

# Opening a Chinese Restaurant in Toronto



**By:**  
**Abhishek Kumar Singh**  
**Coursera Capstone Assignment**

## Table of Contents

1. Introduction .....	4
2. Aim .....	4
3. Data Used.....	5
4. Methodology.....	6
5. Results .....	9
6. Conclusion .....	10

<b>Figure 1 : Toronto Neighborhood details obtained via scraping Wikipedia .....</b>	<b>6</b>
<b>Figure 2 : Clubbing the Neighborhoods on the basis of similar boroughs. ....</b>	<b>6</b>
<b>Figure 3 : Geospatial coordinates of the neighborhoods.....</b>	<b>7</b>
<b>Figure 4 : Map of the neighbprhoods using FOLIUM .....</b>	<b>7</b>
<b>Figure 5 : Estimate of Chinese restaurants in the vicinity of the neighborhood.....</b>	<b>9</b>
<b>Figure 6 : Neighborhoods after clustering .....</b>	<b>10</b>

## 1. Introduction

In the view of the requirements of this course project, the idea of opening a Chinese restaurant has been taken up. It's not that the neighborhoods of Toronto have a dearth of multi-cuisine restaurants but with the help of machine learning techniques and foursquare api, an identification of a feasible opportunity to open a Chinese restaurant will be done.

From a business person's perspective, this task needs to be taken up from very basic. So, we will gather information about multi-cuisine restaurants in Toronto's neighborhood. Then, based on the algorithms like clustering, an appropriate decision can be made.

## 2. Aim

To find the best possible location to open a Chinese restaurant in Toronto's neighborhood.

### 3. Data Used

For the purpose of opening the Chinese restaurant, we need to gather following data:

- The neighborhood of Toronto, we have earlier used this in our assignments.
- Geographical info about the different neighborhoods.
- Information about restaurants, cafes of all cuisine especially Asian, because they are going to be the ones to compete.

Following sources will be used to extract the data:

- Wikipedia
- Geocoder package
- Foursquare API to get the details of neighborhood so that proper clustering can be done.

## 4. Methodology

First we start off by scraping the data from Wikipedia for different regions around Toronto. Once, we processed the data into a data frame we check in for null values in our data and club the regions as per **boroughs** and **neighborhoods**.

	Postal Code	Borough	Neighborhood
0	M1A	NaN	NaN
1	M2A	NaN	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

Figure 1 : Toronto Neighborhood details obtained via scraping Wikipedia

Once, we get the data from the web we start to preprocess it and check it for null values in the columns of **boroughs** and **neighborhood**. After this preprocessing we club the **neighborhoods** together based on **boroughs**. The processed data appears as follows.

	Postal Code	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Figure 2 : Clubbing the Neighborhoods on the basis of similar boroughs.

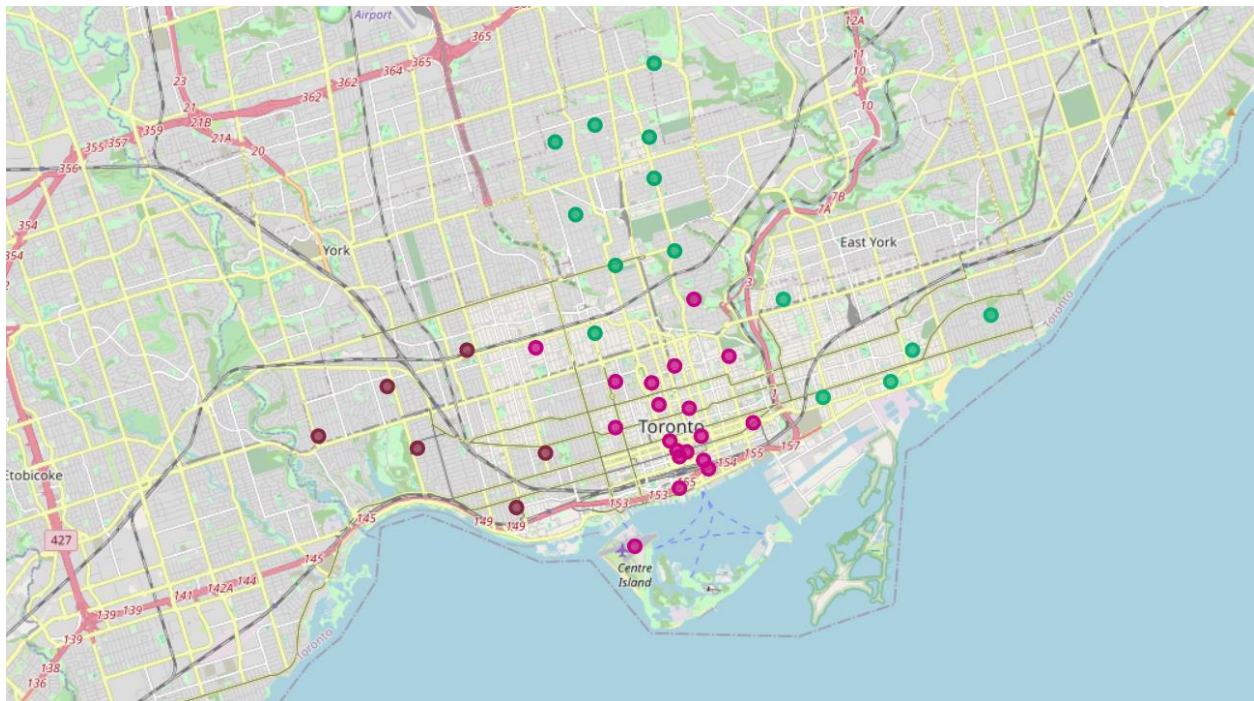


Our next step starts off as reading the geospatial coordinates of the location in Toronto. For the sake of informing, these coordinates were made available from Week 3 Assignment itself. This gives us the longitudes and latitudes of different neighborhoods and accordingly help us in clustering the neighborhoods based on venue types present in their proximity.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

**Figure 3 : Geospatial coordinates of the neighborhoods.**

Next, we map the different locations and color code them so that likewise locations are placed together. We have used Folium for the purpose of creating maps.



