

Segmenting Neighbourhood in Toronto on Restaurant variety

1. Introduction

1.1 Background -

Toronto is big city with multiple Boroughs and each Borough has multiple neighbourhood. It is not easy for a businessman who is looking to start a food or liquor business to select suitable area. Even for tourist or foodies also would like to get areas with good restaurants. Party lover people would look for bar or pub areas for chilling outs.

1.2 Description of problem -

This study is about to find out localities which will be suitable for food or liquor business. Also, it will give good picture to food lover or party lover peoples for their destination for lunch, breakfast, diner or for party.

1.3 Solutions -

We have taken top 10 categories from each neighbour. Out of all top 10 categories food and bar categories are selected. All categories are segregated into Restaurant only, tea, coffee shop or light food shop and bar or pub. We have separated each neighbourhood to few clustered bases on the availability of these super category types. Charts and graph are included to figure out high, moderate or low restaurant or bar zone area which any person can use for their own purpose.

2. Data acquisition and cleaning

2.1 Data Source

All data about Toronto's neighbourhood are sourced from Cousera and location of latitude and longitude also sourced from Cousera. Details of neighbourhood like venue details, venues category etc are sourced from foursquare api(api.foursquare.com).

2.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There is duplicate entries which is removed.

For few borough areas, no neighbourhood is assigned. So those entries are removed as well.

There are multiple neighbourhoods available for same postal code. So all neighbourhoods are combined if they shared same postal code.

Data sample for Toronto neighbourhood

	Postcode	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M6A	North York	Lawrence Heights
6	M6A	North York	Lawrence Manor
7	M7A	Downtown Toronto	Queen's Park

Data sample of coordinate data

	Postal Code	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

2.3 Features Selection

Our intention is to get mainly details on any kind of food venues like restaurants, liquor shops, bar, fast food shop or tea and coffee shop. So other venues details return from foursquare api are excluded from analysis.

Included Lists

```
In [48]: np.unique(np.array(snack_list))
```

```
Out[48]: array(['Bakery', 'Breakfast Spot', 'Bubble Tea Shop', 'Burger Joint',  
               'Burrito Place', 'Cheese Shop', 'Coffee Shop', 'Dessert Shop',  
               'Donut Shop', 'Fried Chicken Joint', 'Ice Cream Shop',  
               'Pizza Place', 'Salad Place', 'Steakhouse', 'Tea Room'],  
              dtype='<U19')
```

```
In [49]: np.unique(np.array(rest_list))
```

```
Out[49]: array(['Airport Food Court', 'American Restaurant', 'Asian Restaurant',  
               'Café', 'Cajun / Creole Restaurant', 'Chinese Restaurant',  
               'Cuban Restaurant', 'Dim Sum Restaurant', 'Diner',  
               'Doner Restaurant', 'Dumpling Restaurant',  
               'Eastern European Restaurant', 'Empanada Restaurant',  
               'Ethiopian Restaurant', 'Fast Food Restaurant',  
               'Food & Drink Shop', 'Greek Restaurant', 'Health Food Store',  
               'Indian Restaurant', 'Italian Restaurant', 'Japanese Restaurant',  
               'Mexican Restaurant', 'Middle Eastern Restaurant',  
               'Portuguese Restaurant', 'Restaurant', 'Seafood Restaurant',  
               'Sushi Restaurant', 'Thai Restaurant',  
               'Vegetarian / Vegan Restaurant', 'Vietnamese Restaurant'],  
              dtype='<U29')
```

```
In [50]: np.unique(np.array(bar_list))
```

```
Out[50]: array(['Bar', 'Beer Bar', 'Beer Store', 'Brewery', 'Cocktail Bar',  
               'Gastropub', 'Gay Bar', 'Juice Bar', 'Liquor Store', 'Nightclub',  
               'Pub', 'Salon / Barbershop', 'Wine Bar'], dtype='<U18')
```

