FINDING THE BEST NEIGHBOURHOOD IN TORONTO TO OPEN A PAKISTANI RESTAURANT

(By: Sarwat Sohail)

INTRODUCTION:

Toronto is the largest and one of the most important cities of Toronto, with a population of 2,731,571 people. Not only is Toronto the most populated city in Canada, it is also the fourth-largest city in all of North America. Toronto's cuisine represents the city's diversity, with different ethnic neighbourhoods throughout the city focusing on different cuisines. A number of culinary festivals take place in Toronto each year. In addition, food tours in Toronto are an increasingly popular way for locals and tourists to explore the food culture of the city.

Toronto is also one of the most popular places in the world for tourists; around about 43.7 million tourists visited Toronto in 2017 alone.

Business Problem:

Being such a vibrant, multicultural hub, Toronto has people of a wide variety of ethnicities, who would undoubtedly enjoy a wide variety of cuisines. For this purpose, our investor has decided to explore its neighbourhoods to see which one is ideal for opening a new restaurant serving mainly Pakistani cuisine.

Interest:

This project can be of interest to a wide variety of investors. The food industry is usually one of the safer choices of business worldwide, and therefore anyone looking to invest their money into a new venture could find this project useful. This project can also be useful to researchers and social scientists looking to get a better grasp on Toronto's dining culture.

DATA ACQUISITION AND CLEANING:

Sources:

The data sources for this notebook are the following:

- The Neighbourhood Profiles from the 2016 Census Data, obtained from the City of Toronto's open data portal: https://open.toronto.ca/dataset/neighbourhood-profiles/
- The location data for the restaurants in the city, obtained from the Foursquare location data through its API.
- Data on the neighbourhoods of Toronto, obtained by scraping this Wikipedia page.

Data Cleaning and Feature Selection:

The neighbourhood data that was scraped was cleaned and transformed before being placed into a table, with each row specifying a neighbourhood, and the columns detailing the neighbourhood name, its borough, latitude, longitude, and the 'unofficial' neighbourhood names that it covers. It is this final table that was imported into our Jupyter Notebook.

For the Neighbourhood Profiles dataset, it was a massive dataset with 2000+ rows detailing information on each neighbourhood, while the columns each specified a neighbourhood. From this entire dataset, we took only one row: the row having the '**Ethnic origin – Pakistan**' characteristic. This is because that is the only data we needed for this project; any other data would only have muddled our results.

The Neighbourhood Profile dataset also had the problem that all of its numerical values were actually in the form of strings that had commas in them. This necessitated us to first remove the commas, then convert each value into integer form for ease of operations.

The json file returned by the Foursquare location data was also whittled down to only the latitude, longitude, venue and venue category for each result; further, the venues were filtered so that the final dataset only contained data on restaurants. After this, the json file was converted into a dataframe.

METHODOLOGY:

EXPLORATORY DATA ANALYSIS:

To explore our data, the first thing we did was to display all the neighbourhoods out on a map of Toronto.

Map of Toronto showing all its neighbourhoods

This showed us that the majority of the neighbourhoods were clusters to the south, with a slight density to the east; otherwise, the neighbourhoods were spread out uniformly all over the city.

Locating all the restaurants in Toronto:

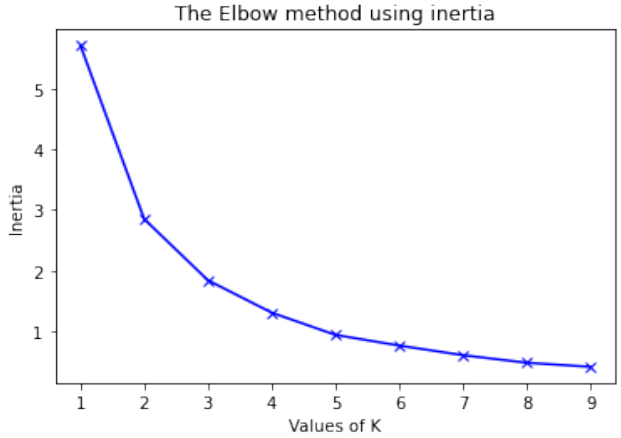
Then, we mapped all the restaurants we'd found upon the map, using different colours for each neighbourhood, to create another map:

