

## **Introduction**

Toronto is the most populous city in Canada with over 6 million people. Being a multicultural and cosmopolitan city, people of varied ethnic backgrounds dwell here. As Toronto is also a business hub, visitors from around the globe arrive here and hang out in restaurants, pubs and shopping malls. With an influx of people, new restaurants pop up on a frequent basis to satiate the visitors' palates.

A study by Ohio State University revealed that nearly 60% of restaurants close or change ownership within the first year and 80% fail within the first 5 years. With the average restaurant profit margins falling between 3-5%, it's vital to set the business up for success from the beginning by having a strategic plan for significant foot fall and minimum competition.

Two important points to consider before setting up a restaurant in Toronto are to understand the ethnic backgrounds of its residents and keep the competition at a bare minimum.

**This project focuses on assisting any party interested in opening a restaurant by:**

- a) Identifying the ethnic backgrounds of residents in a given neighbourhood in Toronto and**
- b) Locating a neighbourhood where the competition of similar restaurants is lowest**

**Any potential entrepreneur looking to open a restaurant in the city of Toronto can be benefited by this analysis** as the objective of this project is to cluster the neighbourhoods based on the ethnicity of its residents and also intend to identify the presence of similar restaurants which would minimize competition.

Based on the food preferences of people living in the vicinity of the restaurant the right type of cuisine can be offered to the customers. When there is limited competition more customers can be expected to visit our restaurants and thereby maximizing business and profits. We use Foursquare API to retrieve and study any existing competitors in the neighbourhood. This is done by filtering the specific list of restaurants from the overall list of venues we obtain from the Foursquare API query.

Hence, it is imperative that anyone interested in opening a new restaurant should study the demographic distribution of the neighbourhood and identify any competition before deciding on the type of restaurant and its location.

## **Data**

Demographic data about Toronto is available on the [www.toronto.ca](http://www.toronto.ca) website.

Data: [https://www.toronto.ca/ext/open\\_data/catalog/data\\_set\\_files/2016\\_neighbourhood\\_profiles.csv](https://www.toronto.ca/ext/open_data/catalog/data_set_files/2016_neighbourhood_profiles.csv)

The above file has comprehensive data of the residents of Toronto city which includes:

- a) Population
- b) Income

- c) Education
- d) Ethnic origin
- e) Languages spoken
- f) Mobility

Below is how the data looks:

	Category	Topic	Data Source	Characteristic	City of Toronto	Agincourt North	Agincourt South-Malvern West	Alderwood	Annex	Banbury-Don Mills	...	Willowdale West	Willow Martin Ric
0	Neighbourhood Information	Neighbourhood Information	City of Toronto	Neighbourhood Number	NaN	129	128	20	95	42	...	37	
1	Neighbourhood Information	Neighbourhood Information	City of Toronto	TSNS2020 Designation	NaN	No Designation	No Designation	No Designation	No Designation	No Designation	...	No Designation	Desig
2	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2016	2,731,571	29,113	23,757	12,054	30,526	27,695	...	16,936	4
3	Population	Population and dwellings	Census Profile 98-316-X2016001	Population, 2011	2,615,060	30,279	21,988	11,904	29,177	26,918	...	15,004	4
4	Population	Population and dwellings	Census Profile 98-316-X2016001	Population Change 2011-2016	4.50%	-3.90%	8.00%	1.30%	4.60%	2.90%	...	12.90%	

For our analysis we need only the 'Ethnic Origin Population' data available under the column 'Characteristic' and the different neighbourhoods available along with their respective populations. The neighbourhoods are provided as columns 6 to 144. Each neighbourhood contains data split on its residents with respect to different ethnicities.

The remaining of the columns in this data are redundant to our analysis and can be dropped.

Only ethnic origins with high number of residents (such as Chinese, English & other European countries, Indian) are retained to simplify our analysis and the rest are left out as their numbers are not quite significant.

### **Foursquare API**

Foursquare API offers location based experiences with diverse information about venues, users, photos and check-ins. Using Foursquare API, a free application providing geographical data, we can retrieve the different venues available in the various neighbourhoods of Toronto city.

