

# BEST LOCATION TO OPEN A RESTAURANT IN TORONTO CITY



# Identifying foot fall is paramount for a restaurateur

- Ethnic backgrounds decide the eating habits and food choice of people
- Competition also drives a restaurant's business
- It is important to determine these two factors among others for a restaurant to be successful

# Data Import and Cleaning

- Demographic data of Toronto city is available on the [www.toronto.ca](http://www.toronto.ca) website
- The raw data has 2384 rows and 145 columns of which only the neighbourhood numbers & names along with population of each neighbourhood based on ethnicity is retained for analysis
- Geographical coordinates for the neighbourhoods in Toronto are obtained from using the Geocoder library
- List of venues in each of the neighbourhoods can be retrieved from Foursquare API



# Methodology

## Data Analysis-1

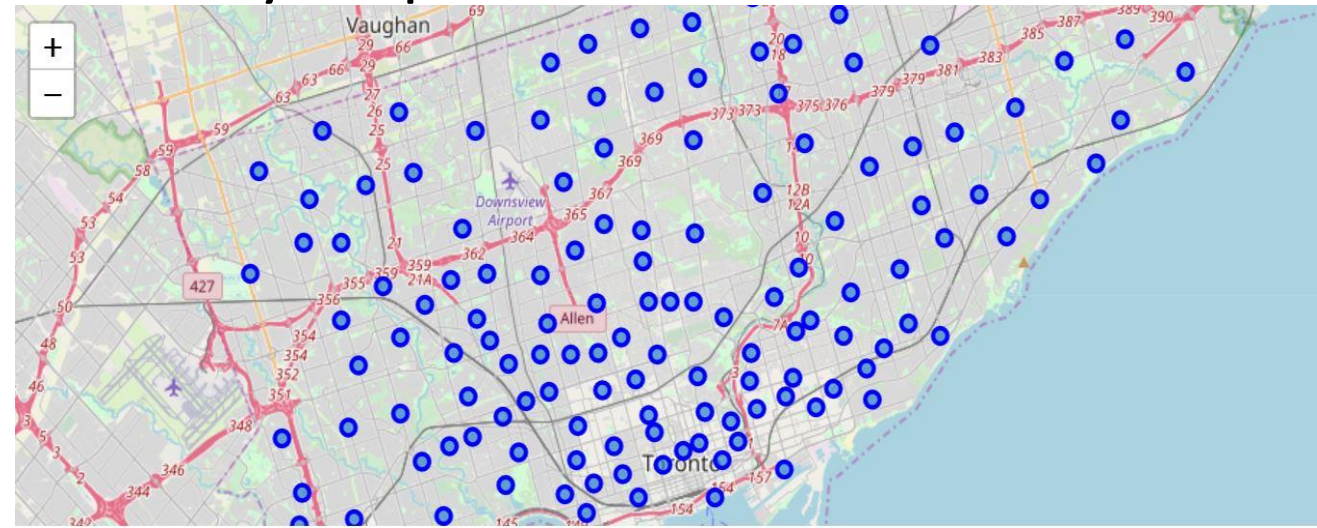
- Obtain each neighbourhoods geographical coordinates

Using geocoder we obtain the latitude and longitude of each neighbourhood

	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

- Map these neighbourhoods onto Toronto city map

We can visualize the distribution of neighbourhoods in blue colour



## Data Analysis-2

- Ethnic origin with maximum count is determined

Using count() function we can determine the number of different ethnic origins in each neighbourhood

