Introduction

This project will analyse the feasibility of opening a Canadian Chinese Restaurant in the city of Toronto for our client. We will make a conclusion on the nominal area where the restaurant has to be opened based on the factors of sustainability of business, accessibility, type of cuisine and exclusivity in the area. This project deals with the step by step process of

Analyzing the problem that we are trying to solve for our client

What is the best approach to solve the problem

What data do we need in order for us to acheive the solution

The ways we can acquire the data that we need to solve the problem

What insights have we infered from the visualization of the data acquired

What statistical methods can be used to leverage the data that we have

What are the predictions and conclusions that have been drawn on the methodology we have applied

Objective

To suggest the best possible location for our client to take advantage of the various factors in a specific locality which is based on a multivariate analysis of the different factors that can influence a business such as

Population Density

Demographics of the locality

Cuisine which are popular in the neighborhood

Exclusivity of the type of cuisine we are offering

Accesibility to the location for the target customers

Data Description

As mentioned above the data must provide us a detailed view on the factors that influence the business of the restaurant. These include Neighborhoods in the city of Toronto, Geospatial coordinates of each neighborhood in order for us to determine the closeness of neighboorhoods, The venues which are located in the city to assess the competitors and the type of cuisines which are popular in the city.

The following include the data sources which are being used:

Toronto Neighborhoods - https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).

Cuisine of Toronto - https://en.wikipedia.org/wiki/Cuisine_in_Toronto (https://en.wikipedia.org/wiki/Cuisine_in_Toronto) to analyse the popular cuisines based on the demographics of Canada

Demographics of Neighborhoods - https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods (https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods) to analyse the taget customers and sustainability of business

Geospatial Data of Nieghboorhoods - http://cocl.us/Geospatial_data (http://cocl.us/Geospatial_data) for the purpose of determining the closeness and ease of accessibility

Fousquare API : "https://developer.foursquare.com/ (https://developer.foursquare.com/)" for venues in each neighborhood.

Methodology

In [2]:

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

def plot(x, y, xlabel, ylabel):
    plt.figure(figsize=(20,10))
    plt.plot(np.arange(2, x), y, 'o-')
    plt.xlabel(xlabel)
    plt.ylabel(ylabel)
    plt.xticks(np.arange(2, x))
    plt.show()
```

In [3]:

```
url='https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
page = requests.get(url)
soup = BeautifulSoup(page.content,'html.parser')
table = soup.table
headers = table.find_all('th')
rows = table.find_all('tr')
dict_df = {}
for i in range(len(headers)):
    col_content = headers[i].text.strip()
    dict_df[col_content] = []
for i in range(1,len(rows)):
    row_content = rows[i].find_all('td')
    for j in range(len(row_content)):
        cell_content = row_content[j].text.rstrip()
        if j == 0:
            dict_df['Postal Code'].append(cell_content)
        if j == 1:
            dict_df['Borough'].append(cell_content)
        if j == 2:
            dict_df['Neighborhood'].append(cell_content)
```

In [4]:

```
df = pd.DataFrame.from_dict(dict_df)
pd.set_option('display.max_columns', None)
df.head()
```

Out[4]:

	Postal Code	Borough	Neighborhood
0	M1A	Not assigned	
1	M2A	Not assigned	
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

Ignoring 'Not assigned' values of 'Borough' columns

In [5]:

```
df = df[df['Borough'] != 'Not assigned']
df.reset_index(drop = True, inplace = True)
df.head()
```

Out[5]:

	Postal Code	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	Downtown Toronto	Queen's Park, Ontario Provincial Government

Calling shape method on the dataframe to retrieve the rows and columns size

In [6]:

```
df.shape
```

Out[6]:

```
(103, 3)
```

Assignment of Latitude and Longitude values to each Postal Code

In [7]:

```
geo_spatial_df = pd.read_csv('https://cocl.us/Geospatial_data')
geo_spatial_df.head()
```

Out[7]:

	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

In [8]:

```
df = pd.merge(df, geo_spatial_df, on='Postal Code')
df.head()
```

Out[8]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

In [1]:

```
import requests # library to handle requests
import pandas as pd # library for data analsysis
import numpy as np # library to handle data in a vectorized manner
import random # library for random number generation

!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim # module to convert an address into latitude and
 longitude values

# libraries for displaying images
from IPython.display import Image
from IPython.core.display import HTML

# tranforming json file into a pandas dataframe library
from pandas.io.json import json_normalize

!conda install -c conda-forge folium=0.5.0 --yes
import folium # plotting library
import numpy as np # library to handle data in a vectorized manner

import pandas as pd # library for data analsysis
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json # library to handle JSON files
from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

# Matplotlib and associated plotting modules
import matplotlib.cm as cm
import matplotlib.colors as colors

# import k-means from clustering stage
from sklearn.cluster import KMeans

print('Libraries imported.')
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.

Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.

Libraries imported.
```

Creating dataframe with boroughs of Toronto

In [10]:

```
neighborhoods = df
neighborhoods = neighborhoods[neighborhoods.Borough.str.contains('Toronto', case=False
)]
neighborhoods.reset_index(drop = True, inplace = True)
print('The dataframe has {} boroughs and {} neighborhoods.'.format(
        len(neighborhoods['Borough'].unique()),
        neighborhoods.shape[0]
    )
)
neighborhoods
```

```
The dataframe has 4 boroughs and 39 neighborhoods.
```

Out[10]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031
5	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
6	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
7	M6G	Downtown Toronto	Christie	43.669542	-79.422564
8	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568
9	M6H	West Toronto	Dufferin, Dovercourt Village	43.669005	-79.442259
10	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752
11	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750
12	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
13	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576
14	M6K	West Toronto	Brockton, Parkdale Village, Exhibition Place	43.636847	-79.428191
15	M4L	East Toronto	India Bazaar, The Beaches West	43.668999	-79.315572
16	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817
17	M4M	East Toronto	Studio District	43.659526	-79.340923
18	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790
19	M5N	Central Toronto	Roselawn	43.711695	-79.416936
20	M4P	Central Toronto	Davisville North	43.712751	-79.390197
21	M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.696948	-79.411307
22	M6P	West Toronto	High Park, The Junction South	43.661608	-79.464763
23	M4R	Central Toronto	North Toronto West, Lawrence Park	43.715383	-79.405678
24	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678
25	M6R	West Toronto	Parkdale, Roncesvalles	43.648960	-79.456325
26	M4S	Central Toronto	Davisville	43.704324	-79.388790
27	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049
28	M6S	West Toronto	Runnymede, Swansea	43.651571	-79.484450

	Postal Code	Borough	Neighborhood	Latitude	Longitude
29	M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160
30	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049
31	M4V	Central Toronto	Summerhill West, Rathnelly, South Hill, Forest...	43.686412	-79.400049
32	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420
33	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529
34	M5W	Downtown Toronto	Stn A PO Boxes	43.646435	-79.374846
35	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.667967	-79.367675
36	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280
37	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
38	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.662744	-79.321558

Retreiving the coordinates of Toronto

In [11]:

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

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

In [12]:

```
# create map of Toronto using latitude and longitude values
map_toronto = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, neighborhood in zip(neighborhoods['Latitude'], neighborhoods['Lo
ngitude'], neighborhoods['Borough'], neighborhoods['Neighborhood']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)

map_toronto
```

Out[12]:

Make this Notebook Trusted to load map: File -> Trust Notebook

In [13]:

```
CLIENT_ID = 'GWNRTXUQI4P0SYKJFJOTDFW4J4SNRW4PCAFNGMBNABQLN4HS' # your Foursquare ID
CLIENT_SECRET = 'VCZFCRKIJV4QVTQGUN5L0TVR5X12JTFXLTGUQAZSYJANYLPD' # your Foursquare Se
cret
VERSION = '20180604'
limit = 100
```

In [14]:

```
def getNearbyVenues(names, latitudes, longitudes, radius=500, limit=100):

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

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret
={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            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_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_
list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

In [15]:

```
nearby_venues = getNearbyVenues(names = neighborhoods['Neighborhood'],
                                 latitudes = neighborhoods['Latitude'],
                                 longitudes = neighborhoods['Longitude']
                                 )
print(nearby_venues.shape)
nearby_venues.head()
```

(1613, 7)

Out[15]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot
3	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
4	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa

Venues per each Neighborhood

In [16]:

```
nearby_venues.groupby('Neighborhood').count()
```

Out[16]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	56	56	56	56	56	56
Brockton, Parkdale Village, Exhibition Place	22	22	22	22	22	22
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	18	18	18	18	18	18
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	15	15	15	15	15	15
Central Bay Street	63	63	63	63	63	63
Christie	16	16	16	16	16	16
Church and Wellesley	79	79	79	79	79	79
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	36	36	36	36	36	36
Davisville North	9	9	9	9	9	9
Dufferin, Dovercourt Village	14	14	14	14	14	14
First Canadian Place, Underground city	100	100	100	100	100	100
Forest Hill North & West, Forest Hill Road Park	4	4	4	4	4	4
Garden District, Ryerson	100	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100	100
High Park, The Junction South	23	23	23	23	23	23
India Bazaar, The Beaches West	22	22	22	22	22	22
Kensington Market, Chinatown, Grange Park	58	58	58	58	58	58
Lawrence Park	3	3	3	3	3	3
Little Portugal, Trinity	44	44	44	44	44	44
Moore Park, Summerhill East	3	3	3	3	3	3
North Toronto West, Lawrence Park	21	21	21	21	21	21

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkdale, Roncesvalles	14	14	14	14	14	14
Queen's Park, Ontario Provincial Government	35	35	35	35	35	35
Regent Park, Harbourfront	45	45	45	45	45	45
Richmond, Adelaide, King	93	93	93	93	93	93
Rosedale	4	4	4	4	4	4
Roselawn	3	3	3	3	3	3
Runnymede, Swansea	38	38	38	38	38	38
St. James Town	79	79	79	79	79	79
St. James Town, Cabbagetown	44	44	44	44	44	44
Stn A PO Boxes	93	93	93	93	93	93
Studio District	40	40	40	40	40	40
Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park	16	16	16	16	16	16
The Annex, North Midtown, Yorkville	21	21	21	21	21	21
The Beaches	5	5	5	5	5	5
The Danforth West, Riverdale	42	42	42	42	42	42
Toronto Dominion Centre, Design Exchange	100	100	100	100	100	100
University of Toronto, Harbord	35	35	35	35	35	35

Unique categories curated from all the returned venues

In [17]:

```
print('There are {} uniques categories.'.format(len(nearby_venues['Venue Category'].uni
que())))
```

```
There are 239 uniques categories.
```

Analyze Each Neighborhood

In [18]:

```
# one hot encoding
toronto_onehot = pd.get_dummies(nearby_venues['Venue Category'], prefix="", prefix_sep=
"")

# add neighborhood column back to dataframe
toronto_onehot['Neighborhood'] = nearby_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = toronto_onehot.columns.tolist()
fixed_columns.insert(0, fixed_columns.pop(fixed_columns.index('Neighborhood')))
toronto_onehot = toronto_onehot.reindex(columns = fixed_columns)

print(toronto_onehot.shape)
toronto_onehot.head()
```

```
(1613, 239)
```

Out[18]:

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	Americai Restauran
0	Regent Park, Harbourfront	0	0	0	0	0	0	0	(
1	Regent Park, Harbourfront	0	0	0	0	0	0	0	(
2	Regent Park, Harbourfront	0	0	0	0	0	0	0	(
3	Regent Park, Harbourfront	0	0	0	0	0	0	0	(
4	Regent Park, Harbourfront	0	0	0	0	0	0	0	(

In [19]:

```
toronto_onehot.shape
```

Out[19]:

```
(1613, 239)
```

Top 5 common venues in each neighborhood

In [20]:

```
grouped_toronto_df = toronto_onehot.groupby('Neighborhood').mean().reset_index()
grouped_toronto_df
```

Out[20]:

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	I
0	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1	Brockton, Parkdale Village, Exhibition Place	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
2	Business reply mail Processing Centre, South C...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.066667	0.066667	0.066667	0.066667	0.133333	0.066667	
4	Central Bay Street	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
5	Christie	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
6	Church and Wellesley	0.012658	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
7	Commerce Court, Victoria Hotel	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
8	Davisville	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
9	Davisville North	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
10	Dufferin, Dovercourt Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
11	First Canadian Place, Underground city	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
12	Forest Hill North & West, Forest Hill Road Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
13	Garden District, Ryerson	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
14	Harbourfront East, Union Station, Toronto Islands	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
15	High Park, The Junction South	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
16	India Bazaar, The Beaches West	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal
17	Kensington Market, Chinatown, Grange Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	Lawrence Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	Little Portugal, Trinity	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	Moore Park, Summerhill East	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	North Toronto West, Lawrence Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
22	Parkdale, Roncesvalles	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
23	Queen's Park, Ontario Provincial Government	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
24	Regent Park, Harbourfront	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25	Richmond, Adelaide, King	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
26	Rosedale	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
27	Roselawn	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
28	Runnymede, Swansea	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
29	St. James Town	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
30	St. James Town, Cabbagetown	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
31	Stn A PO Boxes	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
32	Studio District	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
33	Summerhill West, Rathnelly, South Hill, Forest...	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
34	The Annex, North Midtown, Yorkville	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
35	The Beaches	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
36	The Danforth West, Riverdale	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
37	Toronto Dominion Centre, Design Exchange	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	Neighborhood	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	I
38	University of Toronto, Harbord	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

