The Battle of Neighbourhoods

Open a Mediterranean food restaurant in Toronto

1. Introduction | Business Proposal

Growing immigration from many countries around the world to Canada in the last decade has increased globalization in different cities. Specifically, in the city of Toronto, restaurants of different nationalities have opened, including Chinese, Japanese, Indian, among others.

The objective of this project is to find out if among all these gastronomies of the world, there is a place in Toronto for the famous "Mediterranean gastronomy". It is well known that Mediterranean gastronomy is one of the healthiest in the world because it uses among its main ingredients fresh and quality products, such as vegetables, fruits, virgin olive oil, etc.

Thanks to this project, we will be able to know if opening a Mediterranean cuisine restaurant in Toronto is a viable project or not. On the other hand, the target audience will be all people who have an exquisite palate and want to taste the flavors of Mediterranean cuisine.

2. Data description

We will use the dataset obtained in week. This contains the latitudes, longitudes and zip codes of Canada. This dataset can be found in: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Taking into account the needs of our project, determining factors to establish a restaurant of Mediterranean gastronomy will be:

- location of the neighborhood in the city of Toronto,
- Mediterranean restaurants in the neighborhood, in case there are any.

Foursquare API Data

To get the data for the different neighborhoods of Toronto in Canada, we will use the Foursquare API. As in week 3 of the course, we will get location information of the venues.

The data obtained from Foursquare are: Neighborhood, Neighborhood latitude, Neighborhood Longitude, Venue, Name of the Venue, Venue latitude, Venue longitude, and Venue category.

Libraries:

- Pandas: to create and manipulate dataframes.
- Scikit learn: to use k-means clustering.
- Numpy: to support the creation of multidimensional arrays and vectors.
- Matplotlib: to create plots.
- Geocoder: to retrieve location data.
- Folium: map rendering library.
- JSON: to handle JSON files.
- Requests: to handle requests.

Unsupervised machine learning

For this project we will use the unsupervised learning algorithm k-means to segment and cluster the data and thus obtain enough information to know in which place is more appropriate to open a restaurant of Mediterranean gastronomy.

3. Methodology

Importing required libraries

```
import pandas as pd
import numpy as np
pd.set_option('display.max_columns', 300)
pd.set_option('display.max_rows', 300)

import json # library to handle JSON files

import geocoder
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values

import requests # library to handle requests
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 matplotlib.pyplot as plt

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

import folium # map rendering library

print('Libraries imported.')
```

Reading the data and importing it into the dataframe (Data Collection)

```
df_raw = pd.read_html("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M")[0]
df_raw # The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood
```

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

```
# Printing summary of the df_raw
df_raw.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 3 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Postal Code    180 non-null    object
 1   Borough        180 non-null    object
 2   Neighbourhood  180 non-null    object
dtypes: object(3)
memory usage: 4.3+ KB
```

Only process the cells that have an assigned borough. Ignore cells with a borough that is not assigned.

```
df = df_raw[df_raw['Borough']!='Not assigned']
df
```

	Postal Code	Borough	Neighbourhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

More than one neighborhood can exist in one postal code area. For example, in the table on the Wikipedia page, you will notice that M5A is listed twice and has two neighborhoods: Harbourfront and Regent Park. These two rows will be combined into one row with the neighborhoods separated with a comma.

```
df_pc = df.groupby('Postal Code', sort = False).agg(', '.join)
df_pc.head(10)
```

	Borough	Neighbourhood
Postal Code		
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park, Harbourfront
M6A	North York	Lawrence Manor, Lawrence Heights
M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

If a cell has a borough but a not assigned neighborhood, then the neighborhood will be the same as the borough.

```
df_pc.loc[df_pc['Neighbourhood'] == 'Not assigned', 'Neighbourhood'] = df_pc.loc[df_pc['Neighbourhood'] == 'Not assigned', 'Borou
df_pc.reset_index(inplace = True)
df_pc.head(10)
```