This is useful in Kmeans clustering employed at a later stage

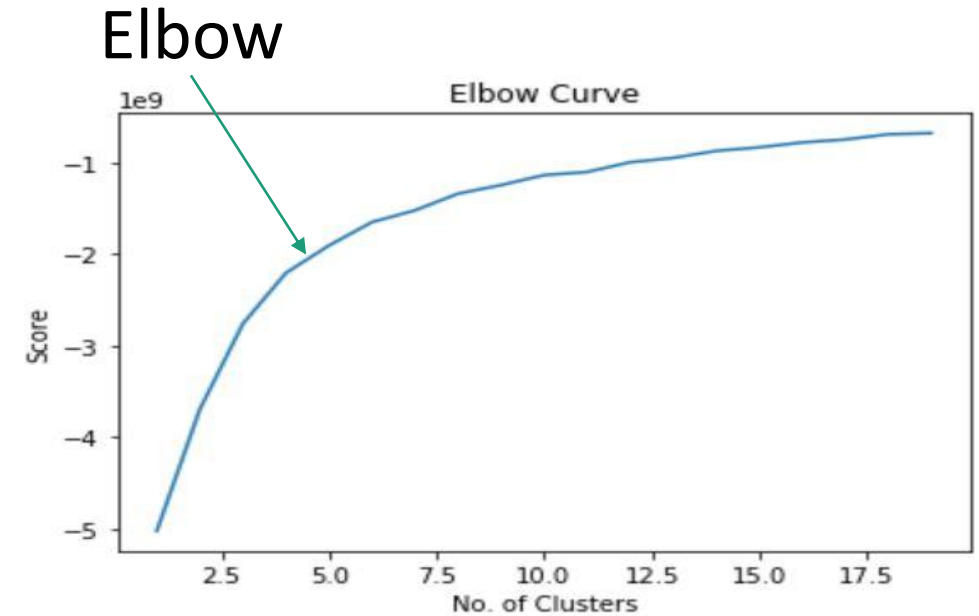
```
----Agincourt North----  
      Origin      Count  
0      Chinese  16950.0  
1    Sri Lankan   2230.0  
2    East Indian   2090.0  
3    Filipino    1465.0  
4    Canadian    1295.0  
5     English     870.0  
6       Tamil     855.0  
7    Jamaican     780.0  
8    Scottish     600.0  
9      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

# K-Means Clustering

- K-means clustering is a commonly used clustering algorithm for unsupervised learning
- Useful to find similar groups in data
- Beneficial in demographic segmentation and anomaly detection
- Using the elbow method we identify the optimal number of clusters

# K-Means Clustering contd...

- At  $k = 5$  we have the elbow
- Once we run the Kmeans cluster each neighbourhood is assigned a cluster

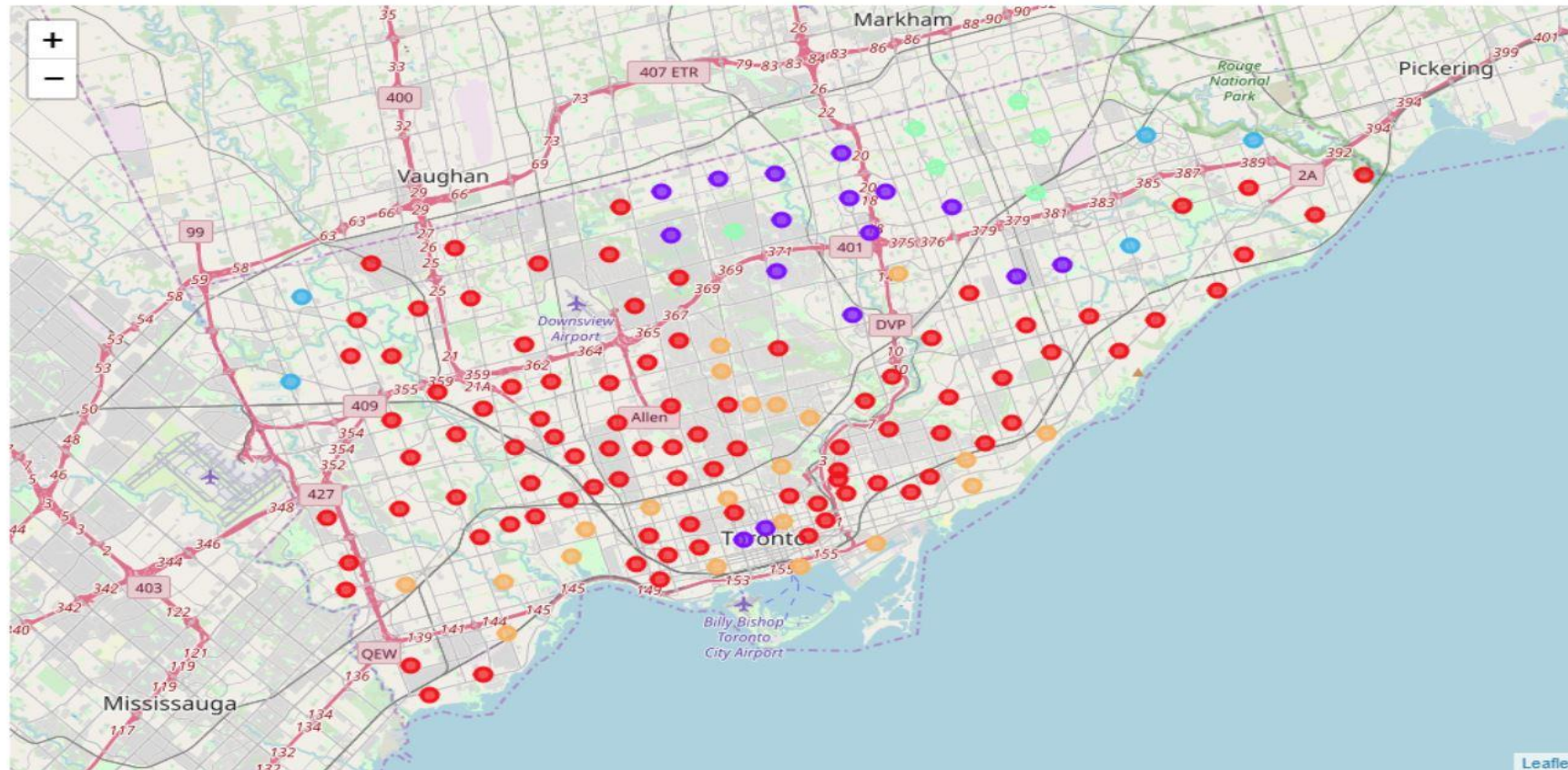


	CDN	Neighborhood	Latitude	Longitude	Cluster_Labels	1st Most Common Ethnic Origin	2nd Most Common Ethnic Origin	3rd Most Common Ethnic Origin	4th Most Common Ethnic Origin	5th Most Common Ethnic Origin	6th Most Common Ethnic Origin	7th Most Common Ethnic Origin	8th Most Common Ethnic Origin	9th Most Common Ethnic Origin
0	129	Aginccourt North	43.80930	-79.26707	3	Chinese	Sri Lankan	East Indian	Filipino	Canadian	English	Tamil	Jamaican	Scottish
1	128	Aginccourt 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



# K-Means Clustering contd...

Each cluster has a different colour on the map





# Evaluation & Discussion

- With the help of Foursquare API we can obtain different venues available in the neighbourhoods
- Then we filter on the required type of restaurants

For example, we can filter for the list of Chinese restaurants in clusters 1 and 3

	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

# Evaluation & Discussion contd...

- The number of Chinese restaurants that already exist in the neighbourhood are counted
- The neighbourhood that has the least number of Chinese restaurants is the most ideal place to open one
- Top 10 neighbourhoods to open a Chinese restaurant can be visualized on the map in blue



# Results

- Using demographic data of Toronto city we can identify the best places to open a new restaurant
- We selected ethnic background and competition as the two main criteria to determine the business for any restaurant
- Any restaurateur and catering business can be beneficial with such kind of study

# Further Scope

- Restaurants can also consider 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 in
- 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.