In [21]:

```
grouped_toronto_df.shape
```

Out[21]:

```
(39, 239)
```

In [22]:

```
num_top_venues = 5

for hood in grouped_toronto_df['Neighborhood']:
    print("----"+hood+"----")
    temp = grouped_toronto_df[grouped_toronto_df['Neighborhood'] == hood].T.reset_index
()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top
_venues))
    print('\n')
```

```
----Berczy Park----
                venue  freq
0         Coffee Shop  0.07
1        Cocktail Bar  0.05
2         Cheese Shop  0.04
3  Seafood Restaurant  0.04
4                 Pub  0.04


----Brockton, Parkdale Village, Exhibition Place----
            venue  freq
0            Café  0.14
1  Breakfast Spot  0.09
2     Coffee Shop  0.09
3   Burrito Place  0.05
4    Intersection  0.05


----Business reply mail Processing Centre, South Central Letter Processing
Plant Toronto----
                venue  freq
0  Light Rail Station  0.11
1         Pizza Place  0.06
2             Butcher  0.06
3          Restaurant  0.06
4       Auto Workshop  0.06


----CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst
Quay, South Niagara, Island airport----
             venue  freq
0  Airport Service  0.13
1            Plane  0.07
2          Airport  0.07
3      Coffee Shop  0.07
4              Bar  0.07


----Central Bay Street----
                venue  freq
0         Coffee Shop  0.17
1  Italian Restaurant  0.06
2                Café  0.06
3      Sandwich Place  0.05
4     Bubble Tea Shop  0.03


----Christie----
           venue  freq
0  Grocery Store  0.25
1           Café  0.19
2           Park  0.12
3     Baby Store  0.06
4     Restaurant  0.06


----Church and Wellesley----
                 venue  freq
0          Coffee Shop  0.08
1  Japanese Restaurant  0.06
2     Sushi Restaurant  0.06
```

```
3           Restaurant  0.04
4              Gay Bar  0.04


----Commerce Court, Victoria Hotel----
         venue  freq
0  Coffee Shop  0.11
1         Café  0.07
2   Restaurant  0.07
3        Hotel  0.05
4          Gym  0.04


----Davisville----
            venue  freq
0     Pizza Place  0.14
1  Sandwich Place  0.08
2    Dessert Shop  0.08
3            Café  0.06
4     Coffee Shop  0.06


----Davisville North----
              venue  freq
0    Breakfast Spot  0.11
1              Park  0.11
2  Department Store  0.11
3    Sandwich Place  0.11
4             Hotel  0.11


----Dufferin, Dovercourt Village----
                       venue  freq
0                   Pharmacy  0.14
1                     Bakery  0.14
2                       Pool  0.07
3  Middle Eastern Restaurant  0.07
4                       Bank  0.07


----First Canadian Place, Underground city----
         venue  freq
0  Coffee Shop  0.10
1         Café  0.08
2        Hotel  0.04
3   Restaurant  0.04
4          Gym  0.04


----Forest Hill North & West, Forest Hill Road Park----
                venue  freq
0  Mexican Restaurant  0.25
1       Jewelry Store  0.25
2    Sushi Restaurant  0.25
3               Trail  0.25
4   Afghan Restaurant  0.00


----Garden District, Ryerson----
                       venue  freq
0             Clothing Store  0.09
```

```
1                Coffee Shop  0.08
2  Middle Eastern Restaurant  0.03
3                       Café  0.03
4         Italian Restaurant  0.03


----Harbourfront East, Union Station, Toronto Islands----
         venue  freq
0  Coffee Shop  0.13
1     Aquarium  0.05
2         Café  0.04
3        Hotel  0.04
4      Brewery  0.03


----High Park, The Junction South----
                venue  freq
0                Café  0.09
1  Mexican Restaurant  0.09
2     Thai Restaurant  0.09
3         Flea Market  0.04
4  Italian Restaurant  0.04


----India Bazaar, The Beaches West----
                  venue  freq
0                  Park  0.09
1        Sandwich Place  0.09
2  Fast Food Restaurant  0.09
3     Food & Drink Shop  0.05
4      Sushi Restaurant  0.05


----Kensington Market, Chinatown, Grange Park----
                   venue  freq
0                   Café  0.09
1            Coffee Shop  0.07
2     Mexican Restaurant  0.05
3  Vietnamese Restaurant  0.05
4                 Bakery  0.05


----Lawrence Park----
               venue  freq
0               Park  0.33