The link to this application is: <https://foursquare.com/developers/apps>

This website is easy to use and has excellent documentation on creating developer account and how to retrieve the location data. Our interest lies in obtaining the list of restaurants and other eateries in various neighbourhoods using Foursquare API. The retrieved data can be used to filter out the type of restaurant we are looking for and then be compared against the existing ethnic population and competitors to make a calculated decision on setting up any new restaurant.

Once the neighbourhood coordinates are provided into the Foursquare API query, we can retrieve a list of venues within a particular radius of that coordinates. An example of the retrieved list of venues is:

```
{
  "meta": {
    "code": 200,
    "requestId": "5ac51d7e6a607143d811cecb"
  },
  "response": {
    "venues": [
      {
        "id": "5642aef9498e51025cf4a7a5",
        "name": "Mr. Purple",
        "location": {
          "address": "180 Orchard St",
          "crossStreet": "btwn Houston & Stanton St",
          "lat": 40.72173744277209,
          "lng": -73.98800687282996,
          "labeledLatlngs": [
            {
              "label": "display",
              "lat": 40.72173744277209,
              "lng": -73.98800687282996
            }
          ]
        },
        "distance": 8,
        "postalCode": "10002",
        "cc": "US",
        "city": "New York",
        "state": "NY",
        "country": "United States",
        "formattedAddress": [
          "180 Orchard St (btwn Houston & Stanton St)",
          "New York, NY 10002",
          "United States"
        ]
      },
      {
        "id": "4bf58dd8d48988d1d5941735",
        "name": "Hotel Bar",
        "pluralName": "Hotel Bars",
        "shortName": "Hotel Bar",
        "icon": {
          "prefix": "https://ss3.4sqi.net/img/categories_v2/travel/hotel_bar_",
          "suffix": ".png"
        },
        "primary": true
      }
    ],
    "venuePage": {
      "id": "150747252"
    }
  }
}
```

---

The above data can be filtered out and only the required type of restaurants be retained for our analysis. Using the coordinates of the neighbourhood and by looking at the existing types of restaurants and their numbers, one can come to a decision whether it's a reasonable location to start a new restaurant. This will aid in running a successful restaurant by maximizing footfall.

## Methodology

The data sources are as follows:

- From the [www.toronto.ca](http://www.toronto.ca) website we can download the demographic data of Toronto city. The data has information on population, income, education, ethnic origin and so on. All these are split against the neighbourhoods
- Latitude and Longitude Coordinates for each of the neighbourhoods are obtained from the geocoder package
- From the Foursquare API we can obtain the list of venues for each of the neighbourhoods by providing the geographical coordinates. We use this feature to get the list of restaurants in each neighbourhood for Toronto city.

### a) Data Analysis

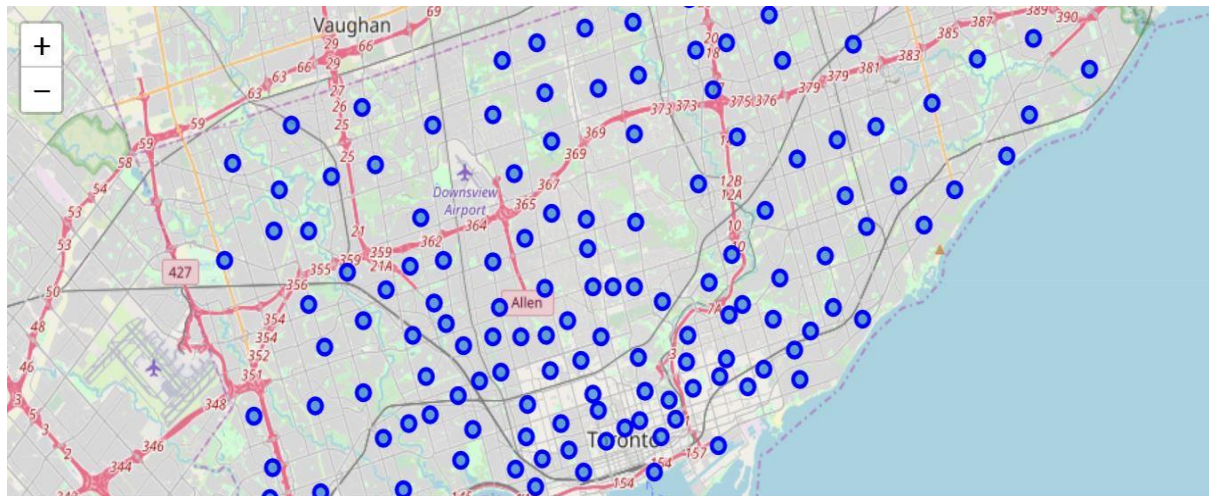
After importing the demographic data of Toronto into a data frame, we only retain the neighbourhood numbers (CDN) and neighbourhood names (Neighborhood) columns. The rest of the data is redundant and unnecessary for our analysis.

