Analysis of South Asian Restaurant Locations in the City of Brampton

Taranveer Birk

July 14, 2019

TABLE OF CONTENTS

1. Introduction

The Greater Toronto Area serves as a home for many immigrants from all parts of the world. The society is a diverse mix as a collective, although many celebrated enclaves of ethnic decent exist throughout the surrounding cities of Toronto. A very pronounced enclave of ethnic people is the South Asian (Afghanistan, Bangladesh, Bhutan, Indian, Maldives, Nepal, Pakistan, Sri Lanka) Community within the city of Brampton. The 2015 Canadian census reported that roughly 44% of residences were of South Asian descent, one of the few Canadian cities to have an ethnic community with the largest general population among all other communities. With such a pronounced presence within a city, this report will target stakeholders that are interested in opening a **South Asian Restaurant** in the **city** of **Brampton (1)**.

2015 Canadian Cencus Population of Brampton Represented by Ethnic Minority Groups

Ethnic Minority Groups	Brampton North		Brampton East		Brampton Central		Brampton South		Brampton West		Brampton Total	
	#	% of Total	#	% of Total	#	% of Total	#	% of Total	#	% of Total	#	% of Total
South Asian	48,935	41%	80,035	66%	26,630	26%	49,950	42%	56,145	43%	261,705	44%
Chinese	2,020	2%	1,235	1%	1,685	2%	1,885	2%	2,130	2%	8,955	2%
Black	14,650	12%	13,775	11%	15,570	15%	13,775	11%	24,405	19%	82,175	14%
Filipino	3,260	3%	1,915	2%	4,955	5%	4,105	3%	5,870	5%	20,100	3%
Latin American	3,075	3%	2,220	2%	3,925	4%	2,110	2%	2,715	2%	14,045	2%
Arab	900	1%	1,515	1%	1,060	1%	1,325	1%	1,250	1%	6,045	1%
South East Asian	1345	1%	1705	1%	1,620	2%	1935	2%	1,820	1%	8,425	1%
West Asain	685	1%	2110	2%	925	1%	660	1%	895	1%	5,275	1%
Korean	85	0%	25	0%	50	0%	125	0%	145	0%	430	0%
Japanese	130	0%	25	0%	155	0%	125	0%	85	0%	520	0%
Visible Minority - Other	3,205	3%	3,875	3%	2,670	3%	2,235	2%	3,970	3%	15,950	3%
Multiple Visible Minorities	1,970	2%	1,620	1%	1,635	2%	1,875	2%	2,490	2%	9,585	2%
Not a Visible Minority	37,695	32%	11,435	9%	40,870	40%	40,215	33%	27,510	21%	157,720	27%
Total	**117,955**	**100%**	**121,490**	**100%**	**101,750**	**100%**	**120,320**	**100%**	**129,430**	**100%**	**590,930**	**100%**

Figure 1. 2015 Canadian census data for the FSA's within the city of Brampton (1)

In order to provide the best context for analysis, this report will source spatial and venue data to better answer some initial questions. First, is the selection of a central point within the city. We'll use the central point to define geometric clusters based on distance away from the center point. Stakeholders will want to compare the current landscape of venues across the city, then look at the penetration of Indian restaurants. Once we've collected the required data, we can map locations and use machine learning techniques to calculate optimal locations based on ideal requirements. Once the analysis is completed, we'll discuss the outcome and see if we can narrow our location list to a few centroids for stakeholders to complete a street level analysis.

2. Data

As stated in the introduction, we require a few external data sources to help with our analysis, they include:

- StatsCan - 2015 Census data, to provide background on the ethnic population distribution of the federal census tracts in Brampton.
- The geocoordinates for the approximate addresses of a central location within the city of Brampton will be obtained using **Google Maps API** reverse geocoding**.**
- Venue data for locational analysis will be obtained using the **Foursquare API.** The venue data will assist in mapping locations and calculating distance from the city centre.

- Addresses of centroid locations created during analysis will also be obtained using the **Google Maps API** for reverse geocoding. Addresses will be used to present optimal locations in the results section of the report.