1        Swim School  0.33
2           Bus Line  0.33
3  Afghan Restaurant  0.00
4      Movie Theater  0.00


----Little Portugal, Trinity----
                           venue  freq
0                            Bar  0.11
1               Asian Restaurant  0.07
2                     Restaurant  0.07
3  Vegetarian / Vegan Restaurant  0.05
4                    Coffee Shop  0.05


----Moore Park, Summerhill East----
```

```
              venue  freq
0       Tennis Court  0.33
1                Gym  0.33
2              Trail  0.33
3             Museum  0.00
4  Martial Arts Dojo  0.00


----North Toronto West,  Lawrence Park----
                 venue  freq
0       Clothing Store  0.10
1          Coffee Shop  0.10
2        Metro Station  0.05
3  Fast Food Restaurant  0.05
4                  Spa  0.05


----Parkdale, Roncesvalles----
                        venue  freq
0              Breakfast Spot  0.14
1                   Gift Shop  0.14
2                     Dog Run  0.07
3                Dessert Shop  0.07
4  Eastern European Restaurant  0.07


----Queen's Park, Ontario Provincial Government----
               venue  freq
0        Coffee Shop  0.23
1    Sushi Restaurant  0.06
2        Yoga Studio  0.03
3            Theater  0.03
4  College Auditorium  0.03


----Regent Park, Harbourfront----
         venue  freq
0  Coffee Shop  0.18
1          Pub  0.07
2       Bakery  0.07
3         Park  0.07
4      Theater  0.04


----Richmond, Adelaide, King----
           venue  freq
0    Coffee Shop  0.11
1           Café  0.05
2     Restaurant  0.04
3          Hotel  0.03
4  Deli / Bodega  0.03


----Rosedale----
              venue  freq
0              Park  0.50
1        Playground  0.25
2             Trail  0.25
3            Museum  0.00
4  Martial Arts Dojo  0.00
```

```
----Roselawn----
                     venue  freq
0              Music Venue  0.33
1                   Garden  0.33
2             Home Service  0.33
3  Mediterranean Restaurant  0.00
4              Men's Store  0.00


----Runnymede, Swansea----
              venue  freq
0              Café  0.08
1       Coffee Shop  0.08
2  Sushi Restaurant  0.05
3       Pizza Place  0.05
4               Pub  0.05


----St. James Town----
                 venue  freq
0          Coffee Shop  0.06
1                 Café  0.06
2         Cocktail Bar  0.05
3            Gastropub  0.04
4  American Restaurant  0.04


----St. James Town, Cabbagetown----
                venue  freq
0         Coffee Shop  0.09
1         Pizza Place  0.05
2                 Pub  0.05
3  Italian Restaurant  0.05
4          Restaurant  0.05


----Stn A PO Boxes----
                 venue  freq
0          Coffee Shop  0.10
1                 Café  0.04
2             Beer Bar  0.03
3   Seafood Restaurant  0.03
4  Japanese Restaurant  0.03


----Studio District----
         venue  freq
0         Café  0.10
1  Coffee Shop  0.08
2    Gastropub  0.05
3       Bakery  0.05
4      Brewery  0.05


----Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park----
          venue  freq
0           Pub  0.12
1   Coffee Shop  0.12
2  Liquor Store  0.06
3    Bagel Shop  0.06
```

```
4    Restaurant  0.06


----The Annex, North Midtown, Yorkville----
               venue  freq
0               Café  0.14
1     Sandwich Place  0.14
2        Coffee Shop  0.10
3           Pharmacy  0.05
4  Indian Restaurant  0.05


----The Beaches----
               venue  freq
0              Trail   0.2
1  Health Food Store   0.2
2                Pub   0.2
3        Coffee Shop   0.2
4  Afghan Restaurant   0.0


----The Danforth West, Riverdale----
                    venue  freq
0        Greek Restaurant  0.21
1      Italian Restaurant  0.07
2             Coffee Shop  0.07
3  Furniture / Home Store  0.05
4          Ice Cream Shop  0.05


----Toronto Dominion Centre, Design Exchange----
           venue  freq
0    Coffee Shop  0.10
1          Hotel  0.06
2           Café  0.06
3     Restaurant  0.04
4  Deli / Bodega  0.03


----University of Toronto, Harbord----
                 venue  freq
0                 Café  0.17
1                  Bar  0.06
2           Restaurant  0.06
3            Bookstore  0.06
4  Japanese Restaurant  0.06
```

