Capstone Project - Clustering Neighborhoods in Downtown Toronto

**The Battle of Neighborhoods (Week 2)**

By Sayed Adel

20 Aug 2020

**Introduction**

Toronto is the capital city of the [Canadian province](https://en.wikipedia.org/wiki/Provinces_and_territories_of_Canada) of [Ontario](https://en.wikipedia.org/wiki/Ontario). With a recorded population of 2,731,571 in 2016, it is the [most populous city in Canada](https://en.wikipedia.org/wiki/List_of_the_100_largest_municipalities_in_Canada_by_population) and the [fourth most populous city in North America](https://en.wikipedia.org/wiki/List_of_North_American_cities_by_population). The city is the anchor of the [Golden Horseshoe](https://en.wikipedia.org/wiki/Golden_Horseshoe), an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of [Lake Ontario](https://en.wikipedia.org/wiki/Lake_Ontario), while the [Greater Toronto Area](https://en.wikipedia.org/wiki/Greater_Toronto_Area) (GTA) proper had a 2016 population of 6,417,516. Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most [multicultural](https://en.wikipedia.org/wiki/Multicultural) and [cosmopolitan](https://en.wikipedia.org/wiki/Cosmopolitanism) cities in the world.

Downtown Toronto is the main [central business district](https://en.wikipedia.org/wiki/Central_business_district) of [Toronto](https://en.wikipedia.org/wiki/Toronto), [Ontario](https://en.wikipedia.org/wiki/Ontario), Canada. Located entirely within the district of [Old Toronto](https://en.wikipedia.org/wiki/Old_Toronto), it is approximately 17 square kilometers in area, bounded by [Bloor Street](https://en.wikipedia.org/wiki/Bloor_Street) to the northeast and Dupont Street to the northwest, [Lake Ontario](https://en.wikipedia.org/wiki/Lake_Ontario) to the south, the [Don Valley](https://en.wikipedia.org/wiki/Don_Valley_River) to the east, and [Bathurst Street](https://en.wikipedia.org/wiki/Bathurst_Street,_Toronto) to the west. It is also the location of the [municipal government of Toronto](https://en.wikipedia.org/wiki/Municipal_government_of_Toronto) and the [Government of Ontario](https://en.wikipedia.org/wiki/Government_of_Ontario).

The area is made up of Canada's largest concentration of skyscrapers and businesses that form Toronto's skyline. Downtown Toronto has the third most skyscrapers in North America exceeding 200 meters (656 ft) in height, behind [New York City](https://en.wikipedia.org/wiki/New_York_City) and [Chicago](https://en.wikipedia.org/wiki/Chicago).

Having that said, opening businesses in big cities is of course a good investment but it would be a bit risky to go directly and open any kind of business without some analysis to take the right decision and to choose the right place and the right kind of business where I will focus on opening a restaurant business in Toronto

I chose Downtown Toronto Borough within Toronto to explore its neighborhoods where I’m going in this project to provide some insight for those who are interested in opening a restaurant business in Downtown Toronto by choosing the best neighborhood(s) and best type of restaurants, in order to increase the probability that the business will succeed

**Data Description**

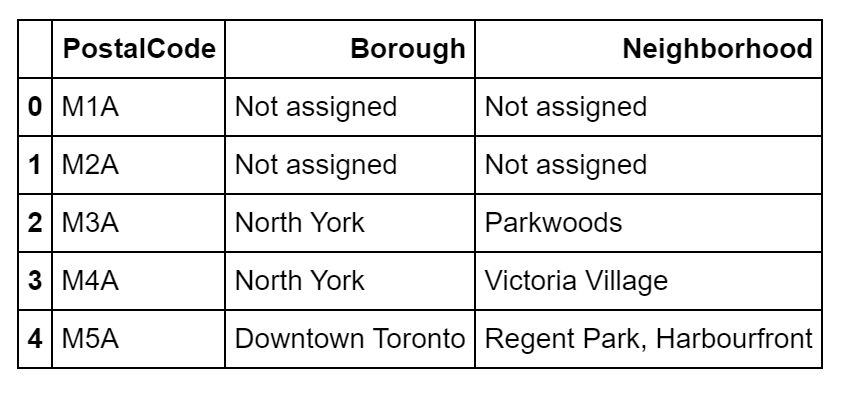
For the purpose of this project I used the below data sources:

