Applied Data Science Capstone Project

Best place to open an Indian Restaurant in the city of Toronto, Canada

Deepthi Chiruvolu

21st May,2021

Business Problem: **To identify the best location to open an Indian Restaurant in the city of Toronto**

Introduction:

Indian immigrants have been a part of life in Canada for over a century. As of 2020, there are more than one million Indian citizens in Canada.

Toronto is a major magnet for Indian immigrants. The city is Canada's economic engine and is also home to some of the continent's top schools and universities. Opening an Indian restaurant in this city should be an exciting idea for business startup.

In this Project, We will identify the best location to setup an Indian restaurant by analyzing the neighborhoods of Toronto and finding out the borough that has the least number of Indian restaurants. As lesser number implies little competition for the startup.

Data Acquisition

Firstly the data related to different locations(postal codes) of the city, Toronto is acquired from wikipedia page

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

The geospatial coordinates (latitudes and longitudes) of different locations are obtained from this .csv file

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv

Four-Square location API has been used to explore and identify the top venues in the neighborhoods

Webscraping and Data cleansing

Data downloaded or scraped from multiple sources were combined into one table.
First, We will scrape the wiki page to get the basic information of Toronto. In that table data, we can see that many postal codes are 'Not assigned' to any neighborhood. As part of data pre-processing, we will go with few assumptions that helps in refining the data

1. we shall ignore these 'Not assigned' postal codes as these missing values may mislead the analysis.
2. We will combine the neighborhoods separated by a comma if there exists multiple neighborhoods for single postal code area. Example: M5A Regent Park,Harbourfront
3. If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

After data preprocessing, we will convert that data into Pandas dataframe for applying machine learning algorithms and for data visualizations.

Then we will append the geospatial cooordinates.csv file to get the corresponding location coordinates (Latitude & longitude) of the postal codes.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
15	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
19	M4E	East Toronto	The Beaches	43.676357	-79.293031
20	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
24	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
25	M6G	Downtown Toronto	Christie	43.669542	-79.422564

Now along with the above neighborhood data, we will leverage FourSquare API. FourSquare API is used to explore these neighborhoods to fetch the details about different types of venues and their location. This data will be used for our analysis and final result. The sample data as below

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

using Folium maps and K-means clustering we can identify clustered neighborhoods on maps .

Methodology

1. we will explore the neighborhoods of Toronto city on maps
2. we will explore venues in those neighborhoods
3. we will list out the venues of category 'restaurant'
4. We will list out the neighborhoods having Indian restaurant and their most common visited venues
5. Clustering the neighborhoods

Data Analysis:

We will scrape the webpage containing Toronto postal codes, convert it into dataframe, append location coordinates. The data frame has 15 boroughs and 103 neighborhoods.

