# Battle of the Neighbourhoods

**Applied Data Science**

**Capstone Project**

**By: Fahamida Haque**

# Outline

- Intoduction and Objectives
- Data
- Methodology
- Results
- Discussion
- Drawbacks
- Conclusion

# Introduction

- Coffee is one of the most widely consumed beverages in the world

- A survey found that almost three quarters of Canadians have consumed coffee in the last day

- Our client is looking to open his coffee shop in Toronto

- Toronto is also an international center of business, culture, and the arts and is one of the most multicultural cities in the world

- The ideal location for this new business would be an area with minimal competition and close to the city center

# Data

- Wikipedia – to obtain postal codes for the city of Toronto
- Geocoder Package  –  used to find the latitudes and longitudes of each neighbourhood
- Foursquare API  –  used to search for the venues in each neighbourhood

# Methodology- Data Cleansing

- Postal codes for the city of Toronto is obtained from Wikipedia.

- Wikipedia table is extracted using Beautiful Soup.

- Data is saved in a Pandas data frame.

- Discard rows without values for borough

- If borough assigned but neighbourhoods is , neighbourhoods = borough

| | Postal Code | Borough | Neighbourhood |
|---|---|---|---|
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights |
| 4 | M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government |

# Methodology – Data Exploration

- Use geocoder to get geographical coordinates for each neighbourghood

- Plot the neighbourghoods using Folium

- Use Foursquare Api to search for 500 venues within 1 km radius for each neighbourhood
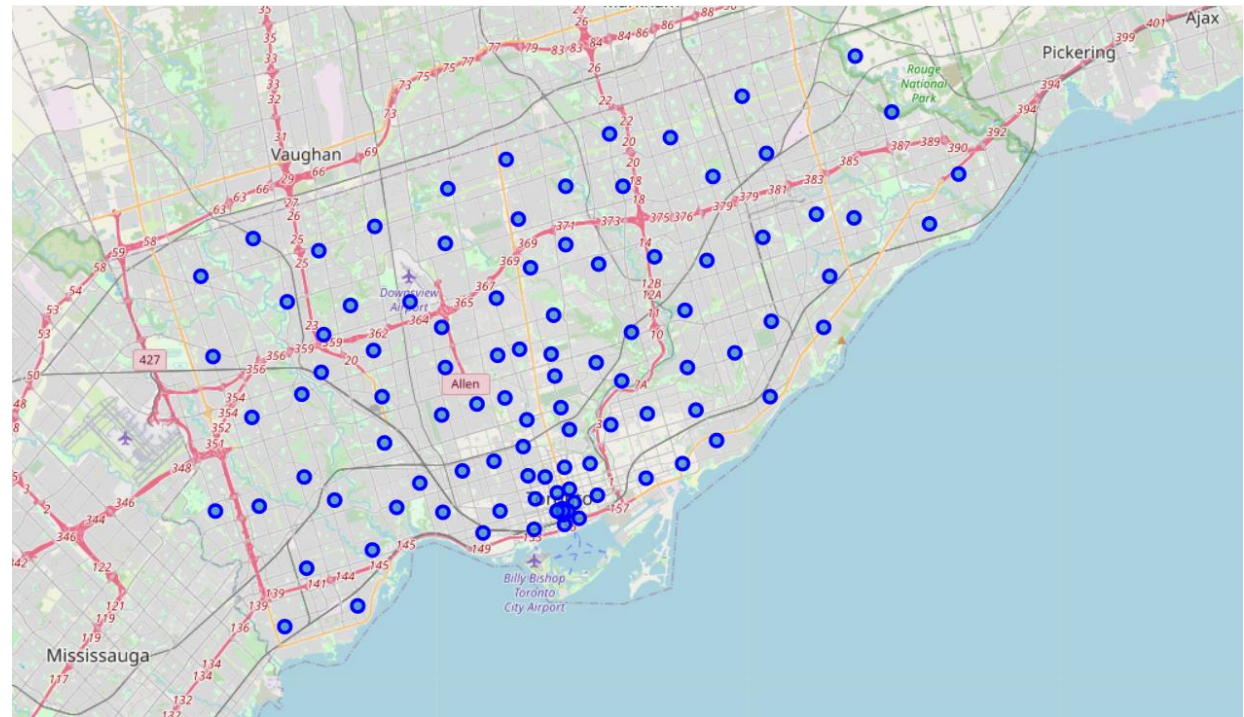
# Methodology – Machine Learning

- Get a list of all the venues
- Find the unique venue categories
- Find average mean of each category in each neighbourhood
- use the k-means clustering method to cluster the neighbourhoods based on their average mean
- Use the elbow point method to find number of clusters
- K = 5

