

Coursera Capstone – The battle of Neighborhood

Recommendation to open a new Indian cuisine restaurant in Toronto **Neighbourhood**



1.0 Introduction:

Toronto is the provincial capital of Ontario and the most populous city in Canada. It is also known as an international center of business, finance, arts, tourism, food and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. This makes Toronto a place of ample business opportunities including food service industry.

1.1 Problem Description:

The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents belong to a visible minority population group. This diversity is reflected in Toronto's ethnic neighborhoods, which include Chinatown, Corso Italia, Little India, Little Italy, Greektown, Koreatown, Little Jamaica etc. Hence there is a huge opportunity for food service industry to invest. This report will identify and recommend the kind of eatery restaurateurs can establish for a profitable engagement.

2.0 Data Description:

As we need to explore, segment and cluster the neighborhoods in the city of Toronto, the Toronto neighborhoods data is the key for this project. Neighborhood data is not widely available on the internet in the structured format, hence we need to scrap it through an existing Wikipedia page that has all the required information. We will also like to obtain the key information like below:

1: The latitude and longitude of the neighborhoods are retrieved using geocoding. The geometric location values are then stored into the initial data frame. Geo data is helpful to visualize the places for Toronto Neighborhoods.

Example of initial data with information.

```
toronto_df = dataframe.merge(geo_data, left_on="Postcode", right_on="Postal Code")
toronto_df.head()
```

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude
0	M3H	North York	Bathurst Manor, Downsview North, Wilson Heights	M3H	43.754328	-79.442259
1	M4A	North York	Victoria Village	M4A	43.725882	-79.315572
2	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	M6K	43.636847	-79.428191
3	M2J	North York	Fairview, Henry Farm, Oriole	M2J	43.778517	-79.346556
4	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	M1V	43.815252	-79.284577

2: Population data is helpful to explore the places for Toronto neighbourhoods and the key feature to find out the places where we have south Asian population.

```
toronto_pop_df.head()
```

	Postcode	Borough_x	Neighbourhood	Latitude	Longitude	Population, 2016	Population density per square kilometre	South Asian	Chinese	Black	Arab	Southeast Asian	West Asian
0	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	29834	14280	5.88	2.65	10.05	0.63	1.78	0.68
1	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576	13826	9601	2.86	8.21	2.39	0.40	0.61	0.54
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636	76716	25823	17.93	13.09	9.55	2.25	2.32	1.80
3	M6S	West Toronto	Runnymede, Swansea	43.651571	-79.484450	10070	6333	2.43	3.62	2.14	0.20	0.50	0.25
4	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	15818	5273	8.28	35.59	4.71	2.56	0.79	3.03

3: The Venue data is another key factor for this project, which is found out by passing the required parameters to the foursquare API and creating another data frame to contain all the venue details along with the respective neighbourhood. We will also need to understand the type of these venues nearby (500 meters) in each of the neighbourhoods.

Example of venue data along with their type and categories in structured format to allow further computation:

```
Places_food .head()
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue ID
0	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	The Abbott	43.637996	-79.430717	Café	4cdeecddf8a4a14330b8dabc
1	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	Caffino	43.639021	-79.425289	Italian Restaurant	4b632632f964a52085662ae3
2	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	Vogue Supper Club	43.636951	-79.425446	Restaurant	4ce8ba46f865370410eab2c4
3	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	Loco Burrito	43.637513	-79.422898	Burrito Place	4ce6bec9948f224ba0cbe55d
4	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	Uma Cafe	43.637016	-79.424048	Café	5137b53ce4b071af38d34174

3.0 Methodology:

Data scrapping from the Wikipedia page that contains the Toronto neighbourhoods has been used and the population data mainly for south Asian people in Toronto neighbourhoods is the key factor for this project.

Furthermore, we need to know the coordinates and locations of this neighbourhoods, and therefore the geocoder API has been used for achieving this objective. This is important so that we can input this information into the location information provider such as Foursquare.com to obtain venue information in these neighbourhoods, and this is precisely what we have done for it in this project.

After fetching the data for all venues of neighbourhoods, I filtered out the specific venues I was interested in (Restaurants and Nightlife venues).