	Postal Code	Borough	Neighbourhood
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

Reading geospatial data

```
df_geo = pd.read_csv('http://cocl.us/Geospatial_data')
df_geo
```

	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

Merging both dfs by Postal Code

```
df_pc = pd.merge(df_pc, df_geo, on = 'Postal Code')
df_pc
```

	Postal Code	Borough	Neighbourhood	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

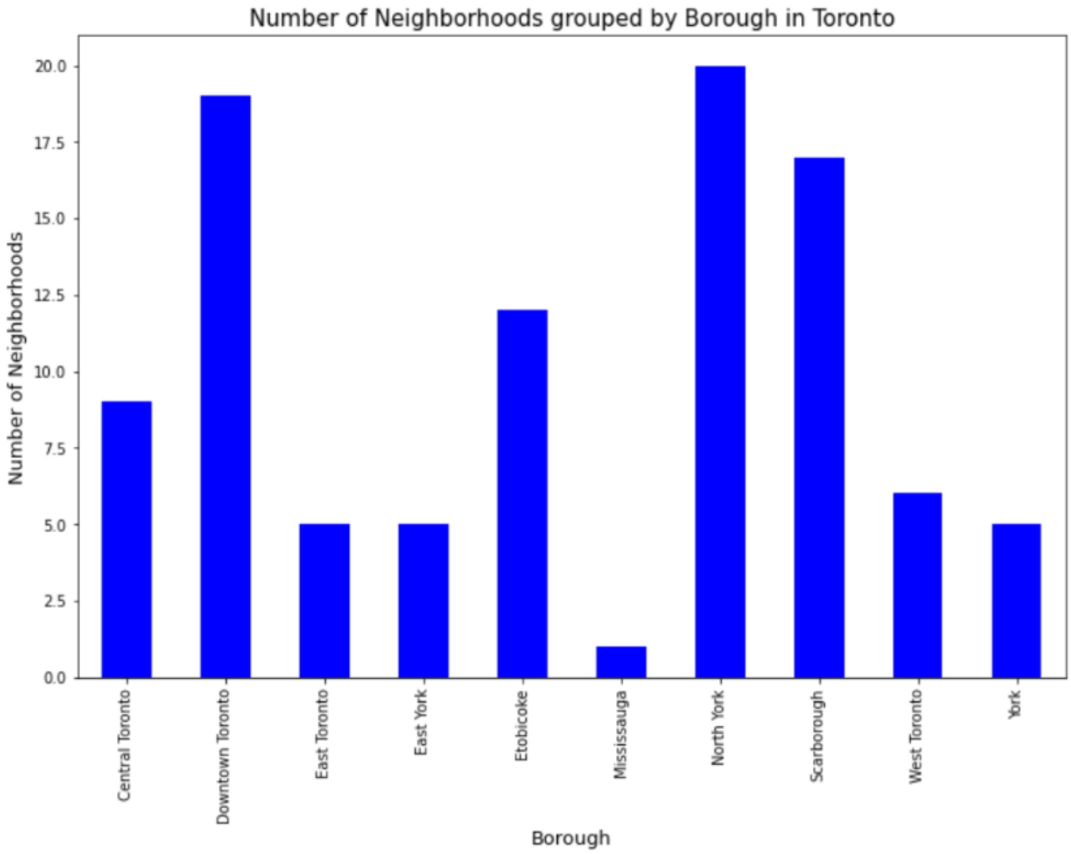
Data Visualization

Exploratory Data Analysis to understand better how many neighborhoods are.

```
neigh_tor = df_pc.groupby('Borough')['Neighbourhood'].nunique()
neigh_tor
```

```
Borough
Central Toronto      9
Downtown Toronto    19
East Toronto         5
East York            5
Etobicoke           12
Mississauga          1
North York          20
Scarborough         17
West Toronto         6
York                 5
Name: Neighbourhood, dtype: int64
```