Map of all the restaurants in Toronto

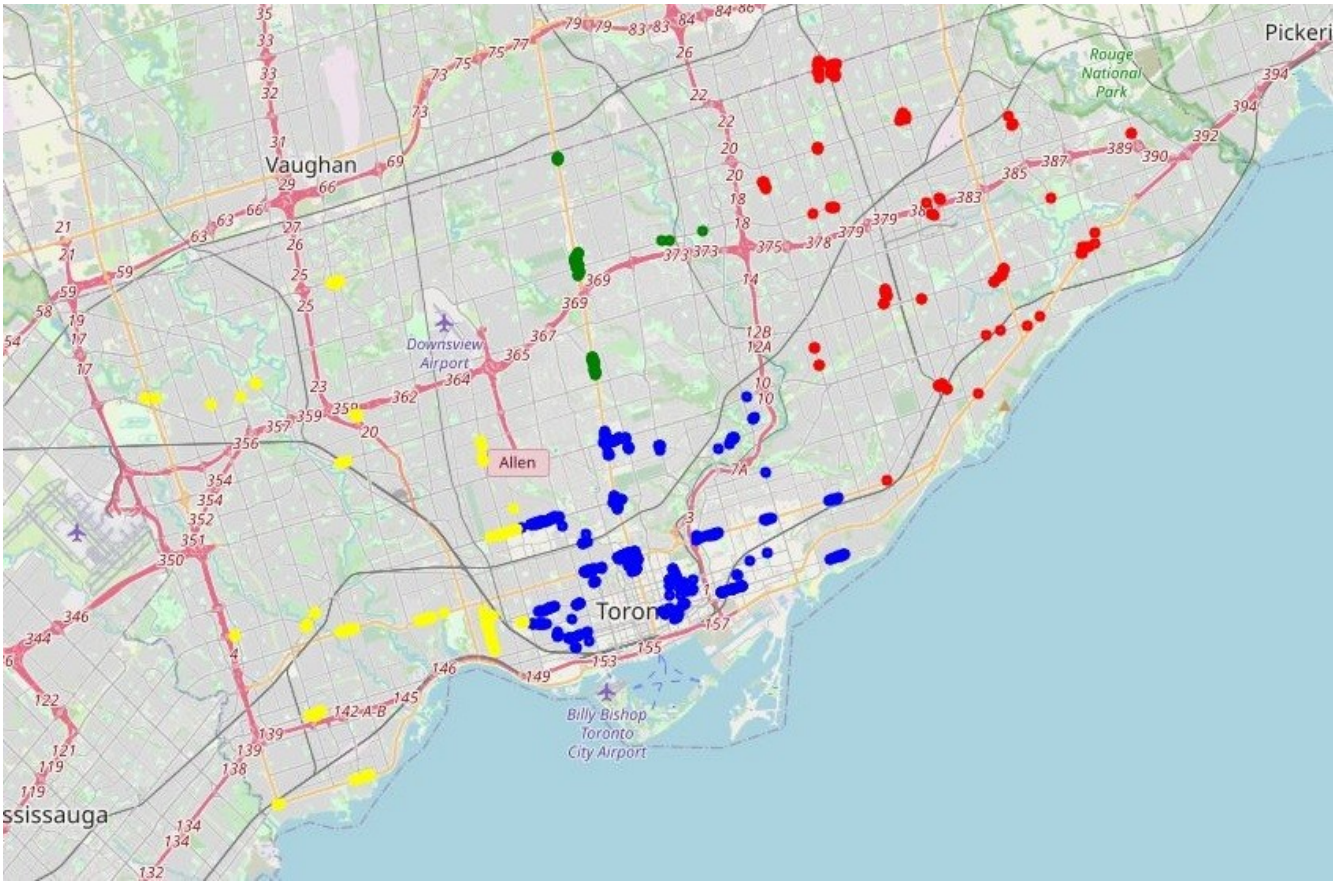
The opaque dots are the restaurants, while the circles represent the central point of the neighbourhoods themselves. Once again, we saw that the majority of the restaurants were located in the south, near the beach-front.

K-means clustering of restaurants:

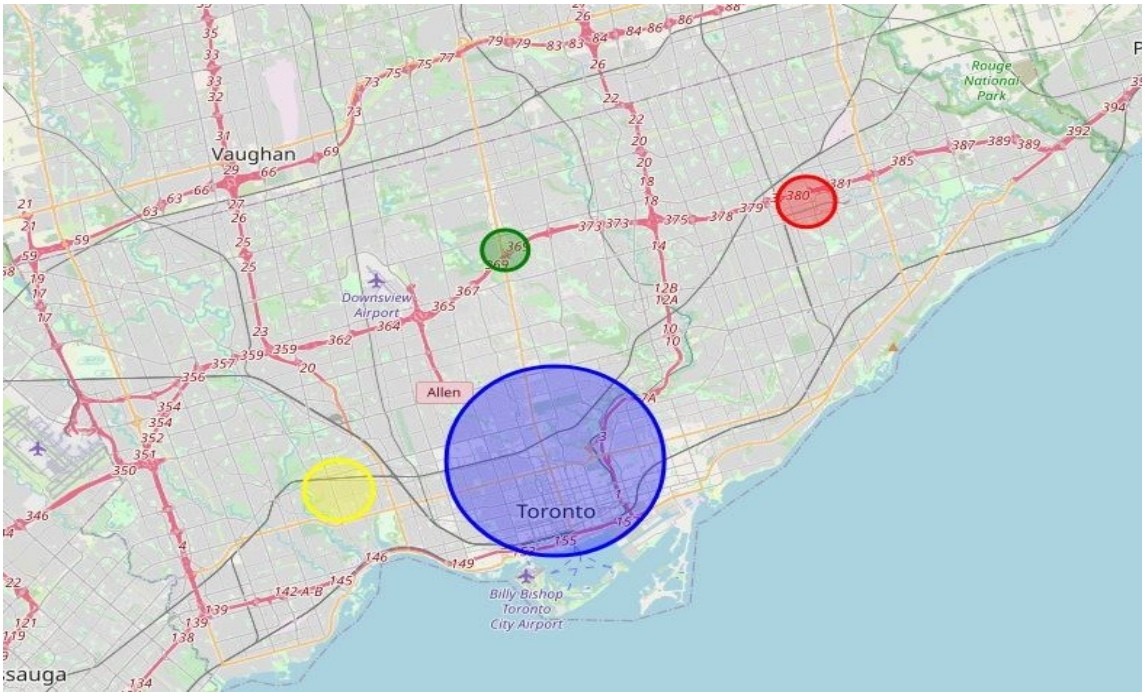
Visually, the only thing we can see is that there is a greater number of restaurants in the south compared to the rest of the city. We clustered the latitudes and longitudes of the data points using k-means, in order to see if there were any significant clusters. We used k-means inertia values in the Elbow Method to calculate the optimum value of k. The results obtained were as follows:



The following line chart showed us that there wasn't as sharp an elbow point as we would have hoped; still the optimum point appeared to be at k =4. So we used k = 4 to generate the clusters for the restaurant in the city:



Restaurants of Toronto colour-coded into clusters

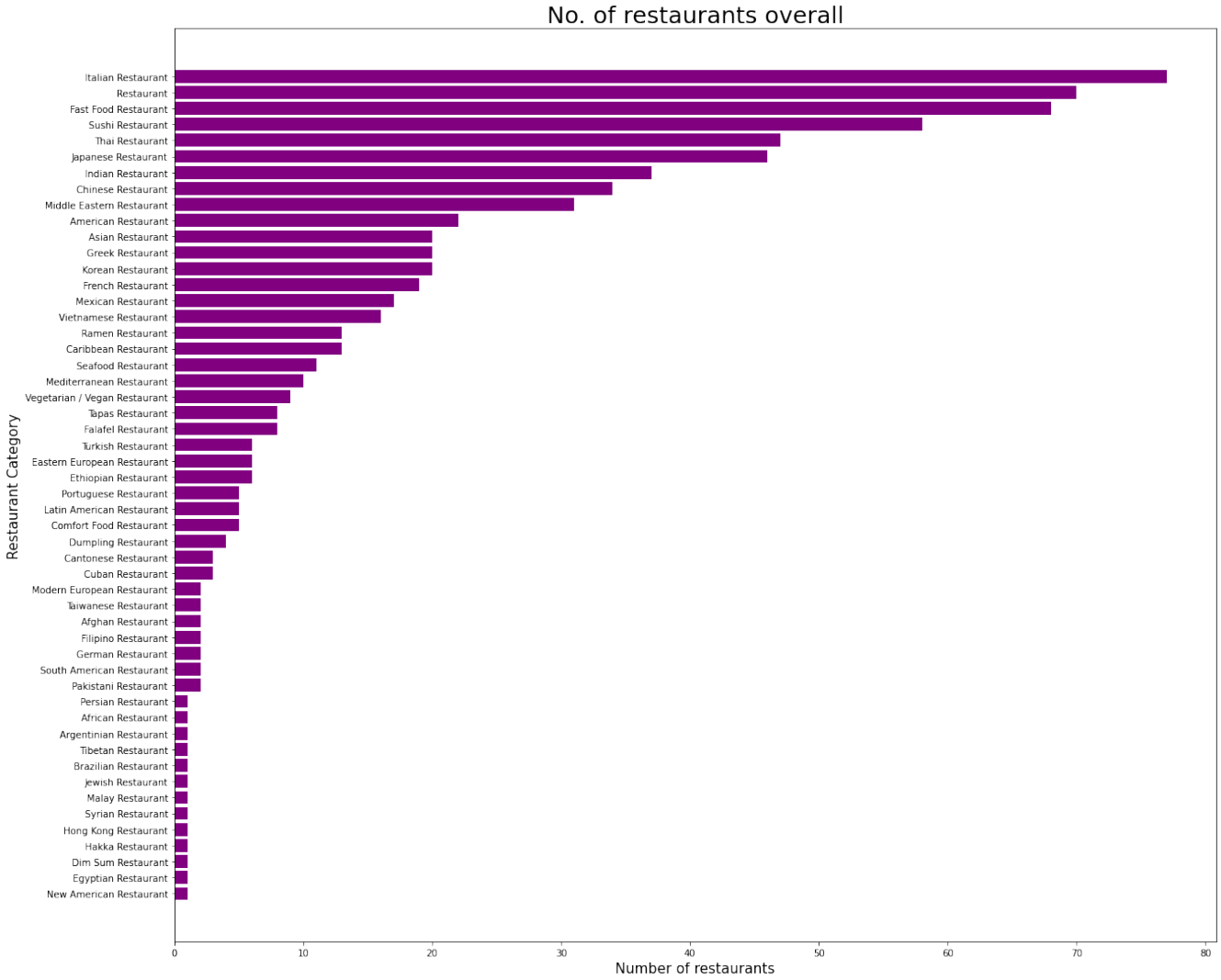


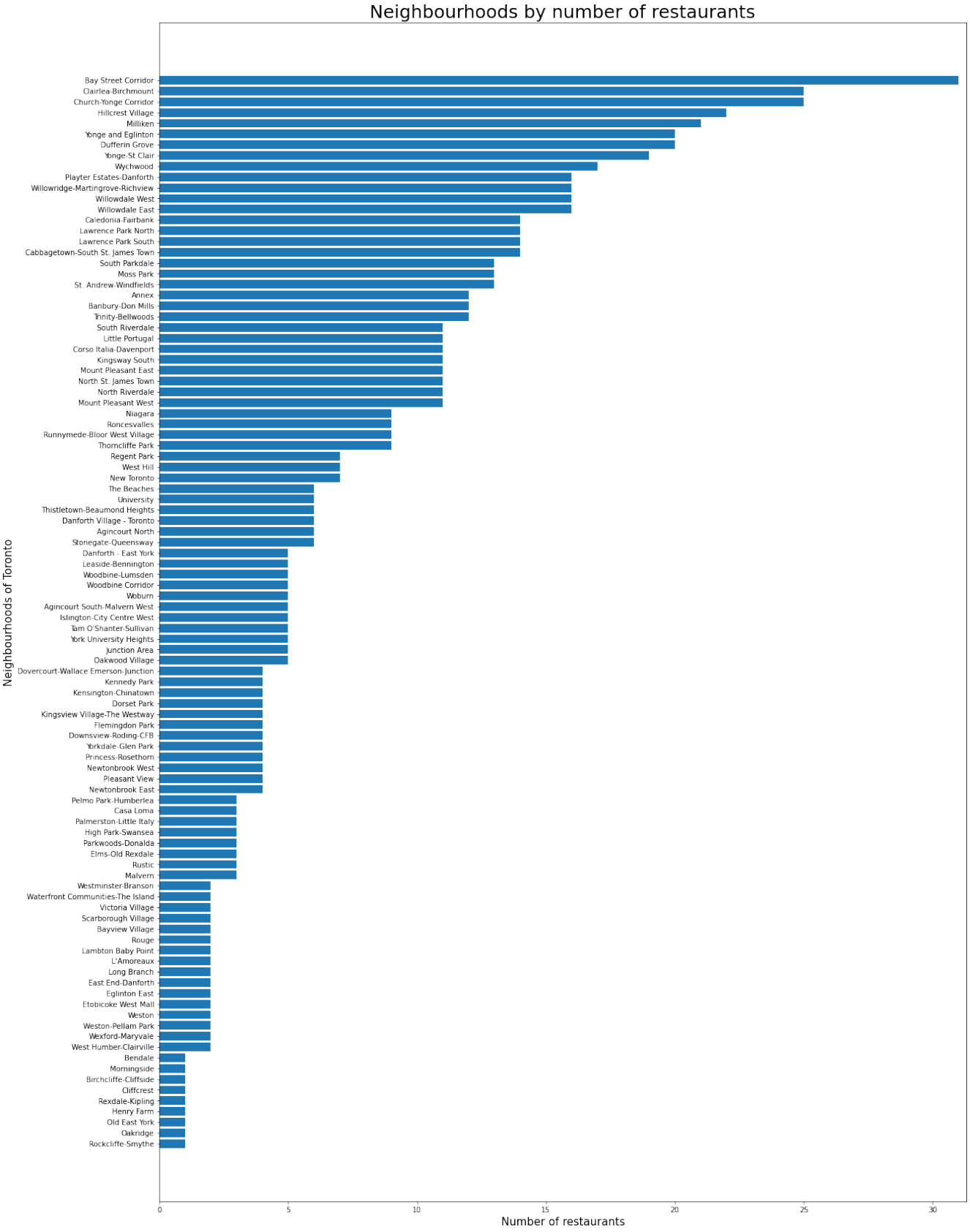
Clusters of restaurants by no. of restaurants

