Segmenting Neighbourhood in Toronto on Restaurant variety

Food zone in Toronto - Introduction

- 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.

Food zone Problem and Solutions

- 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.

- 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.

Data acquisition and cleaning

- 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).
- 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.
- Sample data

	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

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.
- Venues related to airport, music, movie, pharmacy, Department Store, Furnitures, gym , beach etc are excluded

Top Most Venues from each Neighbours

- Only venues which are available to most of the places within neighbours are analysed
- Sample

```
----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
```

10 venues from each neighbourhood

- Top 10 venues which are available in most of the parts are taken into consideration

	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						Airport Food				Rental Car

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

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.

Clustering techniques

- **DBSCAN Clustering techniques**

Clustered into a single segments.

So this one is not suitable for us.

- **Agglomerative Clustering**

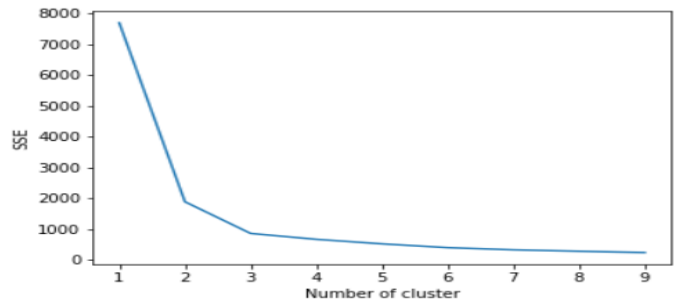
Clustered into a single segments.

Ideal for our solutions

- **Kmeans Clustering**

2 cluster are identified

Not suitable for our solution




```
# 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
```

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

```
agglom.labels_
```

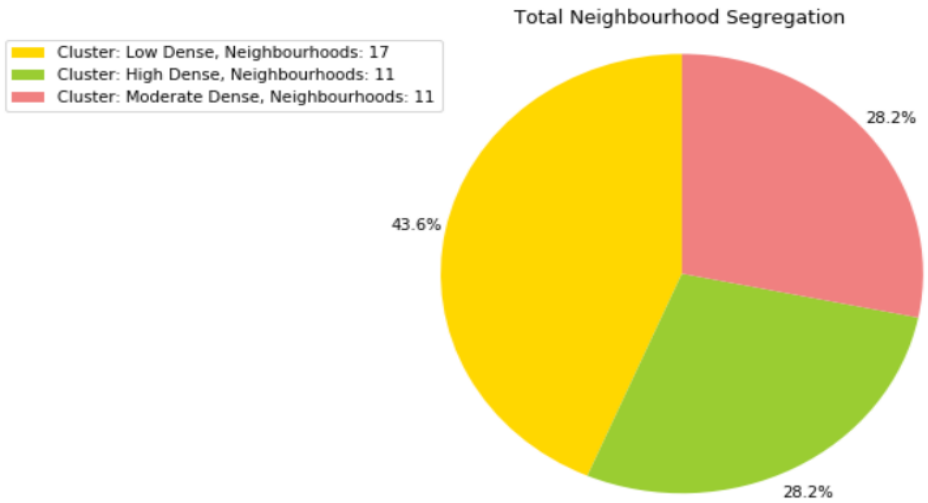
```
array([0, 1, 2, 2, 2, 1, 0, 0, 2, 0, 0, 1, 2, 2, 0, 2, 0, 2, 1, 1, 0, 1,
       2, 1, 2, 2, 2, 1, 2, 2, 1, 0, 0, 0, 1, 2, 2, 2, 1])
```

```
unique, counts = np.unique(agglom.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
```