Discarded lists

```
In [51]: np.unique(np.array(non_rest_list))

Out[51]: array(['Airport', 'Airport Gate', 'Airport Lounge', 'Airport Service',
                'Airport Terminal', 'Aquarium', 'Auto Workshop', 'Baby Store',
                'Beach', 'Bookstore', 'Boutique', 'Bus Line', 'Candy Store',
                'Climbing Gym', 'Clothing Store', 'Comic Shop', 'Cosmetics Shop',
                'Dance Studio', 'Deli / Bodega', 'Department Store',
                'Discount Store', 'Distribution Center', 'Dog Run',
                'Electronics Store', 'Farmers Market', 'Flea Market',
                'Frozen Yogurt Shop', 'Furniture / Home Store', 'Garden',
                'Garden Center', 'Gas Station', 'Gift Shop', 'Grocery Store',
                'Gym', 'Gym / Fitness Center', 'Harbor / Marina', 'Hotel',
                'Intersection', 'Jewelry Store', 'Light Rail Station', 'Lounge',
                'Men's Store', 'Movie Theater', 'Music Venue', 'Neighborhood',
                'Park', 'Pet Store', 'Pharmacy', 'Playground', 'Plaza', 'Pool',
                'Rental Car Location', 'Sandwich Place', 'Scenic Lookout',
                'Shoe Store', 'Spa', 'Speakeasy', 'Sporting Goods Shop', 'Stadium',
                'Summer Camp', 'Supermarket', 'Swim School', 'Theater', 'Trail',
                'Yoga Studio'], dtype='<U22')
```

3. Exploratory Data Analysis

3.1 Combine both neighbourhood and coordinate data

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude
0	M4E	East Toronto	The Beaches	M4E	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West,Riverdale	M4K	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West,India Bazaar	M4L	43.668999	-79.315572
3	M4M	East Toronto	Studio District	M4M	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	M4N	43.728020	-79.388790
5	M4P	Central Toronto	Davisville North	M4P	43.712751	-79.390197
6	M4R	Central Toronto	North Toronto West	M4R	43.715383	-79.405678

3.2 Select only Borough started containing Toronto

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude
0	M4E	East Toronto	The Beaches	M4E	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West,Riverdale	M4K	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West,India Bazaar	M4L	43.668999	-79.315572
3	M4M	East Toronto	Studio District	M4M	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	M4N	43.728020	-79.388790
5	M4P	Central Toronto	Davisville North	M4P	43.712751	-79.390197
6	M4R	Central Toronto	North Toronto West	M4R	43.715383	-79.405678
7	M4S	Central Toronto	Davisville	M4S	43.704324	-79.388790
8	M4T	Central Toronto	Moore Park,Summerhill East	M4T	43.689574	-79.383160
9	M4V	Central Toronto	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	M4V	43.686412	-79.400049
10	M4W	Downtown Toronto	Rosedale	M4W	43.679563	-79.377529
11	M4X	Downtown Toronto	Cabbagetown,St. James Town	M4X	43.667967	-79.367675

3.3 Get venue details

Get venue name, venue category, venue's coordinate

	Postcode	Borough	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	M4E	East Toronto	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	M4E	East Toronto	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	M4E	East Toronto	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
5	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	MenEssentials	43.677820	-79.351265	Cosmetics Shop
6	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Mezes	43.677962	-79.350196	Greek Restaurant
7	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Cafe Fiorentina	43.677743	-79.350115	Italian Restaurant
8	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Dolce Gelato	43.677773	-79.351187	Ice Cream Shop
9	M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	Moksha Yoga Danforth	43.677622	-79.352116	Yoga Studio

3.4 Get counts for total venues from each neighbourhood

Postcode	Borough	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
M4E	East Toronto	The Beaches	4	4	4	4	4	4
M4K	East Toronto	The Danforth West,Riverdale	41	41	41	41	41	41
M4L	East Toronto	The Beaches West,India Bazaar	19	19	19	19	19	19
M4M	East Toronto	Studio District	41	41	41	41	41	41
M4N	Central Toronto	Lawrence Park	3	3	3	3	3	3
M4P	Central Toronto	Davisville North	7	7	7	7	7	7
M4R	Central Toronto	North Toronto West	21	21	21	21	21	21
M4S	Central Toronto	Davisville	38	38	38	38	38	38

3.5 Get frequency of each venues for each neighbourhood

Neighbourhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Workshop	BBQ Joint	Baby Store	Bagel Shop	Ba
0 Adelaide,King,Richmond	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.02	0.0	0.0	0.000000	0.0	0.02	0.0	0.0000	0.000000	0.0	0.000000	0.02
1 Berczy Park	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0	0.0	0.017857	0.0	0.00	0.0	0.0000	0.017857	0.0	0.017857	0.03
2 Brockton,Exhibition Place,Parkdale Village	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0000	0.000000	0.0	0.000000	0.04
3 Business Reply Mail Processing Centre 969	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.0	0.0	0.000000	0.0	0.00	0.0	0.0625	0.000000	0.0	0.000000	0.00

