Capstone Project - The Battle of Neighborhoods

1. Introduction/Business Statement

Toronto, a population around over 6 million is the capital of the province of Ontario. It is a major Canadian city along Lake Ontario's northwestern shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen's Park to 400-acre High Park and its trails, sports facilities and zoo.

According to the 2011 National Household Survey (NHS) at https://en.wikipedia.org/wiki/Demographics_of_Toronto , there are 1,264,395 non-Whites, or 20.2% of Canada's visible minority population, live in the city of Toronto; of this, approximately 70 percent are of Asian ancestry.

A good tasty Asian cuisines such as Nasi Lemak (Malaysia), Tom Yam(Thailand), Roti Canai(Malaysia), Lamb Korma(Indian) and many other good Asian food are definitely among the sought after menus in the correct neighborhood.

The questions for opening an asian restaurant would be:

1. Which area have high concentration of ethnic asian
2. Which area have less asian restaurant
3. What are common venues surrounding the neighbourhood

The problem statement for the project is

1. **Where is the best location to open an asian cuisine restaurant in Toronto**

The stakeholder or the target audience for this would be

1. Any business owner who would want to venture into food industry
2. Food chain business owner who would want to expand their businesses
3. Marketing company in food business industry
4. Fund manager who would want to fund a startup for food business
5. Food lovers themselves who just wish to have a good asian cuisines
6. Tourist who are looking for good asian cuisine

2. Data

We will be using data source below to answer the problem we sough to solve. Explanation of each data is inluded.

1. Data contains postal code, borough and neighborhood for Toronto area.
    1. https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,
    2. Description : This page will contain the table of Toronto neighborhood with postal code, borough and neighbourhood. This is very important data and where we start the project with. We will be scraping the data from the website using BeautifulSoup library and load into dataframe from manipulation

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Ontario Provincial Government

2. Geospatial information for Toronto borough and neighborhood
    1. https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.cs
    2. Description : This data will provide the latitude and longitude information for the neighborhood in Toronto. We will read them from the source (csv format) and load to dataframe and then match them to the Toronto neighborhood data we have above.

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

3. Demography breakdown of Toronto area based on the ethnic
    1. Demographics of Toronto – Wikipedia
    2. Description : This data will provide concentration of ethnic in Toronto areas. We can then zoom into these areas where asian ethnic are high in density.

The top visible-minority groups per **Community Council** (2016 Census) [34] are as follows:

- **Toronto & East York** (847,045): White: 65.3%, Chinese: 8.9%, South Asian: 6.7%, Black: 5.6%
- **North York** (638,100): White: 47.4%, Chinese: 14.0%, South Asian: 8.5%, Filipino: 8.0%, West Asian: 5.3%, Black: 5.2%
- **Scarborough** (623,135): White: 26.5%, South Asian: 25.4%, Chinese: 19.0%, Black: 10.8%, Filipino: 8.4%
- **Etobicoke York** (583,395): White: 48.9%, Black: 15.7%, South Asian: 11.9%, Latin American: 5.6%

Community Council	1996	*Percent change*	2001	*Percent change*	2006	*Percent change*	2011	*Total percent change (1996-2011)*
Toronto & East York	28.7%	*+3.9%*	32.6%	*+0.5%*	33.1%	*0.0%*	33.1%	***+4.4%***
Etobicoke York	31.1%	*+6.0%*	37.1%	*+3.4%*	40.5%	*+2.8%*	43.3%	***+12.2%***
North York	39.7%	*+2.9%*	42.6%	*+5.1%*	47.7%	*+3.5%*	51.2%	***+11.5%***
Scarborough	51.9%	*+8.1%*	60.0%	*+7.4%*	67.4%	*+2.9%*	70.3%	***+18.4%***

4. Foursquare location data via API for identifying and exploring the selected localities and surroundings areas where high in ethnic asian population
    1. Creation of developer account is required to have an access to Foursquare API.

