Data Science Capstone Project – Final Report

Identifying unique clusters of neighbourhoods based upon density of restaurants in Greater Toronto Area (GTA)

Simon Park

May 2019

Contents

Introduction

Toronto is one of the most multicultural cities in the world. According to the most recent census (2016) the city's population is made up of 51 per cent of residents born outside of Canada, and this doesn't even take into consideration the presence of second generation immigrants. Toronto is believed to be home to 230 different nationalities.

Within Toronto over 180 languages and dialects are spoken and according to the 2006 consensus, about 47% of immigrants still practice their mother tongue.

As is common in any large city the immigrant populations have tendencies to reside in neighbourhoods which provide proximity to residents of a similar cultural background. As a result, each Toronto neighbourhood embodies a distinct culture.

Toronto features dozens of recognizable international neighborhoods including Chinatown, Greektown, Roncesvalles Village (Little Poland), Little Italy, Koriatown and Little India.

Toronto's population statics based upon the previous 3 censuses (2006, 2011, 2016) are:

Visible minority and Aboriginal population							
Population group		**Population (2016)**	**% of total population (2016)**	**Population (2011)**	**% of total population (2011)**	**Population (2006)**	**% of total population (2006)**
European (White)		1,282,750	47.7%	1,292,365	50.2%	1,300,330	52.5%
Visible minority group	South Asian	338,965	12.6%	317,100	12.3%	298,370	12%
	Chinese	299,460	11.1%	278,390	10.8%	283,075	11.4%
	Black	239,850	8.9%	218,160	8.5%	208,555	8.4%
	Filipino	152,715	5.7%	132,445	5.1%	102,555	4.1%
	Latin American	77,160	2.9%	71,205	2.8%	64,855	2.6%
	Arab	36,030	1.3%	28,920	1.1%	22,485	0.9%
	Southeast Asian	41,645	1.5%	46,825	1.8%	37,495	1.5%
	West Asian	60,325	2.2%	50,235	2%	42,755	1.7%
	Korean	41,640	1.5%	37,225	1.4%	34,220	1.4%
	Japanese	13,410	0.5%	12,315	0.5%	11,965	0.5%
	Visible minority, n.i.e.	36,975	1.4%	33,670	1.3%	25,195	1%
	Multiple visible minorities	47,675	1.8%	37,920	1.5%	31,100	1.3%
Total visible minority population		1,385,850	51.5%	1,264,395	49.1%	1,162,630	46.9%
Aboriginal group	First Nations	14,380	0.5%	12,990	0.5%	9,130	0.4%
	Métis	7,270	0.3%	4,875	0.2%	3,650	0.1%
	Inuit	275	0%	305	0%	195	0%
	Aboriginal, n.i.e.	645	0%	920	0%	485	0%
	Multiple Aboriginal identities	500	0%	180	0%	145	0%
Total Aboriginal population		23,065	0.9%	19,265	0.7%	13,605	0.5%
Total population		2,691,665	100%	2,576,025	100%	2,476,565	100%

The top 30 languages spoken in Toronto (2016):

Language	Population	%
English	1,375,900	50.9
Cantonese	114,670	4.2
Mandarin	111,405	4.1
Tagalog (Filipino)	83,230	3.1
Spanish	72,850	2.7
Italian	62,640	2.3
Portuguese	59,355	2.2
Tamil	57,535	2.1
Farsi	49,185	1.8
Urdu	37,420	1.4
Russian	36,145	1.3
French	35,440	1.3
Korean	33,665	1.2
Arabic	29,825	1.1
Bengali	28,460	1.1
Greek	27,840	1.0
Gujarati	26,400	1.0
Polish	25,060	0.9
Vietnamese	24,775	0.9
Panjabi (Punjabi)	19,965	0.7
Ukrainian	15,465	0.6
Hindi	15,230	0.6
German	14,515	0.6
Serbian	13,380	0.5
Romanian	12,335	0.5
Hungarian	11,885	0.5
Somali	11,375	0.4
Turkish	8,855	0.3
Albanian	8,495	0.3
Armenian	7,845	0.3

The statistics mentioned up to this point have related only to geography officially designated as Toronto. However, the city has sprawled far beyond its boundaries in recent decades. When speaking of Toronto, we really need to include adjoining communities such as Mississauga, Brampton, Vaughan, Richmond Hill, and Markham, because the city now includes these communities, with residents working and living in any of these communities. To a visitor not concerned with municipal boundaries, the entire region would be simply viewed as Toronto. This collective group of Toronto and is neighboring communities is often referred to as the Greater Toronto Area (GTA).

Historically, when immigrant populations settled in Toronto over the past century they tended to settle in the downtown core. Due to crowding in Toronto proper, and increasing real estate prices, communities began to pop up in the west, north and east, and later generations moved to these newer areas of Toronto and surrounding communities.

The result of this relocation of second and third generation immigrants, as well as more recent immigration patterns, is that individual ethnic communities may now have more than one distinct population grouping within the GTA.

When new communities develop, businesses must transition to serve these populations. By studying the locations of services catering to specific cultural populations, in particular types of restaurants, we may determine that distinct clusters of restaurants exist throughout the GTA. The locations and clusters of restaurants serving specific ethnic food gives us insight into who might be living in these neighbourhoods. For example, this clustering may indicate that Toronto does not only have the historic “Chinatown” district in the downtown core – but does in fact have multiple additional “Chinatowns” throughout the GTA. This same clustering of services may indicate the existence of various population groups possessing multiple pockets.

The purpose of this analysis is to provide an understanding of the composition of the neighbourhoods in Toronto, in order to give urban planners, business owners, and even municipal leaders valuable insight into the successful planning and delivery of future services.

Data Collection

The goal of this analysis project is to attempt to identify neighbourhoods in Toronto and the immediate vicinity which contain a significant density of restaurants of a style and therefore provides clues to the cultural makeup of the community.

Define Neighbourhoods

Our first step in this process will be to identify the unique neighbourhoods to be included in our analysis. Canadian addresses are identified by a postal code of the format:

'LNL NLN'

Where L is a letter in the range A to Z

N is a number in the range 0 to 9

We will use the first 3 positions of postal codes (the LNL) to uniquely name and label the neighbourhoods.

All postal codes within the official municipal boundaries of Toronto begin with the letter 'M'. The web page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M will provide the raw neighbourhood data we require for the City of Toronto portion of our analysis. We will need to manipulate this data into the desired format using the BeautifulSoup Python package.

For the purpose of our analysis we have chosen to expand our areas of interest beyond the municipal boundaries of Toronto. The adjoining communities of Mississauga, Brampton, Vaughan, Richmond Hill, and Markham will also be included. The postal codes for these communities do not begin with the letter 'M' and therefore the information for these communities will not be supplied the https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M pages. These additional postal codes will be identified by manually analyzing the postal codes details of these communities. This process will need to be performed manually because it will require judgement calls on which geographic areas are sufficiently close to Toronto to be appropriate candidates for inclusion. Also, automatic identification of these communities is problematic because they often have been assigned historic names. For example, rather than using the name 'Vaughan' these postal districts have old village names such as 'Concord', 'Maple', 'Thornhill', and 'Woodbridge'. Once identified these additional neighbourhoods will be manually added to the list automated from the https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M page.

Latitude and Longitude of Neighbourhoods

For each neighbourhood previously identified we will require the geographic location coordinates. For the postal codes within the official Toronto boundaries we can obtain a list of their latitude and longitude values from the page http://cocl.us/Geospatial_data. For the additional communities we he chosen to manually add (Mississauga, Brampton, Vaughan, Richmond Hill, and Markham) we will need to manually determine their latitude and longitude and add them to the Toronto list. These additional coordinates can be obtained by performing lookups using a variety of websites including https://www.latlong.net/ .

Geographic location of Toronto

In order to undertake our analysis, we will be making extensive use of mapping and location data. Since our maps will be centered on Toronto we will require the latitude and longitude values of the city. This can be obtain using the Nominatim module we will import using the 'from geopy.geocoders import Nominatim' command.

Venue Data

We will obtain lists of restaurants and their locations by utilizing the Foursquare API. The API will locate restaurant venues of a specified restaurant style which are geographically close to each neighbourhood in our study.

The API call will be performed with the following command:

```
'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&query={}&radius={
}&limit={}'.format(
CLIENT_ID,
CLIENT_SECRET,
VERSION,
lat,
lng,
search_query,
radius,
LIMIT)
```

Where lat = latitude of specific postal neighbourhood
lng = longitude of specific postal neighbourhood
search_query = search string we are attempting to identify.
Examples include 'Chinese Restaurant' or 'Italian Restaurant
radius = venue results will be limited to this many meters from the specified location
LIMIT = the maximum number of venues to be returned in the results

Methodology

The methodology utilized in this data analysis is to identify each unique neighbourhood in the city of Toronto plus additional neighbourhoods bordering Toronto. For each of these neighbourhoods we will determine the names and locations of all restaurants of a specific ethnic type.

According to the most recent census data Chinese are the largest distinct group of immigrants to Toronto and Cantonese and Mandarin are the most common spoken language in the city besides English. Based upon this information the data analysis in this project will be perform utilizing Chinese restaurants. However, the approach and logic could easily be modified to any other type of restaurant.

Any particular style of restaurant (Chinese, Italian, Korean, etc.) will likely occur in virtually every area of Toronto. Since our goal is to identify only restaurants which have a high number of occurrences (density) in a specific neighbourhood we will be counting the number of occurrences of the restaurant style in each neighbourhood. Any neighbourhood which does not contain this density of this restaurant type will be excluded from the analysis – and therefore the restaurants in the neighbourhood will be excluded.

The best way for the recipients of this report to understand the results are by utilizing maps of the GTA which display the key information.

In order to convey this information, the following maps will be generated:

1) Map showing each neighbourhood included in the analysis
2) Map showing the restaurants which were determined to be part of a dense grouping of similar type of restaurants in the vicinity
3) Map showing the clustering of the restaurant as determined by k-means algorithm

Assumptions:

a) A restaurant is considered to be included in a neighbourhood if it's geographic location (latitude and longitude) is within a defined threshold of the center of the neighbourhood. Generally speaking the physical size of neighbourhoods within the city of Toronto are smaller than the size of neighbourhoods north of the city. This is a result of higher population densities within Toronto. In order to best accommodate these size differences, the threshold for inclusion in the neighbourhood has been defined as 1750 meters for Toronto and 2500 meters for north of the city.
b) The minimum density of a restaurant type (Chinese, Italian, Korean, etc.) has been chosen as 25. If a neighbourhood does not contain at least 25 restaurants of the type, then the neighbourhood will be excluded
c) The additional communities of Mississauga, Brampton, Vaughan, Richmond Hill, and Markham have been included in the analysis

Define Neighbourhoods

Neighbourhoods will be identified by the first 3 characters of their postal code.

The postal codes within the boundaries of Toronto will be obtain from the web page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M . This data will be decomposed using the BeautifulSoup Python package.

Since we have expanded our geographic area for analysis to also include Mississauga, Brampton, Vaughan, Richmond Hill, and Markham we need to determine the postal codes of these communities also. This will be achieved by manually listing the appropriate additional postal codes.

Data cleaning will be performed to drop any postal code which does not have a borough. If a value of neighbourhood name does not exist then a value will also be set.

If a postal code contains more than one neighbourhood then the names will be concatenated together.

The two lists will be then be combined into a single postal code list.

These steps will be accomplished by the following Python code:

Get Wikipedia page which contains postal codes for Toronto

```
page = requests.get('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
)
```

Create a BeautifulSoup object which will be used for parsing postal codes from web page

```
soup = BeautifulSoup(page.text, 'html.parser')
```

loop through page for required tags

```
table_Toronto_zipinfo = soup.find('table')
colvals = table_Toronto_zipinfo.find_all('td')
elem_cnt = len(colvals)
postcode = []
borough = []
neighborhood = []

for i in range(0, elem_cnt, 3):
    postcode.append(colvals[i].text.strip())
    borough.append(colvals[i+1].text.strip())
    neighborhood.append(colvals[i+2].text.strip())
```

clean the data by dropping rows where 'Borough' is not assigned

```
df_Toronto_postcodes.drop(df_Toronto_postcodes[df_Toronto_postcodes['Borough'] == 'No
t assigned'].index, inplace=True)
```

if Borough exist but a Neighborhood does not exist then set the Neighborhood to have the same value as the Borough

```
df_Toronto_postcodes.loc[df_Toronto_postcodes.Neighborhood == 'Not assigned', 'Neighborhood'] = df_Toronto_postcodes.Borough
```

group data where more than one Neighborhood has the same Postal Code

```
df_group_Toronto = df_Toronto_postcodes.groupby(['Postcode', 'Borough'])['Neighborhood'].apply(', '.join).reset_index()
```

Add postal codes and neighbourhood descriptions for additional communities in GTA (outside Toronto municipal boundaries)

```
other_gta_postcodes=[
{'Postcode': 'L5A', 'Borough': 'Mississauga', 'Neighborhood': 'Mississauga Valley / East Cooksville'},
{'Postcode': 'L5B', 'Borough': 'Mississauga', 'Neighborhood': 'West Cooksville /Fairview / City Centre / East Creditview'},
{'Postcode': 'L5C', 'Borough': 'Mississauga', 'Neighborhood': 'West Creditview /Mavis / Erindale'},
{'Postcode': 'L5E', 'Borough': 'Mississauga', 'Neighborhood': 'Central Lakeview'},
{'Postcode': 'L5G', 'Borough': 'Mississauga', 'Neighborhood': 'SW Lakeview /Mineola / East Port Credit'},
{'Postcode': 'L5H', 'Borough': 'Mississauga', 'Neighborhood': 'West Port Credit /Lorne Park / EastSheridan'},
{'Postcode': 'L5J', 'Borough': 'Mississauga', 'Neighborhood': 'Clarkson /Southdown'},
{'Postcode': 'L5K', 'Borough': 'Mississauga', 'Neighborhood': 'West Sheridan'},
{'Postcode': 'L5L', 'Borough': 'Mississauga', 'Neighborhood': 'Erin Mills / Western Business Park'},
{'Postcode': 'L5M', 'Borough': 'Mississauga', 'Neighborhood': 'Churchill Meadows /Central Erin Mills / South Streetsville'},
{'Postcode': 'L5N', 'Borough': 'Mississauga', 'Neighborhood': 'Lisgar / Meadowvale'},
{'Postcode': 'L5P', 'Borough': 'Mississauga', 'Neighborhood': 'Mississauga'},
{'Postcode': 'L5R', 'Borough': 'Mississauga', 'Neighborhood': 'West Hurontario / SWGateway'},
{'Postcode': 'L5S', 'Borough': 'Mississauga', 'Neighborhood': 'Cardiff / NE Gateway'},
{'Postcode': 'L4T', 'Borough': 'Mississauga', 'Neighborhood': 'Malton'},
{'Postcode': 'L5T', 'Borough': 'Mississauga', 'Neighborhood': 'Courtney Park / EastGateway'},
{'Postcode': 'L4V', 'Borough': 'Mississauga', 'Neighborhood': 'Wildwood'},
{'Postcode': 'L5V', 'Borough': 'Mississauga', 'Neighborhood': 'East Credit'},
{'Postcode': 'L4W', 'Borough': 'Mississauga', 'Neighborhood': 'Matheson / EastRathwood'},
{'Postcode': 'L5W', 'Borough': 'Mississauga', 'Neighborhood': 'Meadowvale Village / West Gateway'},
{'Postcode': 'L4X', 'Borough': 'Mississauga', 'Neighborhood': 'East Applewood / East Dixie / NE Lakeview'},
{'Postcode': 'L4Y', 'Borough': 'Mississauga', 'Neighborhood': 'West Applewood / West Dixie / NW Lakeview'},
{'Postcode': 'L4Z', 'Borough': 'Mississauga', 'Neighborhood': 'West Rathwood / EastHurontario / SEGateway /Sandalwood'},
```

```
{'Postcode': 'L7A', 'Borough': 'Brampton', 'Neighborhood': 'Brampton West'},
{'Postcode': 'L6S', 'Borough': 'Brampton', 'Neighborhood': 'Brampton North Central'},
{'Postcode': 'L6V', 'Borough': 'Brampton', 'Neighborhood': 'Brampton Central'},
{'Postcode': 'L6W', 'Borough': 'Brampton', 'Neighborhood': 'Brampton Southeast'},
{'Postcode': 'L6X', 'Borough': 'Brampton', 'Neighborhood': 'Brampton Southwest'},
{'Postcode': 'L6Y', 'Borough': 'Brampton', 'Neighborhood': 'Brampton South'},
{'Postcode': 'L6Z', 'Borough': 'Brampton', 'Neighborhood': 'Brampton North Central'},
{'Postcode': 'L4B', 'Borough': 'Richmond Hill', 'Neighborhood': 'Richmond Hill Southe
ast'},
{'Postcode': 'L4C', 'Borough': 'Richmond Hill', 'Neighborhood': 'Richmond Hill Southe
ast'},
{'Postcode': 'L4E', 'Borough': 'Richmond Hill', 'Neighborhood': 'Richmond Hill North/
Oak Ridges / Lake Wilcox /Temperanceville'},
{'Postcode': 'L4S', 'Borough': 'Richmond Hill', 'Neighborhood': 'Richmond Hill Centra
l'},
{'Postcode': 'L6B', 'Borough': 'Markham', 'Neighborhood': 'Markham East'},
{'Postcode': 'L6C', 'Borough': 'Markham', 'Neighborhood': 'Markham Northwest'},
{'Postcode': 'L6E', 'Borough': 'Markham', 'Neighborhood': 'Markham Northeast'},
{'Postcode': 'L6G', 'Borough': 'Markham', 'Neighborhood': 'Markham Northeast'},
{'Postcode': 'L3P', 'Borough': 'Markham', 'Neighborhood': 'Markham Central'},
{'Postcode': 'L3R', 'Borough': 'Markham', 'Neighborhood': 'Markham Outer Southwest'},
{'Postcode': 'L3S', 'Borough': 'Markham', 'Neighborhood': 'Markham Southeast'},
{'Postcode': 'L6A', 'Borough': 'Vaughan', 'Neighborhood': 'Maple'},
{'Postcode': 'L4H', 'Borough': 'Vaughan', 'Neighborhood': 'Woodbridge North'},
{'Postcode': 'L4L', 'Borough': 'Vaughan', 'Neighborhood': 'Woodbridge South'},
{'Postcode': 'L4J', 'Borough': 'Vaughan', 'Neighborhood': 'Thornhill West'},
{'Postcode': 'L4K', 'Borough': 'Vaughan', 'Neighborhood': 'Concord'},
{'Postcode': 'L3T', 'Borough': 'Vaughan', 'Neighborhood': 'Thornhill East'},
]
```

Join Toronto neighbourhoods with the additional GTA neighbourhoods

```
df_group_Toronto=df_group_Toronto.append(other_gta_postcodes_df)
```

Latitude and Longitude of Neighbourhoods

The latitude and longitude for each Toronto postal code will be download from http://cocl.us/Geospatial_data for Toronto postal codes and manually defined for the additional communities.

The latitude and longitude for each postal code for the additional communities will be manually defined.

The two lists will be then be combined into a single postal code list.

These steps will be accomplished by the following Python code:

download CSV file containing Toronto geospatial data

```
!wget -q -O 'Geospatial_Coordinates.csv' http://cocl.us/Geospatial_data
```

Define geospatial data for additional communities in GTA (outside Toronto municipal boundaries)

```
other_gta_coord=[
{'Postal Code': 'L3P' , 'Latitude': 43.881618, 'Longitude':-79.265137},
{'Postal Code': 'L3R' , 'Latitude': 43.841888, 'Longitude':-79.322868},
{'Postal Code': 'L3S' , 'Latitude': 43.843712, 'Longitude':-79.266296},
{'Postal Code': 'L3T' , 'Latitude': 43.82235, 'Longitude':-79.395622},
{'Postal Code': 'L4B' , 'Latitude': 43.85564, 'Longitude':-79.402328},
{'Postal Code': 'L4C' , 'Latitude': 43.87067, 'Longitude':-79.44017},
{'Postal Code': 'L4E' , 'Latitude': 43.93737, 'Longitude':-79.45842},
{'Postal Code': 'L4H' , 'Latitude': 43.829121, 'Longitude':-79.583839},
{'Postal Code': 'L4J' , 'Latitude': 43.813759, 'Longitude':-79.455193},
{'Postal Code': 'L4K' , 'Latitude': 43.81234, 'Longitude':-79.502579},
{'Postal Code': 'L4L' , 'Latitude': 43.796539, 'Longitude':-79.576408},
{'Postal Code': 'L4S' , 'Latitude': 43.89402, 'Longitude':-79.42305},
{'Postal Code': 'L4T' , 'Latitude': 43.717251, 'Longitude':-79.643272},
{'Postal Code': 'L4V' , 'Latitude': 43.698631, 'Longitude':-79.621353},
{'Postal Code': 'L4W' , 'Latitude': 43.638969, 'Longitude':-79.620178},
{'Postal Code': 'L4X' , 'Latitude': 43.618889, 'Longitude':-79.581001},
{'Postal Code': 'L4Y' , 'Latitude': 43.60368, 'Longitude':-79.593529},
{'Postal Code': 'L4Z' , 'Latitude': 43.614052, 'Longitude':-79.647713},
{'Postal Code': 'L5A' , 'Latitude': 43.588001, 'Longitude':-79.607719},
{'Postal Code': 'L5B' , 'Latitude': 43.578651, 'Longitude':-79.631912},
{'Postal Code': 'L5C' , 'Latitude': 43.565239, 'Longitude':-79.652397},
{'Postal Code': 'L5E' , 'Latitude': 43.583511, 'Longitude':-79.563759},
{'Postal Code': 'L5G' , 'Latitude': 43.56361, 'Longitude':-79.583031},
{'Postal Code': 'L5H' , 'Latitude': 43.5406, 'Longitude':-79.611893},
{'Postal Code': 'L5J' , 'Latitude': 43.516109, 'Longitude':-79.632988},
{'Postal Code': 'L5K' , 'Latitude': 43.529228, 'Longitude':-79.662491},
{'Postal Code': 'L5L' , 'Latitude': 43.535118, 'Longitude':-79.693123},
{'Postal Code': 'L5M' , 'Latitude': 43.562309, 'Longitude':-79.721024},
{'Postal Code': 'L5N' , 'Latitude': 43.587849, 'Longitude':-79.760757},
{'Postal Code': 'L5P' , 'Latitude': 43.679218, 'Longitude':-79.630013},
{'Postal Code': 'L5R' , 'Latitude': 43.6022, 'Longitude':-79.66861},
{'Postal Code': 'L5S' , 'Latitude': 43.681541, 'Longitude':-79.675583},
{'Postal Code': 'L5T' , 'Latitude': 43.655689, 'Longitude':-79.671791},
{'Postal Code': 'L5V' , 'Latitude': 43.594349, 'Longitude':-79.690193},
{'Postal Code': 'L5W' , 'Latitude': 43.632359, 'Longitude':-79.71978},
{'Postal Code': 'L6A' , 'Latitude': 43.858509, 'Longitude':-79.508987},
{'Postal Code': 'L6B' , 'Latitude': 43.882778, 'Longitude':-79.22551},
{'Postal Code': 'L6C' , 'Latitude': 43.890339, 'Longitude':-79.336617},
{'Postal Code': 'L6E' , 'Latitude': 43.900269, 'Longitude':-79.266502},
{'Postal Code': 'L6G' , 'Latitude': 43.850609, 'Longitude':-79.332893},
{'Postal Code': 'L6S' , 'Latitude': 43.736549, 'Longitude':-79.731689},
{'Postal Code': 'L6V' , 'Latitude': 43.706749, 'Longitude':-79.762001},
{'Postal Code': 'L6W' , 'Latitude': 43.676311, 'Longitude':-79.73349},
{'Postal Code': 'L6X' , 'Latitude': 43.677891, 'Longitude':-79.795029},
{'Postal Code': 'L6Y' , 'Latitude': 43.652592, 'Longitude':-79.756828},
{'Postal Code': 'L6Z' , 'Latitude': 43.725368, 'Longitude':-79.793167},
{'Postal Code': 'L7A' , 'Latitude': 43.699699, 'Longitude':-79.826889}
]
```

Join Toronto geospatial data with the additional GTA neighbourhoods spatial data

```
geospatial_data=geospatial_data.append(other_gta_coord_df)
```

Merge neighbourhood data with geospatial data

```
df_group_Toronto_with_coord = df_group_Toronto.merge(geospatial_data, on = 'Postcode'
)
```

Geographic location of Toronto

Each of our maps will be centered on Toronto. We will determine the latitude and longitude values of the city by using the Nominatim module.

This will be accomplished by the following Python code:

Determine Latitude and Longitude of Toronto

```
address = 'Toronto, Ontario, Canada'

geolocator = Nominatim(user_agent="Toronto Details")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Toronto are {}, {}.'.format(latitude, longitude)
)
```

Map of Toronto Neighbourhoods

Create map of Toronto with Neighborhoods

The neighbourhood map of Toronto will be generated using the folium library.

This will be accomplished by the following Python code:

```
# create a map of Toronto
map_of_toronto = folium.Map(location = [latitude, longitude], zoom_start = 10)

#add neighborhood markers to the Toronto map
for lat, long, bor, neigh in zip(df_group_Toronto_with_coord['Latitude'], df_group_To
ronto_with_coord['Longitude'],
```

```
                                    df_group_Toronto_with_coord['Borough'], df_group_Tor
onto_with_coord['Neighborhood']):
    label = '{}, {}'.format(neigh, bor)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, long],
        radius = 7,
        popup = label,
        color = 'red',
        fill = True,
        fill_color = 'white',
        fill_opacity = 0.7,
        parse_html = False).add_to(map_of_toronto)

map_of_toronto
```

Venue Data

Venue data will be obtained by making API calls to the Foursquare URL. The API will be called for every one of the neighbourhoods and pass the latitude and longitude of the center of each of these neighbourhoods.

This will be accomplished by the following Python code:

Reusable API call to Foursquare

```
def getNearbyRestaurants(names, latitudes, longitudes, search_query,LIMIT=100):

    if latitude > 43.783:
         radius = 2500
    else:
         radius = 1750


    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secr
et={}&v={}&ll={},{}&query={}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            search_query,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
```

```
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_restaurants = pd.DataFrame([item for venue_list in venues_list for item in
venue_list])
    nearby_restaurants.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_restaurants)
```

Call Foursquare API for each neighbourhood

```
GTA_restaurants = getNearbyRestaurants(names=df_group_Toronto_with_coord['Neighborhoo
d'],
                                       latitudes=df_group_Toronto_with_coord['Latitude'],
                                       longitudes=df_group_Toronto_with_coord['Longitude'
],
                                       search_query='Chinese Restaurant'
                                      )
```

Determine which neighbourhoods and restaurants will be part of the further analysis

We will first determine the number of restaurants of the selected style in each neighbourhood. Any neighbourhoods which do not have sufficient restaurant density (as determined by the threshold set early) will be ignored in our analysis.

Restaurants may be located in close proximity to more than one neighbourhood. We will therefore need to remove any duplicate listings for the same restaurant.

This will be accomplished by the following Python code:

Count the number of restaurants (of selected restaurant type) in each neighbourhood

```
Neighborhood_counts = pd.DataFrame(GTA_restaurants['Neighborhood'].value_counts())
```

We will ignore any neighbourhoods which do not have a sufficient number (density) of restaurants of the selected style

```
Dense_Neighborhood = Neighborhood_counts.loc[Neighborhood_counts['RestaurantCount'] >
= 25]
```

In the earlier logic we have found all restaurants within a stated radius of the center of a postal code zone. Some zones (particularly in the downtown core) can be close together. This may result in the same restaurant being selected by multiple zones. In the next step we will drop any duplicate rows.

```
GTA_restaurants_to_map.drop_duplicates(subset=['Venue', 'Venue Latitude'], inplace=Tr
ue)
```

Generate map of Toronto with all restaurants which occur in a postal zone with a density of the restaurant type

The folium library will be utilized again to produce this map.

Use k-means algorithm to determine the number of restaurant clusters which exist

We will be using the k-means method of clustering the restaurant groupings. The code will be executed in a loop to determine the optimal k value based upon the sum of squared distances. By viewing an elbow graph, we will determine the preferred value of k.

This will be accomplished by the following Python code:

Use k-means algorithm to determine the number of restaurant clusters which exist

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```

```
from sklearn.preprocessing import StandardScaler
X = locations_to_cluster.values[:,0:2]
X = np.nan_to_num(X)
Clus_dataSet = StandardScaler().fit_transform(X)
```

```
Sum_of_squared_distances = []
K = range(1,15)
for k in K:
    km = KMeans(n_clusters=k)
    km = km.fit(Clus_dataSet)
    Sum_of_squared_distances.append(km.inertia_)
```

Create Elbow Graph

```
plt.plot(K, Sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('Sum_of_squared_distances')
plt.title('Elbow Method For Optimal k')
plt.show()
```

Perform clustering with optimal K value

```
clusterNum = ?
k_means = KMeans(init = "k-means++", n_clusters = clusterNum, n_init = 12)
k_means.fit(X)
labels = k_means.labels_
```

Assign cluster number to each restaurant

```
clustered_restaurants_to_map = pd.merge(GTA_restaurants_to_map, locations_to_cluster,
how='left', left_on=['Venue Latitude','Venue Longitude'], right_on = ['Venue Latitude
','Venue Longitude'])
```

Generate map of Toronto with all restaurants clustered

The folium library will be utilized again to produce this map.

Results

The city of Toronto was divided into 103 unique neighbourhoods

```
In [10]: df_group_Toronto.shape
Out[10]: (103, 3)
```

An additional 47 neighbourhoods outside of Toronto were added

```
In [12]: other_gta_postcodes_df.shape
Out[12]: (47, 3)
```

The full list of neighbourhoods to be included in our analysis is:

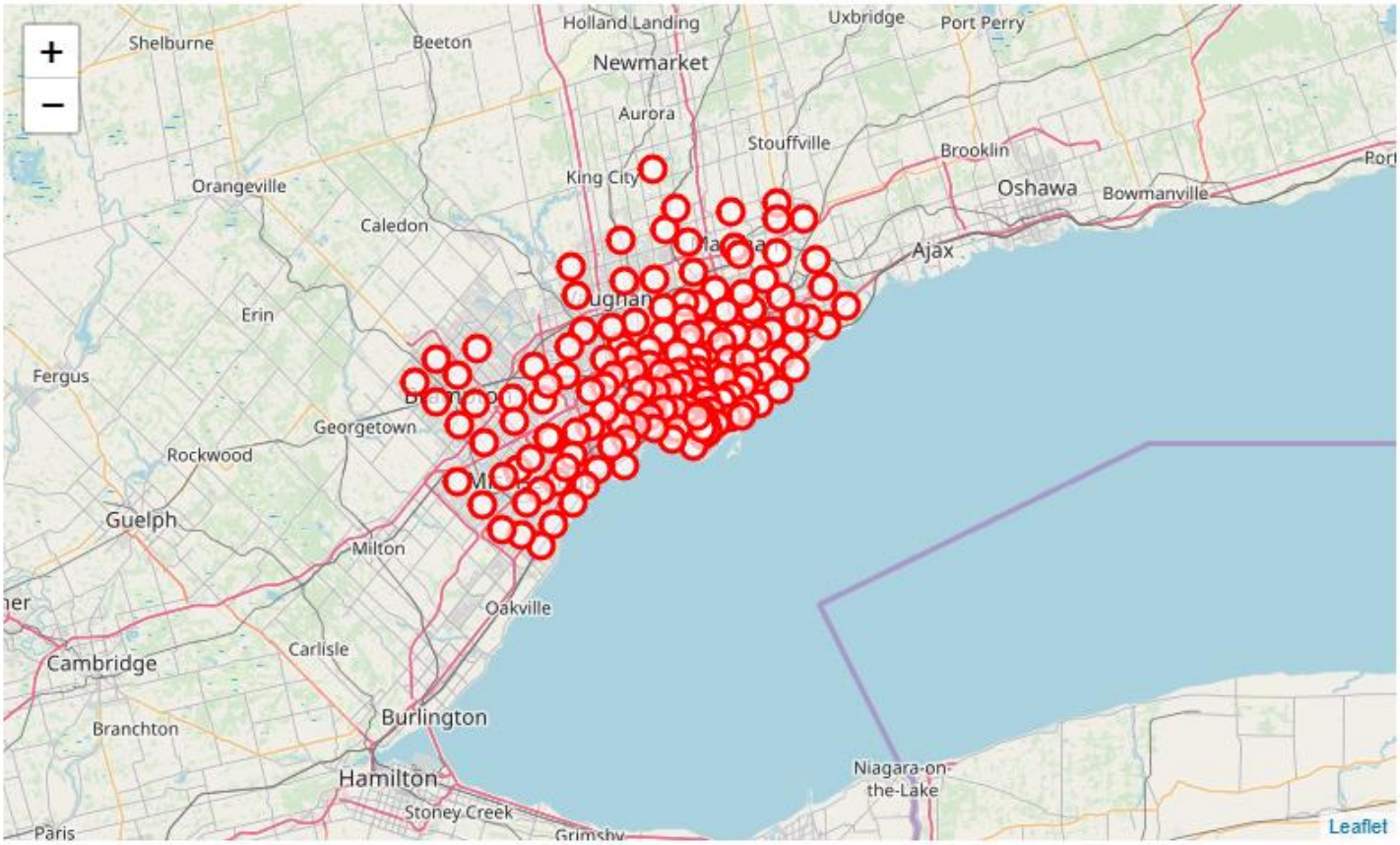
Postcode	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
M1G	Scarborough	Woburn	43.770992	-79.216917
M1H	Scarborough	Cedarbrae	43.773136	-79.239476
M1J	Scarborough	Scarborough Village	43.744734	-79.239476
M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304
M1R	Scarborough	Maryvale, Wexford	43.750072	-79.295849
M1S	Scarborough	Agincourt	43.794200	-79.262029
M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302
M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577
M1W	Scarborough	L'Amoreaux West	43.799525	-79.318389
M1X	Scarborough	Upper Rouge	43.836125	-79.205636
M2H	North York	Hillcrest Village	43.803762	-79.363452
M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556
M2K	North York	Bayview Village	43.786947	-79.385975
M2L	North York	Silver Hills, York Mills	43.757490	-79.374714
M2M	North York	Newtonbrook, Willowdale	43.789053	-79.408493
M2N	North York	Willowdale South	43.770120	-79.408493
M2P	North York	York Mills West	43.752758	-79.400049
M2R	North York	Willowdale West	43.782736	-79.442259
M3A	North York	Parkwoods	43.753259	-79.329656
M3B	North York	Don Mills North	43.745906	-79.352188
M3C	North York	Flemingdon Park, Don Mills South	43.725900	-79.340923