```
targets = ['Restaurant', 'Joint', 'Place', 'Theater', 'Music', 'Club', 'club', 'Cocktail', 'cocktail', 'Roof', \
          'Café', 'Lounge', 'Pub', 'pub', 'House', 'Cafe', 'Wine', 'Food', \
          'Nightlife', 'Party']
```

```
matching = [s for s in uniqueList if any(xs in s for xs in targets)]
len(matching)
```

75

```
mask = toronto_venues['Venue Category'].isin(matching)
Places_food = toronto_venues[mask]
Places_food.reset_index(inplace = True)
Places_food.drop('index', axis = 1, inplace = True)
Places_food.shape
```

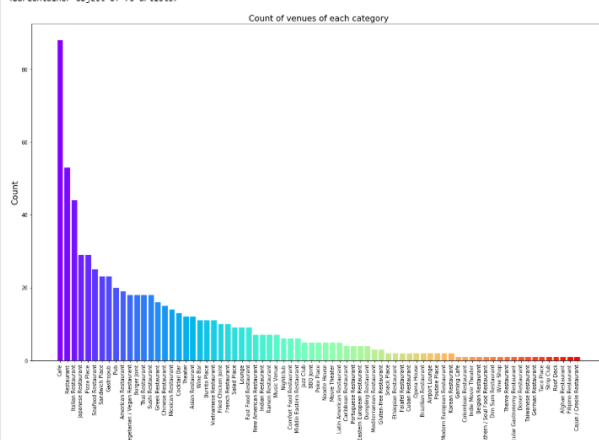
(723, 8)

Places_food.head()

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue ID
0	Adelaide, King, Richmond	43.650571	-79.384568	Rosalinda	43.650252	-79.385156	Vegetarian / Vegan Restaurant	5aff06ca6e4650002cc6286b
1	Adelaide, King, Richmond	43.650571	-79.384568	Cafe Landwer	43.648753	-79.385367	Café	5b6c842bc36588002c80a934
2	Adelaide, King, Richmond	43.650571	-79.384568	Canadian Opera Company	43.650660	-79.386242	Opera House	4ad4c062f964a520bf720e3
3	Adelaide, King, Richmond	43.650571	-79.384568	Noodle King	43.651706	-79.383046	Asian Restaurant	4b8d5856f964a520f4f532e3
4	Adelaide, King, Richmond	43.650571	-79.384568	Estiatorio Volos	43.650329	-79.384533	Greek Restaurant	4e048b98315168be7fd7ee4d

```
venue_distribution_res = filtered['Venue Category'].value_counts()
colors = cm.rainbow(np.linspace(0, 1, len(venue_distribution_res.index)))
plt.figure(figsize = (20, 12))
plt.xticks(rotation = 90)
plt.xlabel('Venue Category', fontsize = 16)
plt.ylabel('Count', fontsize = 16)
plt.title('Count of venues of each category', fontsize = 16)
plt.bar(venue_distribution_res.index, venue_distribution_res.values, color = colors)
```

<BarContainer object of 75 artists>



In above figure, we can see the filtered venues for restaurants and different food places in Toronto Neighbourhoods.

This project is specially to find out the place where we can open an Indian restaurant based on the south Asian population in Toronto neighbourhood, so I filter out the venues accordingly.