3.6 Get most top common venues from each neighbourhood

----Adelaide,King,Richmond----		
	venue	freq
0	Coffee Shop	0.07
1	Restaurant	0.05
2	Thai Restaurant	0.04
3	Café	0.04
4	Steakhouse	0.03
----Berczy Park----		
	venue	freq
0	Coffee Shop	0.09
1	Cocktail Bar	0.05
2	Cheese Shop	0.04
3	Farmers Market	0.04
4	Bakery	0.04

3.7 Identify top 10 venues from each neighbourhood

	Neighbourhood	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	Coffee Shop	Restaurant	Thai Restaurant	Café	Bar	Steakhouse	Sushi Restaurant	Hotel	Bakery	Gastropub
1	Berczy Park	Coffee Shop	Cocktail Bar	Café	Farmers Market	Bakery	Restaurant	Beer Bar	Cheese Shop	Seafood Restaurant	Beach
2	Brockton,Exhibition Place,Parkdale Village	Café	Breakfast Spot	Coffee Shop	Climbing Gym	Burrito Place	Japanese Restaurant	Italian Restaurant	Restaurant	Stadium	Intersection
3	Business Reply Mail Processing Centre 969 Eastern	Yoga Studio	Auto Workshop	Garden Center	Garden	Fast Food Restaurant	Farmers Market	Light Rail Station	Comic Shop	Pizza Place	Restaurant
	CN Tower,Bathurst Quay,Island						Agent Exam				Beach Park

3.8 Categorize each venue

Out of top 10 venues of each neighbourhoods, select restaurants or bars or foods venues and drop others.

Categorize those venues mainly into

- Restaurants
- Small restaurants which covers breakfast points, ice-cream shop, tea and coffee shop
- Bar, pub or liquor shops

	Neighbourhood	Restaurant_yes	Bar_yes	Snack_yes
0	Adelaide,King,Richmond	27	8	20
1	Berczy Park	12	7	11
2	Brockton,Exhibition Place,Parkdale Village	6	1	6
3	Business Reply Mail Processing Centre 969 Eastern	2	1	2
4	CN Tower,Bathurst Quay,Island airport,Harbourf...	1	1	1
5	Cabbagetown,St. James Town	13	5	8
6	Central Bay Street	23	6	27
7	Chinatown,Grange Park,Kensington Market	27	13	21
8	Christie	6	1	1
9	Church and Wellesley	27	11	17
10	Commerce Court,Victoria Hotel	33	10	20
11	Davisville	14	1	8
12	Davisville North	1	0	1

4. Predictive Modelling

As our aim is to find out suitable zones for restaurants business or finding out good place for food lover to go out, we will cluster out neighbours into few segments base on current venues. So clustering model is suitable for us.

We will cluster base on few clustering techniques and identify suitable one.

4.1 DBSCAN Clustering techniques

All neighbours are clustered into a single segments. So this one is not suitable for us.

```
: # dbscan
#toronto_grouped_clustering_dbs = toronto_grouped.drop('Neighbourhood', 1)
toronto_grouped_clustering_dbs = toronto_super_total.drop('Neighbourhood', 1)
db_scan=DBSCAN(eps=0.15,min_samples=10).fit(toronto_grouped_clustering_dbs)

: unique, counts = np.unique(db_scan.labels_, return_counts=True)
  for i in range(0,len(unique)):
    print('Label: {}, Count: {}'.format(unique[i],counts[i]))
```