In addition to collecting the data in our notebook, it will be merged and processed to complete a clustering analysis based on the requirements of an ideal space within the city.

A. Assigning Neighborhood Candidates

First, we need to find a central location within the city, one which can cover a large portion of populated land in the city when scaled out by its radius. The location selected is **'140 Kennedy Rd N, Brampton, Ontario'** this is located close to the downtown core. The 9km radius assigned covers most of Brampton **excluding the east end** past Airport Rd.. In order to begin assigning our neighborhood centroids, we need to connect the Google Maps API to acquire the geo-coordinates using our address.

We'll then use the longitude and latitude values of our city centre to create identical circular grid cells which cover our area of interest. We will look to maximize the surface area within city limits, while keeping our total centers to less than 400 candidates. The code allows for adjusting the radius of both the total surface area and each centroid. This is important to adjust according to city limits, city planning and population. We then use a Folium Map to visualize the distribution of centroids. The centroids were adjusted to be larger to compensate for the sparse distribution of land use in the primarily suburban city.

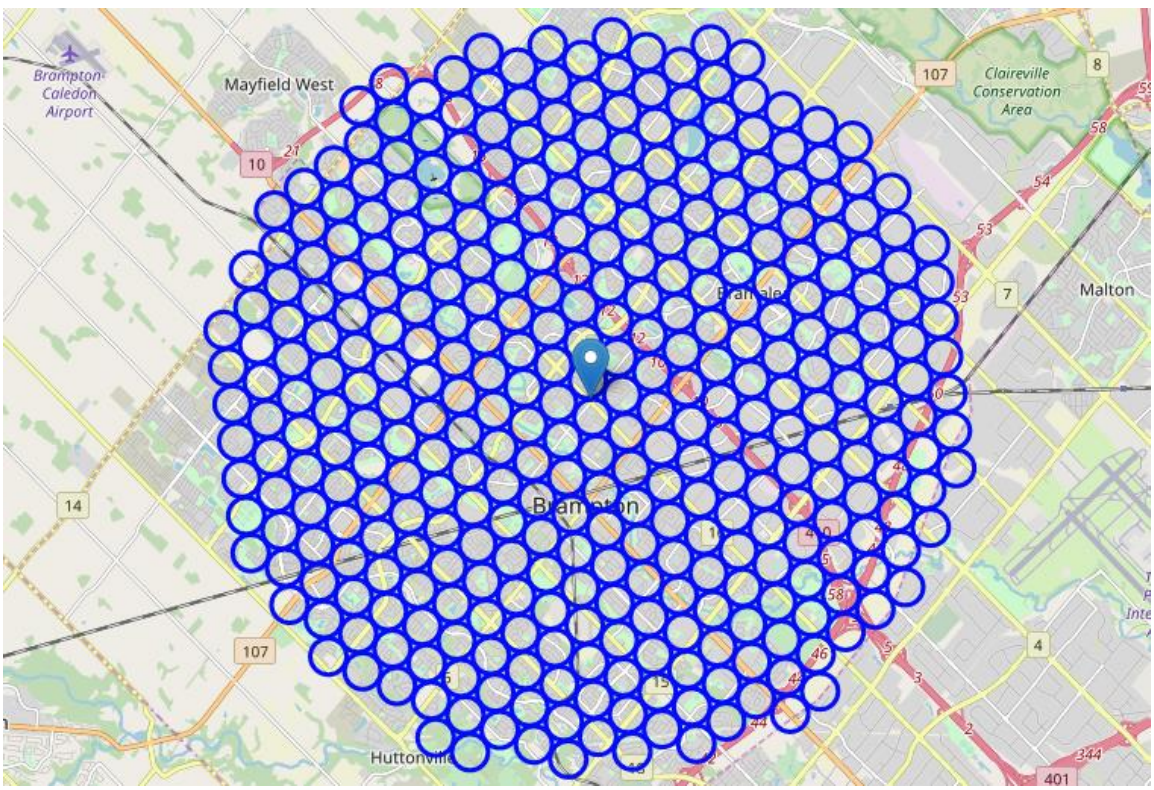

Figure 2. Centroid clusters mapped around a 7km our selected 'City Center' in Brampton (2)

B. Google Map API

The geo-coordinates produced for each centroid on our map are reverse geo-coded to obtain a list of addresses sourced using the **Google Map API**. Once we've extracted a clean list, we'll create a Dataframe which contains the address, Lat, Long, X, Y and Distance from City Center.

