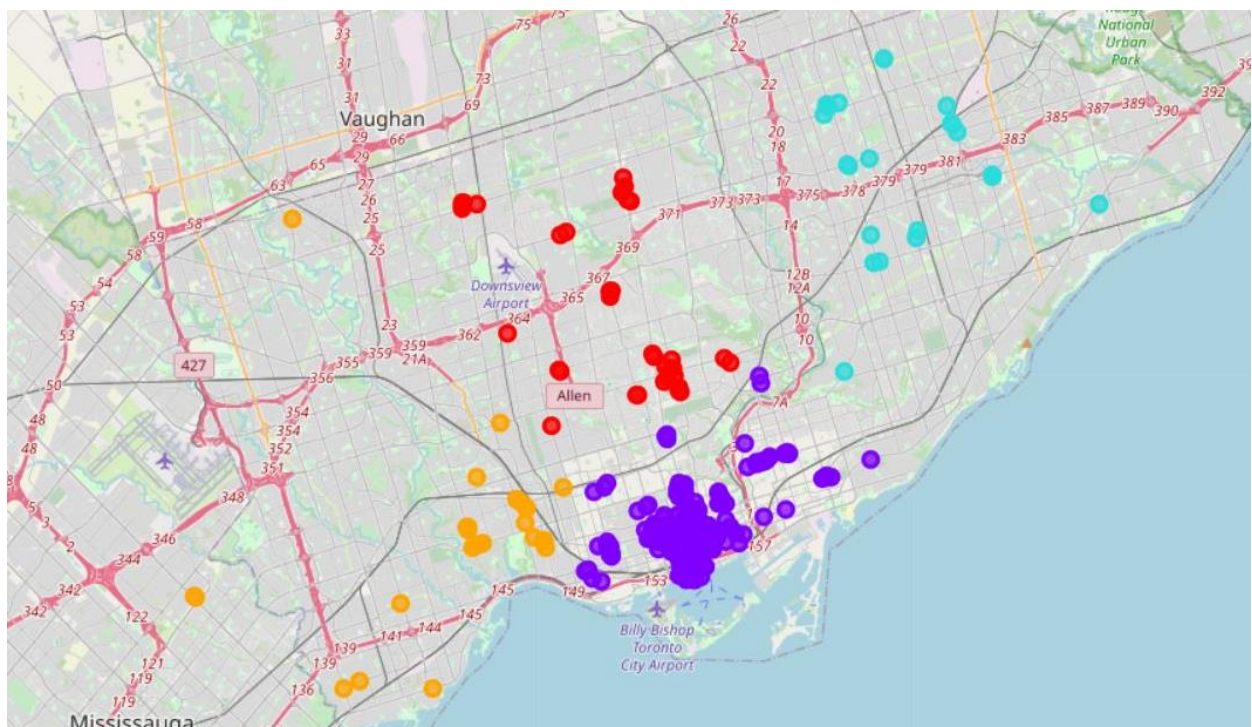


Figure 3 Clustered Visual of restaurants

Conclusion and Future Scope

Purpose of this project was to find a good location to open an Indian restaurant. From research we found that an ideal place to open a restaurant is where there is average no. of competitors. Based on this assumption the entire project has been completed and recommended neighborhoods have been produced.

For further analysis, we can also include data like, Indian population, crime rate, labour wage etc.