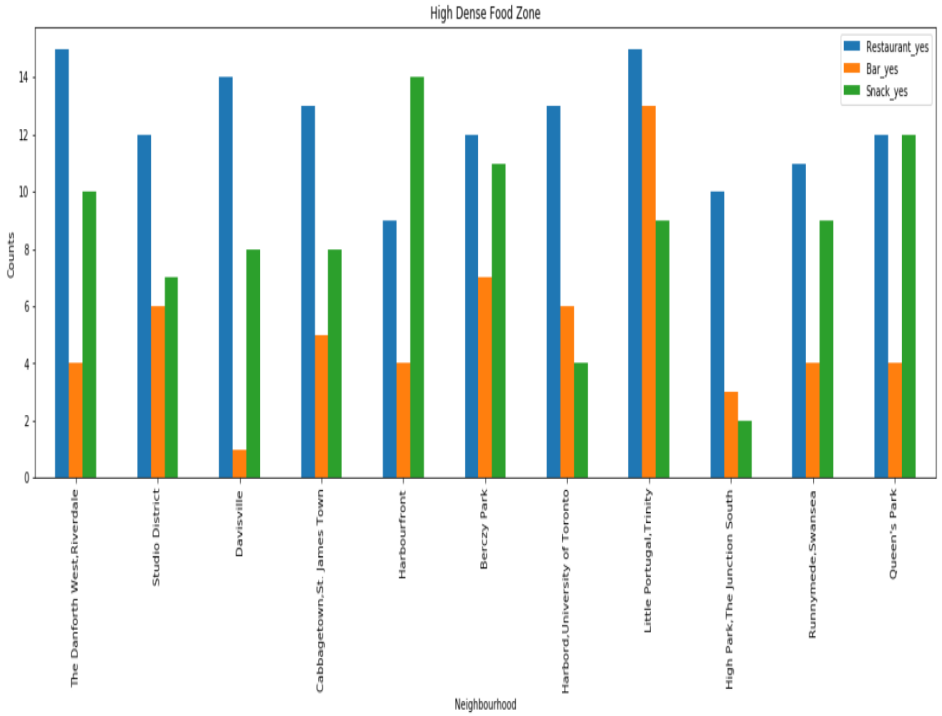
Solutions

- 3 Cluster are identified as below
- ✓ Cluster-0: High Dense Restaurant zone
- ✓ Cluster-1: Moderate Dense Restaurant zone
- ✓ Cluster-2: Low Dense Restaurant zone



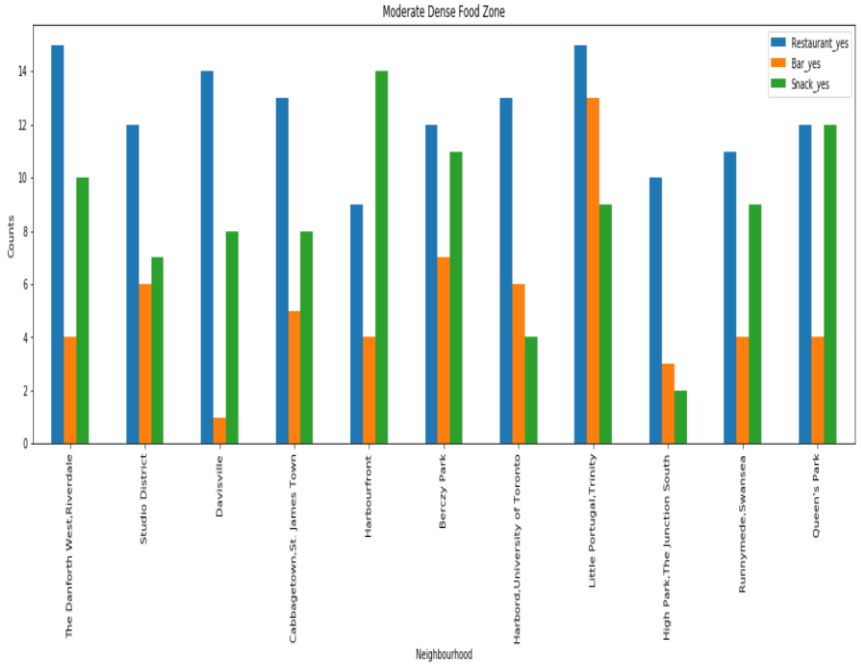
High Dense restaurant zone

- 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



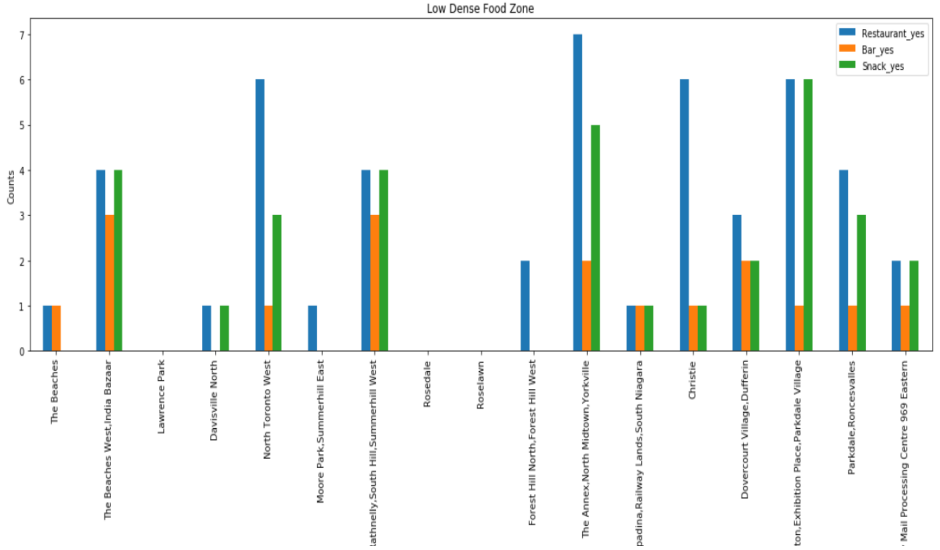
Moderate Dense Area

- 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.



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.



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.

Thank You ☺