Location Data Analysis to Open a Fitness Centre

Final Capstone Project

AGENDA

1. Introduction
2. Data Processing
3. Methodology
4. Results and Discussion
5. Conclusion

1. Introduction

Background and Business Problem

- According to 2019 ParticipACTION Adult Report Card conducted by Canadian Fitness and Lifestyle Research Institute, 29% of Canadian adults (aged 18 to 79 years old) had low active lifestyle while 18% of them were observed having sedentary lifestyle.

- Reducing sedentary behaviors and increasing physical activity can benefit to many individuals, especially working adults who spend most of their time in the office throughout the day.

- In recent years, the commitment in promoting physical activity to get healthier lifestyle was initiated by the government and different organizations.

- Toronto, being the international centre for businesses and crowded city with working professionals, is a good place to consider opening a fitness centre for those who would like to promote active lifestyle among Canadians.

Problem Statement – To choose one of the best neighborhoods in Toronto to open a yoga studio or fitness centre to promote healthy lifestyle

2. Data Processing

Data Sources and Acquisition

- To get the list of all the neighborhoods in Toronto, Wikipedia page "List of postal codes of Canada: M" (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) is scraped.
- Foursquare API is used to extract the venue data such as names, categories, latitude, longitude.
- The required category IDs (i.e., office, gym/fitness centre, yoga studio in this project) is retrieved from Foursquare developer website at https://developer.foursquare.com/docs/build-with-foursquare/categories/.
- If the name of neighborhood is 'not assigned', it is replaced with the name of borough.
- After cleaning the data, there are a total of 15 boroughs and 103 neighborhoods.

3. Methodology

Based on the business problem of this project, the data will be explored and analyzed in the following steps.

- explore the nearby offices, gyms and yoga studios of all the neighborhood in each borough
- decide which borough(s) have more offices and popular among working professionals
- delve into each neighborhood of selected borough
- cluster the neighborhoods of selected borough by predictive modelling and recommend the location