5. Foursquare API can be accessed as below
    1. url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{}&v={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, radius, LIMIT)
    2. Credentials and parameters must be supplied for each of the request such as
        1. CLIENT_ID = 'xxx' # your Foursquare ID
        2. CLIENT_SECRET = 'xxx' # your Foursquare Secret
        3. ACCESS_TOKEN = 'xxx' # your FourSquare Access Token
        4. VERSION = '20180604'
    3. Foursquare location data will be critical in exploring the localities and find the concentration of asian food outlet or restaurant in those searched areas.
    4. Hence, we would leverage Foursquare location data to:
        1. identify the location of most dense Asian origin population in Toronto
        2. identify the top 10 most common venues for neighborhood with dense Asian origin
        3. Less dense Asian restaurant in the dense Asian origin area

3. Methodology

1. We created a dataframe called df_neigh from scraped Toronto postal code , Borough and Neighboorhood data from Wikipedia page. It consists of 103 rows for all the neighborhood. All 'Not assigned' postal code from the webpage are ignored.

2. The next step is to populate the dataframe with latitude and longitude information. This is gathered from our second source of data ie Geospatial datasheet. The data is loaded from csv formated source put into dataframe. The data is then joined with df_neigh with the same PostalCode to create the main data for Toronto neighborhood with latitude and longitude information. The dataframe is called Toronto_data. Toronto_data.tail(5).

	PostalCode	Borough	Neighborhood	Latitude	Longitude
98	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944
99	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
100	M7Y	East Toronto	Enclave of M4L	43.662744	-79.321558
101	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.636258	-79.498509
102	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.628841	-79.520999

3. We would now want to add demographic data to reflect the population density of Asian ehtnic in Toronto neighborhood. The data available in the wikipedia is extracted and saved in csv format locally. We then load this csv data to a dataframe. Now, we have a good information of selecting the best neighborhood with high ethnic Asian population in Toronto area. We call this df_demographic. There are about 7 neighborhood with high Asian ethnic density in Toronto. Scarborough being the highest percentage of 19% population consisting of Asian ethnic.

	Borough	Race1	Race2	Race3	Race4	Race5	Race6
0	Toronto,East York	White: 65.3%	Chinese: 8.9%	South Asian: 6.7%	Black: 5.6%	NaN	NaN
1	North York	White: 47.4%	Chinese: 14.0%	South Asian: 8.5%	Black: 5.2%	Filipino: 8.0%	West Asian: 5.3%
2	Scarborough	White: 26.5%	Chinese: 19.0%	South Asian: 25.4%	Black: 10.8%	Filipino: 8.4%	NaN
3	Etobicoke,York	White: 48.9%	Black: 15.7%	South Asian: 11.9%	, Latin American: 5.6%	NaN	NaN

```
df_demographic['Borough']
```

```
0    Toronto,East York
1           North York
2          Scarborough
3       Etobicoke,York
Name: Borough, dtype: object
```

4. Master Toronto neighborhood data is then filtered with this high ethnic Asian populated area. There are about only 83 neighborhood identified to have high ethnic Asian population .