We see the four clusters, depicting regions were restaurants are in the highest quantity. As per our previous observations, the cluster in the south is the largest, with a secondary cluster to the left, and smaller clusters to the east and north.

Further analysis on already existing restaurants:

We then divided the data points in the restaurant dataframe by neighbourhood as well as cuisine, to get a better idea of which neighbourhoods were popular spots for restaurants, and where Pakistani restaurants stood in the mix. The results were as follows:


No. of restaurants overall
Italian Restaurant
Restaurant
Fast Food Restaurant
Sushi Restaurant
Thai Restaurant
Japanese Restaurant
Indian Restaurant
Chinese Restaurant
Middle Eastern Restaurant
American Restaurant
Asian Restaurant
Greek Restaurant
Korean Restaurant
French Restaurant
Mexican Restaurant
Vietnamese Restaurant
Ramen Restaurant
Caribbean Restaurant
Seafood Restaurant
Mediterranean Restaurant
Vegetarian / Vegan Restaurant
Tapas Restaurant
Falafel Restaurant
Turkish Restaurant
Eastern European Restaurant
Ethiopian Restaurant
Portuguese Restaurant
Latin American Restaurant
Comfort Food Restaurant
Dumpling Restaurant
Cantonese Restaurant
Cuban Restaurant
Modern European Restaurant
Taiwanese Restaurant
Afghan Restaurant
Filipino Restaurant
German Restaurant
South American Restaurant
Pakistani Restaurant
Persian Restaurant
African Restaurant
Argentinian Restaurant
Tibetan Restaurant
Brazilian Restaurant
Jewish Restaurant
Malay Restaurant
Syrian Restaurant
Hong Kong Restaurant
Hakka Restaurant
Dim Sum Restaurant
Egyptian Restaurant
New American Restaurant
Restaurant Category
0
10
20
30
40
50
60
70
80
Number of restaurants

