

# Battle of Neighborhoods



APPLIED DATA SCIENCE CAPSTONE PROJECT

BY MUSSARRAT JAHAN MOURI

# Introduction (Business Problem)

- ▶ The aim of this project is to build up a model, a search algorithm that compares and analyzes randomly chosen neighborhoods of the city of Toronto
- ▶ To observe the neighborhood similarities and dissimilarities
- ▶ Thus to help stakeholders to gain better and deeper understanding of their destinations and to make better decisions

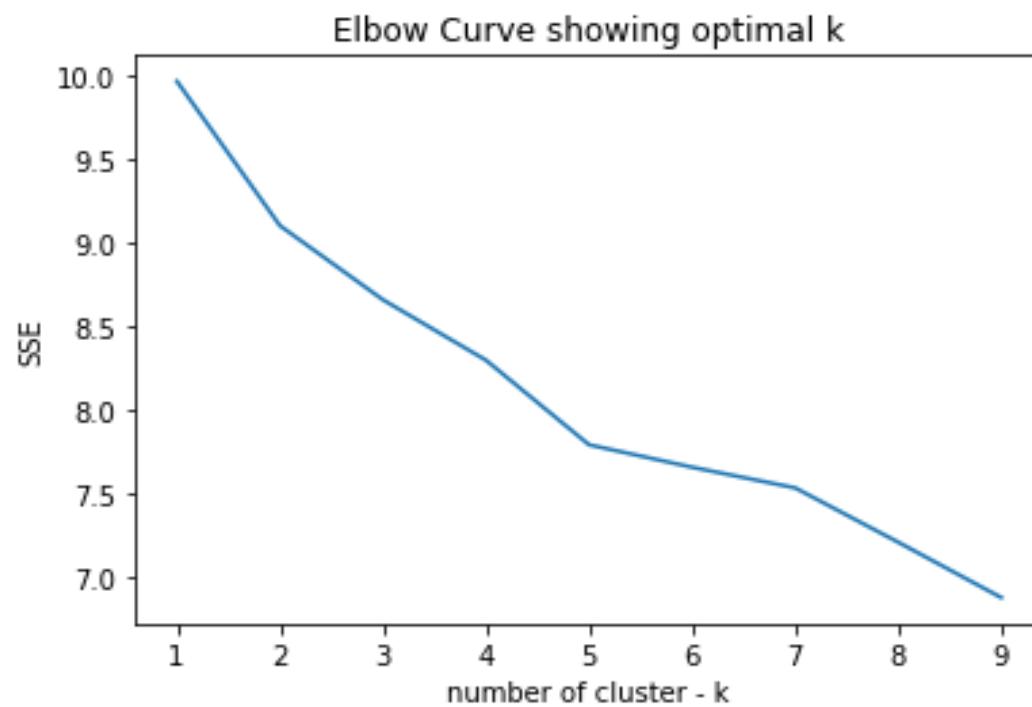
# Data – acquisition and cleaning

- ▶ Initial dataset is collected by web scraping, from Wikipedia
- ▶ Geospatial data and population counts are collected from various publicly available sources. Those data was combined into .csv file and after cleaning conveniently, converted into pandas data frame
- ▶ Foursquare API is used as prime data gathering source. The retrieved data is processed adequately to meet all the requirements