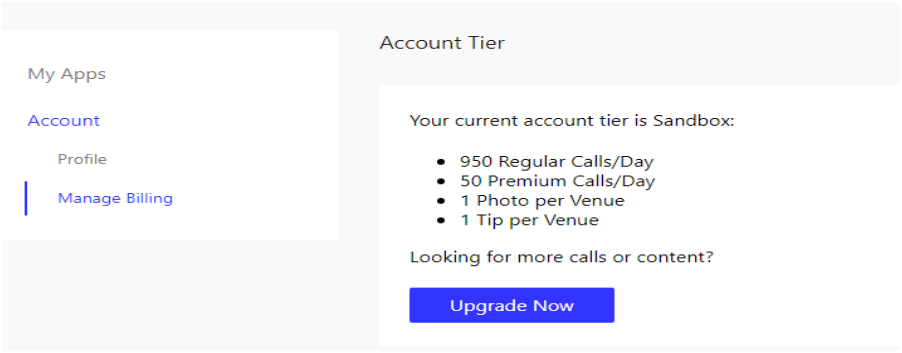
```
#will search for Asian restaurant nearby neighborhood
search_query = 'Asian'
radius = 1000  # set to 1000m
LIMIT = 100 # set to 100 venues
print(search_query + ' .... OK!')
for i in range(len(Toronto_data)):
    url_x = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oauth_t
    results_x = requests.get(url_x).json()
    try:
        num_venues = len(results_x['response']['venues']) #get number of venues
    except:
        num_venues = 0
    print("<>",i,"<>",num_venues,"<>",Toronto_data.loc[i,'Latitude'],"<>",Toronto_data.loc[i,'Longitude
    Toronto_data.loc[i,'AsianRestaurant'] = num_venues
    print('There are {} popular spots around {} in {}.'.format(num_venues,Toronto_data.loc[i,'Neighborh
#dataframe2 = json_normalize(results_x['response']['venues'])
```

```
gh.
<> 56 <> 0 <> 43.7527583 <> -79.4000493
There are 0 popular spots around York Mills West in North York.
<> 57 <> 0 <> 43.696319 <> -79.5322424
There are 0 popular spots around Westmount in Etobicoke.
<> 58 <> 1 <> 43.7500715 <> -79.2958491
There are 1 popular spots around Wexford, Maryvale in Scarborough.
```

```
Toronto_data.shape
```

```
(83, 6)
```

5. We would use Foursquare API for collecting location data from the web. Developer account has to be created first to access the API. Free developer account entitled us to have 950 regular calls/day & additional rights as below.



6. Now we get the get the list of Asian restaurant in Asian ethnic populated

neighborhood

1. url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oauth_token={}&v={}&query={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, neighborhood_latitude, neighborhood_longitude, ACCESS_TOKEN, VERSION, search_query, LIMIT)

```
#final_df = df.sort_values(by=['2'], ascending=False)
Toronto_data.sort_values('AsianRestaurant',ascending=False).head(20)
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude	AsianRestaurant
22	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	12.0
77	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	10.0
28	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	10.0
37	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576	9.0
8	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	9.0
41	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	8.0
65	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	8.0
79	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	7.0
14	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	7.0
72	M5W	Downtown Toronto	Enclave of M5E	43.646435	-79.374846	6.0
63	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049	6.0
33	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752	4.0
62	M1S	Scarborough	Agincourt	43.794200	-79.262029	4.0

7. Geopy.Nominatim is being used to locate the latitude and logitude on Toronto

```
address = 'Toronto, Ontario'

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

```
The geograpical coordinate of Toronto are 43.6534817, -79.3839347.
```

8. We use K-means clustering algorithm to cluster Toronto venues category .For this project, we would use K equal to 5 and train it with Toronto_group dataframe. This algo will choose the similarity(closeness) of the neighborhood based on the category and cluster them together.

	Neighborhood	Venue Category_Accessories Store	Venue Category_Adult Boutique	Venue Category_Airport	Venue Category_Airport Food Court	Venue Category_Airport Gate	Ve Category_Air Lou
0	Agincourt	0.000000	0.000000	0.000000	0.0000	0.0000	0.0
1	Alderwood, Long Branch	0.000000	0.000000	0.000000	0.0000	0.0000	0.0
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000	0.000000	0.000000	0.0000	0.0000	0.0
3	Bayview Village	0.000000	0.000000	0.000000	0.0000	0.0000	0.0
4	Bedford Park, Lawrence Manor East	0.000000	0.000000	0.000000	0.0000	0.0000	0.0

4. Result/Findings

1. 7 identified borough with high density of Asian ethnic . Scarborough being the most dense with more than 50% of ethnic Asian population.