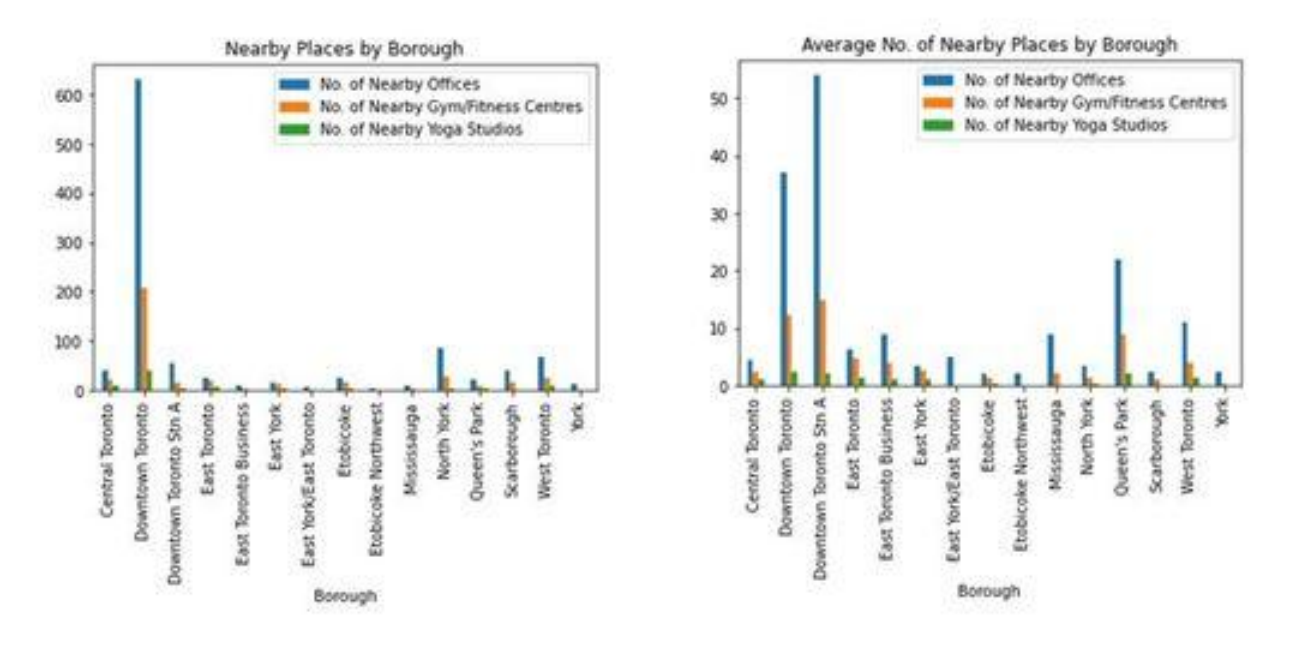
3. Methodology

Exploratory Data Analysis

	Borough	No. of Nearby Offices	No. of Nearby Gym/Fitness Centres	No. of Nearby Yoga Studios
0	Central Toronto	39	21	9
1	Downtown Toronto	631	208	39
2	Downtown Toronto Stn A	54	15	2
3	East Toronto	25	19	5
4	East Toronto Business	9	4	1

Total no. of nearby offices, gyms and yoga studios by Borough

No. of nearby places and average no. of nearby places by Borough



3. Methodology

Location Data Analysis

'Downtown Toronto' is the most popular among working professional with more than 600 nearby offices.

```
In [17]: df_downtown = df_toronto[df_toronto['Borough'] == 'Downtown Toronto'].reset_index(drop=True)
         df_downtown
```

Out[17]:

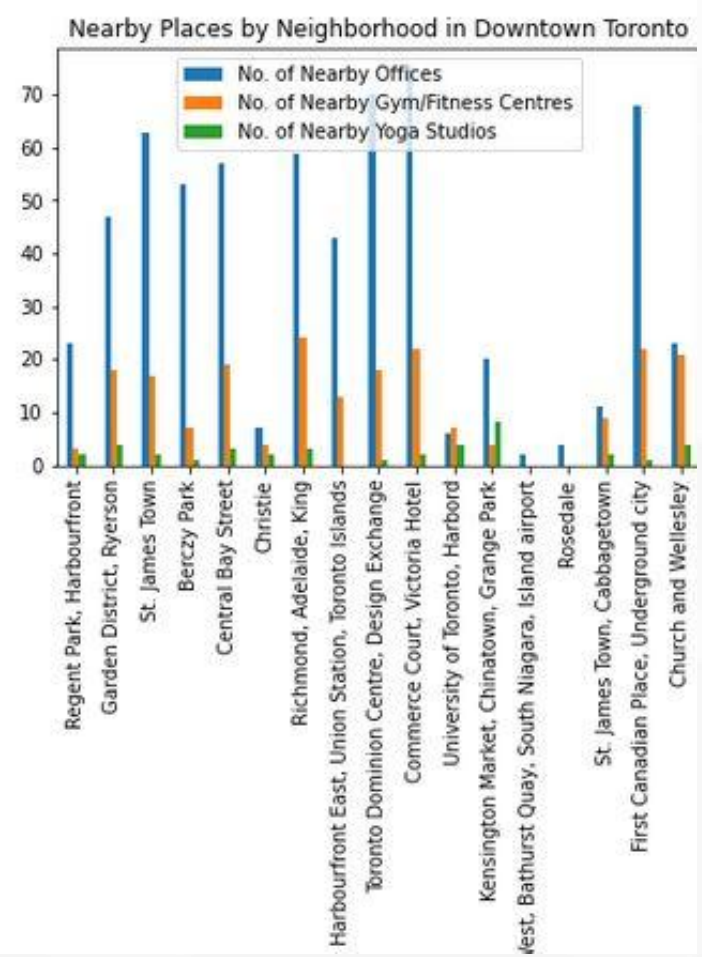
	PostalCode	Borough	Neighbourhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
2	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
3	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
4	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
5	M6G	Downtown Toronto	Christie	43.669542	-79.422564
6	M5H	Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568
7	M5J	Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752
8	M5K	Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576
9	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817
10	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049
11	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049
12	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.628947	-79.394420
13	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529
14	M4X	Downtown Toronto	St. James Town, Cabbagetown	43.667967	-79.367675
15	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280
16	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160