We can visualize these neighborhoods on the Toronto map.
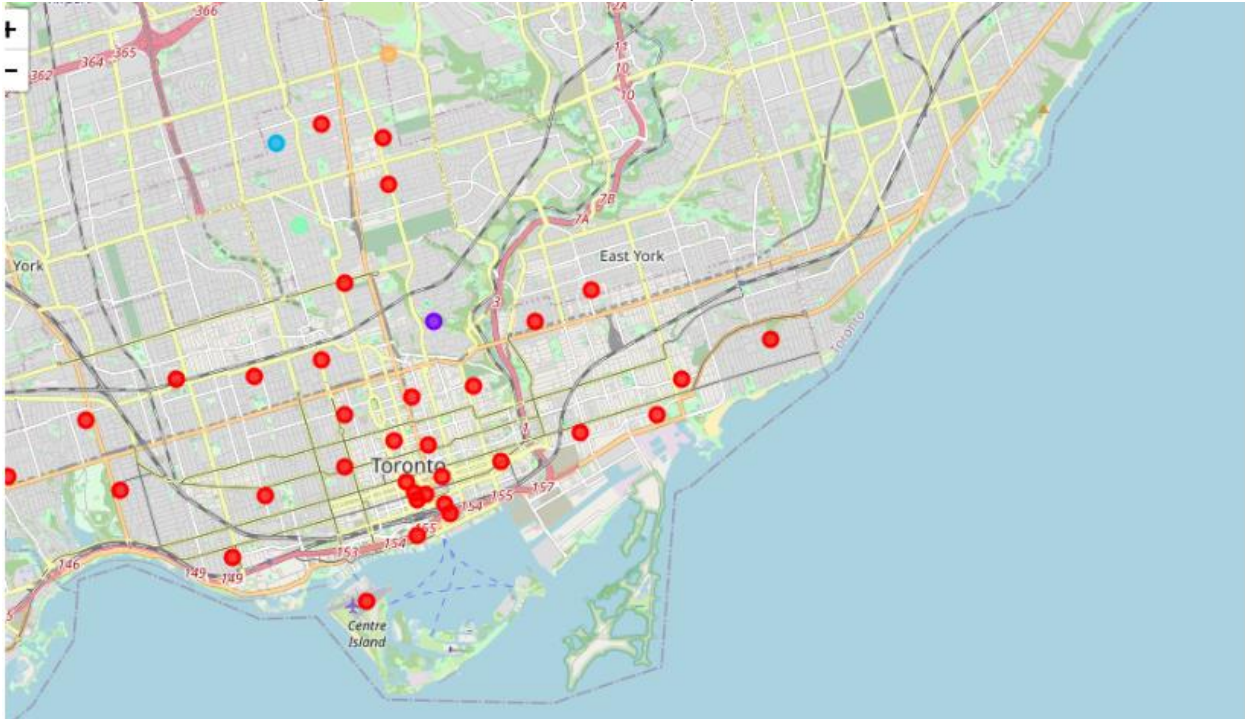
Using Foursquare API, we will explore the venues and venue categories in those neighborhoods.

We have got a total of 2136 venues in the Toronto city for the given latitude and longitude coordinates within a radius of 500meters. There are 275 unique venue categories.

Out of these, 487 venues are related to restaurants and among those 487 venues only 12 venues are of 'Indian Restaurant'.

We will find the top 10 most common venues in those neighborhoods.

	Neighborhood	Venue Category	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bedford Park, Lawrence Manor East	Indian Restaurant	Coffee Shop	Restaurant	Sandwich Place	Italian Restaurant	Comfort Food Restaurant	Indian Restaurant	Pizza Place	Pub	Café	Butcher
1	Berczy Park	Indian Restaurant	Coffee Shop	Cocktail Bar	Bakery	Seafood Restaurant	Restaurant	Beer Bar	Farmers Market	Pharmacy	Cheese Shop	Café
2	Central Bay Street	Indian Restaurant	Coffee Shop	Café	Italian Restaurant	Sandwich Place	Bubble Tea Shop	Burger Joint	Japanese Restaurant	Salad Place	Pizza Place	Discount Store
3	Church and Wellesley	Indian Restaurant	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Gay Bar	Restaurant	Yoga Studio	Mediterranean Restaurant	Men's Store	Fast Food Restaurant	Dance Studio
4	Davisville	Indian Restaurant	Pizza Place	Sandwich Place	Dessert Shop	Café	Gym	Coffee Shop	Sushi Restaurant	Italian Restaurant	Park	Thai Restaurant

Let's plot the count of Indian restaurants in the neighborhoods containing those 12 venues.