```
cluster_toronto_food=toronto_final_food_df[["Indian Restuarants","South Asian"]].astype(float)
```

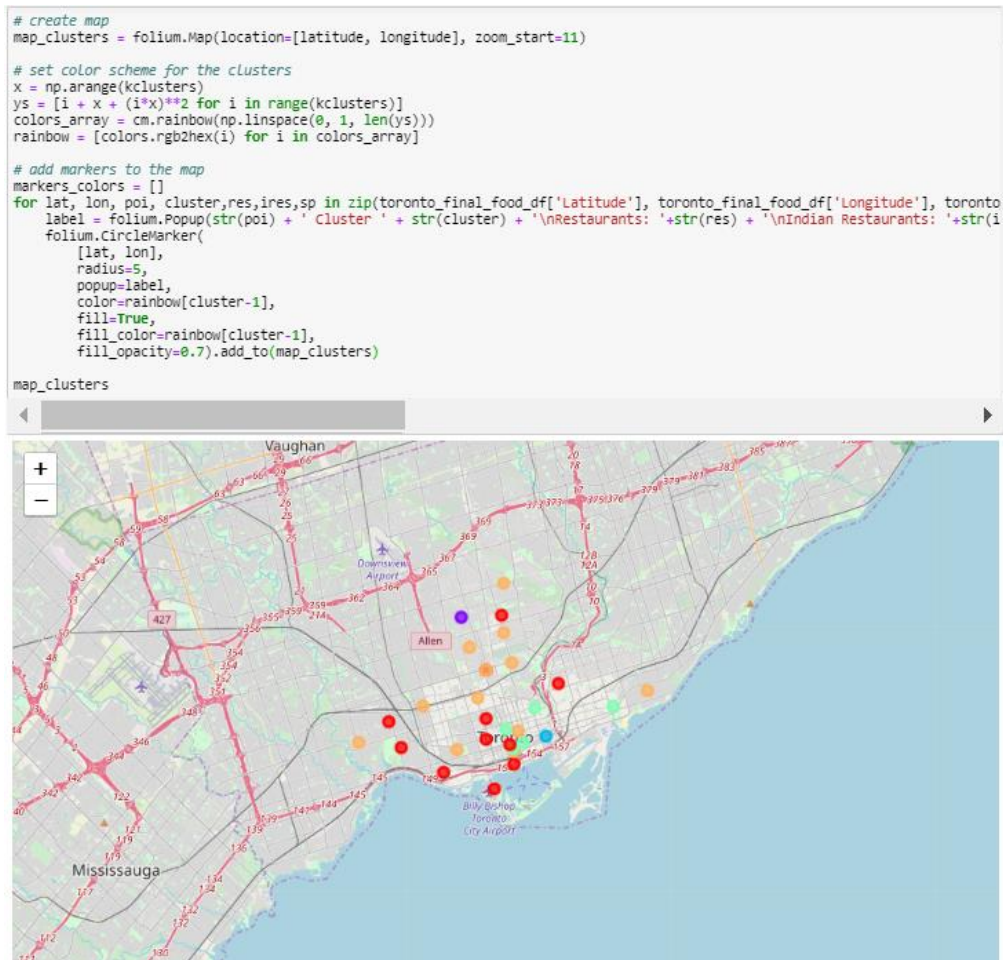
```
cluster_toronto_food
```

	Indian Restuarants	South Asian
0	1.0	5.76
1	0.0	2.89
2	0.0	4.24
3	0.0	3.71
4	0.0	2.86
5	1.0	5.58
6	0.0	5.22
7	0.0	3.77
8	0.0	4.46
9	1.0	5.88
10	0.0	5.88
11	1.0	12.38
12	0.0	2.95
13	0.0	2.25
14	0.0	5.19
15	0.0	41.79
16	0.0	4.78
17	0.0	6.41
18	0.0	17.93
19	0.0	2.43
20	1.0	2.99

After pulling all the data together in a single data frame, I ran a clustering analysis to group similar neighbourhoods together.

4.0 Results:

The below cluster shown in light blue are the most ideal spots to open an Indian restaurant.



This is based on the high south Asian population and the almost complete absence of other Indian and south Asian Restaurants

The ideal neighbourhoods for our new location are:

- Harbourfront
- Roselawn

5.0 Discussion:

We identified that from the total set of venues, majority of them were Cafes. So if a person love particular cuisine (Indian) it is not necessary that he will get it easily.

After picking the best neighbourhood, there's a lot more work that needs to be done to actually find the best location. However, this project gives a nice start to the process and narrowed down a very long list (from 30 to 2 choices)

The correct set of data is very important for any such projects and mostly it is difficult to find the required data even in the internet.

6.0 Conclusion:

With that, we have concluded that the best recommendation for Indian food industry to first offer their services in Toronto will be Harbourfront and Roselawn with the key factors to consider the high population of south Asians,

It is also recommended to the food industry to re-run this data science program to get the updated result and use the result into consideration as part of the business growth plan in selecting the next neighbourhood to offer their services. This is critical not only to make sure that they got the updated result for better decision making, but also to make sure that they can re-validate the findings from this project. Finally, thank you for the opportunity in this project and we wish you the best success in your business.