All neighbourhoods in Downtown Toronto

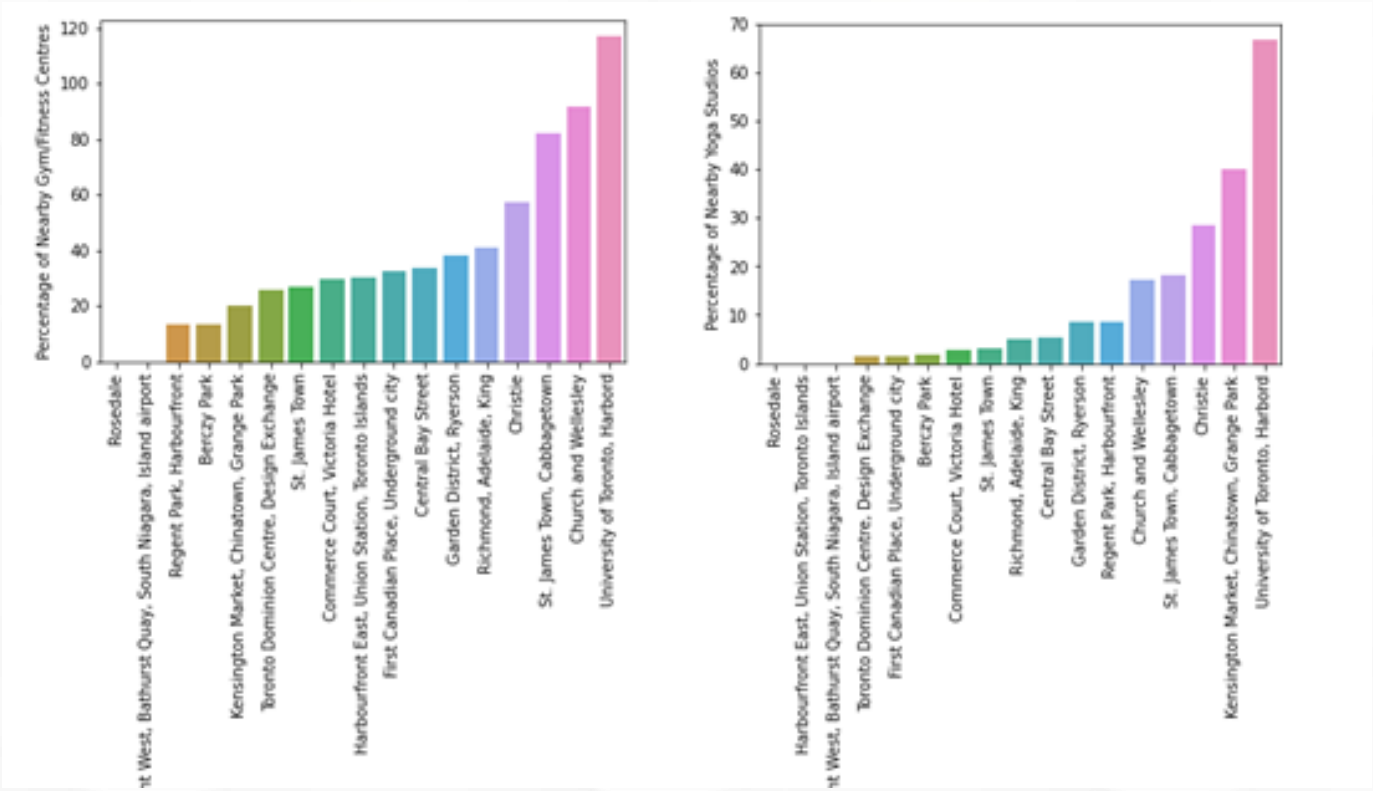
Folium map of neighbourhoods in Downtown Toronto