Let's also plot the total number of restaurants in those neighborhoods.
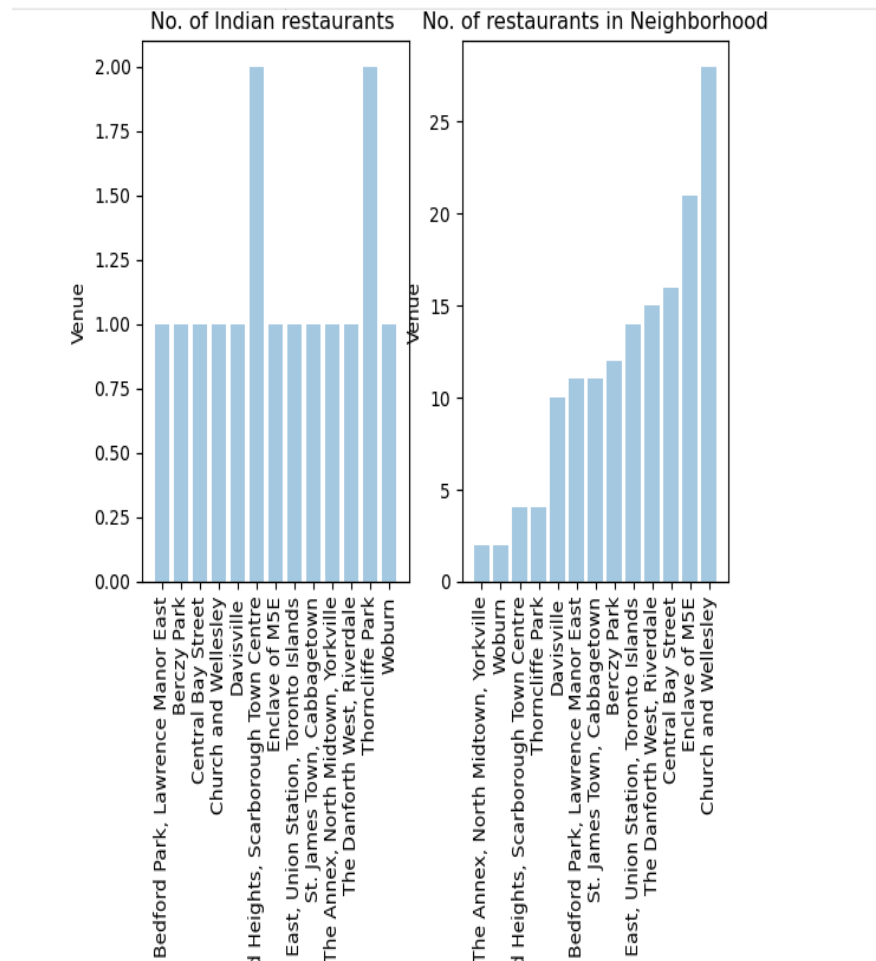
Results:

From the above bar graph we can observe that ' Dorset Park, Wexford Heights, Scarborough Town'and ' Thorncliffe Park' have two Indian restaurants . These two neighborhoods also has the 1st most common venue as 'Indian restaurant' but these two neighborhoods has a total of 4 restaurants out of which 2 are Indian , so these two cannot be the best place to start the restaurant.

The last 3 neighborhoods 'Central Bay Street', 'Enclave of M5E','Church and Wellesley' have a total of 16,21 and 28 restaurants respectively out of which 1 is an Indian restaurant. As these neighborhoods already have a lot of restaurants, these cannot be good place to start a new one.

'Davisville', 'Bedford Park, Lawrence Manor East', 'St. James Town, Cabbagetown'

has optimum number of restaurants with only 1 Indian restaurant. So these three neighborhoods could be the best places to start an Indian Restaurant.

Cluster Neighborhoods

K-Means algorithm is one of the most common clustering method of unsupervised learning.

I have analyzed the K-Means with elbow method and found the optimum k of the K-Means as 3. I have run k-means to cluster the neighborhood into 3 clusters.
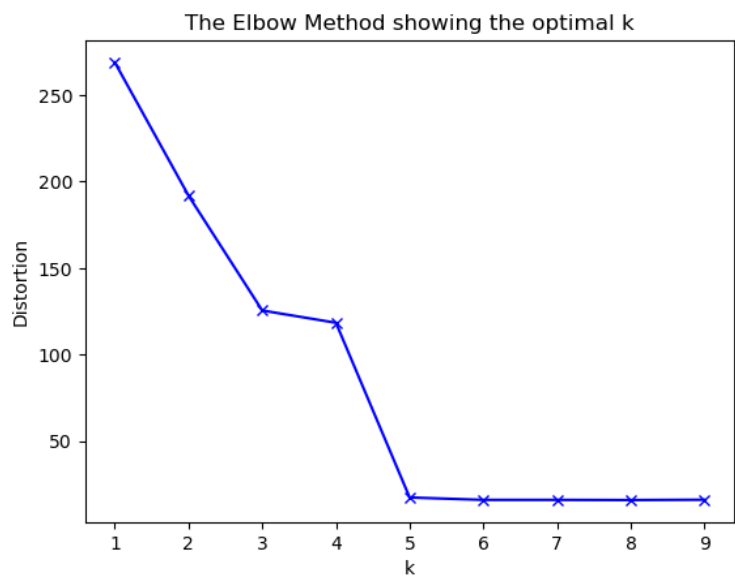
I have run k-means to cluster the neighborhood into 3 clusters.
Examined each cluster and determined the discriminating venue categories that distinguish each cluster.
Based on those defining categories, we have assigned a name to each cluster.

