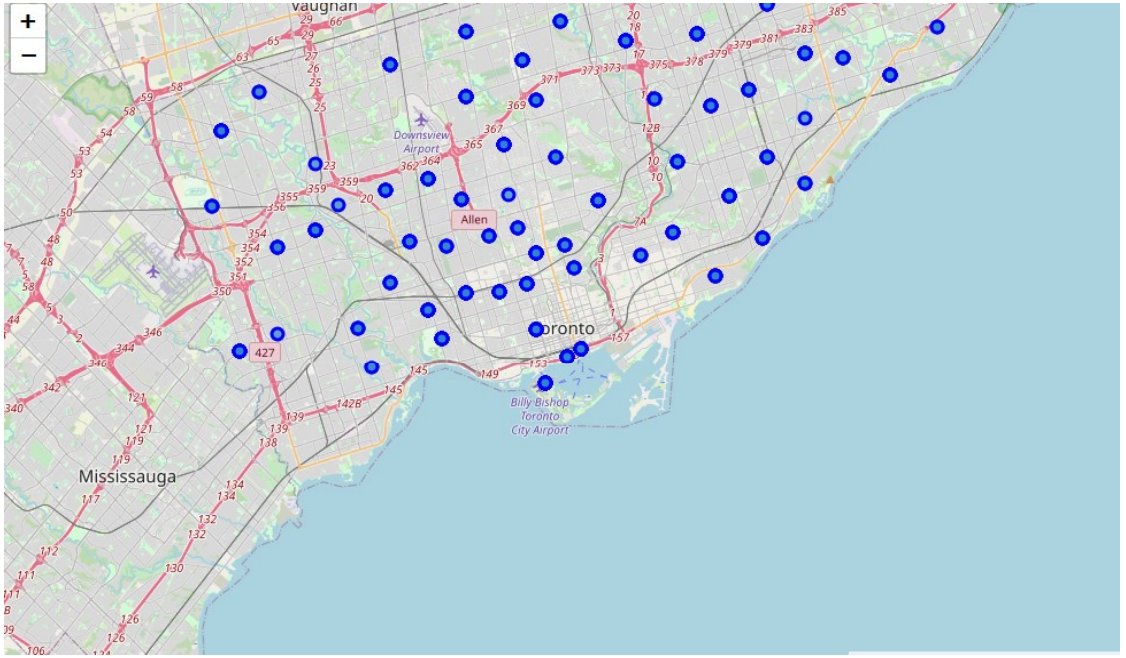
# WHERE TO START A NEW GYM BUSINESS IN THE CITY OF TORONTO

## Business Problem

Which of the neighborhoods of Toronto have no gym



## Data Sources

- Wikipedia page

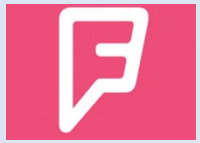
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

- Csv geographical coordinates file

http://cocl.us/Geospatial_data

- FourSquare API

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}



# Methodology

Scraped Wikipedia page

Wrangled the data

Cleaned it

Read it into a pandas dataframe

Merged Coordinates

FourSquare API for gym venues

Modeling

Visualization

| | Cluster Labels | Neighborhood | Neighborhood Latitude | Neighborhood Longitude |
|---|---|---|---|---|
| **0** | 4 | Agincourt | 43.794200 | -79.262029 |
| **1** | 4 | Agincourt | 43.794200 | -79.262029 |
| **2** | 4 | Agincourt | 43.794200 | -79.262029 |
| **3** | 4 | Agincourt | 43.794200 | -79.262029 |
| **4** | 2 | Bathurst Manor, Wilson Heights, Downsview North | 43.754328 | -79.442259 |
| **...** | ... | ... | ... | ... |
| **753** | 4 | Woodbine Heights | 43.695344 | -79.318389 |
| **754** | 2 | York Mills West | 43.752758 | -79.400049 |
| **755** | 2 | York Mills West | 43.752758 | -79.400049 |
| **756** | 2 | York Mills West | 43.752758 | -79.400049 |
| **757** | 2 | York Mills West | 43.752758 | -79.400049 |