	CDN	Neighborhood
0	129	Agincourt North
1	128	Agincourt South-Malvern West
2	20	Alderwood
3	95	Annex
4	42	Banbury-Don Mills

Using the Geocoder package we obtain the latitude and longitude coordinates for each of the neighbourhoods in the above column. This along with the neighbourhood number and name is stored in a data frame.

	CDN	Neighborhood	Latitude	Longitude
0	129	Agincourt North	43.80930	-79.26707
1	128	Agincourt South-Malvern West	43.78735	-79.26941
2	20	Alderwood	43.60496	-79.54116
3	95	Annex	43.66936	-79.40280
4	42	Banbury-Don Mills	43.74041	-79.34852

Using the folium library, we create a map of Toronto city having all the neighbourhoods. The blue circles represent neighbourhoods on the map.



Our dataset covers the entire city of Toronto and there is no missing data.

The next task for us is to find out the top ethnic origins for each of the neighbourhoods and sort it in descending order of population. This is used to cluster on the demographic data

```
----Agincourt North----
      Origin      Count
0      Chinese  16950.0
1    Sri Lankan   2230.0
2    East Indian   2090.0
3    Filipino    1465.0
4    Canadian    1295.0
5    English     870.0
6    Tamil       855.0
7    Jamaican    780.0
8    Scottish    600.0
9    Irish       425.0
```

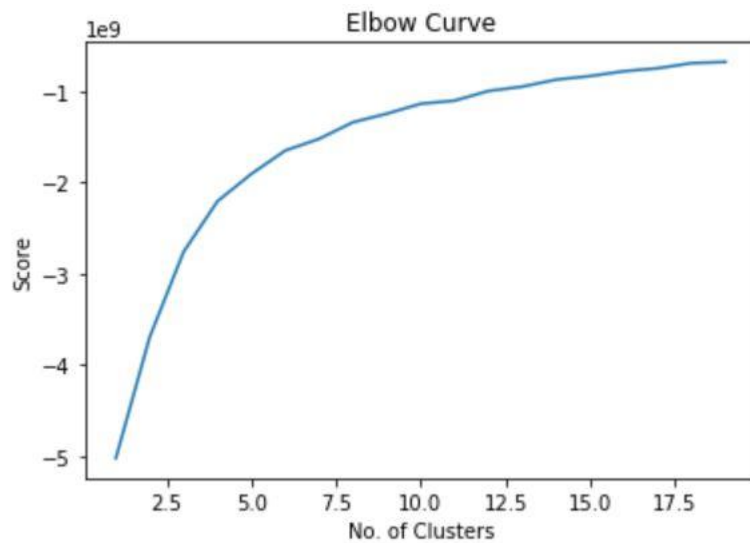
## b) K-Means Algorithm

K-Means clustering is most commonly used for data clustering analysis such as population distribution and anomaly detection. In our analysis, we aim to cluster the neighbourhoods in Toronto city and understand the population distribution.

One hot encoding is not necessary in this analysis as the demographic data is not categorical but numerical values.

**Elbow method:** In the K-Means algorithm, the number of clusters is to be fed in manually. To determine the best number of clusters the elbow method is employed. Multiple K-Means algorithms are run and the sum of squared distances of each observation to the nearest cluster centre is calculated. Higher the sum of the squared distance, farther the data points globally are from their cluster centres. If we set a high value of k (for instance k= no. of observations) each data point forms a cluster of its own and the sum of squared distance results in zero. Such a clustering is not recommended.

