# Capstone: The Battle of Neighborhood

Recommending Indian restaurant in Toronto, CA

By: Siddharth Sharma

# Introduction

- This project deals with segmenting the neighborhood of toronto on the basis of the similarities of the locations and clustering the locations on the basis of the competition with other Indian restaurants.

## Problem Statement

Given the location data of the neighborhood of the Toronto, Canada, one would like to cluster the similar neighborhoods according to the frequency of Indian restaurants and on the basis of analysis we would like to recommend best neighborhoods to start a new Indian restaurant.

# Data

**To solve this problem, I will need below data:**

- **List of neighborhoods in Toronto, Canada. “ https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M”**
- **Latitude and Longitude of these neighborhoods.**
- **Venue data related to Indian restaurants. This will help us find the neighborhoods that are most suitable to open a Indian restaurant.**

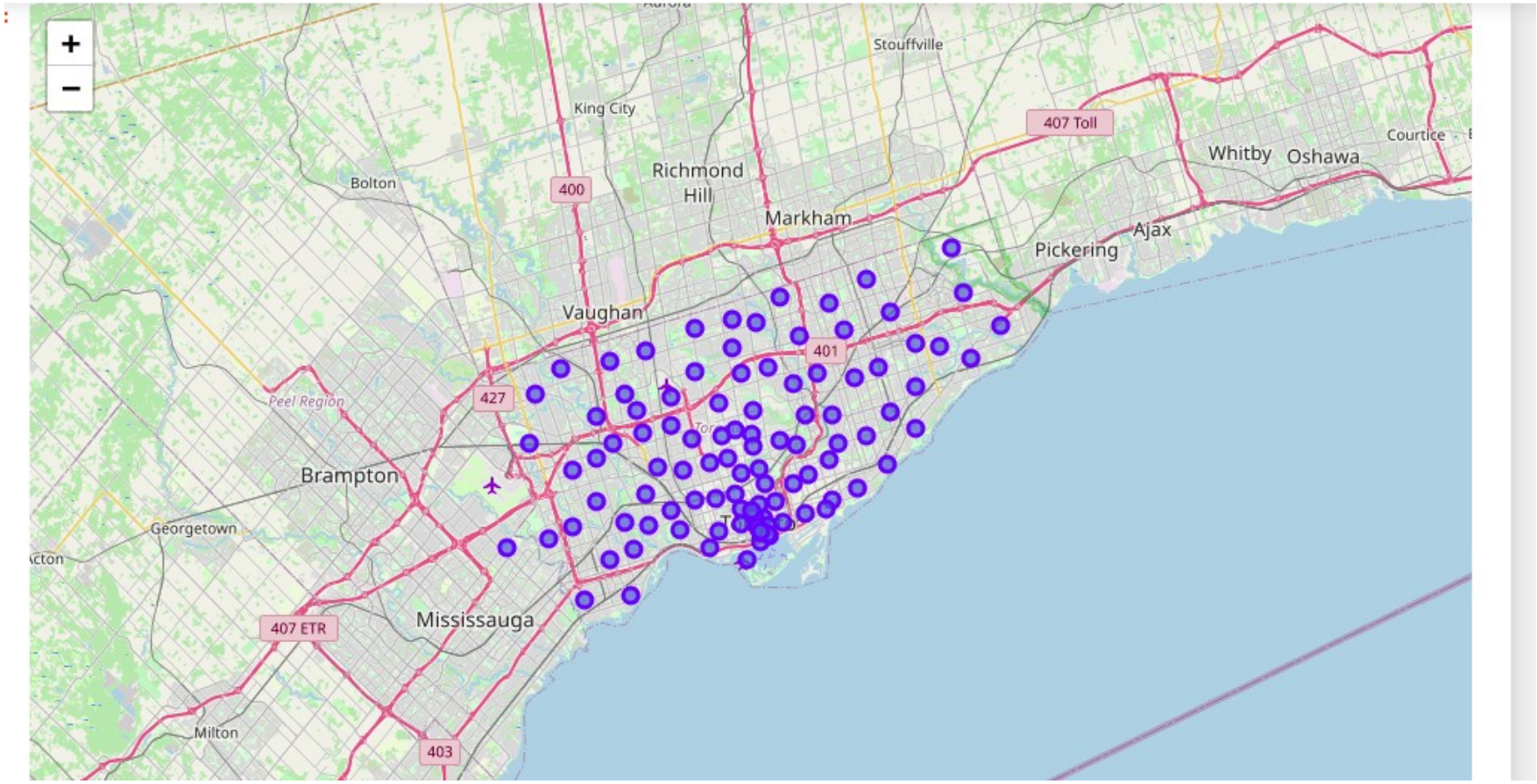
- **Extracting the data**
  - **Scrapping of Toronto neighborhoods via Wikipedia**
  - **Getting Latitude and Longitude data of these neighborhoods via Geocoder package**
  - **Using Foursquare API to get venue data related to these neighborhoods.**

