

Finding the ideal Neighborhood in Toronto for starting an Indian Restaurant

Introduction/Business Problem

The objective of this study is to find an ideal Neighbourhood in Toronto to start an Indian Restaurant, based on the location data extracted using Foursquare API, and then their visualization through Folium. Toronto, being the most populous city and a major financial hub of Canada, attracts residents from all over the world both via employment opportunities and tourism. With nearly half, 46%, of its population coming from other countries. More than 200 different ethnic groups call Toronto home, and they bring with them over 140 languages. It is also one of those cities housing a wide selection of restaurants from cuisines all over the world, as we are going to see in this study.

Here, we deal with a hypothetical situation, of solving queries of someone planning to start an Indian restaurant in Toronto. They have approached us, to conduct a preliminary study of feasible neighborhoods for this new venture. Here, we use the business concept of proximity to competition, to shortlist ideal neighborhoods. Competition can be good, in industries where comparison shopping is popular. (That's why competing retail businesses, such as fast-food restaurants, antique shops and clothing stores tend to cluster together.) You may also catch the overflow from existing businesses, or customers who wish to try a different version of the same product, in our case, a new cuisine of food. Hence, in this project, we try to find the neighborhoods around Toronto, which has a clustering of Restaurants, but none of them being an Indian Restaurant. This will avoid competing with the same cuisine, one that's hard to pull off with a new restaurant.

Target Audience

Someone who wishes to get into the restaurant business by starting an Indian restaurant in Toronto.

Data

To solve this problem, we will use the following data:

- List of neighborhoods in Toronto, Canada.
- Latitude and Longitude of these neighborhoods.
- Venue data of restaurants. This will help us find the neighborhoods that have a cluster of restaurants located in them. This data would also help us identify the cuisine of the restaurant, including Indian.

Data Sourcing

1. The Toronto neighborhoods data is obtained from the Wikipedia page:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

example of data:

Postal Code	Borough	Neighborhood
M4A	North York	Victoria Village

2. The latitude and longitude data for the above neighborhoods:

http://cocl.us/Geospatial_data

Example of data

Postal Code	Latitude	Longitude
M1B	43.806686	-79.194353

3. Data about the restaurants near these neighborhoods will be collected via Foursquare API and the explore endpoint.

Methodology

The course of this project would involve accessing the data of different venues existing within a set radius from each neighborhood in the list. The collection of data about the different kinds of venues would be done using **Foursquare api**, and then only the categories containing the word “Restaurant” will be extracted from them. After this, we remove the data of Indian Restaurants from the dataframe, and then aggregate the other restaurant values together, to represent the relative presence of restaurants near each neighborhood. Then we add the Indian restaurant column back into the dataframe beside the aggregate value of restaurants of all other cuisines for comparison. Finally, we sort the dataframe in the decreasing order of aggregate restaurant value, and then pick the neighborhoods at the top, with a 0 value for Indian restaurants column

We now examine in detail, the methodology followed for the project and its different stages

The data from the Wikipedia page

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M was read into a pandas dataframe using read_html. Thereafter, the rows having “Not assigned” for Boroughs were removed from the dataset. Multiple neighborhoods having the same postal code were grouped together with a comma separator. For rows which had missing neighborhood field but a non-empty Borough field, the Borough was used as the neighborhood.

The latitude and longitude data for each Neighborhood was read from

http://cocl.us/Geospatial_data into a separate dataset, and later merged with the first dataset using the merge function.

We chose to work with only Neighborhoods that were in Toronto, so we selected the Boroughs with the name Toronto in it from the dataframe and marked all of them into a map of Toronto to visualize them.

As a next step, we used the Foursquare API and the explore endpoint to get the details of venues which were in a radius of 1 km from each Neighborhood. 3189 such venues were acquired via the Foursquare API. We then performed a one hot encoding for the venue categories, and then restricted the dataframe to show only the restaurants among the venues. This included a wide list of cuisines all around the world.

```
[16]: sample=toronto_grouped.filter(regex='Restaurant')
```

```
[17]: sample.head()
```

```
[17]:
```

	American Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cantonese Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	...	Sust Restauran
0	0.010000	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.01	0.0	...	0.00000
1	0.010000	0.0	0.0	0.0	0.0	0.0	0.01	0.00	0.01	0.0	...	0.00000
2	0.020408	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.0	...	0.04081
3	0.000000	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.0	...	0.00000
4	0.010000	0.0	0.0	0.0	0.0	0.0	0.00	0.01	0.00	0.0	...	0.03000

