The Battle of Neighborhoods

To find similarities in neighborhoods of New York and Toronto

Date : 6th Feb 2021

1. **Introduction**

New York and Toronto are two extremely popular cities in their respective countries. In this project we will highlight the borough and neighborhoods of these respective cities , obtain graphical data of the venues around to have an idea as to the similarity in terms of venues and hence the locality by itself.

* 1. **Business focus**

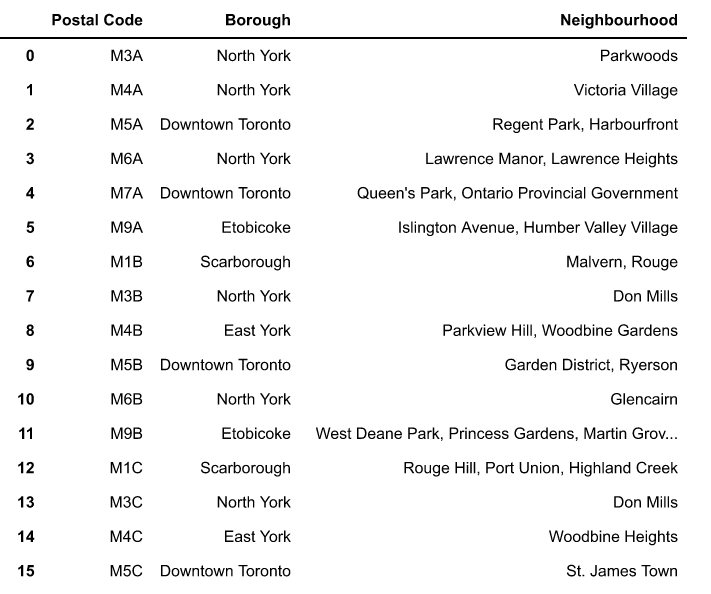
We can figure out the neighborhoods in these cities. We will figure out the similarity in terms of venues. We will get to know the frequency of their usage to give us an idea as to what kind of business are really popular in their neck of the woods.

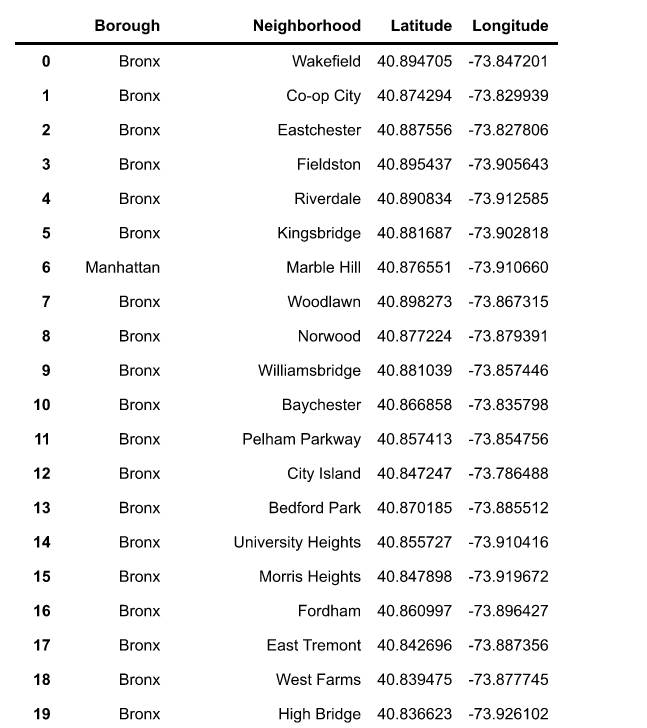
1. **Data**

Our sources of data are obtained from internet (Wikipedia, websites). We are also using the Foursquare map services to explore and cluster our neighborhoods.

