

Capstone Project: The Battle of Neighbourhoods – Toronto

1. Introduction / Business Problem

Identify the business problem:

Toronto is the largest city in Canada with an estimated population of more than 6 million, so it is a good place for entrepreneurs eager to open new businesses that become profitable and drive a great number of customers.

The goal of this project is to use the Foursquare API to provide the location data and clustering of venues information to find out what would be the best neighbourhoods in downtown Toronto to open a new gym venue.

Everybody knows that regular physical activity improves your health and well-being. It has benefits for all ages and abilities including reducing your risk of developing many chronic diseases, increasing your energy, your self-esteem and improving your sleep. So, nowadays, people are going to gyms more frequently due to the increasing interest in a healthy lifestyle.

Target audience:

This project is aimed to entrepreneurs or business owners who want to open a new gym in downtown Toronto or grow their current business in a profitable location where little or no gyms exist today.

2. Data

For the purpose of this project, we will need to have the list of Neighbourhoods of Toronto city, as well as the geographical location of the Toronto's Neighbourhoods and the venue data of gyms through Foursquare API.

Here are the details:

* Wikipedia page with the Neighbourhoods and postal codes of Toronto city ([https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada: M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M))

List of postal codes of Canada: M

From Wikipedia, the free encyclopedia

This is a list of [postal codes in Canada](#) where the first letter is M. Postal codes beginning with M are located within the city of [Toronto](#) in the province of [Ontario](#). Only the first three characters Sortation Area.

[Canada Post](#) provides a free postal code look-up tool on its website,^[1] via its applications for such smartphones as the [iPhone](#) and [BlackBerry](#),^[2] and sells hard-copy directories and CD-ROM which allow customers to properly match addresses and postal codes. Hard-copy directories can also be consulted in all post offices, and some libraries.

Toronto - 103 FSAs [edit]

Note: There are no rural FSAs in Toronto, hence no postal codes should start with M0. However, the postal code M0R 8T0 is assigned to an [Amazon](#) warehouse in Mississauga, suggesting th FSA for high volume addresses.

Postal Code ↕	Borough ↕	Neighbourhood ↕
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
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
M8A	Not assigned	Not assigned
M9A	Etobicoke	Islington Avenue, Humber Valley Village
M1B	Scarborough	Malvern, Rouge
M2B	Not assigned	Not assigned
M3B	North York	Don Mills
M4B	East York	Parkview Hill, Woodbine Gardens
M5B	Downtown Toronto	Garden District, Ryerson
M6B	North York	Glencairn
M7B	Not assigned	Not assigned
M8B	Not assigned	Not assigned
M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale

The Wikipedia site provide the following information: postal code, borough and the name of the neighbourhoods existing in Toronto.

We need to scrape the data from the Wikipedia table into a data frame as below:

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

* the geographical coordinates of each postal code through reading the CSV file (https://cocl.us/Geospatial_data) and transforming it into a Pandas data frame.

Postal Code	Latitude	Longitude
M1B	43.80669	-79.1944
M1C	43.78454	-79.1605
M1E	43.76357	-79.1887
M1G	43.77099	-79.2169
M1H	43.77314	-79.2395
M1J	43.74473	-79.2395

	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

* explore and get the venues data using Foursquare API:

For retrieving the name, location and category of venues in Toronto, we need to create an account within Foursquare API and use the Client_ID and Client_Secret to pull the necessary data.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
...
4640	Business reply mail Processing Centre, South	43.663744	-79.331550	TTC Russell Station	43.664000	-79.331550	Light Rail Station

3. Methodology

3.1 Exploratory Data Analysis

When getting the data from Wikipedia source, we need to make the following transformations:

- Ignore data where a Borough is 'not assigned'.
- If a cell has a Borough but a 'not assigned' Neighbourhood, then the Neighbourhood will be the same as the Borough.
- Group the data by PostalCode and Borough

