**Capstone Project - The Battle of Neighborhoods**

**Introduction:**

In this project, I will use the data location Foursquare API and some of the machine learning algorithms and data science to solve a business problem. The main objective of the project is to use these available resources for starting a new Restaurants and/or food market business in another city. In the first step, will describe the problem and then data preparation for the analysis then the results and conclusion.

**A description of the problem and a discussion of the background**

Toronto, city, capital of the province of Ontario, southeastern Canada. It is the most populous city in Canada, a multicultural city, and the country’s financial and commercial center. Also, Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world. So, it considered one of the most preferred destinations for the visitors and tourists to come from inside and outside of Canada and also the immigrant and newcomers from around the globe choose this city as their new home, especially the people from the middle east and/or Arabic people that come from all countries of the Arab world. There is a lot of the population from the middle east/Arabic, so it is a good place to start a new restaurant and/or food market.

Therefore, the main idea is to make a study and analysis and to decide whether it's good to make this business or not.

That is, to use data science methods and the data collected from the Toronto city such as the population distribution and tourist places that attract the visitors and how it's close in location to the city center or downtown and to decide where to open it.

**Who would be interested in this project?**

Middle east and the Arab community that looking out to find places and neighborhoods with different options for the Middle East and the Arab food market and restaurants.

Also, New Businessmen that want to find a good location to open a Middle East restaurant in Toronto and its neighborhoods. This work analysis will give them some information about the Toronto area and start their business.

**A description of the data and how it will be used to solve the problem.**

First will start with the data of Toronto city that give information about the city and its neighborhoods from the Wikipedia web site (List of postal codes of Canada: M)

(https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M).

This gives all the information about the neighborhoods present in Toronto, also the postal code, borough & the name of all the neighborhoods.

Also, using (http://cocl.us/Geospatial\_data) a link to a CSV file that has the geographical coordinates of each postal code.

In addition, using the Wikipedia web (https://en.m.wikipedia.org/wiki/Demographics\_of\_Toronto#Ethnic\_diversity) the "Demographics of Toronto” that used to find the neighborhoods that are densely populated of middle east/Arab people as it might be helpful in identifying the appropriate place to open this new business.

Also, the Wikipedia site (https://en.wikipedia.org/wiki/Arab\_Canadians) shows the distribution of the Arab population of Canada that will help us to choose the right location.

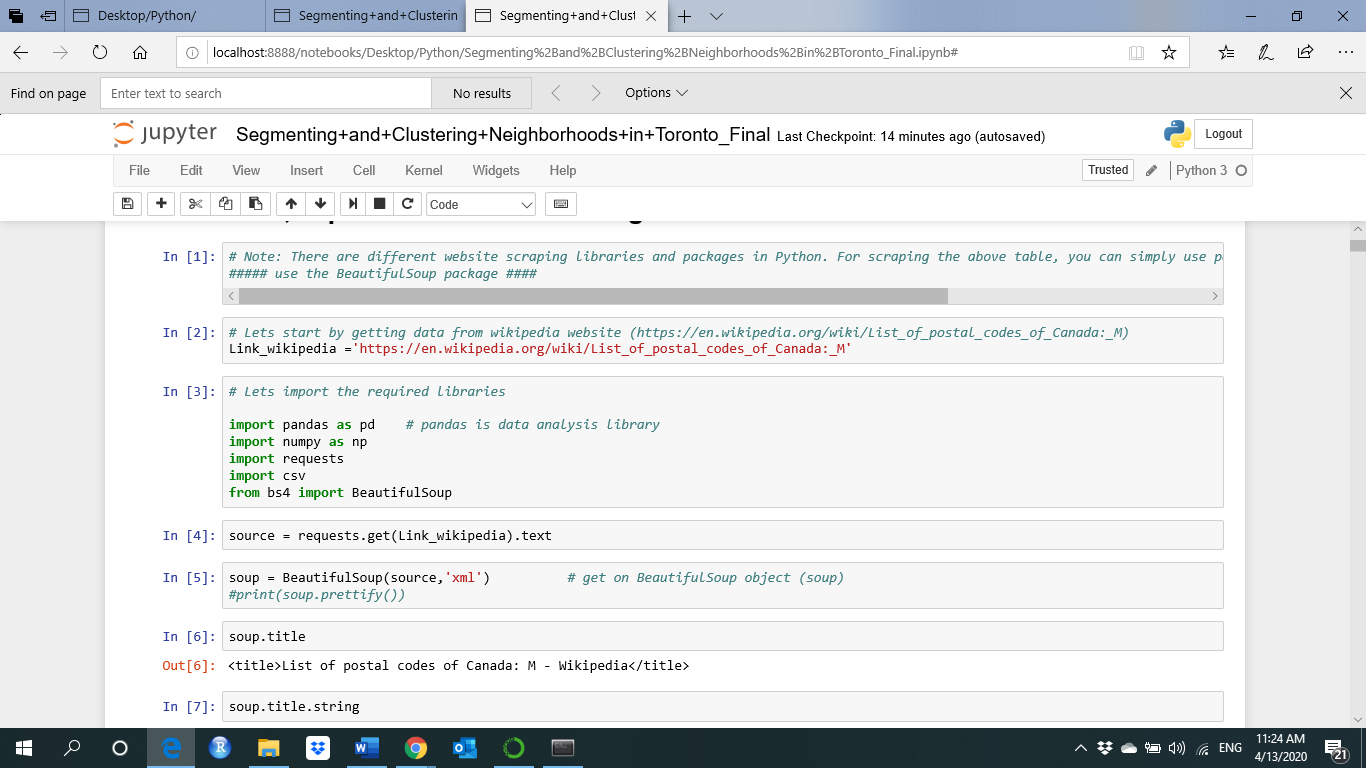
Finally, using the Foursquare API ((https://developer.foursquare.com/docs) that gives venue information such as names, categories, and locations (latitude and longitude).