M3H	North York	Bathurst Manor, Downsview North, Wilson Heights	43.754328	-79.442259
M3J	North York	Northwood Park, York University	43.767980	-79.487262
M3K	North York	CFB Toronto, Downsview East	43.737473	-79.464763
M3L	North York	Downsview West	43.739015	-79.506944
M3M	North York	Downsview Central	43.728496	-79.495697
M3N	North York	Downsview Northwest	43.761631	-79.520999
M4A	North York	Victoria Village	43.725882	-79.315572
M4B	East York	Woodbine Gardens, Parkview Hill	43.706397	-79.309937
M4C	East York	Woodbine Heights	43.695344	-79.318389
M4E	East Toronto	The Beaches	43.676357	-79.293031
M4G	East York	Leaside	43.709060	-79.363452
M4H	East York	Thorncliffe Park	43.705369	-79.349372
M4J	East York	East Toronto	43.685347	-79.338106
M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
M4M	East Toronto	Studio District	43.659526	-79.340923
M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
M4P	Central Toronto	Davisville North	43.712751	-79.390197
M4R	Central Toronto	North Toronto West	43.715383	-79.405678
M4S	Central Toronto	Davisville	43.704324	-79.388790
M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160
M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049
M4W	Downtown Toronto	Rosedale	43.679563	-79.377529
M4X	Downtown Toronto	Cabbagetown, St. James Town	43.667967	-79.367675
M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
M5A	Downtown Toronto	Harbourfront, Regent Park	43.654260	-79.360636
M5B	Downtown Toronto	Ryerson, Garden District	43.657162	-79.378937
M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
M5H	Downtown Toronto	Adelaide, King, Richmond	43.650571	-79.384568
M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752
M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576
M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817
M5M	North York	Bedford Park, Lawrence Manor East	43.733283	-79.419750
M5N	Central Toronto	Roselawn	43.711695	-79.416936
M5P	Central Toronto	Forest Hill North, Forest Hill West	43.696948	-79.411307
M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678
M5S	Downtown Toronto	Harbord, University of Toronto	43.662696	-79.400049
M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049
M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420
M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846
M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280
M6A	North York	Lawrence Heights, Lawrence Manor	43.718518	-79.464763
M6B	North York	Glencairn	43.709577	-79.445073
M6C	York	Humewood-Cedarvale	43.693781	-79.428191
M6E	York	Caledonia-Fairbanks	43.689026	-79.453512
M6G	Downtown Toronto	Christie	43.669542	-79.422564
M6H	West Toronto	Dovercourt Village, Dufferin	43.669005	-79.442259
M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750
M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191

M6L	North York	Downsview, North Park, Upwood Park	43.713756	-79.490074
M6M	York	Del Ray, Keelesdale, Mount Dennis, Silverthorn	43.691116	-79.476013
M6N	York	The Junction North, Runnymede	43.673185	-79.487262
M6P	West Toronto	High Park, The Junction South	43.661608	-79.464763
M6R	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325
M6S	West Toronto	Runnymede, Swansea	43.651571	-79.484450
M7A	Queen's Park	Queen's Park	43.662301	-79.389494
M7R	Mississauga	Canada Post Gateway Processing Centre	43.636966	-79.615819
M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558
M8V	Etobicoke	Humber Bay Shores, Mimico South, New Toronto	43.605647	-79.501321
M8W	Etobicoke	Alderwood, Long Branch	43.602414	-79.543484
M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
M8Y	Etobicoke	Humber Bay, King's Mill Park, Kingsway Park So...	43.636258	-79.498509
M8Z	Etobicoke	Kingsway Park South West, Mimico NW, The Queen...	43.628841	-79.520999
M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
M9B	Etobicoke	Cloverdale, Islington, Martin Grove, Princess ...	43.650943	-79.554724
M9C	Etobicoke	Bloordale Gardens, Eringate, Markland Wood, Ol...	43.643515	-79.577201
M9L	North York	Humber Summit	43.756303	-79.565963
M9M	North York	Emery, Humberlea	43.724766	-79.532242
M9N	York	Weston	43.706876	-79.518188
M9P	Etobicoke	Westmount	43.696319	-79.532242
M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724
M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437
M9W	Etobicoke	Northwest	43.706748	-79.594054
L5A	Mississauga	Mississauga Valley / East Cooksville	43.588001	-79.607719
L5B	Mississauga	West Cooksville /Fairview / City Centre / East...	43.578651	-79.631912
L5C	Mississauga	West Creditview /Mavis / Erindale	43.565239	-79.652397
L5E	Mississauga	Central Lakeview	43.583511	-79.563759
L5G	Mississauga	SW Lakeview /Mineola / East Port Credit	43.563610	-79.583031
L5H	Mississauga	West Port Credit /Lorne Park / EastSheridan	43.540600	-79.611893
L5J	Mississauga	Clarkson /Southdown	43.516109	-79.632988
L5K	Mississauga	West Sheridan	43.529228	-79.662491
L5L	Mississauga	Erin Mills / Western Business Park	43.535118	-79.693123
L5M	Mississauga	Churchill Meadows /Central Erin Mills / South ...	43.562309	-79.721024
L5N	Mississauga	Lisgar / Meadowvale	43.587849	-79.760757
L5P	Mississauga	Mississauga	43.679218	-79.630013
L5R	Mississauga	West Hurontario / SWGateway	43.602200	-79.668610
L5S	Mississauga	Cardiff / NE Gateway	43.681541	-79.675583
L4T	Mississauga	Malton	43.717251	-79.643272
L5T	Mississauga	Courtney Park / EastGateway	43.655689	-79.671791
L4V	Mississauga	Wildwood	43.698631	-79.621353
L5V	Mississauga	East Credit	43.594349	-79.690193
L4W	Mississauga	Matheson / EastRathwood	43.638969	-79.620178
L5W	Mississauga	Meadowvale Village / West Gateway	43.632359	-79.719780
L4X	Mississauga	East Applewood / East Dixie / NE Lakeview	43.618889	-79.581001
L4Y	Mississauga	West Applewood / West Dixie / NW Lakeview	43.603680	-79.593529
L4Z	Mississauga	West Rathwood / EastHurontario / SEGateway /Sa...	43.614052	-79.647713
L7A	Brampton	Brampton West	43.699699	-79.826889
L6S	Brampton	Brampton North Central	43.736549	-79.731689
L6V	Brampton	Brampton Central	43.706749	-79.762001
L6W	Brampton	Brampton Southeast	43.676311	-79.733490