Plotting the data



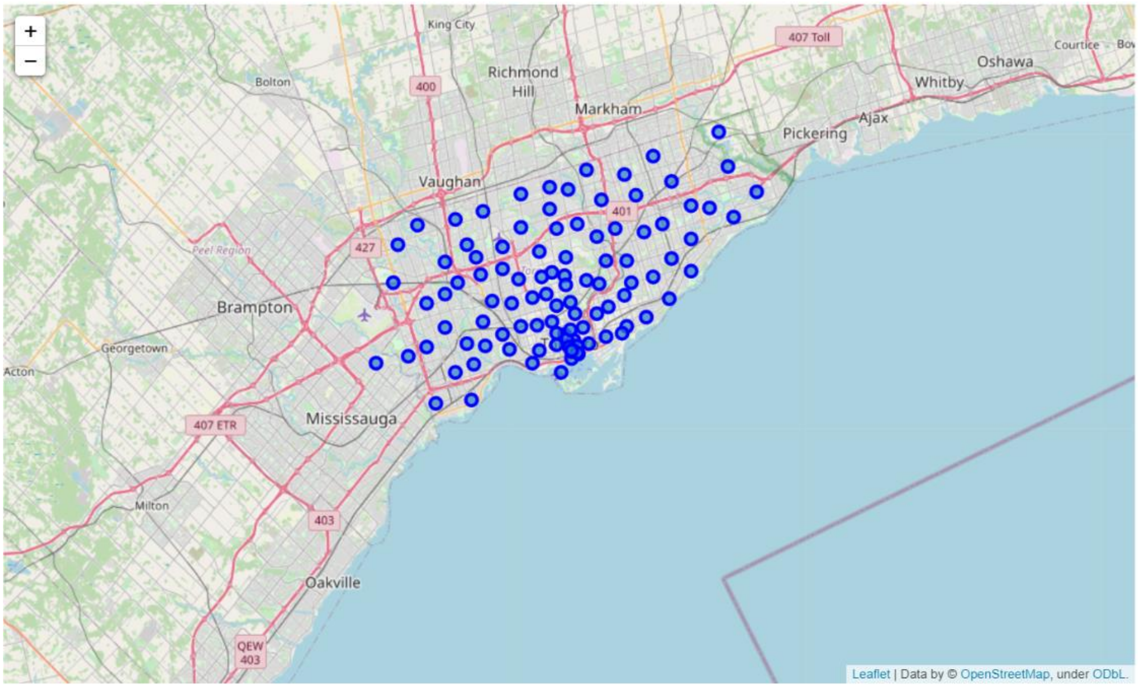
North York is the borough with the greatest number of neighborhoods. Downtown Toronto is the borough with the second largest number of neighborhoods.