**Data Analysis Process**

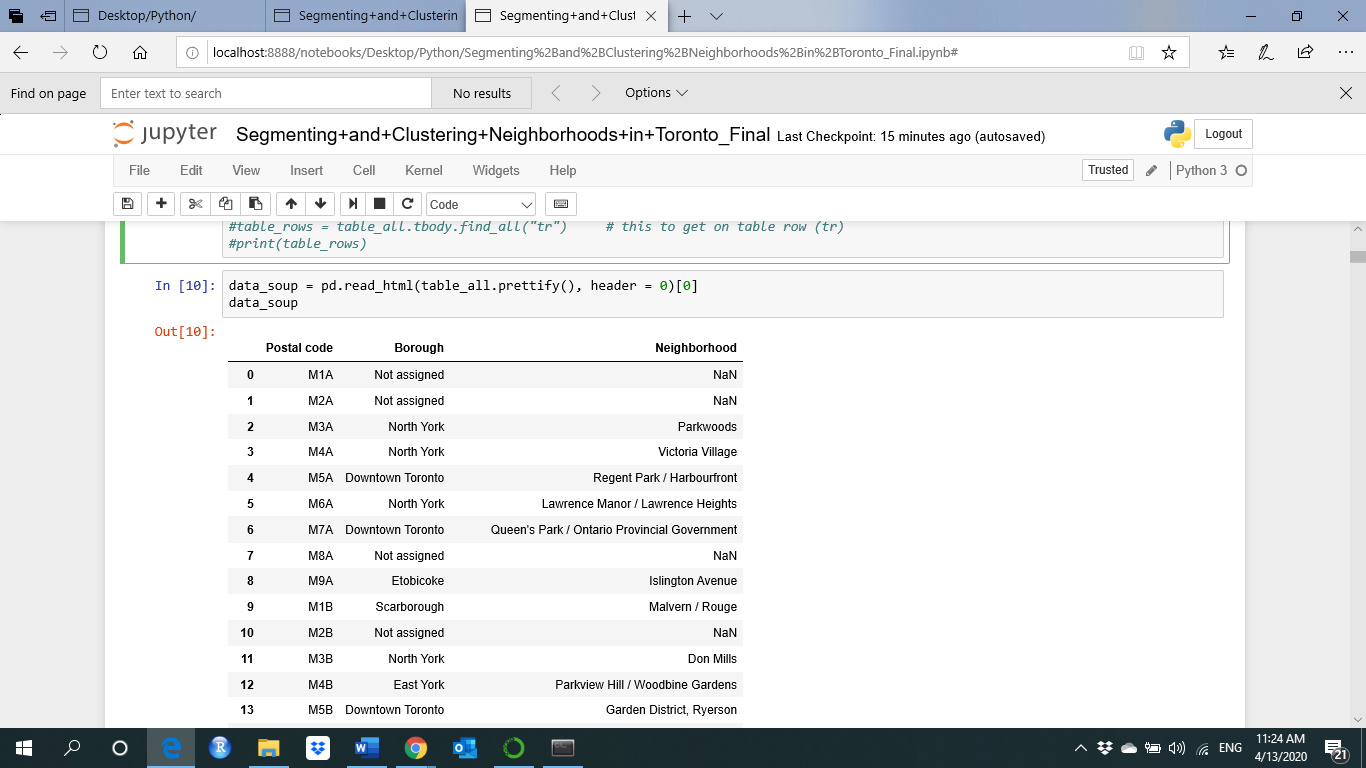
**Diagram of Analysis Process**

**Dara Preprocessing Stage**

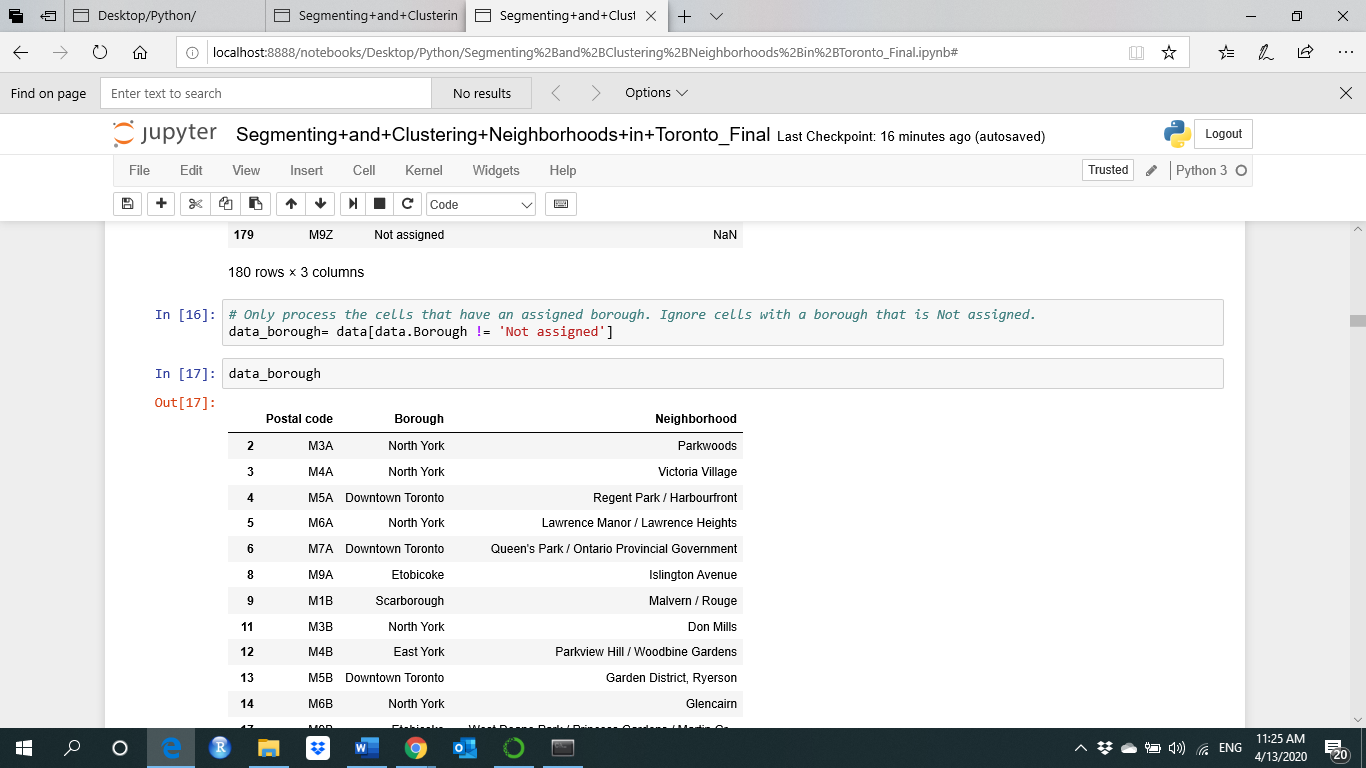
Below show the data preprocessing steps and some of the results obtain from the analysis work. Scraping Toronto Neighborhoods Table from Wikipedia website to start working on the data



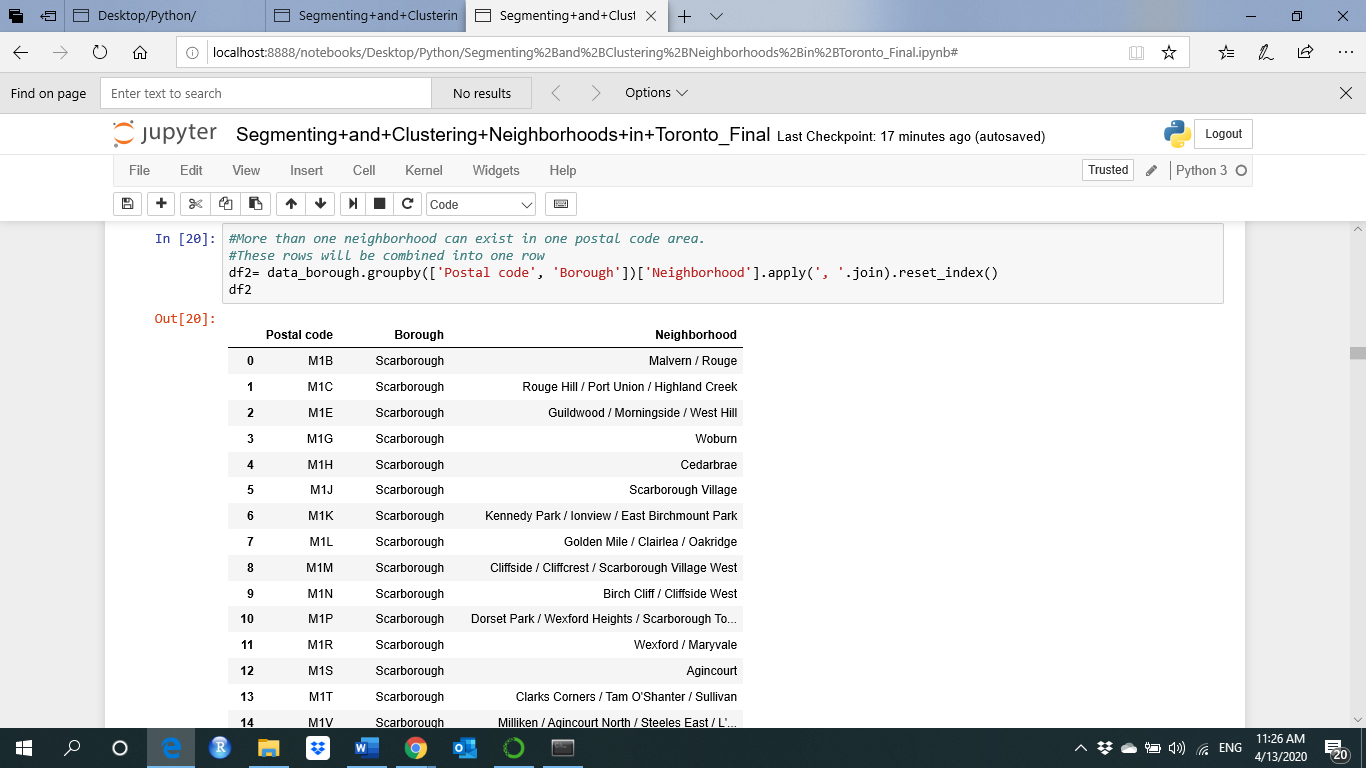
Get on the Toronto neighborhoods table after scraping the data from Wikipedia website



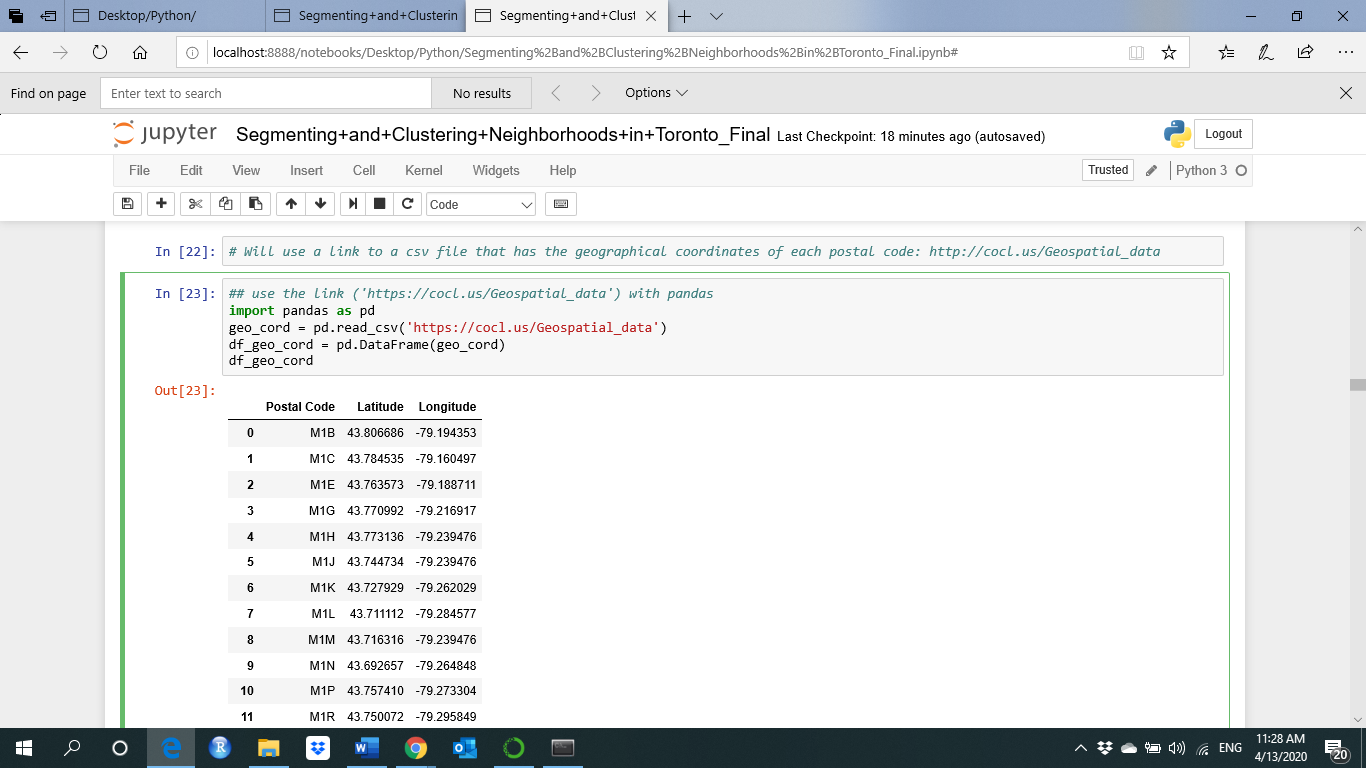
Only process the cells that have an assigned borough. Ignore cells with a borough that is Not assigned as given in the below result.



