Coursera Capstone Project - Opening a chinese restaurant

## **Introduction**

### Background

Toronto is recognized as the most diverse city in the world. About 51% of the residents belong to the visible minority, out of which 12.5% are chinese, this makes it a good place to start a chinese restaurant.

To make it successful, we need to find the best neighborhood with less competition and more demand of chinese food.

### Problem

Data that might contribute to determining the best neighborhood for opening a chinese restaurant might include venues in each geographical location, corresponding to a unique postal code. This project aims to predict the best location for opening a chinese restaurant based on this data.

### Interest

Obviously those people/ businessmen would be interested in this project who would like to open a chinese restaurant or buy and establish any chinese restaurant franchise. Others who would be interested could be people who would like to find a good place to eat chinese food.

## **Data acquisition and cleaning**

### Data Sources

### Wikipedia

List of postal codes, boroughs and neighborhoods of Toronto from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

This contains the list of Toronto, its boroughs and neighborhoods which will be identified with postal codes. I will clean and extract only those neighborhoods which are in Toronto borough.

### Geospatial data

I will use a csv file from '<https://cocl.us/Geospatial_data>' to get latitude and longitude coordinates of each neighborhood. This will help me in finding the exact location of venues.

### Foursquare API

I will use the Foursquare API in listing the most popular venues of each location and in determining the location where a chinese restaurant has more demand and less competition.

### Data cleaning

First, data was scraped from wikipedia and was stored into a data frame. The parsed data was formatted into three columns of postal code, borough and neighborhood. There were some missing values in the borough, such rows were dropped, while those rows where neighborhood values were missing , they were given the same value as borough.

Data for latitude and longitude coordinates was also downloaded and combined into one table with postal code, borough and neighborhood columns.

### Feature selection

After data cleaning, I selected only the data related to Toronto, as I was interested in finding the suitable location in the borough of Toronto. So I defined a new dataframe with only those rows that contain the word ‘Toronto’ in their borough. I came up with a 38 rows data frame with boroughs ‘Downtown Toronto’, ‘East Toronto’, ‘West Toronto’ and ‘Central Toronto’.

## **Methodology**

### Foursquare API

I defined the Foursquare API credentials and then got the venues of each neighborhood in Toronto at 500 km radius.

### One hot encoding

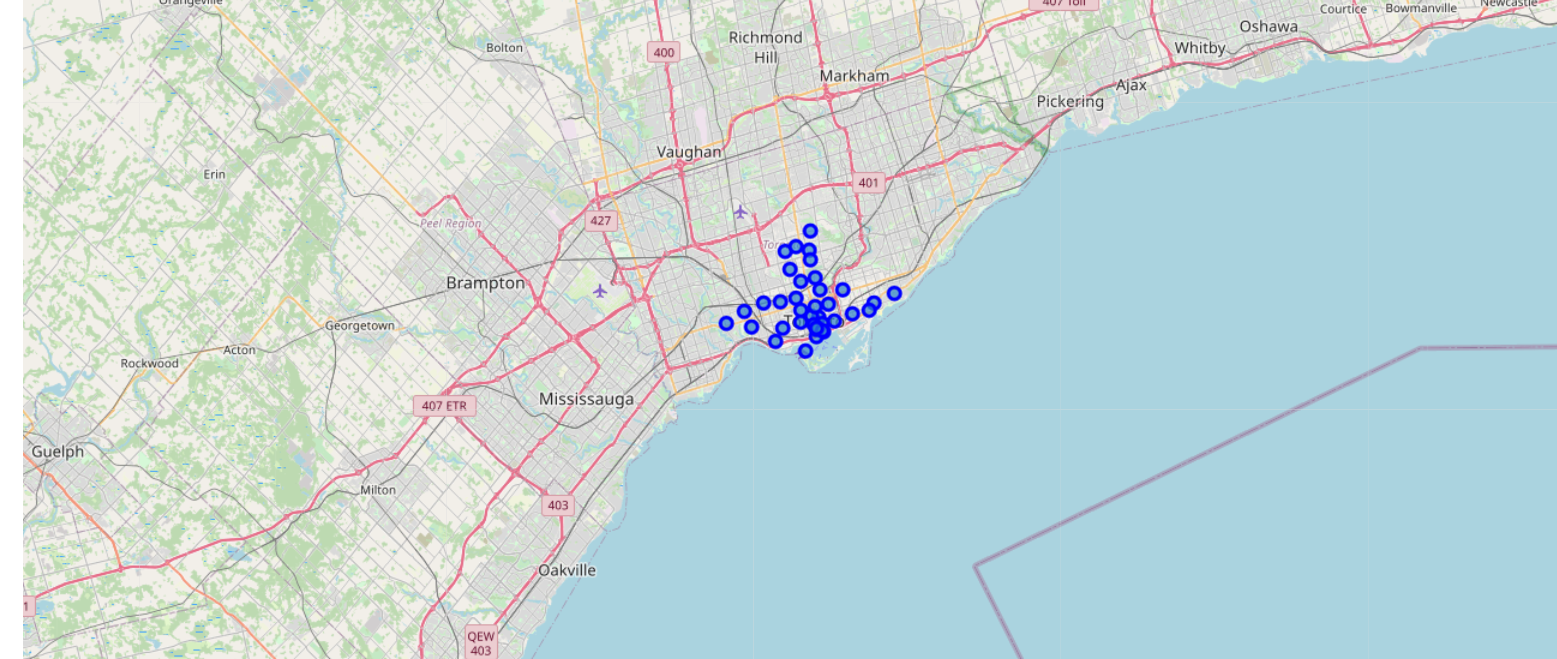
I also used one hot encoding to represent the categorical data as it is more expressive as machine learning algorithms work with binary values.