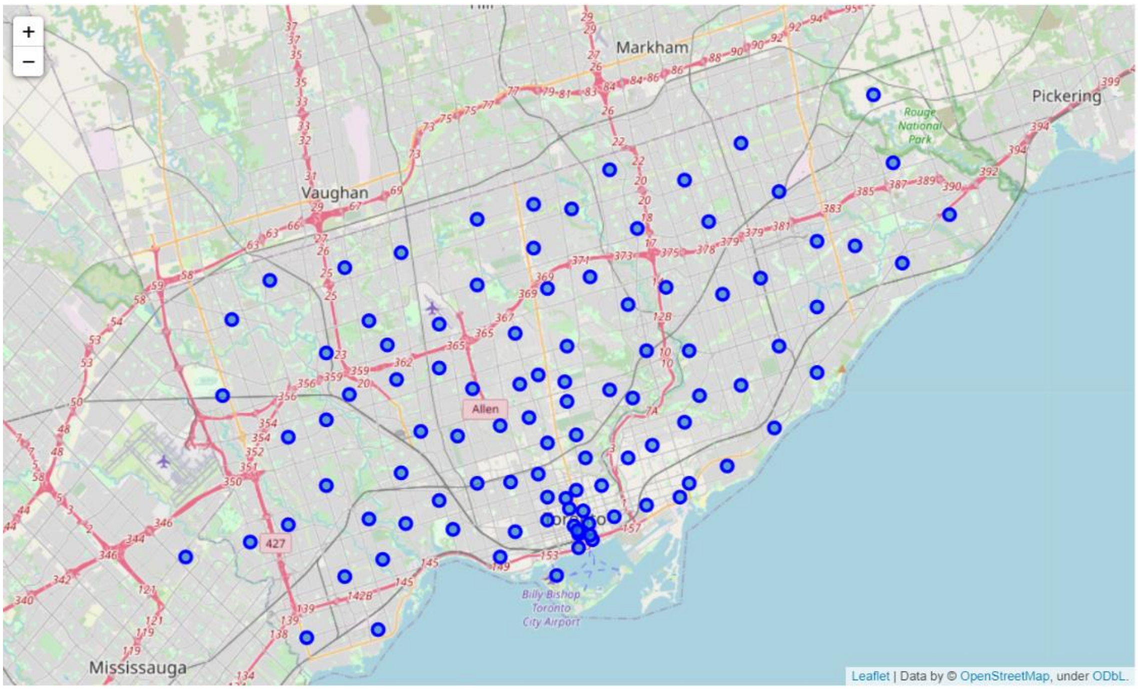
Using Foursquare API

```
address = 'Toronto, Canada'

geolocator = Nominatim(user_agent="canada_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.
```

Create a map of Toronto with neighborhoods superimposed on top





Setting Foursquare Credentials

```
CLIENT_ID = 'NGLJLENJFP4RXWM4X04A2A0DH3OGECSQDTY1NPNU3G2IVZHC' # your Foursquare ID
CLIENT_SECRET = 'JV4YR0MZIQKGUQ1DLCPN02SDQEY1K45TIWBLRP445DQTG20K' # your Foursquare Secret
VERSION = '20210202' # Foursquare API version
LIMIT = 100 # A default Foursquare API limit value

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: NGLJLENJFP4RXWM4X04A2A0DH3OGECSQDTY1NPNU3G2IVZHC
CLIENT_SECRET:JV4YR0MZIQKGUQ1DLCPN02SDQEY1K45TIWBLRP445DQTG20K
```

Explore Neighborhoods in Toronto

Printing 10 rows of toronto_venues.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Brookbanks Pool	43.751389	-79.332184	Pool
2	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
5	Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
6	Victoria Village	43.725882	-79.315572	Pizza Nova	43.725824	-79.312860	Pizza Place
7	Regent Park, Harbourfront	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
8	Regent Park, Harbourfront	43.654260	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
9	Regent Park, Harbourfront	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center

Let's check how many venues were returned for each neighborhood.

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

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Agincourt	4	4	4	4	4	4
Alderwood, Long Branch	7	7	7	7	7	7
Bathurst Manor, Wilson Heights, Downsview North	23	23	23	23	23	23
Bayview Village	4	4	4	4	4	4
Bedford Park, Lawrence Manor East	24	24	24	24	24	24
Berczy Park	55	55	55	55	55	55
Birch Cliff, Cliffside West	4	4	4	4	4	4

Analyze Each Neighborhood

Let's check for Mediterranean restaurants.

```
toronto_onehot[toronto_onehot['Mediterranean Restaurant'] == 1].sum()
```

```
Light Rail Station          0
Lingerie Store              0
Liquor Store                0
Lounge                      0
Luggage Store               0
Market                      0
Martial Arts School         0
Massage Studio              0
Medical Center              0
Mediterranean Restaurant    6
```

There are 6 Mediterranean restaurants. Sounds interesting.

Grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each category

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

	Neighborhood	Yoga Studio	Accessories Store	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Art Gallery	Art Museum
0	Agincourt	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
1	Alderwood, Long Branch	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
2	Bathurst Manor, Wilson Heights, Downsview North	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
3	Bayview Village	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000
4	Bedford Park, Lawrence Manor East	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.041667	0.000000	0.00	0.000000	0.000000

Printing each neighborhood along with the top 5 most common venues

```
num_top_venues = 5

for hood in toronto_grouped['Neighborhood']:
    print("----"+hood+"----")
    temp = toronto_grouped[toronto_grouped['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')
```

```
----Agincourt----
                      venue  freq
0        Chinese Restaurant  0.25
1                    Lounge  0.25
2            Breakfast Spot  0.25
3  Latin American Restaurant  0.25
4               Yoga Studio  0.00
```

Inserting that into a pandas *dataframe*

```
# Sorting the venues in descending order.

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]
```

```
# Creating the new dataframe and display the top 10 venues for each neighborhood.

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))
```

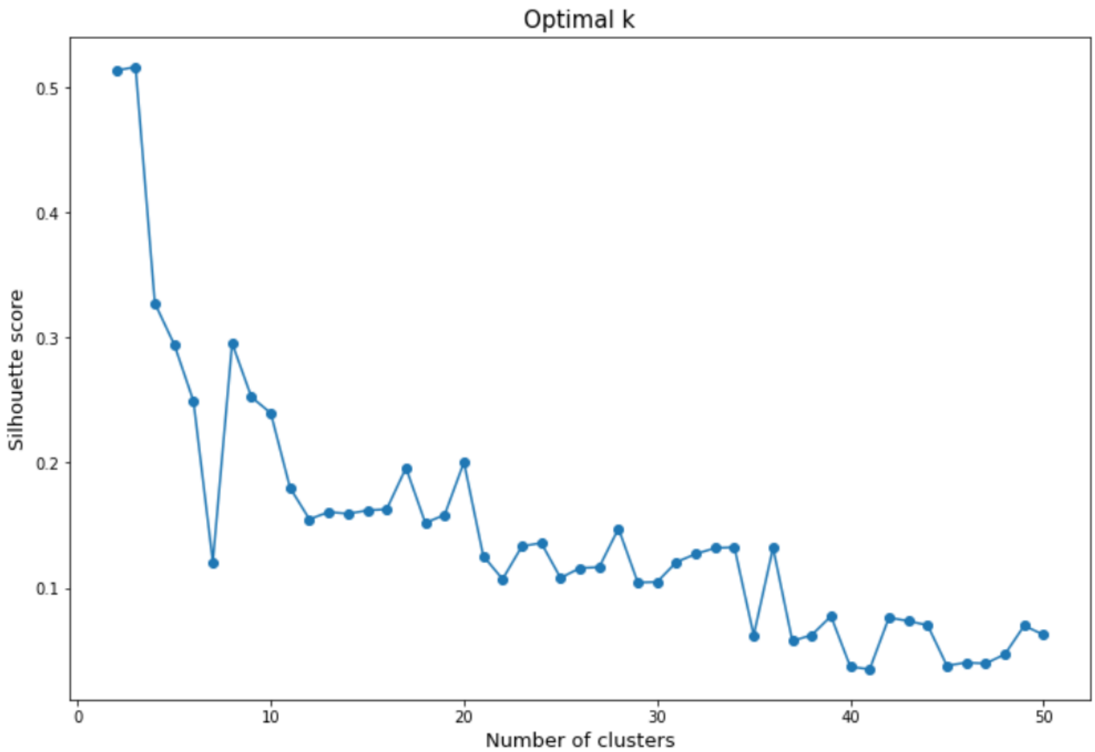
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Chinese Restaurant	Lounge	Breakfast Spot	Latin American Restaurant	Yoga Studio	Motel	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant	Modern European Restaurant
1	Alderwood, Long Branch	Pizza Place	Pharmacy	Pub	Coffee Shop	Skating Rink	Gym	Mediterranean Restaurant	Men's Store	Mexican Restaurant	Movie Theater
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Pizza Place	Diner	Middle Eastern Restaurant	Shopping Mall	Pharmacy	Mobile Phone Shop	Sandwich Place	Fried Chicken Joint
3	Bayview Village	Café	Japanese Restaurant	Bank	Chinese Restaurant	Music Venue	Museum	Movie Theater	Motel	Moroccan Restaurant	Monument / Landmark