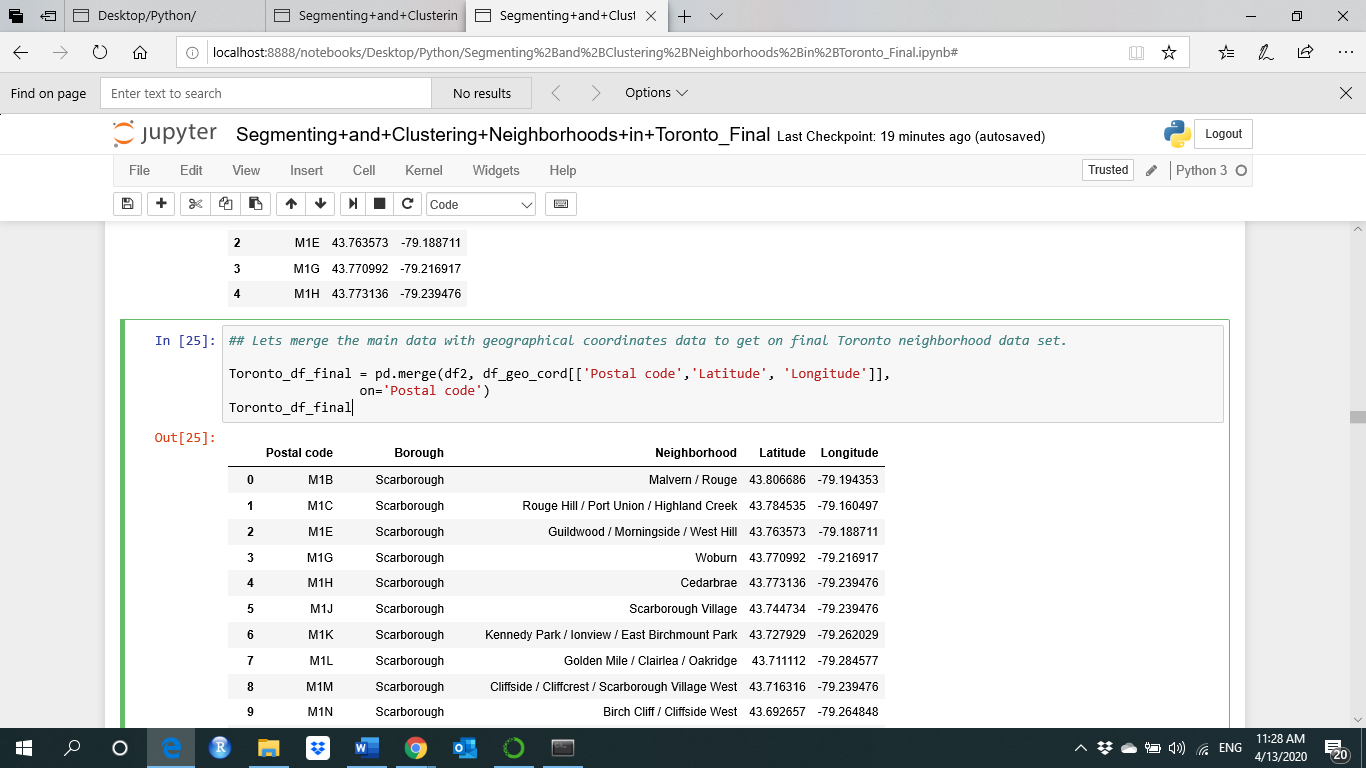
More than one neighborhood can exist in one postal code area. These rows will be combined into one row as shown below.



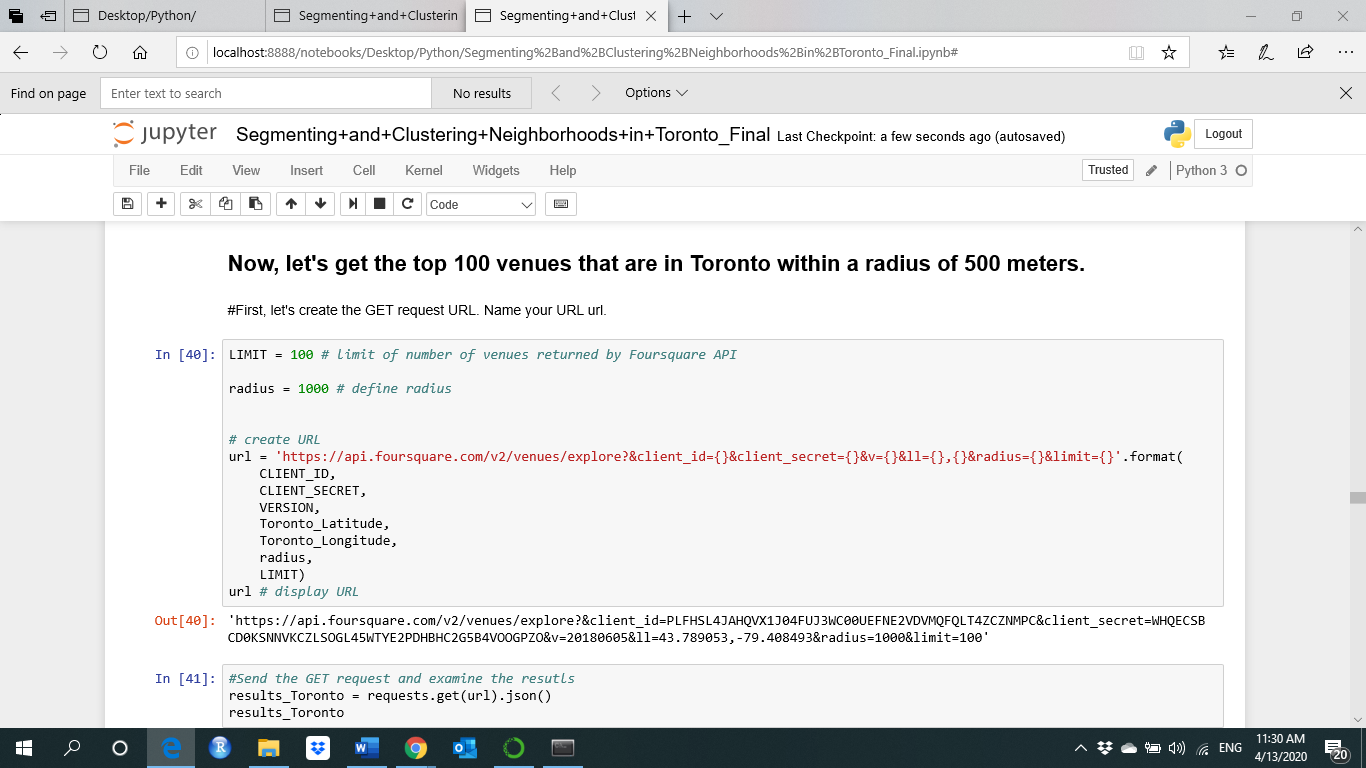
Now, use the Geocoder package or the CSV file to create a data frame with latitude and the longitude coordinates of each neighborhood as given in the result below.