Distortion Score Elbow for KMeans Clustering

- - - elbow at $k = 5$, $score = 0.013$
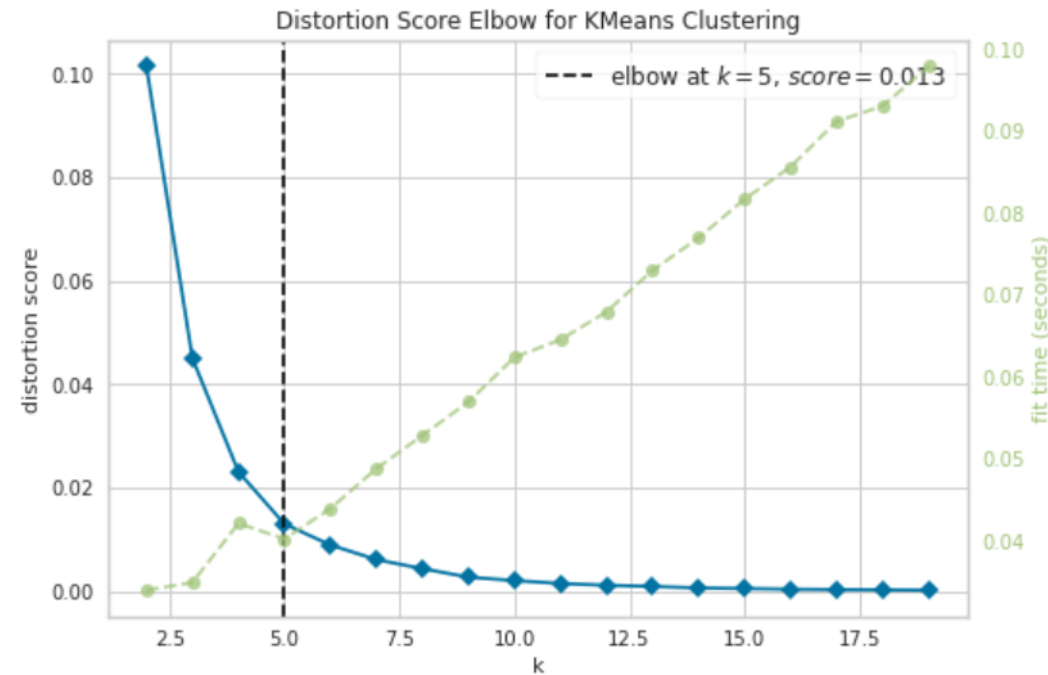
# Methodology – Data Analysis

- Plot all the cluster labels from each neighbourhood onto the map of Toronto.

- Plot the number of neighbourhoods per cluster

- Plot the average number of coffee shops in each cluster

- Analyze the plots and choose a cluster with a low density of coffee shops

- Search cluster to find a neighbourhood that is closest to the city's downtown
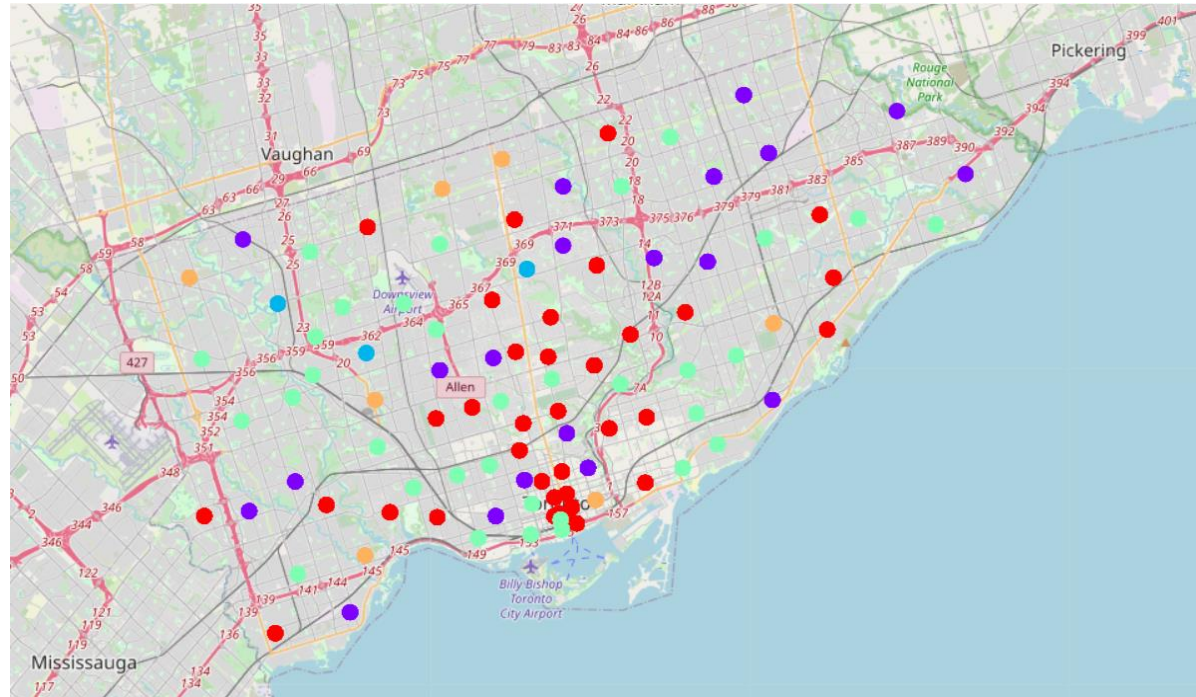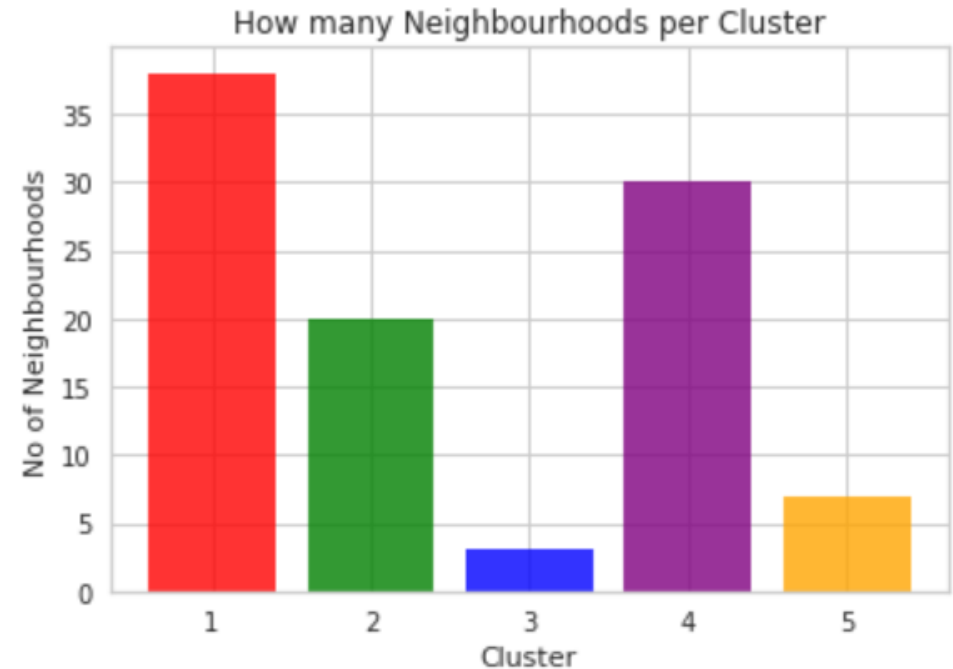
# Results

- Cluster 1 (k = 0) is represented by the red points
- Cluster 2 (k = 1) are represented by the points
- Cluster 3 (k = 2) is represented by the blue points
- Cluster 4 (k = 3) is represented by the purple points
- Cluster 5 (k = 4) is represented by the orange points
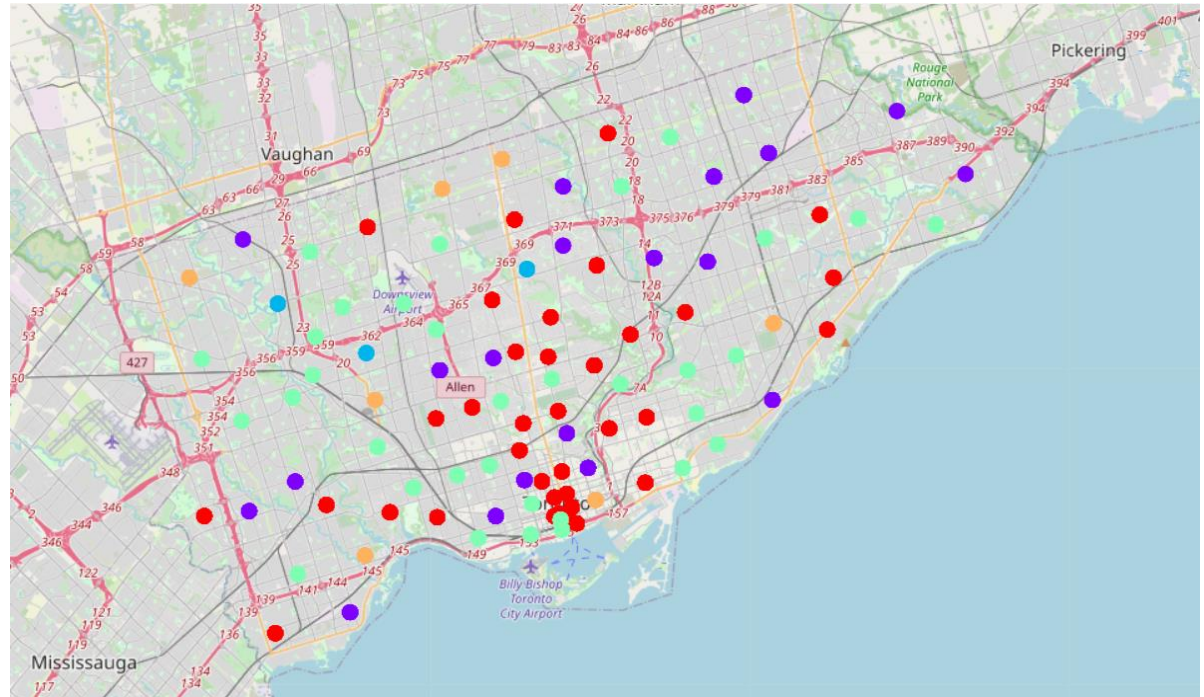
# Results - Continued

1. Cluster 1 - 38 neighbourhoods
2. Cluster 4 – 30 neighbourhoods
3. Cluster 2 – 20 neighbourhoods
4. Cluster 5 – 7 neighbourhoods
5. Cluster 3 – 3 neighbourhoods



How many Neighbourhoods per Cluster

# Results - Continued

1. Cluster 2 (0.018)
2. Cluster 4 (0.059)
3. Cluster 1 (0.092)
4. Cluster 5 (0.136)
5. Cluster 3 (0.222)



Average number of Coffee Shops per Cluster
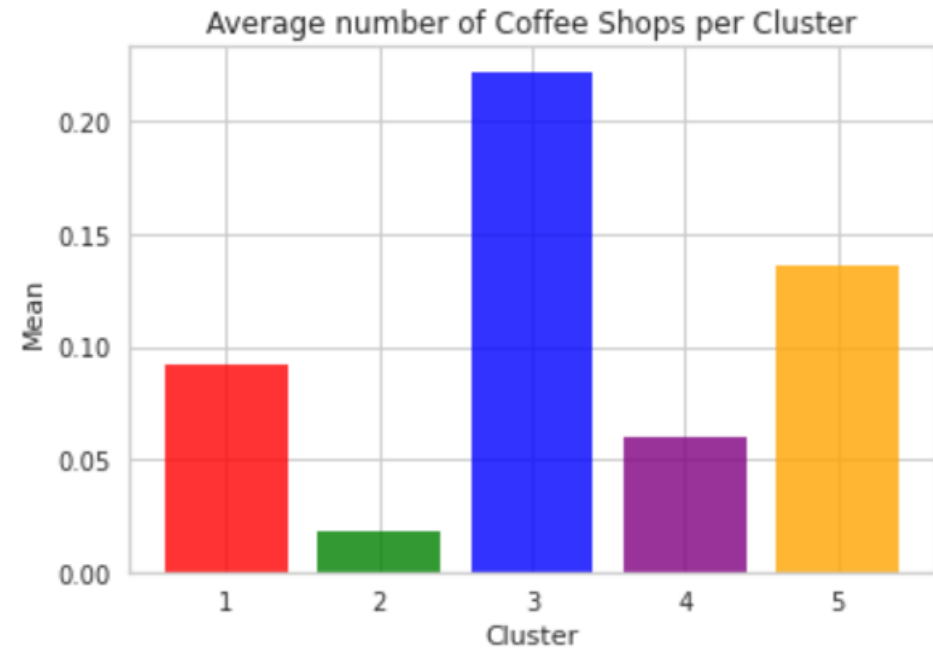
# Results - Continued

- We can see that clusters 1, 2 and 4 have the lowest density of coffee shops with the greatest number of neighbourhoods

- Clusters 3 and 5 have the highest density of coffee shops with the least number of neighbourhoods

- Our coffee shop will be either in cluster 1, 2 or 4 since a low density of coffee shops indicates that there will be less competition with similar businesses

- Clusters 1, 2 and 4 contain neighbourhoods in Downtown Toronto

- Ideal locations from cluster 1, near Ryerson and central Bay street

- Ideal locations from cluster 2 include Adelaide and King

- Ideal locations from cluster 4 include St James and Rosedale

# Discussion

- All data used in our analysis comes from the Foursquare API

- Using more databases would have resulted in a more complete recommendation

- Elimination of some neighbourhoods during data cleansing, excluded them from the analysis

- Analyzing more neighbourhoods would result in a better recommendation of our final location

# Conclusion

- We found that the downtown neighbourhoods in clusters 1, 2 and 4 are the best locations to open the coffee shop

- Ideal locations from cluster 1, near Ryerson and central Bay street

- Ideal locations from cluster 2 include Adelaide and King

- Ideal locations from cluster 4 include St James and Rosedale