The Battle of the Neighborhoods Report

Index

The Battle of the Neighborhoods Report

1.0 Introduction/Business Problem

The City of Toronto, is one of the most popular cities and one of the most populated cities in Canada. It is diverse and is the financial capital of Canada. There is tremendous interest for moving to the new houses in this city. I work for a realtor firm, one of our clients is looking to buy a property in this beautiful city. the client is retired couple. They are trying to find the best location to chose for their new home after retirement.

Various requirements that client requested for the surrounding of the area where they are looking to buy their new home:

-Parks

-Pet friendly places

-Drug Store Restaurants

-Coffee shops

-Trails

-Music friendly

The above are only some of the factors that can affect the location of the new home. There are some other factors such as hospital, neighbourhood. The insights derived from this analysis will give good understanding of the best fit for this client based on their requirements.

2.0 Data

- Data City of interest in this project: Toronto
- Following datasets for analysing Toronto city

Data 1

- Data from this location: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario. Only the first three characters are listed, corresponding to the Forward Sortation Area.

This dataset provides the list of Borough and neighbourhood associated with the postal codes of the city of Toronto.

We will use this data to categorize Boroughs and neighbourhood base on postal codes and then will merge with the foursquare data.

Below is an example of the data from Wikipedia link mentioned above:

26]:

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge,Malvern
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union
2	M1E	Scarborough	Guildwood,Morningside,West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Data 2

Toronto city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provide venues information for each neighborhood.

We will use the Foursquare API to explore neighborhoods in Toronto city. We will use this data to figure out which neighbourhood has as many as amenities and facilities that fits with the client's requirement.

We will cluster this data using neighbourhood to figure out which area best fits with the client requirements to figure out location to buy a new house

Below is an example of the data from foursquare: We will use this data to categorize Boroughs and neighbourhood base on postal codes and then will merge with the foursquare data
Below is an example of the data from Foursquare API. geographical data were also for Toronto from http://cocl.us/Geospatial_data.

Out[46]:

	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

3.0 Methodology

Objective of this study is to find the best neighborhood of Toronto city to buy a new house for retired client as per their requirement. This project will help us to define first footstep i.e. to find good location to start looking for a new house in quiet and nature blessed surrounding for client.

Toronto city has

Exploratory Data Analysis:

Toronto city geographical coordinates data

1. Various python libraries were imported to enable various data science functions
2. Imported data from Wikipedia URL related to city of Toronto https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
3. Imported geographical data for Toronto from http://cocl.us/Geospatial_data

Out[138]:

	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

4. Transformed the data of nested python dictionaries into pandas data frame
5. This data was used to get venues data from foursquare database using foursquare credentials

6. Data was formatted and edited as per the requirement of this project

7. Venue related information were converted into binary form (Yes/No (0/1) using dummies.

Out[115]:

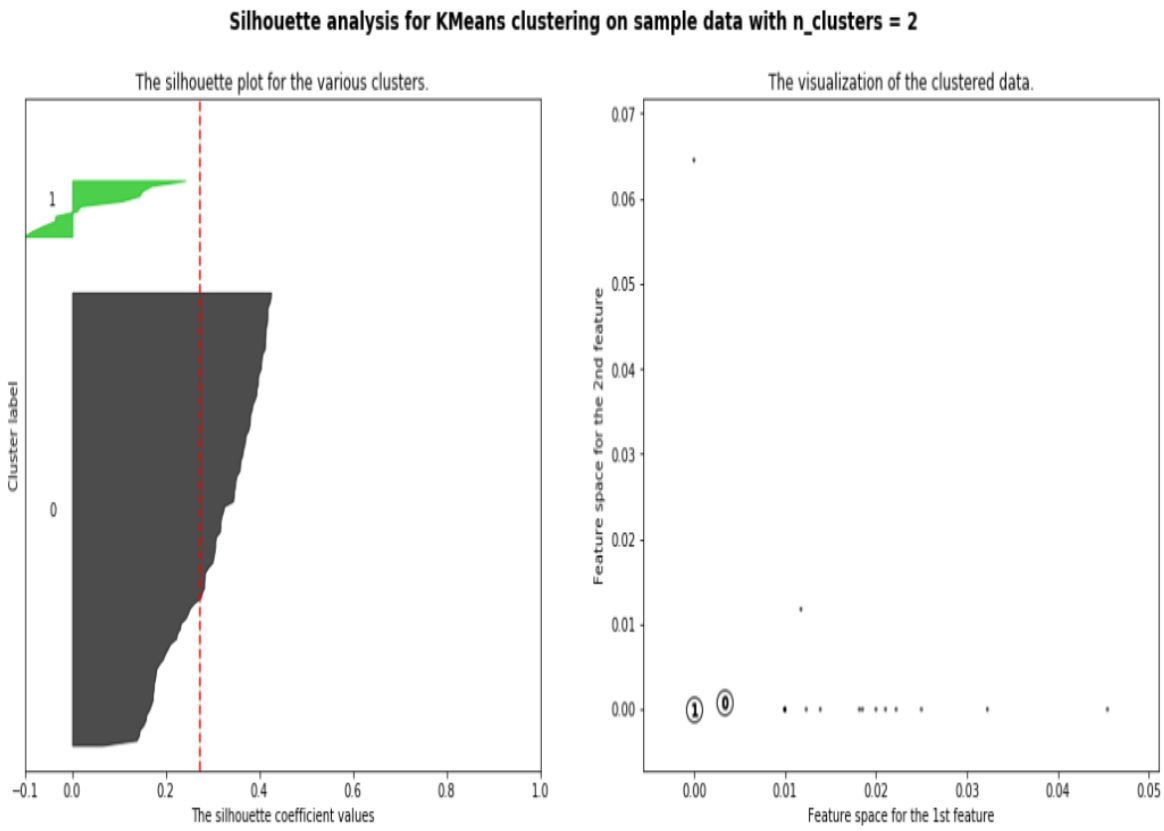
	Yoga Studio	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium	Arcade	Argentinian Restaurant	A Galle
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

