Introduction

In this article, I am going to provide you my exploratory analysis of neighborhoods in Canada. For my analysis, I have used the data provided for Toronto, as one of the biggest Canadian cities with a huge potential of opening a business.

**A. Introduction, Business Issue, Data Description.**

**A.1. Introduction**

Canada is very attractive for foreigners, due to its high living standards, strong economy, strong market and soft regulations for immigrants.

According to Wikipedia, Toronto is a capital of Canadian province of Ontario. Toronto is the most populated city in Canada with 2,731,000mln people registered in 2016. Important to emphasize, that the GDP of Toronto is CA$385.1billion and GDP per capita is CA$57.004, which are very high and that fact makes the city very attractive for foreign investors.

**A.2. Business Issue**

My client is an investor from Kazakhstan, who wants to move to Canada and continue his Coffee business in Toronto. My client has Coffee shop network located in CIS are and wants to expand on Canadian market. His Coffee shop called “Salem” is a shop, which collected coffee recipes from all over the world.

His main issue is that there is a risk that Canadians may not understand his coffee and due to the strong competition, our investor can easily lose his money.

I am going to use exploratory analysis to see the most suitable location among all neighborhoods.

**A.2. Data Description**

To analyze the problem, we will need the data listed below:

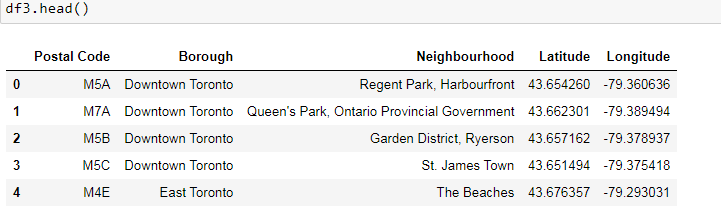
* Wikipedia page of Toronto:

<https://en.wikipedia.org/wiki/Toronto>

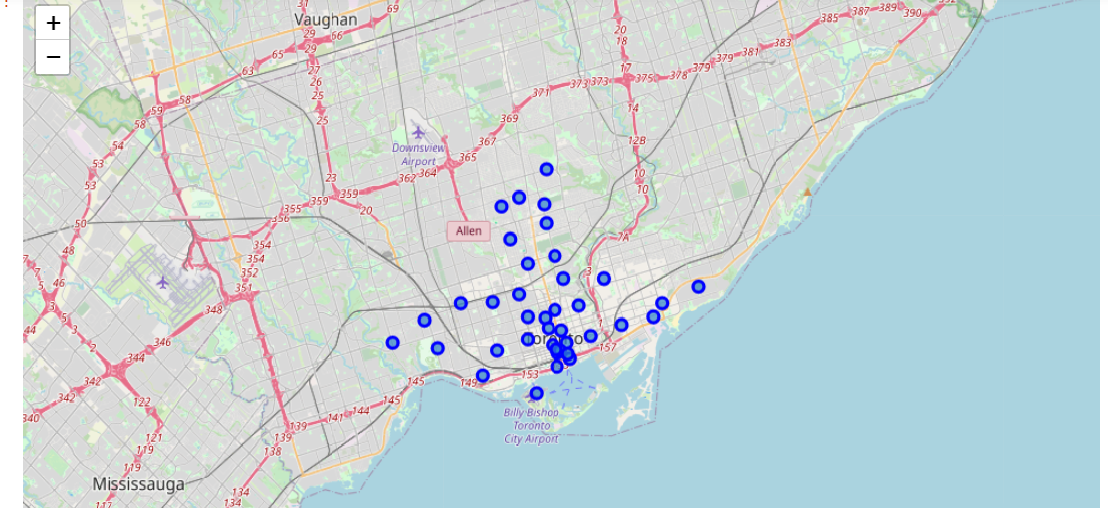
* List of Postal Codes in Canada (Neighborhoods) <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* List of Latitude and Longitude data per each neighborhood: <https://cocl.us/Geospatial_data>
* Foursquare API, to get the most common venues of given Borough of Toronto

**B. Background Explanation**

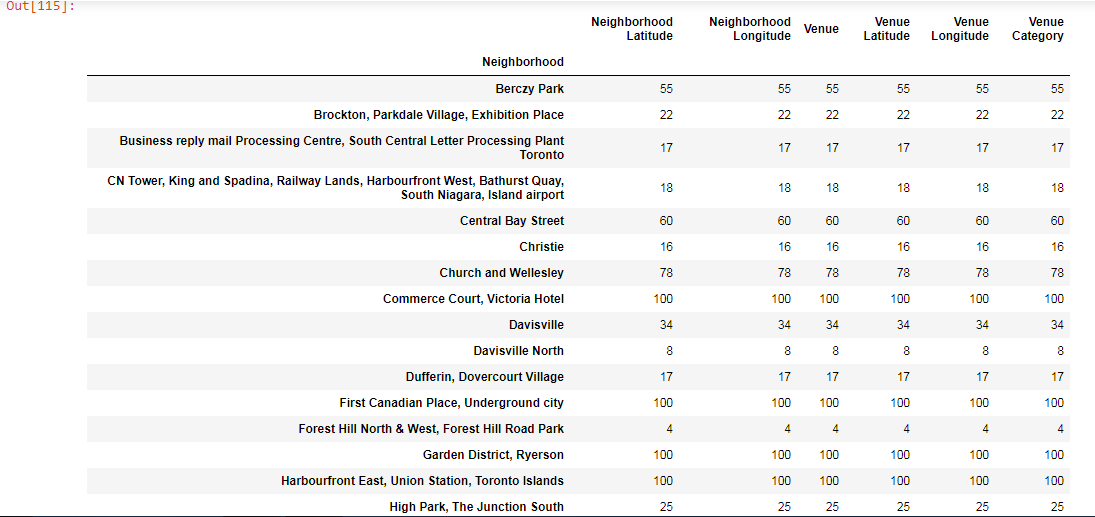
Here is my Master data, list of each neighborhood with its coordinates.



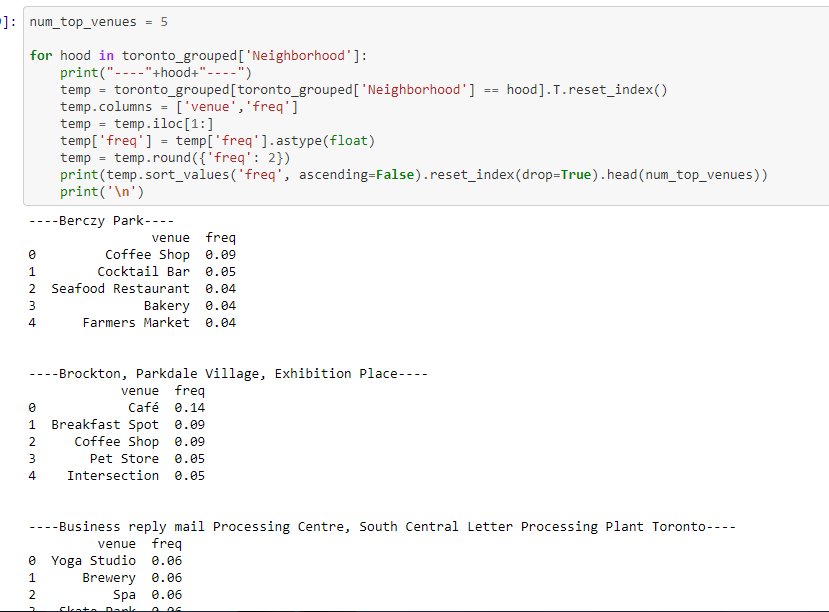
Here is an initial visualization of neighborhoods in Toronto with a help of **Folium** library.



With a help of Foursquare API, we are able to see the amount of venues per each neighborhood. There are 236 unique categories in a table:



Our aim is to see the neighborhoods with a lowest concentration of Coffee Shops in order to have less competition in an area. To do this we first have to see the grouped rows per each neighborhood and the frequency of occurrence per each category.

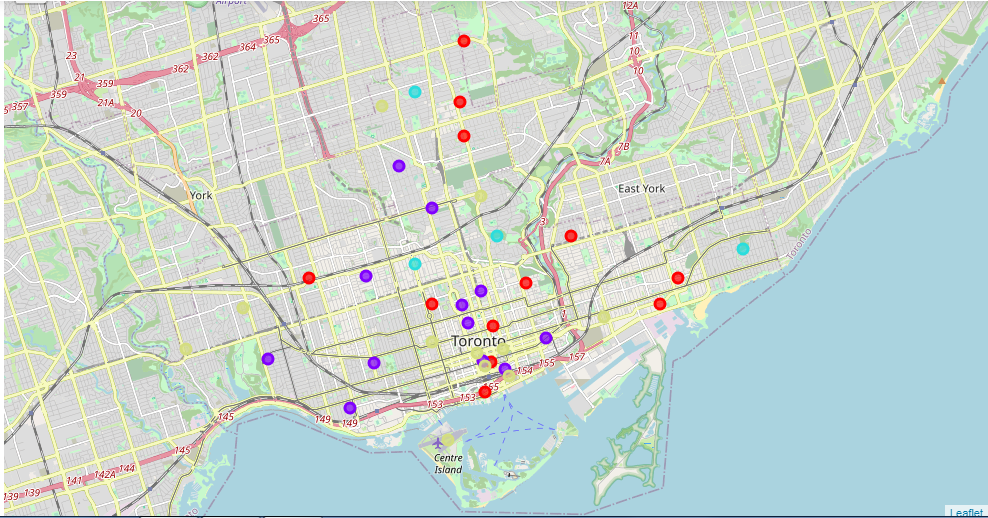




As a result, we can see the table of 10 most popular venues. Luckily, for our investor we see, on the table and from the previous code, we can see that Coffee Shops are very popular in most of neighborhoods, which means, there is a high demand for them.

Now our main aim is to find out, which area will be more potential to open a Coffee shop. For this investigation, we have to run the Clustering process. In this case, I used K means clustering to group the data in our table and find out, which neighborhoods have a lowest concentration of Coffee Shops.



Now we have clustering results, we can visualize them using folium map.

**Cluster 0:** contains area with no Coffee Shops **(Red Color)**

**Cluster 1:** Highest amount of Coffee Shops **(Purple Color).**

**Cluster2:** Low amount, low competition **(Mint green color).**

**Cluster3:** Medium Amount of Coffee Shops **(Light Brown color)**

**Result**

Based on Analysis above, we can see that the best neighborhoods for opening the Coffee Shop are Cluster labels 0 and 2.

**C. Conclusion**

Nowadays, many people are traveling and migrating from place to place. Mainly people are focusing on big cities to grow and develop by either opening a business or seeking for a job. With nowadays opportunities and huge amount of data from sources, which are either reliable or not, people still have a risk of being mistaken. Especially it is harmful for people with families, who are seeking for better life.

We can see that Data Analysis is very helpful in case of predicting the place for opening a business. The model can be used in other cases as well. For instance, model may be used to find a proper location for opening a shopping mall, restaurant, drugstore etc.

I hope this analysis may be useful for future predictions.

Feel free to find a code on a github repository: <https://github.com/jaksylyk327/Capstone-Project-ZZ/blob/main/Final%20Capstone%20Project.ipynb>

Zhaksylyk Zamanbek