DATA SCIENCE CAPSTONE

BATTLE OF NEIGHBOURHOOD

Introduction

Problem Description:

When one changes the city, he/she wants to be in the same environment in which he/she is currently residing. Like, if he/she is living in a family, he/she wants to be in the society in which most of the families are living. Say this is you and you live on the west side of the city of Toronto in Canada. You love your neighborhood, mainly because of all the great amenities and other types of venues that exist in the neighborhood, such as gourmet fast food joints, pharmacies, parks, grad schools. Now say you receive a job offer from a great company on the other side of the city with great career prospects. However, given the far distance from your current place you unfortunately must move if you decide to accept the offer. Wouldn't it be great if you're able to determine neighborhoods on the other side of the city that are the same as your current neighborhood, and if not, perhaps similar neighborhoods that are at least closer to your new job?

So, we need to compare different neighborhoods in terms of a service, search for potential explanation of
1) why a neighborhood is popular,
2) the cause of complaints in another neighborhood,
3) or anything else related to neighborhoods.

I will segment it into different neighborhoods using the geographical coordinates of the center of each neighborhood, and then using a combination of location data and machine learning, I will group the neighborhoods into clusters.

Target audience:

I believe this is for every person who is shifting from one place to another. As the need for a job everyone must travel and wants to be in a place where he got all the amenities. After a long hour of job, he needs that the daily needs should be nearer as much it can be. So, predicting the Battle of neighborhood is correct for description for this project.

Dataset:

1) Using Canada Dataset: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and get latitude, longitude from https://cocl.us/Geospatial_data'

2) Finding based on address the nearby restaurants using Foursquare API. By using this API we will get all the venues in each neighborhood. We can filter these venues to get restaurants based on input we have provided.

Analysis from the Dataset:

	Postcode	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	Harbourfront

Number of Neighbourhood
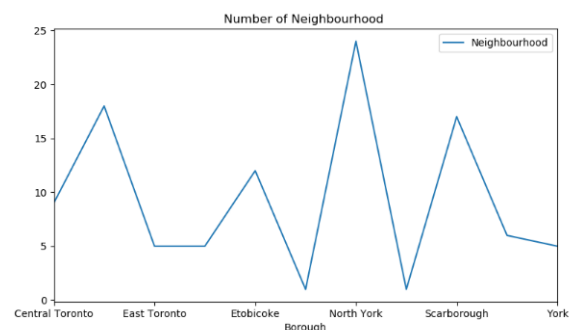
Using Foursquare API:

▶ Using Foursquare API, we input the address where we are residing or from where to we search the restaurant.

▶ Using API, we get the latitude and longitude of the given location

▶ Now, input the type of restaurant you want to search.

▶ You will get all the restaurants in radius of 500.

▶ We get the rating of those restaurants and sort them.

▶ Recommend the restaurant which is having the highest rating.

Let's visualize the restaurants that are nearby

```
[19]:  dataframe_filtered.name
```

```
[19]:  0            Indian Biriyani House
       1            Indian Biriyani House
       2          Touch - Indian Cuisine
       3                   Indian Flavour
       4             Ram's Indian kitchen
       5      Tamarind: The Indian Kitchen
       6           Chadani Indian Cuisine
       Name: name, dtype: object
```

[27]:

	name	lat	lng	categories	Rating	Pricing	Likes	distance
1	Indian Biriyani House	43.655120	-79.386645	Indian Restaurant	7.1	2.0	5.0	514
2	Touch - Indian Cuisine	43.649869	-79.378218	Indian Restaurant	0.0	2.0	2.0	445
5	Tamarind: The Indian Kitchen	43.646859	-79.378707	Indian Restaurant	0.0	1.0	2.0	609
0	Indian Biriyani House	43.650050	-79.380662	Indian Restaurant	0.0	2.0	1.0	255
3	Indian Flavour	43.655649	-79.384119	Indian Restaurant	0.0	2.0	0.0	508
4	Ram's Indian kitchen	43.648026	-79.379819	Indian Restaurant	0.0	2.0	0.0	452
6	Chadani Indian Cuisine	43.649153	-79.374814	Indian Restaurant	0.0	2.0	0.0	730

Results:
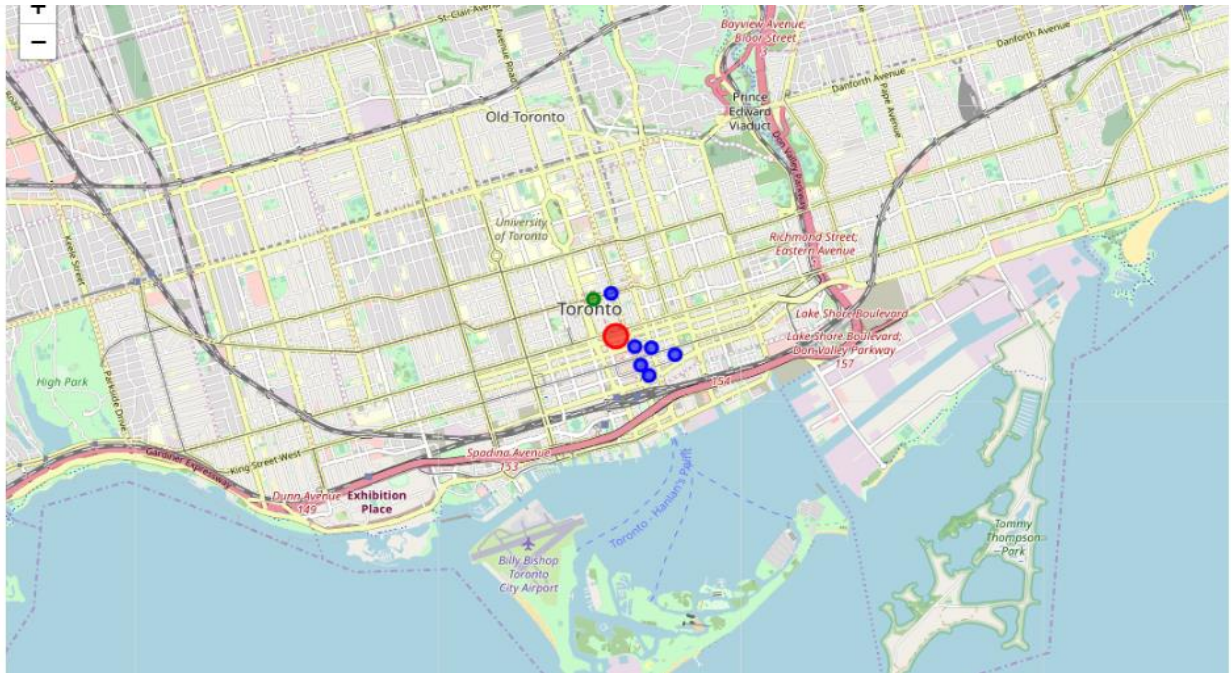
Summary:

First requirement is solved by applying K-means clustering to perform city segmentation. With the neighborhoods being clustered into clusters, this task becomes finding the neighborhoods within the same cluster
The Second and Third requirement can be solved by providing venue and ratings information from Foursquare API.

Conclusion:

The Green Dot on the map shows the Restaurant having the highest rating and recommended. And the blue dots are the other restaurants that are having less rating from the recommended one.