The location recommendation to new restaurants

1. Introduction

Location is the most important factor of success to open a new restaurant. There are some key indexes to evaluate if the location is good or not:

- How's the neighborhood in terms of population to check if the restaurant would meet local consumers' needs
- What's the current competition situation within the location, in terms of the number of restaurants, cuisines offering, comments, etc.
- What's the future opportunity within the location, considering local development plan

This project will conduct an analysis on Toronto neighborhood to set up a model for investors to make decision on how to select the location for their new restaurants.

2. Data Acquisition

2.1 Here is the data list used for this analysis:

- List of postal codes of Toronto from WIKIPEDIA
- Geospatial Coordinates CSV file for Toronto postal codes from http://cocl.us/Geospatial_data
- Foursquare API

2.2 The process of data cleaning on postal codes of Toronto:

- Read the HTML text using BeautifulSoup library of the list of postal codes of Toronto
- Group neighborhood with the same postal code
- Drop data that without info. of Borough

Here is the dataframe built up by pandas.

	Postal code	Borough	Neighborhood
0	M1B	Scarborough	Malvern / Rouge
1	M1C	Scarborough	Rouge Hill / Port Union / Highland Creek
2	M1E	Scarborough	Guildwood / Morningside / West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

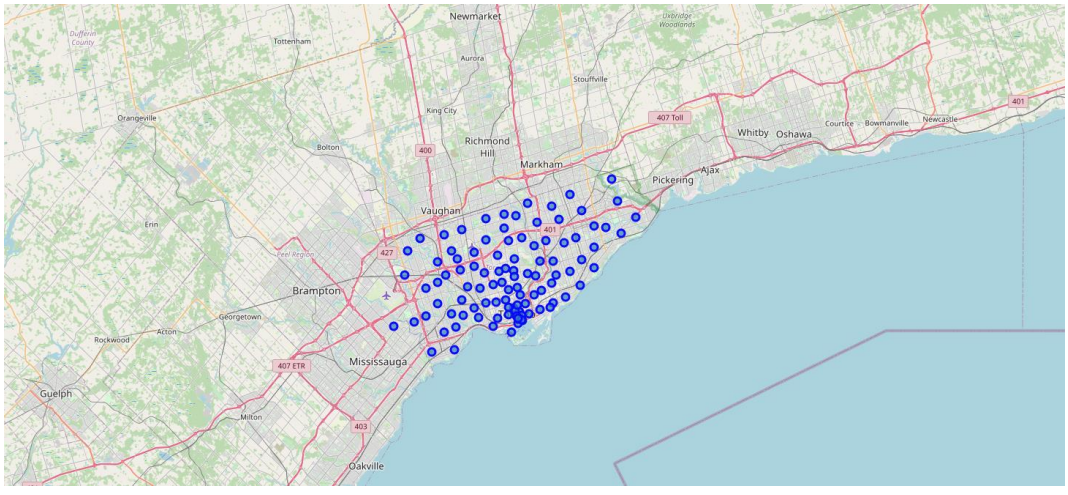
2.3 Merge Geo-spatial coordinates into the dataframe

- Read the Geospatial Coordinates CSV file from website: http://cocl.us/Geospatial_data
- Merge it into Postal Code dataframe by matching with Postal code column

Here is the merged dataset.

	Postal code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park / Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor / Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.662301	-79.389494

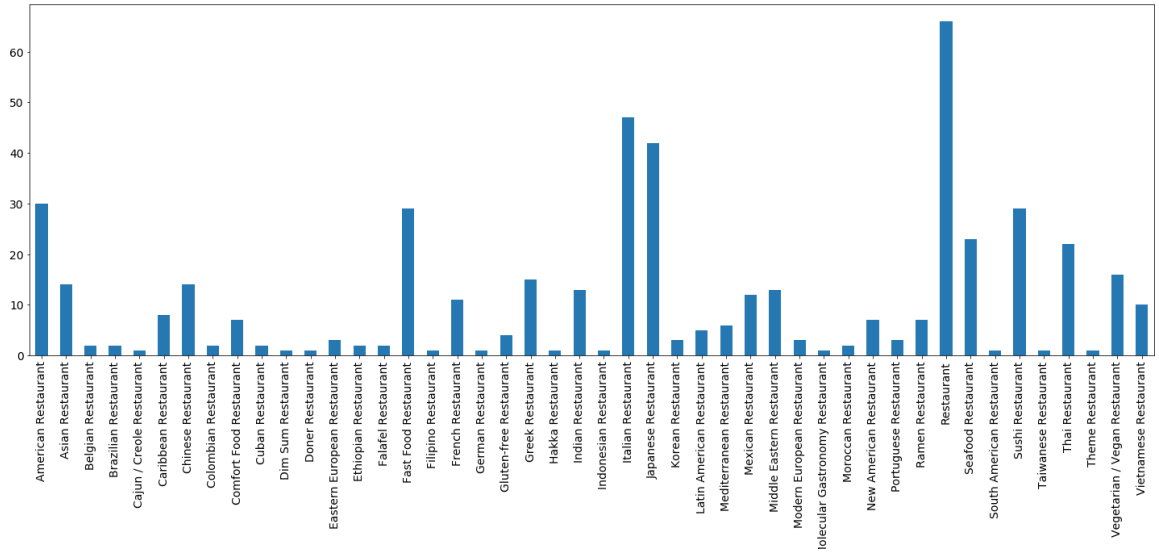
And then I plotted a map of Toronto by using Folium library:



2.4 Explore the neighborhoods and segment by using Foursquare API

By linking with Foursquare API, we can check out the top100 restaurants in each neighborhoods within a radius of 500 metres, with the info. of restaurant name and category. And then we grouped them by neighborhood.

To make the data clear, we only selected those categories with ‘restaurant’ in category name. Here is the distribution of all types of restaurant at Toronto.



And we can also know how many restaurants in different category within each neighborhood.

	Neighborhood	Venue Category_American Restaurant	Venue Category_Asian Restaurant	Venue Category_Belgian Restaurant	Venue Category_Brazilian Restaurant	Venue Category_Cajun / Creole Restaurant	Venue Category_Caribbean Restaurant	Venue Category_Chinese Restaurant	Venue Category_Colombian Restaurant	Venue Category_Comfort Food Restaurant	...	Venue Category_Ramen Restaurant	Venue Category_Restaurant	Ca
209	Toronto Dominion Centre / Design Exchange	1	0	0	0	0	0	0	0	0	...	0	0	
329	Canada Post Gateway Processing Centre	1	0	0	0	0	0	0	0	0	...	0	0	
23	Garden District, Ryerson	1	0	0	0	0	0	0	0	0	...	0	0	
269	Cliffside / Cliffcrest / Scarborough Village West	1	0	0	0	0	0	0	0	0	...	0	0	
255	Commerce Court / Victoria Hotel	1	0	0	0	0	0	0	0	0	...	0	0	

5 rows × 47 columns

3. Data Analysis

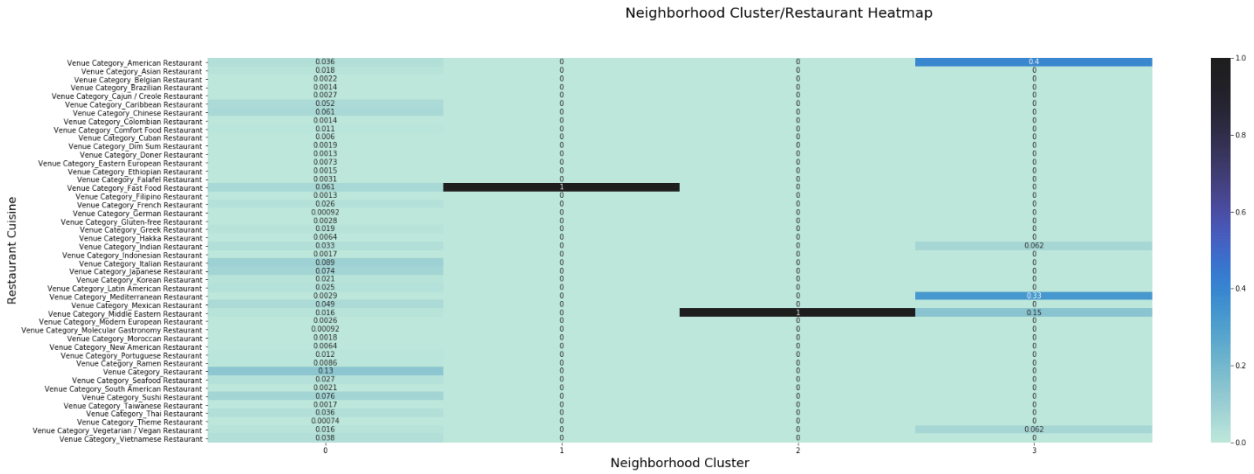
3.1 Model selection

I use K-means clustering and popularity recommendation filtering aproach to make the data analysis. With K=4, I split neighborhood to 4 clusters.

```
Cluster Labels
0    54
1     5
2     2
3     4
Name: Neighborhood, dtype: int64
```

3.2 Result section

In this stage, I merged cluster labels to the dataset, and then grouped the dataset by cluster. The we got the heatmap of restaurants in each cluster.



4. Conclusion

By using the heatmap above, we can check the frequency of a certain type of restaurant in each cluster. For example, Middle Eastern restaurant frequency is very high in cluster2 than others, which means if someone want to start a Middle Eastern restaurant in the neighborhood of cluster2, the competition would be more fierce than on other areas.

```
[9]: toronto_clusters_trans.loc['Venue Category_Middle Eastern Restaurant'].plot(kind='bar')
t[139]: <matplotlib.axes._subplots.AxesSubplot at 0x7f16a3347400>
```