

Capstone Project – Neighborhood Recommendation for a Real Estate Company.

Business Problem:

Identifying the best location for a new Apartment Complex based on walking scores.

Problem Description and Stakeholders:

A real estate company based in the United States, is thinking about expanding their business to Canada. This company is known for affordable high rise's located in areas with high walking scores. They market their apartments for people who prefer using public transportation/or are in densely populated areas with minimal parking space.

A good walking score essentially means that the Apartment is in walking distance to a lot of amenities like restaurant's, public transportation, super markets, parts etc.

Since it's the first time they are entering Canadian market, they decided to pick Toronto as their first hub.

In this project, I will work with FourSquare API's and the neighborhoods in Toronto, to come up with recommendations on the best locations for an apartment complex.

Dataset:

For this project, I worked on getting a list of all the Neighborhoods in the city of Toronto. I utilized the Beautiful Soup web scraper to parse the list of zip codes and neighborhoods from Wikipedia. Once I got the data into a list, I utilized the panda's library to put that data in a Data Frame.

I then performed data cleaning operations, to filter unnecessary information, fill missing values and complete the data frame.

Here's how the Data frame looks like:

	PostalCode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M5A	Downtown Toronto	Regent Park
6	M6A	North York	Lawrence Heights

Since FourSquare requires the latitude and longitude, I used a CSV file provided by [cocl.us](#) which contains the location coordinates for all the postal codes.

	PostalCode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Now we have two sets of data frames one with location data and one with neighborhood information. I merged both the data frames to have a single unified list.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

At this stage, since I want to focus only on the Borough's in Toronto, I filter the data. My final data frame looks as follows:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188

	PostalCode	Borough	Neighborhood	Latitude	Longitude
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Now I have a list of all the Neighborhood's with a Borough in Toronto. This is my dataset for the analysis. Using Folium, a visualization library, I can plot all the Neighborhoods on a Map and prep the data for clustering.

Using FourSquare API's I will perform exploratory analysis to identify the features that are most suitable for an apartment complex and then work on clustering and segmentation to identify potential zones.

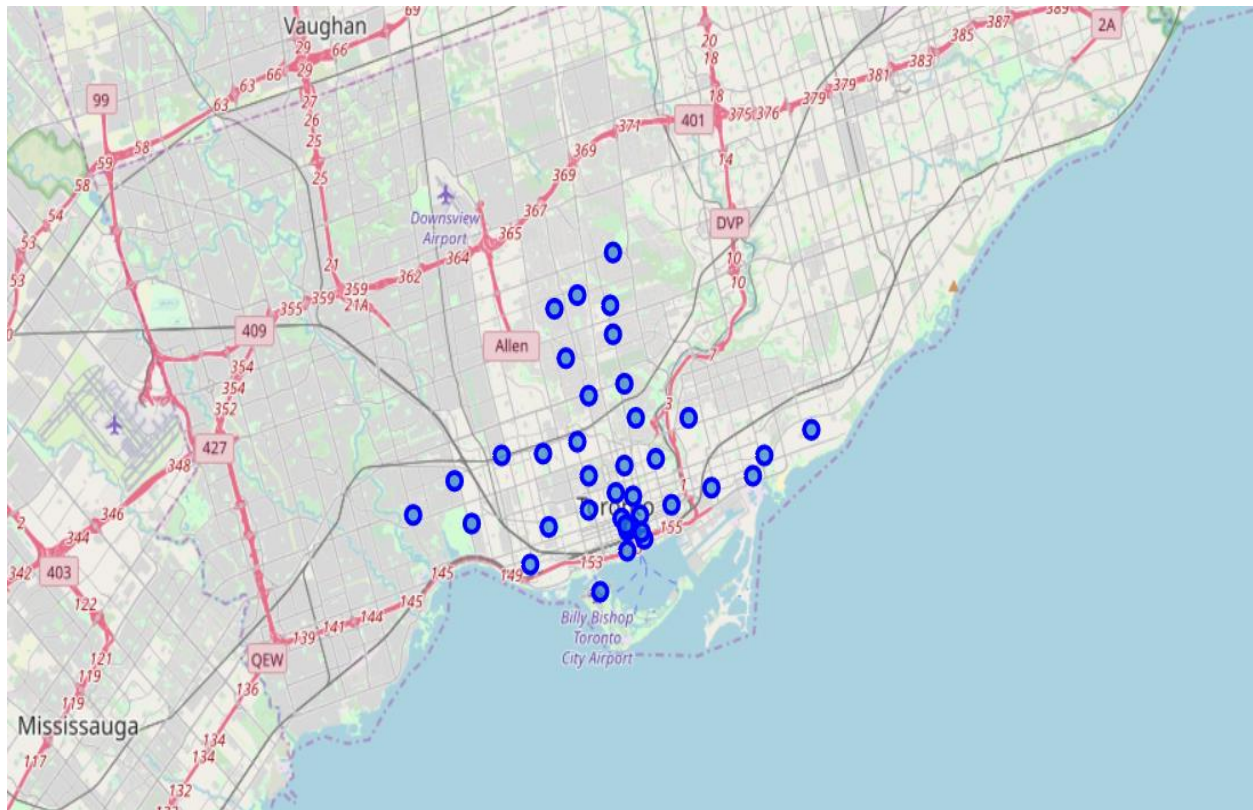
Methodology:

From the previous section, we can see that I have a dataframe containing a list of all the Boroughs in the city of Toronto. This listing contains the postal codes, borough, neighborhood along with it's latitude and longitude coordinates.

Since our stakeholder's want to know the best location to construct an Apartment complex with a good walking score, we need get details about the venues present in the Neighborhoods. To do this, I have relied on FourSquare, a city guide which let's you explore neighborhoods and get a list of all the available venues.

Using the geolocator library, I retrieved the geographical coordinates of Toronto city, and plotted it on a map. I utilized Folium to generate the map and visualize the coordinates.

I then plotted all the neighborhoods in Toronto based on their latitude and longitude on to the map of Toronto. Please see the image below for details.



Using the FourSquare API's I got a list of all the venues in Toronto based on their geographical coordinates and assigned it to a pandas dataframe. Here's a sample from the output.

The Beaches
The Danforth West, Riverdale
The Beaches West, India Bazaar
Studio District
Lawrence Park
Davisville North
North Toronto West
Davisville
Moore Park, Summerhill East
Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West
Rosedale
Cabbagetown, St. James Town
Church and Wellesley
Harbourfront, Regent Park
Ryerson, Garden District
St. James Town
Berczy Park

Central Bay Street

Adelaide, King, Richmond

Harbourfront East, Toronto Islands, Union Station

Design Exchange, Toronto Dominion Centre

The Annex, North Midtown, Yorkville

The dataframe contains the Neighborhood, geolocation details along with the venues present and their location info along with the category the venue belongs to.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
1	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
2	The Beaches	43.676357	-79.293031	Starbucks	43.678798	-79.298045	Coffee Shop
3	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
4	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood

There are a total of 839 records in the dataframe containing a list of all the venues. For the next step, I decided to count the number of venues in each Neighborhood.

Davisville	30
Studio District	30
St. James Town	30
Central Bay Street	30

Chinatown, Grange Park, Kensington Market	30
Adelaide, King, Richmond	30
Harbourfront, Regent Park	30
Design Exchange, Toronto Dominion Centre	30
Stn A PO Boxes 25 The Esplanade	30
Cabbagetown, St. James Town	30

Based on this list, I was to get a good idea about which neighborhoods have more venues/user engagement and would be a good fit for the Apartment complex.

Next I decided to group the dataframe based on the neighborhoods. The sample output was as follows:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Adelaide, King, Richmond	30	30	30	30	30	30
Berczy Park	30	30	30	30	30	30
Brockton, Exhibition Place, Parkdale Village	19	19	19	19	19	19
Business Reply Mail Processing Centre 969 Eastern	18	18	18	18	18	18
CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara	13	13	13	13	13	13
Cabbagetown, St. James Town	30	30	30	30	30	30

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Central Bay Street	30	30	30	30	30	30
Chinatown, Grange Park, Kensington Market	30	30	30	30	30	30
Christie	16	16	16	16	16	16
Church and Wellesley	30	30	30	30	30	30

I then calculated the number of unique category of venues. The result was 189. This meant that across all these neighborhoods there were 189 unique venues. This information was the first step in coming up with a list of the most popular venues.

