Assignment Battle of Neighborhoods:

Opening a Korean Restaurant either in Toronto or in New York City?

Problem Description

At a food fair in Seoul/Korea a chef who runs a restaurant chain, famous for homemade food style throughout the country, has received an offer from an investor to open a restaurant either in Toronto or in NYC.

The main target group for the restaurant are Koreans have been living in Toronto and NYC and missing the traditional homemade food style. There are Korean restaurants available in both cities but customized to the overall populations eating habits. For sure all other people are highly welcome too.

The outcome of this project should give to the investor and chef a suggestion about the right choice of the city between Toronto and NYC, furthermore the right choice for Neighborhood and restaurant class in order to get the aimed business case.

The goal of this project is not to grant the success of the restaurant, but to support the strategy deployment based on the ratio no. of restaurant to Koreans, being in the centroid among all existing Korean restaurants and deciding the restaurant class based on the prosperity of the selected Neighborhood. The Strategy itself is under the responsibility of the project clients namely investor and chef.

Describing the Data

- The ratio of number of restaurants per Korean will be taken into account for the city decision
    - Number of restaurants from Foursquare
    - Population of Korean community in Toronto and NYC from Wikipedia

Approach:

Population of Koreans in Toronto and NYC received from Wikipedia tables based on Census surveys.

- Raw data for NYC in Wikipedia (based on US Census 2015):https://en.wikipedia.org/wiki/List_of_U.S._cities_with_significant_Korean-American_populations

	Rank	City	State	Korean-Americans 2010	Percentage 2010	Korean-Americans 2015	Percentage 2015
0	1	Los Angeles	California	108282	2.9%	110679	2.8%
1	2	Honolulu	Hawaii	22179	2.3% [19]	20729	2.1%
2	3	Anaheim	California	6575	2.0%	6696	1.9%
3	4	San Jose	California	12409	1.3%	12939	1.3%
4	5	New York City	New York	96741	1.2%[7]	91729	1.1%
5	6	Aurora	Colorado	3646	1.2%	3379	1.0%
6	7	Anchorage	Alaska	3251	1.2%	3799	1.3%

- Raw data for Toronto in Wikipedia (based on Toronto Census 2016): https://en.wikipedia.org/wiki/Demographics_of_Toronto

	Ethnic groups in the Toronto CMA (2016)Source: Focus on Geography Series, 2016 Census; Toronto, (CMA) - Ontario	Ethnic groups in the Toronto CMA (2016)Source: Focus on Geography Series, 2016 Census; Toronto, (CMA) - Ontario.1	Population	%
0	Ethnic group	White	2804630	47.8
1	Ethnic group	South Asian	973225	16.6
2	Ethnic group	Chinese	631045	10.8
3	Ethnic group	Black	442020	7.5
4	Ethnic group	Filipino	254480	4.3
5	Ethnic group	Latin American	132950	2.3
6	Ethnic group	West Asian	123755	2.1
7	Ethnic group	Arab	105610	1.8
8	Ethnic group	Southeast Asian	83540	1.4
9	Ethnic group	Korean	69670	1.2
10	Ethnic group	Japanese	20650	0.4
11	Ethnic group	Multiple minorities	97185	1.7

Number of Korean Restaurants:

- **Toronto**:

Postcode table which includes Boroughs and Neighborhoods from Wikipedia (https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&diff=942851379&oldid=942655599) preprocessed and merged with geo data table for Toronto (http://cocl.us/Geospatial_data).

```
#sorting of column order
Toronto_geo=df_merge_Toronto[['Postcode','Borough','Neighbourhood','Latitude','Longitude']]
Toronto_geo.head()
```

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

The merged data frame used as input to request venue data from Foursquare.

The output data frame from Foursquare filtered to Korean Restaurant and counted the number of these.

```
Filter_Korean_Restaurant_Tr=toronto_venues[(toronto_venues['Venue Category']=='Korean Restaurant')]
Filter_Korean_Restaurant_Tr
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
12	Woburn	43.770992	-79.216917	Korean Grill House	43.770812	-79.214502	Korean Restaurant
879	St. James Town	43.651494	-79.375418	Bob's Bulgogi	43.651908	-79.377523	Korean Restaurant
1062	Central Bay Street	43.657952	-79.387383	Mo'Ramyun	43.656148	-79.392282	Korean Restaurant
1901	Little Portugal, Trinity	43.647927	-79.419750	OddSeoul	43.646192	-79.419601	Korean Restaurant

```
Num_KoreanRest_Toronto=Filter_Korean_Restaurant_Tr['Venue Category'].count()
Num_KoreanRest_Toronto
```

```
4
```

```
print("Number of Korean Restaurants in Tororntо listed in Foursquare:",Num_KoreanRest_Toronto)
```

```
Number of Korean Restaurants in Tororntо listed in Foursquare: 4
```

- **New York City**

Geo data for NYC Neighborhoods downloaded from source (json file `newyork_data.json from https://cocl.us/new_york_dataset`) given in the Coursera LAB: Segmenting and Clustering Neighborhoods in NYC

After preprocessing the raw data from json file the data frame for NYC Neighborhoods geo data was obtained.

```
NYC_geo=neighborhoods
NYC_geo.head(15)
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756

This data frame used as input to request venue data from Foursquare.

The data frame received from Foursquare filtered to Korean Restaurant and counted the number of these.

```
NYC_venues.head()
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

```
Filter_Korean_Restaurant_NY=NYC_venues[(NYC_venues['Venue Category']=='Korean Restaurant')]
Filter_Korean_Restaurant_NY
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1358	Brighton Beach	40.576825	-73.965094	Cafe At Your Mother-in-Law	40.577575	-73.963449	Korean Restaurant
1570	Prospect Heights	40.676822	-73.964859	Kimchi Grill	40.675130	-73.963245	Korean Restaurant
1661	Williamsburg	40.707144	-73.958115	Dotory	40.707730	-73.955779	Korean Restaurant
1695	Bushwick	40.698116	-73.925258	Kichin	40.697706	-73.927023	Korean Restaurant
2305	Park Slope	40.672321	-73.977050	Hanki Everyday Korean	40.670208	-73.978944	Korean Restaurant
2307	Park Slope	40.672321	-73.977050	Haenyeo	40.675163	-73.981100	Korean Restaurant

```
Num_KoreanRest_NYC=Filter_Korean_Restaurant_NY['Venue Category'].count()
Num_KoreanRest_NYC
```

```
68
```

```
print("Number of Korean Restaurants in NYC listed in Foursquare:",Num_KoreanRest_NYC)
```

```
Number of Korean Restaurants in NYC listed in Foursquare: 68
```

- **Decision of the target city:**

After having obtained the number of Korean restaurants in Toronto and NYC, the ratio per Korean population was calculated.

```
#Ratio of Korean Restaurant per Korean in Toronto
KR_Rest_Ratio_Toronto=Num_KoreanRest_Toronto/KRpopToronto
KR_Rest_Ratio_Toronto
```

```
5.741352088416822e-05
```

```
#Ratio of Korean Restaurant per Korean in NYC
KR_Rest_Ratio_NYC=Num_KoreanRest_NYC/KRpopNYC
KR_Rest_Ratio_NYC
```

```
0.0007413140882381799
```

Thru a simple if - else function the target city was decided as Toronto.

Selection of the City based on the ratio of Number of KR restaurants per Korean

```
if KR_Rest_Ratio_NYC > KR_Rest_Ratio_Toronto:
    print ('Location for the new Korean Restaurant should be in Toronto')
else:
    print ('Location for the new Korean Restaurant should be in New York City')
```

```
Location for the new Korean Restaurant should be in Toronto
```

Target Neighborhood for the new restaurant in Toronto:

The inventor and the chef of the restaurant have given the goal to settle the new restaurant as the geographical centroid among the existing Korean restaurants.

Reason for that is, that the distance to the new restaurant should be reasonable to the potential guests who are acquainted to visit existing Korean restaurant.

