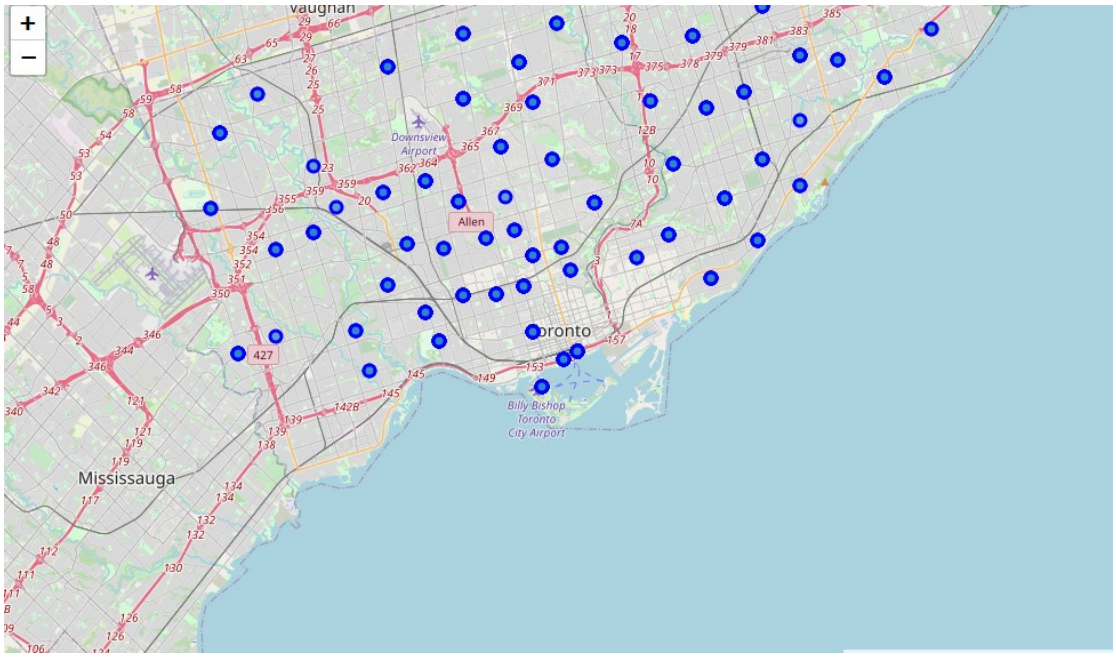
WHERE TO START A NEW GYM BUSINESS IN THE CITY OF TORONTO

BUSINESS PROBLEM-INTRODUCTION

The city of Toronto is the most populated city in Canada and one of the most populated cities in North America. As it is a very developed city, so is the business case. That's why any new investment that wants to move to Toronto needs to use market-based insights that will help them understand the business environment, allowing for a strategy to reduce risk. In this case I am exploring the suitable neighborhood in Toronto for a new gym.



DATA SECTION

1. A Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
2. Csv file that has the geographical coordinates of each postal code: http://cocl.us/Geospatial_data
3. FourSquare API to collect data about locations of gyms in the city of Toronto. https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

METHODOLOGY

First I have scraped Wikipedia page and wrangled the data, cleaned it, and then read it into a pandas dataframe. My main target here was to access which neighborhood in the city has no gym so there it is more possiple that it will attract customers. I used the FourSquare API through the venues channel. I made clusters of neighborhoods using kmeans so the future investor will choose between them. I used follium library to visualize the neighborhoods on the Toronto map.

```
: ##### 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)
```

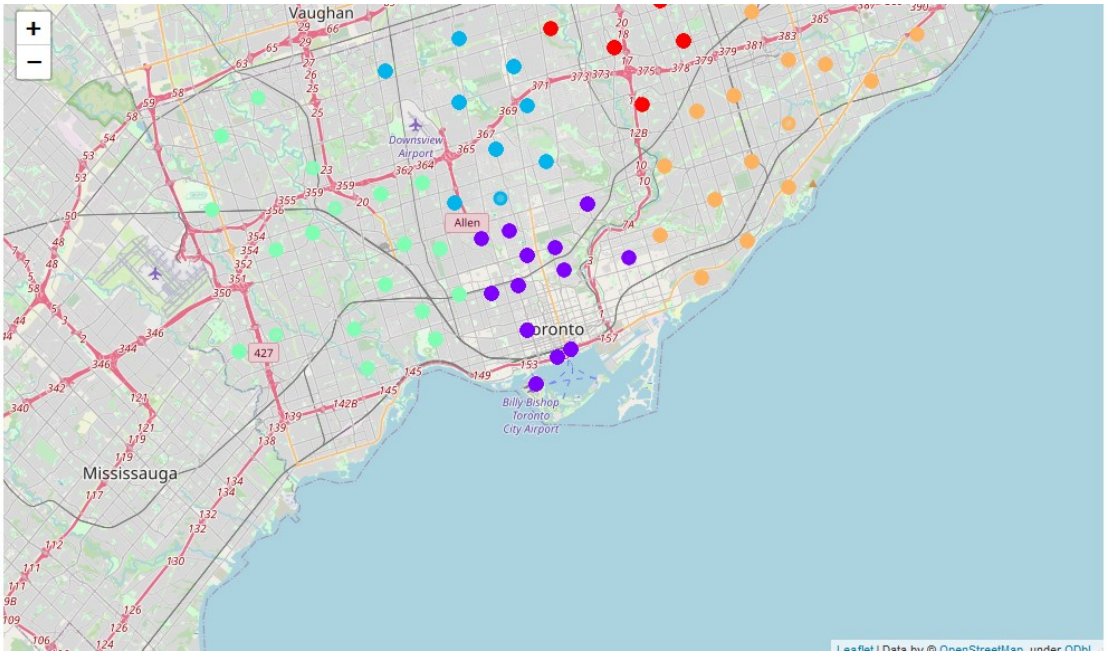
```
: # add clustering labels
  plot_gym.insert(0, 'Cluster Labels', kmeans.labels_)
```

```
: plot_gym
```

	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

RESULTS

I created 5 regions in the city of Toronto with multiple neighborhoods without a gym.



RECOMMENDATIONS

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 but 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.

https://ckan0.cf.opendata.inter.prod-toronto.ca/dataset/business-improvement-areas
https://www.arcgis.com/apps/webappviewer/index.html?id=1535b9fca54f46b3954bca6aaf3ab3f5

CONCLUSION

We can rely on the results quoted before but also later we should consider more data to reinforce our choices.