C. Four Squares API

Using existing Foursquare credentials, we want to leverage the venue database to acquire a list and coordinates of all restaurants and Indian restaurants within the city of Brampton. Once we've acquired this data, we can look at a few simple metrics that describe the restaurant landscape within the city.

1. Total number of Restaurants,
2. Total number of South Asian Restaurants
3. Percentage of South Asian Restaurants
4. Average number of Restaurants in each neighborhood centroid

In order to only select restaurant venues, the FourSquare API allows for specific category ID's to be filters when completing a data pull. This is very useful in limiting our data set, and defining categorical variables that can be visualized with their locations.

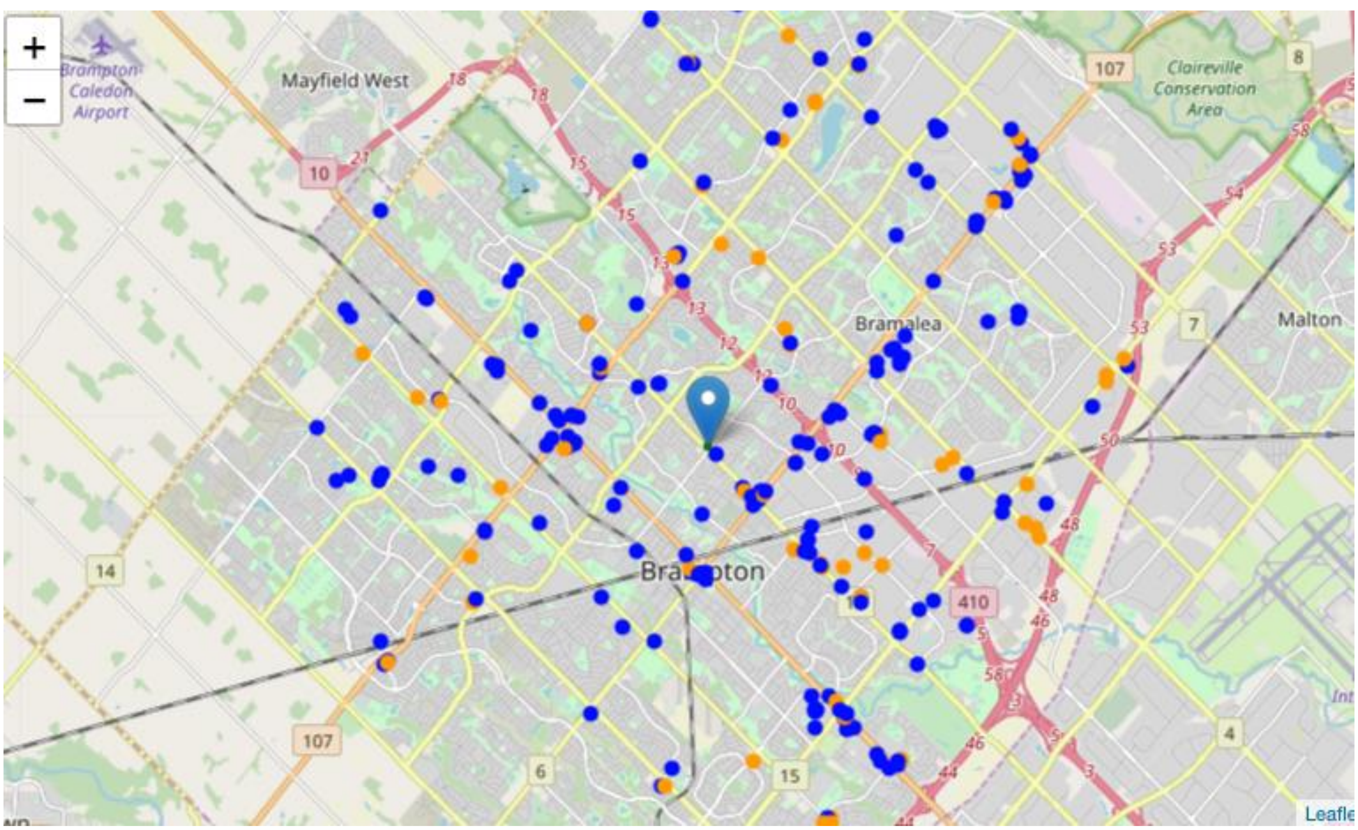

Figure 3. Restaurants (blue points) and South Asian Restaurants (orange point) mapped across the city of Brampton (2)

	Address	Latitudes	Longitudes	X	Y	Distance from center	Restaurants in area	Distance to southasian restaurant
0	7400 Bramalea Rd, Mississauga, ON L5S 1X1, Canada	43.691866	-79.676063	-5.300573e+06	1.054055e+07	9787.742334	0	2152.718466
1	7505 Bramalea Rd, Mississauga, ON L5S 1C4, Canada	43.697953	-79.677052	-5.299593e+06	1.054055e+07	9539.281944	0	2294.871229
2	Express Toll Route, ON L6T, Canada	43.704040	-79.678042	-5.298613e+06	1.054055e+07	9387.049590	0	1364.129106
3	20 Melanie Dr, ON L6T 4K8, Canada	43.710129	-79.679033	-5.297633e+06	1.054055e+07	9335.753853	1	432.528572
4	2550 Steeles Ave E, ON L6T, Canada	43.716217	-79.680023	-5.296653e+06	1.054055e+07	9387.049590	2	275.609551

Figure 4. Locations dataframe created for our centroid locations with distances. (2)

3. Methodology

The objective will be to use the retrieved data to create and map clusters that have a low density of Indian restaurants within a particular radius. Based on our city center location, we will work with roughly a 7km radius around the desired point. This should be sufficient surface coverage for a densly suburban population. With the combination of centroid location data and the FourSquare venue data, we can use Folium Maps to explore densities of restaurant throughout the city.