L6X	Brampton	Brampton Southwest	43.677891	-79.795029
L6Y	Brampton	Brampton South	43.652592	-79.756828
L6Z	Brampton	Brampton North Central	43.725368	-79.793167
L4B	Richmond Hill	Richmond Hill Southeast	43.855640	-79.402328
L4C	Richmond Hill	Richmond Hill Southeast	43.870670	-79.440170
L4E	Richmond Hill	Richmond Hill North/Oak Ridges / Lake Wilcox /...	43.937370	-79.458420
L4S	Richmond Hill	Richmond Hill Central	43.894020	-79.423050
L6B	Markham	Markham East	43.882778	-79.225510
L6C	Markham	Markham Northwest	43.890339	-79.336617
L6E	Markham	Markham Northeast	43.900269	-79.266502
L6G	Markham	Markham Northeast	43.850609	-79.332893
L3P	Markham	Markham Central	43.881618	-79.265137
L3R	Markham	Markham Outer Southwest	43.841888	-79.322868
L3S	Markham	Markham Southeast	43.843712	-79.266296
L6A	Vaughan	Maple	43.858509	-79.508987
L4H	Vaughan	Woodbridge North	43.829121	-79.583839
L4L	Vaughan	Woodbridge South	43.796539	-79.576408
L4J	Vaughan	Thornhill West	43.813759	-79.455193
L4K	Vaughan	Concord	43.812340	-79.502579
L3T	Vaughan	Thornhill East	43.822350	-79.395622

The following map shows each of these neighbourhoods:



The Foursquare API was called for each of these neighbourhoods and it was determine that the following neighbourhoods contain more than the threshold number(25) of Chinese Restaurants:

Neighborhood	RestaurantCount
Queen's Park	100
Adelaide, King, Richmond	100
Ryerson, Garden District	100
First Canadian Place, Underground city	100
Chinatown, Grange Park, Kensington Market	100
Harbord, University of Toronto	100
Central Bay Street	100
Design Exchange, Toronto Dominion Centre	96

Neighborhood	RestaurantCount
Commerce Court, Victoria Hotel	92
Church and Wellesley	71
St. James Town	67
Agincourt North, L'Amoreaux East, Milliken, Steeles East	61
Stn A PO Boxes 25 The Esplanade	45
Berczy Park	41
Harbourfront East, Toronto Islands, Union Station	39
Richmond Hill Southeast	39
Agincourt	38
Markham Northeast	34
Cabbagetown, St. James Town	27

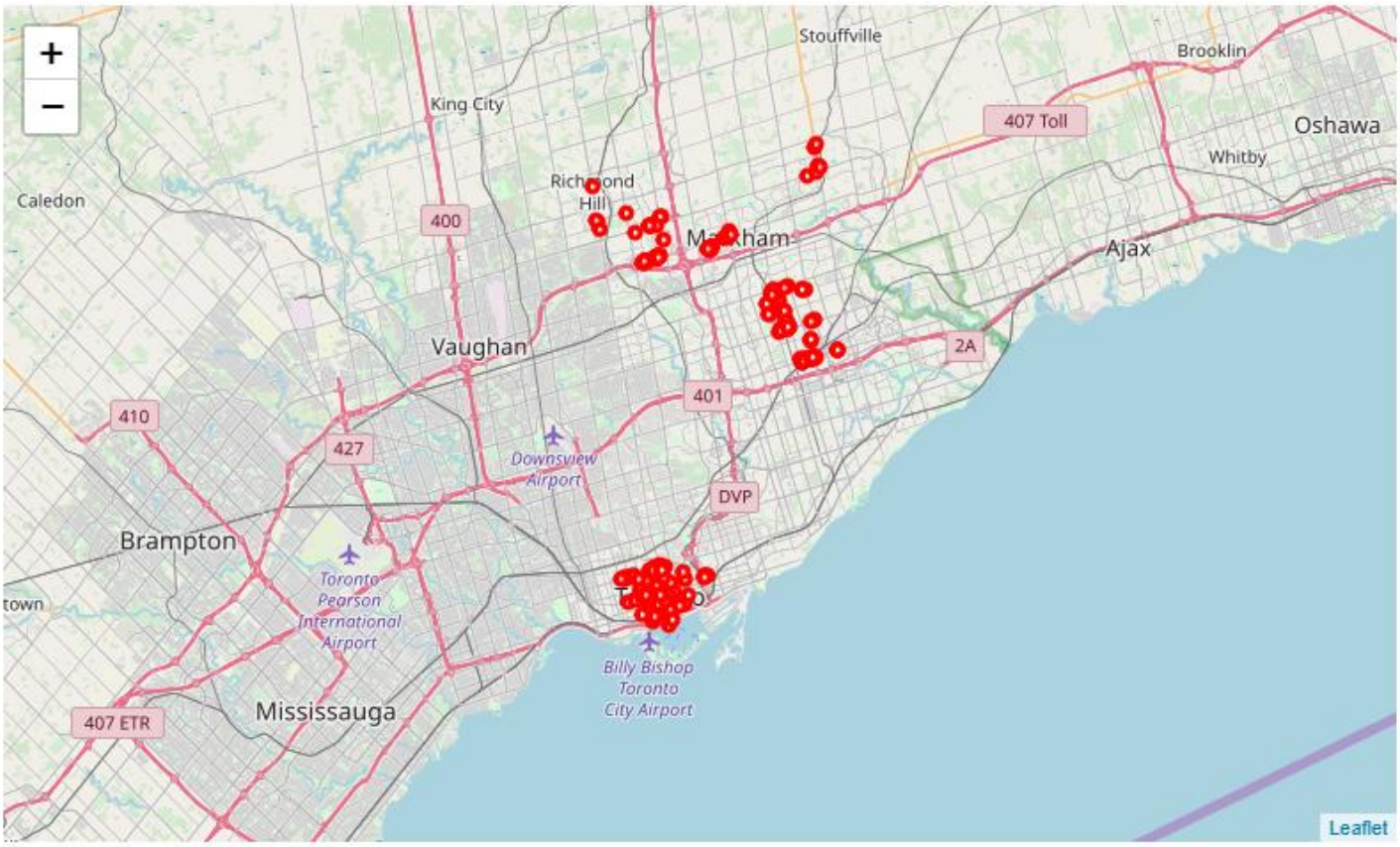
Only 19 of the 150 neighbourhoods contained at least 25 Chinese Restaurants.