Inserting the top 10 venues into a dataframe

In [23]:

```python
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]
```

In [24]:

```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = grouped_toronto_df['Neighborhood']

for ind in np.arange(grouped_toronto_df.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(grouped_toron
to_df.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[24]:

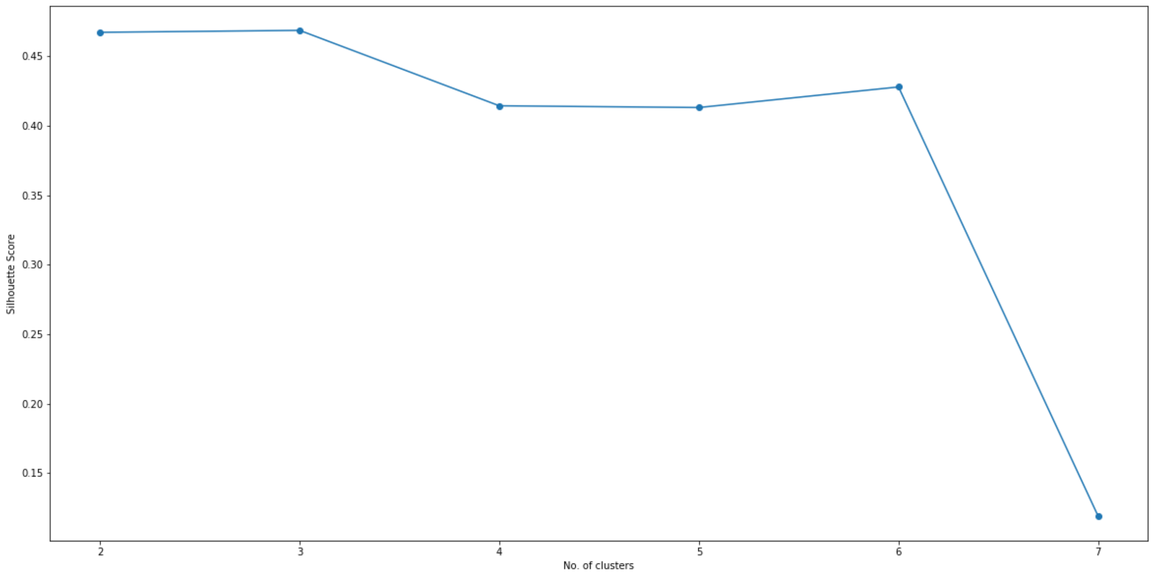
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
0	Berczy Park	Coffee Shop	Cocktail Bar	Pub	Café	Bakery	Restaurant	Cheese Shop
1	Brockton, Parkdale Village, Exhibition Place	Café	Coffee Shop	Breakfast Spot	Grocery Store	Bakery	Performing Arts Venue	Pet Store
2	Business reply mail Processing Centre, South C...	Light Rail Station	Auto Workshop	Park	Pizza Place	Recording Studio	Restaurant	Butcher
3	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Harbor / Marina	Bar	Plane	Coffee Shop	Rental Car Location	Sculpture Garden
4	Central Bay Street	Coffee Shop	Italian Restaurant	Café	Sandwich Place	Burger Joint	Japanese Restaurant	Department Store

Cluster Neighborhoods

In [26]:

```
max_range = 8

toronto_grouped_clustering = grouped_toronto_df.drop('Neighborhood', 1)

from sklearn.metrics import silhouette_samples, silhouette_score

indices = []
scores = []

for kclusters in range(2, max_range) :

    # Run k-means clustering
    n_clusters = kclusters
    kmc = toronto_grouped_clustering
    kmeans = KMeans(n_clusters = kclusters, init = 'k-means++', random_state = 0).fit_p
redict(kmc)