Heat mapping and markers will be used to narrow down specific neighborhoods of interest. Ideally, we want to reduce our radius from 10km to in-between 2-3km, and would preferably be close to our city center. We will refer to this area as the **ROI Center**, and apply the necessary parameters to view on the map. After we've defined our area of analysis, we can add new centroid clusters at a lower granularity. These centroids can be restricted by distance from other restaurants or Indian restaurants, this will allow us to prepare for **k-means clustering** in our final stage of analysis.

Based on the thresholds the stakeholders determine for the basic location requirements, we want to avoid being 600 meters in distance to any other Indian restaurant and no more than one restaurant in a 250-meter radius. These requirements are fitting, as commercial real-estate is sparsely spread out around major intersections. When viewing on a map which has both ROI centroids and heatmap density of existing restaurants, we can begin visualizing small clusters of centroids. At this point we can utilize the k-means clustering algorithm to create zones of interest based on general neighborhood/addresses. They will be ranked according to overall distance to our city center.

4. Analysis

Once we've loaded our locations dataframe, we should have sufficient data to:

1. Create a heatmap of restaurants and South Asian locations throughout the city.
2. Create a heatmap which focuses on a promising radius close to our city center.
3. Generate candidate locations within ROI area with applied restrictions.
4. Use K-Means clustering to cluster our ROI candidate locations, narrow down top candidates.
5. Present top locations as markers on the final map.

Much of the interpretation of good/bad areas to explore will come from visually analyzing our maps. Inputs will be adjusted to assume optimal values for each of our targeted assessments.

Figure 5. Folium Heatmap of restaurant locations surrounding our 'City Center' (3)

Our first HEATMAP above serves as a view of restaurant clusters in a 7km radius of our 'City Center'. At this point we should narrow down an attractive area with a low density of heat markers to analyze deeper

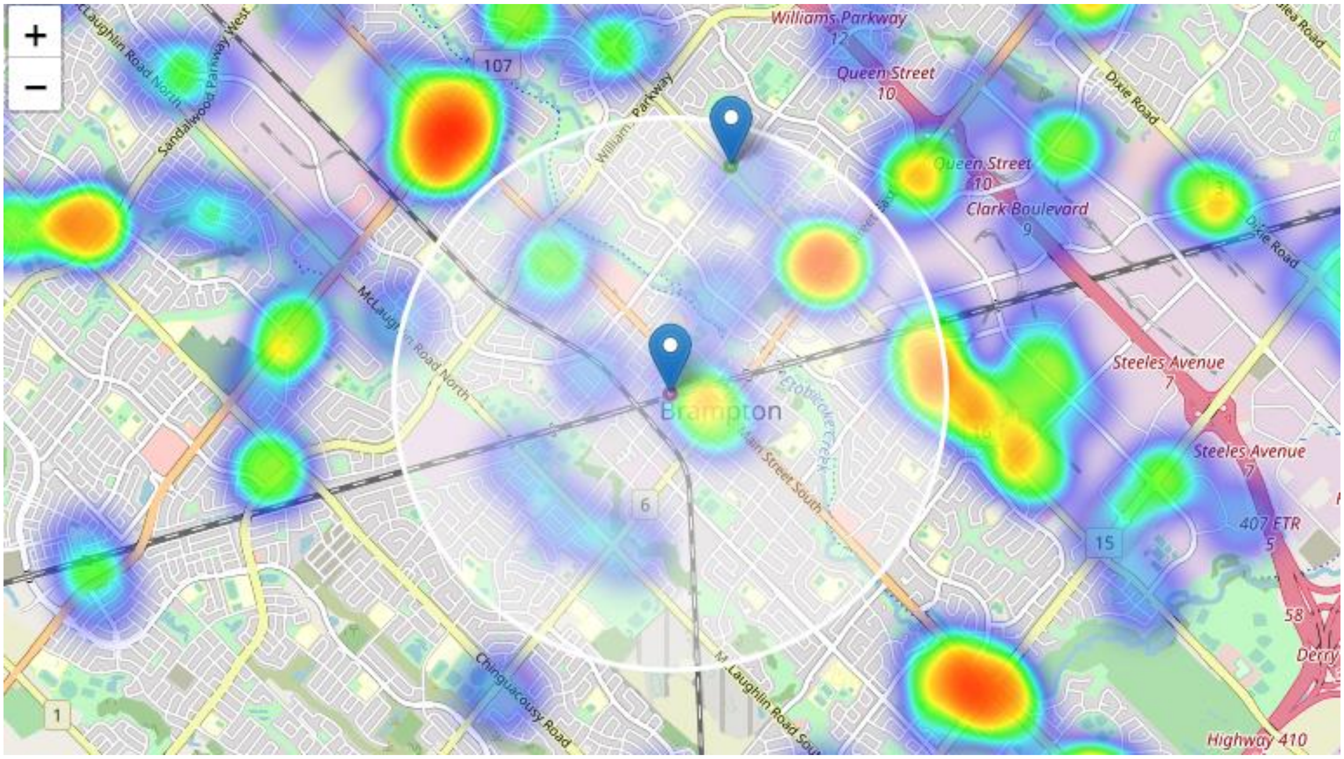

Figure 6. Our 'City Center' and 'RIO Center' mapped with markers alongside heat density of restaurants. (3)

Our second HEATMAP above has been narrowed down 2.2km radius area of interest. We've seen an attractive area of interest which is directly south (true north compass) of our 'City Center' marker, and we'll focus our clustering analysis here. We can refer to this area as the 'ROI Center'.