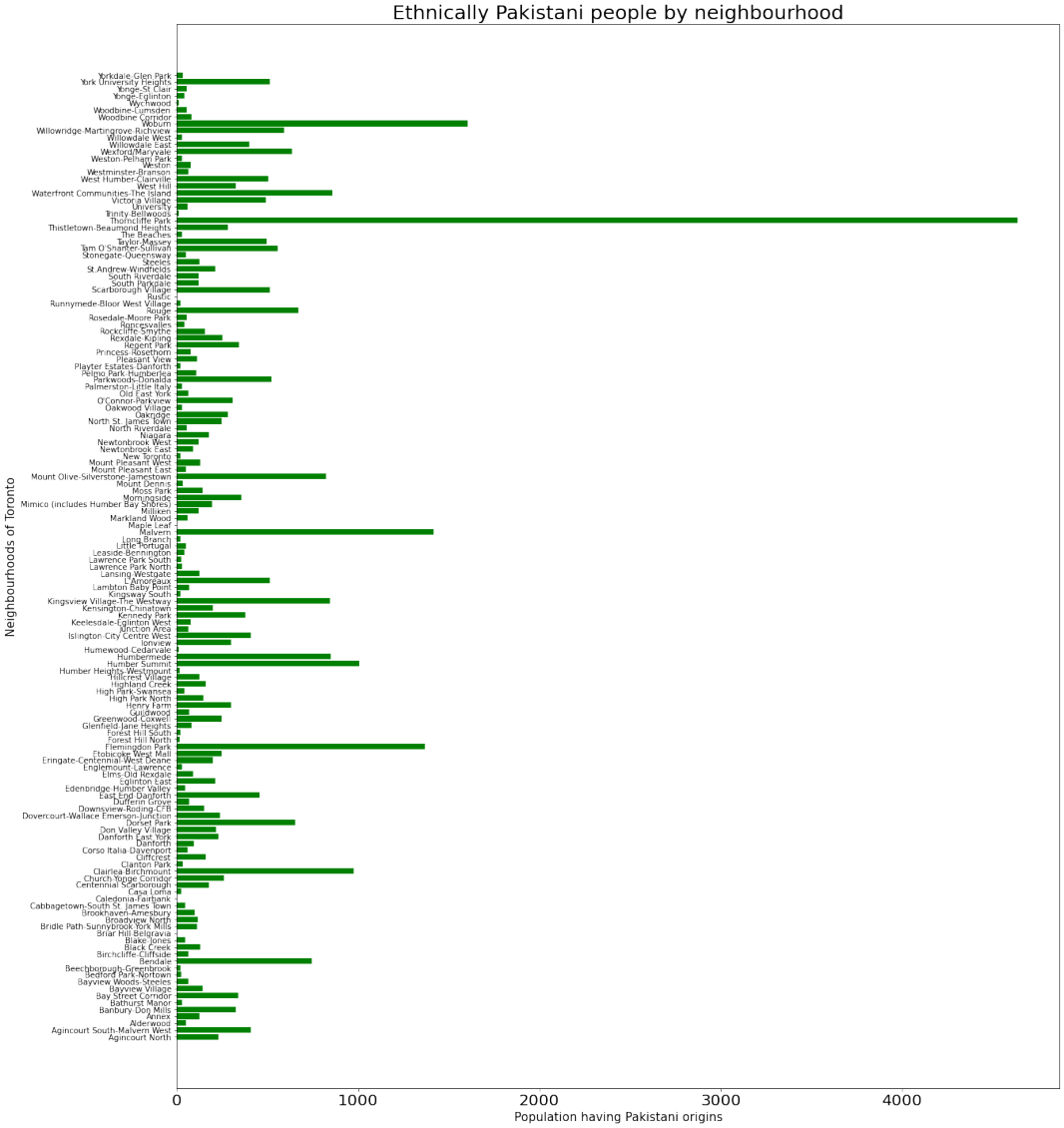

Neighbourhoods by number of restaurants
Neighbourhoods of Toronto
Bay Street Corridor
Clairlea-Birchmount
Church-Yonge Corridor
Hillcrest Village
Milliken
Yonge and Eglinton
Dufferin Grove
Yonge-St.Clair
Wychwood
Playter Estates-Danforth
Willowridge-Martingrove-Richview
Willowdale West
Willowdale East
Caledonia-Fairbank
Lawrence Park North
Lawrence Park South
Cabbagetown-South St. James Town
South Parkdale
Moss Park
St. Andrew-Windfields
Annex
Banbury-Don Mills
Trinity-Bellwoods
South Riverdale
Little Portugal
Corso Italia-Davenport
Kingsway South
Mount Pleasant East
North St. James Town
North Riverdale
Mount Pleasant West
Niagara
Roncesvalles
Runnymede-Bloor West Village
Thorncliffe Park
Regent Park
West Hill
New Toronto
The Beaches
University
Thistletown-Beaumond Heights
Danforth Village - Toronto
Agincourt North
Stonegate-Queensway
Danforth - East York
Leaside-Bennington
Woodbine-Lumsden
Woodbine Corridor
Woburn
Agincourt South-Malvern West
Islington-City Centre West
Tam O'Shanter-Sullivan
York University Heights
Junction Area
Oakwood Village
Dovercourt-Wallace Emerson-Junction
Kennedy Park
Kensington-Chinatown
Dorset Park
Kingsview Village-The Westway
Flemingdon Park
Downsview-Roding-CFB
Yorkdale-Glen Park
Princess-Rosethorn
Newtonbrook West
Pleasant View
Newtonbrook East
Pelmo Park-Humberlea
Casa Loma
Palmerston-Little Italy
High Park-Swansea
Parkwoods-Donalda
Elms-Old Rexdale
Rustic
Malvern
Westminster-Branson
Waterfront Communities-The Island
Victoria Village
Scarborough Village
Bayview Village
Rouge
Lambton Baby Point
L'Amoreaux
Long Branch
East End-Danforth
Eglinton East
Etobicoke West Mall
Weston
Weston-Pellam Park
Wexford-Maryvale
West Humber-Clairville
Bendale
Morningside
Birchcliffe-Cliffside
Cliffcrest
Rexdale-Kipling
Henry Farm
Old East York
Oakridge
Rockcliffe-Smythe
0
5
10
15
20
25
30
Number of restaurants

The first chart showed us one very important thing: that there are only two Pakistani restaurants in Toronto. The second chart showed us that the majority of the restaurants in the city are to the south, meaning that restaurants near those neighbourhoods are likelier to find success.

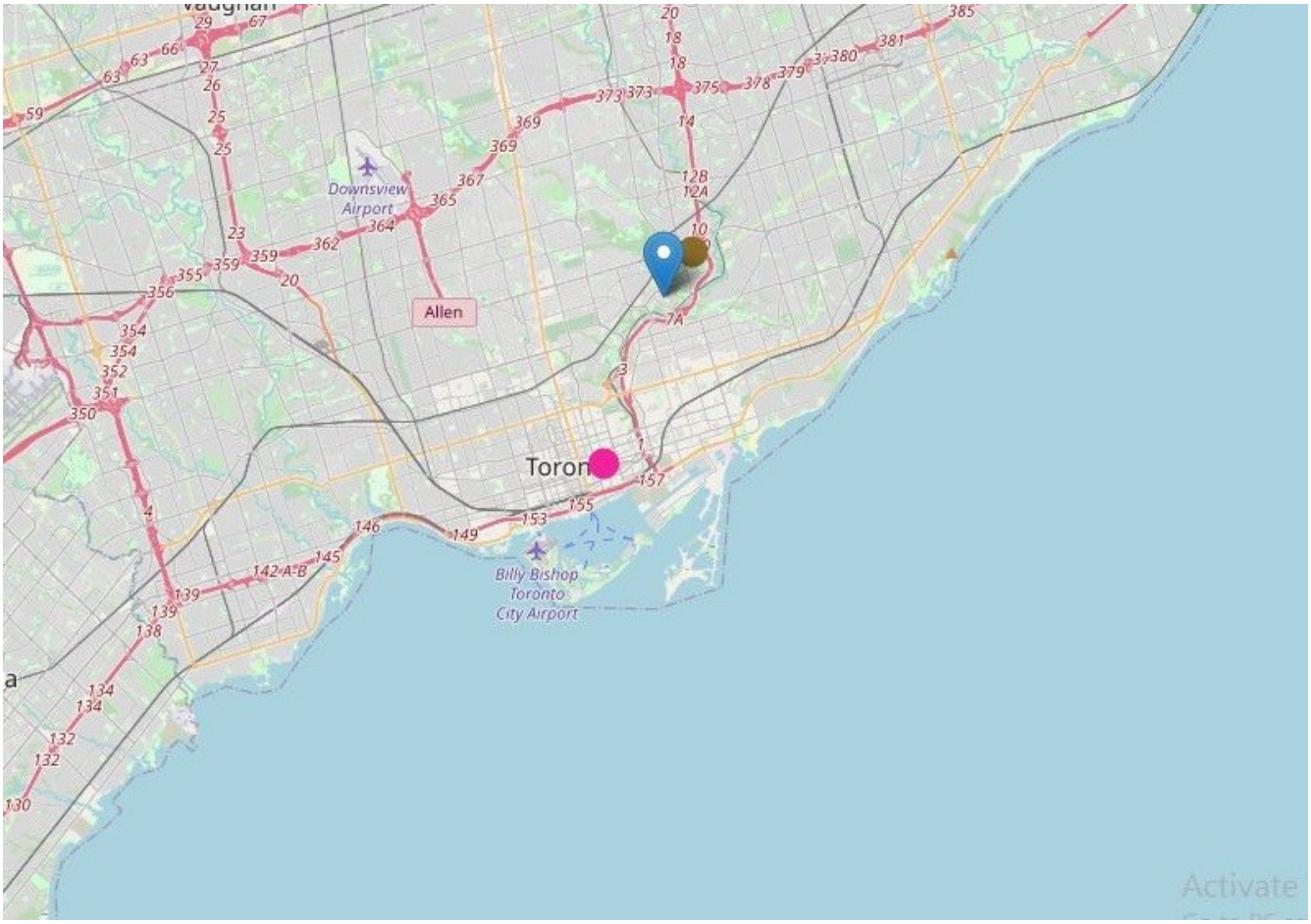
The low number of restaurants serving Pakistani cuisine could be due to many factors, but it does indicate that for whatever reason, they are present in low numbers. A good way to proceed was to find the neighbourhoods having a high number of people who were ethnically Pakistani, as they would be more likely to appreciate, and thus patronize, a Pakistani restaurant. With this, we turned to our next analysis.