# Final Data

| | Postal code | Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|---|
| 0 | M1B | Scarborough | Malvern , Rouge | 43.8067 | -79.1944 |
| 1 | M1C | Scarborough | Rouge Hill , Port Union , Highland Creek | 43.7845 | -79.1605 |
| 2 | M1E | Scarborough | Guildwood , Morningside , West Hill | 43.7636 | -79.1887 |
| 3 | M1G | Scarborough | Woburn | 43.771 | -79.2169 |
| 4 | M1H | Scarborough | Cedarbrae | 43.7731 | -79.2395 |
| 5 | M1J | Scarborough | Scarborough Village | 43.7447 | -79.2395 |
| 6 | M1K | Scarborough | Kennedy Park , Ionview , East Birchmount Park | 43.7279 | -79.262 |
| 7 | M1L | Scarborough | Golden Mile , Clairlea , Oakridge | 43.7111 | -79.2846 |
| 8 | M1M | Scarborough | Cliffside , Cliffcrest , Scarborough Village West | 43.7163 | -79.2395 |
| 9 | M1N | Scarborough | Birch Cliff , Cliffside West | 43.6927 | -79.2648 |
| 10 | M1P | Scarborough | Dorset Park , Wexford Heights , Scarborough To... | 43.7574 | -79.2733 |
| 11 | M1R | Scarborough | Wexford , Maryvale | 43.7501 | -79.2958 |
| 12 | M1S | Scarborough | Agincourt | 43.7942 | -79.262 |
| 13 | M1T | Scarborough | Clarks Corners , Tam O'Shanter , Sullivan | 43.7816 | -79.3043 |
| 14 | M1V | Scarborough | Milliken , Agincourt North , Steeles East , L'... | 43.8153 | -79.2846 |
| 15 | M1W | Scarborough | Steeles West , L'Amoreaux West | 43.7995 | -79.3184 |
| 16 | M1X | Scarborough | Upper Rouge | 43.8361 | -79.2056 |

# Methodology

- Using Folium to map all the neighborhoods in Toronto



# Methodology

- **Majorly relied on Foursquare API to retrieve all venues of each neighborhoods, then group by each neighborhoods and to count how many venues before filter top 10 most common venue types of each neighborhoods**

(2145, 7)

Out[22]:

| | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Malvern , Rouge | 43.806686 | -79.194353 | Wendy's | 43.807448 | -79.199056 | Fast Food Restaurant |
| 1 | Rouge Hill , Port Union , Highland Creek | 43.784535 | -79.160497 | Royal Canadian Legion | 43.782533 | -79.163085 | Bar |
| 2 | Rouge Hill , Port Union , Highland Creek | 43.784535 | -79.160497 | SEBS Engineering Inc. (Sustainable Energy and ... | 43.782371 | -79.156820 | Construction & Landscaping |
| 3 | Guildwood , Morningside , West Hill | 43.763573 | -79.188711 | G & G Electronics | 43.765309 | -79.191537 | Electronics Store |
| 4 | Guildwood , Morningside , West Hill | 43.763573 | -79.188711 | Big Bite Burrito | 43.766299 | -79.190720 | Mexican Restaurant |

# Methodology

- **One hot coding for presence and absence of venues in the neighborhood.**

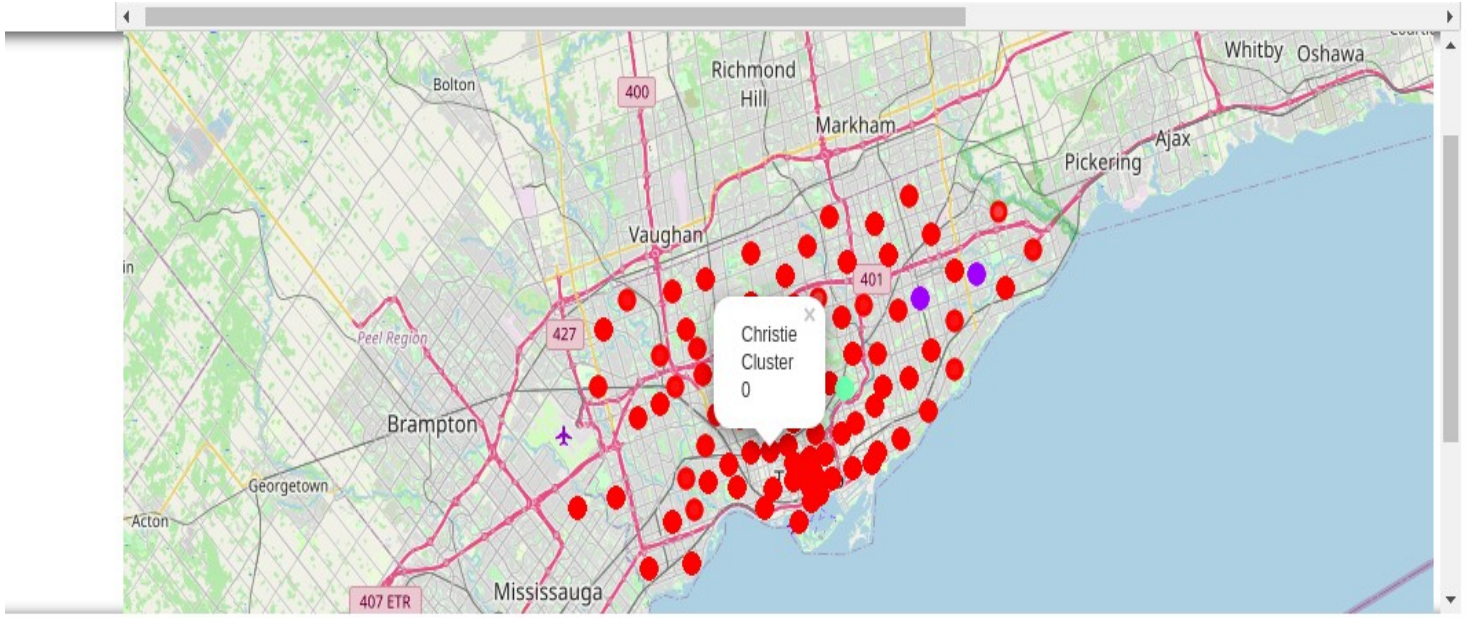
| Neighborhood | New American Restaurant | Nightclub | Noodle House | Office | Opera House | Optical Shop | Organic Grocery | Other Great Outdoors | Outdoor Sculpture | Park | Performing Arts Venue | Pet Store | Pharmacy | Pizza Place | Plane | Playground | Plaza |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Malvern , Rouge | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Rouge Hill , Port Union , Highland Creek | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Rouge Hill , Port Union , Highland Creek | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Guildwood , Morningside , West Hill | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

```
In [26]: toronto_onehot.shape
```

```
Out[26]: (2143, 269)
```

# Results

- K mean clustering is used to cluster the neighborhood based on disimilarities. Three clusters were made



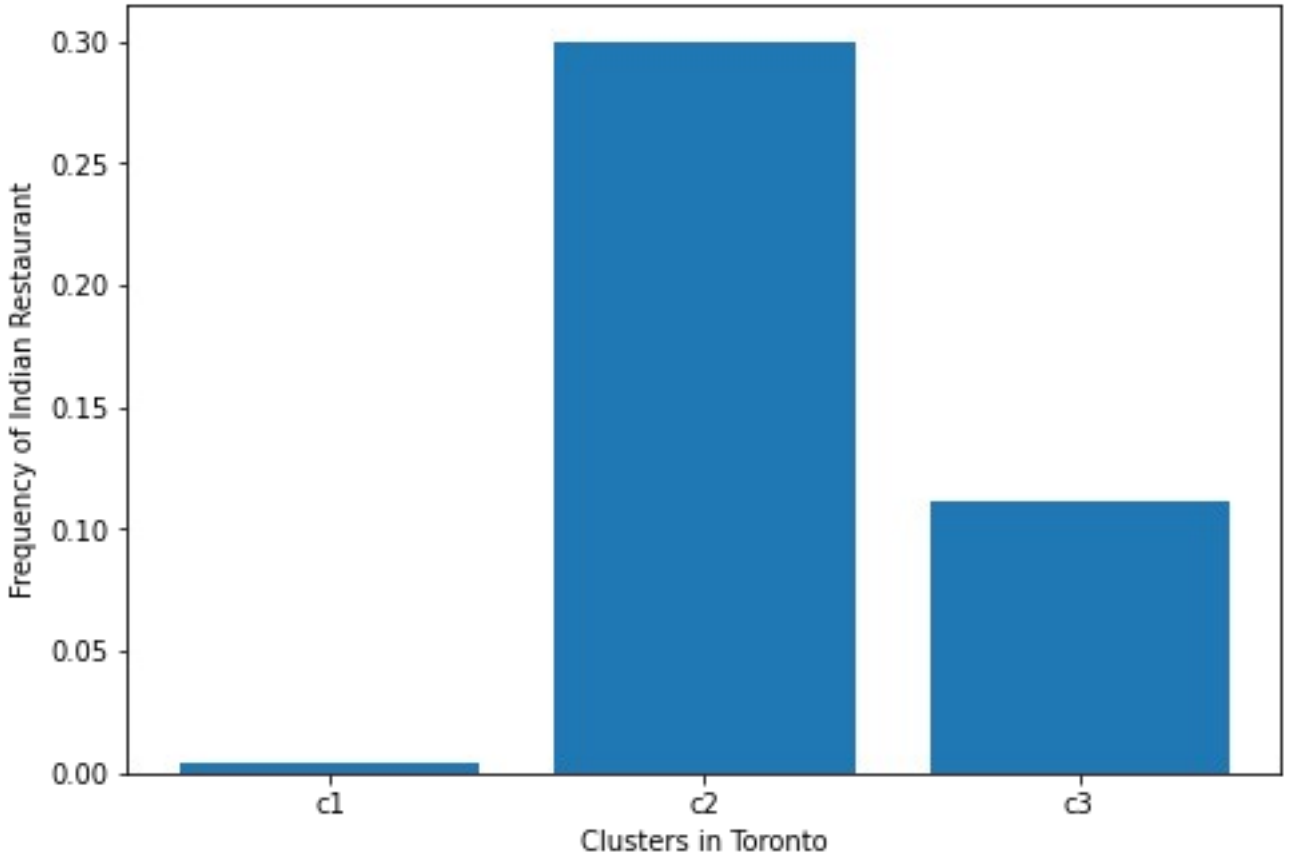

| | Cluster 1 | Cluster 2 | Cluster 3 |
|---|---|---|---|
| Mean | 0.003 | 0.3 | 0.11 |

# Results

Cluster 1 : Has minimum number of Indian restaurants

Cluster 2: Has highest amount of Indian Restaurants

Cluster 3: In between



# Discussion and Conclusion

**Cluster 1** has the least amount of indian restaurants, whereas **Cluster 2** has the highest amount of indian restaurants. Through this data analysis, we can recommend to choose any neighborhood from the **cluster 1** to start a new Indian restaurant, as there is very less competition and a demand for Indian restaurant.