# Modeling

- Cluster Algorithm
  - K-means

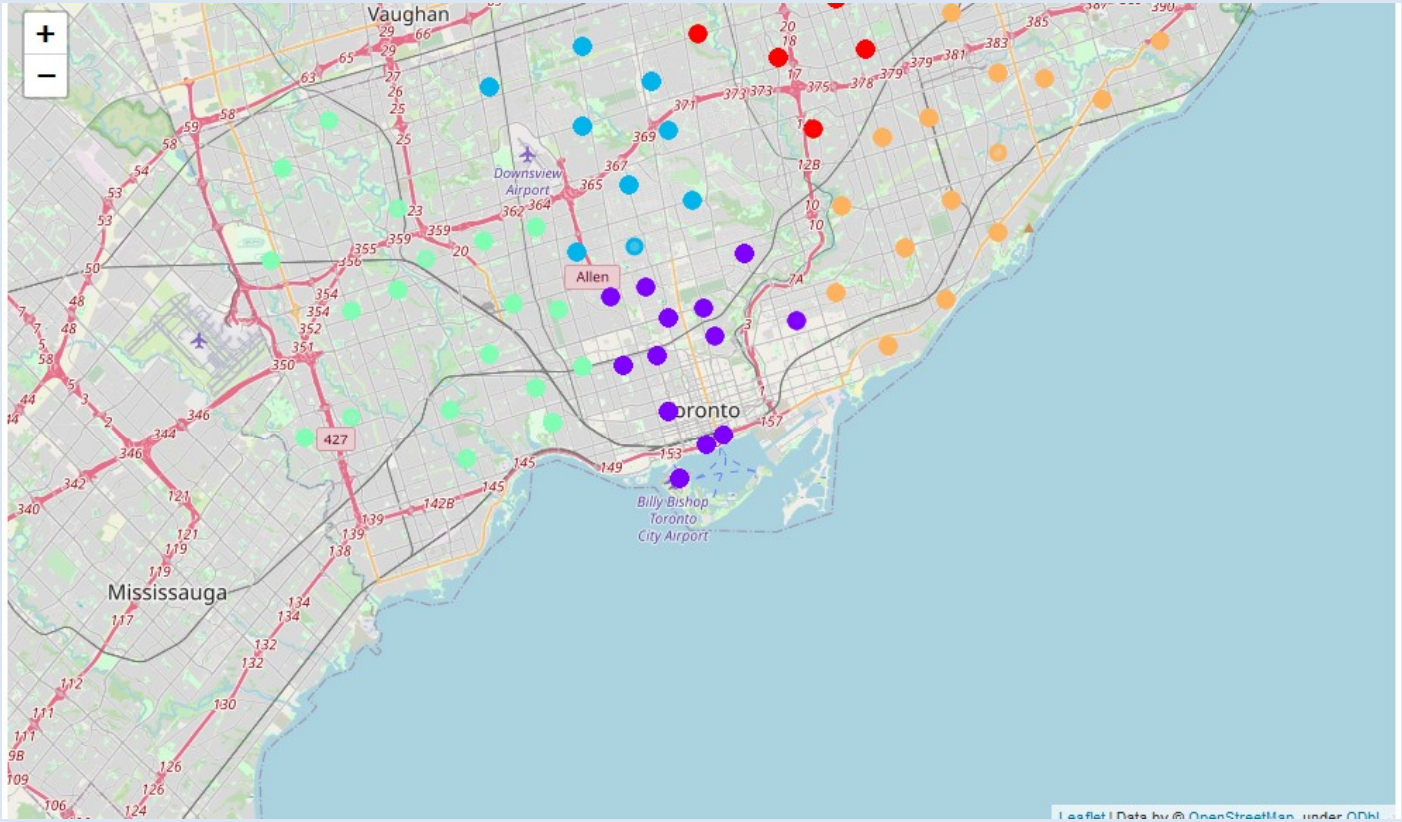
```
#### set number of clusters
k = 5
toronto_clustering = plot_gym.drop('Neighborhood', 1)

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

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([4, 4, 4, 4, 2, 2, 2, 2, 2, 2], dtype=int32)
```

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

```
plot_gym
```

# Clusters on map



# CONCLUSION

- Most people usually prefer to go to a gym that is close to their work or their home.
- The criteria for opening a new gym here depend on the lack of a gym in the neighborhoods
- We should also take into account the Business Improvement Area (BIA) (association of commercial property owners and tenants within a defined area who work in partnership with the City to create thriving, competitive, and safe business areas that attract shoppers, diners, tourists, and new businesses)
- The density of the population in each region etc
- We should consider more data to reinforce our choices