Venue suggestion on opening a start-up restaurant in Toronto

Introduction

- Toronto is the most populous city in Canada. Toronto would be the first choice if planning to open a business in Canada.
- Food is an amazing deal in the city of Toronto. It comprises a mixture of cuisines, extending around the world.
- It is the most multicultural city in the country.
- Brand New places for eating are continually opening, and chefs continue to try hard to come up with new and contemporary ideas.

Business Problem

The main goal of this project is to help public to identify which area of toronto consist of most concentrate of restaurant. By the result giving out in this project, area will be suggested for those who want to open a restaurant business in Toronto. The suggested area is based on the condition that have fewer candidates (other restaurants).

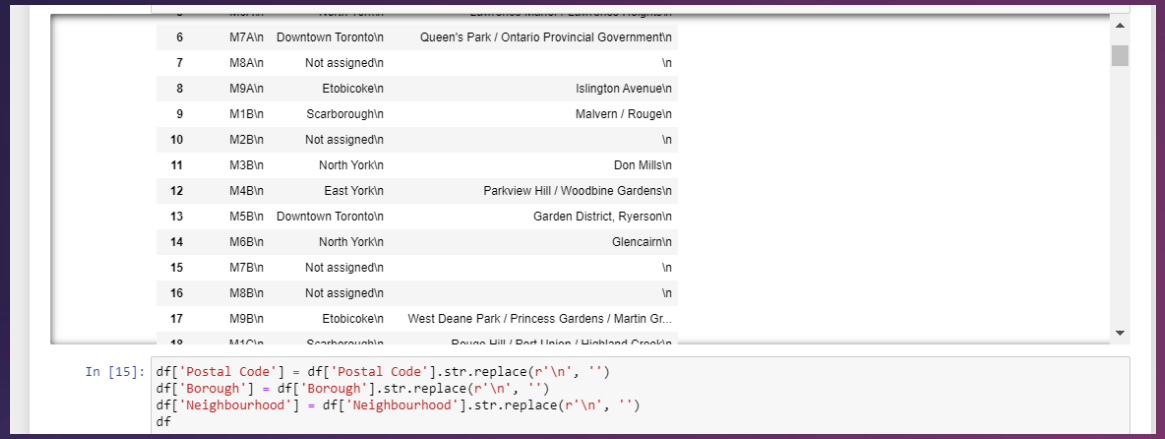
Data Selection

The list of Canada postal code is obtained from wikipedia in order to get the neighbourhood of Canada.

- Canada postal code source : https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Canada neighbourhood geolocation data from the file : C:/Users/WIN10/Desktop/Geospatial_Coordinates.csv

Beautiful soup is used to read the web page. The geolocation data is used to locate all the canada neighbourhood obtained from wikipedia. The reason to get the neighbourhood geolocation is to match with the Foursquare location data to perform the analysis. From there, the venue data of the respecting neighbourhood can be obtained.

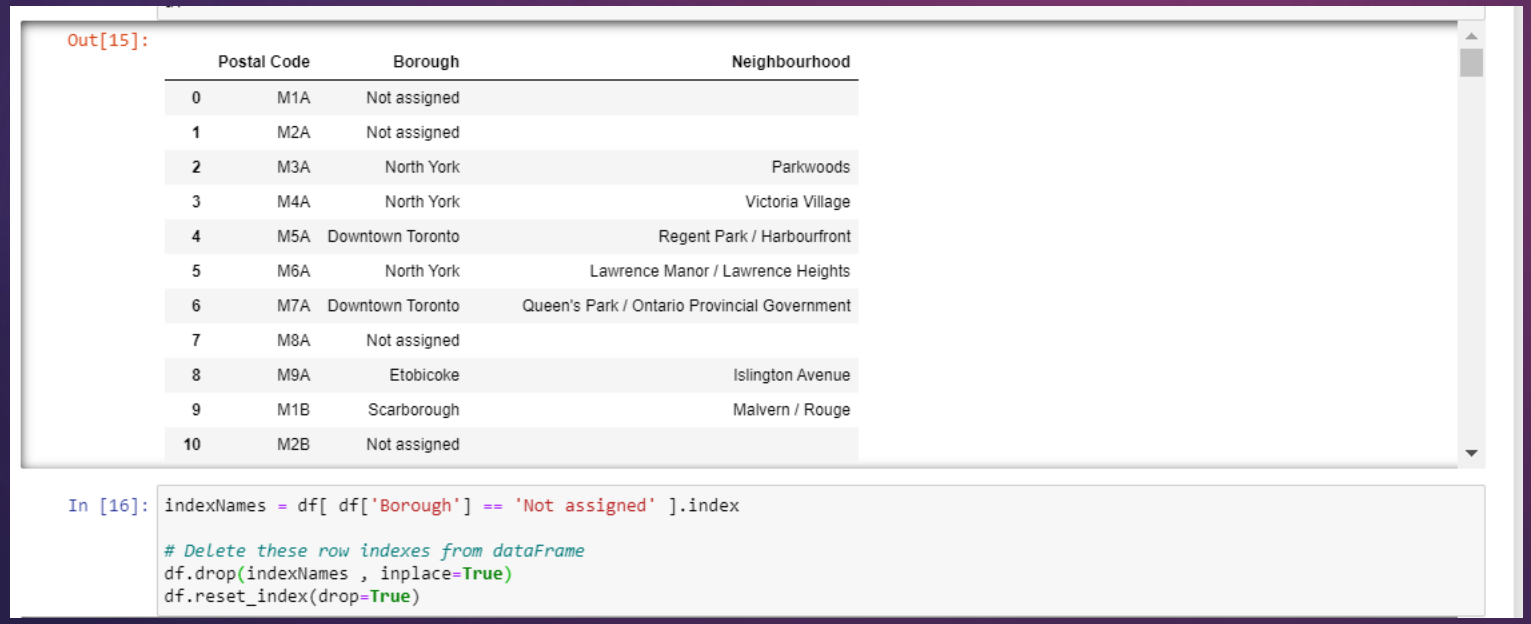
Cleansing of Data



6	M7A\n	Downtown Toronto\n	Queen's Park / Ontario Provincial Government\n
7	M8A\n	Not assigned\n	\n
8	M9A\n	Etobicoke\n	Islington Avenue\n
9	M1B\n	Scarborough\n	Malvern / Rouge\n
10	M2B\n	Not assigned\n	\n
11	M3B\n	North York\n	Don Mills\n
12	M4B\n	East York\n	Parkview Hill / Woodbine Gardens\n
13	M5B\n	Downtown Toronto\n	Garden District, Ryerson\n
14	M6B\n	North York\n	Glencairn\n
15	M7B\n	Not assigned\n	\n
16	M8B\n	Not assigned\n	\n
17	M9B\n	Etobicoke\n	West Deane Park / Princess Gardens / Martin Gr...

```
In [15]: df['Postal Code'] = df['Postal Code'].str.replace(r'\n', '')
         df['Borough'] = df['Borough'].str.replace(r'\n', '')
         df['Neighbourhood'] = df['Neighbourhood'].str.replace(r'\n', '')
         df
```

Remove the "\n" in the dataframe by using the code showed in the picture.



Out[15]:

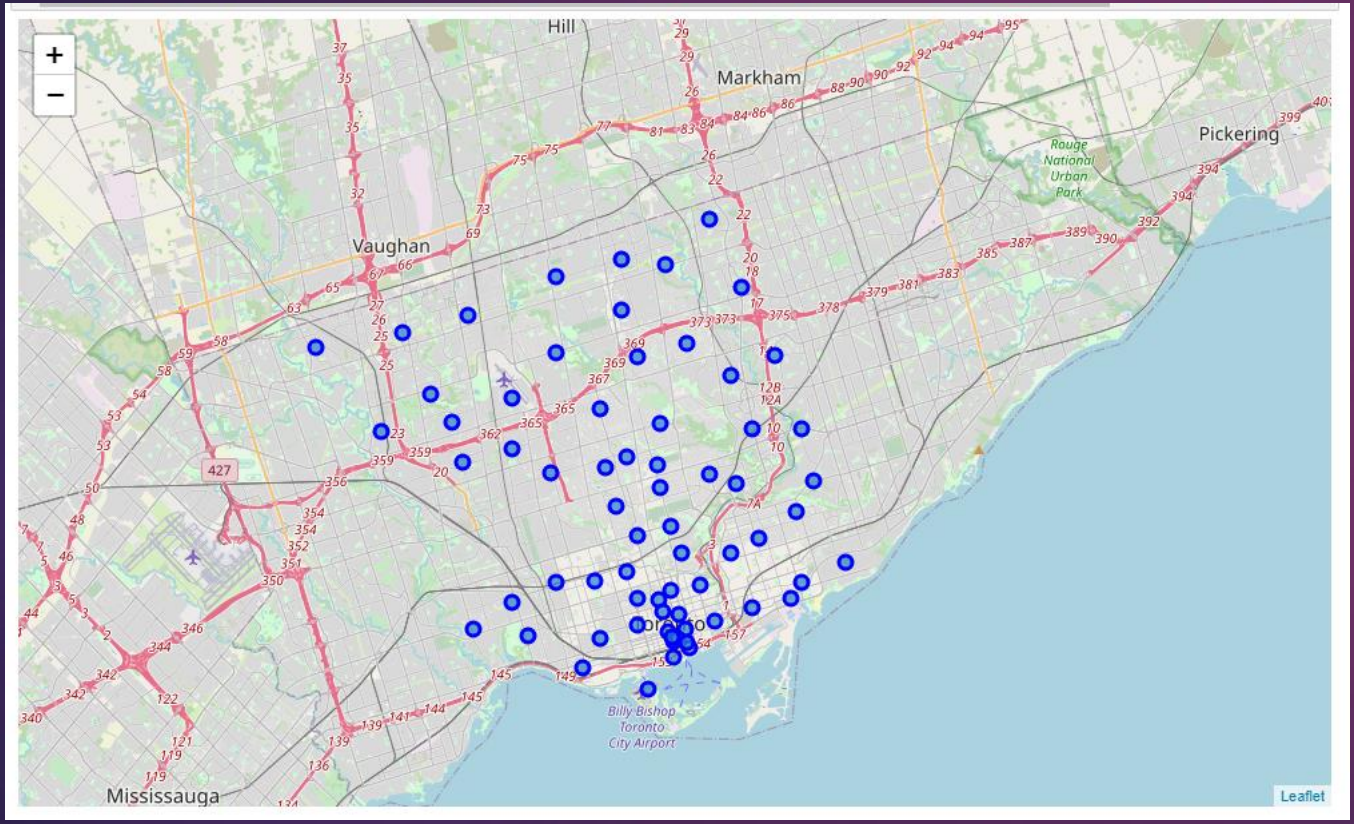
	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	
1	M2A	Not assigned	
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park / Harbourfront
5	M6A	North York	Lawrence Manor / Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government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

```
In [16]: indexNames = df[ df['Borough'] == 'Not assigned' ].index

         # Delete these row indexes from dataFrame
         df.drop(indexNames , inplace=True)
         df.reset_index(drop=True)
```

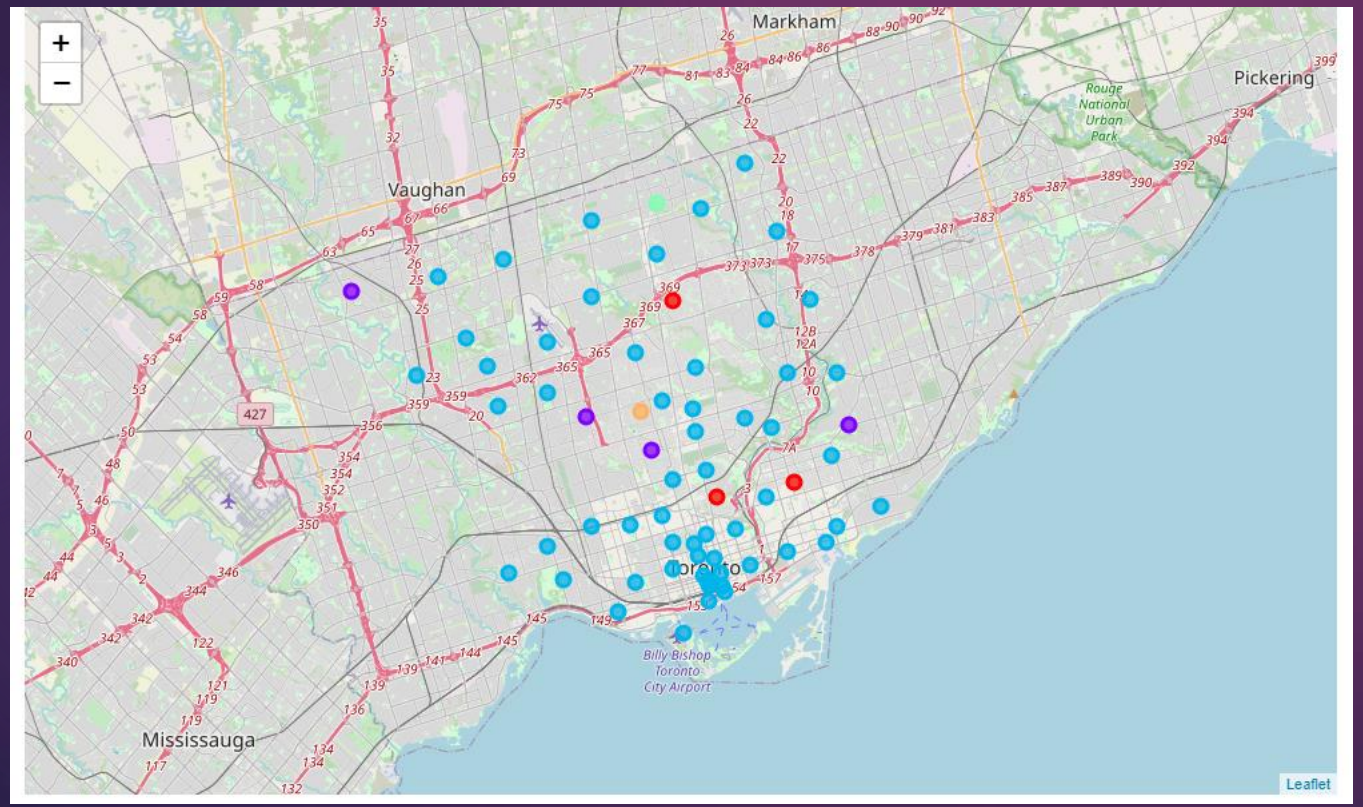
Remove all the rows with Borough = Not assigned to obtain all relevant data only.

Results



- Neighbourhoods of Toronto are plotted in the map. All the location data (longitude and latitude) are obtained in a csv file : Geospatial_Coordinates.csv.
- This map is plotted for the fundamental of the analysis with Foursquare data.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bathurst Manor / Wilson Heights / Downsview North	Coffee Shop	Bank	Sushi Restaurant	Middle Eastern Restaurant	Pharmacy
1	Bayview Village	Chinese Restaurant	Café	Japanese Restaurant	Bank	Women's Store
2	Bedford Park / Lawrence Manor East	Thai Restaurant	Coffee Shop	Italian Restaurant	Restaurant	Sandwich Place
3	Berczy Park	Coffee Shop	Cocktail Bar	Seafood Restaurant	Farmers Market	Café
4	Brockton / Parkdale Village / Exhibition Place	Café	Coffee Shop	Breakfast Spot	Bakery	Convenience Store



- The top 5 most common venue in each neighbourhood are found.
- All the neighbourhoods are classified into 5 clusters.
  Cluster 0 : Park
  Cluster 1 : Park, Pizza Place
  Cluster 2 : Not suggested (concentrated)
  Cluster 3 : Gym
  Cluster 4 : Garden

Suggestion

Throughout the cluster presented above, cluster 2 is not suggestted to open a restaurant business because there are too many candidates around which may make less profit than other clusters. The suggested clusters are cluster 0, 1 , 3 and 4. The listed clusters are lacked of restaurants in the respecting neighbourhood which may attract the consumers around to the new start-up restaurant.

Discussion

Foursquare is a good platform to obtain venue data to perform similiar analysis. The weakness would be some location which cannot be find in Foursquare and return with no venue data. This may affect effectiveness of your data set.

Conclusion

This project can conclude that Toronto is a city full of different type of restaurant in cluster 2. If the business owner need to expand or start up a new restaurant business in Toronto, cluster 0, 1, 3 and 4 will be suggested.