* Information about Toronto from <https://en.wikipedia.org/wiki/Toronto>
* Information about Downtown Toronto from <https://en.wikipedia.org/wiki/Downtown_Toronto>
* List of Borough and Neighborhoods by postal code from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* Geographical Coordinates of each postal code from [http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data)
* **Foursquare API** to get the most common venues for the neighborhoods in Downtown Toronto

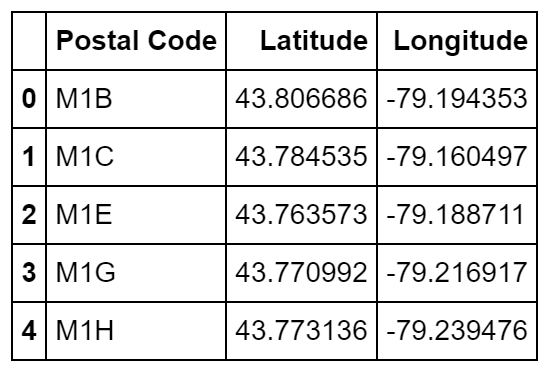
**Methodology**

I used GitHub as my version control and repository environment and also Jupyter Notebook to write my code and conduct my analysis. I’m going to mention other methodologies and libraries as long as I describe my work here. Let’s start the journey!

First of all, I got Toronto’s Boroughs and Neighborhoods by Postal Code from the mentioned page in the data section from Wikipedia and used Pandas Library to manipulate the data



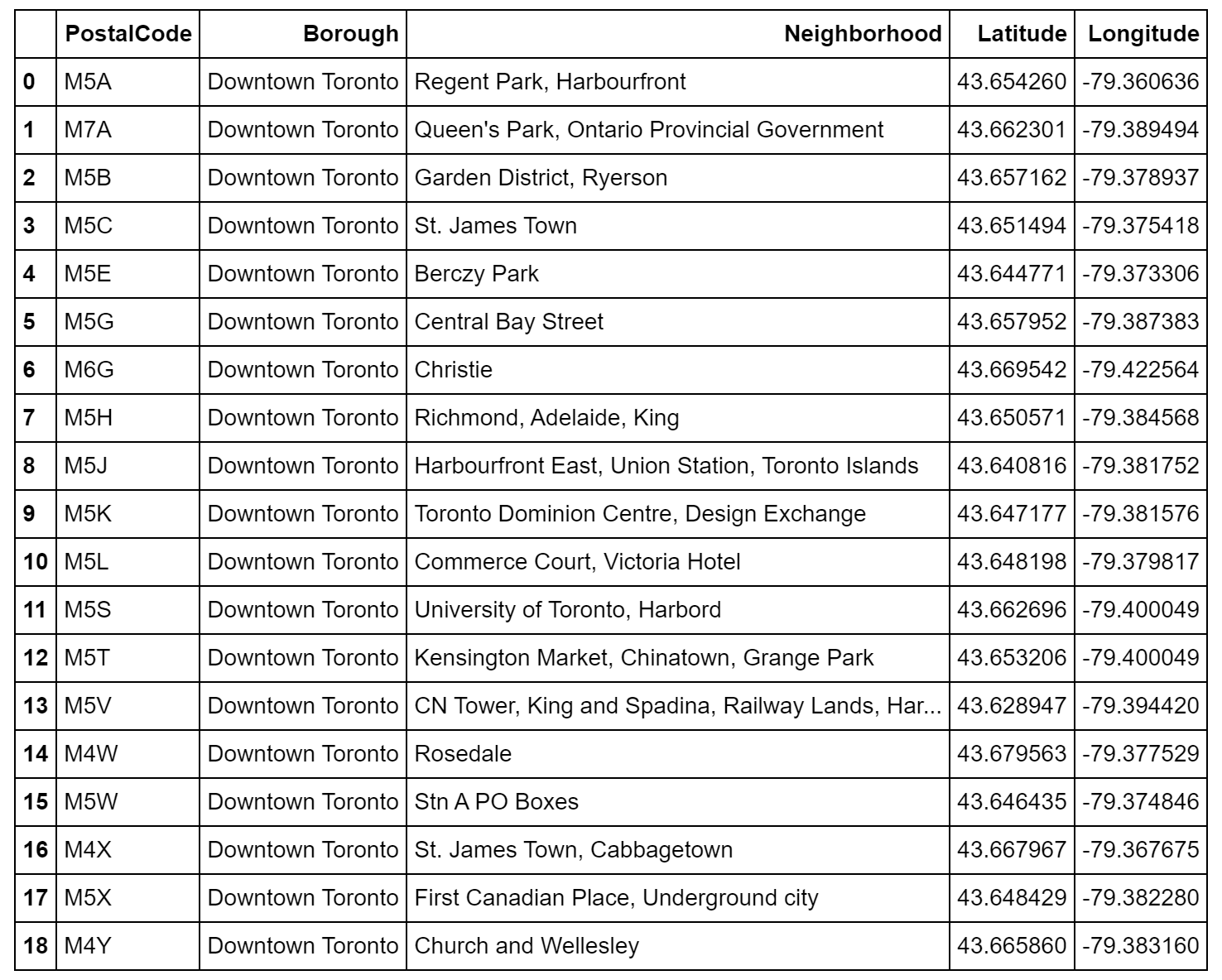
Then I cleaned the data by removing the “Not assigned” entries and imported the Geographical Coordinates of each postal code from [http://cocl.us/Geospatial\_data](https://cocl.us/Geospatial_data)



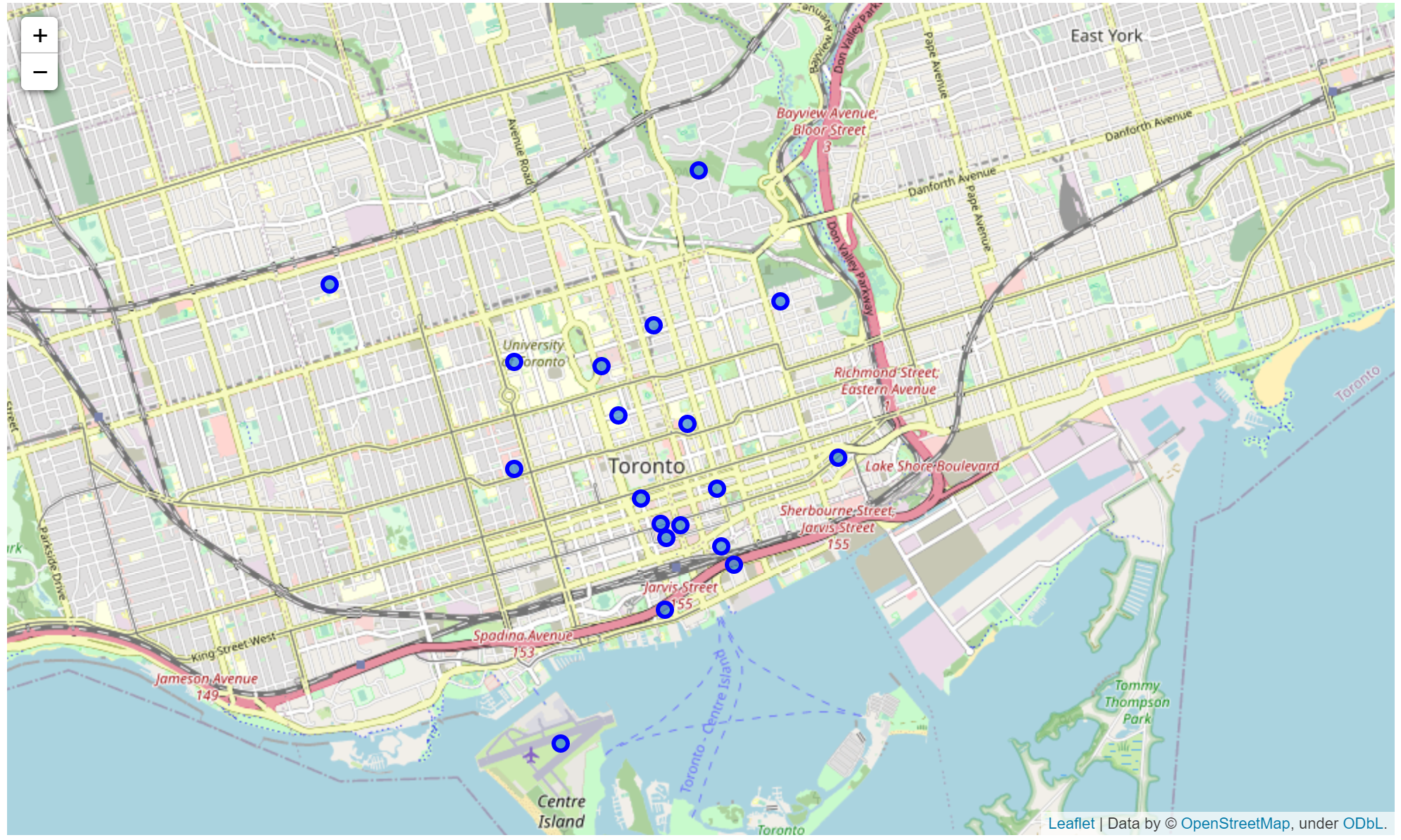
Then I merged both data frames, below is the result



The data frame consists of 10 Boroughs and 103 Neighborhoods. Then I removed all other Boroughs and kept “Downtown Toronto” only



As we can see it resulted in 19 rows of Neighborhoods by postal codes. I used Folium library to create a map of Downtown Toronto with neighborhoods superimposed on top

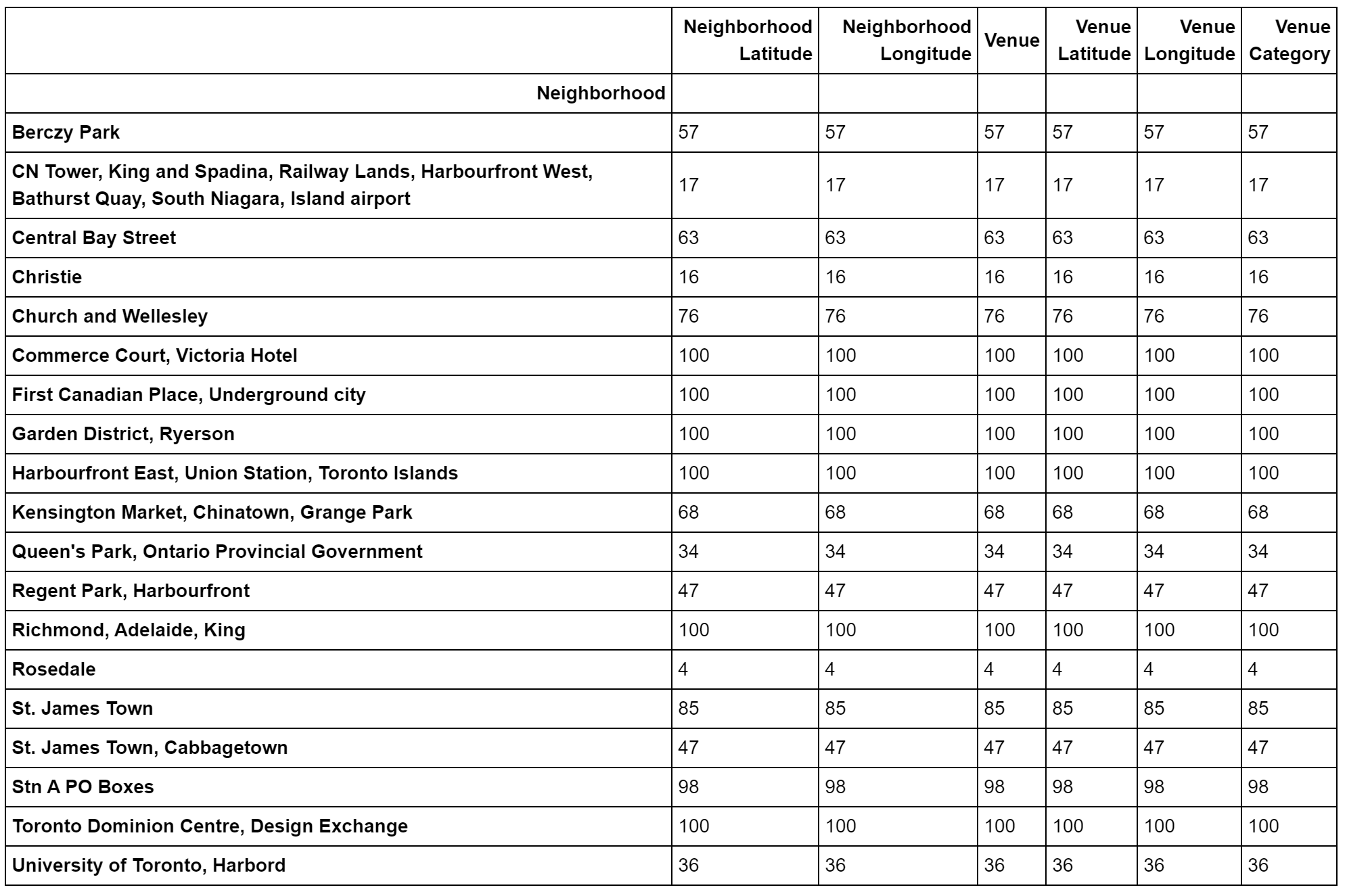


Now we are going to use Foursquare API to get the top 100 venues in Downtown Toronto’s Neighborhoods within a radius of 1000 meters



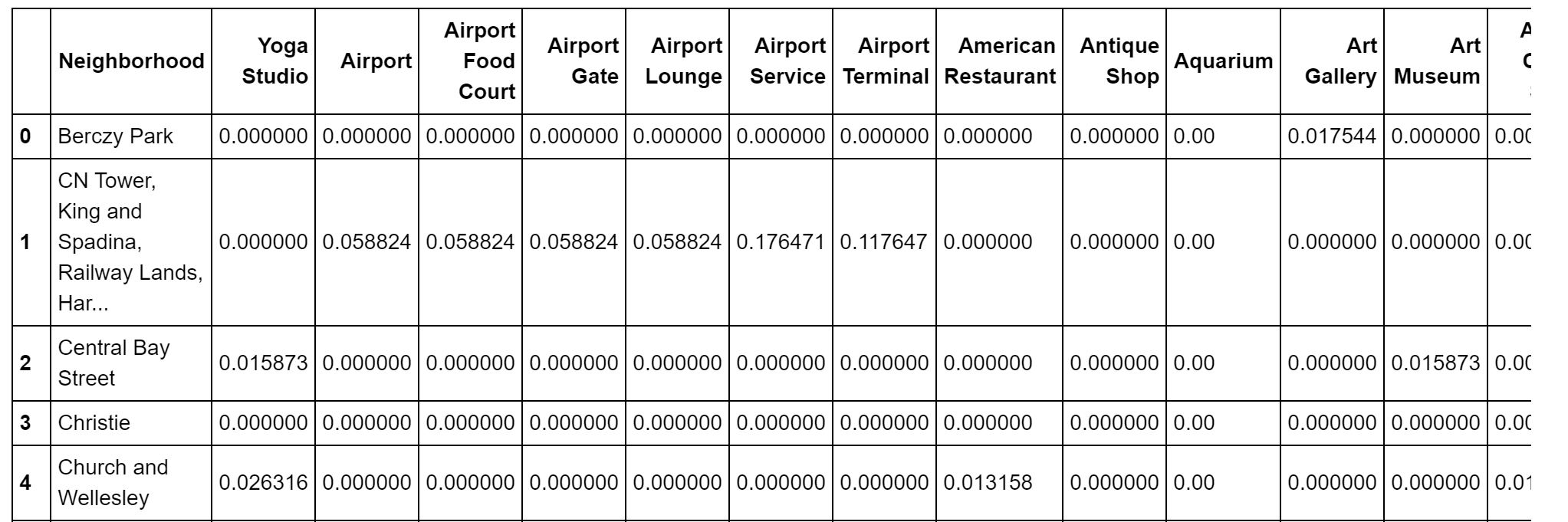
As we can see we found 1248 venues

Checking how many venues were returned for each neighborhood



We found that there are 207 unique categories can be curated from all the returned venues.