- **Calculation of the centroid:**

  With the location data (lat, lng) of the existing 4 restaurants listed in Foursquare the centroid was calculation with below function.

	latitude	longitude
12	43.770812	-79.214502
879	43.651908	-79.377523
1062	43.656148	-79.392282
1901	43.646192	-79.419601

```
import math

x = 0.0
y = 0.0
z = 0.0

for i, coord in KR_Rest_geo_Tr.iterrows():
    latitude = math.radians(coord.latitude)
    longitude = math.radians(coord.longitude)

    x += math.cos(latitude) * math.cos(longitude)
    y += math.cos(latitude) * math.sin(longitude)
    z += math.sin(latitude)

total = len(KR_Rest_geo_Tr)

x = x / total
y = y / total
z = z / total

central_longitude = math.atan2(y, x)
central_square_root = math.sqrt(x * x + y * y)
central_latitude = math.atan2(z, central_square_root)

mean_location = {
    'latitude': math.degrees(central_latitude),
    'longitude': math.degrees(central_longitude)
    }
```

```
mean_location
```

```
{'latitude': 43.68129289571633, 'longitude': -79.35104579977836}
```

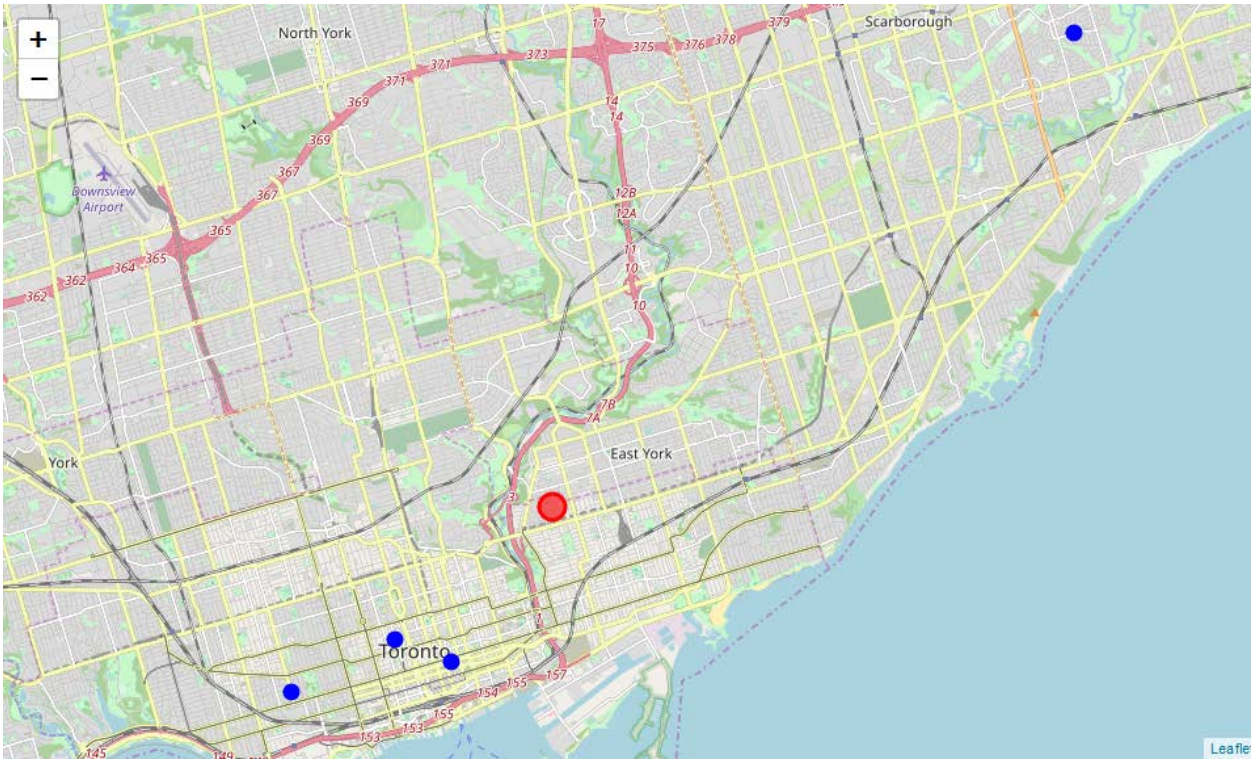
With the location data of the centroid the Neighborhood was determined by use of Geopy

```
#location for new Korean restaurant as ceontroid of exisiting Korean restaurants
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="Toronto_new_KR_Rest")
location = geolocator.reverse("43.68129289571633, -79.35104579977836")
print('Location for new Korean restaurant in Toronto as centroid of existing Korean restaurant will be in:',location.address)
```

```
Location for new Korean restaurant in Toronto as centroid of existing Korean restaurant will be in: 1002, Logan Avenue,
```