5 rows × 53 columns

```
[18]: sample.shape
```

```
[18]: (39, 53)
```

We have captured 53 different cuisines of restaurants, which are active around 39 Neighborhoods of Toronto. Then, we dropped the column containing the Indian restaurants, did a horizontal sum to get an aggregate value of all other cuisines and then added the Neighborhood column and Indian Restaurant column back to the dataframe to generate the following

```
[30]: sample2
```

```
[30]:
```

	Neighborhood	Aggregate Restaurants	Indian Restaurants
0	Berczy Park	0.180000	0.000000
1	Brockton, Parkdale Village, Exhibition Place	0.250000	0.020000
2	Business reply mail Processing Centre, South C...	0.204082	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.000000
4	Central Bay Street	0.240000	0.000000
5	Christie	0.310000	0.020000
6	Church and Wellesley	0.230000	0.010000
7	Commerce Court, Victoria Hotel	0.230000	0.000000
8	Davisville	0.340000	0.030000
9	Davisville North	0.250000	0.000000
10	Dufferin, Dovercourt Village	0.233333	0.000000

The next step was to sort the dataframe in the decreasing order of the aggregate Restaurant score, and then look for rows at the top having a 0 value in the Indian restaurant column. This gave us the following result

```
[31]: #Sorting the database in decreasing order of the aggregate values
sample2.sort_values(by=['Aggregate Restaurants'], inplace=True, ascending=False)
```

```
[32]: sample2.head()
```

```
[32]:
```

	Neighborhood	Aggregate Restaurants	Indian Restaurants
8	Davisville	0.340000	0.030000
36	The Danforth West, Riverdale	0.320000	0.010000
19	Little Portugal, Trinity	0.320000	0.000000
5	Christie	0.310000	0.020000
30	St. James Town, Cabbagetown	0.297297	0.027027

Of the top 5 rows, we see that Little Portugal is tied with the 2nd spot in the aggregate restaurant spot, and there are no Indian restaurants near the Neighborhood.

Results

From the above data frame, let's consider the top 5 rows, we see that that Davisville has the highest clustering of restaurants around, and it also has the highest number of Indian restaurants around. In fact, all neighborhoods in the top 5, except Little Portugal has an Indian restaurant operating nearby.

We can also see that Little Portugal and the Danforth West has the same value for aggregate restaurants, making it a tie for second place, but makes it a better option to start our Indian restaurant. Hence, we can focus more on the Little Portugal neighborhood

Since we have found that Little Portugal has no Indian restaurants nearby, we would like to verify whether the neighborhood is welcoming for restaurants of different cuisine. Unless the neighborhood and the residents are open to try out new cuisines, we won't benefit much from opening our restaurant in this neighborhood.

For knowing this, we retrieve every restaurant in Little Portugal region by their cuisine type. This is done by calling the column names of restaurant cuisines of Little Portugal which had a non-zero value. This operation resulted in the following output

```
[36]: i=0
      for column in sample3.columns:
          if sample3.iloc[0,i]!=0:
              print(column)
          i+=1
```

```
Neighborhood
American Restaurant
Asian Restaurant
Cuban Restaurant
Dumpling Restaurant
French Restaurant
Greek Restaurant
Italian Restaurant
Japanese Restaurant
Korean Restaurant
Malay Restaurant
New American Restaurant
Restaurant
Seafood Restaurant
Tapas Restaurant
Thai Restaurant
Vegetarian / Vegan Restaurant
Vietnamese Restaurant
```

We can see that the cuisines around Little Portugal are infact cosmopolitan and we have a good chance of securing business if we select **Little Portugal** to open our Indian Restaurant.

Discussions and Recommendations

This analysis is purely based on the proximity of other restaurants and their cuisine types. Although this analysis provides a preliminary idea of selection of neighborhood, further study has to be carried out before finalizing. Factors like rent, availability of space, traffic and busy hours can influence the business of a restaurant to a great extent, and these have to be taken into consideration before finalizing. Further analysis can include the population density, spending capacity of residents of each neighborhood etc to take more leverage of the available data. Additionally, the analysis was limited to a radius of 1 km, this can be tweaked for varying the results.

Conclusion

In this project, we utilised the data obtained through Foursquare API to decide a Neighborhood in Toronto for starting an Indian Restaurant. Based on the Analysis, we have selected Little Portugal as a viable option for further analysis and a potential location for the starting of the new Indian Restaurant.

References

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

http://cocl.us/Geospatial_data

<https://pandas.pydata.org/docs/>

<https://developer.foursquare.com/docs/places-api/endpoints/>