We plot the number of clusters against a score to find out the number of clusters that best fits our analysis.



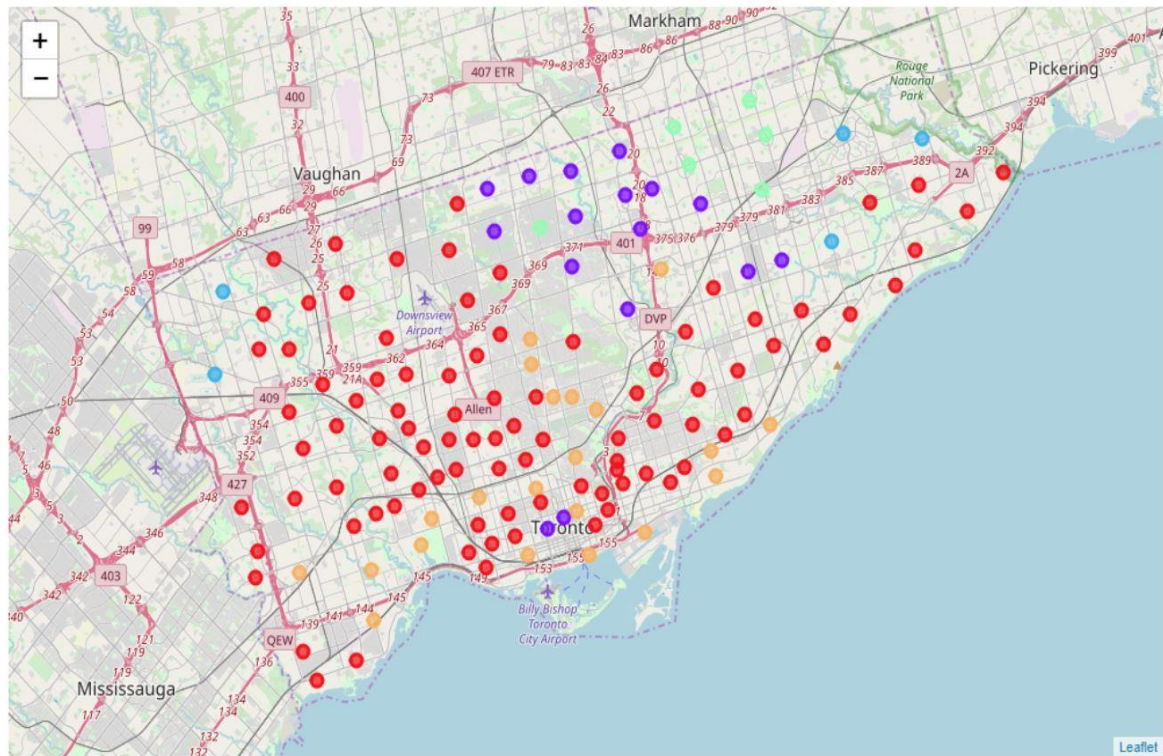
Clearly, an elbow is visible at K=5. Hence we choose 5 as the number of clusters for our K-Means algorithm.

After we run the K-Means algorithm and the clustering is done, each neighbourhood is set into one of the 5 clusters. The same can be seen in the image below, cluster numbers for each of the neighbourhoods are visible under the column Cluster\_Labels.

CDN	Neighborhood	Latitude	Longitude	Cluster_Labels	1st Most Common Ethnic Origin	2nd Most Common Ethnic Origin	3rd Most Common Ethnic Origin	4th Most Common Ethnic Origin	5th Most Common Ethnic Origin	6th Most Common Ethnic Origin	7th Most Common Ethnic Origin	8th Most Common Ethnic Origin	9th Most Common Ethnic Origin
0	129 Agincourt North	43.80930	-79.26707	3	Chinese	Sri Lankan	East Indian	Filipino	Canadian	English	Tamil	Jamaican	Scottish
1	128 Agincourt South-Malvern West	43.78735	-79.26941	3	Chinese	East Indian	Filipino	Sri Lankan	Canadian	English	Scottish	Jamaican	Italian
2	20 Alderwood	43.60496	-79.54116	0	English	Canadian	Irish	Scottish	Italian	Polish	German	French	Ukrainian
3	95 Annex	43.66936	-79.40280	4	English	Irish	Scottish	Canadian	German	French	Polish	Chinese	Italian
4	42 Banbury-Don Mills	43.74041	-79.34852	1	Chinese	English	Irish	Canadian	Scottish	East Indian	German	Filipino	Polish

The same can be visualized on a map using the folium library. Each neighbourhood is assigned to a cluster. The different colours for each neighbourhood represent the cluster it belongs to.