	Borough	Race1	Race2	Race3	Race4	Race5	Race6
0	Toronto,East York	White: 65.3%	Chinese: 8.9%	South Asian: 6.7%	Black: 5.6%	NaN	NaN
1	North York	White: 47.4%	Chinese: 14.0%	South Asian: 8.5%	Black: 5.2%	Filipino: 8.0%	West Asian: 5.3%
2	Scarborough	White: 26.5%	Chinese: 19.0%	South Asian: 25.4%	Black: 10.8%	Filipino: 8.4%	NaN
3	Etobicoke,York	White: 48.9%	Black: 15.7%	South Asian: 11.9%	, Latin American: 5.6%	NaN	NaN

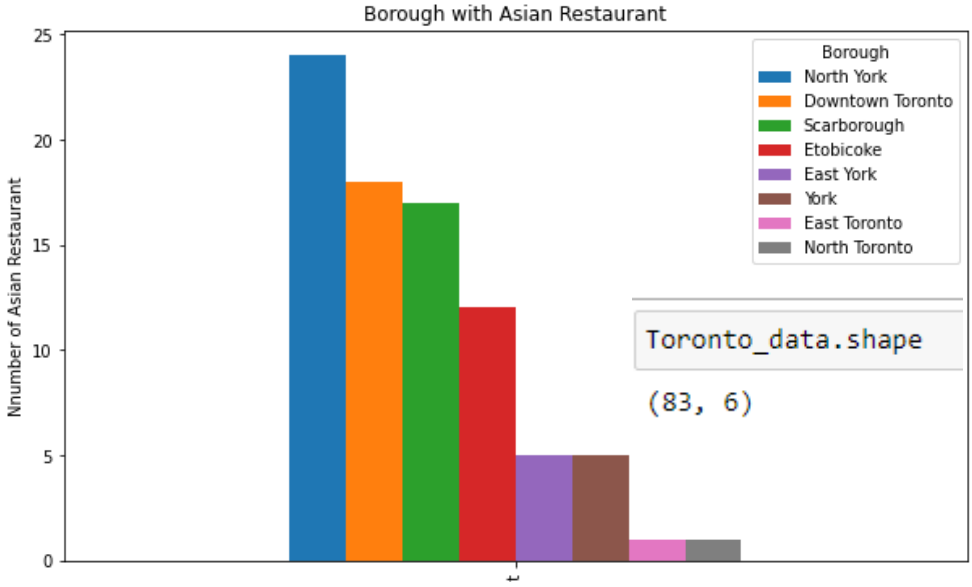
```
df_demographic['Borough']
```

```
0    Toronto,East York
1           North York
2          Scarborough
3       Etobicoke,York
Name: Borough, dtype: object
```

Table shows high density of Asian ethnic in Toronto

2. We have only 83 total neighborhood with high density of Asian ethnic based on available demographic data. North York having the highest number of dense asian ethnic population with 24 neighborhood followed by burough of Downtown Toronto and Scarborough.

```
Text(0.5, 1.0, 'Borough with Asian Restaurant')
```



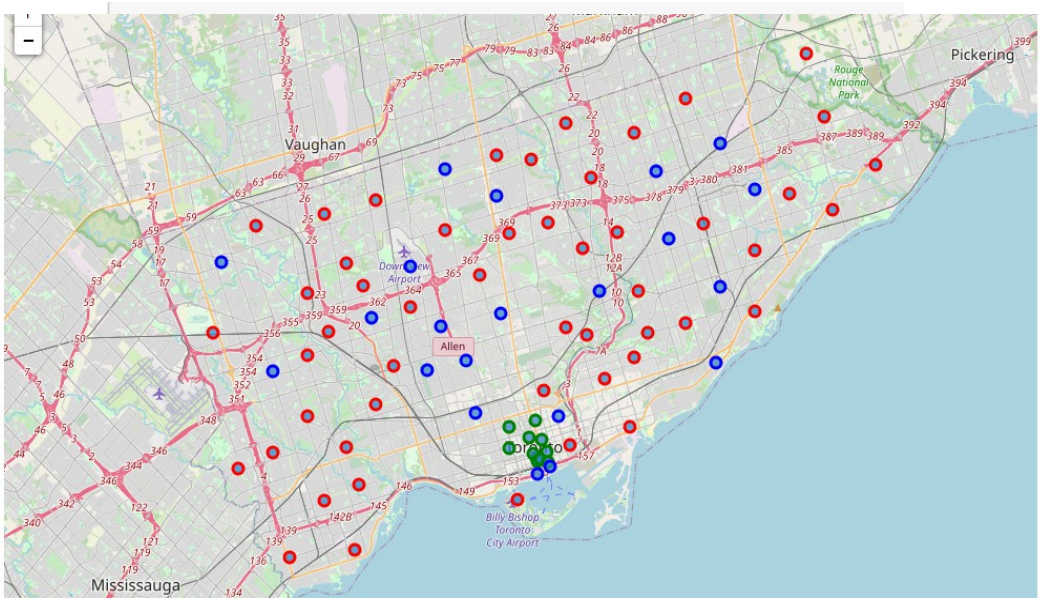
Histogram shows the number of dense Asian ethnic neighborhood in Toronto

```
T_data = Toronto_data[['Borough','Neighborhood',]].groupby('Borough').count()
T_data.sort_values(['Neighborhood'], ascending=False, axis=0, inplace=True)
T_data
```

