IBM Applied Data Science Capstone

Opening a Gym in Toronto

BY ABHIJITH V

Introduction

❑ To analyse and select best locations in Toronto to open a new Gym.

❑ Where and How would you recommend that they open it?

❑ Regional clustering of venue information

Data

❏Data required

- ➢List of neighborhoods in Toronto

- ➢Latitude and longitude coordinates of the neighborhoods

- ➢Venue data, particularly data related to Gym

❏ Sources of data

- ➢Wikipedia page for neighborhoods

- ➢Geocoder package for latitude and longitude coordinates

- ➢Foursquare API for venue data

Methodology

❑Web scraping Wikipedia page for neighborhoods list of Toronto

❑Get latitude and longitude coordinates using Geocoder Package

❑Use Foursquare API to get venue data

❑Group neighborhood data by each venue category

❑Filter venue category by Gym

❑Perform clustering on the data by using k-means clustering

❑Visualize the clusters in a map using Folium

Results

❑ Categorized the neighborhoods into 3 clusters :

➢Cluster 0: Neighborhoods with highest number of Gyms

➢Cluster 1: Neighborhoods with least number to no existence of Gyms

➢Cluster 2: Neighborhoods with second highest concentration of Gyms

➢Cluster 3: Neighborhoods with less number to no existence of Gyms
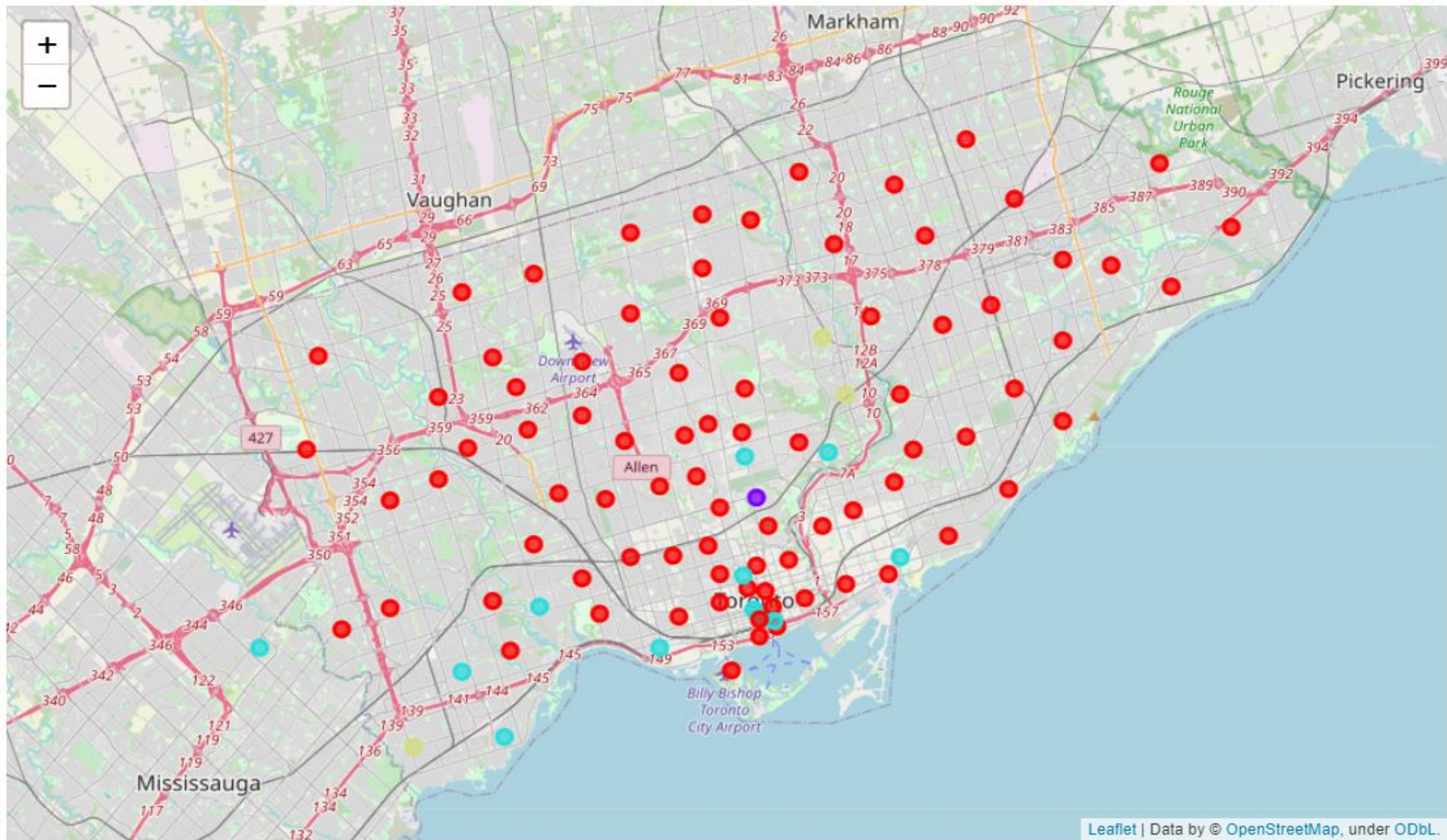
Recommendations

❑ Open new Gyms in neighborhoods in cluster 1 with little to no competition

❑ Can also open in neighborhoods in cluster 3 with moderate competition if have unique selling propositions to stand out from the competition

❑ Avoid neighborhoods in cluster 0, already high concentration of shopping malls and intense competition.

Conclusion

❑ The neighborhoods in cluster 1 are the most preferred locations to open a new Gym in Toronto

❑ Findings of this project will definitely help the business owners and entrepreneurs to start a new Gym or establish a new gym outlet.