For analyzing the Neighborhoods, I grouped the rows by neighborhood and by taking the mean of the frequency of occurrence of each category



Now we can print the 7 most common venues for each Neighborhood’s group in order to have an idea about the mot common venues and which Neighborhoods have most of the restaurants as long as the type of the restaurants



Let’s now print the 10 most common venues for each Neighborhood’s group

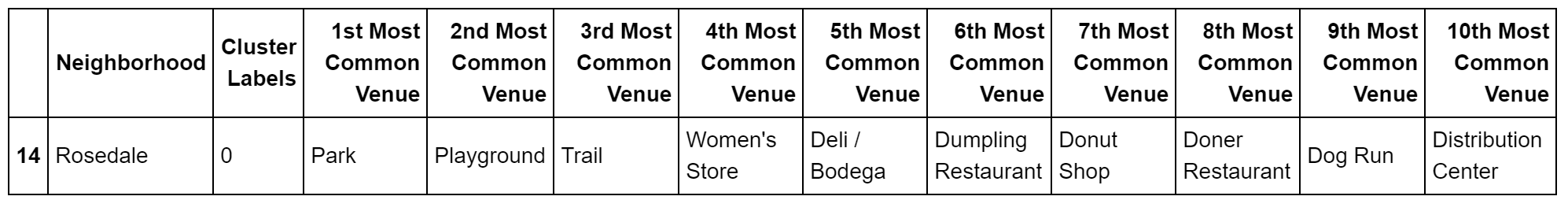


The next step would be clustering and showing the map. For clustering I used the K-means algorithm for clustering the Neighborhoods as it’s a well know algorithm for unsupervised Machine Learning model for clustering. Below the map shows 5 clusters



Let’s now examine the clusters and give them proper names

Cluster 1: Entertainment Venues



Cluster 2: Intensive Cafe Venues



Cluster 3: Versatile Venues



Cluster 4: Transportation Venues



Cluster 5: Restaurants Venues



**Results**

From the above clustered map we can see the distribution of the 5 clusters and from the examination of the clusters in the above tables we can have an idea of the common venues in the 5 clusters where we named them according to our observation, Entertainment Venues Cluster, Intensive Cafe Venues Cluster, Versatile Venues Cluster, Transportation Venues Cluster and the Restaurants Venues Cluster

**Discussion**

As I mentioned in the introduction Toronto is a big city as well as its Downtown Toronto Borough and starting any business aimed to succeed would be challenging as if you open something new or of a real need would mostly lead the stakeholders to excel in their business.

Along they journey in this project I got the data from different resources and prepared them to produce 5 clusters of Downtown Toronto to predict the best places to open new restaurants, as well as what type of restaurants. I used Pandas, NumPy, folium libraries and K-Means unsupervised clustering algorithm for my analysis and claustration.

When we look at the results, we can see that there are many Italian, Japanese, American restaurants as an example in many Neighborhoods; therefore, starting a *French or Lebanese* restaurant would be a good choice or any restaurant in Cluster 1 and 4

**Conclusion**

It’s normal that people aim at big cities and places where there is a big potential for a successful business but before that they need to do some analysis and conduct a feasibility studies in order to start their business on a strong basis that leads the to success and good profitability.

We saw in the above pages how using data analysis and Machine Learning is a powerful tool to provide insights and knowledge in any discipline that leads to a wise decision

**References**

[1] <https://en.wikipedia.org/wiki/Toronto>

[2] https://en.wikipedia.org/wiki/Downtown\_Toronto

[3] <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

[4] <http://cocl.us/Geospatial_data>

[5] Foursquare API, Foursquare.com