Data Science Project_JQL

Introduction

Toronto is a fast-growing city in Canada, more and more people plan to start their own business there. Hypothetically, there might not so many gyms in Toronto, an entrepreneur probably wants to take the good opportunity to start his own gym in Toronto. For the entrepreneur, finding a suitable location to open a gym is a big decision. The project aims to create analysis and provide solutions for the business problem: where is the most suitable location to open a gym for an entrepreneur in the Toronto area?

Data

To analyze and solve the business problem, the data we will use is as follows,

1.https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
http://cocl.us/Geospatial_data

2. Foursquare API Data

Methodology

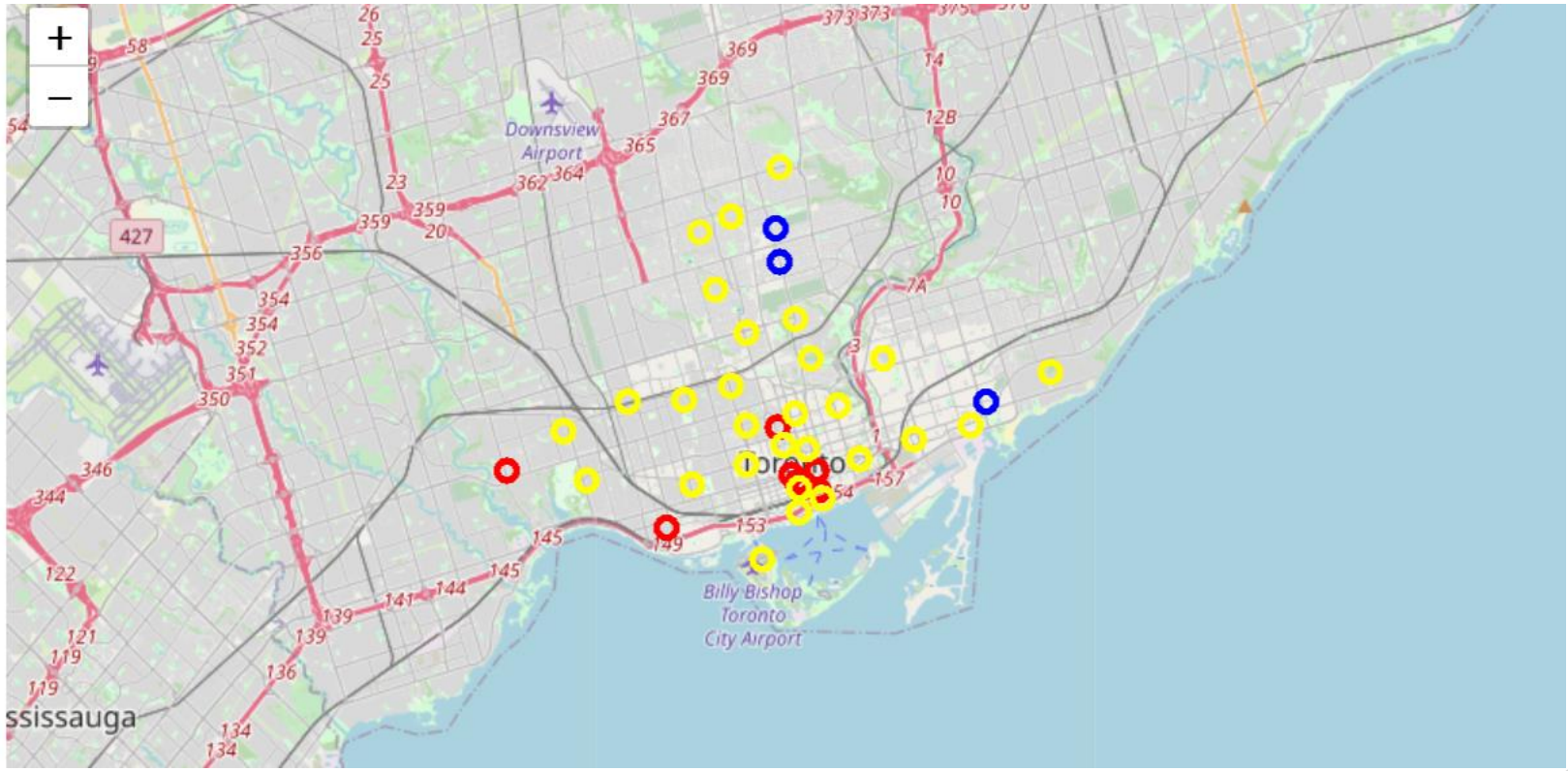
The K-means clustering algorithm

The neighborhoods in the Toronto area were categorized into 3 clusters.
They are cluster 0 (red), cluster 1 (yellow), cluster 2 (blue)

Results

See the 3 clusters on the map.

[56]:



Discussion

In cluster 0 (red), the neighborhoods have a normal number of gyms. In cluster 1 (yellow), the neighborhoods have the lowest number of gyms. In cluster 2 (blue), the neighborhoods have the highest number of gyms. Most of the gyms are near Davisville and Davisville North. I would suggest opening a gym in cluster 0 (red) or cluster 1 (yellow), also, not too far from downtown, some places like Church and Wellesley, Harbourfront, St. James Town, and Stn A PO Boxes could be good choices.

Conclusion

This project is to give suggestions for people who would like to open a gym in the Toronto area. Some methods or tools like web scraping, Foursquare API, Folium are utilized in this project. Also, by using the K-means clustering algorithm, the neighborhoods of Toronto area are separated into 3 different clusters for exploring and analyzing the data.