Label: -1, Count: 39

4.2 Agglomerative Clustering

Agglomerative clustering identified that neighbours can be segmented into 3 cluster. This is a potential solution we are looking for. We will see another technique and figure out which one is ideal.

```
aggom_cluster = 3
toronto_grouped_clustering_agm = toronto_super_total.drop('Neighbourhood', 1)
aggglom = AgglomerativeClustering(n_clusters=aggom_cluster, linkage='complete').fit(toronto_grouped_clustering_agm)
```

aggglom.labels_

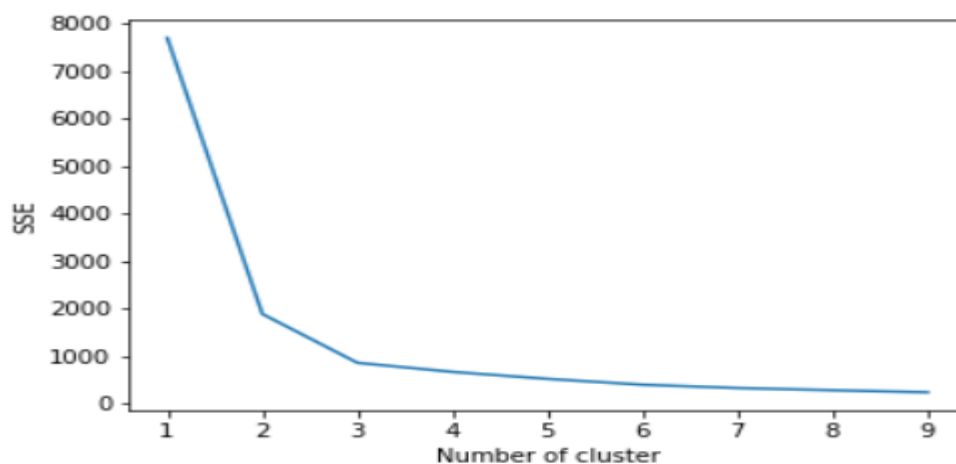
```
2]: array([0, 1, 2, 2, 2, 1, 0, 0, 2, 0, 0, 1, 2, 2, 0, 2, 0, 2, 1, 1, 0, 1,
          2, 1, 2, 2, 2, 1, 2, 2, 1, 0, 0, 0, 1, 2, 2, 2, 1])
```

```
unique, counts = np.unique(aggglom.labels_, return_counts=True)
for i in range(0,len(unique)):
    print('Label: {}, Count: {}'.format(unique[i],counts[i]))
```

Label: 0, Count: 11
Label: 1, Count: 11
Label: 2, Count: 17

4.3 Kmeans Clustering

Base on Kmeans cluster 2 cluster are identified.



```
unique, counts = np.unique(kmeans.labels_, return_counts=True)
for i in range(0,len(unique)):
    print('Label: {}, Count: {}'.format(unique[i],counts[i]))
```

Label: 0, Count: 38

Label: 1, Count: 1

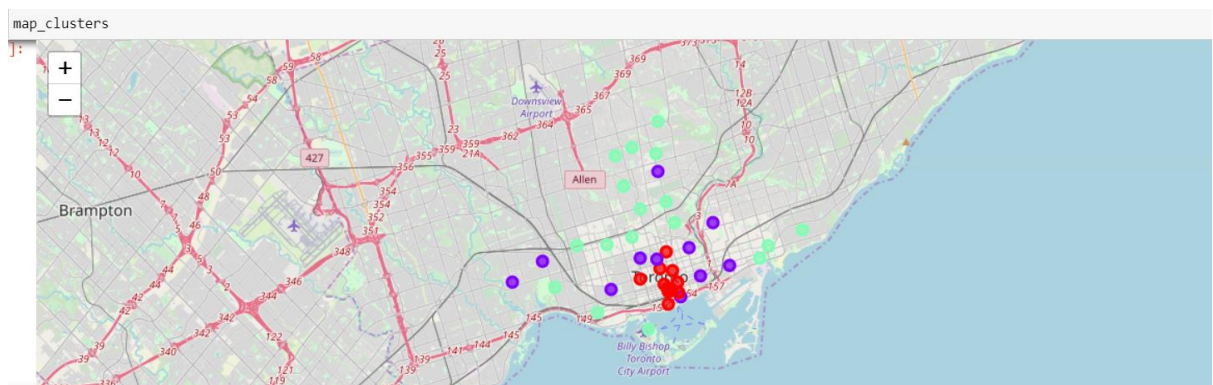
So base on above clustering it is obvious that **Agglomerative Clustering** is better suitable here.