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

Then we combine this dataframe (df) with the geographical data that includes latitude and longitude information (geo_df) based on PostalCode, obtaining the toronto.df dataframe:

```
# Merge both dataframes on PostalCode
toronto = pd.merge(df, geo_df, on='PostalCode')
toronto.head()
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	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

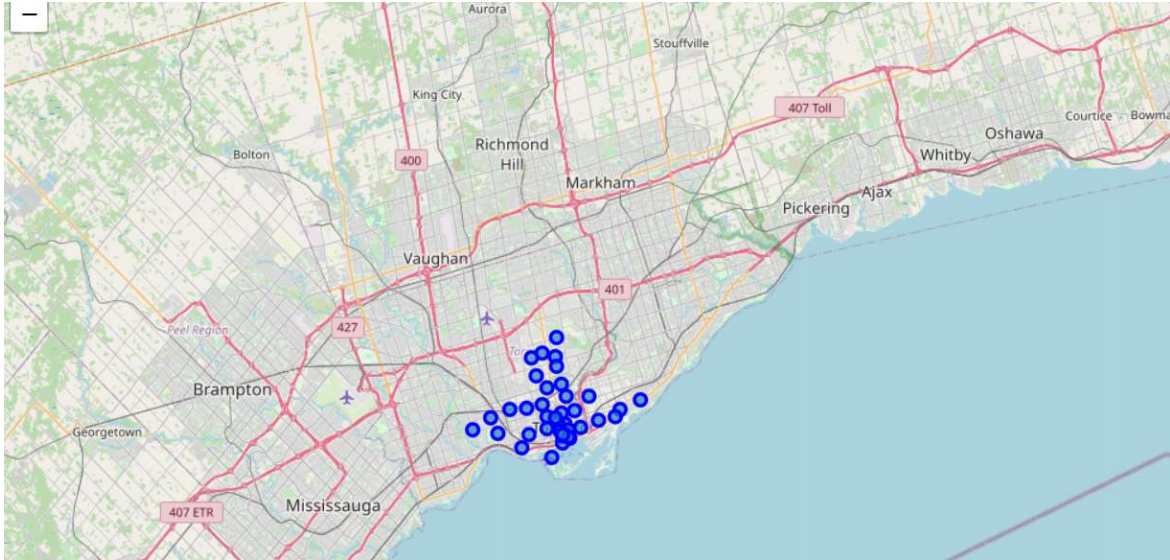
Afterwards, we filter out the data to Boroughs containing the word “Toronto” in their names, which represents the downtown Toronto.

We create a folium map to have a visual representation:

```
# Create a map of Toronto to plot the Boroughs with "Toronto" in its name
latitude = 43.651070
longitude = -79.347015

map_toronto = folium.Map(location=[latitude, longitude], zoom_start=10)

for lat, lng, borough, neighborhood in zip(toronto_central['Latitude'], toronto_central['Longitude'], toronto_central['Borough'],
                                           toronto_central['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)
```



Now using Foursquare API we explore the Neighbourhoods and get the nearby venues within a 500-meters radius

```
print(toronto_central_venues.shape)
print('There are {} uniques categories.'.format(len(toronto_central_venues['Venue Category'].unique())))
toronto_central_venues
```

(1624, 7)
There are 235 uniques categories.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant
...
1619	Business reply mail Processing Centre, South C...	43.662744	-79.321558	TTC Russell Division	43.664908	-79.322560	Light Rail Station
1620	Business reply mail Processing Centre, South C...	43.662744	-79.321558	Jonathan Ashbridge Park	43.664702	-79.319898	Park
1621	Business reply mail Processing Centre, South C...	43.662744	-79.321558	Olliffe On Queen	43.664503	-79.324768	Butcher
	Business reply mail Processing Centre, South						

Then we group the venues by Neighbourhood:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Berczy Park	55	55	55	55	55	55
Brockton, Parkdale Village, Exhibition Place	23	23	23	23	23	23
Business reply mail Processing Centre, South Central Letter Processing Plant Toronto	16	16	16	16	16	16
CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport	16	16	16	16	16	16
Central Bay Street	68	68	68	68	68	68
Christie	16	16	16	16	16	16
Church and Wellesley	75	75	75	75	75	75
Commerce Court, Victoria Hotel	100	100	100	100	100	100
Davisville	33	33	33	33	33	33
Davisville North	9	9	9	9	9	9
Dufferin, Dovercourt Village	13	13	13	13	13	13
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. Rverson	100	100	100	100	100	100

Then we start to analyse each Neighbourhood, firstly using one-hot encoding format and secondly grouping the data by Neighbourhood and by the mean of the frequency of occurrence of each category:

Analyze each Neighborhood

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

# add Borough column to the dataframe
toronto_onehot['Neighborhoods'] = toronto_central_venues['Neighborhood']

# move Neighborhood to the first column
fixed_columns = [toronto_onehot.columns[-1]] + list(toronto_onehot.columns[:-1])
toronto_onehot = toronto_onehot[fixed_columns]

print(toronto_onehot.shape)
toronto_onehot.head()

(1624, 236)
```