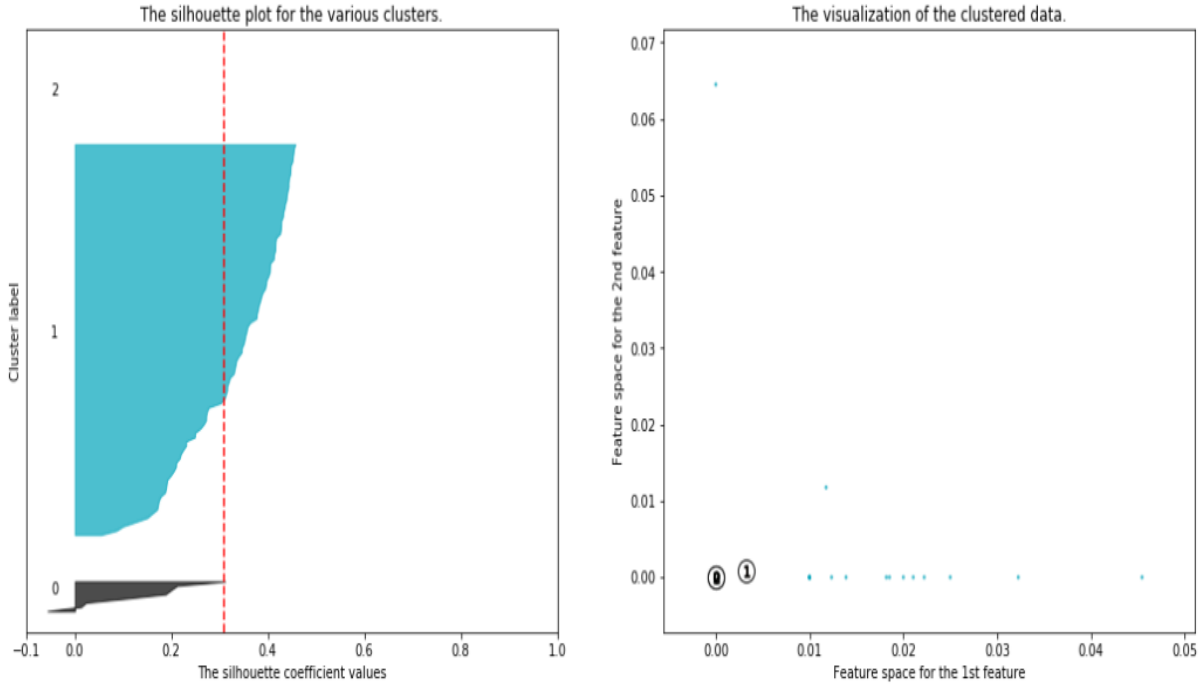
8. Data were grouped by neighbourhood

	Neighborhood	Yoga Studio	Accessories Store	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium	Arcade	A R
0	Adelaide,King,Richmond	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.02	0.0	0.0	0.0	0.0	0.0	
1	Agincourt	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	
2	Agincourt North,L'Amoreaux East,Milliken,Steel...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	
3	Albion Gardens,Beaumond Heights,Humbergate,Jam...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	
4	Alderwood,Long Branch	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.0	0.0	0.0	0.0	0.0	