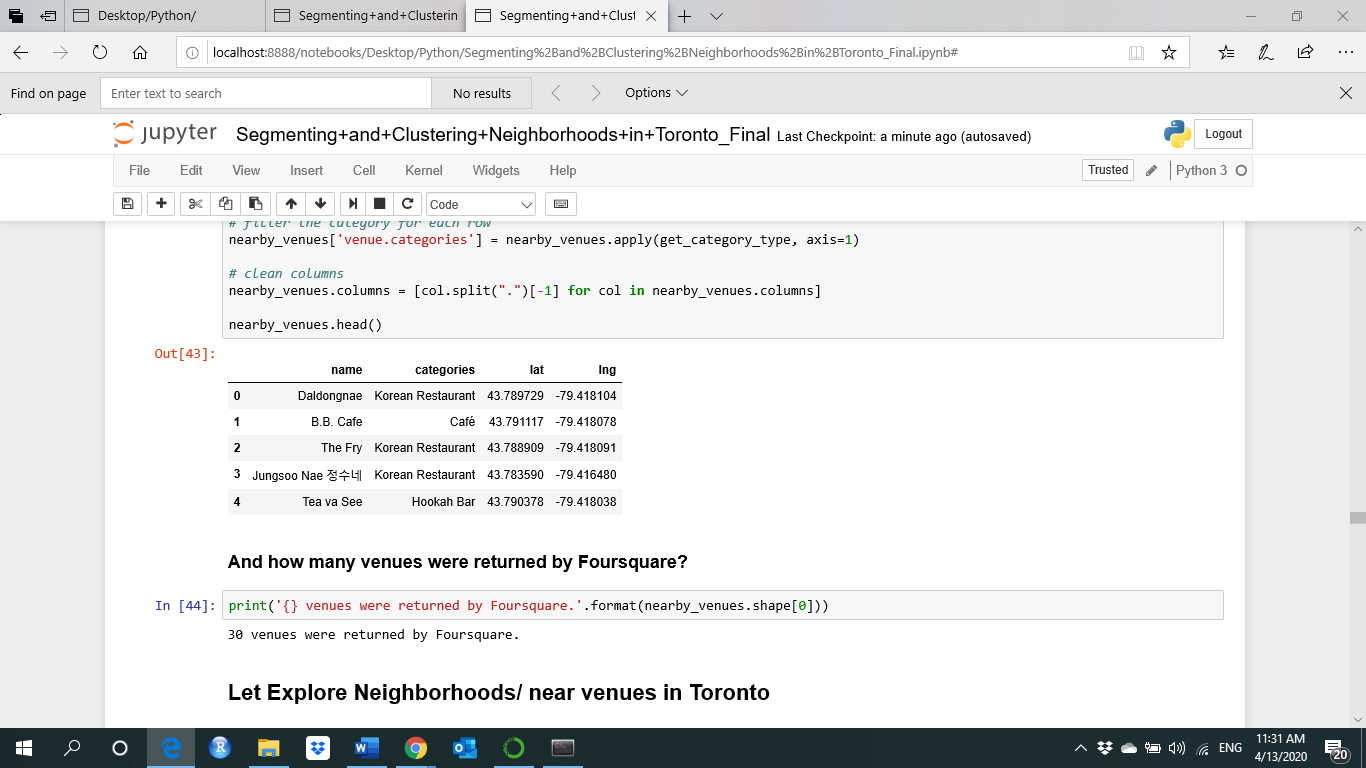


Let's merge the main data with geographical coordinates data to get on the final Toronto neighborhood data set as below.

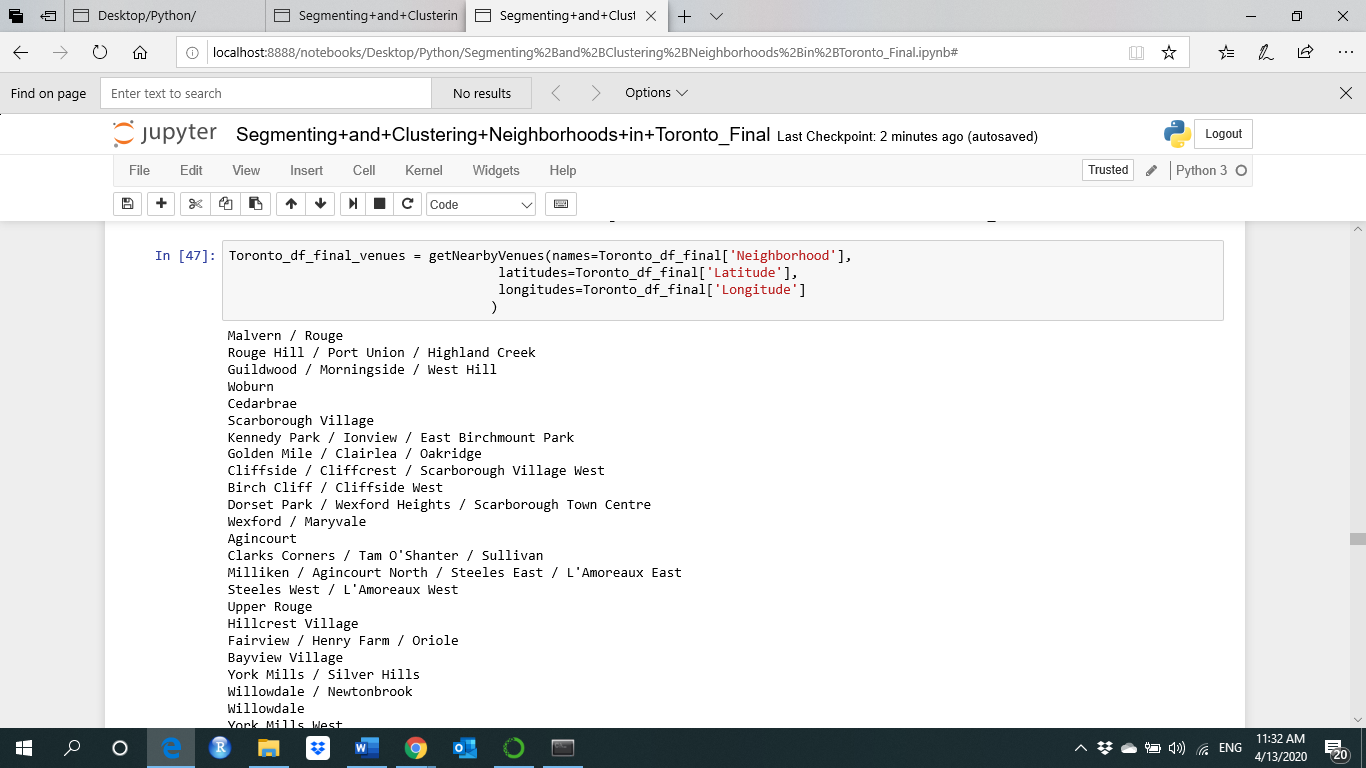


Now, let's get the top 100 venues that are in Toronto within a radius of 1000 meters as below.