Analysing neighbourhoods having a higher number of Pakistanis:

For this analysis, we used the data we obtained from the neighbourhood profiles dataset. Plotting the numbers on a bar chart produced the following result:



This showed us that the majority of neighbourhoods have a low Pakistani population – less than 1000, in fact. Only Thorncliffe Park stood out, and even then it only has around 4000+ people of Pakistani origin. We searched for the location of Thorncliffe Park on the map and plotted it with the locations of the two Pakistani restaurants, to obtain the following result:

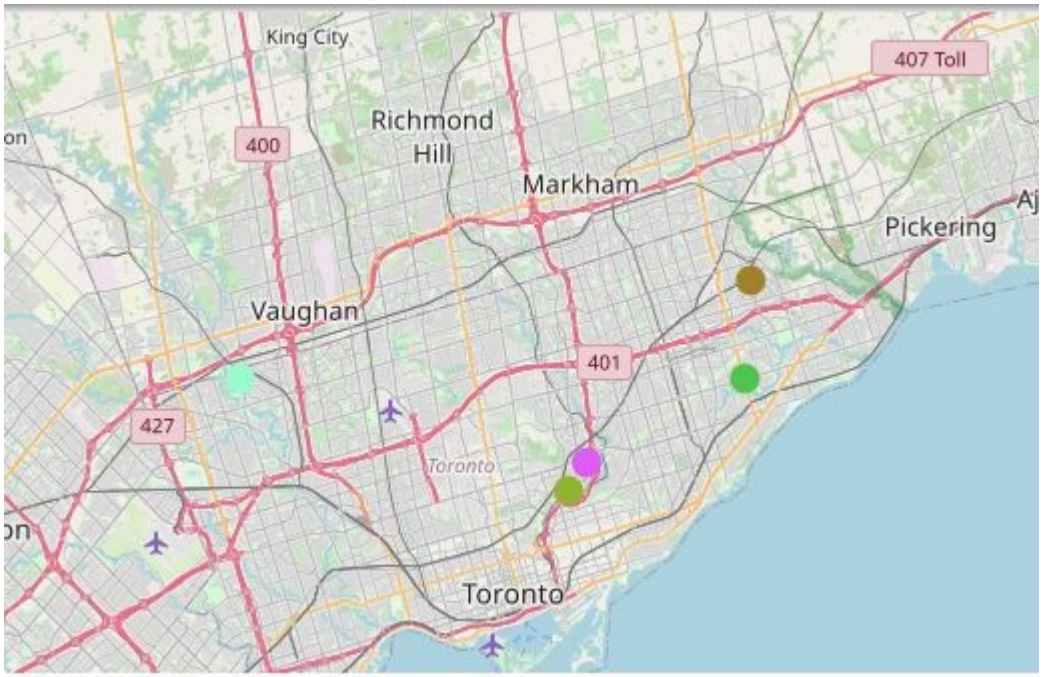


Location of Thorncliffe Park against the locations of Pakistani restaurants

This showed us an unsurprising result: that the two Pakistani restaurants are located around the neighbourhood with the greatest number of Pakistanis. This seems to confirm our initial assessment, that a restaurant close to a neighbourhood with a large number of people having Pakistani origins will help our restaurant be successful.

Neighbourhoods with people having Pakistani origins:

We filtered the dataframe to find the neighbourhoods having more than a 1000 people of Pakistani origin. There were five of these neighbourhoods: Thorncliffe Park, Woburn, Malvern, Flemingdon Park and Humber Summit. Of these, Flemingdon Park already has a Pakistani restaurant; however, it has relatively few restaurants, a total of four in number. We shall plot their locations on the map to see where they are exactly:



Five neighbourhoods with greatest number of people having Pakistani origin

The five dots show us the five neighbourhoods having the highest number of people with Pakistani origins. With our data analysed thus far, we move forward towards a solution.

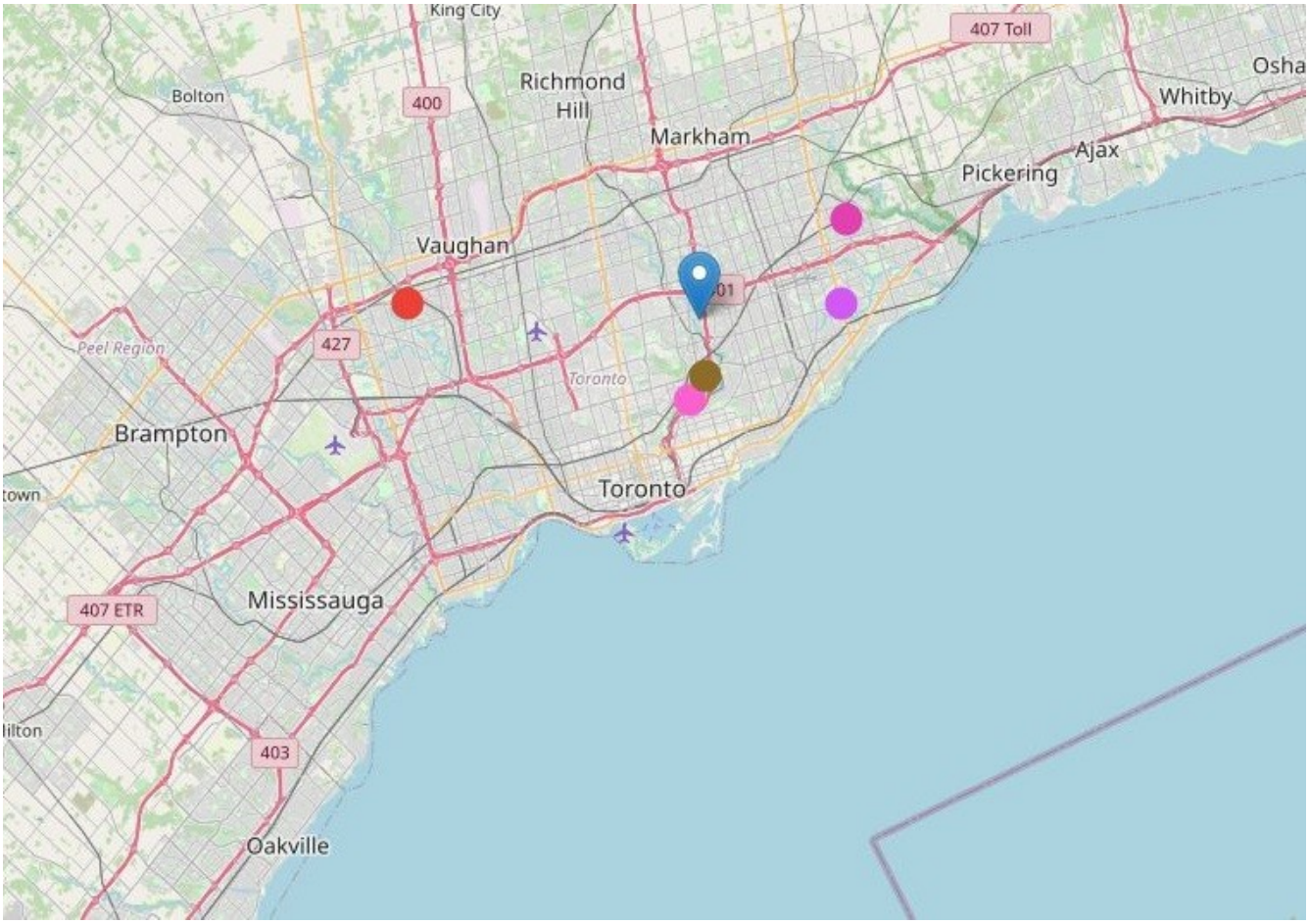
PREDICTIVE ANALYSIS:

Finding the optimum spot:

The first thing that we do is to find the centre point between all five locations. I used the Numpy library to average out their latitudes and longitudes, resulting in a set of coordinates that were directly in the centre of all five neighbourhoods. The coordinates were these:

```
[43.75041642000001, -79.33947255999999]
```

And the place they pinpointed to on the map was the following:



Central Point between all five neighbourhoods

The pin marker shows the centre point, while the dots are the five neighbourhoods. However, three of the locations (Humber Summit, Malvern and Woburn) are far away from each other; consequently, the location we've pinpointed is far away from all the neighbourhoods, and thus its success is likely to be lowered.

Using K-means clustering on the five neighbourhoods:

To get better results, we cluster the five neighbourhoods we've obtained. We use K-means clustering using the latitudes and longitudes of their locations once again, using k = 3 as it is visually obvious that there are three clusters. We get the following three coordinates:

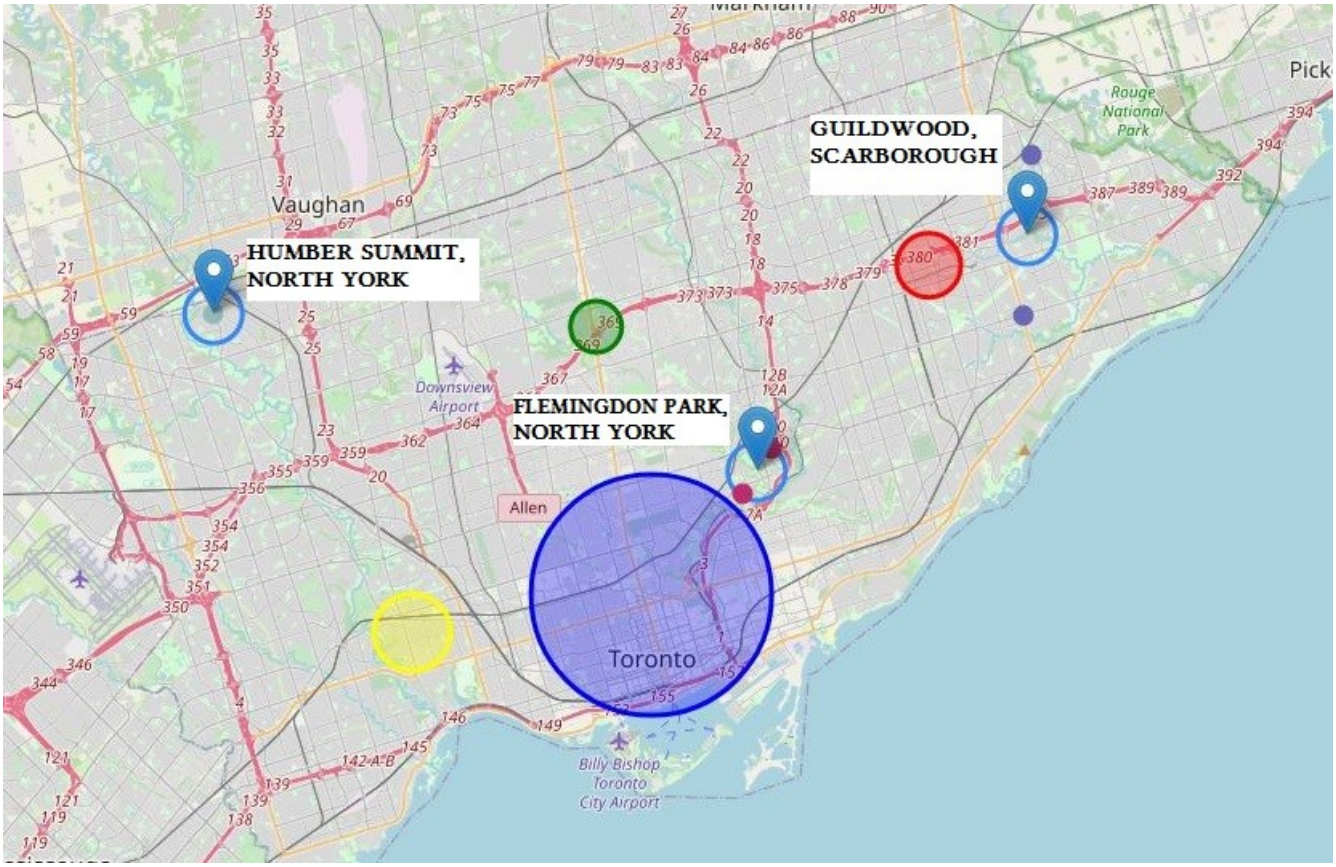
```
[43.711492, -79.339306]
[43.784510, -79.223496]
[43.760078, -79.571760]
```