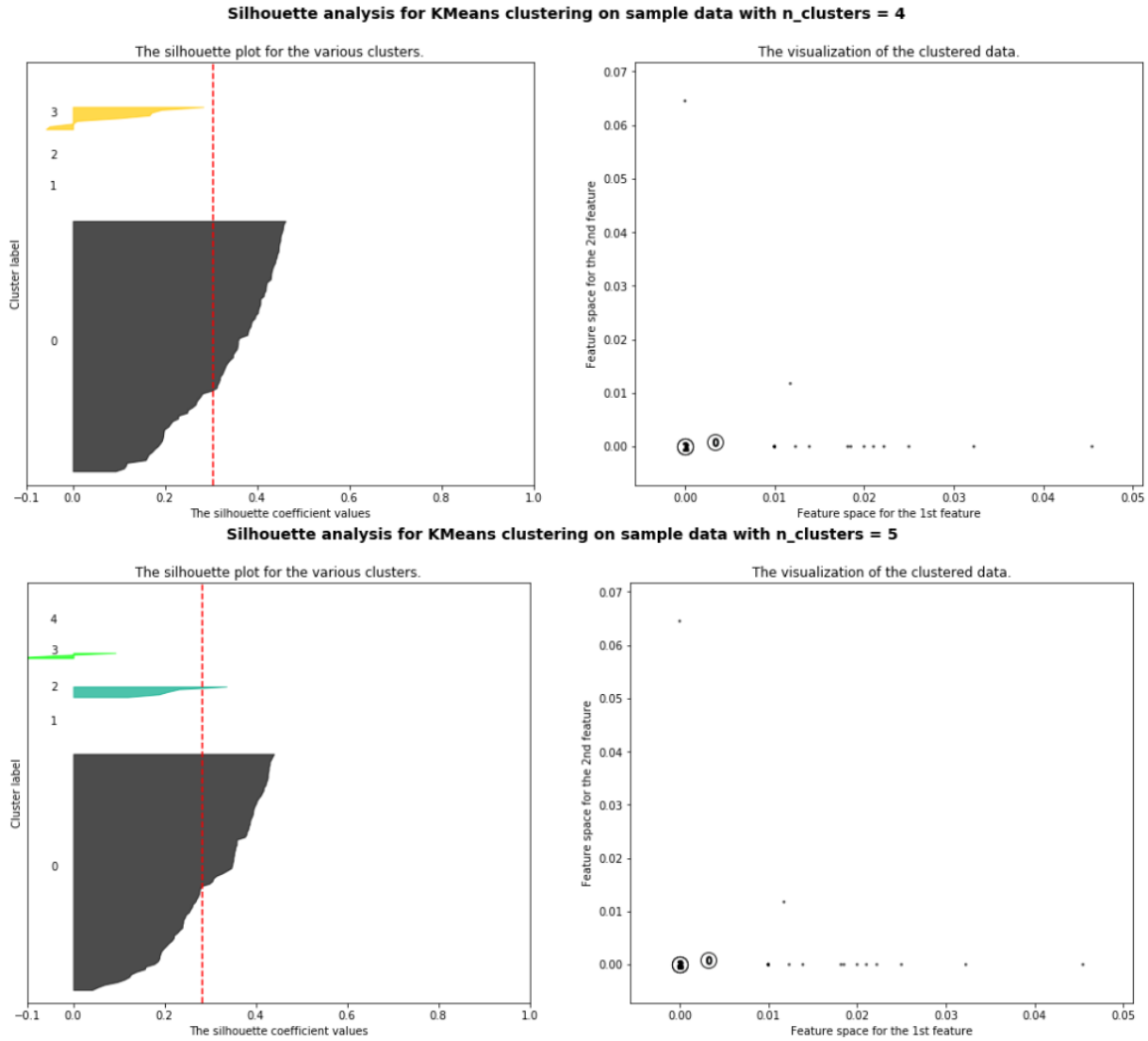
9. Top 5 most venues for each neighbourhood and 1st to 10th most common venues were identified for each neighbourhood

10. Executed analysis using KMeans clustering on sample data based on neighbourhoods and divided in 3 clusters


Silhouette analysis for KMeans clustering on sample data with n_clusters = 2
The silhouette plot for the various clusters.
Cluster label
1
0
−0.1
0.0
0.2
0.4
0.6
0.8
1.0
The silhouette coefficient values
The visualization of the clustered data.
Feature space for the 2nd feature
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.00
0.01
0.02
0.03
0.04
0.05
Feature space for the 1st feature