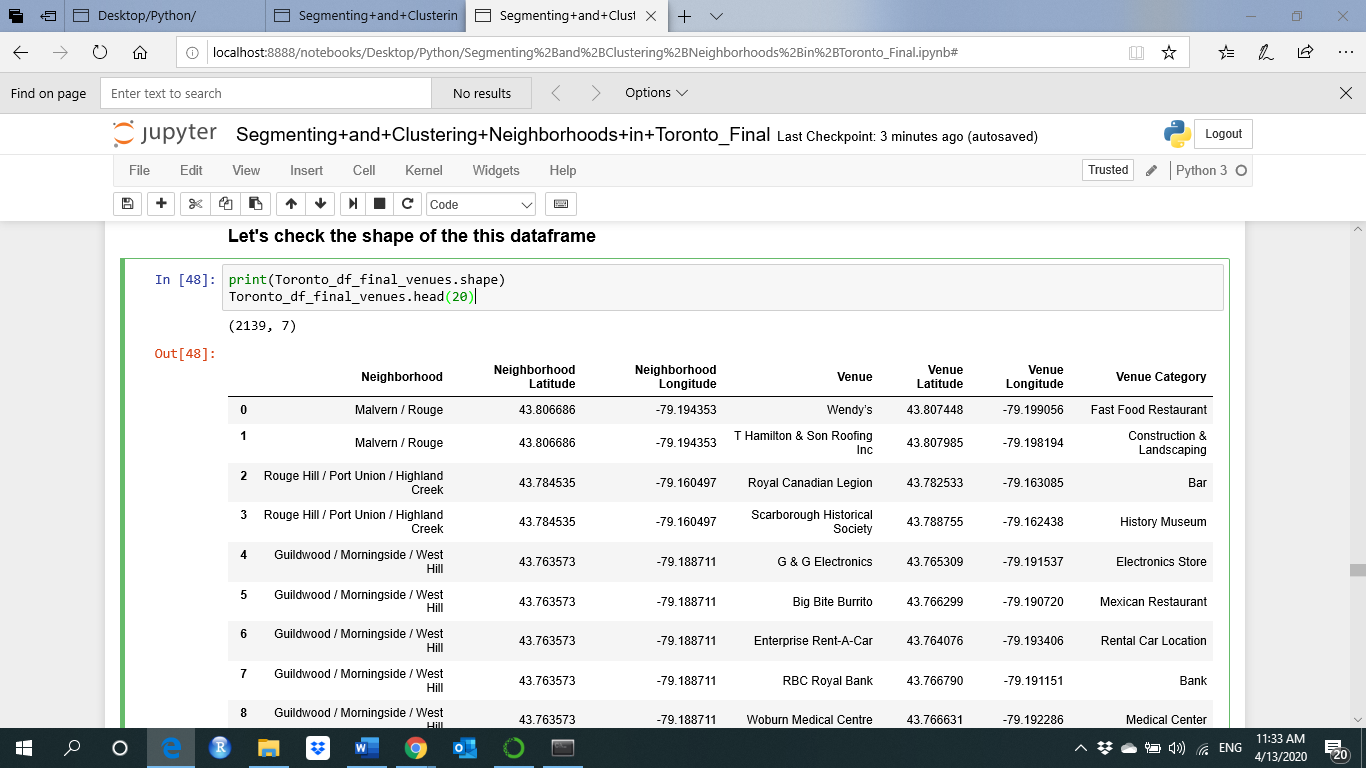




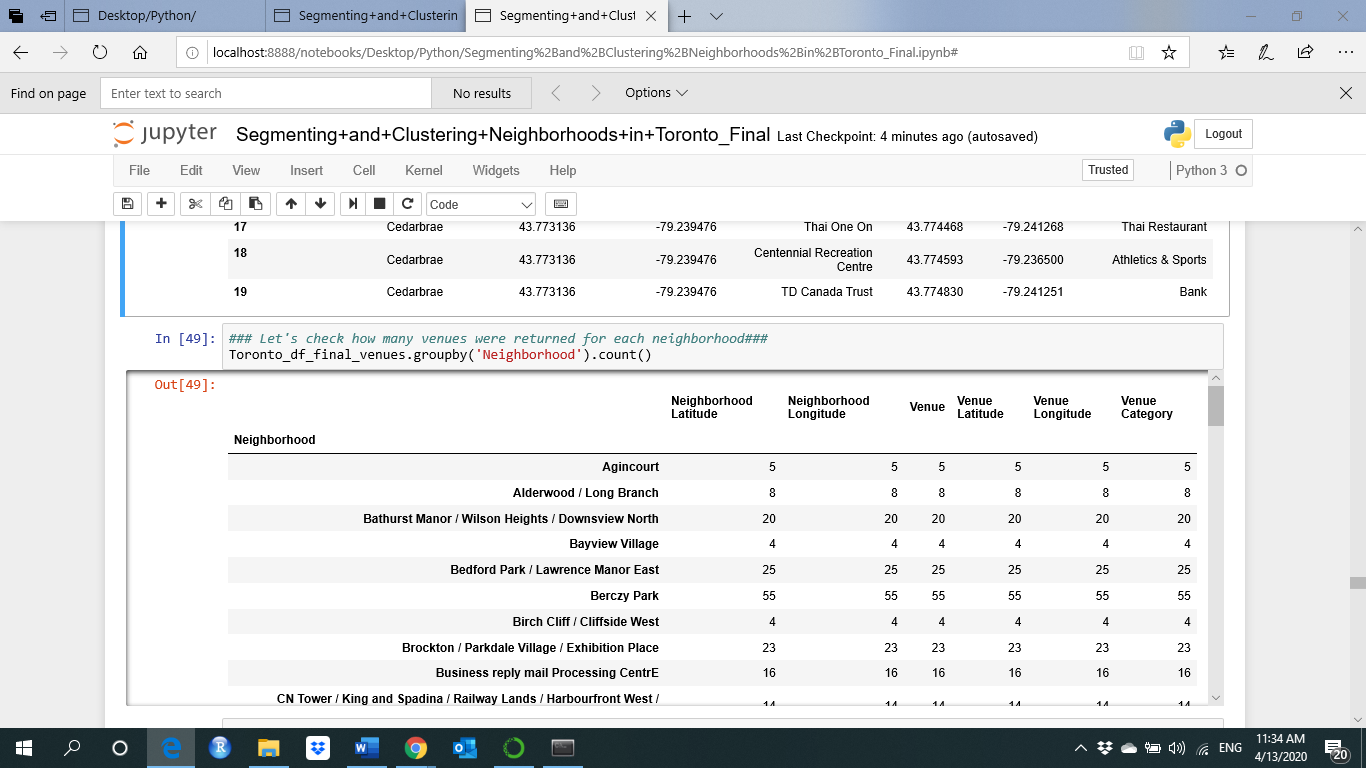
To get nearby venues in Toronto as shown



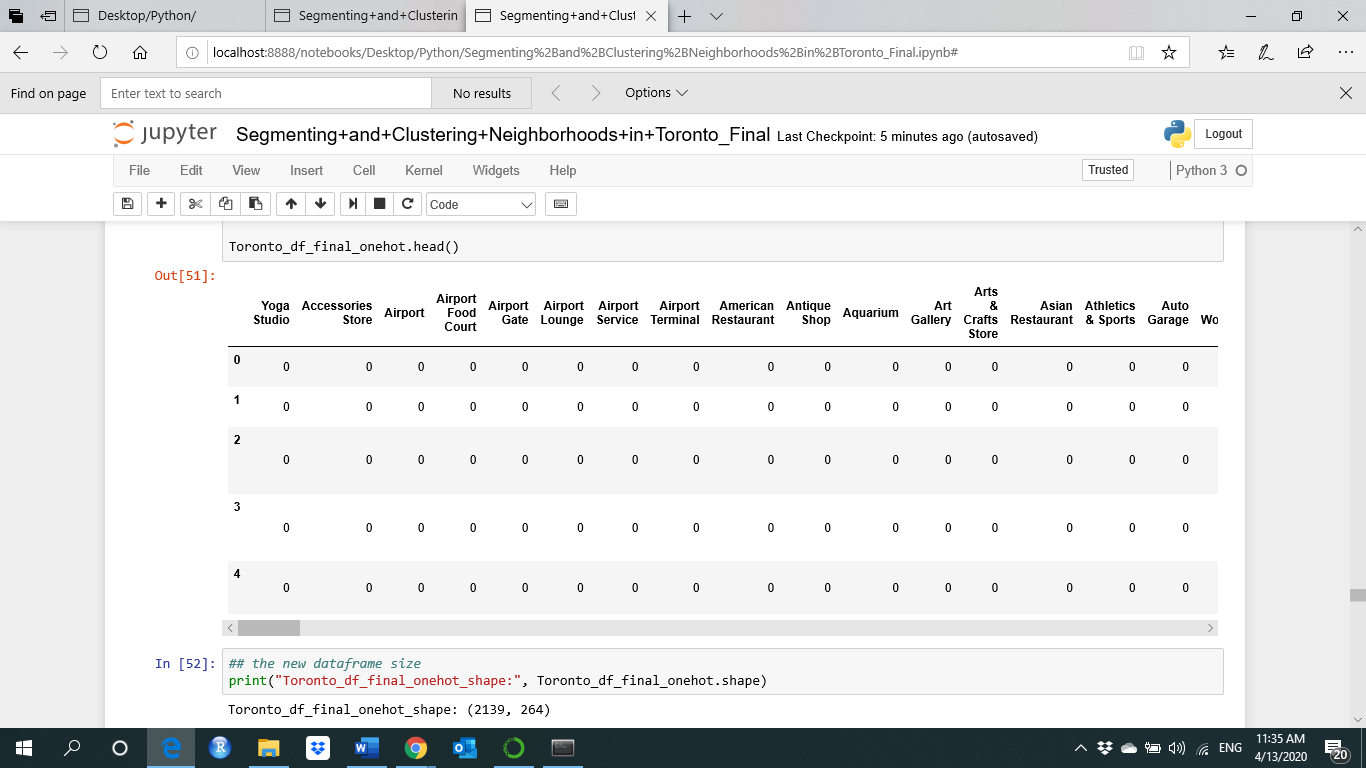
The Neighborhood name and Neighborhood Latitude, Neighborhood Longitude with the corresponding Venue and Venue Category as display below.