Note: the maximum number of restaurants counted for a neighbourhood will be limited to 100 based upon the parameters utilized for the Foursquare API calls.

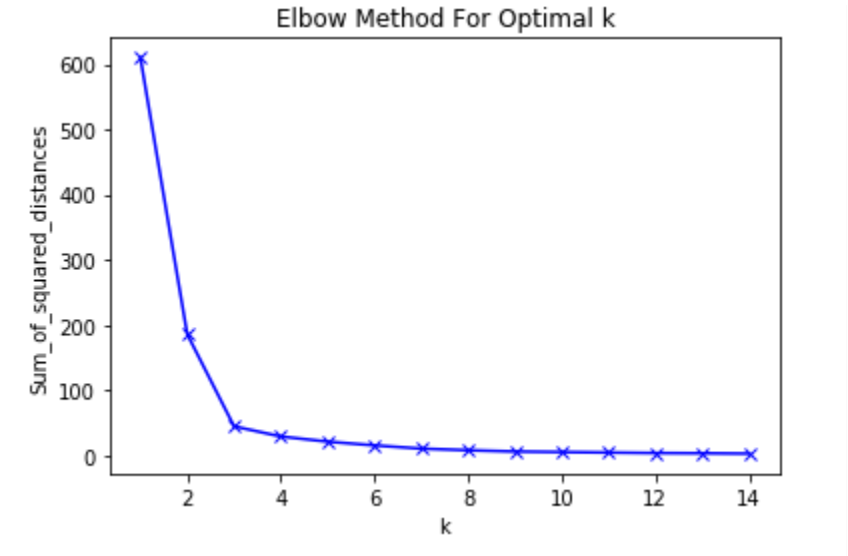
In these neighbourhoods 305 unique restaurants were identified by Foursquare

```
In [47]: GTA_restaurants_to_map.shape
Out[47]: (305, 3)
```

We can visualize the locations of these 305 restaurants by the following map:

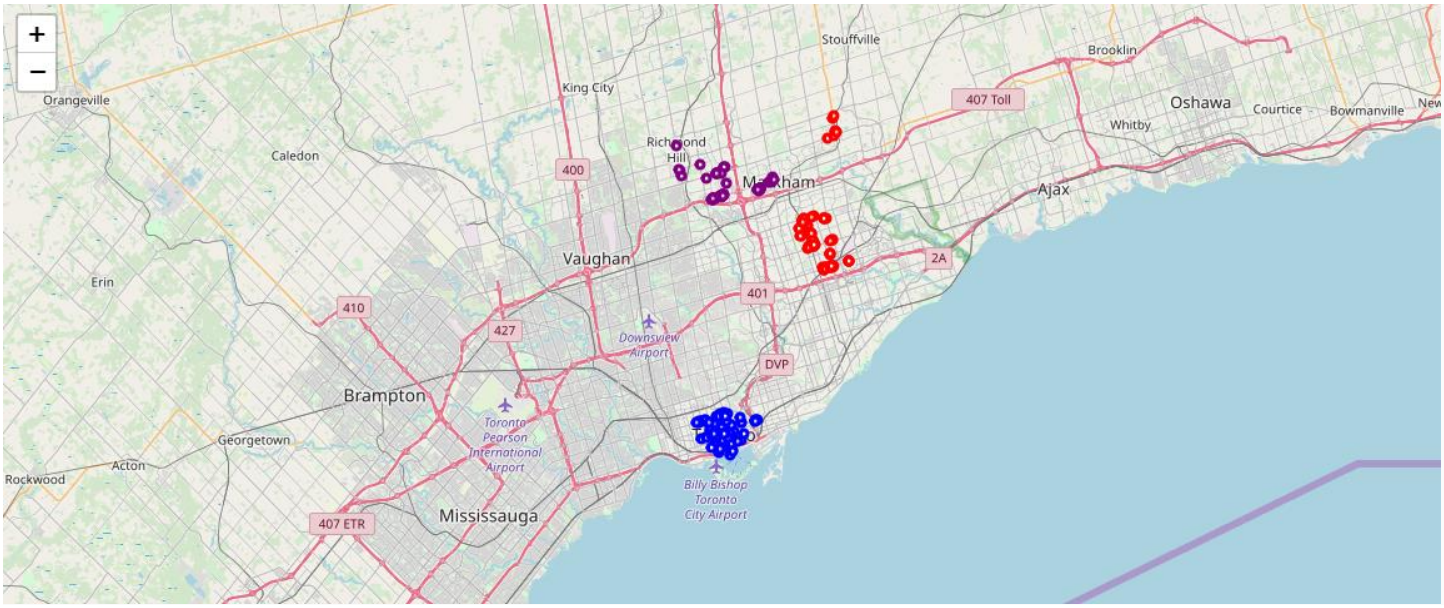


We wish to identify the number of unique groups (clusters) of Chinese Restaurants which exist in the GTA. By utilizing the k-means algorithm on the restaurant data and generating an elbow graph we determine that 3 clusters is the optimal number of clusters.



When the k-means algorithm is executed with a value of k=3 we are able to clusters each restaurant into these three groups.

The following map allows us to visualize the clustering of Chinese restaurants which exist in the GTA.



Based upon the clustering indicated in the above map we can observe that the greater Toronto area now includes not only the original Chinatown in downtown Toronto – but two additional Chinatowns have developed in the suburbs around the city.

Discussion

The approach of analyzing the makeup of services (in our case restaurants) to suggest the ethnic populations of these neighbourhoods appear to be viable. Based upon my own knowledge of Toronto and surrounding communities the observed clustering does coincide with population patterns.

In order to fine-tune the results, we would probably want to further analyze the choices made in the assumptions. In particular:

a) A restaurant is considered to be included in a neighbourhood if it's geographic location (latitude and longitude) is within a defined threshold of the center of the neighbourhood. Generally speaking the physical size of neighbourhoods within the city of Toronto are smaller than the size of neighbourhoods north of the city. This is a result of higher population densities within Toronto. In order to best accommodate these size differences, the threshold for inclusion in the neighbourhood has been defined as 1750 meters for Toronto and 2500 meters for north of the city.

   The distances to be considered (1750 meters for Toronto and 2500 meters for north of the city) may not be the ideal values. This is particularly true north of the city where due to lower populations density the geographic size of the neighbourhoods is significantly larger than inside the Toronto borders.

b) The minimum density of a restaurant type (Chinese, Italian, Korean, etc.) has been chosen as 25. If a neighbourhood does not contain at least 25 restaurants of the type, then the neighbourhood will be excluded

   The density of 25 restaurants may not be the optimal value.

c) The additional communities of Mississauga, Brampton, Vaughan, Richmond Hill, and Markham have been included in the analysis

   Perhaps additional communities should be considered.

Conclusion

The analysis of the density of restaurant types within neighbourhoods does appear to provide insight into understanding the composition of the neighbourhoods in Toronto. This determination would provide urban planners, business owners, and even municipal leaders valuable insight into the successful planning and delivery of future services to these communities.