The data is obtained from (Wikipedia for Toronto and a json file for NY from the coursework). This data is further cleaned to exclude any ‘not assigned’ rows, merge same postal codes with different neighborhoods and so on. Further We will prepare a data frame to analyse by means of bar graph. Also explore the clusters around and prepare maps accordingly.

* 1. **Toronto and NY with their boroughs and neighborhoods**

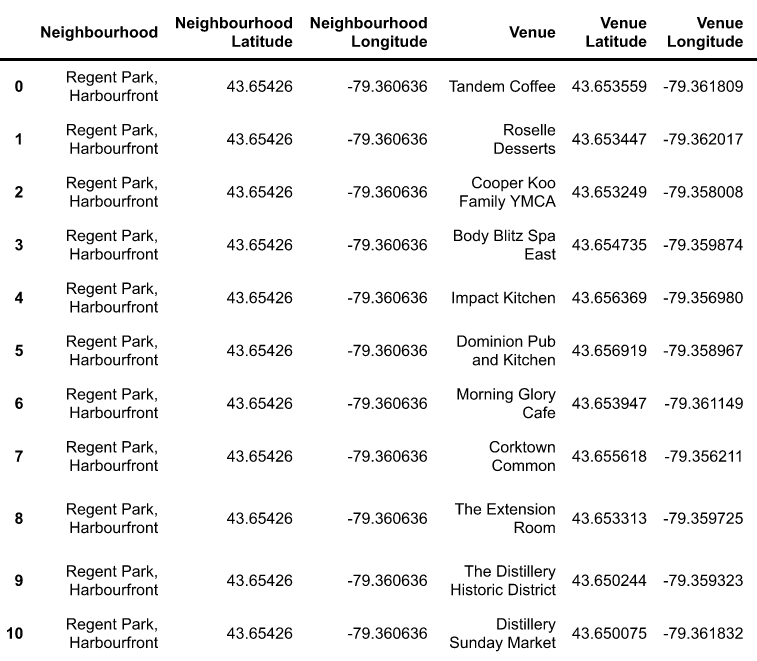




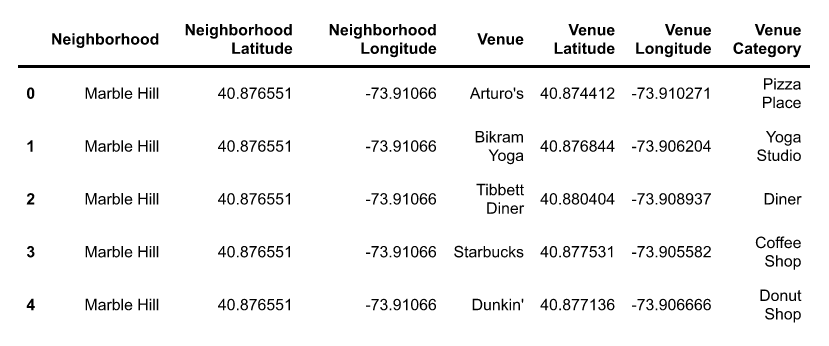
* 1. **Using Foursquare**

Foursquare is a tool/portal with its set of services meant for obtaining geographical data. We use Foursquare pretty much to obtain the coordinates on these cities and their respective localities.

*Toronto data*



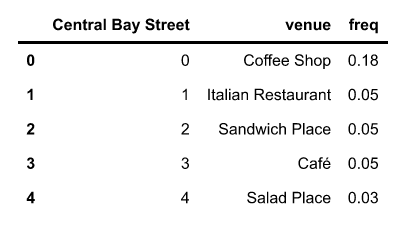
*NY data*



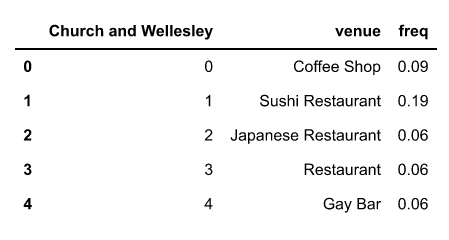
The data was also used to find out the venues around, the frequency of visits on these to gauge what was more popular (eg a Café, Restaurant and so on). Further besides preparing data frames for 5 venues based on 5 neighborhoods, bar graphs were prepared to know what kind of venues were frequented and hence what was popular.