Silhouette analysis for KMeans clustering on sample data with n_clusters = 3
The silhouette plot for the various clusters.
Cluster label
2
1
0
−0.1
0.0
0.2
0.4
0.6
0.8
1.0
The silhouette coefficient values
The visualization of the clustered data.
Feature space for the 2nd feature
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.00
0.01
0.02
0.03
0.04
0.05
Feature space for the 1st feature

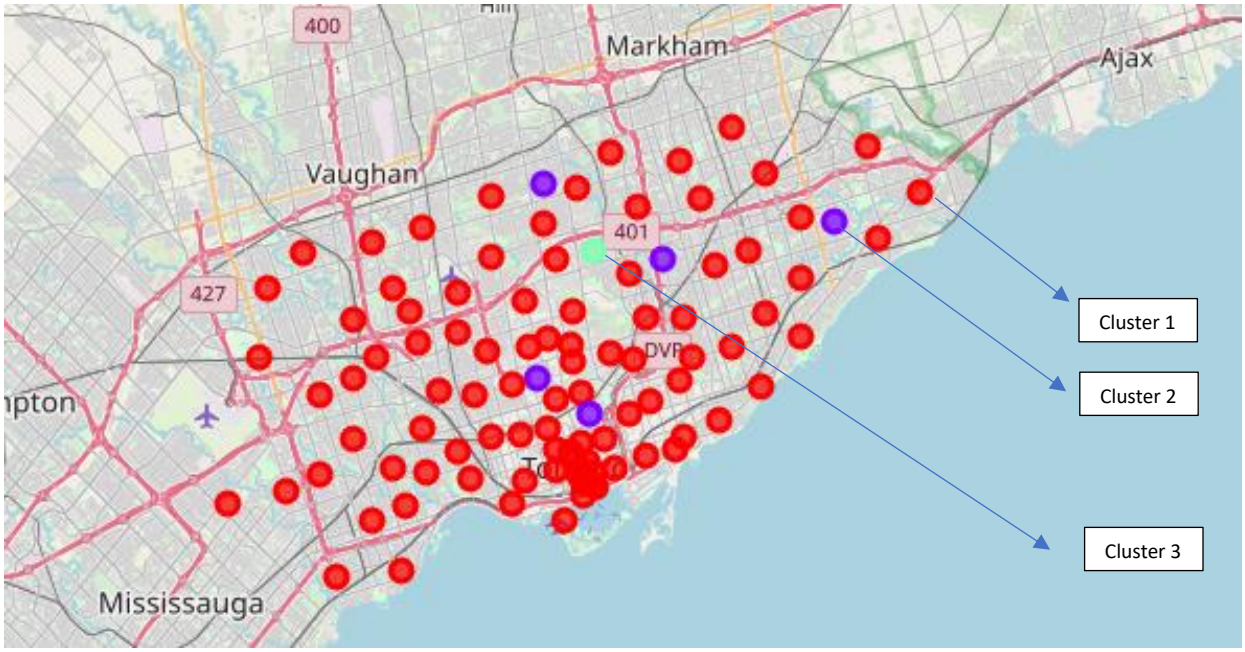
The Battle of the Neighborhoods Report



11. Merged merge grouped venue data with Toronto demographic data to add latitude/longitude for each neighborhood and added cluster labels

	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
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353	0.0	Fast Food Restaurant	Hobby Shop	Business Service	Coffee Shop	Spa	Bus Station	Construction & Landscaping	Drugstore
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	0.0	Breakfast Spot	Burger Joint	Bar	Women's Store	Dive Bar	Dog Run	Doner Restaurant	Donut Shop
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711	0.0	Pizza Place	Fast Food Restaurant	Thrift / Vintage Store	Park	Rental Car Location	Moving Target	Bus Line	Electronics Store
3	M1G	Scarborough	Woburn	43.770992	-79.216917	1.0	Coffee Shop	Park	Business Service	Women's Store	Drugstore	Diner	Discount Store	Dive Bar
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0.0	Bakery	Coffee Shop	Indian Restaurant	Thai Restaurant	Caribbean Restaurant	Gas Station	Fried Chicken Joint	Athletics & Sports

12. Neighbourhoods are grouped into 3 clusters based on the most common venues



Out[149]:

	Cluster1	Cluster2	Cluster3
0	Rouge,Malvern		
1	Highland Creek,Rouge Hill,Port Union		
2	Guildwood,Morningside,West Hill		
3		Woburn	
4	Cedarbrae		
5	Scarborough Village		
6	East Birchmount Park,Ionview,Kennedy Park		
7	Clairlea,Golden Mile,Oakridge		
8	Cliffcrest,Cliffside,Scarborough Village West		
9	Birch Cliff,Cliffside West		

13. Various Python libraries were used for analysis:

```
import numpy as np
import pandas as pd
import json
from geopy.geocoders import Nominatim
import requests
from pandas.io.json import json_normalize
import matplotlib.cm as cm
import matplotlib.colors as colors
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import folium
from bs4 import BeautifulSoup etc.
```

4.0 Results and Discussions

Cluster 1

```
df1=toronto_cl1.loc[toronto_cl1['Cluster Labels'] == 0,toronto_cl1.columns[[2] + list(range(5, toronto_cl1.shape[1]))]]
df1
```

[146]:

	Neighborhood	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
0	Rouge,Malvern	0	Fast Food Restaurant	Hobby Shop	Business Service	Coffee Shop	Spa	Bus Station	Construction & Landscaping	Drugstore	Discount Store	Dive Bar
1	Highland Creek,Rouge Hill,Port Union	0	Breakfast Spot	Burger Joint	Bar	Women's Store	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Drugstore	Dumpling Restaurant
2	Guildwood,Morningside,West Hill	0	Pizza Place	Fast Food Restaurant	Thrift / Vintage Store	Park	Rental Car Location	Moving Target	Bus Line	Electronics Store	Mexican Restaurant	Breakfast Spot
4	Cedarbrae	0	Bakery	Coffee Shop	Indian Restaurant	Thai Restaurant	Caribbean Restaurant	Gas Station	Fried Chicken Joint	Athletics & Sports	Asian Restaurant	Chinese Restaurant
5	Scarborough Village	0	Fast Food Restaurant	Coffee Shop	Pizza Place	Convenience Store	Farm	Dumpling Restaurant	Discount Store	Dive Bar	Dog Run	Doner Restaurant

Neighbourhoods in cluster 1 has many fast food and speciality restaurants and coffee shops. Pretty much clustered through out the whole city of Toronto. There are very less parks and natural places in this neighbourhood. These places seem to be well populated and business friendly.

Cluster 2

```
df2=toronto_cl1.loc[toronto_cl1['Cluster Labels'] == 1,toronto_cl1.columns[[2] + list(range(5, toronto_cl1.shape[1]))]]
df2
```

147]:

	Neighborhood	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
3	Woburn	1	Coffee Shop	Park	Business Service	Women's Store	Drugstore	Diner	Discount Store	Dive Bar	Dog Run	Doner Restaurant
20	Newtonbrook,Willowdale	1	Park	Coffee Shop	Music Venue	Trail	Drugstore	Diner	Discount Store	Dive Bar	Dog Run	Doner Restaurant
24	Parkwoods	1	Park	Food & Drink Shop	Pet Store	Women's Store	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Dog Run	Doner Restaurant
49	Rosedale	1	Park	Playground	Trail	Drugstore	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Dog Run	Doner Restaurant
63	Forest Hill North,Forest Hill West	1	Park	Jewelry Store	Trail	Sushi Restaurant	Women's Store	Drugstore	Diner	Discount Store	Dive Bar	Dog Run

Neighbourhoods in cluster contains many parks and trails. There are only few assorted neighbourhoods in this cluster. These places seem to be less populated and quiet.

Cluster 3

```
df3=toronto_cl1.loc[toronto_cl1['Cluster Labels'] == 2,toronto_cl1.columns[[2] + list(range(5, toronto_cl1.shape[1]))]]
df3
```

[148]:

	Neighborhood	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
19	Silver Hills,York Mills	2	Cafeteria	College Rec Center	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Drugstore

Cluster 3 represents only unique neighbourhood with unique venues which are not common in cluster 1 and cluster 3.

5.0 Conclusion:

As per the above analysis, it is clear that the neighbourhoods listed in cluster 2 best suites for client to chose a location that meets client's requirement since parks, coffee shops, drug stores and trails are the amongst the top 5 most common venues around this neighbourhood.

Woburn

Newtonbrook

Willowdale

Parkwoods

Rosedale

Forest Hill North

Forest Hill West

However, noticed the Hospital is not a most common venue in any of these neighbourhood.

According to this analysis it is evident that the neighbourhood of Newtonbrook and Willowdale (North York/Postal code:M2) is the most fit location for this client since all venues that client would want for their new house location are common in this neighbourhood

The best suited location to buy new house for client:

Newtonbrook and Willowdale (North York/Postal code:M2)