Added cluster to each neighbourhood

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude	Cluster Labels	Restaurant_yes	Bar_yes	Snack_yes
0	M4E	East Toronto	The Beaches	M4E	43.676357	-79.293031	2	1	1	0
1	M4K	East Toronto	The Danforth West,Riverdale	M4K	43.679557	-79.352188	1	15	4	10
2	M4L	East Toronto	The Beaches West,India Bazaar	M4L	43.668999	-79.315572	2	4	3	4
3	M4M	East Toronto	Studio District	M4M	43.659526	-79.340923	1	12	6	7
4	M4N	Central Toronto	Lawrence Park	M4N	43.728020	-79.388790	2	0	0	0

5 Solutions

5.1 Map visualization



5.2 Analysis of cluster

Cluster-0

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude	Cluster Labels	Restaurant_yes	Bar_yes	Snack_yes
12	M4Y	Downtown Toronto	Church and Wellesley	M4Y	43.665860	-79.383160	0	27	11	17
14	M5B	Downtown Toronto	Ryerson,Garden District	M5B	43.657162	-79.378937	0	24	6	21
15	M5C	Downtown Toronto	St. James Town	M5C	43.651494	-79.375418	0	28	9	15
17	M5G	Downtown Toronto	Central Bay Street	M5G	43.657952	-79.387383	0	23	6	27
18	M5H	Downtown Toronto	Adelaide,King,Richmond	M5H	43.650571	-79.384568	0	27	8	20
19	M5J	Downtown Toronto	Harbourfront East,Toronto Islands,Union Station	M5J	43.640816	-79.381752	0	18	8	25
20	M5K	Downtown Toronto	Design Exchange,Toronto Dominion Centre	M5K	43.647177	-79.381576	0	23	10	27

Cluster -1

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude	Cluster Labels	Restaurant_yes	Bar_yes	Snack_yes
1	M4K	East Toronto	The Danforth West,Riverdale	M4K	43.679557	-79.352188	1	15	4	10
3	M4M	East Toronto	Studio District	M4M	43.659526	-79.340923	1	12	6	7
7	M4S	Central Toronto	Davisville	M4S	43.704324	-79.388790	1	14	1	8
11	M4X	Downtown Toronto	Cabbagetown,St. James Town	M4X	43.667967	-79.367675	1	13	5	8
13	M5A	Downtown Toronto	Harbourfront	M5A	43.654260	-79.360636	1	9	4	14
16	M5E	Downtown Toronto	Berczy Park	M5E	43.644771	-79.373306	1	12	7	11
25	M5S	Downtown Toronto	Harbord,University of Toronto	M5S	43.663806	-79.400040	1	13	6	4

Cluster -2

	Postcode	Borough	Neighbourhood	Postal Code	Latitude	Longitude	Cluster Labels	Restaurant_yes	Bar_yes	Snack_yes
0	M4E	East Toronto	The Beaches	M4E	43.676357	-79.293031	2	1	1	0
2	M4L	East Toronto	The Beaches West,India Bazaar	M4L	43.668999	-79.315572	2	4	3	4
4	M4N	Central Toronto	Lawrence Park	M4N	43.728020	-79.388790	2	0	0	0
5	M4P	Central Toronto	Davisville North	M4P	43.712751	-79.390197	2	1	0	1
6	M4R	Central Toronto	North Toronto West	M4R	43.715383	-79.405678	2	6	1	3
8	M4T	Central Toronto	Moore Park,Summerhill East	M4T	43.689574	-79.383160	2	1	0	0
9	M4V	Central Toronto	Don Mills,Forest Hill,St. Patrick's,South Hill	M4V	43.688440	-79.400040	2	4	2	4