Cluster Neighborhoods

The first step to use k-means clustering is to determine the number of clusters to use.

```
from sklearn.metrics import silhouette_score
k = 50
k_values = []
sc = []

# kmeans++ for a better initialization of centroids
for k in range(2, k+1):
    df_cl = toronto_grouped.drop('Neighborhood', axis = 1)
    kmeans = KMeans(n_clusters = k, init = 'k-means++', random_state = 40).fit_predict(df_cl)

    score = silhouette_score(df_cl, kmeans, metric = 'euclidean', random_state = 0)
    k_values.append(k)
    sc.append(score)
```

Plotting Silhouette score with different number of clusters




```
print(k_values[0:10])
print(sc[0:10])
```

```
[2, 3, 4, 5, 6, 7, 8, 9, 10, 11]
[0.5137198031502376, 0.5166286648657628, 0.3275463434614149, 0.29474050061096707, 0.24927807804855537, 0.12048578322839436, 0.2
9557129636684626, 0.2525273926619521, 0.23975826603652947, 0.17989451390694086]
```

Taking into account the results of the silhouette score, we will use 3 clusters for our clustering model as this is the best result obtained.

```
# Run k-means to cluster the neighborhood into 5 clusters.

# set number of clusters
kclusters = 3

toronto_grouped_clustering = toronto_grouped.drop('Neighborhood', axis = 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_
```

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1,
       1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 2, 1, 1, 0, 1, 1, 1, 0,
       1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       0, 1, 1, 1, 1, 1, 1, 0])
```

Creating a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

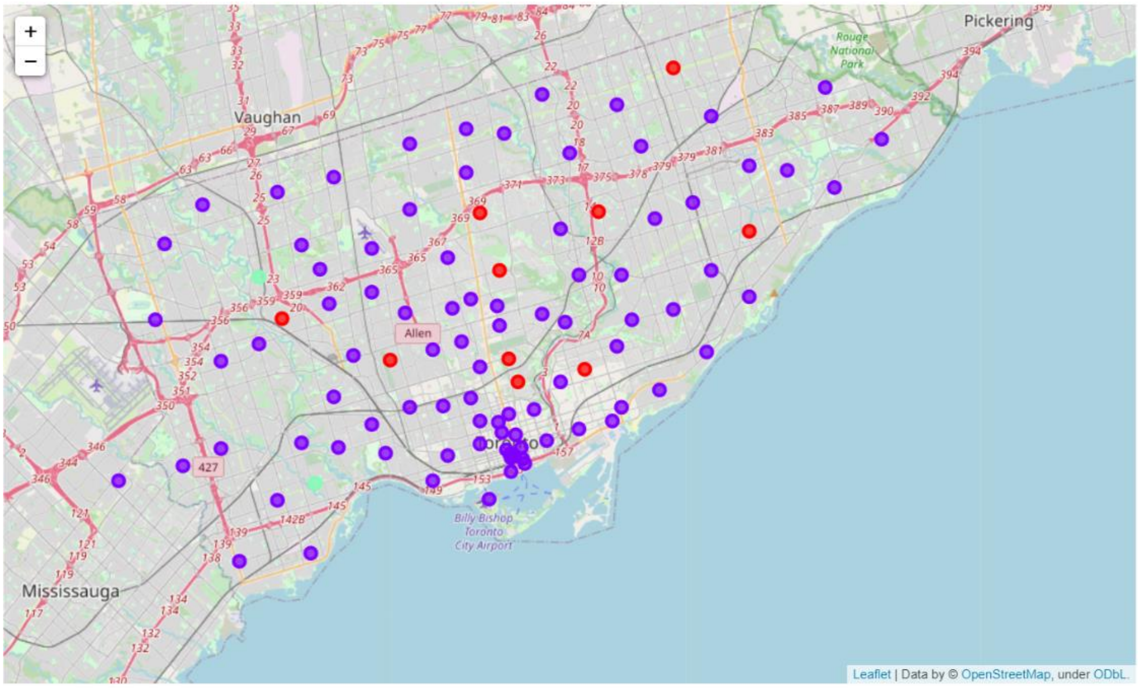
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

toronto_merged = df_pc

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
toronto_merged = toronto_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighbourhood')
```