Created a bar chart of "Number of Venues in Each Cluster"

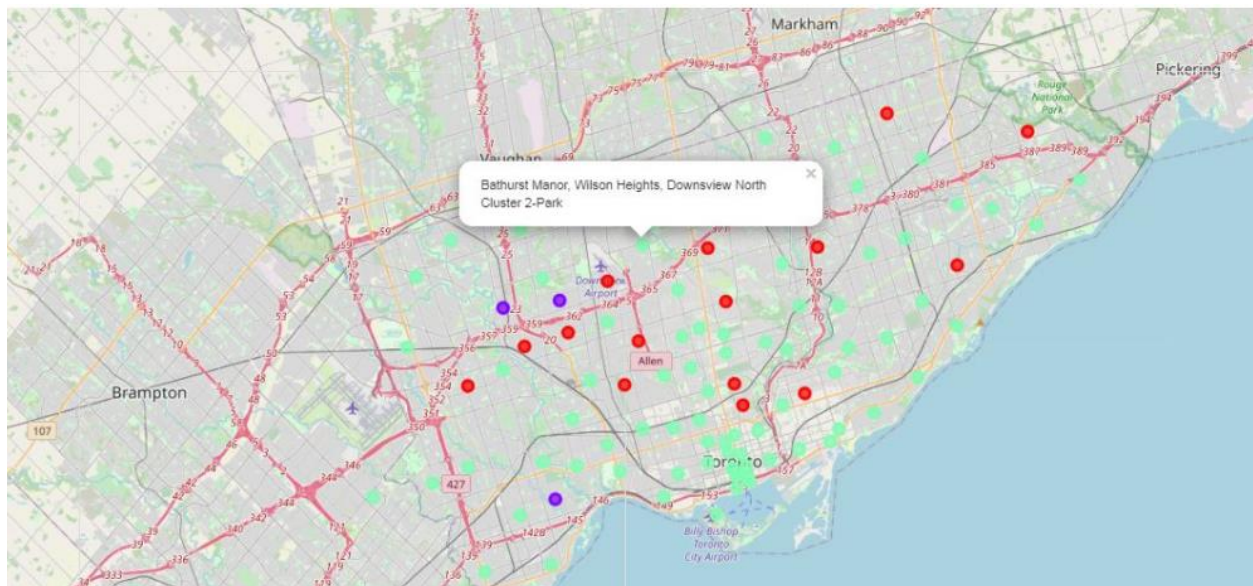When we examine above graph we can label each cluster as follows:

Cluster 0 : "Baseball Field"

Cluster 1 : "Coffee Shop"

Cluster 2 : "Park"

We can now assign those new labels to clusters.

Finally, let's visualize the resulting clusters

Discussion

To start a new restaurant, we have to consider many things like the number of restaurants in that borough, distance of that borough to city centre etc. Here we have taken a sample data for just a pair of location coordinates of the city and has just done a sample analysis with some assumptions.

While setting a restaurant based on particular cuisine, we have to consider the population data of those who are interested in that particular cuisine in that particular location. As I have not considered census data, I assumed that the neighborhoods having Indian restaurants will be the locations where people are interested in that particular cuisine, so I have done my final analysis on the neighborhoods having Indian Restaurants and concluded that 3 locations:

'Davisville', 'Bedford Park, Lawrence Manor East', 'St. James Town, Cabbagetown'

as the best places to start a new one.

I have clustered the information on the Toronto map. I used the K-means algorithm as part of this clustering study.

For more detailed and accurate guidance, the data set can be expanded and different approaches can be tried in clustering and classification studies.

One more thing to be noted is the venue details will change everyday.

Conclusion

I have found out the best neighborhoods to start an Indian restaurant in Toronto for a pair of location coordinates. I concluded the analysis by visualizing the data and clustering information on the Toronto map.

This kind of analysis will be helpful to investors, immigrants, tourists and many more.