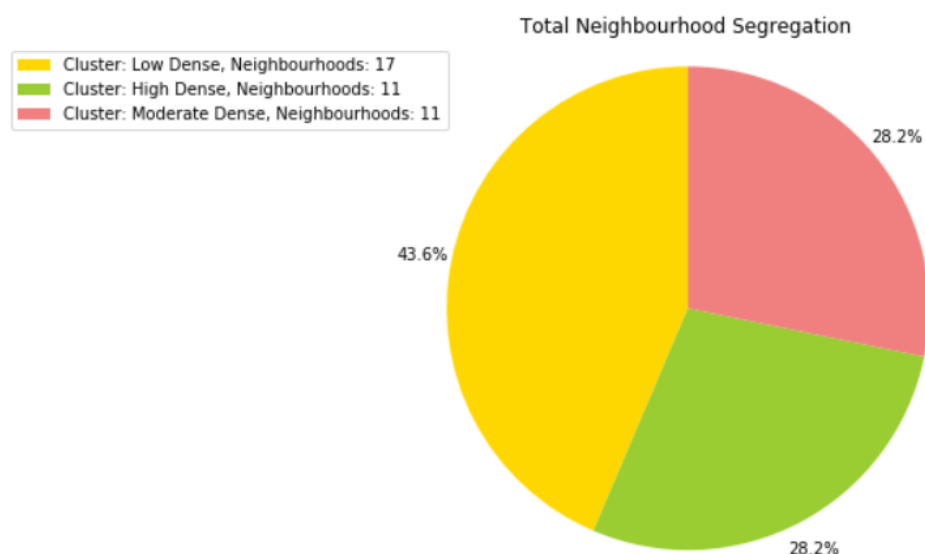
From above below points are obvious

Cluster-0: High Dense Restaurant zone

Cluster-1: Moderate Dense Restaurant zone

Cluster-2: Low Dense Restaurant zone

Below pie chart will depicts the % of each zone

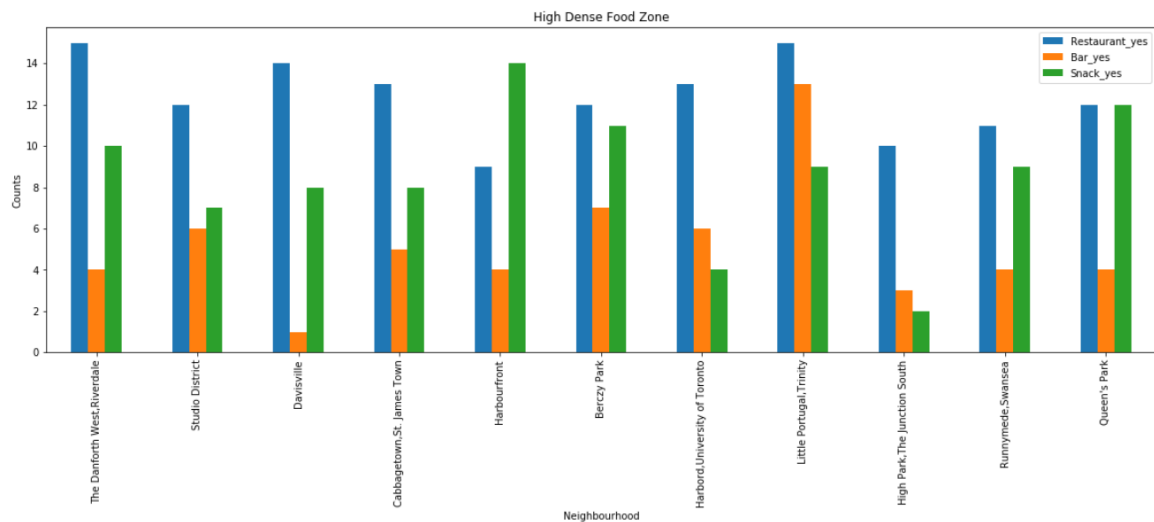


5.3 High Dense restaurant zone

Businessman looking for opening business can consider these area for opening restaurants as most of the people prefer to go those area.

Food lover also would like to go for this zone due to high availability of restaurants or bar or mini restaurants point