3. Methodology

Location Data Analysis



No. of nearby offices, gyms and yoga studios in Downtown Toronto



Percentage of nearby gyms and yoga studios in Downtown Toronto

3. Methodology

Predictive Modelling

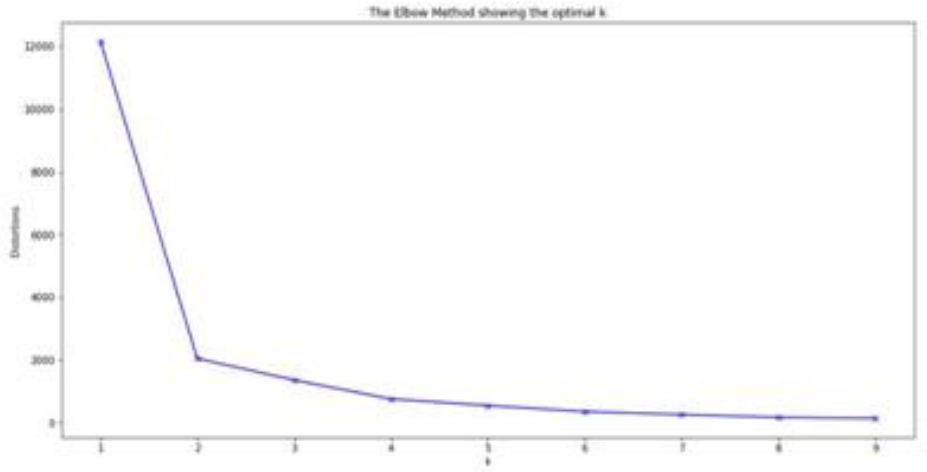
Clustering Neighborhood by K-means clustering

To cluster the neighborhoods, unsupervised machine learning technique K-means algorithm is adopted in this project. First, the elbow method is used to identify the optimal k value in a given dataset. As shown in the figure, the best k value seems to be 3 after analyzing elbow method using distortions.

```
nearby_downtown = nearby_downtown.drop(['Borough', 'Neighborhood'], 1)

distortions=[]
for k in range(1,10):
    km=KMeans(n_clusters=k)
    km.fit(nearby_downtown)
    distortions.append(km.inertia_)

plt.figure(figsize=(16,8))
plt.plot(range(1,10), distortions, 'bx-')
plt.xlabel('k')
plt.ylabel('Distortions')
plt.title('The Elbow Method showing the optimal k')
plt.show()
```



3. Methodology

Predictive Modelling

Clustering Neighborhood by K-means clustering

After clustering with k=3, each cluster is examined to identify and recommend the most promising neighborhood to open a yoga studio or a fitness centre.

Cluster 0

```
In [30]: nearby_downtown.loc[nearby_downtown['Cluster Labels'] == 0]
```

Out[30]:

	Neighborhood	Cluster Labels	No. of Nearby Offices	No. of Nearby Gym/Fitness Centres	No. of Nearby Yoga Studios
1	Garden District, Ryerson	0	47	18	4
3	Berczy Park	0	53	7	1
7	Harbourfront East, Union Station, Toronto Islands	0	43	13	0

```
In [31]: nearby_downtown.loc[nearby_downtown['Cluster Labels'] == 0].mean()
```

```
Out[31]: Cluster Labels                  0.000000
         No. of Nearby Offices          47.666667
         No. of Nearby Gym/Fitness Centres    12.666667
         No. of Nearby Yoga Studios      1.666667
         dtype: float64
```

Cluster 1

```
In [32]: nearby_downtown.loc[nearby_downtown['Cluster Labels'] == 1]
```

Out[32]:

	Neighborhood	Cluster Labels	No. of Nearby Offices	No. of Nearby Gym/Fitness Centres	No. of Nearby Yoga Studios
0	Regent Park, Harbourfront	1	23	3	2
5	Christie	1	7	4	2
10	University of Toronto, Harbord	1	6	7	4
11	Kensington Market, Chinatown, Grange Park	1	20	4	8
12	CN Tower, King and Spadina, Railway Lands, Har...	1	2	0	0
13	Rosedale	1	4	0	0
14	St. James Town, Cabbagetown	1	11	9	2
16	Church and Wellesley	1	23	21	4

```
In [33]: nearby_downtown.loc[nearby_downtown['Cluster Labels'] == 1].mean()
```

```
Out[33]: Cluster Labels                  1.00
         No. of Nearby Offices          12.00
         No. of Nearby Gym/Fitness Centres     6.00
         No. of Nearby Yoga Studios      2.75
         dtype: float64
```

Cluster 2

```
In [34]: nearby_downtown.loc[nearby_downtown['Cluster Labels'] == 2]
```

Out[34]:

	Neighborhood	Cluster Labels	No. of Nearby Offices	No. of Nearby Gym/Fitness Centres	No. of Nearby Yoga Studios
2	St. James Town	2	63	17	2
4	Central Bay Street	2	57	19	3
6	Richmond, Adelaide, King	2	59	24	3
8	Toronto Dominion Centre, Design Exchange	2	70	18	1
9	Commerce Court, Victoria Hotel	2	75	22	2
15	First Canadian Place, Underground city	2	68	22	1

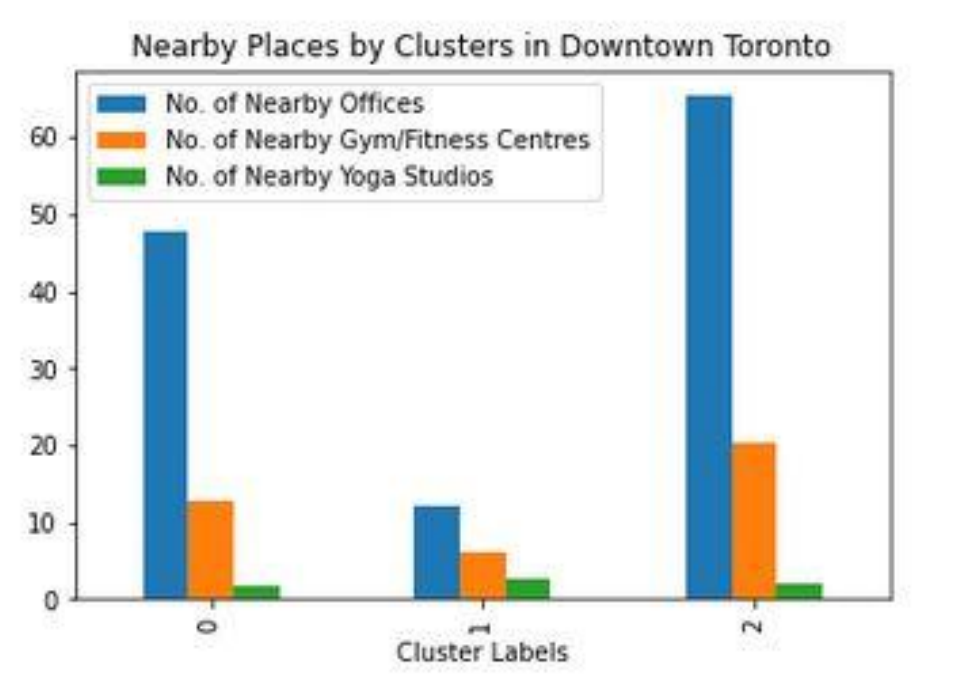
```
In [35]: nearby_downtown.loc[nearby_downtown['Cluster Labels'] == 2].mean()
```

```
Out[35]: Cluster Labels                  2.000000
         No. of Nearby Offices          65.333333
         No. of Nearby Gym/Fitness Centres    20.333333
         No. of Nearby Yoga Studios      2.000000
         dtype: float64
```