➔ Location for new Korean restaurant in Toronto as centroid of existing Korean restaurant will be in: 1002, Logan Avenue, Greektown, East York, Toronto–Danforth, Old Toronto, Toronto, Golden Horseshoe, Ontario, M4K 1P3, Canada

The existing Korean restaurants (blue bubbles) according to Foursquare and the place for the new Korean restaurant (red bubble) were visualized in a folium map:



The final question is, whether the restaurant style should be more in a simple and affordable style or rather in a high class style?

Decision of the restaurant class

The investor and the chef want to decide about the restaurant class (menu, price strategy, interior, size etc) based on the average income of the selected Neighborhood relative to the other Neighborhoods of Toronto.

In order to answer this question the average income of each Neighborhood in Toronto were evaluated and compared. The date were received from the Demographics table available in Wikipedia (https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods) that is based on Canadian Census.

```
#download demographics table for Toronto from wikipedia
source_demographics_Toronto=pd.read_html("https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods")
```

```
avrg_income_Toronto=source_demographics_Toronto[1]
```

```
avrg_income_Toronto=avrg_income_Toronto[['Name','Average Income']]
avrg_income_Toronto.head()
```

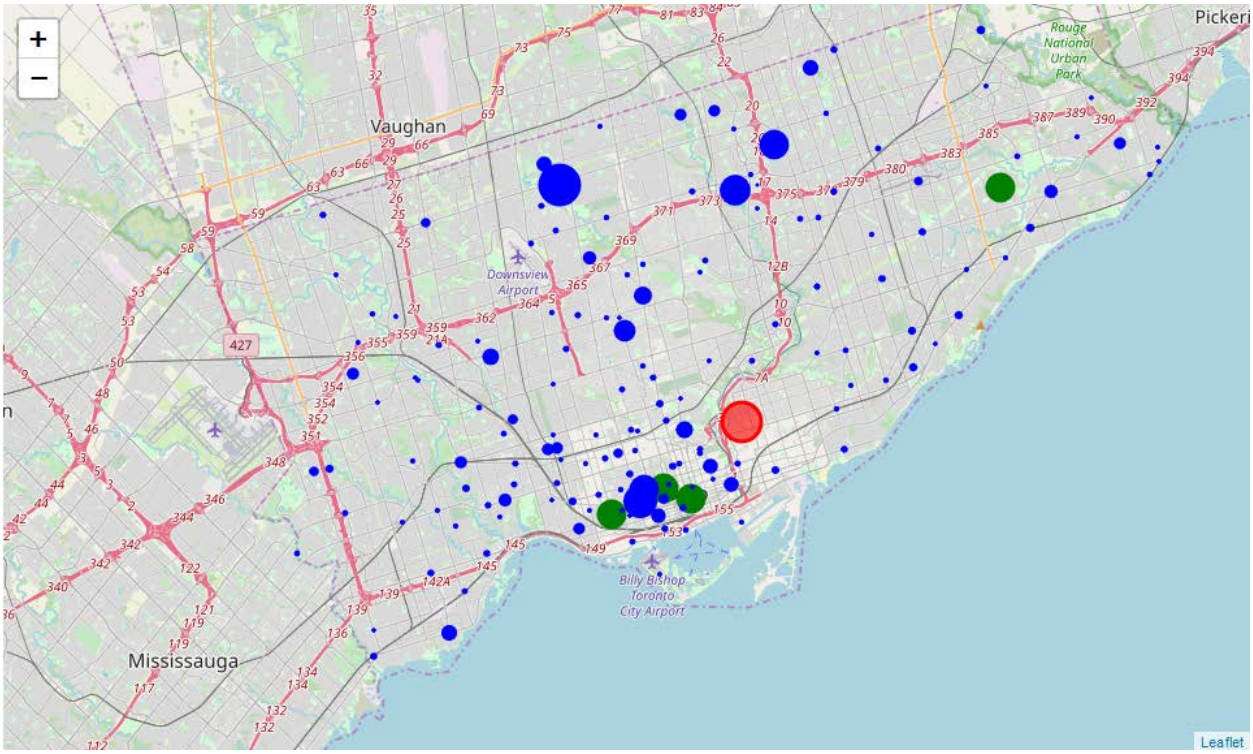
	Name	Average Income
0	Toronto CMA Average	40704
1	Agincourt	25750
2	Alderwood	35239
3	Alexandra Park	19687
4	Allenby	245592