	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theater	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant	Win Bz
0	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	0	0	
1	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	
2	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	
3	The Beaches	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	
4	The Danforth West, Riverdale	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	

Activate Windows

Let's group data by Neighborhood and by the mean of the frequency of occurrence of each category

```
toronto_grouped = toronto_onehot.groupby(['Neighborhoods']).mean().reset_index()
```

```
print(toronto_grouped.shape)
```

```
toronto_grouped.head()
```

(39, 236)

	Neighborhoods	Afghan Restaurant	Airport	Airport Food Court	Airport Gate	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	...	Theater	Theme Restaurant	Toy / Game Store	Trail	Train Station	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant
0	Berczy Park	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.018182	0.0	0.0
1	Brockton, Parkdale Village, Exhibition Place	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
2	Business reply mail Processing Centre, South C...	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
3	CN Tower, King and Spadina, Railway Lands, Har...	0.0	0.0625	0.0625	0.0625	0.125	0.125	0.0625	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0
4	Central Bay Street	0.0	0.0000	0.0000	0.0000	0.000	0.000	0.0000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.014706	0.0	0.0

Finally, we get the average of gyms by Neighbourhood. Remember that the goal of this project is to find out the best Neighbourhood to open a new gym in downtown Toronto:

```
gym_data = toronto_grouped[['Neighborhoods', 'Gym']]
gym_data.head()
```

	Neighborhoods	Gym
0	Berczy Park	0.000000
1	Brockton, Parkdale Village, Exhibition Place	0.043478
2	Business reply mail Processing Centre, South C...	0.000000
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000
4	Central Bay Street	0.000000

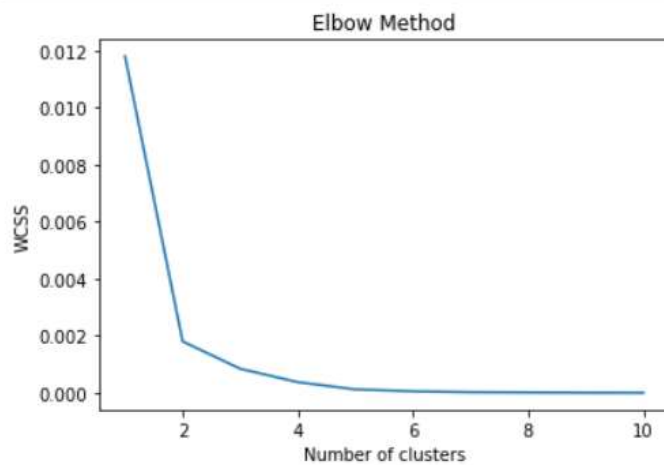
3.2 Clustering Model

We want to cluster the Neighbourhoods based on the criteria of having a similar number of gyms in them. To do so we use the K-Means method. But how to know which is the optimal number of clusters? Using the Elbow method, which consists of training multiple models using different number of clusters and storing the value of inertia_property (wcoss) every time.

```
wcss = []

for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=100, n_init=10, random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

plt.plot(range(1, 11), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```



We can conclude that the optimum K value is 5, so we will have 5 clusters.

So we run K-Means clustering with this value and create a new data frame including the clusters:

```
### Create a new dataframe that includes the clusters
toronto_merged = gym_data.copy()

toronto_merged['Cluster Labels'] = kmeans.labels_

toronto_merged.head()
```

	Neighborhoods	Gym	Cluster Labels
0	Berczy Park	0.000000	0
1	Brockton, Parkdale Village, Exhibition Place	0.043478	4
2	Business reply mail Processing Centre, South C...	0.000000	0
3	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0
4	Central Bay Street	0.000000	0

Next, we merge this data frame with the one containing the venues (toronto_central_venues):

```
### Merge toronto_merged with toronto_venues
toronto_merged = toronto_merged.join(toronto_central_venues.set_index('Neighborhood'), on='Neighborhoods')

print(toronto_merged.shape)
toronto_merged.head()
```

(1624, 9)

	Neighborhoods	Gym	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berczy Park	0.0	0	43.644771	-79.373306	The Keg Steakhouse + Bar - Esplanade	43.646712	-79.374768	Restaurant
0	Berczy Park	0.0	0	43.644771	-79.373306	LCBO	43.642944	-79.372440	Liquor Store
0	Berczy Park	0.0	0	43.644771	-79.373306	Fresh On Front	43.647815	-79.374453	Vegetarian / Vegan Restaurant
0	Berczy Park	0.0	0	43.644771	-79.373306	Goose Island Brewhouse	43.647329	-79.373541	Beer Bar
0	Berczy Park	0.0	0	43.644771	-79.373306	Hockey Hall Of Fame (Hockey Hall of Fame)	43.646974	-79.377323	Museum


We create a folium map to plot the clusters and Neighbourhoods:

```
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=10)

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]

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

map_clusters
```



4. Results and Discussion

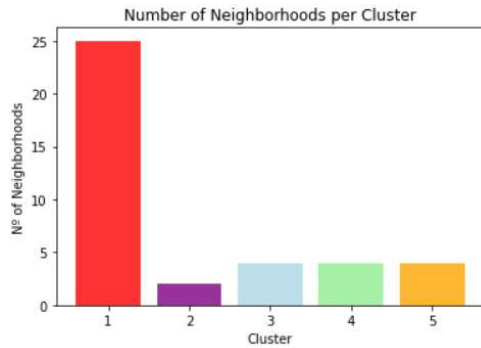
We have a total of five clusters (0, 1, 2, 3, 4) and we get how many neighbourhoods there are in each cluster.

```

objects = (1, 2, 3, 4, 5)
y_pos = np.arange(len(objects))
performance = gym_data['Cluster Labels'].value_counts().to_frame().sort_index(ascending=True)
perf = performance['Cluster Labels'].tolist()
plt.bar(y_pos, perf, align='center', alpha=0.8, color=['red', 'purple', 'lightblue', 'lightgreen', 'orange'])
plt.xticks(y_pos, objects)
plt.ylabel('Nº of Neighborhoods')
plt.xlabel('Cluster')
plt.title('Number of Neighborhoods per Cluster')

plt.show()

```



Then we analyse each cluster.

Cluster1 – colour red

This cluster encompasses the Boroughs of East and Downtown Toronto, where the main Neighbourhoods are The Beaches, Kensington Market, Chinatown, Grange Park, Regent Park and Harbourfront. The total number of Neighbourhoods within this cluster are 25.

In this cluster, there was no gym venues.

```
cluster1_df.sort_values(['Gym'], ascending=False)
```

	Borough	Neighborhood	Gym	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	East Toronto	The Beaches	0.0	0	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
421	Downtown Toronto	Kensington Market, Chinatown, Grange Park	0.0	0	43.653206	-79.400049	Global Cheese	43.654623	-79.400606	Cheese Shop
414	Downtown Toronto	Kensington Market, Chinatown, Grange Park	0.0	0	43.653206	-79.400049	Little Pebbles	43.654883	-79.400264	Coffee Shop
415	Downtown Toronto	Kensington Market, Chinatown, Grange Park	0.0	0	43.653206	-79.400049	El Rey	43.652764	-79.400048	Cocktail Bar
416	Downtown Toronto	Kensington Market, Chinatown, Grange Park	0.0	0	43.653206	-79.400049	Millie Creperie	43.654994	-79.399829	Dessert Shop
...
209	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
210	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
211	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant
212	Downtown Toronto	Regent Park, Harbourfront	0.0	0	43.654260	-79.360636	Soulpepper Theatre	43.650780	-79.357615	Theater
626	East Toronto	Business reply mail Processing Centre, South C...	0.0	0	43.662744	-79.321558	Ashbridges Bay Skatepark	43.662548	-79.315631	Skate Park

Cluster2 – colour purple

This cluster includes Central Toronto and East Toronto as Boroughs and there are two Neighbourhoods in it. The average of gym venues is 0.06 and the total number of these venues is 3.