## Evaluation & Results

- a) **Clustering:** Let us try to understand the ethnic background of people living in each clusters by studying them. This will give us a fair idea of the kind of restaurants one can think of opening to suit the residents' palate.

### Cluster 0:

Indicated by red colour on the map, this cluster is predominantly inhabited by English & other Europeans and Canadians. They are the dominant population in Toronto city and dwell in central, southern and eastern Toronto.

Anyone interested in opening an Italian restaurant, English pub or fish & chips, or a European style restaurant can focus on these neighbourhoods.

CDN	Neighborhood	Latitude	Longitude	Cluster_Labels	1st Most Common Ethnic Origin	2nd Most Common Ethnic Origin	3rd Most Common Ethnic Origin	4th Most Common Ethnic Origin	5th Most Common Ethnic Origin	6th Most Common Ethnic Origin	7th Most Common Ethnic Origin	8th Most Common Ethnic Origin
2	20 Alderwood	43.604960	-79.541160	0	English	Canadian	Irish	Scottish	Italian	Polish	German	French
5	34 Bathurst Manor	43.763780	-79.454770	0	Filipino	Russian	Italian	Polish	Canadian	Jewish	Chinese	English
9	39 Bedford Park-Nortown	43.730660	-79.424500	0	Canadian	Polish	English	Russian	Scottish	Irish	Filipino	Jewish
10	112 Beechborough-Greenbrook	43.693110	-79.478310	0	Portuguese	Jamaican	Italian	Canadian	Filipino	East Indian	English	Spanish

## Cluster 1:

People of Chinese, Filipino and Korean backgrounds reside in these neighbourhoods. These are represented by violet colour on the map and living mostly in the northern parts of the city of Toronto

An entrepreneur who wishes to open a Chinese style, south Asian restaurant should put his money in these neighbourhoods.

CDN	Neighborhood	Latitude	Longitude	Cluster_Labels	1st Most Common Ethnic Origin	2nd Most Common Ethnic Origin	3rd Most Common Ethnic Origin	4th Most Common Ethnic Origin	5th Most Common Ethnic Origin	6th Most Common Ethnic Origin	7th Most Common Ethnic Origin	8th Most Common Ethnic Origin	9th Mos Common Ethni Origin	
4	42	Banbury-Don Mills	43.74041	-79.34852	1	Chinese	English	Irish	Canadian	Scottish	East Indian	German	Filipino	Polish
6	76	Bay Street Corridor	43.65771	-79.38618	1	Chinese	English	Canadian	Scottish	Irish	East Indian	Korean	German	French
7	52	Bayview Village	43.77710	-79.37957	1	Chinese	Iranian	Korean	English	Canadian	Irish	Scottish	East Indian	Filipino
8	49	Bayview Woods-Steeles	43.79485	-79.38222	1	Chinese	Iranian	English	Canadian	Korean	Polish	Irish	Scottish	Russian

## Cluster 2:

Indian origin people reside in these neighbourhoods which are represented by sky blue clusters on the map. These neighbourhoods are in the outskirts of the city.

Any Indian style restaurant can be started in these neighbourhoods as we can expect decent demand here.

