Coursera Capstone project report

Coursera IBM Data Science Certification

Akshay C A

Oct -12th 2019

Contents

The Battle of Neighbourhoods

Introduction/Business Problem

Define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute.

Toronto, one of the famous places in world which is diverse and multicultural. I'm planning to move into Toronto but I'm not sure of the exact neighborhood which would be a best fit for me. I would like to explore how much they are similar or dissimilar neighborhoods are aspects from a tourist point of view regarding food, accommodation, beautiful places, and many more.

You should be able to choose, compare different neighborhoods in terms of a service, search for potential explanation of why a neighborhood is popular etc., . Hence the name of the capstone project will be the **Battle of the neighborhoods.**

Data section

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea.

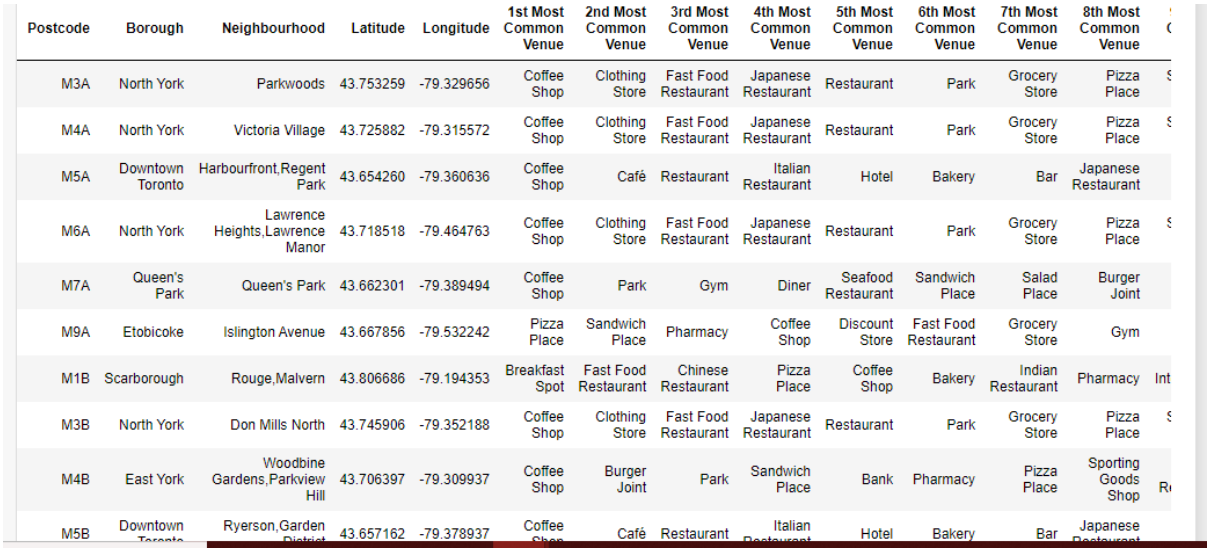
In order to explore the similar or dissimilar in aspects of the neighborhoods, I would need **Foursquare location data** to fetch the Venue Category and Boroughs of Toronto.

We will segment it into different neighborhoods using the geographical coordinates of the center of each neighborhood, and then using a combination of location data and machine learning.

Building a recomendation system for finding best clusters of neighborhood based on certain criteria is valuable analytical problem that perfectly fits into Clustering type of Data Science problems which could be solved by unsupervised learning algorithms.