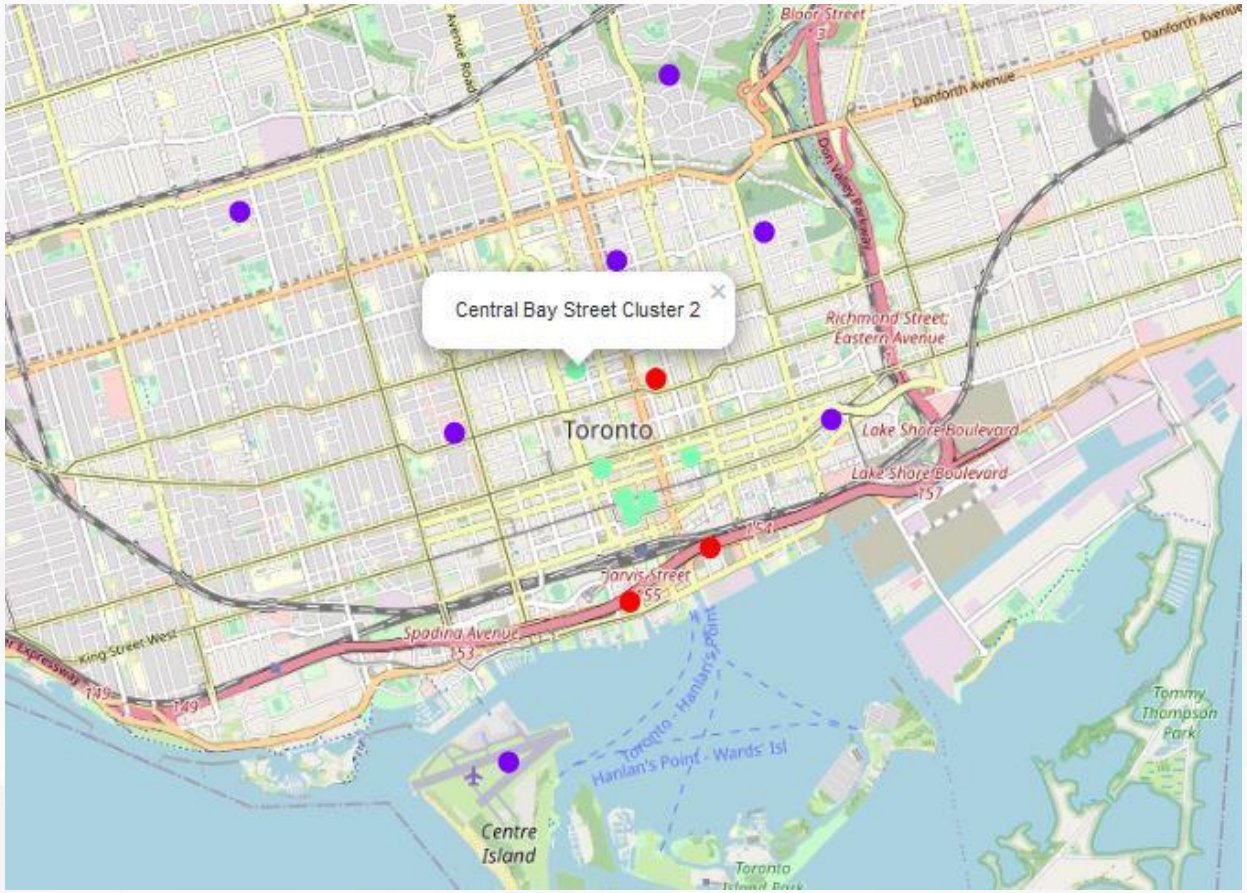
3. Methodology

Predictive Modelling

Clustering Neighborhood by K-means clustering



Bar chart of nearby places by clusters



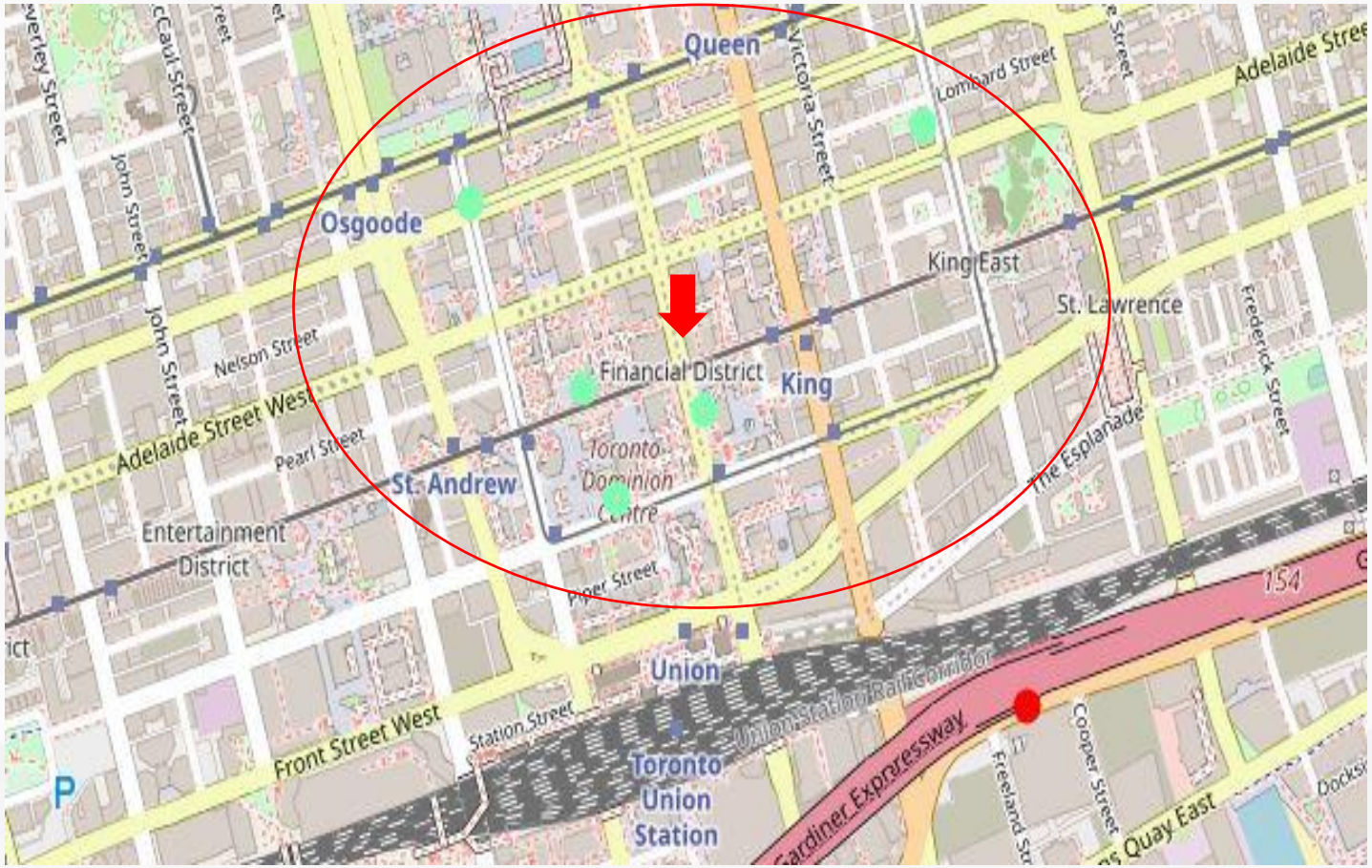
Folium map of each cluster in Downtown Toronto

Cluster 0 Cluster 1 Cluster 2

3. Methodology

Predictive Modelling

A closer look at Cluster 2



Financial District, Toronto

Three of the six neighborhoods in Cluster 2 are in Financial District.

Finacial District, Toronto is considered to be the main business district with many office towers and centre of Canada's financial industry.

Total no. of offices near Financial District: **72**

Total no. of gym/fitness centres near Financial District: **19**

Total no. of yoga studio near Financial District: **1**

4. Results and Discussion

- After exploring the nearby places in each borough, Downtown Toronto has the most number of offices and gyms nearby, followed by North York and West Toronto. Thus, Downtown Toronto is focused for more in-depth analysis.

- The analysis shows that there are many neighborhood candidates in Downtown Toronto. With the help of predictive modelling, specifically K-means clustering helps to identify the different clusters of each neighborhood.

- After analyzing with k=3, Cluster 0 represents the middle cluster with average no. of 48 offices, 13 gym/fitness centres and 2 yoga studios nearby. Cluster 1 has the least average no. of places with 12 offices, 6 gym/fitness centres and 3 yoga studios. Meanwhile, Cluster 2 stands out with the highest average no. of nearby places with 65 offices, 20 gym/fitness centres and 2 yoga studios.

- The optimal places to open a business will be different based on business problem and decision criteria. Here, I would recommend to explore more on Financial District area (Cluster 2) in Downtown Toronto.

- According to data analysis, there are 72 nearby offices within 500 metres of Financial District, in which there are 19 gym/fitness centres and only 1 yoga studios. Thus, starting a yoga studio in the area would not be a bad idea.

5. Conclusion

- The nearby offices, gyms and yoga studios of all the neighborhoods in each borough in Toronto was explored.
- The borough with more offices and popular among working professionals was selected and delved into each neighborhoods of selected borough which was Downtown Toronto.
- The predictive modelling helped to identify the three clusters of Downtown Toronto and recommended the location.
- Hopefully, this project could be the starting point for more detailed analysis and insightful recommendations for future studies of data science with different business ideas.

- Thank You -