*Toronto data*



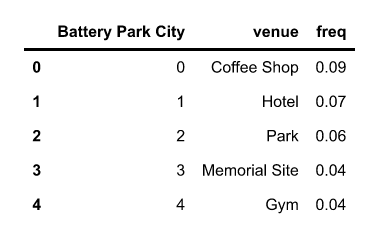








*NY data*

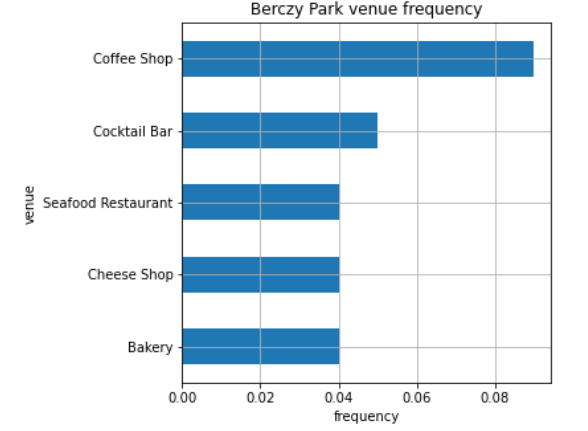


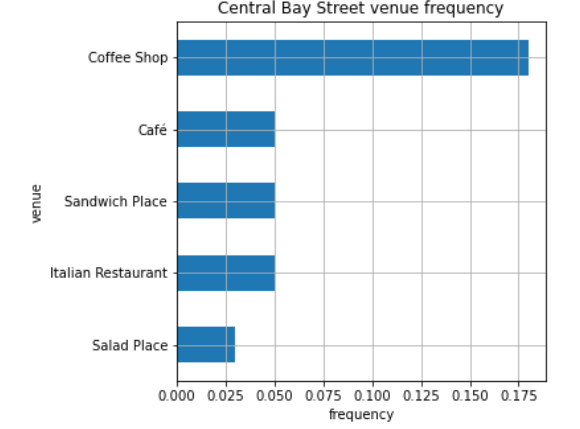


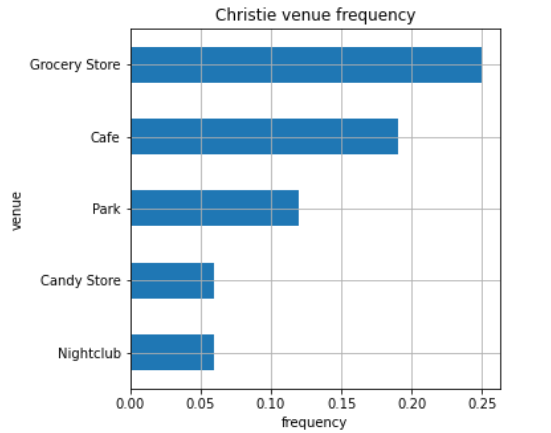
1. **Methodology**
   1. **Exploration**

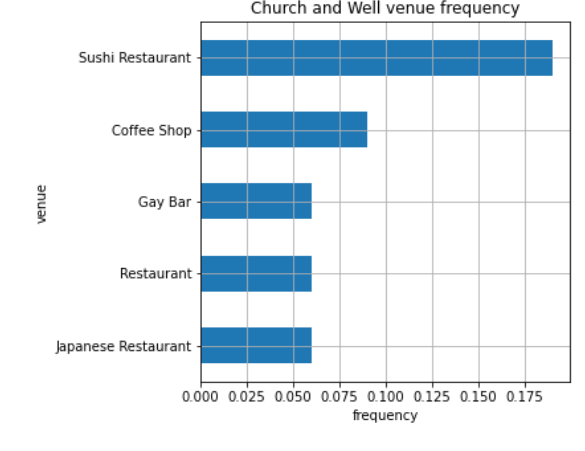
We would use information on 5 venues in 5 neighborhoods, use their frequency to prepare bar charts. The number of venues and neighborhoods can be increased to give a more in depth picture though the data obtained for 5 venues (from 5 neighborhoods) does suffice initially.