Data Analysis

After performing several data transformation our Dataframe would look like the below screenshot:



Postcode	Borough	Neighbourhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	
M3A	North York	Parkwoods	43.753259	-79.329656	Coffee Shop	Clothing Store	Fast Food Restaurant	Japanese Restaurant	Restaurant	Park	Grocery Store	Pizza Place	S
M4A	North York	Victoria Village	43.725882	-79.315572	Coffee Shop	Clothing Store	Fast Food Restaurant	Japanese Restaurant	Restaurant	Park	Grocery Store	Pizza Place	S
M5A	Downtown Toronto	Harbourfront,Regent Park	43.654260	-79.360636	Coffee Shop	Café	Restaurant	Italian Restaurant	Hotel	Bakery	Bar	Japanese Restaurant	
M6A	North York	Lawrence Heights,Lawrence Manor	43.718518	-79.464763	Coffee Shop	Clothing Store	Fast Food Restaurant	Japanese Restaurant	Restaurant	Park	Grocery Store	Pizza Place	S
M7A	Queen's Park	Queen's Park	43.662301	-79.389494	Coffee Shop	Park	Gym	Diner	Seafood Restaurant	Sandwich Place	Salad Place	Burger Joint	
M9A	Etobicoke	Islington Avenue	43.667856	-79.532242	Pizza Place	Sandwich Place	Pharmacy	Coffee Shop	Discount Store	Fast Food Restaurant	Grocery Store	Gym	
M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	Breakfast Spot	Fast Food Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bakery	Indian Restaurant	Pharmacy	Int
M3B	North York	Don Mills North	43.745906	-79.352188	Coffee Shop	Clothing Store	Fast Food Restaurant	Japanese Restaurant	Restaurant	Park	Grocery Store	Pizza Place	S
M4B	East York	Woodbine Gardens,Parkview Hill	43.706397	-79.309937	Coffee Shop	Burger Joint	Park	Sandwich Place	Bank	Pharmacy	Pizza Place	Sporting Goods Shop	R
M5B	Downtown Toronto	Ryerson,Garden District	43.657162	-79.378937	Coffee Shop	Café	Restaurant	Italian Restaurant	Hotel	Bakery	Bar	Japanese Restaurant	

After the Exploratory analysis if was found that there are 276 unique categories.

And Borough along with the top 5 most common venues

```
----Central Toronto----
              venue  freq
0       Coffee Shop  0.07
1       Pizza Place  0.06
2    Sandwich Place  0.06
3              Park  0.05
4  Sushi Restaurant  0.04


----Downtown Toronto----
                venue  freq
0         Coffee Shop  0.10
1               Café  0.05
2  Italian Restaurant  0.03
3          Restaurant  0.03
4               Hotel  0.03


----East Toronto----
                venue  freq
0         Coffee Shop  0.07
1    Greek Restaurant  0.07
2  Italian Restaurant  0.05
3               Café  0.04
4      Ice Cream Shop  0.04
```

```
----East York----
                venue  freq
0          Coffee Shop  0.06
1         Burger Joint  0.05
2                 Park  0.05
3  Sporting Goods Shop  0.04
4                 Bank  0.04


----Etobicoke----
            venue  freq
0     Pizza Place  0.12
1  Sandwich Place  0.07
2        Pharmacy  0.05
3     Coffee Shop  0.05
4  Discount Store  0.04
```

Machine Learning – KMeans Clustering

A Clustering Algorithm tries to analyse natural groups of data on the basis of some similarity. It locates the centroid of the group of data points. To carry out effective clustering, the algorithm evaluates the distance between each point from the centroid of the cluster.

K-means Clustering will group these locations of maximum prone areas into clusters and define a cluster center for each clusters. These Clusters centers are the centroids of each cluster and are at a minimum distance from all the points of a particular cluster.

Code:

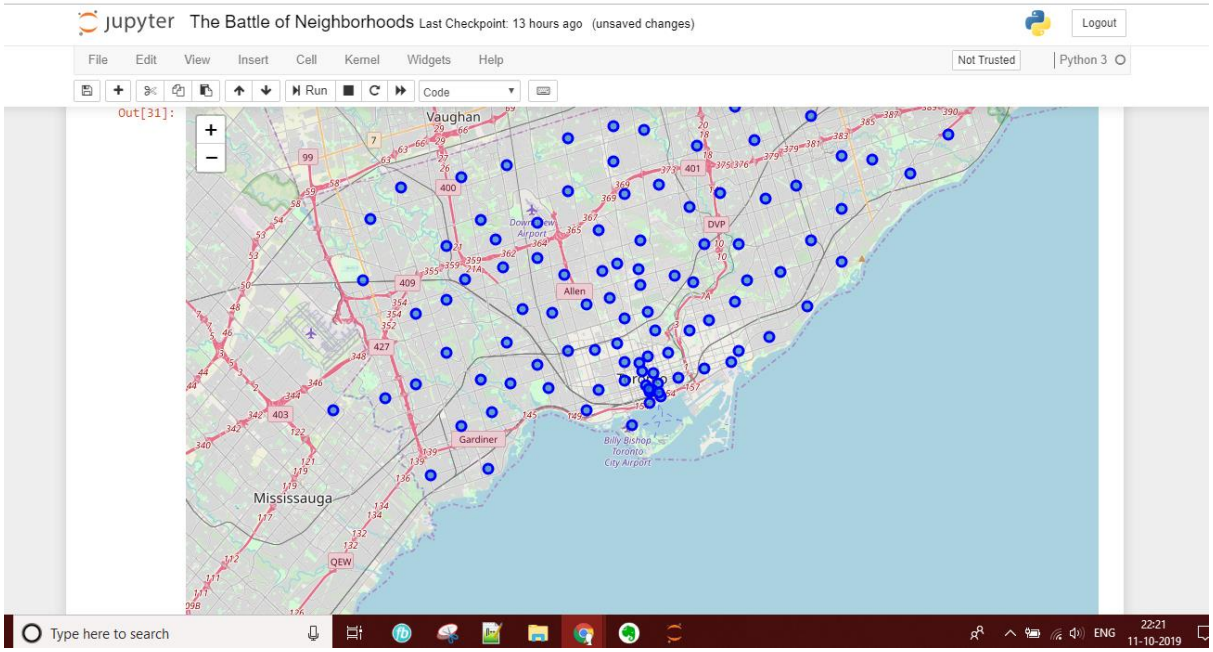
Clustering Borough

In [67]:
```
from sklearn.cluster import KMeans

kclusters = 5

toronto_grouped_clustering = toronto_wc.drop('Borough', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[1:10]
```

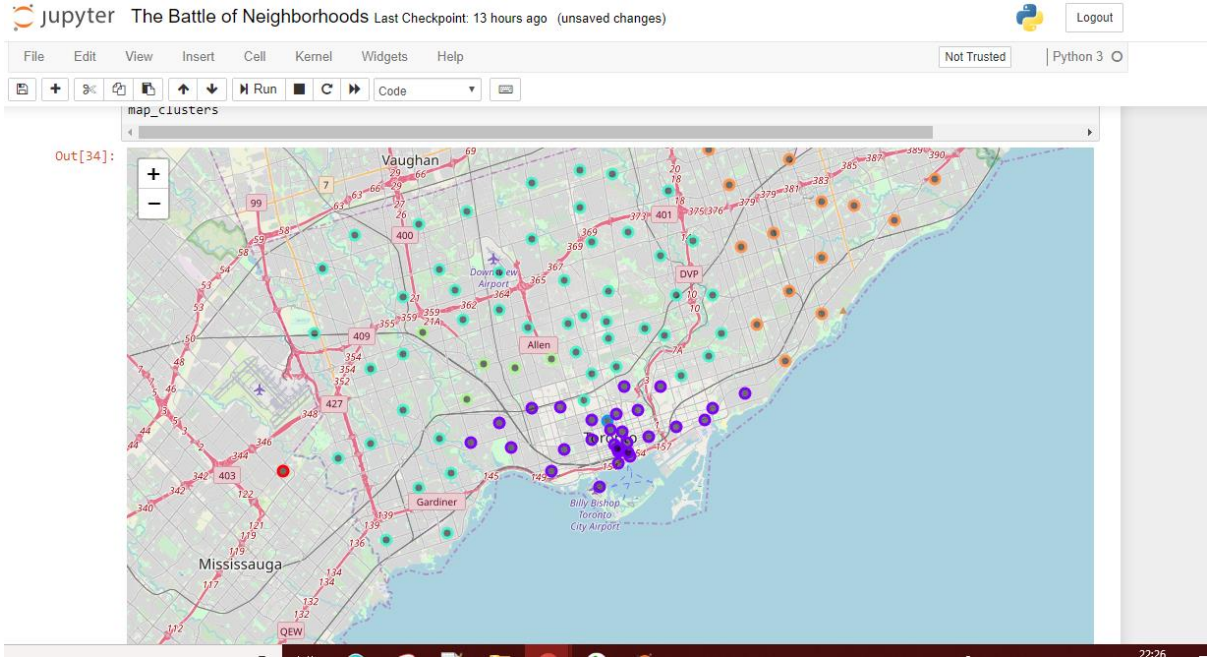
Out[67]:
```
array([0, 0, 4, 4, 1, 0, 3, 4, 0])
```