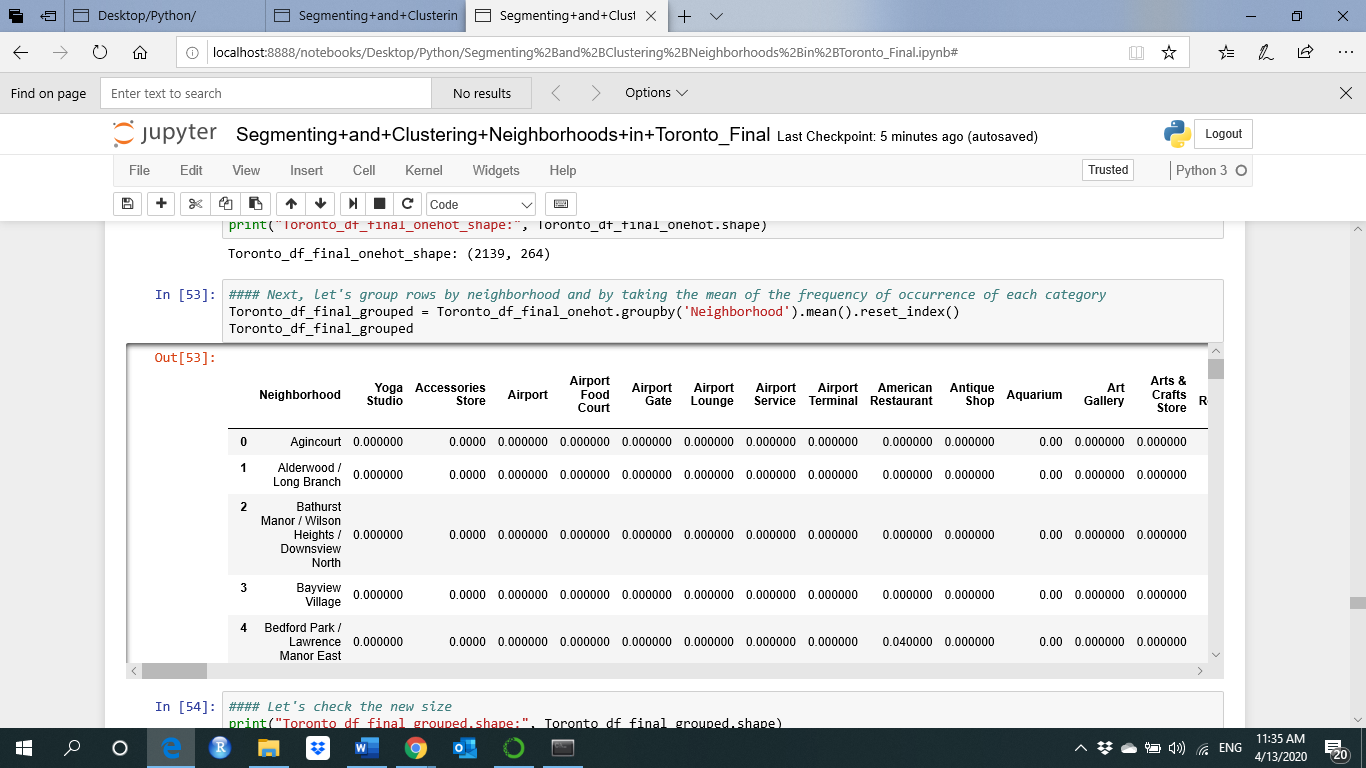
below show how many venues were returned for each neighborhood.



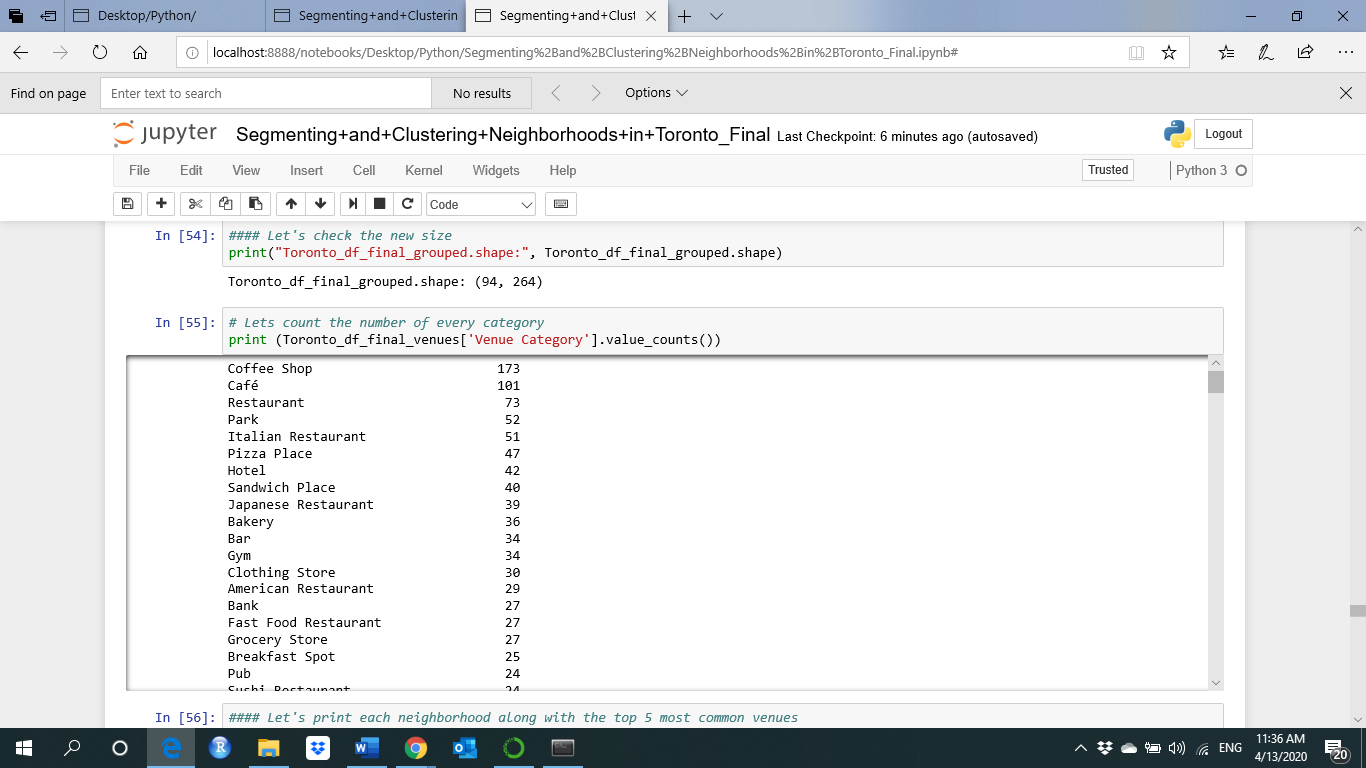
There are more than 262 unique categories and it should be one of them are Middle east/Arab restaurant as shown:



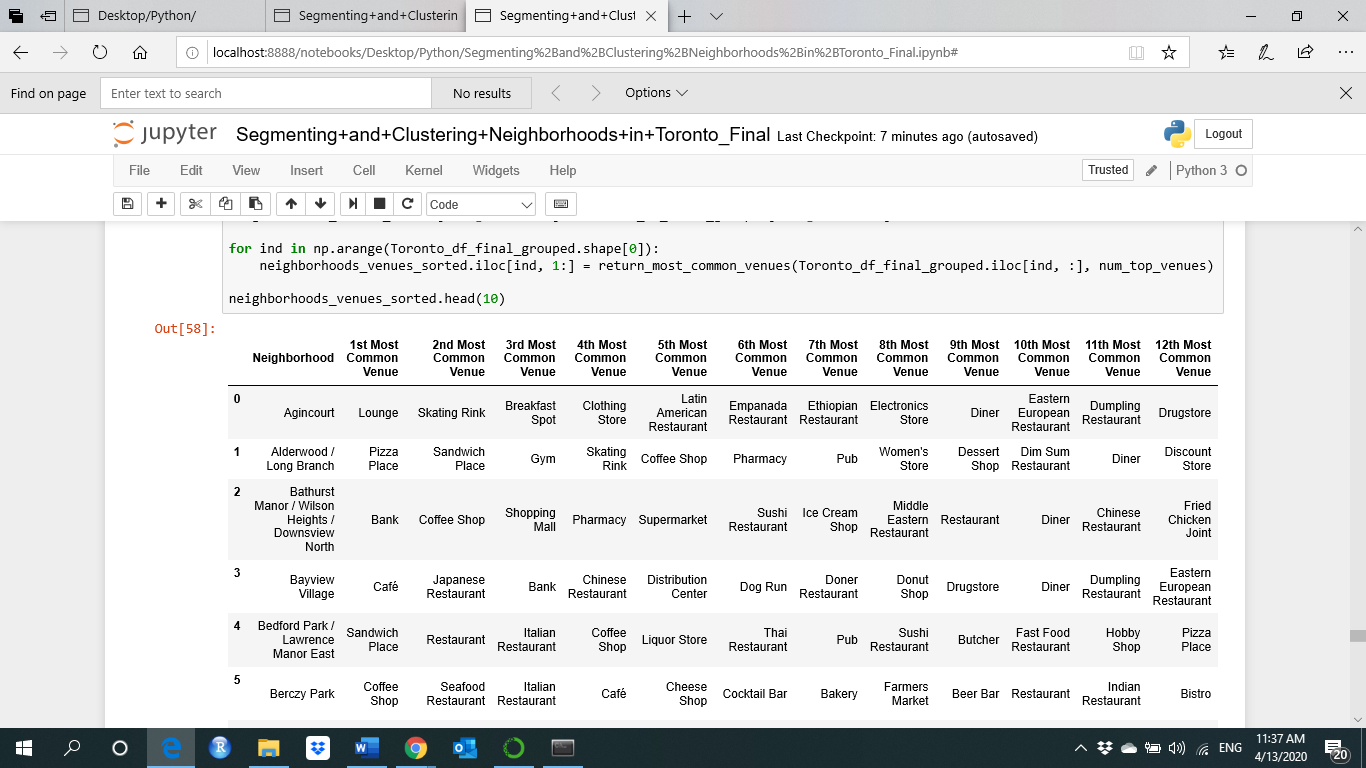
Below grouping the rows by neighborhood and by taking the mean of the frequency of occurrence of each category



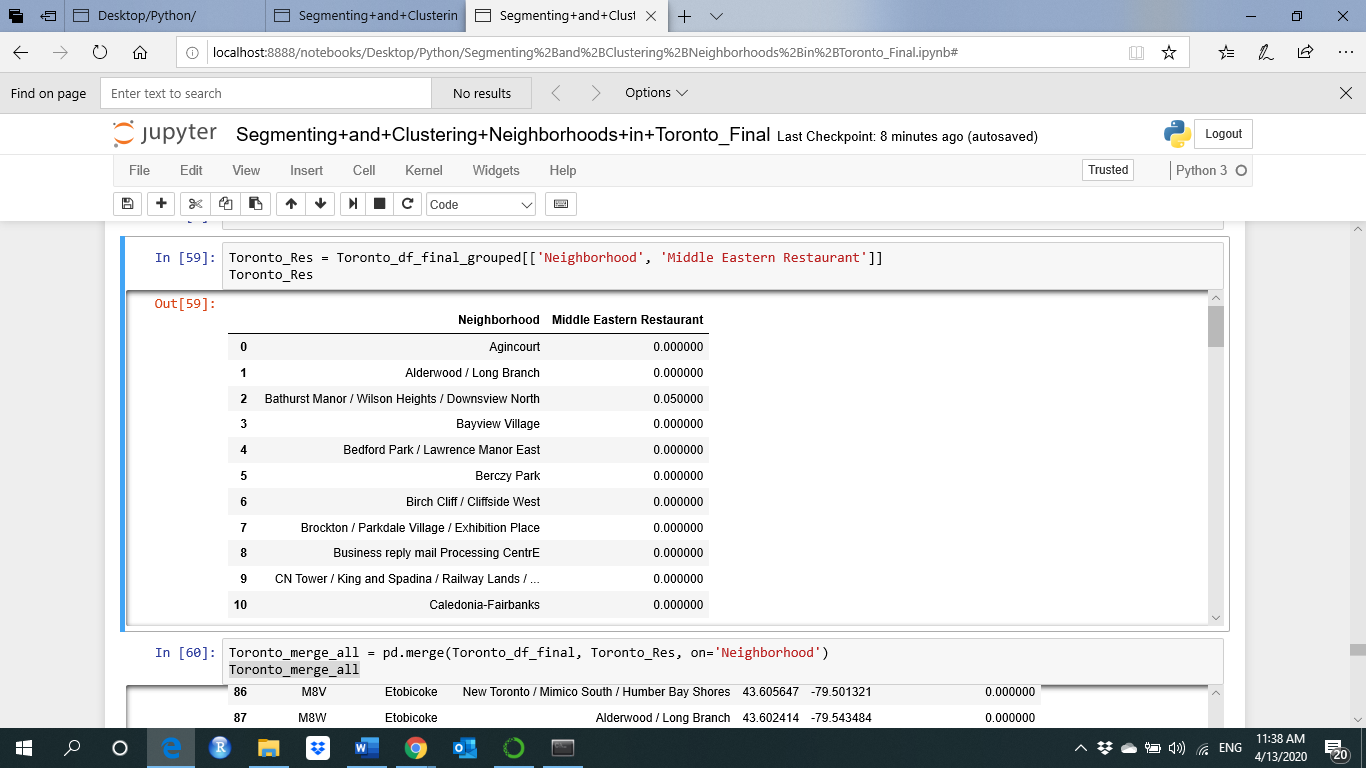
The count of the number of every category as shown below



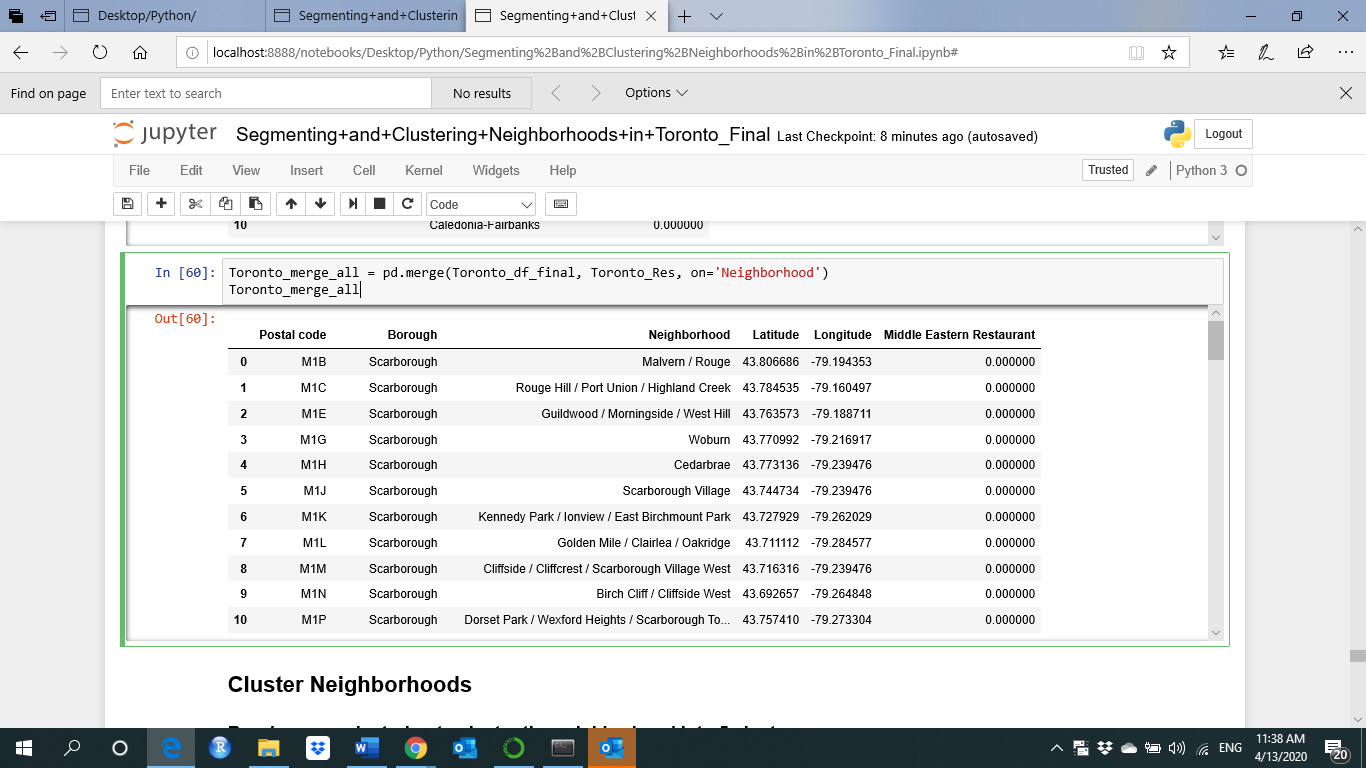
The below results show a new data frame that displays the top 20 venues for each neighborhood.



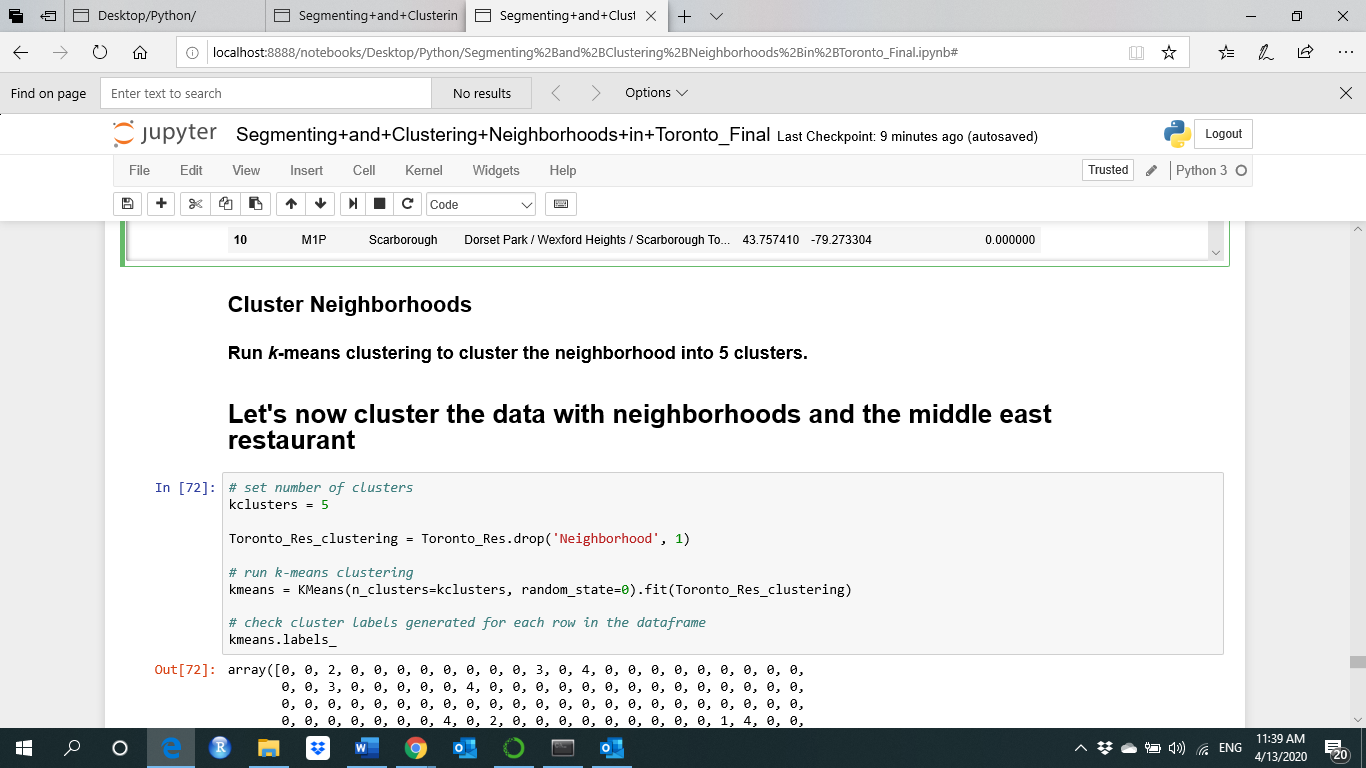
Each Neighborhood with the Middle Eastern Restaurant shown below.



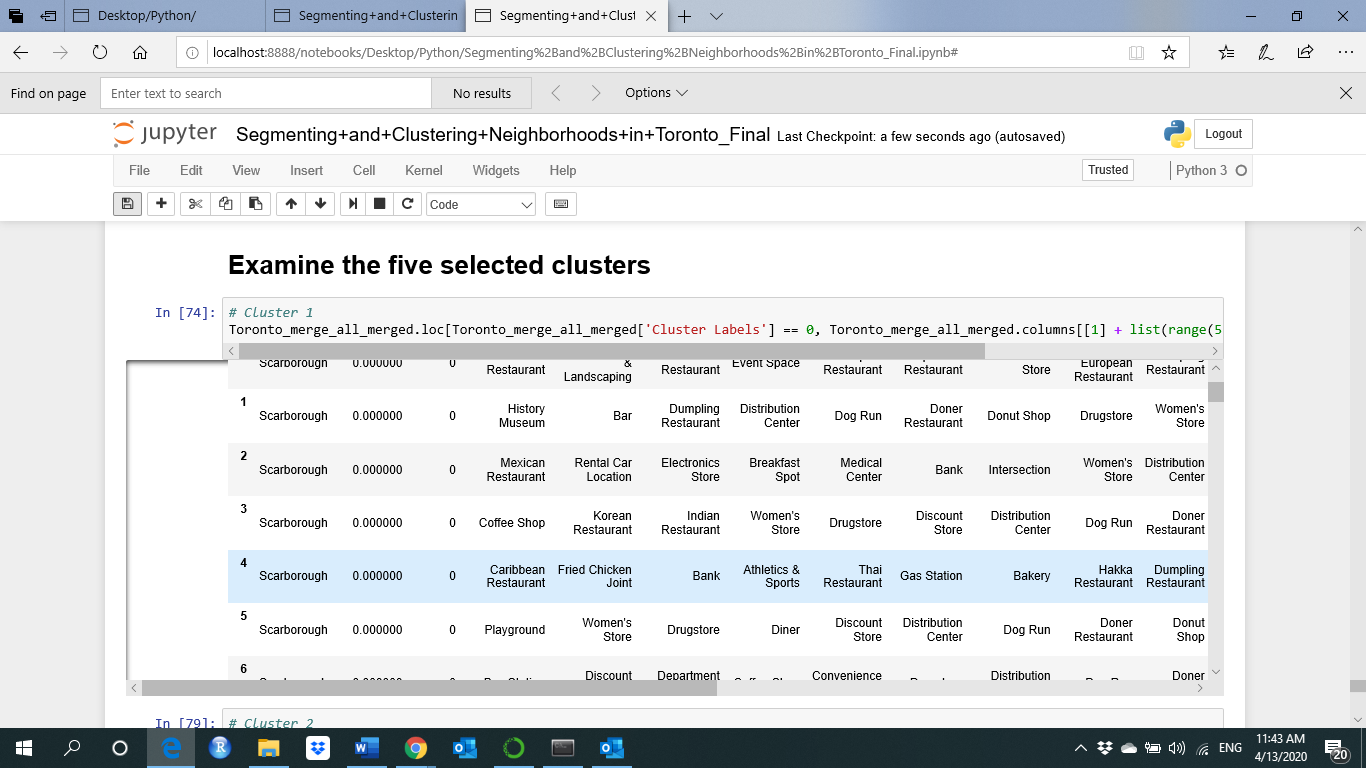
Merge the borough and the neighborhood with the corresponding latitude and longitude and the Middle Eastern Restaurant shown below.



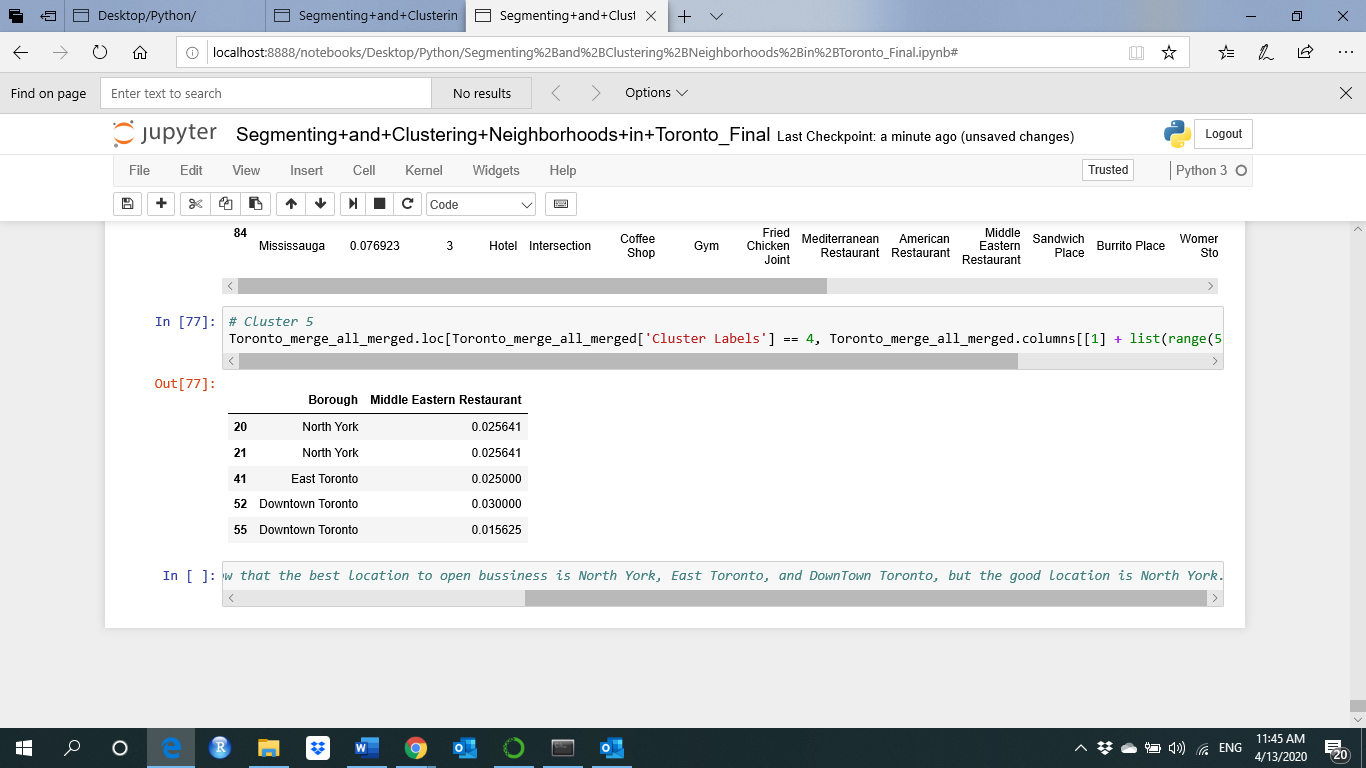
Clustering the data



Let Examine some of the selected clusters, first one is cluster number 1



Another example, the last one (cluster number = 5)



**Results**

Some of the obtained results as shown in the above section. We can see from the results, and for example that Cluster 5 contains all the neighborhoods which are highly populated with Middle Eastern Restaurants such as North York, East Toronto, and Downtown Toronto, and some other neighborhood, so its good idea not to make the business in these neighborhoods and build it in another place such as Scarborough or Mississauga because it contains a high population.

**Discussion and Conclusions**

In this project, we used real data to make an analysis to solve a business problem. Several python libraries and machine learning algorithms used to solve the problem.

In this first analysis and primary investigation, I think and from the first try that the obtained results show that it is good to start the middle east and Arab food market and restaurant in the Scarborough or Mississauga neighborhoods and far away from North York, East Toronto, and Downtown Toronto.

In future work, we will make more study and analysis on the city of Toronto i.e. by applying the same process with deeper analysis to the educational institutes and universities such as the University of Toronto and its good to open this business close to this university or others by checking the international student's population number because a lot of international students from middle east and Arab world studying and living in Toronto. Not just for the city of Toronto and maybe go deeper analysis to all the Ontario province with more usage of the population distribution data and to get on more accurate results.

References

1. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
2. <http://cocl.us/Geospatial_data>
3. <https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity>
4. <https://developer.foursquare.com/docs>