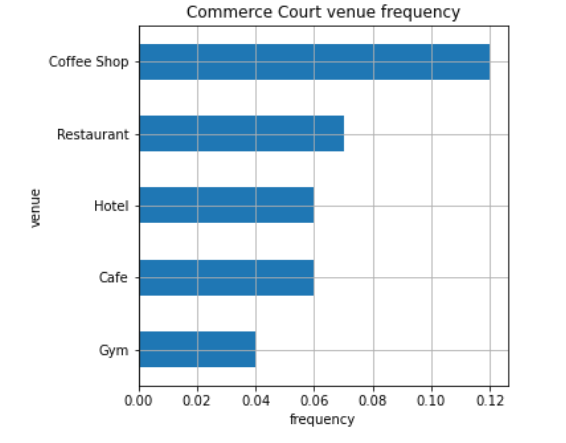
*Downtown Toronto neighbourhood bar graphs*



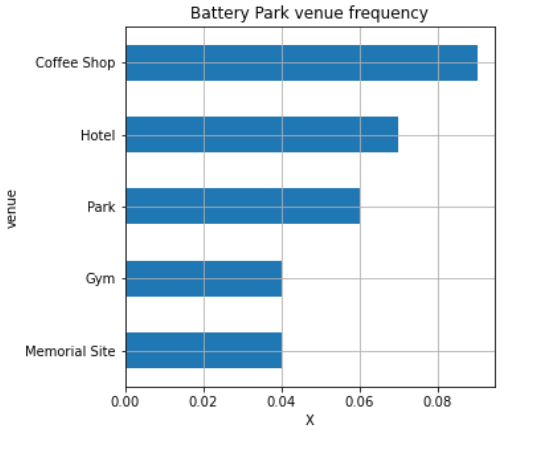


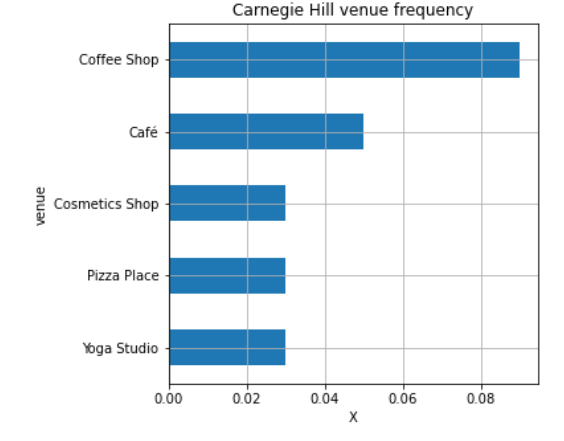


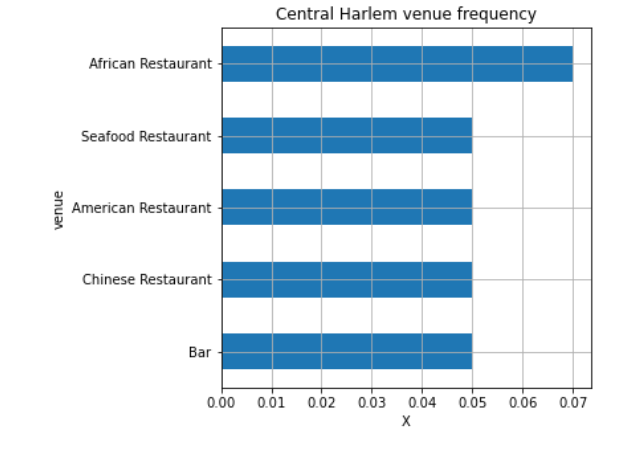


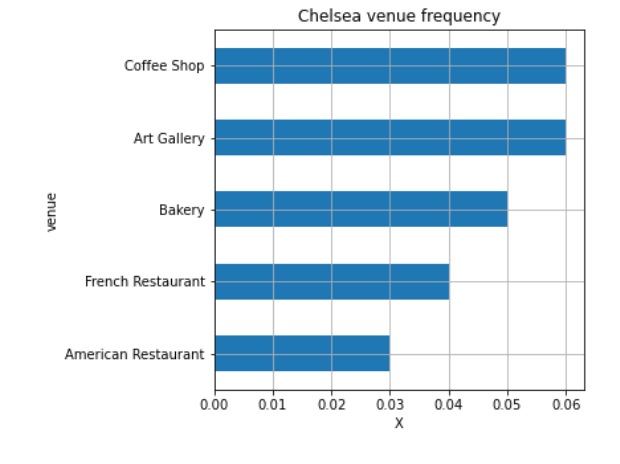


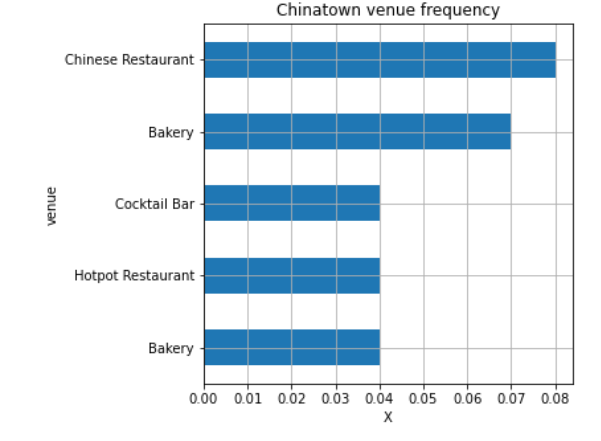
*Manhattan neighbourhood bar graphs*







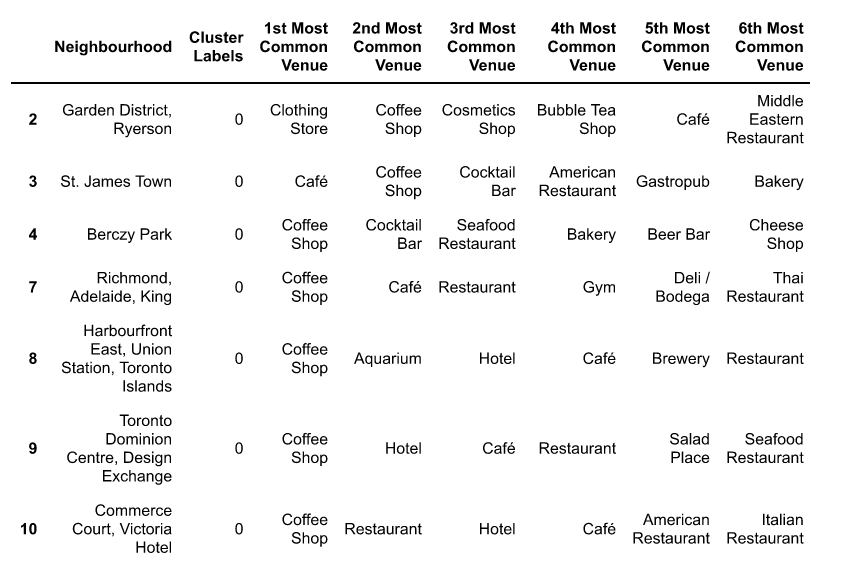




* 1. **Clustering**

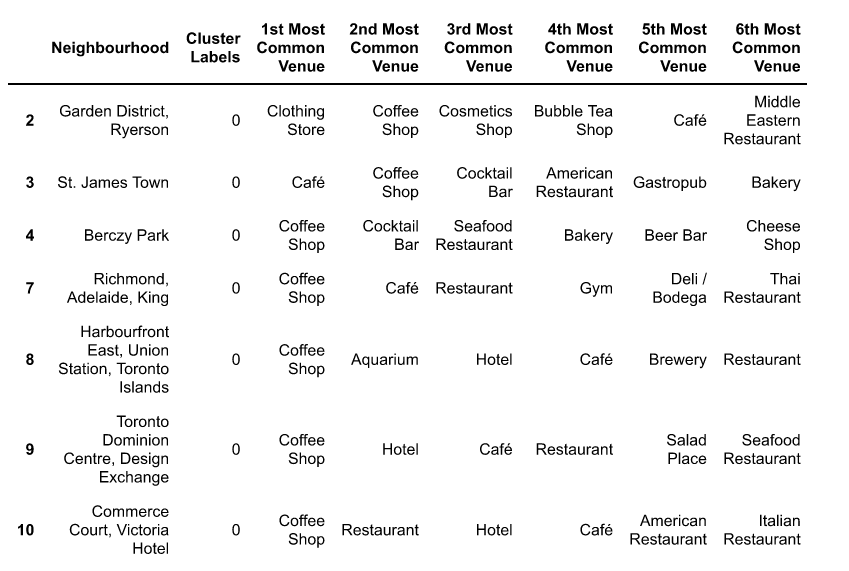
Using k-means to cluster the neighborhoods into 5 clusters. Preparing data frames for the respective neighborhoods and the top 10 venues.

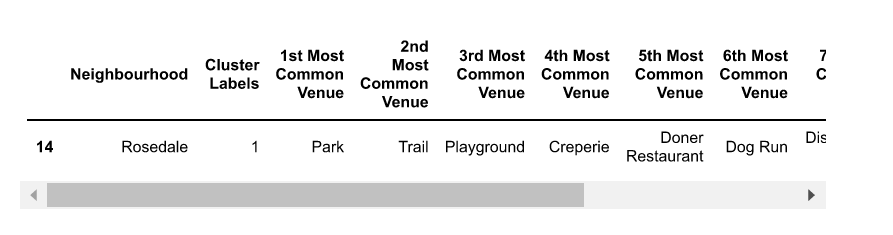
For clustering, one hot encoding is done to give binary values to venue categories. Data is grouped by borough names to find out how many venues of each category exists in these boroughs. The frequency of the respective or more popular venues in these neighborhoods is seen in accordance.

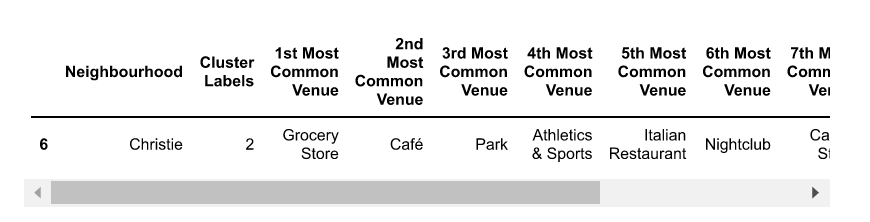


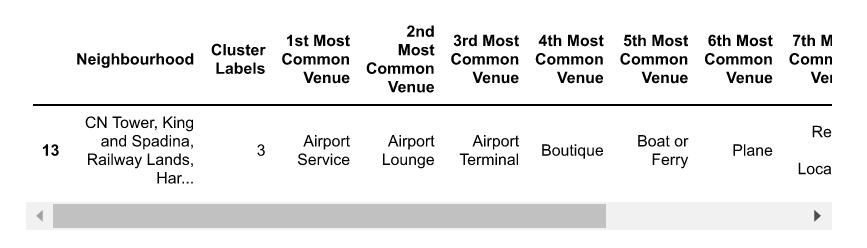
The clusters are further broken down into 5 across neighborhoods (Toronto and NY). We get an idea of the most common venues in each of these clusters. Like what is popular and accordingly the next one and so on.

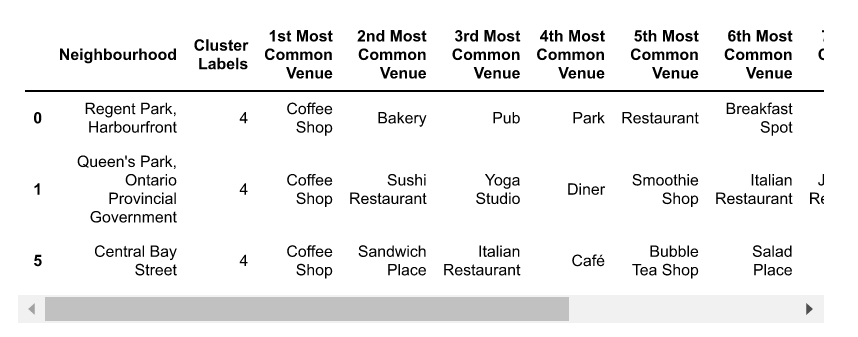
*Toronto neighbourhood*





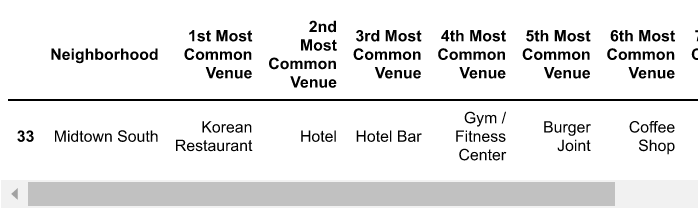






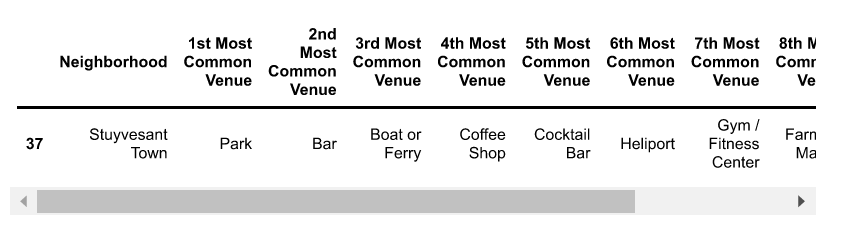
*Manhattan neighbourhood*











1. **Results**
   1. Analysis of NY and Toronto data to find out the similarities. As can be seen above we get an idea of the venues frequent in five random neighborhoods. Café, restaurants, parks, gym are pretty much the norm.

What can be observed is that ‘food’ and its avenues are the most sought after eg Coffee shops/Café’s/Restaurants shine the most in these neighborhoods. Augmenting to their similarities we can see that parks, gym are the other segments that are pretty much the norm. If we had to pull out a larger sample of neighborhoods, we are mostly likely to come up with representation on similar lines.

1. **Conclusion**
   1. From the above data it can be concluded that comparing neighborhoods of two metros (which are also the financial districts of respective countries), one can consider that the most common venues are coffee shops, cafes and restaurants followed by parks, gyms as such. This gives us an idea of the social space around and gives us the prominent venues. A more accurate measure of this data would obviously be to collect more samples in terms of neighborhoods, venues and compare them in case needs to be done.

**References:**

1. NY data “https://cf-courses-data.s3.us.cloud-object-storage.appdoma

in.cloud/IBMDeveloperSkillsNetwork-DS0701EN SkillsNetwork/labs/newyork\_data.json”

1. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
2. Foursquare portal for geographical data.

<https://foursquare.com/>