

FINDING TORONTO NEIGHBOURHOOD FOR OPENING OF AN INDIAN RESTAURANT

CAPSTONE PROJECT – IBM DATA SCIENCE CERTIFICATION

By JYOTI DASH

The objective of this study is to find an ideal Neighborhood in Toronto to start an Indian Restaurant, based on the location data extracted using Foursquare API, and then their visualization through Folium.

**Introduction/
Business
Problem**
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.

Data Requirement / Sourcing

- List of neighbourhoods in Toronto, Canada.
- Latitude and Longitude of the neighbourhoods. This data will be collected in collaboration with FOURSQUARE API.
- Venue data of restaurants. This will help us find the neighbourhoods 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 neighbourhoods details were collected from Wikipedia page:

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

2. Venue & Lat-Long data:

FOURSQUARE API

Data Cleaning

The following steps were done to extract the useful data

1. 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.
2. 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.
3. 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.
4. 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.
5. We have captured 53 different cuisines of restaurants, which are active around 39 Neighborhoods of Toronto.

Data Analysis

A one hot encoding for the venue categories, and a condition to show only the restaurants among the venues gave the following output including a wide list of cuisines all around the world.

	American Restaurant	Asian Restaurant	Belgian Restaurant	Brazilian Restaurant	Cajun / Creole Restaurant	Cantonese Restaurant	Caribbean Restaurant	Chinese Restaurant	Comfort Food Restaurant	Cuban Restaurant	...	Sushi Restaurant
0	0.010000	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.01	0.0	...	0.000000
1	0.010000	0.0	0.0	0.0	0.0	0.0	0.01	0.00	0.01	0.0	...	0.000000
2	0.020408	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.0	...	0.040816
3	0.000000	0.0	0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.0	...	0.000000
4	0.010000	0.0	0.0	0.0	0.0	0.0	0.00	0.01	0.00	0.0	...	0.030000

5 rows × 53 columns

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

	Neighborhood	Aggregate Restaurants	Indian Restaurants
0	Berczy Park	0.180000	NaN
1	Brockton, Parkdale Village, Exhibition Place	0.250000	NaN
2	Business reply mail Processing Centre, South C...	0.204082	NaN
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	NaN
4	Central Bay Street	0.240000	NaN
5	Christie	0.310000	NaN
6	Church and Wellesley	0.230000	NaN
7	Commerce Court, Victoria Hotel	0.220000	NaN
8	Davisville	0.350000	NaN
9	Davisville North	0.250000	NaN
10	Dufferin, Dovercourt Village	0.230769	NaN
11	First Canadian Place, Underground city	0.220000	NaN
12	Forest Hill North & West, Forest Hill Road Park	0.215686	NaN
13	Garden District, Ryerson	0.270000	NaN
14	Harbourfront East, Union Station, Toronto Islands	0.150000	NaN
15	High Park, The Junction South	0.210000	NaN

Data Analysis

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

	Neighborhood	Aggregate Restaurants	Indian Restaurants
8	Davisville	0.350000	NaN
19	Little Portugal, Trinity	0.320000	NaN
36	The Danforth West, Riverdale	0.310000	NaN
5	Christie	0.310000	NaN
33	Summerhill West, Rathnelly, South Hill, Forest...	0.307692	NaN
30	St. James Town, Cabbagetown	0.297297	NaN
20	Moore Park, Summerhill East	0.285714	NaN
13	Garden District, Ryerson	0.270000	NaN
32	Studio District	0.270000	NaN
34	The Annex, North Midtown, Yorkville	0.270000	NaN

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.
- 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

```
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 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

- we utilized the data obtained through the 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 study 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://developer.foursquare.com/docs/places-api/endpoints/>

Thank You