```
cluster2_df.sort_values(['Gym'], ascending=False)
```

	Borough	Neighborhood	Gym	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
26	Central Toronto	Davisville	0.060606	1	43.704324	-79.388790	Marcheleo's Gourmet Marketplace	43.708041	-79.392195	Gourmet Shop
39	Central Toronto	Davisville	0.060606	1	43.704324	-79.388790	Viva Napoli	43.705752	-79.389125	Pizza Place
29	Central Toronto	Davisville	0.060606	1	43.704324	-79.388790	Thobors Boulangerie Patisserie Café	43.704514	-79.388616	Café
30	Central Toronto	Davisville	0.060606	1	43.704324	-79.388790	Marigold Indian Bistro	43.702881	-79.388008	Indian Restaurant
31	Central Toronto	Davisville	0.060606	1	43.704324	-79.388790	Zee Grill	43.704985	-79.388476	Seafood Restaurant
32	Central Toronto	Davisville	0.060606	1	43.704324	-79.388790	Starbucks	43.705923	-79.389548	Coffee Shop
33	Central Toronto	Davisville	0.060606	1	43.704324	-79.388790	June Rowlands Park	43.700517	-79.389189	Park
Central										

Cluster3 – colour lightblue

In this cluster there are 4 Neighbourhoods and a total of 6 gym venues in it.

```
cluster3_df.sort_values(['Gym'], ascending=False)
```

	Borough	Neighborhood	Gym	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
246	Downtown Toronto	Queen's Park, Ontario Provincial Government	0.030303	2	43.662301	-79.389494	Queen's Park	43.663946	-79.392180	Park
196	West Toronto	Runnymede, Swansea	0.030303	2	43.651571	-79.484450	Awai	43.650412	-79.478477	Vegetarian / Vegan Restaurant
211	West Toronto	Runnymede, Swansea	0.030303	2	43.651571	-79.484450	The Coffee Bouquets	43.648785	-79.485940	Coffee Shop
210	West Toronto	Runnymede, Swansea	0.030303	2	43.651571	-79.484450	(The New) Moksha Yoga Bloor West	43.648658	-79.485242	Yoga Studio
208	West Toronto	Runnymede, Swansea	0.030303	2	43.651571	-79.484450	Supper Solved	43.648781	-79.485233	Restaurant
...
139	Downtown Toronto	Stn A PO Boxes	0.020833	2	43.646435	-79.374846	NAMI	43.650853	-79.375887	Japanese Restaurant
140	Downtown Toronto	Stn A PO Boxes	0.020833	2	43.646435	-79.374846	The Old Spaghetti Factory	43.646964	-79.374403	Italian Restaurant
Downtown										

Cluster4 – colour lightgreen

This cluster is about Downtown Toronto Borough and has 4 Neighbourhoods as well as cluster3. The highest average of gyms in this cluster is 0.013 and there is a total of 4 gyms.