The average income values were preprocessed to a range of 0 to 1. The scaled average income values are used for bubble size to visualize these in a folium map over the neighborhoods of Toronto.

```
#scale average income to range of 0 to 1 for radius of bubble in folium map
from sklearn.preprocessing import MinMaxScaler
data = avrg_income_Tr_geo2[['Average Income']].values
scaler = MinMaxScaler(feature_range=(0, 1), copy=True)
scaler.fit(data)
scaler.fit_transform(data)
output=scaler.transform(data)
```

```
df_avrginc_sc = pd.DataFrame({'scaled_avrginc': output[:, 0]})
df_avrginc_sc.head()
```

	scaled_avrginc
0	0.021145
1	0.053356
2	0.000564
3	0.767419



The blue bubbles indicate the average income per Neighborhood, the green bubbles the place of the existing Korean restaurant listed in Foursquare and the red one the target place for the new Korean restaurant.

Describing the Methodology:

For obtaining data from Wikipedia e.g. for number of Korean population and average income the scraping function of pandas pd.read_html was used. Other functions as from beautifulSoup could be used though. However the use of the pandas function is quite simply to my perspective.

As can be seen, the Census Data for number of Korean population and average incomes are from few years back. Due to the fact the decision process is depending on relative considerations and e.g. the difference of the number of Korean population between NYC and Toronto is quite high, it does not play a big role at the end.

In order to obtain the number of Korean restaurant in NYC and Toronto the venue data supplier Foursquare was used. Other service supplier as e.g. google, Tripadvisor and further could be used as well.

The calculation of the geographical centroid of the four Korean restaurants from Foursqaure in Toronto a formula from stackoverflow was used. By visualization of the location of the four restaurants and the calculated centroid a plausibility check was done in a folium map.

In order to decide about the restaurant class the average income of the target neighborhood was to be compared relatively with that of the other neighborhoods. The average income of each Toronto neighborhood was obtained from Wikipedia based on Canada Census. In order to visualize the average income of each neighborhood in a folium map, the local data (lat, lng) was to be obtained. This was done by use of geolocator function of Geopy.

The average income values were preprocessed to a range of 0 to 1 by use of preprocessing MinMaxScaler function of the machine learning library scikit learn. This was done in order to visualize the average income as the bubble size in a folium map.

Results

Due the difference of ratio of Korean restaurant to Korean population in NYC and Toronto is significant, the decision for Toronto is quite clear.

The plausibility check of the calculated centroid by visual check in a folium map shows that calculation method worked fine. However, the latitude and longitude as geo data for the centroid does not mean that restaurant has to be exactly at this point. It shall define the Neighborhood where the centroid is. In this case it is the Neighboorhood ‘East-York’.

The foil map with bubbles for average income shows that wealth in Toronto is concentrated in a few neighborhoods. The rest of the neighborhoods seem to be below the overall Toronto average.

Conclusion

This report shall only give an idea how such a decision process by means of data analysis could look like. The considered parameters are by far not enough for a thorough evaluation and final Decision.

Here are few more parameters recommended to take into account to the decision process. These will be reported to the clients.

Such as…

- looking to different database for number of Korean restaurants in addition to Foursquare. The 4 in Foursquare seem to be too less taking into account the population. In google and Tripadvisor are plenty of more Korean restaurants listed.
- for decision of the suitable neighborhood not only the centroid but also other facts to be considered e.g. rental costs, vacant space for restaurant, public transportation and further
- eating habits and preferences of Koreans and others, but also a comparison of spent budget for eating outside between NYC and Toronto etc.
- not only the average income of the neighborhoods to be considered for restaurant class but also that of the Koreans because they are the declared target group and for sure further facts such as the restaurant classes in the surrounding area and that of the existing Korean restaurants in the city.

It is obvious not every needed information are freely available in the internet. It will be told to the clients that e.g. a dedicated survey has to be made in order to obtain the information required for the entire decision process.