We used the geocoders reverse coding to obtain their addresses. The snapshot of the table given below gives the details of each location:

	Cluster Labels	Optimum Point Latitude	Optimum Point Longitude	Neighbourhood	Borough
0	1	43.711492	-79.339306	Flemingdon Park	North York
1	2	43.784510	-79.223496	Scarborough—Guildwood	Scarborough
2	0	43.760078	-79.571760	Humber Summit	North York

Plotting the final three points on the map:

We plot these points out onto the map along with the five neighbourhoods, as well as the cluster of restaurants we created earlier:



Optimum locations to open a Pakistani restaurant

The opaque circles represent the five neighbourhoods that have the highest population of people having Pakistani origins. They match colours with other neighbourhoods close to them (ie., in their cluster).

The pin markers represent the optimum points to open a restaurant for each cluster of neighbourhoods.

The translucent circles represent clusters of restaurants, with the size of the circles representing the number of restaurants close by.

RESULTS:

So we have found three locations in three different neighbourhoods where we could possibly open a restaurant serving Pakistani cuisine. These are as follows:

- **NEIGHBOURHOOD 1: HUMBER SUMMIT, NORTH YORK:**

It has the advantage of being right in the centre of the neighbourhood, as well as being located in an area with relatively few restaurants. However, it has the disadvantage that the ethnic Pakistani population there is not very large, and the lesser number of restaurants, as well as its distance from any of the restaurant clusters, may indicate a disinclination of the locals to eat out.

- **NEIGHBOURHOOD 2: FLEMINGDON PARK, NORTH YORK:**

It has the advantage of being in a neighbourhood having a relatively large Pakistani population, as well as having the neighbourhood with the largest population of people with Pakistani origins (Thorncliffe Park) close by. It is also the closest to the largest restaurant cluster, meaning that there is a good possibility of getting customers who might be willing to try a new restaurant. That latter advantage, however, may also be a disadvantage, in that our new restaurant may face tough competition from established restaurants, including the only two Pakistani restaurants in Toronto, which are also close by.

- **NEIGHBOURHOOD 3: GUILDWOOD, SCARBOROUGH:**

It has the advantage of being between two neighbourhoods with a relatively high population of people with Pakistani origins. It also has the advantage of being a moderate distance from a restaurant cluster, meaning that the people of the neighbourhoods might have a propensity to eat out, and would likely appreciate a new restaurant opening nearby. It has the fewest disadvantages - its biggest disadvantage would be that the area is 'average' for restaurants, and could be a hit or a miss depending on the locals' propensity for the cuisine. Nevertheless, this location is the likeliest for a new restaurant, and should be considered first by potential investors.

DISCUSSION:

Our analysis revealed that there are 744 restaurants in Toronto, only two of them serve Pakistani cuisine. Also, most of the restaurants are to the south of the city, with smaller clusters to the east, possibly for the reason that sea is to the south, and these dense clusters in the south may represent tourist or general entertainment areas, where people might be more likely to get out. Keeping these points in mind, we proceeded with two assumptions: that the people most likely to give a new Pakistani restaurant a try would be people of Pakistani origin, or people who are generally looking to eat out and may try new things. Thus, we searched for places closest to the neighbourhoods having people of Pakistani origin. We obtained three possible locations. Two of the three places we discovered were close to restaurant clusters (i.e., in the east and south of the city). The third was in a place with a relative scarcity of restaurants, most of them scattered, where trying to move it close to a restaurant cluster would have likely skewed our data.

Of course, this was only an initial analysis. Further research needs to be done to determine whether these locations are truly optimum for our purposes; socioeconomic factors such as income and age, as well as residential/commercial make-up of the neighbourhoods could well be key to deciding whether or not they are ideal locations for our purpose.

CONCLUSION:

We started this project to find the best neighbourhood to open Pakistani restaurant in Toronto. We used the census data obtained from the City of Toronto's open data portal, as well as location data from Foursquare. We also used the geoocoders API to obtain addresses and geospatial coordinates. After cleaning and processing the data, it was grouped, filtered, and then clustered by means of the K-Means module from the sklearn library, with the results plotted onto a map of Toronto to pinpoint the ideal locations for the restaurant.

By analysing the ethnic make-up of the neighbourhoods, as well identifying the locations of already existing restaurants in the city, we were able to narrow our search down to three possible candidates: Guildwood, Flemingdon Park, and Humber Summit, each of which has its own advantages and disadvantages. Further investigation needs to be done in order to make a final decision.