Borough	Neighborhood
North York	24
Downtown Toronto	18
Scarborough	17
Etobicoke	12
East York	5
York	5
East Toronto	1
North Toronto	1

Borough with dense Asian ethnic community

3. Visualization of asian restaurant in asian dense populated area. 5 or more asian restaurant located within 500 meter of the neighborhood is colored in green. While 5 and less is in blue. It is obvious that high concentrated Asian restaurant is in Downtown Toronto borough with 101 Asian restaurant



Map show the neighborhod with number of Asian restaurant.
Green dot represent > 5, blue > 0 & < 6 while red = 0

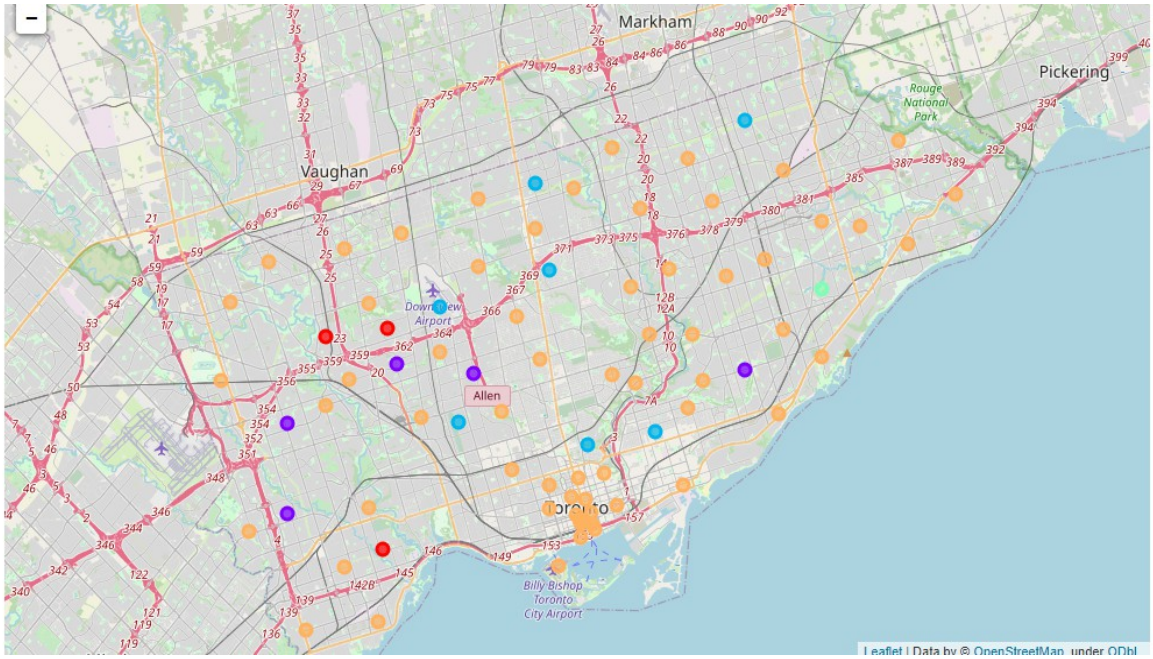
4. Downtown Toronto borough being the most dense Asian restaurant with Central Bay, First Canadian Place, Undergound City, Richmond, Adelaide & King neighborhood having 10 or more restaurant.

```
#final_df = df.sort_values(by=['2'], ascending=False)
Toronto_data.sort_values('AsianRestaurant',ascending=False).head(20)
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude	AsianRestaurant
22	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	12.0
77	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	10.0
28	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	10.0
37	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576	9.0
8	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	9.0
41	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	8.0
65	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	8.0
79	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160	7.0
14	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	7.0
72	M5W	Downtown Toronto	Enclave of M5E	43.646435	-79.374846	6.0
63	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049	6.0
33	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752	4.0
62	M1S	Scarborough	Agincourt	43.794200	-79.262029	4.0

Number of Asian restaurant in the neighborhood

5. Clustering of Toronto neighborhood based on categorical venues as below. Cluster 4 being dominant with most number of neighborhood.



Map show the clustering of neighborhood by venues category

6. Most common venues for Scarborough, North York, East York and East Toronto