CDN	Neighborhood	Latitude	Longitude	Cluster_Labels	1st Most Common Ethnic Origin	2nd Most Common Ethnic Origin	3rd Most Common Ethnic Origin	4th Most Common Ethnic Origin	5th Most Common Ethnic Origin	6th Most Common Ethnic Origin	7th Most Common Ethnic Origin	8th Most Common Ethnic Origin	9th Most Common Ethnic Origin	
73	132	Malvern	43.81024	-79.22034	2	East Indian	Sri Lankan	Filipino	Chinese	Jamaican	Canadian	English	Tamil	Guyane
81	2	Mount Olive-Silverstone-Jamestown	43.74721	-79.58826	2	East Indian	Iraqi	Jamaican	Canadian	Somali	Italian	Assyrian	Sri Lankan	Oth Afric. origin
105	131	Rouge	43.80766	-79.17405	2	East Indian	Sri Lankan	Canadian	Filipino	Jamaican	English	Chinese	Tamil	Scotti
124	1	West Humber-Clairville	43.71455	-79.59258	2	East Indian	Jamaican	Canadian	Filipino	Italian	Punjabi	English	Guyanese	Chine



### Cluster 3:

Chinese origin residents live here in these neighbourhoods which are in the north eastern side of Toronto. Fluorescent green clusters represent these neighbourhoods.

A Chinese or any Asian restaurant would be the best bet in these neighbourhoods.

	CDN	Neighborhood	Latitude	Longitude	Cluster_Labels	1st Most Common Ethnic Origin	2nd Most Common Ethnic Origin	3rd Most Common Ethnic Origin	4th Most Common Ethnic Origin	5th Most Common Ethnic Origin	6th Most Common Ethnic Origin	7th Most Common Ethnic Origin	8th Most Common Ethnic Origin	9th M Common Eth Ori
0	129	Agincourt North	43.80930	-79.26707	3	Chinese	Sri Lankan	East Indian	Filipino	Canadian	English	Tamil	Jamaican	Scottish
1	128	Agincourt South-Malvern West	43.78735	-79.26941	3	Chinese	East Indian	Filipino	Sri Lankan	Canadian	English	Scottish	Jamaican	Italian
66	117	L'Amoreaux	43.79726	-79.31220	3	Chinese	East Indian	Canadian	Sri Lankan	Filipino	English	Armenian	Jamaican	Scottish
76	130	Milliken	43.82280	-79.27694	3	Chinese	Sri Lankan	East Indian	Filipino	Canadian	Tamil	Jamaican	English	Vietnamese

### Cluster 4:

English and other European ethnic people dwell in these neighbourhoods. On the map these clusters can be seen in orange colour. These neighbourhoods are in the central and southern parts of Toronto city.

English, Irish and other European culinary dine-ins would be safe to open in these neighbourhoods.

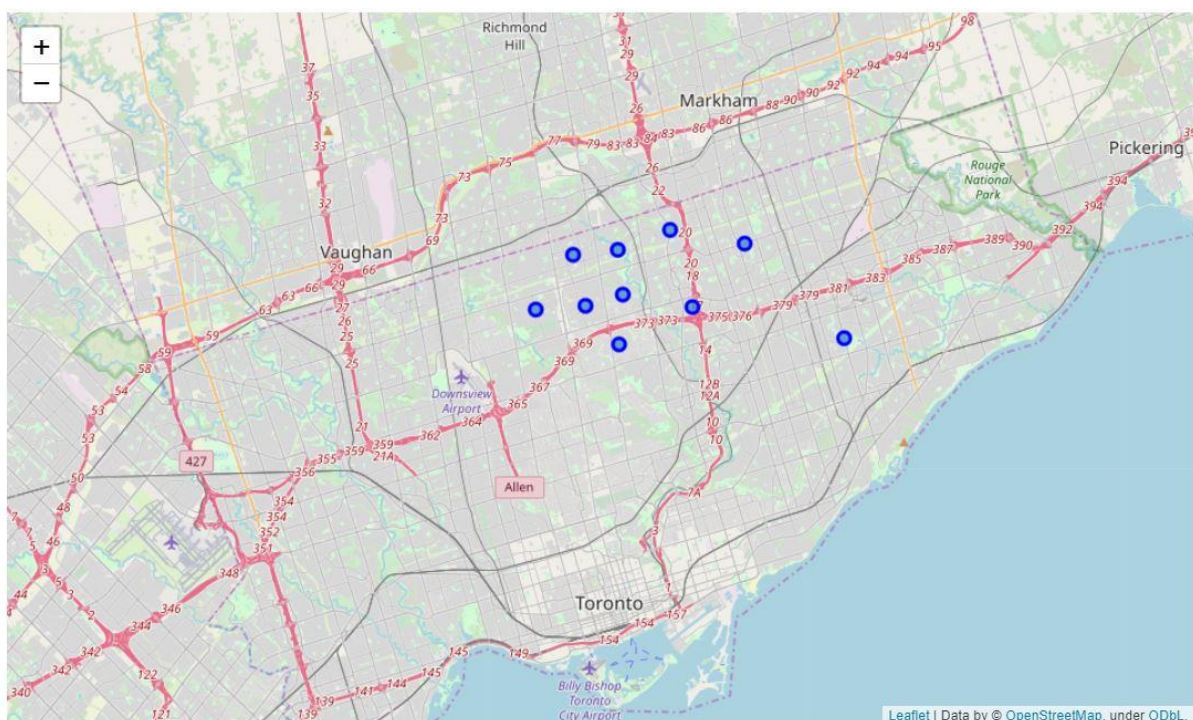
	CDN	Neighborhood	Latitude	Longitude	Cluster_Labels	1st Most Common Ethnic Origin	2nd Most Common Ethnic Origin	3rd Most Common Ethnic Origin	4th Most Common Ethnic Origin	5th Most Common Ethnic Origin	6th Most Common Ethnic Origin	7th Most Common Ethnic Origin	8th Most Common Ethnic Origin	9th Most Common Ethnic Origin	
	3	95	Annex	43.66936	-79.40280	4	English	Irish	Scottish	Canadian	German	French	Polish	Chinese	Italian
	12	122	Birchcliffe-Cliffside	43.69472	-79.26460	4	English	Irish	Canadian	Scottish	French	German	Chinese	Italian	Filipino
	23	75	Church-Yonge Corridor	43.66024	-79.37868	4	English	Irish	Scottish	Chinese	Canadian	French	German	East Indian	Italian
	32	93	Dovercourt-Wallace Emerson-Junction	43.66604	-79.43687	4	Portuguese	English	Canadian	Irish	Scottish	Chinese	Italian	German	French

