

# The prospect of opening a restaurant supply store in Toronto

---

Harun A.R.



# Introduction/Business Problem

---

- This study is to help people planning to open a restaurant supply store in Toronto using cluster analysis by considering the data on the number of restaurants in a cluster



# DATA PREPARATION

---

- Data provided at [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data) and
  - With limit venue = 200
  - Radius Each neighborhood = 500 KM
- [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

# METHODOLOGY

Step by step after download data:

1. Describe Toronto Borough

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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



# METHODOLOGY (2)

Step by step after download data:

## 2. Get all of the Venues (Neighborhood)

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Agincourt	4	4	4	4	4	4
Alderwood, Long Branch	7	7	7	7	7	7
Bathurst Manor, Wilson Heights, Downsview North	21	21	21	21	21	21
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	23	23	23	23	23	23
Berczy Park	57	57	57	57	57	57
Birch Cliff, Cliffside West	4	4	4	4	4	4
Brockton, Parkdale Village, Exhibition Place	25	25	25	25	25	25
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	16	16	16	16	16	16
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	17	17	17	17	17	17
Caledonia-Fairbanks	4	4	4	4	4	4
Canada Post Gateway Processing Centre	13	13	13	13	13	13
Cedarbrae	9	9	9	9	9	9
Central Bay Street	62	62	62	62	62	62
Christie	16	16	16	16	16	16

Result = 94 Neighborhood and 273 uniques categories

## METHODOLOGY (3)

---

Step by step after download data:

3. List "Restaurant"

Result : 54 contain Restaurant

4. Find n\_cluster with best Silhouette Coefficient

**For n\_clusters=2, The Silhouette Coefficient is 0.7347296621494414**

For n\_clusters=3, The Silhouette Coefficient is 0.6588996717779296

For n\_clusters=4, The Silhouette Coefficient is 0.6273022537903404

For n\_clusters=5, The Silhouette Coefficient is 0.5022744542968325

For n\_clusters=6, The Silhouette Coefficient is 0.4551689583117958

For n\_clusters=7, The Silhouette Coefficient is 0.4537558398237324

For n\_clusters=8, The Silhouette Coefficient is 0.4537100467653391

For n\_clusters=9, The Silhouette Coefficient is 0.4530656616555734



## METHODOLOGY (3)

---

Step by step after download data:

5. Count of Neighborhood with Restaurant

Result = From 94 Neighborhood

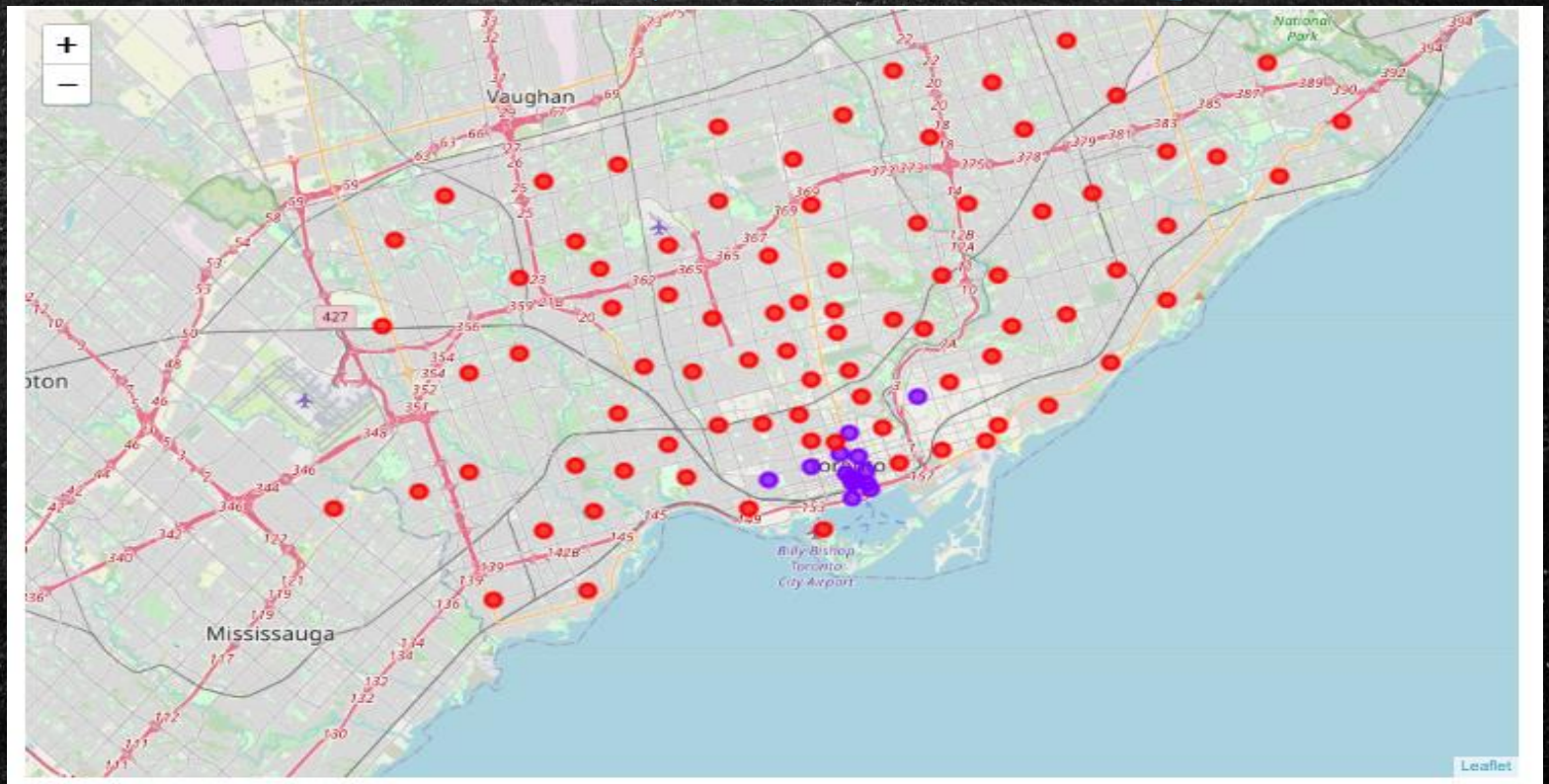
6. Total Merged Jin 2 DataFrame = 103 Neighborhood

6. K-Means Clustering (2 Cluster)

```
array([0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0,  
       0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,  
       0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0,  
       0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0,  
       0, 0, 0, 0, 0, 0], dtype=int32)
```



# Discussion and Result



Red = Cluster 1, Indicate High Prospective  
(84 neighborhood)

Bluer = Cluster 0, Indicate low Prospective  
(14 neighborhood)