Drop NAN values

```
toronto_merged_clean = toronto_merged.dropna(subset=['Cluster Labels'])
toronto_merged_clean
```

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	M3A	North York	Parkwoods	43.753259	-79.329656	0.0	Food & Drink Shop	Pool	Park	Yoga Studio	Moroccan Restaurant	Monument / Landmark
1	M4A	North York	Victoria Village	43.725882	-79.315572	1.0	Pizza Place	Portuguese Restaurant	Coffee Shop	Hockey Arena	Yoga Studio	Modern European Restaurant
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636	1.0	Coffee Shop	Bakery	Park	Pub	Breakfast Spot	Theater
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763	1.0	Clothing Store	Furniture / Home Store	Women's Store	Coffee Shop	Boutique	Gift Shop
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial	43.662301	-79.389494	1.0	Coffee Shop	Sushi Restaurant	Yoga Studio	Beer Bar	Restaurant	Burrito Place

Finally, let's visualize the resulting clusters



Examine Clusters

Cluster 1

```
toronto_merged_clean[toronto_merged_clean['Cluster Labels'] == 0].nunique()
```

```
Postal Code               10
Borough                    6
Neighbourhood             10
Latitude                  10
Longitude                 10
Cluster Labels             1
1st Most Common Venue      6
2nd Most Common Venue      8
3rd Most Common Venue      7
4th Most Common Venue      6
5th Most Common Venue      5
6th Most Common Venue      5
7th Most Common Venue      6
8th Most Common Venue      6
9th Most Common Venue      6
10th Most Common Venue     7
dtype: int64
```

Borough	Neighbourhood	Cluster Labels	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	8th Most Common Venue	9th Most Common Venue
North York	Parkwoods	0.0	Food & Drink Shop	Pool	Park	Yoga Studio	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant	Modern European Restaurant	Mobile Phone Shop
York	Caledonia-Fairbanks	0.0	Park	Women's Store	Pool	Miscellaneous Shop	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant	Modern European Restaurant	Mobile Phone Shop
Scarborough	Scarborough Village	0.0	Playground	Convenience Store	Medical Center	Mediterranean Restaurant	Men's Store	Mexican Restaurant	Middle Eastern Restaurant	Miscellaneous Shop	Music Venue
East York	East Toronto, Broadview North (Old East York)	0.0	Park	Intersection	Convenience Store	Yoga Studio	Mobile Phone Shop	Motel	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant
Central Toronto	Lawrence Park	0.0	Bus Line	Swim School	Park	Yoga Studio	Mobile Phone Shop	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant	Modern European Restaurant

Cluster 2

```
toronto_merged_clean[toronto_merged_clean['Cluster Labels'] == 1].nunique()
```

```
Postal Code               88
Borough                   10
Neighbourhood             84
Latitude                  88
Longitude                 67
Cluster Labels             1
1st Most Common Venue     34
2nd Most Common Venue     52
3rd Most Common Venue     52
4th Most Common Venue     54
5th Most Common Venue     45
6th Most Common Venue     55
7th Most Common Venue     50
8th Most Common Venue     47
9th Most Common Venue     46
10th Most Common Venue    50
dtype: int64
```