- b) Competition:** Now that we have an overview of the ethnic backgrounds of people living in different neighbourhoods in Toronto city, we can now hunt for a location where the competition is minimum.

As an example, let's assume we want to open a Chinese restaurant. A safe place to open this restaurant would be at a location where there are least number of Chinese restaurants existing. To determine the existing number of Chinese restaurants at a location, we obtain the existing venues at a location using the Foursquare API, filter only the Chinese restaurants.

	CDN	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	
	4	129	43.80930	-79.26707	Congee Town 太皇名粥	43.809035	-79.267634	Chinese Restaurant
	27	129	43.80930	-79.26707	Kin Kin Bubble Tea Co	43.807852	-79.270296	Chinese Restaurant
	31	128	43.78735	-79.26941	Congee Me 小米粥鋪	43.787392	-79.268387	Chinese Restaurant
	32	128	43.78735	-79.26941	Asian Legend 味香村	43.788068	-79.266768	Chinese Restaurant
	37	128	43.78735	-79.26941	Perfect Chinese Restaurant 雅瓊海鮮酒家	43.787774	-79.270294	Chinese Restaurant

We can also count the number of Chinese restaurants present in each of the neighbourhoods and choose the one where there are least number of restaurants as a safe bet for opening a new restaurant.



## Discussion

Some of the safe places to open a Chinese style restaurant in Toronto are St.Andrew-Windfields, Willodale West and East, Bayview Woods-Steeles in that order. Our map below represents these neighbourhoods in blue circles.

Similarly we can identify the best places to open an Indian cuisine restaurant, an English or Irish pub, Italian & other European style restaurants and the list can go on.

## **Results**

With the help of geographic and demographic data and by using K-Means clustering we were able to identify the best place to open a particular type of restaurant in any given neighbourhood. This was done by determining the existing ethnic population and the prevalent competition.

Restaurant owners and food catering services can make use of such an analysis.

## **Future Scope**

We can also try to set up restaurants based on the income and expenditure habits of people residing in any given neighbourhood. For example, if people have a high income and expenditure habit, one can think of opening a fine dine restaurant rather than a budget cafe.

We can carry out similar analysis to identify locations to start hospitals, schools, shopping malls and so on. Housing and real estate prices can be determined using such algorithms.