Below chart will show counts of venues for each high dense neighbour



Below venues are top venues in high dense neighbours

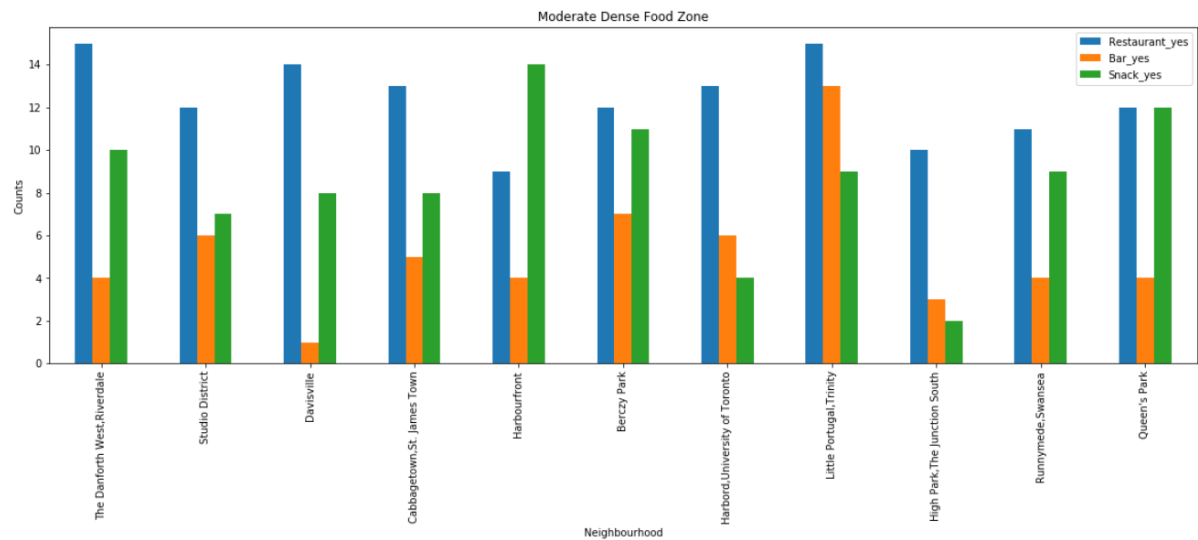


5.4 Moderate Dense Area

These neighbours have few restaurants. Few businessmen can prefer this areas base on business requirements.

People who wants to avoid dense zone can go out for party or lunch or dinner here.

Counts of venues from each neighbour



Top venues

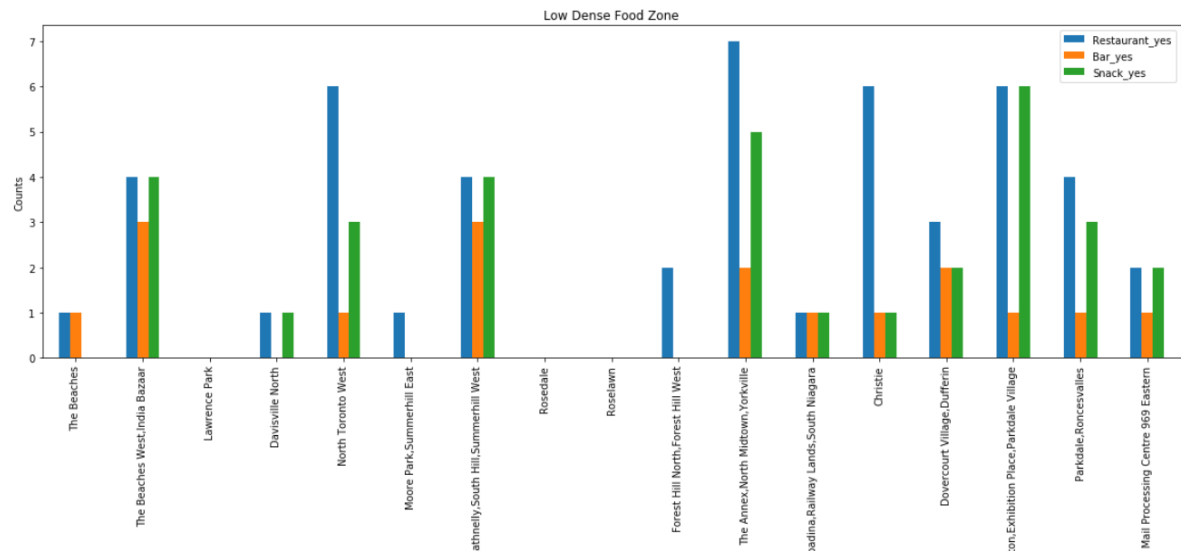


5.5 Low Dense area

There are still scope of improvements for restaurants or liquor business. Person who wants to take risk can open business considering it will grow in future.

Food lover wouldn't like this zone as less availability of food points or liquor points.

Counts of venues from each neighbour



Trending venues



6 Conclusion

In this study, I have analysed and categorized areas base on availability of restaurants, bars or tea & coffee shop or small food shop. People who are interested in business or like to go for party or outing for diner or lunch or breakfast use this analysis and can decide where to go base on their preferences. Even people who would like to start a business can decide where to start business base on their business plan.