	Latitude	Longitude	X	Y	Restaurants nearby	Distance to South Asian restaurant
0	43.684365	-79.735351	-5.300953e+06	1.054752e+07	5	344.478096
1	43.686104	-79.735638	-5.300673e+06	1.054752e+07	6	127.015049
2	43.687842	-79.735924	-5.300393e+06	1.054752e+07	8	226.434876
3	43.689581	-79.736210	-5.300113e+06	1.054752e+07	8	468.272709
4	43.691320	-79.736497	-5.299833e+06	1.054752e+07	4	737.621791

Figure 7. Dataframe of ROI centroids with location and distance metrics. (3)

Now we can create centroids within our 'ROI radius' to prepare for clustering. In total we have 526 candidate neighborhood centers generated in our 'ROI radius'. We'll then load the data into a new (ROI Locations) dataframe which calculates additional metrics for spatial analysis, including count of nearby restaurants and distance from South Asian Restaurant. This metrics will help in our clustering analysis.

```
good_res_count = np.array((df_roi_locations['Restaurants nearby']<=1))
print('Locations with no more than 1 restaurants nearby:', good_res_count.sum())

good_ind_distance = np.array(df_roi_locations['Distance to southasian restaurant']>=600)
print('Locations with no South Asian restaurants within 600m:', good_ind_distance.sum())

good_locations = np.logical_and(good_res_count, good_ind_distance)
print('Locations with both conditions met:', good_locations.sum())

df_good_locations = df_roi_locations[good_locations]
```

```
Locations with no more than 1 restaurants nearby: 414
Locations with no South Asian restaurants within 600m: 454
Locations with both conditions met: 410
```

Figure 8. Stakeholder location restrictions to apply to our cluster analysis.

We can utilize our new 'ROI Locations' dataframe to produce marker mapping conditions. These conditions will allow us to restrict marker placement on areas which have less than 1 restaurant nearby and 600 meters away from any South Asian restaurants. This will suffice for our competition requirements set by our stakeholders.

```
from sklearn.cluster import KMeans

number_of_clusters = 35

good_xys = df_good_locations[['X', 'Y']].values
kmeans = KMeans(n_clusters=number_of_clusters, random_state=0).fit(good_xys)

cluster_centers = [xy_to_lonlat(cc[0], cc[1]) for cc in kmeans.cluster_centers_]
```

Figure 9. SciKit Learn package for the k-means clustering algorithm used.

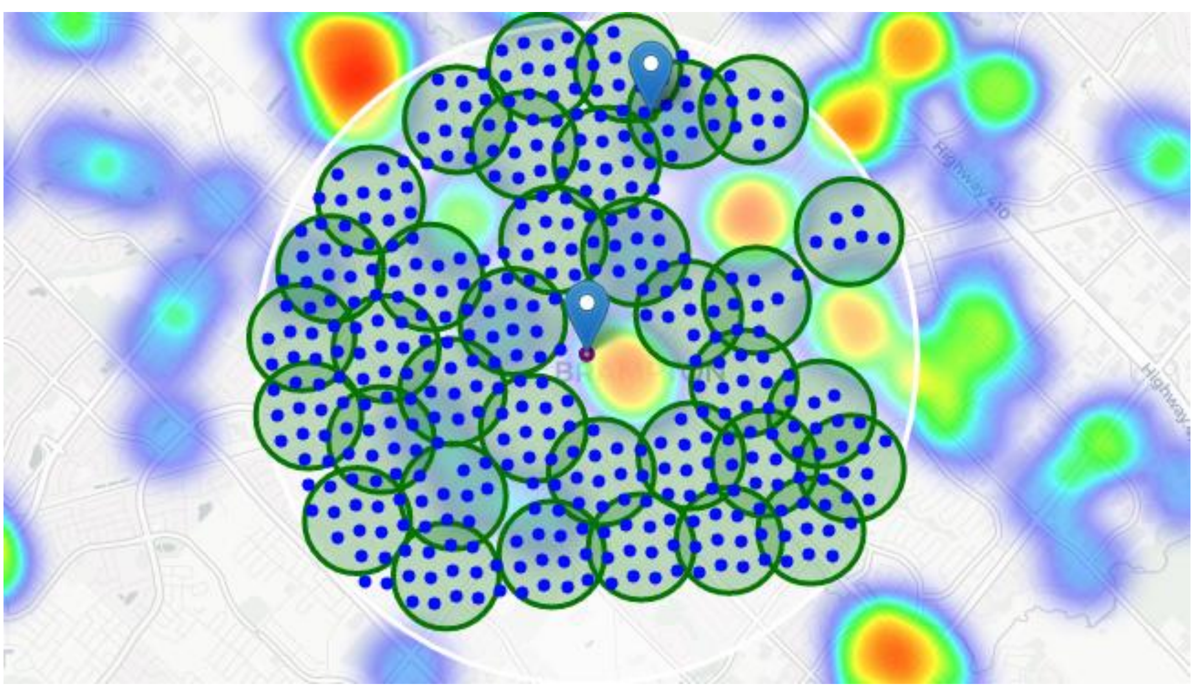

Figure 10. Final 35 clusters of 'good locations' which match our stakeholder requirements.

Through the use of the k-means clustering algorithm, we're able to create 35 cluster centers that group our 'good locations' into larger cohorts. At this point we should have enough supporting material to review street level mapping and find good commercial areas within our ROI circle.

5. Results and Discussion

We began our analysis by exploring the distribution of restaurants within a 7km radius of our selected city center. Through the use of the FourSquare API we were able to determine that there is a significant presence of South Asian restaurants within our radius, approximately 24.27% as reported by our FourSquare data. When we consider the 44% of South Asian residences within the entire city, there is definitely opportunity to for more south Asian restaurants.

```
import numpy as np

print('Total number of restaurants:', len(restaurants))
print('Total number of southasian restaurants:', len(southasian_restaurants))
print('Percentage of southasian restaurants: {:.2f}%'.format(len(southasian_restaurants) / len(restaurants) * 100))
print('Average number of restaurants in neighborhood:', np.array([len(r) for r in location_restaurants]).mean())
```

```
Total number of restaurants: 239
Total number of southasian restaurants: 58
Percentage of southasian restaurants: 24.27%
Average number of restaurants in neighborhood: 0.489010989010989
```

Figure 11. South Asian restaurant penetration in Brampton statistics (3).

To provide a better understanding of our area of analysis, it's important to understand that the city of Brampton is one of the more densely populated suburban cities within the country. Although there is a significant amount residential property already assigned throughout, many commercial lots exist around major intersections. Depending on the neighborhood; commercial lots can consist of strip plazas, indoor malls, commercial units or single structure buildings. Based on the average density of restaurants calculated, its likely to believe that much of the consumers will require transportation and possibly parking space to visit the location.

After effectively mapping our locations and calculating the relevant distance measurements, our heat map analysis of the city allowed us to narrow down to a smaller radius close to the city center. We refer to this area as our 'ROI Center', with a 2.2km radius. This area is directly south of our existing city center, and appears to have a low density of south Asian restaurants in a competitive distance. The area also encompasses a large portion on Main street, which is considered the downtown core. Another area of consideration was North/East Brampton, known for its strong south Asian presence and dense residential zoning. Most commercial lots consist of small to medium size strip plaza's. For the purpose of our analysis, the 'ROI Center' location was chosen as the 'South Brampton' area due its close proximity to the 'City Center'.

Once we were able to cluster zones closer to the street level using the k-means algorithm, we explored plausible locations that within our 'ROI zone' that would meet our stakeholder requirements. Recommended locations were analyzed and verified visually as commercially zoned lots. Although these locations meet our competitive distance requirements, there may be more research required to assess other optimization factors not explored in this analysis. In total there were 6 locations selected from our analysis:

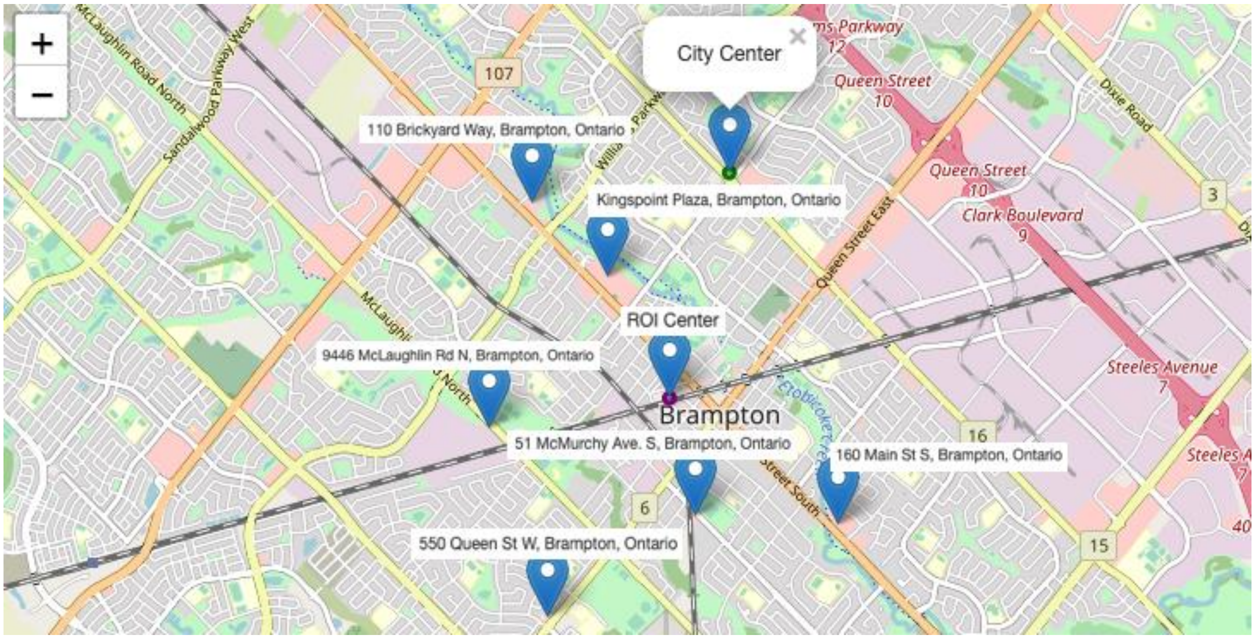

Figure 12. Our top 6 ROI locations selected after analysis.

Location 1: 110 Brickyard Way, Brampton, Ontario: 43.700982, -79.778761

- *Small commercial plaza located directly off Main street, close to two major intersections.*

Location 2: Kingspoint Plaza, Brampton, Ontario: 43.695593, -79.769250

- *Larger commercial strip plaza located directly off Main street, closest to our city center.*

Location 3: 9446 McLaughlin Rd N, Brampton, Ontario: 43.684786, -79.783054

- *Small commercial plaza located directly off McLaughlin street, mixed planning in surrounding area*

Location 4: 51 McMurchy Ave. S, Brampton, Ontario: 43.678406, -79.762528

- *Small commercial plaza located off of a inner city street, directly beside a train track. Mostly surrounded by residential land.*

Location 5: 160 Main St S, Brampton, Ontario: 43.677839, -79.748198

- *Medium sized commercial plaza located off of Main Street South, between Queen Street and Steeles.*

Location 6: 550 Queen St W, Brampton, Ontario: 43.671132, -79.777226

- *Small commercial plaza located off of Queen Street, this would be the further proposed location from the 'City Center'.*

6. Conclusion

Our stakeholders requested an analysis of the city of Brampton's restaurant landscape, in order to find ideal commercial location for a south Asian restaurant based on specific requirements. By establishing the correct steps to collect, process and analyze FourSquare locational data in a python environment, we were able to conclusively explore the restaurant distribution across the city. Once our clustering of 'ROI centroids' was complete, it was much easier to explore at street level within a 2.2 km radius. The Folium map layer distinguishes between deferent land use types (residential, industrial, commercial ect.) and with the cluster mapping; top ROI locations we're quickly narrowed down for this analysis. Ideally this final step would be more efficiently completed if code automated selecting commercial zones only. This can be something to explore in order to better automate the overall project.

Our strongest point of reference for the success of a South Asian restaurant in Brampton, is that the city has a very prominent South Asian community (1). An interesting comparison would be to complete a similar ethnic restaurant/population ratio analysis with other cities within the Greater Toronto Area. There are likely very strong insights to be gained in terms of understanding the potential success south Asian restaurants can have within its primary ethnic enclave, as compared to the success of other ethnic restaurants in their respective enclaves.

7. Data Sources

1. StatsCan https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm

2. Google Geocoding API

3. FourSquare Developer API