Using the `get_dummies` method provided by the pandas library, I created a binary encoding of all the venues. This helped me explore various neighborhoods and look into which venues were present and absent.

I then grouped this list by Neighborhoods, and calculated the average mean. This helped me get a statistical estimation of the spread of venues.

Next I decided to determine the top 10 venues in each neighborhood based on the frequency of their occurrence. The sample output for one of the Neighborhoods is presented below.

```

----Adelaide, King, Richmond----
      venue  freq
0      Steakhouse  0.10
1  Asian Restaurant  0.07
2          Café  0.07
3          Hotel  0.07
4          Lounge  0.03
5      Food Court  0.03
6  Monument / Landmark  0.03
7      Noodle House  0.03
8      Opera House  0.03
9      Pizza Place  0.03

```

Using this frequency table, I then determined the Top 10 venues for each Neighborhood.

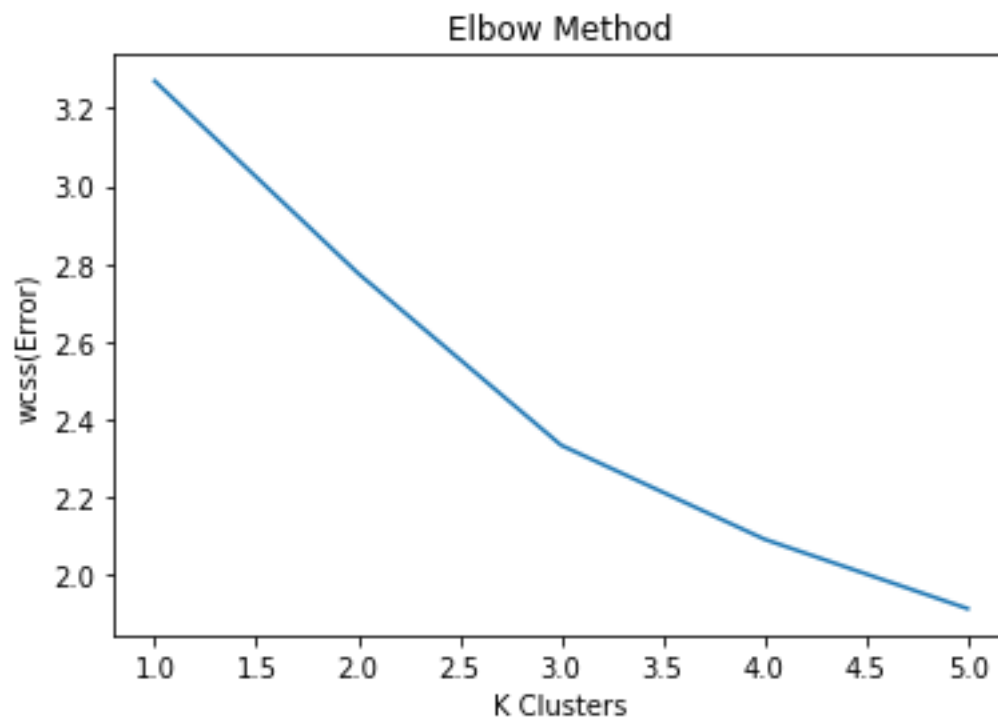
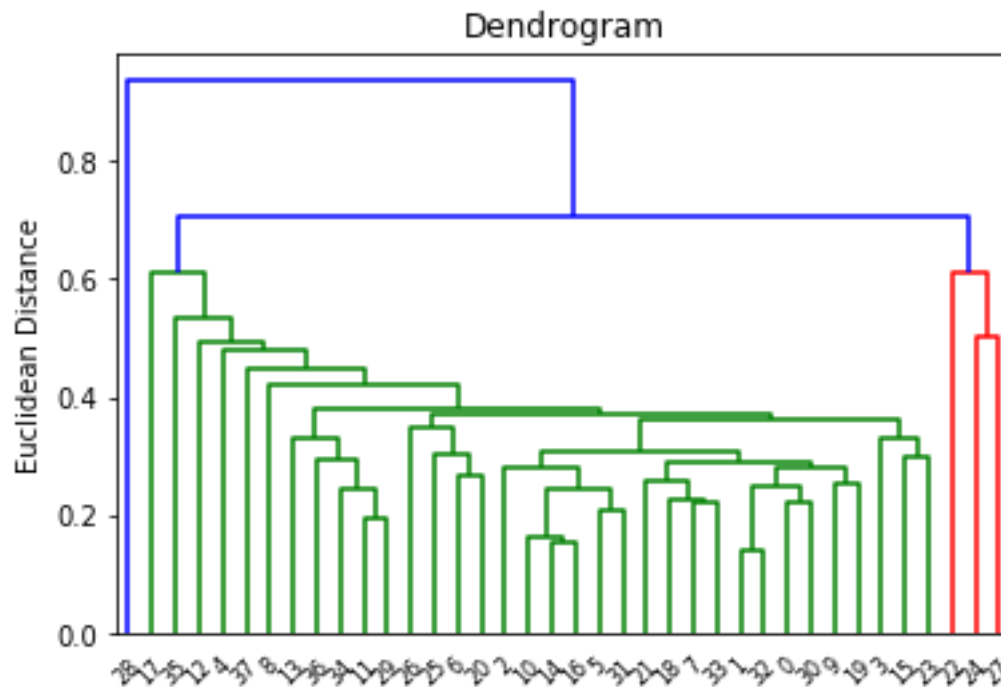
Out[17]:

	Neighborhood	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
0	Adelaide, King, Richmond	Steakhouse	Hotel	Asian Restaurant	Café	Coffee Shop	Seafood Restaurant	Bar	Plaza	Speakeasy	Pizza Place
1	Berczy Park	Café	Seafood Restaurant	Cocktail Bar	Farmers Market	Bakery	Belgian Restaurant	Restaurant	Liquor Store	Bistro	Beer Bar
2	Brockton, Exhibition Place, Parkdale Village	Breakfast Spot	Coffee Shop	Café	Gym	Convenience Store	Bar	Burrito Place	Stadium	Restaurant	Italian Restaurant
3	Business Reply Mail Processing Centre 969 Eastern	Light Rail Station	Butcher	Skate Park	Spa	Brewery	Farmers Market	Fast Food Restaurant	Restaurant	Recording Studio	Burrito Place
4	CN Tower, Bathurst Quay, Island airport, Harbo...	Airport Lounge	Airport Service	Airport Terminal	Harbor / Marina	Airport	Airport Food Court	Airport Gate	Boat or Ferry	Sculpture Garden	Boutique

Next I grouped all the Neighborhoods together based on the borough they belonged to and determined the top 8 venues for each Brorough.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	1	Coffee Shop	Health Food Store	Trail	Pub	Cuban Restaurant	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	1	Greek Restaurant	Italian Restaurant	Ice Cream Shop	Yoga Studio	Indian Restaurant	Pizza Place	Cosmetics Shop	Pub
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572	1	Gym	Pub	Sandwich Place	Brewery	Burger Joint	Burrito Place	Fast Food Restaurant	Fish & Chips Shop
43	M4M	East Toronto	Studio District	43.659526	-79.340923	1	Café	Coffee Shop	American Restaurant	Bakery	Italian Restaurant	Middle Eastern Restaurant	Ice Cream Shop	Sandwich Place
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	1	Bus Line	Gym / Fitness Center	Park	Swim School	Dance Studio	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant

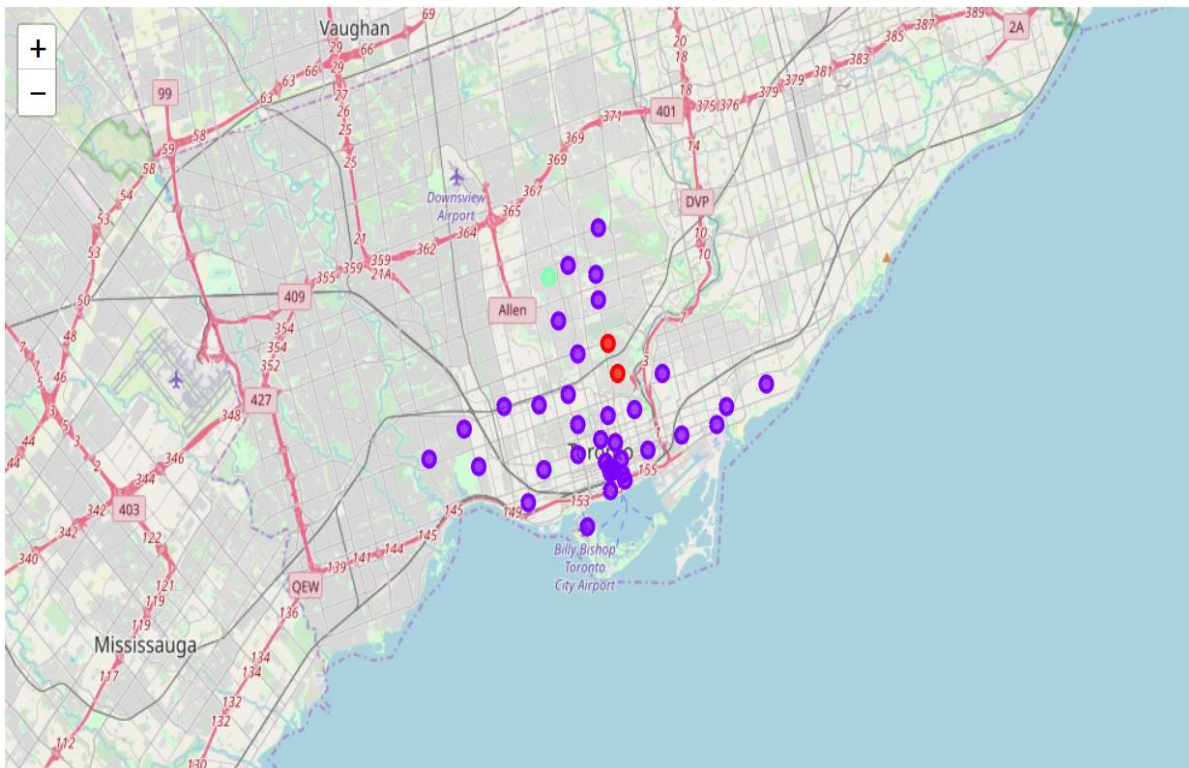
Using K-Means clustering, I then grouped all these Borough's into multiple clusters. I calculated the optimal K-Value using Agglomerative calculation and the elbow method.



I then assigned all the cluster labels to the Neighborhoods and created a new dataframe.

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	1	Coffee Shop	Health Food Store	Trail	Pub	Cuban Restaurant	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	1	Greek Restaurant	Italian Restaurant	Ice Cream Shop	Yoga Studio	Indian Restaurant	Pizza Place	Cosmetics Shop	Pub
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572	1	Gym	Pub	Sandwich Place	Brewery	Burger Joint	Burrito Place	Fast Food Restaurant	Fish & Chips Shop
43	M4M	East Toronto	Studio District	43.659526	-79.340923	1	Café	Coffee Shop	American Restaurant	Bakery	Italian Restaurant	Middle Eastern Restaurant	Ice Cream Shop	Sandwich Place
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	1	Bus Line	Gym / Fitness Center	Park	Swim School	Dance Studio	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant

Out [23]:



I then examined the generated clusters for further evaluation.

Examining some of the Clusters

```
In [68]: Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 0, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.shape[1]))]]
```

Out[68]:

	Borough	Cluster Labels	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
44	Central Toronto	0	Bus Line	Park	Swim School	Dance Studio	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant	Dog Run	Discount Store	Diner
50	Downtown Toronto	0	Park	Playground	Trail	Wings Joint	Cuban Restaurant	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant	Dog Run	Discount Store
64	Central Toronto	0	Bus Line	Park	Trail	Jewelry Store	Sushi Restaurant	Dance Studio	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant	Dog Run

```
In [69]: Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 1, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.shape[1]))]]
```

Out[69]:

	Borough	Cluster Labels	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
48	Central Toronto	1	Playground	Tennis Court	Cuban Restaurant	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant	Dog Run	Discount Store	Diner	Dessert Shop
63	Central Toronto	1	Garden	Home Service	Dance Studio	Falafel Restaurant	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant	Dog Run	Discount Store	Diner

	Borough	Cluster Labels	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
37	East Toronto	2	Coffee Shop	Health Food Store	Trail	Pub	Wings Joint	Dance Studio	Ethiopian Restaurant	Eastern European Restaurant	Dumpling Restaurant	Dog Run
41	East Toronto	2	Greek Restaurant	Ice Cream Shop	Italian Restaurant	Brewery	Bookstore	Restaurant	Spa	Juice Bar	Diner	Dessert Shop
42	East Toronto	2	Park	Ice Cream Shop	Pub	Brewery	Sandwich Place	Burger Joint	Burrito Place	Fast Food Restaurant	Fish & Chips Shop	Italian Restaurant
43	East Toronto	2	Café	Coffee Shop	American Restaurant	Bakery	Italian Restaurant	Comfort Food Restaurant	Cheese Shop	Bookstore	Sandwich Place	Juice Bar
45	Central Toronto	2	Gym	Pizza Place	Grocery Store	Park	Breakfast Spot	Burger Joint	Hotel	Food & Drink Shop	Sandwich Place	Deli / Bodega
46	Central Toronto	2	Coffee Shop	Sporting Goods Shop	Clothing Store	Shoe Store	Sandwich Place	Salon / Barbershop	Rental Car Location	Chinese Restaurant	Park	Yoga Studio
47	Central Toronto	2	Sandwich Place	Dessert Shop	Coffee Shop	Italian Restaurant	Café	Pizza Place	Sushi Restaurant	Pharmacy	Costume Shop	Brewery
49	Central Toronto	2	Coffee Shop	Pub	Pizza Place	Fried Chicken Joint	Medical Center	Sports Bar	Bagel Shop	Supermarket	Sushi Restaurant	Light Rail Station
51	Downtown Toronto	2	Restaurant	Coffee Shop	Café	Italian Restaurant	Gift Shop	Japanese Restaurant	Indian Restaurant	Pub	Bakery	Bank
52	Downtown Toronto	2	Gay Bar	Japanese Restaurant	Coffee Shop	Bubble Tea Shop	Salon / Barbershop	Burger Joint	Restaurant	Ramen Restaurant	Pub	Pizza Place
53	Downtown Toronto	2	Coffee Shop	Bakery	Park	Mexican Restaurant	Café	Pub	Breakfast Spot	Historic Site	Theater	Farmers Market
54	Downtown Toronto	2	Café	Clothing Store	Plaza	Japanese Restaurant	Sandwich Place	Diner	Ramen Restaurant	Burger Joint	Burrito Place	Shopping Mall
55	Downtown Toronto	2	Gastropub	Coffee Shop	Restaurant	Hotel	Italian Restaurant	Japanese Restaurant	Poke Place	Diner	Performing Arts Venue	Café

click to scroll output; double click to hide

Results:

Clustering helped me in assigning all the Boroughs to groups and determine the most popular venues across all the neighborhoods.

Based on the above information, and all the previous analysis, I determined that for an Apartment complex to have a good walking score, and good renter/owner engagement, the following venues were most necessary.

1. Café's
2. Grocery Store's.
3. Hiking Trails/Parks
4. Fitness Clubs.
5. Restaurants.

Based on this information, I revisited all the grouped Neighborhoods, and determined that the following neighborhoods offered the best possible locations to build an apartment complex.

1. The Danforth West, Riverdale
2. Studio District
3. Commerce Court, Victoria Hotel
4. Business Reply Mail Processing Centre 969 Eastern
5. Harbourfront, Regent Park

Discussions and Conclusions:

The dataset and the analysis I have done so far, can be expanded upon in many ways based on the requirements. If the Apartment will be targeting Families, then we can focus our efforts on getting school districts and areas with parks.

Another observation was that some neighborhoods are densely populated while some are sparse. Analysis can be done to determine why certain neighborhoods don't flourish and what can be done to improve them. This project gave me a lot of exposure to Data Science and Machine learning and I look forward to continuing this learning process.