After grouping the data by neighborhood, I created another data frame with only columns of neighborhood and Chinese restaurant.

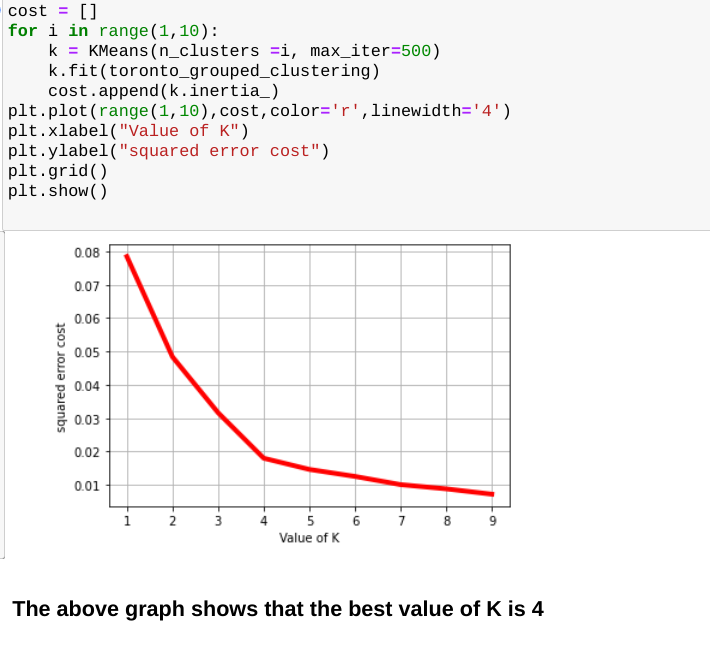
I merged that data frame with the Toronto dataframe with only rows that have chinese restaurant.

### Exploratory Data Analysis

I, first, visualized the above data on map using folium, all the neighborhoods having chinese restaurants, since combining geospatial information with data gives a good context for insight.

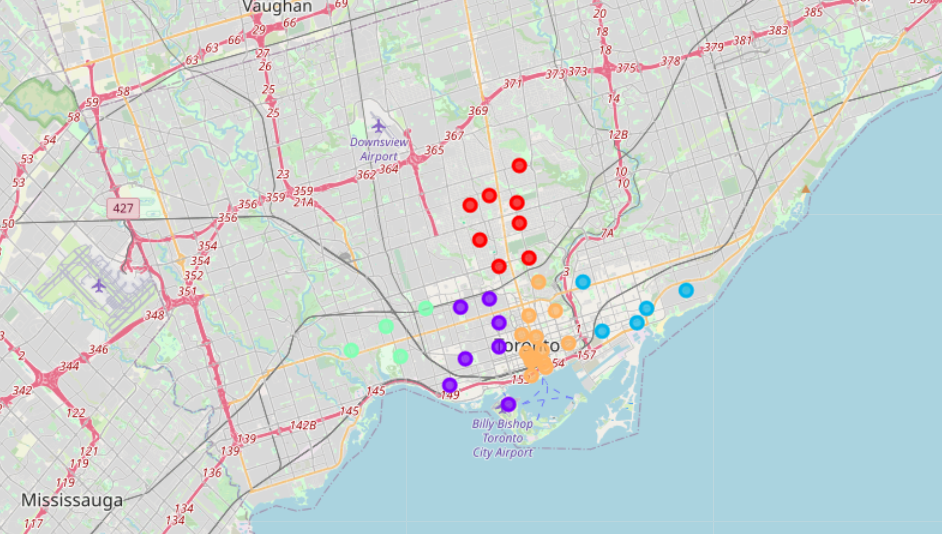


Next, I used k-means clustering to group the neighborhood. In doing so, I used the elbow method with mean squared error to get the optimal value of k.



I used this value in clustering the neighborhood of Toronto with Chinese restaurants.

I used folium again to visualize these clusters on a map.



## **Result**

## I examined each cluster one by one:

## **Discussion**

I could easily derive the following results by examining the clusters:

* Cluster 1,2 and 3 do not have any chinese restaurants which means that people in that area do not like the taste of it.
* North Toronto West has the highest number of chinese restaurants, it could be due to people having the taste of chinese food or more chinese people living in that area.
* Cluster 4 has the most number of chinese restaurants in the area, which could be due to high demand.
* Chinese restaurant is the least popular food in clusters 1, 2 and 3
* Toronto Dominion,Harbourfront, Garden District have average but the same number of chinese restaurants around.

## **Conclusion**

I would conclude here by saying that it would be best to open the chinese restaurant in Downtown Toronto M5J (Harbourfront East, Union Station, Toronto Island) as many people would be visiting this area and also because there would be less competition, which would increase the chances of making it successful.