Visualization

Visualization of Toronto's Borough



Clusters visualization of Toronto's Borough



Results and Conclusion:

Now let's try to fetch **insights** from the data.

The following are the highlights of the 5 clusters above:

Cluster #0

Most common venues: Restaurants and Coffee Shop

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Cluster Labels
0	Central Toronto	Coffee Shop	Sandwich Place	Pizza Place	Park	Café	Sushi Restaurant	Gym	Dessert Shop	Restaurant	Clothing Store	0
1	Downtown Toronto	Coffee Shop	Café	Restaurant	Italian Restaurant	Hotel	Bakery	Bar	Japanese Restaurant	Park	Seafood Restaurant	0
2	East Toronto	Coffee Shop	Greek Restaurant	Italian Restaurant	Café	Ice Cream Shop	Brewery	Yoga Studio	American Restaurant	Pizza Place	Bakery	0
6	North York	Coffee Shop	Clothing Store	Fast Food Restaurant	Japanese Restaurant	Restaurant	Park	Grocery Store	Pizza Place	Sandwich Place	Bank	0
9	West Toronto	Bar	Café	Coffee Shop	Bakery	Italian Restaurant	Restaurant	Breakfast Spot	Men's Store	Pizza Place	French Restaurant	0

Cluster #1

Most common venues: Hotels and Gym/Fitness center

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Cluster Labels
5	Mississauga	Hotel	Coffee Shop	Gym / Fitness Center	Mediterranean Restaurant	Fried Chicken Joint	Middle Eastern Restaurant	Sandwich Place	American Restaurant	Burrito Place	Drugstore	1

Cluster #2

Most common venues: Park, Convenience Store and Check Cashing Service

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Cluster Labels
10	York	Park	Convenience Store	Check Cashing Service	Trail	Restaurant	Caribbean Restaurant	Bus Line	Sandwich Place	Field	Fast Food Restaurant	2

Cluster #3

Most common venues: Park and Gym

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Cluster Labels
7	Queen's Park	Coffee Shop	Park	Gym	Diner	Seafood Restaurant	Sandwich Place	Salad Place	Burger Joint	Burrito Place	Café	3

Cluster #4

Most common venues: Fast Food Restaurants

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Cluster Labels
3	East York	Coffee Shop	Burger Joint	Park	Sandwich Place	Bank	Pharmacy	Pizza Place	Sporting Goods Shop	Indian Restaurant	Gym	4
4	Etobicoke	Pizza Place	Sandwich Place	Pharmacy	Coffee Shop	Discount Store	Fast Food Restaurant	Grocery Store	Gym	Bakery	Beer Store	4
8	Scarborough	Breakfast Spot	Fast Food Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bakery	Indian Restaurant	Pharmacy	Intersection	Sandwich Place	4

My personal preference would be a home around Fast Food Restaurants so Cluster #4 Neighborhoods - East York, Etobicoke and Scarborough would be best for me :)

In conclusion, this project would have had better results if there were more available data in terms of actual land pricing data within the area, public transportantion access and allowance of more venues exploration with the Foursquare (limited venues for free calls).