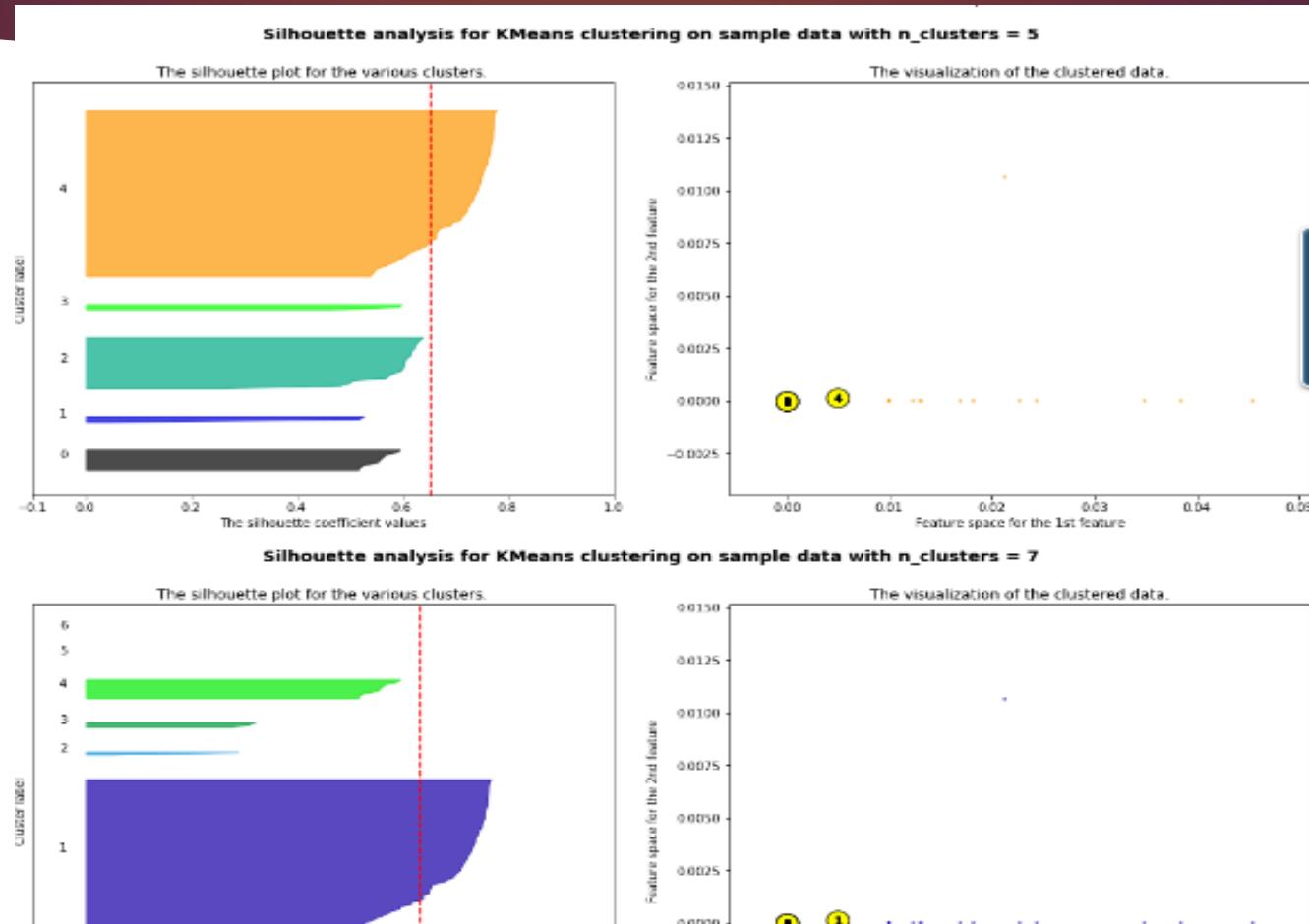
# Clustering

- ▶ To carry out the extensive comparative analysis of randomly chosen neighborhoods, **K-means clustering** is used to form the clusters of different categories of places residing in and around the neighborhoods.
- ▶ However determining the optimal number of clusters in a data set is a fundamental issue in partitioning clustering, such as K-means clustering in our case, which requires the user to specify the number of clusters k to be generated. There are many methods to determine the optimal number. These methods include direct methods and statistical testing methods. In this project the **direct methods** were considered.
- ▶ **Elbow method** and **Average Silhouette** method are used to calculate the optimal number of k

# Elbow Method



# Average Silhouette Method



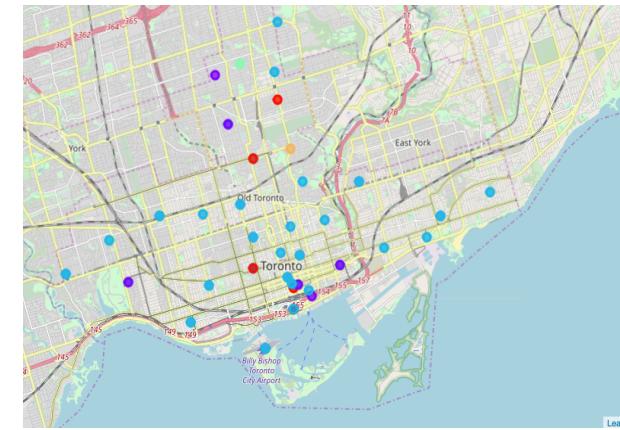
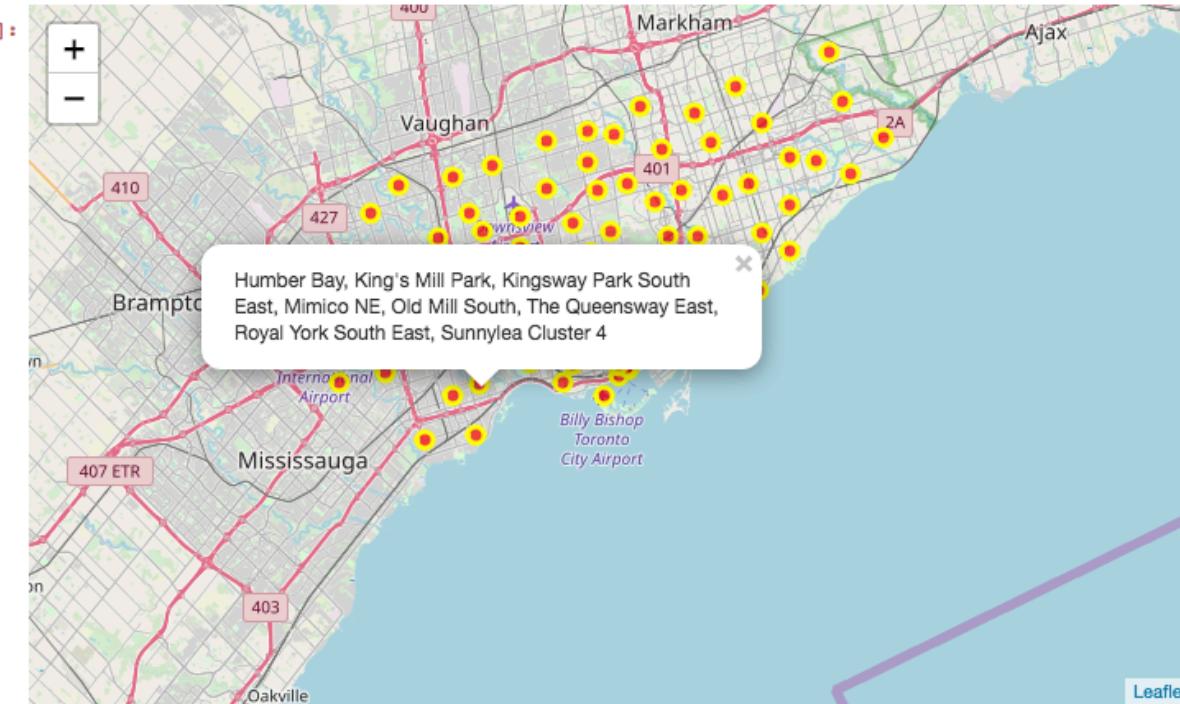
# K-means Clustering

- The value of k is set to 5

```
new_toronto=toronto_merged.set_index("Neighborhood",drop=True)  
new_toronto.head(10)
```

	PostalCode	Borough	Latitude	Longitude	Population	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
Neighborhood													
Rouge, Malvern	M1B	Scarborough	43.806686	-79.194353	66108	4	Fast Food Restaurant	Coffee Shop	Spa	Bus Station	Hobby Shop	Construction & Landscaping	Women's Store
Highland Creek, Rouge Hill, Port Union	M1C	Scarborough	43.784535	-79.160497	35626	4	Breakfast Spot	Bar	Burger Joint	Dumpling Restaurant	Discount Store	Dive Bar	Dog Run
Guildwood, Morningside, West Hill	M1E	Scarborough	43.763573	-79.188711	46943	0	Pizza Place	Fast Food Restaurant	Grocery Store	Breakfast Spot	Moving Target	Electronics Store	Fried Chicken Joint
Woburn	M1G	Scarborough	43.770992	-79.216917	29690	1	Park	Coffee Shop	Convenience Store	Business Service	Event Space	Ethiopian Restaurant	Dessert Shop
Cedarbrae	M1H	Scarborough	43.773136	-79.239476	24383	4	Coffee Shop	Indian Restaurant	Bakery	Thai Restaurant	Gym / Fitness	Fried Chicken Joint	Flower Shop

# Visualization (using Folium)



# Analysis, Comparison and Results

- ▶ Each cluster of neighborhoods are examined carefully

#Examine #Cluster 5													#Examine #Cluster 1														
	Borough	Population	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		Borough	Population	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	Scarborough	66108	4	Fast Food Restaurant	Coffee Shop	Spa	Bus Station	Hobby Shop	Construction & Landscaping	Women's Store	Donut Shop	Diner	Discount Store	2	Scarborough	46943	0	Pizza Place	Fast Food Restaurant	Grocery Store	Breakfast Spot	Moving Target	Electronics Store	Fried Chicken Joint	Rental Car Location	Thrift / Vintage Store	Greek Restaurant
1	Scarborough	35626	4	Breakfast Spot	Bar	Burger Joint	Dumpling Restaurant	Discount Store	Dive Bar	Dog Run	Doner Restaurant	Donut Shop	Drugstore	5	Scarborough	36699	0	Fast Food Restaurant	Women's Store	Convenience Store	Coffee Shop	Pizza Place	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Dog Run
4	Scarborough	24383	4	Coffee Shop	Indian Restaurant	Bakery	Thai Restaurant	Gym / Fitness Center	Fried Chicken Joint	Flower Shop	Chinese Restaurant	Rental Car Location	Caribbean Restaurant	8	Scarborough	22913	0	Furniture / Home Store	Chinese Restaurant	Wings Joint	Burger Joint	Dim Sum Restaurant	Diner	Discount Store	Dive Bar	Dog Run	Doner Restaurant
6	Scarborough	48434	4	Discount Store	Coffee Shop	Sandwich Place	Light Rail Station	Department Store	Convenience Store	Chinese Restaurant	Intersection	Grocery Store	Metro Station	11	Scarborough	29858	0	Pizza Place	Burger Joint	Coffee Shop	Middle Eastern Restaurant	Seafood Restaurant	Bakery	Korean Restaurant	Fish Market	Intersection	Convenience Store
7	Scarborough	35081	4	Intersection	Diner	Coffee Shop	Bus Line	Bakery	Bus Station	Fast Food Restaurant	Park	Metro Station	Soccer Field	13	Scarborough	34588	0	Pharmacy	Shopping Mall	Pizza Place	Chinese Restaurant	Italian Restaurant	Sandwich Place	Bus Stop	Thai Restaurant	Fried Chicken Joint	Seafood Restaurant
9	Scarborough	22136	4	Skating Rink	Bank	Café	Diner	Discount Store	General Entertainment	Thai Restaurant	Park	College Stadium	Eastern European Restaurant	14	Scarborough	54680	0	Chinese Restaurant	Pharmacy	BBQ Joint	Pizza Place	Park	Noodle House	Caribbean Restaurant	Shop & Service	Fast Food Restaurant	Bakery
10	Scarborough	45571	4	Electronics Store	Indian Restaurant	Fast Food Restaurant	Wings Joint	Gym / Fitness Center	Vietnamese Restaurant	Coffee Shop	Pet Store	Bakery	Chinese Restaurant	15	Scarborough	48471	0	Fast Food Restaurant	Grocery Store	Chinese Restaurant	Pharmacy	Indian Restaurant	Burger Joint	Cosmetics Shop	American Restaurant	Other Great Outdoors	Sandwich Place

# Analysis, Comparison and Results

- ▶ A table is set up to compare the neighborhoods

```
#creating cluster table
cluster_t=pd.DataFrame({"Cluster1":df3["Borough"],
                        "Cluster2":df4["Borough"],
                        "Cluster3":df5["Borough"],
                        "Cluster4":df6["Borough"],
                        "Cluster5":df7["Borough"]
})
```

```
cluster_t = cluster_t.replace(np.nan, '', regex=True)
cluster_t
```

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
0					Scarborough
1					Scarborough
2	Scarborough				
3			Scarborough		
4					Scarborough
5	Scarborough				
6					Scarborough
7					Scarborough
8	Scarborough				
9					Scarborough

# Analysis, Comparison and Results

- The comparison model works for 2 or more randomly chosen neighborhoods. Here is a result with 3 random neighborhood names:

```
Nbd1=input("Enter the Neighborhood: ")  
Enter the Neighborhood: Rosedale  
  
Nbd2=input("Enter the Neighborhood: ")  
Enter the Neighborhood: Woburn  
  
Nbd3=input("Enter the Neighborhood: ")  
Enter the Neighborhood: Humber Summit  
  
venue_comparison=new_toronto.loc[[Nbd1,Nbd2,Nbd3]].T  
venue_comparison
```

Neighborhood	Rosedale	Woburn	Humber Summit
PostalCode	M4W	M1G	M9L
Borough	Downtown Toronto	Scarborough	North York
Latitude	43.6796	43.771	43.7563
Longitude	-79.3775	-79.2169	-79.566
Population	14561	29690	11950
Cluster Labels	1	1	3
1st Most Common Venue	Park	Park	Bakery
2nd Most Common Venue	Playground	Coffee Shop	Pizza Place



From the comparison among “Rosedale”, “Woburn” and “Humber Summit” I could retrieve the postal code, name of the boroughs, geospatial data of the neighborhoods, population count, cluster labels and the top 10 most common venue categories.

# Discussion

- ▶ This comparison model clearly shows the expected outcomes. The project aimed to build up a model that can compare two or more randomly chosen neighborhoods of the city of Toronto. The comparison is carried out using K-means clustering algorithm to cluster the neighborhoods based on its venue categories. This model could be helpful for stakeholders to gain more insights about individual neighborhood or to compare chosen neighborhoods.

# Conclusion

- ▶ In this project, a problem or the need of an application or one platform that would help stakeholders to understand a country, state, city or its neighborhoods better is introduced. It is also mentioned that this would require a search algorithm that usually would return the requested features such as population rate, median housing price, school ratings, crime rates, weather conditions, recreational facilities, socio-economical distributions etc. But this project specifically focused only on the neighborhood analysis in means of simple comparison of geospatial data, population counts and the top most common venues based on venue categories. This leaves us to an open end to elaborate the search algorithms by adding more features or feature selection in the future. Nonetheless refining and improving the algorithm further.

# References

- ▶ [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)
- ▶ <https://www.statcan.gc.ca>
- ▶ [https://en.wikipedia.org/wiki/Demographics\\_of\\_Toronto\\_neighbourhoods](https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods)
- ▶ [https://ipfs.io/ipfs/QmXoypizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uc0/wiki/Demographics\\_of\\_Toronto.html](https://ipfs.io/ipfs/QmXoypizjW3WknFiJnKLwHCnL72vedxjQkDDP1mXWo6uc0/wiki/Demographics_of_Toronto.html)
- ▶ [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- ▶ [Foursquare](#)

