Capstone - Project - The Battle of Neighborhoods

1.Introduction:

Toronto is the capitol city of Ontario. It is the most populated city in Canada. It is the center for international business, finances, arts and cultural and it is recognized as multicultural city in the world. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. Toronto has the largest Indian population in Canada. There are numerous opportunities for entrepreneurs to start a new business.

1.1 Problem:

The objective of this project is to find a best neighborhood in Toronto to start a new **Indian restaurant** with a help of data science. We analyze the neighborhoods in Toronto to identify the most profitable area since the success of the restaurant depends on the people and ambience. In this project, we are trying to find a best place to open a new Indian restaurant in Toronto, Canada. Since there are lots of restaurants in Toronto we will try to detect the locations that are not already crowded with restaurants and focusing mainly in areas where no Indian restaurant in the neighborhood. This is will help stakeholders or new entrepreneur who are interested to open new Indian restaurant in Toronto, Canada. Our ultimate goal find a place to start a new Indian restaurant where it yields more profit to the entrepreneur.

1.2 Target Audience:

1. Entrepreneur who wants to start a new business especially to open an Indian restaurant in Toronto and this will be an useful guide to start or expand restaurants targeting the Indian crowd.
2. Indian people who wants to finds a neighborhood with lots of option for Indian Restaurant.
3. This analysis definitely provides all the details require to open a new restaurants and their pros and cons of this business.

2 Data

2.1. Data source:

1. Dataset containing the neighborhood and borough of Toronto can be imported from the Wikipedia page "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M" and read into data frame.
2. **Foursquare** is used to find the number of restaurant and their location in each neighborhood, which helps us to find the best location for new Indian restaurants.
3. The geographical coordinates of all neighborhoods in Toronto can be obtained by using "https://cocl.us/Geospatial data"

2.2Data cleaning:

We downloaded the dataset from the Wikipedia page "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M" and clean and then scrape the Wikipedia page and wrangle the data and read into new data frame. Data frame consists of three columns Postal code, Borough, Neighborhoods.

Out[78]:

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government
7	M8A	Not assigned	Not assigned
8	M9A	Etobicoke	Islington Avenue, Humber Valley Village
9	M1B	Scarborough	Malvern, Rouge
10	M2B	Not assigned	Not assigned

The data set contain Borough and neighborhood and their postal code. But some Borough and neighborhood are not assigned. So, we must remove non-assigned values in the table and more than one neighborhood can exist in one postal code. We must combine all the neighborhood with respect to each postal code separated by comma.

Out[81]:

	Postal Code	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

From geospatial we downloaded all the latitudes and longitudes for each Neighborhood. Our next step is to combine the neighborhood dataset and coordinates dataset to get complete Toronto dataset.

Out[86]:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

3.Methodlogy:

3.1 Foursquare:

Foursquare is a technology company that built a massive dataset of location data. What is interesting about Foursquare is that they were very smart about building their dataset. They actually crowd-sourced their data and had people use their app to build their dataset and add venues and complete any missing information they had in their dataset. We create a uniform resource identifier, or URI, and append it with extra parameters depending on the data that we are seeking from the database. Any call request we make is composed of, we can call this base URI, which is api.foursquare.com/v2, and we can request data about venues, users, or tips. But, every time we make a call request, you have to pass our developer account credentials, which are

your Client ID and Client Secret as well as what is called the version of the API, which is simply a date.

3.2 Explore Neighborhood:

By using Foursquare API, we will retrieve all the nearby venues information in Toronto like Restaurants, cafe, park etc. By combing all the venues, we will get the dataset with nearest venues in the neighborhoods. We want all the nearby venues for each neighborhood within a radius of 1000.

Then we merge all the nearby venues data with neighborhood data with venue category to form the nearest venue dataset for each neighborhood.

[19]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	56	56	56	56	56	56
Brockton, Parkdale Village, Exhibition Place	22	22	22	22	22	22
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	15	15	15	15	15	15
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	16	16	16	16	16	16
Central Bay Street	61	61	61	61	61	61
Christie	15	15	15	15	15	15
Church and Wellesley	80	80	80	80	80	80
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	32	32	32	32	32	32
Davisville North	9	9	9	9	9	9
Dufferin, Dovercourt Village	17	17	17	17	17	17
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North & West, Forest Hill Road Park	4	4	4	4	4	4
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
High Park, The Junction South	25	25	25	25	25	25

3.3 Machine Learning:

Categorical Data in the dataset is transformed into Numerical data by **One hot coding technique**. For each neighborhood, we convert individual venues into frequency to find how many venues are located in each neighborhood.

We created a new data frame that store the neighborhood and their mean frequency of Indian restaurants in the neighborhood.

[12]:

	Yoga Studio	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Baby Store	Bagel Shop	Baker
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

3.4 K mean Clustering:

We use **K mean Clustering** method to cluster all the neighborhoods based on the average of Indian restaurants in that neighborhood. Here we set K = 3. Neighborhoods that has a similar mean frequency of Indian Restaurants were divided into 3 clusters and labelled from 0. We then merge the cluster table with venue data which would be the basis for analyzing the best opportunities to open Indian restaurant in Toronto.