    # Gets the score for the clustering operation performed
    score = silhouette_score(kmc, kmeans)

    # Appending the index and score to the respective lists
    indices.append(kclusters)
    scores.append(score)
    print ("For n_clusters = {}, silhouette score is {})".format(n_clusters, score))
plot(max_range, scores, "No. of clusters", "Silhouette Score")
```

```
For n_clusters = 2, silhouette score is 0.46730366957828845)
For n_clusters = 3, silhouette score is 0.46871205819908196)
For n_clusters = 4, silhouette score is 0.41445561687931315)
For n_clusters = 5, silhouette score is 0.4132031209543001)
For n_clusters = 6, silhouette score is 0.42799408206832257)
For n_clusters = 7, silhouette score is 0.11914821395717534)
```



In [27]:

```
toronto_grouped_clustering = grouped_toronto_df.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=3, random_state=0).fit(toronto_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Out[27]:

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

In [28]:

```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

toronto_merged = neighborhoods

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborho
od
toronto_merged = toronto_merged.join(neighborhoods_venues_sorted.set_index('Neighborhoo
d'), on='Neighborhood')

toronto_merged.head() # check the last columns!
```

Out[28]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3 C
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	0	Coffee Shop	Park	
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0	Coffee Shop	Sushi Restaurant	
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	0	Clothing Store	Coffee Shop	Re
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	0	Coffee Shop	Café	
4	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Pub	Health Food Store	

In [29]:

```
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(toronto_merged['Latitude'], toronto_merged['Longitude'], toronto_merged['Neighborhood'], toronto_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[29]:

Make this Notebook Trusted to load map: File -> Trust Notebook

Examine Clusters

Cluster 1

In [30]:

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0, toronto_merged.columns[[1] +
list(range(5, toronto_merged.shape[1]))]]
```

Out[30]:

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Mos Commo Venu
0	Downtown Toronto	0	Coffee Shop	Park	Bakery	Pub	Theater	Breakfas Spc
1	Downtown Toronto	0	Coffee Shop	Sushi Restaurant	Yoga Studio	Creperie	Beer Bar	Smoothi Sho
2	Downtown Toronto	0	Clothing Store	Coffee Shop	Middle Eastern Restaurant	Japanese Restaurant	Italian Restaurant	Cosmetic Sho
3	Downtown Toronto	0	Coffee Shop	Café	Cocktail Bar	Gastropub	American Restaurant	Restaurar
4	East Toronto	0	Pub	Health Food Store	Coffee Shop	Trail	Yoga Studio	Discour Stor
5	Downtown Toronto	0	Coffee Shop	Cocktail Bar	Pub	Café	Bakery	Restaurar
6	Downtown Toronto	0	Coffee Shop	Italian Restaurant	Café	Sandwich Place	Burger Joint	Japanes Restaurar
7	Downtown Toronto	0	Grocery Store	Café	Park	Restaurant	Italian Restaurant	Baby Stor
8	Downtown Toronto	0	Coffee Shop	Café	Restaurant	Clothing Store	Hotel	Gyr
9	West Toronto	0	Pharmacy	Bakery	Grocery Store	Music Venue	Pool	Middl Easter Restaurar
10	Downtown Toronto	0	Coffee Shop	Aquarium	Café	Hotel	Italian Restaurant	Frie Chicke Joir
11	West Toronto	0	Bar	Restaurant	Asian Restaurant	Men's Store	Café	Vegetarian Vega Restaurar
12	East Toronto	0	Greek Restaurant	Italian Restaurant	Coffee Shop	Furniture / Home Store	Restaurant	Ice Crear Sho
13	Downtown Toronto	0	Coffee Shop	Café	Hotel	Restaurant	American Restaurant	Seafoo Restaurar
14	West Toronto	0	Café	Coffee Shop	Breakfast Spot	Grocery Store	Bakery	Performin Arts Venu
15	East Toronto	0	Park	Sandwich Place	Fast Food Restaurant	Pet Store	Pub	Burrito Plac
16	Downtown Toronto	0	Coffee Shop	Restaurant	Café	Hotel	American Restaurant	Gyr
17	East Toronto	0	Café	Coffee Shop	Brewery	Gastropub	Bakery	America Restaurar
20	Central Toronto	0	Gym	Hotel	Pizza Place	Department Store	Sandwich Place	Breakfas Spc
21	Central Toronto	0	Jewelry Store	Trail	Mexican Restaurant	Sushi Restaurant	Yoga Studio	Discour Stor
22	West Toronto	0	Café	Mexican Restaurant	Thai Restaurant	Grocery Store	Furniture / Home Store	Fast Foo Restaurar