**Figure 4 : Map of the neighborhoods using FOLIUM**

Using the Foursquare APIs we then move on to calculate the details of all neighborhoods. For this purpose, we have set the limit to 100 results and radius of exploration as 500 meters.

We then make use of **K-means clustering** to group the neighbors together and identify feasible opportunity for a restaurant.



## 5. Results

After obtaining the venue details using foursquare api, we make use of one hot encoding to identify the features present in different neighborhoods. This resembles as follows.

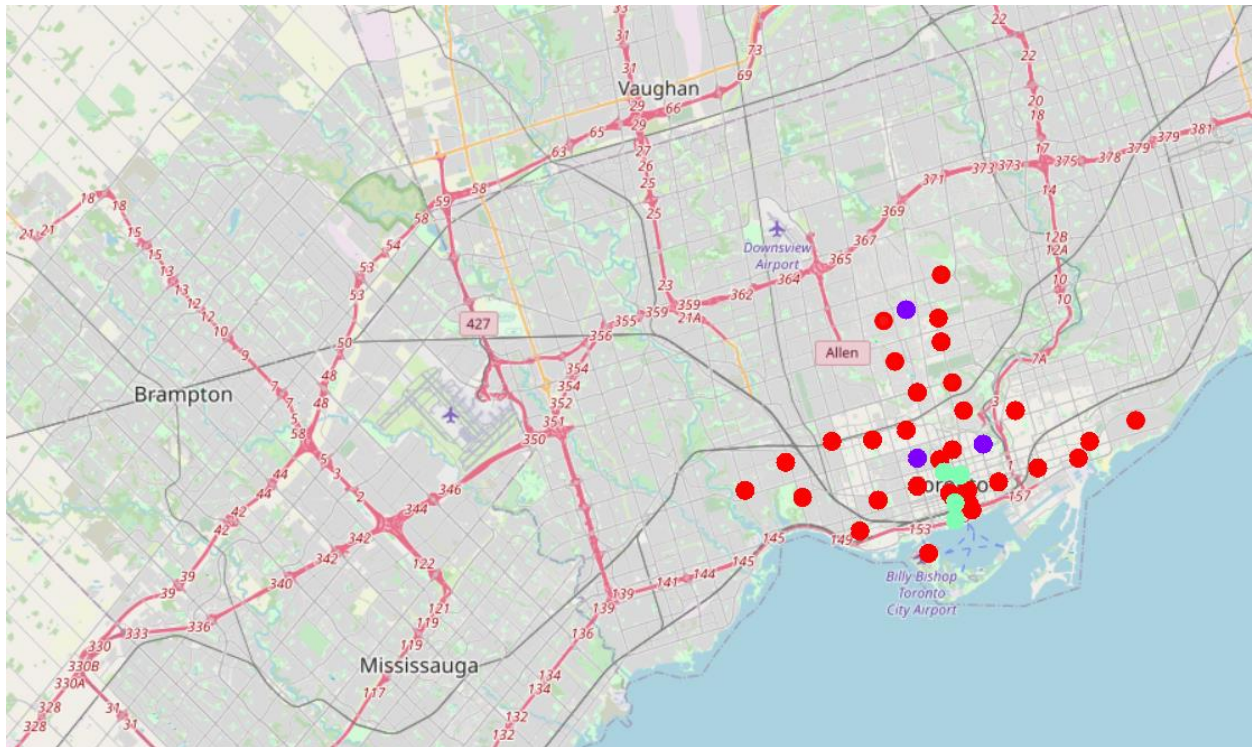
	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant
0	The Beaches	0	0	0	0	0	0	0
1	The Beaches	0	0	0	0	0	0	0
2	The Beaches	0	0	0	0	0	0	0
3	The Beaches	0	0	0	0	0	0	0
4	The Danforth West, Riverdale	0	0	0	0	0	0	0

We had the dataset that had 237 features as columns and the number of neighborhoods were 1711.

We then go on to cluster our neighborhoods into 3 clusters using K-means clustering algorithm. We identified the neighborhoods with Chinese restaurant in the vicinity.

	Neighborhoods	Chinese Restaurant
0	Berczy Park	0.000000
1	Brockton, Parkdale Village, Exhibition Place	0.000000
2	Business reply mail Processing Centre, South C...	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
4	Central Bay Street	0.015625

Figure 5 : Estimate of Chinese restaurants in the vicinity of the neighborhood



**Figure 6 : Neighborhoods after clustering**

## 6. Conclusion

- Most of the Chinese restaurants is in cluster 2 followed by cluster 3.
- So, the business can start off by competing with Chinese restaurants in cluster 3 and then go on exploring in the cluster 1.
- Another strategy would be starting off the restaurant in cluster 1 and create a new market for itself.