**TOP RESTAURANTS TO TARGET FOR DELIVERY SERVICES IN TORONTO**

**TONY D. SUH**

**May 12, 2020**

**1. INTRODUCTION**

*1.1 Problem Statement:*  
With the ongoing fight to limit the spread of COVID-19 virus, a lot of restaurants are either closed or operating a take-out service. Furthermore, COVID-19 has impacted the Canadian economy by increasing unemployment rates while businesses in various sectors are closed until further notice. In this tough period, it would be great if Torontonians can have access to their favourite restaurants with delivery services like UBER Eats, FOODORA, DoorDash, and etc.

If we can determine top restaurants in Toronto by FourSquare API and we can provide insights as to 1) which restaurants Torontonians love the most, 2) which restaurants should be linked to a delivery service, and 3) which restaurants should the delivery service providers target to strengthen their services.

*1.2 Client Base:*

a. Restaurants in Toronto - Make an attempt to keep small and medium business owners in Toronto alive.

b. Food Delivery Services in Toronto - Determining which restaurants are popular by users and partnering with those restaurants to provide a delivery service, it will only make the life of Torontonians much better.

c. People in Toronto - People need to eat to survive and there is no such thing as a free lunch. Might as well pay for a meal that is highly rated and delicious!

*1.3. Analytic Approach*

For achieving the grouping as stated in the problem statement, I can make use of k-means clustering. K -means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

I will be using k-means clustering to group neighborhoods in Toronto and find out top 10 venues in each neighborhood.

**2. DATA PREPARATION**

*2.1 Data Requirements:*

a. Neighborhood data of Toronto ([Wikipedia](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M))

b. FourSquare API for location data of each neighborhood in Toronto

*2.2. Data Collection:*

a. Web scrap Neighborhood data of Toronto ([Wikipedia](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)) using Beautiful Soup python library to map Toronto neighborhoods by Postal Code and Borough.

b. Obtain latitude and longitude of each postal code by using ‘<https://cocl.us/Geospatial_data>’.

c. Use foursquare API calls to get the location details corresponding to each neighborhood:

# create the API request URL

*url = 'https://api.foursquare.com/v2/venues/explore?&client\_id={}&client\_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(*

*CLIENT\_ID,*

*CLIENT\_SECRET,*

*VERSION,*

*lat,*

*lng,*

*radius,*

*LIMIT)*

*# make the GET request*

*results = requests.get(url). json()["response"]['groups'][0]['items']*

*2.3 Data Understanding*

To the data collected in the prior stage, I will explore the data and understand the following:

a. Explore the neighborhoods in Toronto

b. Analyze each neighborhood.

c. Explore the venues corresponding to any one of the neighborhoods.

*2.4. Data Preparation*

This stage involves exploring the data further and making sure that it is in the right format for k-means clustering that I selected in the analytic approach stage.

a. If I see any gap in the data, then I will go back to the data collection stage and collect any missing data.

b. If I see any erroneous data / unnecessary data, then I will filter and remove those data to prepare the final data which will be used in clustering.

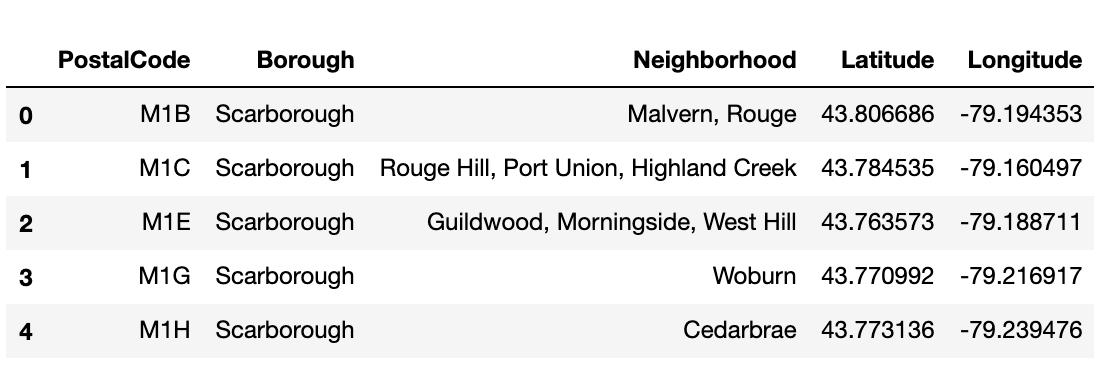
At the end of this stage, I should have the dataset ready to be fed into the clustering analysis.

**3. DATA ANALYSIS**

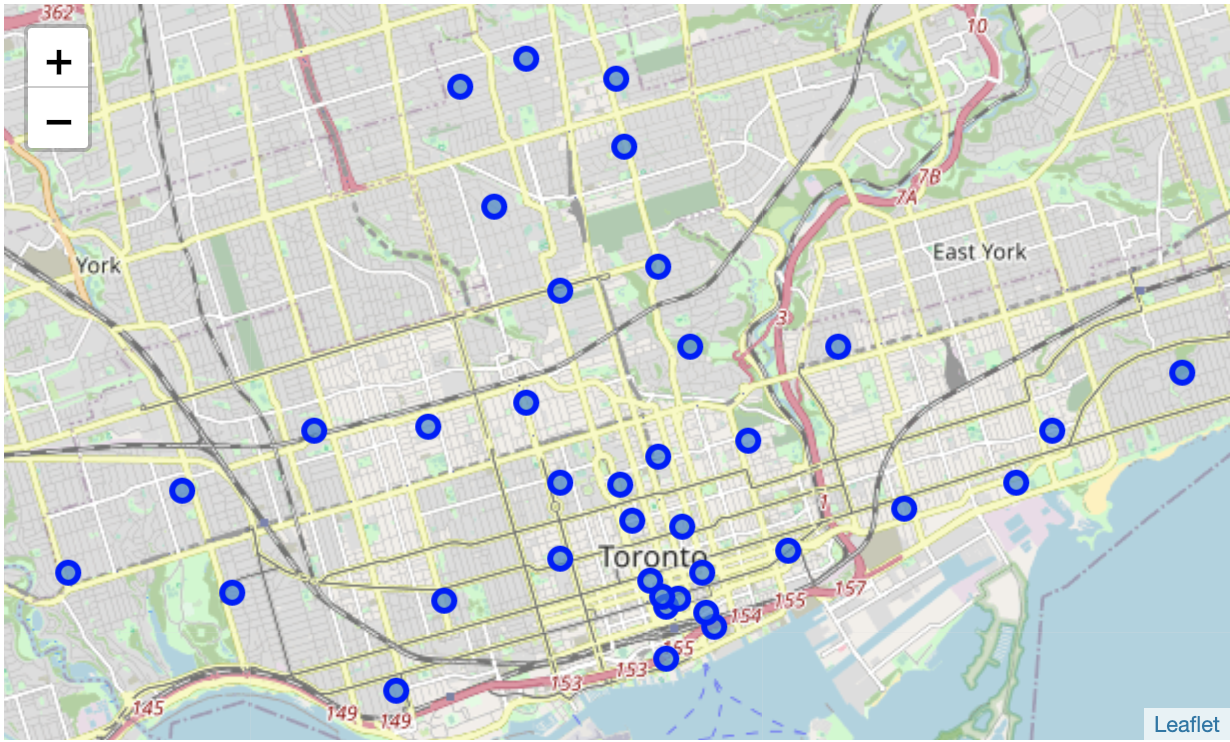
Preparing neighborhood data

Extract the table with postal code, borough and neighborhood data from the Wikipedia link ‘https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M’ using Beautiful Soup library and save it in a csv file. Create a data frame from the csv file created in step 1 and filter rows with borough='Not assigned'. If a postal code has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.Group neighborhood based on postal code and borough. It is represented by the code below: 

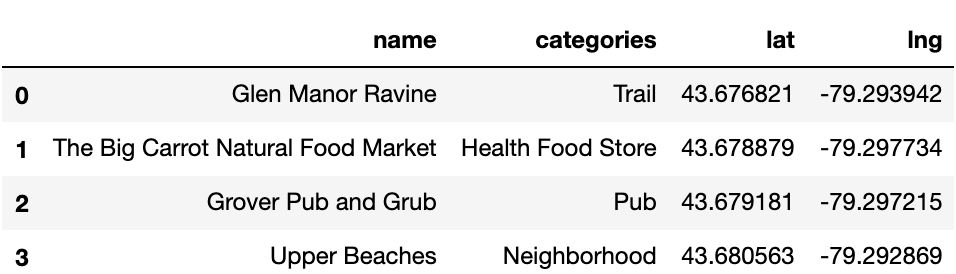
Read the geo spatial data from ‘https://cocl.us/Geospatial\_data’ and merge it with the data frame created in step d. The resulting data frame looks something like:



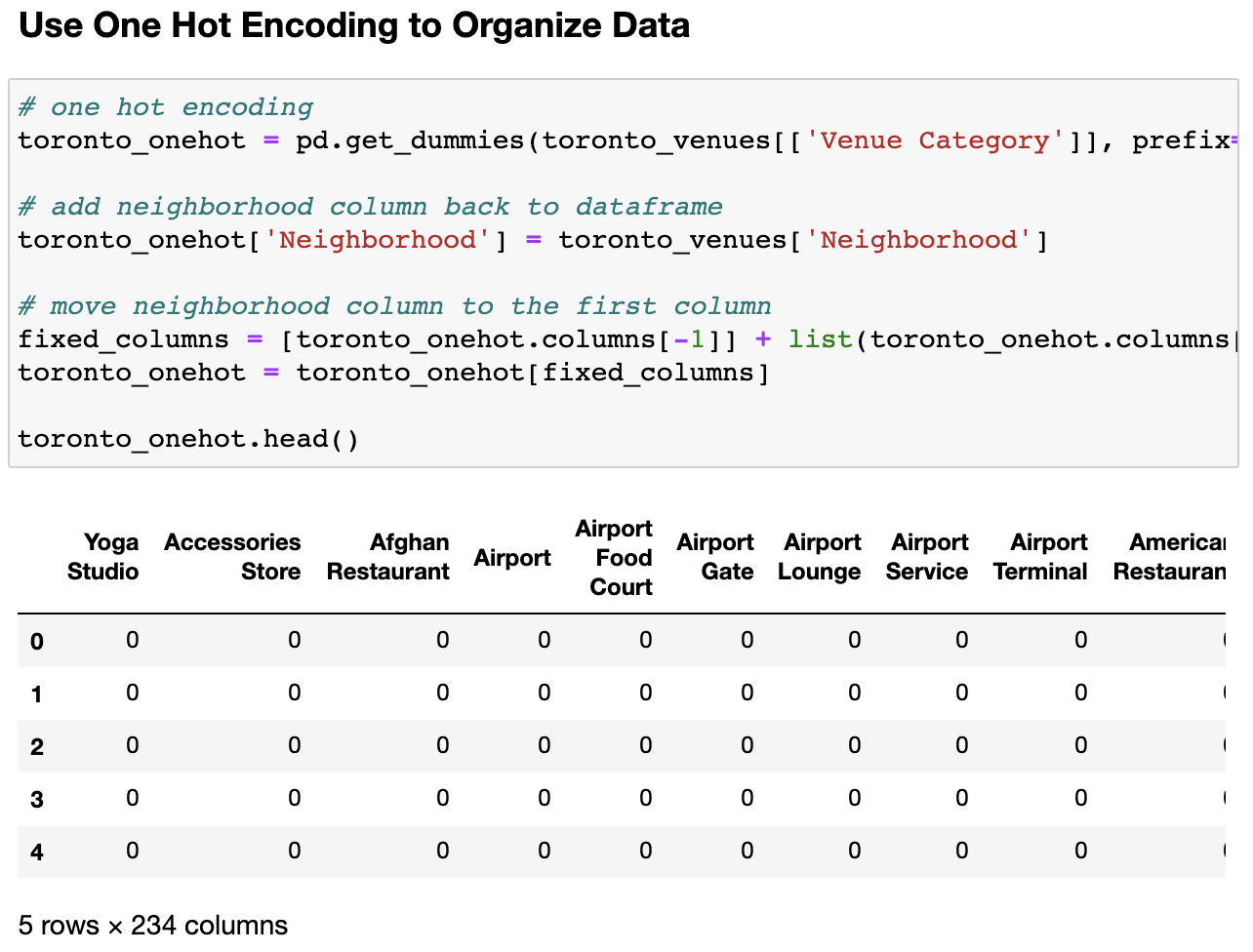
Create a map showing the neighborhoods in Toronto



Test the Foursquare API to get venues for one neighborhood, “The Beaches”. The venues details collected using the neighborhood data and Foursquare API for “The Beaches” neighborhood is as below:



Apply One hot encoding and group rows on neighborhood by taking mean, the resulting data would look something like the below table, where the point of interest is that there are 38 neighborhoods and 234 unique categories.



Find the top 10 common venues in each neighborhood:

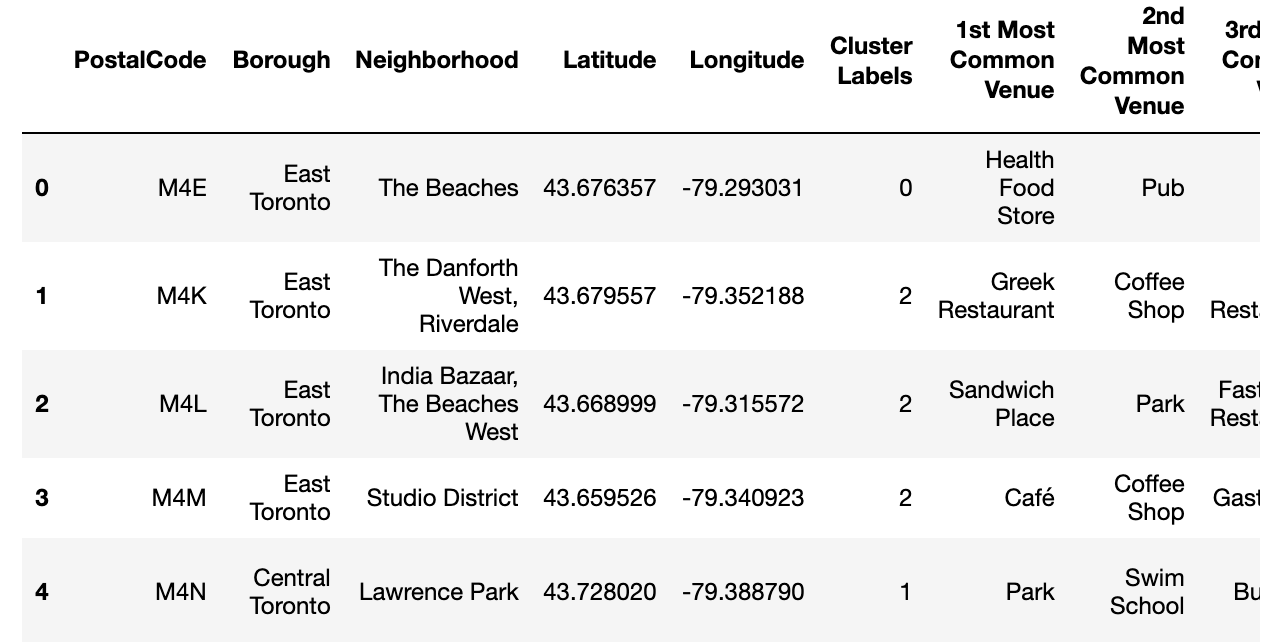
Below table shows how the data looks for all the neighborhoods with a sample of 5 neighborhoods.

Now that I have the required top 10 data ready, I will proceed to the clustering of the data thus prepared in the above section.

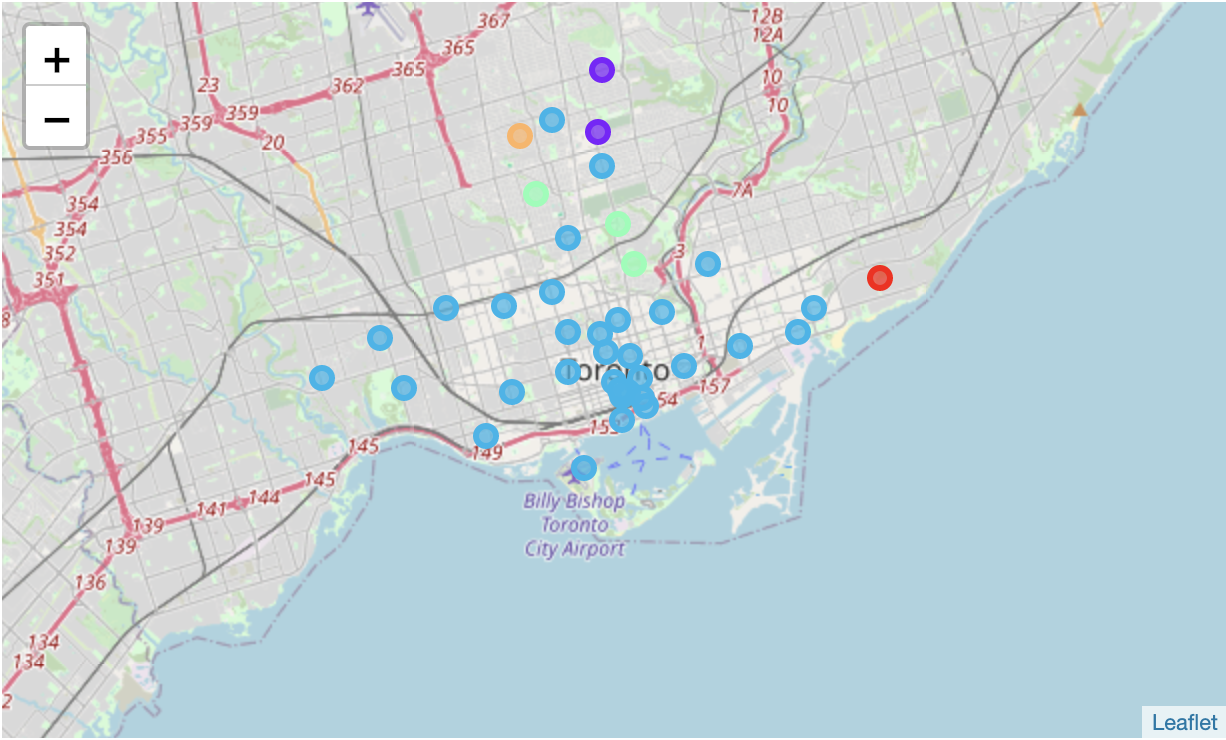
**4. Cluster Neighborhoods**

I will use K-Means clustering for grouping the data of the neighborhoods in Toronto. K -means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided.

I created 5 clusters for the 38 neighborhood groups indicating the top 10 venues. The clustered data look like the table as shown below:



I created matplotlib folium maps to visualize the clusters.

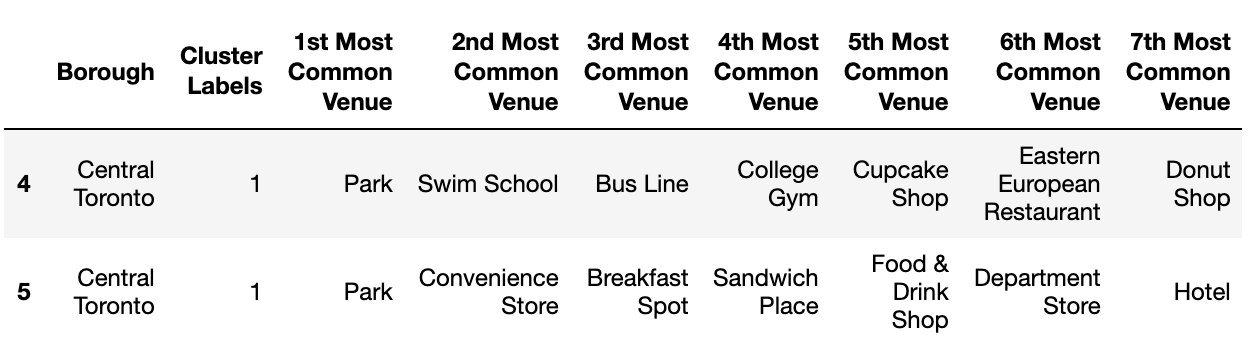


Then I examined the data in the 5 clusters that are created.

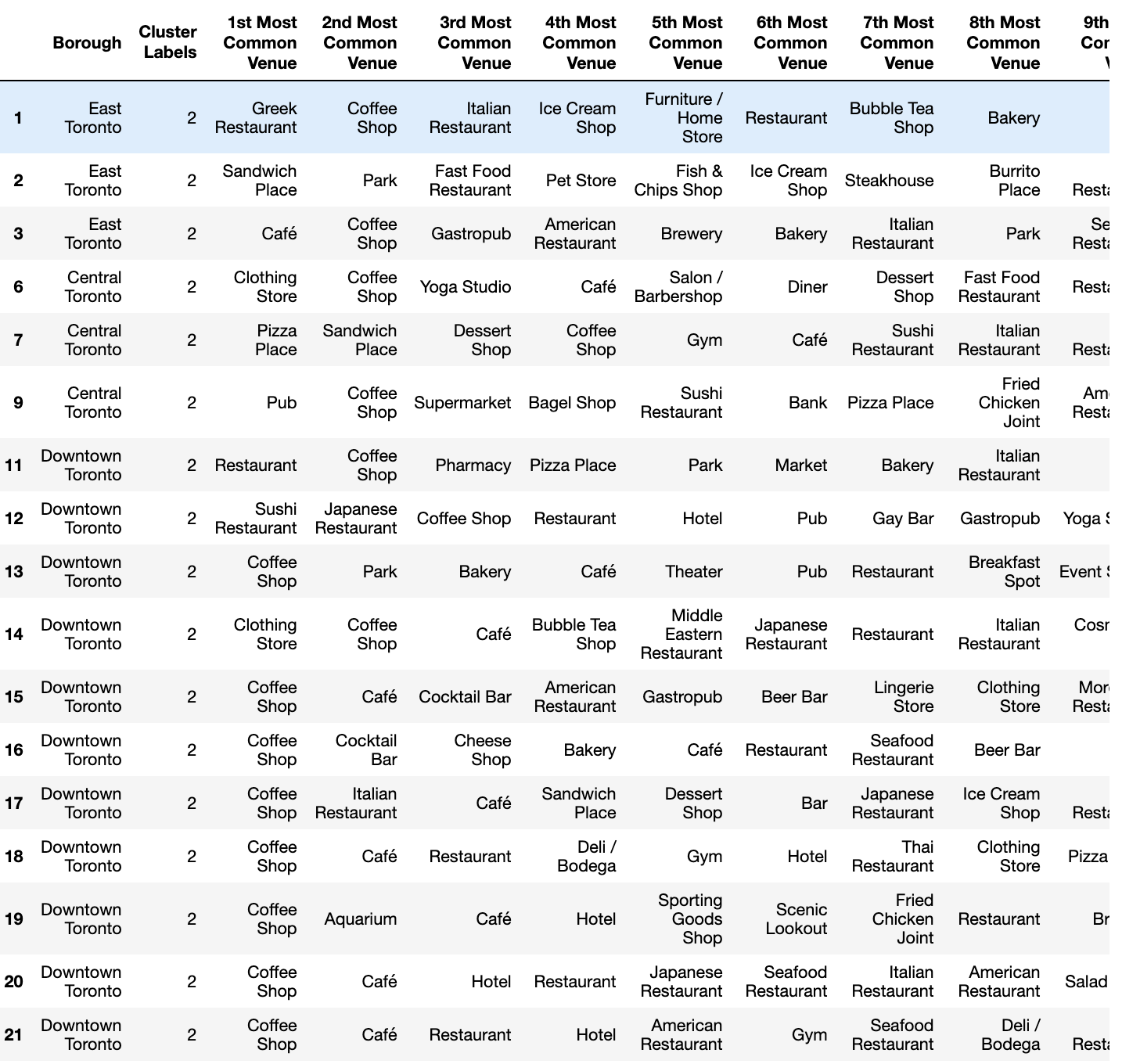
Cluster 1:



Cluster 2:



Cluster 3:



Cluster 4:



Cluster 5:



**5. CONCLUSION**

In this study, I created clusters of neighborhoods in Toronto and mapped the top 10 restaurants with each corresponding neighborhood (some filters need to be applied). Delivery services and Torontonians who are tired of the eateries they have in their neighborhoods can now look towards this data to get a refreshment of foods that they might enjoy. Furthermore, delivery services who have not partnered with the restaurants that are highly sought out may want to gain a competitive advantage against their competition by partnering an exclusive contract with the top eateries in various neighborhoods to strengthen their delivery platform.

**6. REFERENCES**

1. <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
2. <https://cocl.us/Geospatial_data>
3. <https://foursquare.com/>