	Borough	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Mos Commo Venu
23	Central Toronto	0	Clothing Store	Coffee Shop	Yoga Studio	Sporting Goods Shop	Salon / Barbershop	Caf
24	Central Toronto	0	Café	Sandwich Place	Coffee Shop	BBQ Joint	Liquor Store	Pu
25	West Toronto	0	Breakfast Spot	Gift Shop	Dessert Shop	Movie Theater	Eastern European Restaurant	Dog Ru
26	Central Toronto	0	Pizza Place	Sandwich Place	Dessert Shop	Sushi Restaurant	Gym	Italia Restaurar
27	Downtown Toronto	0	Café	Bakery	Restaurant	Italian Restaurant	Japanese Restaurant	Ba
28	West Toronto	0	Café	Coffee Shop	Diner	Pizza Place	Pub	Tea Roor
29	Central Toronto	0	Gym	Trail	Tennis Court	Dessert Shop	Falafel Restaurant	Event Spac
30	Downtown Toronto	0	Café	Coffee Shop	Mexican Restaurant	Vietnamese Restaurant	Bakery	Ba
31	Central Toronto	0	Coffee Shop	Pub	Bank	Restaurant	Liquor Store	Supermarke
32	Downtown Toronto	0	Airport Service	Harbor / Marina	Bar	Plane	Coffee Shop	Rental Ca Locatio
34	Downtown Toronto	0	Coffee Shop	Café	Restaurant	Pub	Seafood Restaurant	Japanes Restaurar
35	Downtown Toronto	0	Coffee Shop	Italian Restaurant	Pizza Place	Pub	Café	Baker
36	Downtown Toronto	0	Coffee Shop	Café	Japanese Restaurant	Gym	Hotel	Restaurar
37	Downtown Toronto	0	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Restaurant	Gay Bar	Caf
38	East Toronto	0	Light Rail Station	Auto Workshop	Park	Pizza Place	Recording Studio	Restaurar

Cluster 2

In [31]:

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1, toronto_merged.columns[[2] +
list(range(6, toronto_merged.shape[1]))]]
```

Out[31]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
18	Lawrence Park	Park	Swim School	Bus Line	Yoga Studio	Diner	Falafel Restaurant	Event Space
33	Rosedale	Park	Playground	Trail	Yoga Studio	Donut Shop	Diner	Discount Store

Cluster 3

In [32]:

```
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2, toronto_merged.columns[[2] +
list(range(6, toronto_merged.shape[1]))]]
```

Out[32]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
19	Roselawn	Garden	Home Service	Music Venue	Dessert Shop	Falafel Restaurant	Event Space	Ethiopian Restaurant

Results

The results of the methodology that we have implemented through K-means shows that the neighborhoods can be clustered into 3 clusters and most of neighboorhoods which are concentrated towards the first cluster.

As we recall the factors that we are basing our solution upon they clearly indicate that cluster 1 is the most effective group for us to take advantage of the features offred by these neighboorhoods.

Population Density - The population density is comparably higher than the other neighboorhoods and this can help us in the scalability of the restaurant. Due to the high population and the multicultural demographics which are represented by these neighborhoods, A new restaurant with a unique cuisine is well suited in an urban and diverse demographic rather than a homogenous ethnic population.

Cuisine and Exclusivity - A restuarant which offers a Canadian Chinese cuisine is well suited in neighboorhoods where there are no other competitors as the cuisine is unique in it's type and the popularity of the most common venues can help us in bringing some new to the market with high customer base.

Recommendations

The recommendation which can be deduced is that a neighborhood which is part of Cluster 1 is the goto option for our client as the restaruant is offering a unique cuisine into a market where there aren't many competitors and the customer base of the cluster 1 neighborhoods help in creating a effective feedback system for improving the business model of the restaurant.

Conclusion

This project has analyzed the business concerns that restaurant owners have when trying to venture into a new area where the sustainability of business has to evaluated properly before venturing into an area without prior footprint.

This mechanism of levaraging the data and use of inferential statistical methods helps us in evaluating these situations effectively enough to predict the right approach while establishing a business venture.

In [ ]: