1.   INTRODUCTION

A major coffee shop chain ("Company"), which is currently operating in Athens has been partially acquired (80% of the shares) from a Canadian Group ("Group"). The coffee shop chain has more than 100 shops operating successfully in Greece. According to information provided by the Company, key factor for the success of its retail shops, is the "ideal" location selection of its shops.

One of its most profitable coffee shop (highest sales, gross profit and net profit margins) is operating in the Neighbourhood of Kolonaki in the centre of Athens.

The Group is considering starting to expand its businesses abroad. In specific the management of the Group is considering opening coffee shops in USA & Canada. For that purpose, they need to commit and "risk" substantial capital expenditures and opex.

Before starting to commit any funds for expanding its businesses abroad, the management of the Group has decided to start with a "pilot project" (opening only one shop in one country).

The main purpose of the pilot project is to figure out if the successful "Greek coffee shop model" could be replicated in Canada and the USA, without any major adjustments on the marketing "mix" & policy or alternatively if the existing business & marketing model must be adjusted substantially in order to penetrate the foreign markets.

As a first step the Group has decided that it will be more efficient to start with Canada/Toronto since the Group is based in Canada and it has sufficient local know how & it would be more cost effective to start the "pilot project" in Canada instead of USA .

The second Step is to identify the characteristics/venue categories of the Greek Neighbourhood , where the Company's existing most successful shop is located, i.e Kolonaki, Greece.

The next step is to identify one or several Boroughs/Neighbourhoods of Toronto Centre, with similar venue categories & characteristics to Kolonaki Athens. By identifying the common venues of Kolonaki-Athens and Toronto Centre, the Company will be able to open a shop in Toronto, which has similar venues categories to Athens Centre.

For preparing the above the Company will use the services of a Data scientist.

After the Data scientist suggest to the Company the relevant Boroughs/Neighbourhoods of Toronto Centre, which are similar to Kolonaki Athens, the Company will use its own additional "selection criteria ", which are currently not disclosed to the Data scientist, in order to select the final Borough/Neighbourhood , where the new coffee shop will operate.

Identifying the Neighbourhoods of Toronto Central which share similar characteristics to Athens ,   is critical for the following indicative reasons :

1. Company success is strongly correlated with proper "location selection"
2. In case of failure of the "new " Toronto coffee shop to achieve sales target, such failure will not primarily be attributed to "wrong" location selection, since both Athens shop and Canadian shop will have similar venue categories. Indicatively the Company may attribute a potential failure to "price policy" of the new shop or "low acceptance" of the Toronto retail clients for the specific coffee taste or ineffective business promotion, or lack of brand awareness etc.

Depending on the success or not of the new pilot shop the Company will decide if expansion with the existing business model is making economic sense or if further adjustments of the business model are necessary.

For executing the above steps the Company is asking for the professional services of a Data Scientist

The data scientist has decided to utilize Foursquare in order to explore the venue categories near the Athens shop. Then the Data scientist will use public available information from several sources (see next section) in order to find the Boroughs and Neighbourhoods in Toronto Center.

In addition the Data scientist will utilize again Foursquare in order to find venue categories for each Borough and Neighbourhood in Toronto centre.

Finally he/she will create a cluster (using KMeans from sklearn cluster) of Boroughs and Neighbourhoods in Toronto, which share similar venue categories to Athens Greece and he/she will deliver the findings of his/her work to the management of the Company.

2. DATA (Source)

For the specific project assigned by the Group we need the following data:

- Coordinates of the existing location in Kolonaki Athens and Coordinates of Toronto Centre. In that respect we will use from "geopy.geocoders" , "Nominatim", which converts an address into latitude and longitude values :

```
# Lets now find the coordinates of Clients existing business in Athens, Patriarchou Ioakim street.

address = 'patriarchou ioakim , Athens'
geolocator = Nominatim()
location = geolocator.geocode(address)
kolonaki_latitude = location.latitude
kolonaki_longitude = location.longitude
print('The geograpical coordinate of {}  Athens-Kolonaki  home are {}, {}.'.format(address, kolonaki_latitude, kolonaki_longitud
print('The exact location of our address is :  {}.'.format(location))
```

- Venues for Athens and Toronto city near our targeted Borough/neighbourhood. In that respect we will utilize "Foursquare":

```
# Access the Foursquare website by using ID and s

CLIENT_ID = 'GN1PKJNKKDJYVYW4V2CDOWTQWBW0TAE5GJJ3
CLIENT_SECRET = 'XGU2XZYWMZS4T2ZXQERKZWN3TT0BUQ1F
VERSION = '20190330' # Foursquare API version
```

- We will use Beautiful soup in order to get from Wikipedia postal code, Neighbourhood and Borough of Toronto :

```
[22]: # target url for getting the postcodes of Canada
      url = 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
      url
[22]: 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'

[23]: # Get access to wikipedia through Beautifulsoup
      html_content = requests.get(url).text
      soup = BeautifulSoup(html_content, "lxml")

[24]: # print soup in order to find table
      #print(soup.prettify())

[25]: # find and print table using soup
      table = soup.find('table',class_='wikitable sortable')
      # table
```

- Using wget we download from cocl.us /Geospatial data the coordinates of Toronto and the relevant postal code :

```
n [39]: # Download coordinates and postal codes for Toronto from web using wget

        import wget
        url_get = wget.download( 'http://cocl.us/Geospatial_data/toronto_coordinates.csv')
        coordinates = pd.read_csv(url_get)
        coordinates.head()
```

	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

- Using Foursqure and with some data adjustments we get a list of Boroughs & neighbourhoods with the most common venues in Toronto Centre:

	PostalCode	Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
27	M5V	DowntownToronto	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Service	Airport Lounge	Airport Terminal	Boutique	Plane
32	M6J	WestToronto	Little Portugal,Trinity	Bar	Men's Store	Coffee Shop	Restaurant	Asian Restaurant
26	M5T	DowntownToronto	Chinatown,Grange Park,Kensington Market	Café	Chinese Restaurant	Vietnamese Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant
3	M4M	EastToronto	Studio District	Café	Coffee Shop	Brewery	Gastropub	Bakery
34	M6P	WestToronto	High Park,The Junction South	Café	Mexican Restaurant	Bar	Thai Restaurant	Diner

3. **Methodology (for all the below steps you can get a more detailed description and access to the python code using the following link:** https://github.com/stelios72/week-5-battle-of-Neighbourhood/blob/488370a866474e0238ab04c6aaaebd67053deba1/capstone%20week%205%20final%20.ipynb

As a first step we import the below libraries:

- numpy, library to handle data in a vectorized manner
- pandas , library for data analysis
- json_normalize , transform JSON file into a pandas dataframe
- json , library to handle JSON files
- requests, library to handle requests
- geopy.geocoders, Nominatim , convert an address into latitude and longitude values
- folium, map rendering library
- Matplotlib and associated plotting modules
- k-means from clustering stage
- wget , download utility
- Beautifulsoup , for getting data from Wikipedia (our case)

We then get the Coordinates of the existing coffee shop located in Kolonaki Athens using "geopy.geocoders" , "Nominatim" . Based on those we get all venues near our existing shop utilizing Foursquare. In our case we have limited the number of venues to 200 and we have used a radius of 1km. We then group the results in order to get the frequency of the number of categories per venue as per below table:

	name
categories	
American Restaurant	1
Art Museum	1
Bar	4
Bistro	1
Bookstore	2
Boutique	5
Cafeteria	1
Café	11
Cheese Shop	1
Chocolate Shop	1
Clothing Store	1
Cocktail Bar	2
Coffee Shop	3
Cosmetics Shop	1
Cupcake Shop	2
Deli / Bodega	1
Dessert Shop	5
Electronics Store	1
French Restaurant	2
Furniture / Home Store	1
Gastropub	1
Gourmet Shop	3
Greek Restaurant	4
Grilled Meat Restaurant	1
Gym / Fitness Center	2
Hill	1
History Museum	2
Hotel	3
Hotel Bar	1
Italian Restaurant	4
Kafenio	1
Mediterranean Restaurant	3
Modern European Restaurant	1
Movie Theater	1
Museum	2
Nightclub	2
Pastry Shop	1
Peruvian Restaurant	1
Playground	1
Plaza	2
Restaurant	2
Sandwich Place	1

Using folium, we will also map the results of venues near our existing shop in Athens as following:
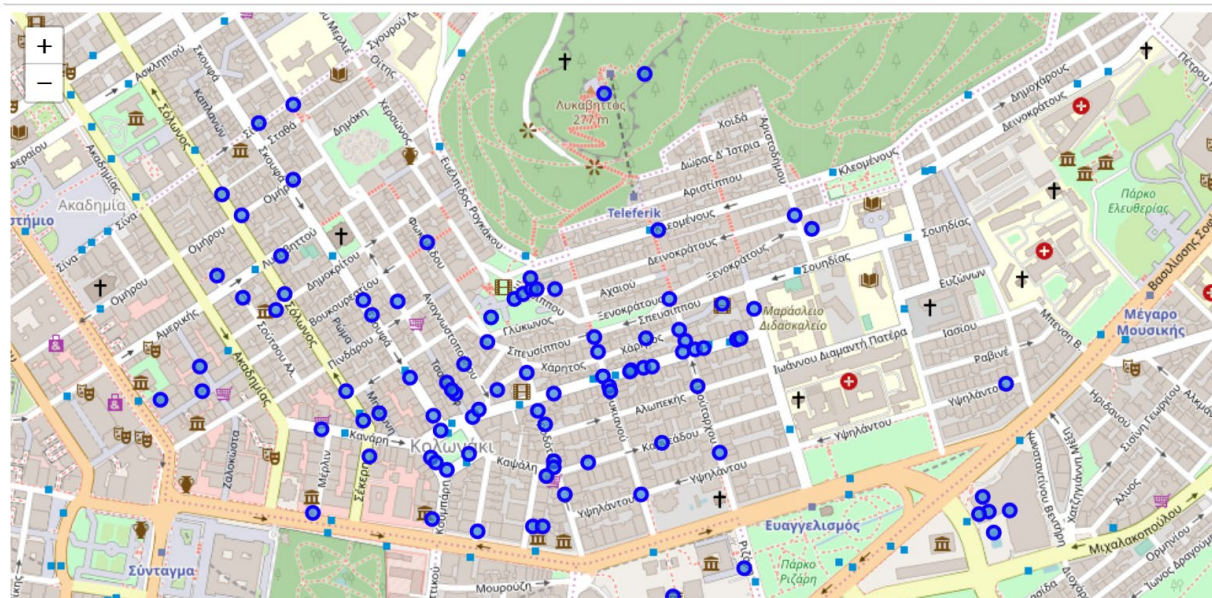
From the above table and map we can easily conclude that our existing most successful coffee shop in Athens is located in an area, where there are many coffee shops/ cafeterias and restaurants. It seems that the specific Neighbourhood is a "busy" area of the centre of Athens. We can also assume that the area where the shop of our client is located, has many visitors per day (due to high number of cafeterias and restaurants) , so it is important that the "new" shop in Toronto, will be located in an area which shares similar characteristics to Athens-kolonaki.

Our task as a Data scientist is to find Boroughs/Neighbourhoods in Toronto centre, where they appear to have similarities to Athens-Kolonaki.

In order to find such a Neighbourhood in Toronto centre we start getting some information from Wikipedia.

In specific we can get the postal code, Neighbourhoods and Boroughs of Toronto using BeautifulSoup:

```
[22]: # target url for getting the postcodes of Canada
url = 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
url
[22]: 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
[23]: # Get access to wikipedia through Beautifulsoup
html_content = requests.get(url).text
soup = BeautifulSoup(html_content, "lxml")
[24]: # print soup in order to find table
#print(soup.prettify())
[25]: # find and print table using soup
table = soup.find('table',class_='wikitable sortable')
# table
```

Using pandas we convert the above data into pandas data frame and we get the below indicative table:

```
# Convert above data to pandas dataframe
df = pd.DataFrame(info, columns=['Postal Code', 'Borough', 'Neighbourhood'])
df
```

	Postal Code	Borough	Neighbourhood
0	None	None	None
1	M1A	Not assigned	Not assigned
2	M2A	Not assigned	Not assigned
3	M3A	North York	Parkwoods
4	M4A	North York	Victoria Village
5	M5A	Downtown Toronto	Harbourfront
6	M6A	North York	Lawrence Heights
7	M6A	North York	Lawrence Manor
8	M7A	Downtown Toronto	Queen's Park
9	M8A	Not assigned	Not assigned
10	M9A	Queen's Park	Not assigned

Using wget we download from cocl.us /Geospatial_data the coordinates of Toronto and the relevant postal code :

```
n [39]: # Download coordinates and postal codes for Toronto from web using wget
import wget
url_get = wget.download( 'http://cocl.us/Geospatial_data/toronto_coordinates.csv')
coordinates = pd.read_csv(url_get)
coordinates.head()
```

ut[39]:

	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

We will then merge the two tables in order to combine the info related to neighbourhood, Borough, Postal Code, and coordinates and after doing some fine tuning in the data we get the final table and map of Toronto centre /Neighbourhoods   (using folium):

```
# Limit Borough's to 'EastToronto', 'CentralToronto', 'DowntownToronto', 'WestToronto' na
dff = ['EastToronto', 'CentralToronto', 'DowntownToronto', 'WestToronto']
df6 = df5[df5['Borough'].isin(dff)].reset_index(drop=True)
df6.head()
```

	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M4E	EastToronto	The Beaches	43.676357	-79.293031
1	M4K	EastToronto	The Danforth West,Riverdale	43.679557	-79.352188
2	M4L	EastToronto	The Beaches West,India Bazaar	43.668999	-79.315572
3	M4M	EastToronto	Studio District	43.659526	-79.340923
4	M4N	CentralToronto	Lawrence Park	43.728020	-79.388790

In [36]:
```
# map the above Borough's
mapTorontoCenter = folium.Map(location=[latitudeToronto, longitudeToronto], zoom_start=12)

for lat, long, post, borough, neigh in zip(df6['Latitude'], df6['Longitude'], df6['PostalCode'], df6['Borough'], df6['Neighbourh
    label = "{} ({}): {}".format(borough, post, neigh)
    popup = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, long],
        radius=5,
        popup=popup,
        color='blue',
        fill=True,
        fill_color='#cc3139',
        fill_opacity=0.8,
        ).add_to(mapTorontoCenter)

mapTorontoCenter
```

Out[36]:
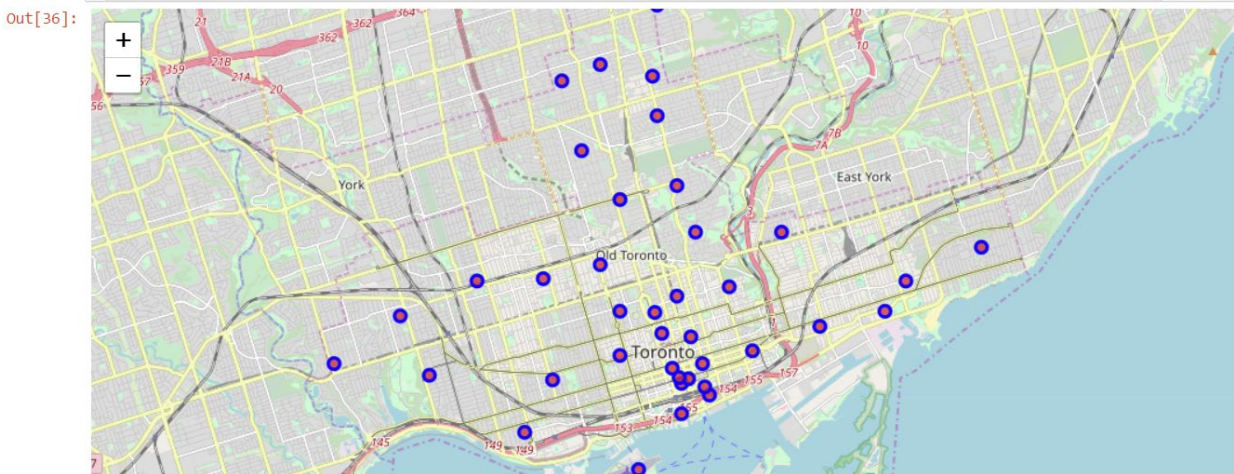
Now that we have mapped all Toronto centre Neighbourhoods, we will utilize once more Foursquare in order to get the 5 most common venue categories in each Neighbourhood. An indicative table of our results is presented below:

	PostalCode	Borough	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
27	M5V	DowntownToronto	CN Tower,Bathurst Quay,Island airport,Harbourf...	Airport Service	Airport Lounge	Airport Terminal	Boutique	Plane
32	M6J	WestToronto	Little Portugal,Trinity	Bar	Men's Store	Coffee Shop	Restaurant	Asian Restaurant
26	M5T	DowntownToronto	Chinatown,Grange Park,Kensington Market	Café	Chinese Restaurant	Vietnamese Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant
3	M4M	EastToronto	Studio District	Café	Coffee Shop	Brewery	Gastropub	Bakery
34	M6P	WestToronto	High Park,The Junction South	Café	Mexican Restaurant	Bar	Thai Restaurant	Diner

Since our target in to find a neighbourhood/borough of Toronto centre, which has similar characteristics to Athens-Kolonaki we will utilize from "sklearn.cluster" "KMeans". K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity.

In our case we will use 7 clusters a summary of which is presented below:

	PostalCode	Borough	Neighbourhood	Latitude	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	M4N	CentralToronto	Lawrence Park	43.728020	-79.388790	0	Park	Bus Line	Swim School	Yoga Studio	Dim Sum Restaurant
3	M4M	EastToronto	Studio District	43.659526	-79.340923	1	Café	Coffee Shop	Brewery	Gastropub	Bakery
36	M6S	WestToronto	Runnymede,Swansea	43.651571	-79.484450	1	Café	Pizza Place	Coffee Shop	Sushi Restaurant	Italian Restaurant
6	M4R	CentralToronto	North Toronto West	43.715383	-79.405678	1	Clothing Store	Coffee Shop	Sporting Goods Shop	Yoga Studio	Salon / Barbershop
19	M5J	DowntownToronto	Harbourfront East,Toronto Islands,Union Station	43.640816	-79.381752	1	Coffee Shop	Aquarium	Café	Hotel	Fried Chicken Joint
18	M5H	DowntownToronto	Adelaide,King,Richmond	43.650571	-79.384568	1	Coffee Shop	Café	Bar	Steakhouse	Asian Restaurant
21	M5L	DowntownToronto	Commerce Court,Victoria Hotel	43.648198	-79.379817	1	Coffee Shop	Café	Hotel	Restaurant	Seafood Restaurant
20	M5K	DowntownToronto	Design Exchange,Toronto Dominion Centre	43.647177	-79.381576	1	Coffee Shop	Café	Hotel	Steakhouse	Restaurant
15	M5C	DowntownToronto	St. James Town	43.651494	-79.375418	1	Coffee Shop	Café	Restaurant	Cocktail Bar	Hotel
29	M5X	DowntownToronto	First Canadian Place,Underground city	43.648429	-79.382280	1	Coffee Shop	Café	Restaurant	Steakhouse	Gym
28	M5W	DowntownToronto	Stn A PO Boxes 25 The Esplanade	43.646435	-79.374846	1	Coffee Shop	Café	Seafood Restaurant	Cocktail Bar	Hotel
14	M5B	DowntownToronto	Ryerson,Garden District	43.657162	-79.378937	1	Coffee Shop	Clothing Store	Cosmetics Shop	Café	Japanese Restaurant
16	M5E	DowntownToronto	Berczy Park	43.644771	-79.373306	1	Coffee Shop	Cocktail Bar	Bakery	Cheese Shop	Café
37	M7A	DowntownToronto	Queen's Park	43.662301	-79.389494	1	Coffee Shop	Gym	Park	Fried Chicken Joint	Salad Place
17	M5G	DowntownToronto	Central Bay Street	43.657952	-79.387383	1	Coffee Shop	Italian Restaurant	Sandwich Place	Japanese Restaurant	Burger Joint
12	M4Y	DowntownToronto	Church and Wellesley	43.665860	-79.383160	1	Coffee Shop	Japanese Restaurant	Restaurant	Sushi Restaurant	Gay Bar

				Latitude	Longitude						
13	M5A	DowntownToronto	Harbourfront	43.654260	-79.360636	1	Coffee Shop	Park	Bakery	Pub	Mexican Restaurant
11	M4X	DowntownToronto	Cabbagetown,St. James Town	43.667967	-79.367675	1	Coffee Shop	Pharmacy	Café	Italian Restaurant	Pub
7	M4S	CentralToronto	Davisville	43.704324	-79.388790	1	Dessert Shop	Sandwich Place	Gym	Coffee Shop	Italian Restaurant
1	M4K	EastToronto	The Danforth West,Riverdale	43.679557	-79.352188	1	Greek Restaurant	Coffee Shop	Italian Restaurant	Furniture / Home Store	Ice Cream Shop
30	M6G	DowntownToronto	Christie	43.669542	-79.422564	1	Grocery Store	Café	Park	Nightclub	Candy Store
5	M4P	CentralToronto	Davisville North	43.712751	-79.390197	1	Gym	Breakfast Spot	Hotel	Food & Drink Shop	Department Store
33	M6K	WestToronto	Brockton,Exhibition Place,Parkdale Village	43.636847	-79.428191	1	Nightclub	Coffee Shop	Café	Breakfast Spot	Gym
9	M4V	CentralToronto	Deer Park,Forest Hill SE,Rathnelly,South Hill,...	43.686412	-79.400049	1	Pub	Coffee Shop	Light Rail Station	American Restaurant	Sushi Restaurant
24	M5R	CentralToronto	The Annex,North Midtown,Yorkville	43.672710	-79.405678	1	Sandwich Place	Café	Coffee Shop	American Restaurant	Pub
2	M4L	EastToronto	The Beaches West,India Bazaar	43.668999	-79.315572	1	Sandwich Place	Gym	Sushi Restaurant	Brewery	Burger Joint
0	M4E	EastToronto	The Beaches	43.676357	-79.293031	2	Neighborhood	Other Great Outdoors	Coffee Shop	Health Food Store	Trail
23	M5P	CentralToronto	Forest Hill North,Forest Hill West	43.696948	-79.411307	3	Park	Jewelry Store	Trail	Sushi Restaurant	Yoga Studio
10	M4W	DowntownToronto	Rosedale	43.679563	-79.377529	3	Park	Playground	Trail	Yoga Studio	Department Store
27	M5V	DowntownToronto	CN Tower,Bathrust Quay,Island airport,Harbourf...	43.628947	-79.394420	4	Airport Service	Airport Lounge	Airport Terminal	Sculpture Garden	Boat or Ferry
32	M6J	WestToronto	Little Portugal,Trinity	43.647927	-79.419750	4	Bar	Asian Restaurant	Restaurant	Coffee Shop	Vietnamese Restaurant
34	M6P	WestToronto	High Park,The Junction South	43.661608	-79.464763	4	Café	Mexican Restaurant	Bar	Thai Restaurant	Bookstore
25	M5S	DowntownToronto	Harbord,University of Toronto	43.662696	-79.400049	4	Café	Restaurant	Bakery	Bar	Bookstore
26	M5T	DowntownToronto	Chinatown,Grange Park,Kensington Market	43.653206	-79.400049	4	Café	Vietnamese Restaurant	Bar	Chinese Restaurant	Vegetarian / Vegan Restaurant
35	M6R	WestToronto	Parkdale,Roncesvalles	43.648960	-79.456325	4	Gift Shop	Bookstore	Dessert Shop	Eastern European Restaurant	Italian Restaurant
31	M6H	WestToronto	Dovercourt Village,Dufferin	43.669005	-79.442259	4	Pharmacy	Bakery	Middle Eastern Restaurant	Bank	Supermarket
38	M7Y	EastToronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	4	Smoke Shop	Auto Workshop	Skate Park	Brewery	Spa
8	M4T	CentralToronto	Moore Park,Summerhill East	43.689574	-79.383160	5	Playground	Summer Camp	Tennis Court	Yoga Studio	Dessert Shop

Red: 0, Purple: 1, Blue: 2, Turquoise: 3, Green: 4, Yellow green: 5, Orange: 6.
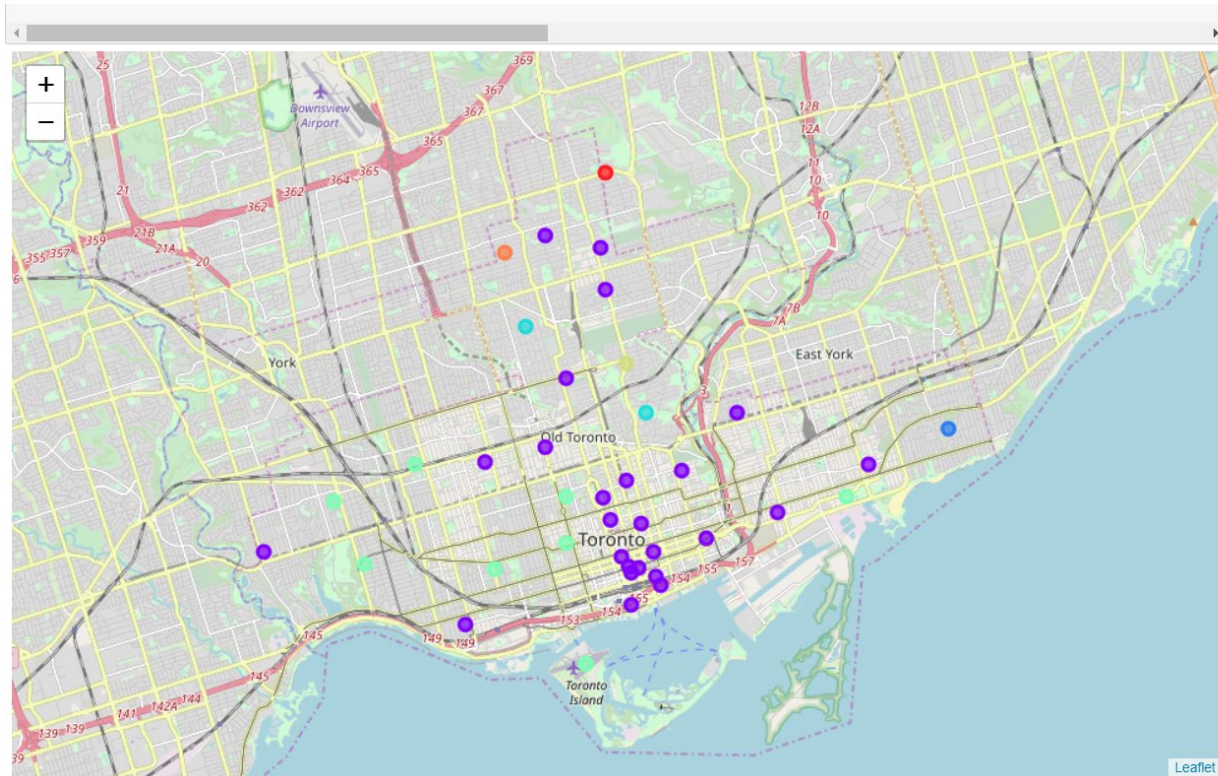
4.   Results section where you discuss the results

From the last Table and map and the table that we have previously presented and refers to Athens Kolonaki venues we can conclude that cluster 1 (purple) and cluster 4 (green) are the "candidate" neighbourhoods that present similarities to Kolonaki Athens.

In specific both clusters appear to have many coffee shops and restaurants similarly to Kolonaki Athens. In addition similarly to Kolonaki Neighbourhood, both clusters are in Toronto Centre.

Considering the above we will suggest as potential candidate Neighbourhoods cluster 1 & 4

5.   Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

As already mentioned cluster 1 and 4 are potential candidates for opening a "pilot" coffee shop on the centre of Toronto Canada.

We strongly believe that before deciding which neighbourhood is the best for opening a pilot store further investigation & information are necessary in order to come to a conclusion.

 Indicatively the following criteria should be considered:

- Real estate prices of each neighbourhood i.e cost of renting a store
- Pedestrian zones which usually increase the number of potential clients
- Licensing procedure restrictions if any
- Company internal factors (which as discussed are not currently disclosed to us)

5.   Conclusion section

Considering our suggestions for further investigation mentioned in section 4, we suggest cluster 1 & 4 (Borough's/Neighbourhood's) as potential candidates for opening a "pilot" coffee shop in the centre of Toronto Canada.