Borough	Neighbourhood	Cluster Labels	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	8th Most Common Venue
North York	Victoria Village	1.0	Pizza Place	Portuguese Restaurant	Coffee Shop	Hockey Arena	Yoga Studio	Modern European Restaurant	Movie Theater	Motel
Downtown Toronto	Regent Park, Harbourfront	1.0	Coffee Shop	Bakery	Park	Pub	Breakfast Spot	Theater	Café	Farmers Market
North York	Lawrence Manor, Lawrence Heights	1.0	Clothing Store	Furniture / Home Store	Women's Store	Coffee Shop	Boutique	Gift Shop	Miscellaneous Shop	Vietnamese Restaurant
Downtown Toronto	Queen's Park, Ontario Provincial Government	1.0	Coffee Shop	Sushi Restaurant	Yoga Studio	Beer Bar	Restaurant	Burrito Place	Bar	Nightclub
Scarborough	Malvern, Rouge	1.0	Fast Food Restaurant	Print Shop	Miscellaneous Shop	Motel	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant	Modern European Restaurant

Cluster 3

```
toronto_merged_clean.loc[toronto_merged_clean['Cluster Labels'] == 2, toronto_merged_clean.columns[[1] + [2] + list(range(5, tor
```

	Borough	Neighbourhood	Cluster Labels	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
57	North York	Humberlea, Emery	2.0	Baseball Field	Yoga Studio	Mobile Phone Shop	Motel	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant	Modern European Restaurant	Miscellaneous Shop	Market
101	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	2.0	Baseball Field	Yoga Studio	Mobile Phone Shop	Motel	Moroccan Restaurant	Monument / Landmark	Molecular Gastronomy Restaurant	Modern European Restaurant	Miscellaneous Shop	Market

4. Results

The results show 3 clusters in the city of Toronto in Canada. We have searched for the 10 most common venues for each of the clusters obtained. It can be observed that some clusters are more suitable for having a restaurant. The results show that for cluster 1 there are 9 observations. These observations belong to 6 boroughs and we see that in none of the 3 most common venues a restaurant appears.

Cluster 2 is the cluster with the highest number of observations and therefore also has the highest number of restaurants among the 10 most common venues. Finally, cluster 3 has the lowest number of observations. Therefore, the number of restaurants is also much lower compared to the other two clusters. However, the top 2 venues show that baseball field and yoga studio are the most common.

5. Discussion

Analyzing the 3 clusters we can conclude that clusters 1 and 2 are the most appropriate when opening a restaurant, specifically, in our case, a Mediterranean food restaurant.

For cluster 1, we observed Mediterranean food restaurants in the boroughs of Scarborough and Central Toronto, so establishing our restaurant in those boroughs would be the most appropriate for our business. In this cluster, up to the most common 5 venues, there are almost no restaurants, this could be due to the fact that people are more sporty as can be seen in different venues, such as: Park, tennis court, pool, and yoga studio.

Cluster 2 is the cluster that contains the most neighborhoods in the city of Toronto, with a total of 84 neighborhoods. In particular, the Hillcrest Village neighborhood, belonging to the borough of North York, has Mediterranean food restaurants in the third place. Therefore, we can think that the people of that neighborhood have a good reception to restaurants of this style. It is worth mentioning that the boroughs of Etobicoke, Scarborough, and York would also be good options to open the restaurant as they also have Mediterranean cuisine.

Finally, cluster 3 would not be an option when it comes to opening our restaurant because as seen in these neighborhoods, they are more inclined to healthy living as the most common venues. Although it seems that Moroccan food has also had a good acceptance.

6. Conclusion

We have successfully analyzed the neighborhoods of Toronto, Canada, to determine which would be the most suitable neighborhood to open our Mediterranean cuisine restaurant. Taking into account the analysis and the results obtained, the most propitious neighborhood is Hillcrest Village, located in the borough of North York, since it has a restaurant of this style among the 3 most common venues.

On the other hand, as a future work, it would be good to know other factors such as associated costs or transportation. These were not considered because they were out of scope.