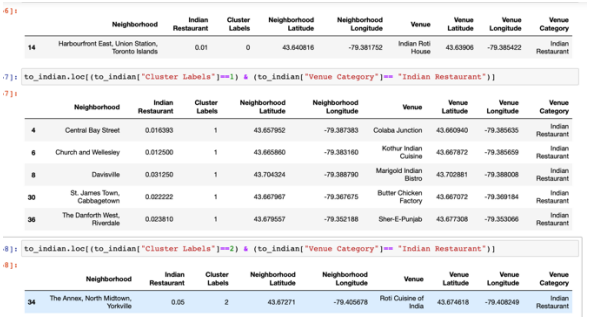
	Neighborhood	Indian Restaurant
0	Berczy Park	0.000000
1	Brockton, Parkdale Village, Exhibition Place	0.000000
2	Business reply mail Processing Centre, South C...	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
4	Central Bay Street	0.016393
5	Christie	0.000000
6	Church and Wellesley	0.012500
7	Commerce Court, Victoria Hotel	0.000000
8	Davisville	0.031250
9	Davisville North	0.000000
10	Dufferin, Dovercourt Village	0.000000
11	First Canadian Place, Underground city	0.000000
12	Forest Hill North & West, Forest Hill Road Park	0.000000
13	Garden District, Ryerson	0.000000
14	Harbourfront East, Union Station, Toronto Islands	0.010000
15	High Park, The Junction South	0.000000
16	India Bazaar, The Beaches West	0.000000
17	Kensington Market, Chinatown, Grange Park	0.000000
18	Lawrence Park	0.000000
19	Little Portugal, Trinity	0.000000

3.5 Data analysis:

The cluster 1 and 3 has the least number of Indian restaurants. The neighborhoods in the cluster 1 are Harbourfront East, Union Station, Toronto Islands and the neighborhoods in cluster 3 are The Annex, North Midtown, Yorkville. The cluster 2 has highest number of neighborhoods with highest number of Indian restaurants.

3.5.1: cluster 0:

The cluster 0 has Harbourfront East, Union Station, Toronto Islands neighborhoods with least number of Indian restaurants.



```
[6]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
14	Harbourfront East, Union Station, Toronto Islands	0.01	0	43.640816	-79.381752	Indian Roti House	43.63906	-79.385422	Indian Restaurant

```
[7]: to_indian.loc[(to_indian["Cluster Labels"]==1) & (to_indian["Venue Category"]== "Indian Restaurant")]
```

```
[7]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Central Bay Street	0.016393	1	43.657952	-79.387383	Colaba Junction	43.660940	-79.385635	Indian Restaurant
6	Church and Wellesley	0.012500	1	43.665860	-79.383160	Kothur Indian Cuisine	43.667872	-79.385659	Indian Restaurant
8	Davisville	0.031250	1	43.704324	-79.388790	Marigold Indian Bistro	43.702881	-79.388008	Indian Restaurant
30	St. James Town, Cabbagetown	0.022222	1	43.667967	-79.367675	Butter Chicken Factory	43.667072	-79.369184	Indian Restaurant
36	The Danforth West, Riverdale	0.023810	1	43.679557	-79.352188	Sher-E-Punjab	43.677308	-79.353066	Indian Restaurant

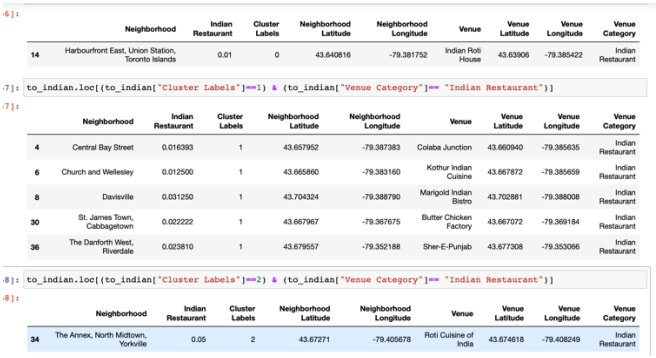
```
[8]: to_indian.loc[(to_indian["Cluster Labels"]==2) & (to_indian["Venue Category"]== "Indian Restaurant")]
```

```
[8]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
34	The Annex, North Midtown, Yorkville	0.05	2	43.67271	-79.405678	Roti Cuisine of India	43.674618	-79.408249	Indian Restaurant

3.5.2: cluster 1:

The cluster 1 has highest number of neighborhoods with highest number of Indian restaurants. The neighborhoods are Central Bay Street, church wellesly, Davis Ville, St. Jamestown, Cabbage town, The Danforth West, Riverdale. The highest average of Indian restaurants is 0.3125.



```
[6]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
14	Harbourfront East, Union Station, Toronto Islands	0.01	0	43.640816	-79.381752	Indian Roti House	43.63906	-79.385422	Indian Restaurant

```
[7]: to_indian.loc[(to_indian["Cluster Labels"]==1) & (to_indian["Venue Category"]== "Indian Restaurant")]
```

```
[7]:
```

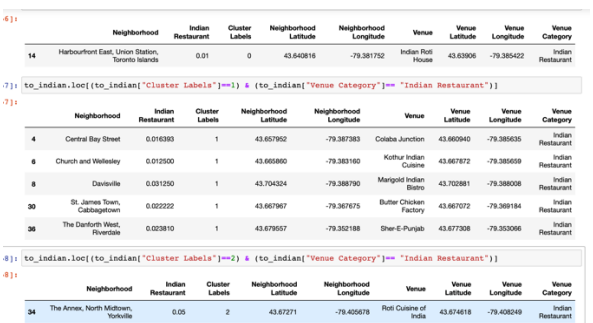
	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Central Bay Street	0.016393	1	43.657952	-79.387383	Colaba Junction	43.660940	-79.385635	Indian Restaurant
6	Church and Wellesley	0.012500	1	43.665860	-79.383160	Kothur Indian Cuisine	43.667872	-79.385659	Indian Restaurant
8	Davisville	0.031250	1	43.704324	-79.388790	Marigold Indian Bistro	43.702881	-79.388008	Indian Restaurant
30	St. James Town, Cabbagetown	0.022222	1	43.667967	-79.367675	Butter Chicken Factory	43.667072	-79.369184	Indian Restaurant
36	The Danforth West, Riverdale	0.023810	1	43.679557	-79.352188	Sher-E-Punjab	43.677308	-79.353066	Indian Restaurant

```
[8]: to_indian.loc[(to_indian["Cluster Labels"]==2) & (to_indian["Venue Category"]== "Indian Restaurant")]
```

```
[8]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
34	The Annex, North Midtown, Yorkville	0.05	2	43.67271	-79.405678	Roti Cuisine of India	43.674618	-79.408249	Indian Restaurant

3.5.3: cluster 2 The cluster 2 has The Annex, North Midtown, Yorkville with least number of Indian restaurants.



```
[6]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
14	Harbourfront East, Union Station, Toronto Islands	0.01	0	43.640816	-79.381752	Indian Roti House	43.63906	-79.385422	Indian Restaurant

```
[7]: to_indian.loc[(to_indian["Cluster Labels"]==1) & (to_indian["Venue Category"]== "Indian Restaurant")]
```

```
[7]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
4	Central Bay Street	0.016393	1	43.657952	-79.387383	Colaba Junction	43.660940	-79.385635	Indian Restaurant
6	Church and Wellesley	0.012500	1	43.665860	-79.383160	Kothur Indian Cuisine	43.667872	-79.385659	Indian Restaurant
8	Davisville	0.031250	1	43.704324	-79.388790	Marigold Indian Bistro	43.702881	-79.388008	Indian Restaurant
30	St. James Town, Cabbagetown	0.022222	1	43.667967	-79.367675	Butter Chicken Factory	43.667072	-79.369184	Indian Restaurant
36	The Danforth West, Riverdale	0.023810	1	43.679557	-79.352188	Sher-E-Punjab	43.677308	-79.353066	Indian Restaurant

```
[8]: to_indian.loc[(to_indian["Cluster Labels"]==2) & (to_indian["Venue Category"]== "Indian Restaurant")]
```

```
[8]:
```

	Neighborhood	Indian Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
34	The Annex, North Midtown, Yorkville	0.05	2	43.67271	-79.405678	Roti Cuisine of India	43.674618	-79.408249	Indian Restaurant

4.Result and Discussion:

4.1 Result:

Thus, we reached the final analysis to open a new restaurant in Toronto, Canada. We looked into all neighborhoods and analyzed Indian population and number restaurants in each neighborhoods. Toronto has 11 boroughs but Central Toronto, Downtown Toronto, East Toronto, East York, North York & Scarborough boroughs have a high number of Indian restaurants.

We found that Downtown Toronto, Central Toronto, East York are populated with Indian restaurants' we will ignore those neighborhoods.

Thus, North York and east Toronto would be the best place to start new Indian restaurants in Toronto, Canada.Theresfore, it is better to open new restaurants in Scarborough because it has high number of Indian population and low competition.

4.2 **Discussion**:

According our final analysis, Scarborough would be the best and least competition place for new Indian restaurants and also it is densely populated by Indians which provide high customer possibility. Definitely this place would be the best place to open new Indian restaurants in Toronto, Canada

5. Conclusion:

In conclusion, we had an opportunity to work on real business problem in data science in IBM data science professional certificate course. We utilized many python libraries and foursquare to explore all neighborhoods. We also visualized maps and different plots using Matplotlib and folium to picture a map. We can use this technique to analysis any scenario such opening new grocery shops etc. This project helps us to understand more about data science.