```
cluster4_df.sort_values(['Gym'], ascending=False)
```

	Borough	Neighborhood	Gym	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Downtown Toronto	Church and Wellesley	0.013333	3	43.665860	-79.383160	Rolltation	43.669388	-79.386566	Sushi Restaurant
48	Downtown Toronto	Church and Wellesley	0.013333	3	43.665860	-79.383160	The Blake House	43.664468	-79.377471	American Restaurant
55	Downtown Toronto	Church and Wellesley	0.013333	3	43.665860	-79.383160	Cawthra Square Dog Park	43.666583	-79.380040	Dog Run
54	Downtown Toronto	Church and Wellesley	0.013333	3	43.665860	-79.383160	Woody's and Sailor	43.664390	-79.380361	Gay Bar
53	Downtown Toronto	Church and Wellesley	0.013333	3	43.665860	-79.383160	Wish	43.668759	-79.385694	Restaurant
...
173	Downtown Toronto	Garden District, Ryerson	0.010000	3	43.657162	-79.378937	The Black Canary Espresso Bar	43.657029	-79.381385	Café

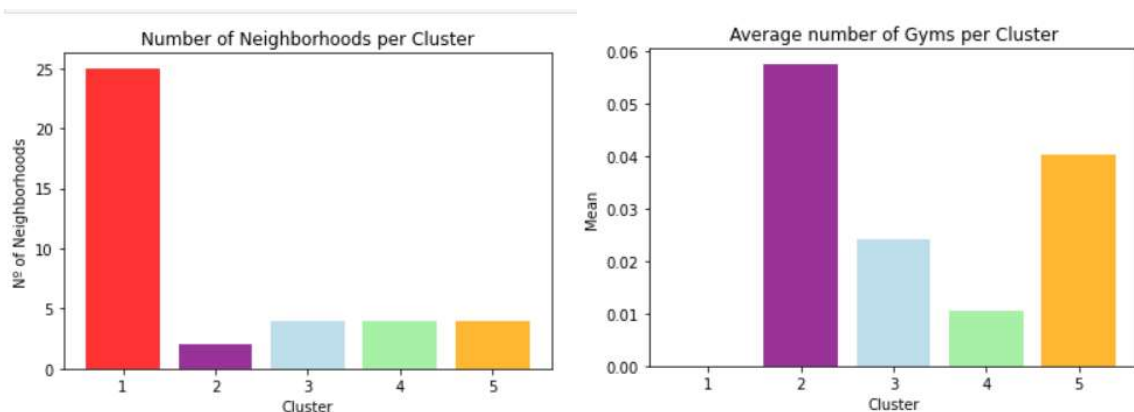
Cluster5 – colour orange

In this cluster there are also 4 Neighbourhoods in Borough like West Toronto and Downtown Toronto. The highest average of gyms in this cluster is 0.043 and there are 13 gyms in total.

```
cluster5_df.sort_values(['Gym'], ascending=False)
```

	Borough	Neighborhood	Gym	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
322	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.043478	4	43.636847	-79.428191	Lamport Stadium	43.638778	-79.423534	Stadium
311	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.043478	4	43.636847	-79.428191	Gardiner Expy & Dufferin St	43.633800	-79.425460	Intersection
300	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.043478	4	43.636847	-79.428191	Dunn Milk Variety	43.633667	-79.431549	Convenience Store
301	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.043478	4	43.636847	-79.428191	Joe Rockhead's Climbing Gym	43.636342	-79.423582	Climbing Gym
302	West Toronto	Brockton, Parkdale Village, Exhibition Place	0.043478	4	43.636847	-79.428191	RED Nightclub	43.637794	-79.423759	Nightclub
...
106	Downtown Toronto	Commerce Court, Victoria Hotel	0.040000	4	43.648198	-79.379817	Prairie Girl Bakery	43.648332	-79.382305	Cupcake Shop
105	Downtown Toronto	Commerce Court, Victoria Hotel	0.040000	4	43.648198	-79.379817	Metropolitan Resto Bar	43.650062	-79.377181	Italian Restaurant
...	Downtown

5. Conclusion



In this project, we manage to cluster the downtown of Toronto city using demographic data by Neighbourhoods, which help us identify which were the most suitable for opening a new gym.

Most of the gym venues in downtown Toronto city are in cluster2 represented by the purple colour. The Neighbourhoods located in Central Toronto that have the highest average of gyms are Davisville, India Bazaar and The Beaches West.

Even though cluster1 has the larger number of Neighbourhoods (25), there is no gym within it, so this is a great opportunity to open a new one, with no competition. So Neighbourhoods like The Beaches, Kensington Market, Chinatown, Grange Park, Regent Park, Harbourfront would be the ideal location for the opening of a gym.

We must also point out that there is at least one drawback in this analysis and it is the fact of only using the data that come from Foursquare API, without looking for richer information like population of different Neighbourhoods to have one more element to make the decision.