```
]: Toronto_merged[Toronto_merged['Borough'] == "Scarborough"]['1st Most Common Venue'].unique()

]: array(['Venue Category_Fast Food Restaurant', 'Venue Category_Bar',
       'Venue Category_Donut Shop', 'Venue Category_Coffee Shop',
       'Venue Category_Fried Chicken Joint', 'Venue Category_Playground',
       'Venue Category_Hobby Shop', 'Venue Category_Bakery',
       'Venue Category_Motel', 'Venue Category_College Stadium',
       'Venue Category_Indian Restaurant', 'Venue Category_Auto Garage',
       'Venue Category_Skating Rink', 'Venue Category_Park'], dtype=object)
```

```
Toronto_merged[Toronto_merged['Borough'] == "North York"]['1st Most Common Venue'].unique()

array(['Venue Category_Bus Stop', 'Venue Category_Portuguese Restaurant',
       'Venue Category_Clothing Store',
       'Venue Category_Japanese Restaurant', 'Venue Category_Restaurant',
       'Venue Category_Fast Food Restaurant', 'Venue Category_Bank',
       'Venue Category_Massage Studio', 'Venue Category_Airport',
       'Venue Category_Grocery Store', 'Venue Category_Basketball Court',
       'Venue Category_Pizza Place', 'Venue Category_Park',
       'Venue Category_Food Truck', 'Venue Category_Sandwich Place',
       'Venue Category_Baseball Field', 'Venue Category_Ramen Restaurant',
       'Venue Category_Athletics & Sports', 'Venue Category_Pharmacy'],
      dtype=object)
```

```
: Toronto_merged[Toronto_merged['Borough'] == "East York"]['1st Most Common Venue'].unique()

: array(['Venue Category_Pizza Place', 'Venue Category_Spa',
       'Venue Category_Coffee Shop', 'Venue Category_Indian Restaurant',
       'Venue Category_Park'], dtype=object)

: Toronto_merged[Toronto_merged['Borough'] == "East Toronto"]['1st Most Common Venue'].unique()

: array(['Venue Category_Light Rail Station'], dtype=object)
```

Extraction show Scarborough and North York common venues

5. Discussion

1. Downtown Toronto having the most of Asian restaurant based on Foursquare location data with 101 then followed by Scarborough with 10. The rest of the neighborhood much lesser with East Toronto and East York having none of Asian restaurant around.
2. Scarborough, NorthYork having around 10 and less Asian restaurant in the neighborhood are the best location for setting up new Asian restaurant
3. As for the data that we have, East Toronto is the least prefered place to setup a restaurant due to single common venues.

6. Conclusion

- Based on the data, findings and discussion highligted above, the recommend ation for the best location for new Asian restaurant are in Scarborough and North York due to many common venues and with none or less having an asian restaurant.