What best place to start a new healthy food restaurant in Toronto?

Khadija Bouagoun

April 2021

1 Introduction

1.1 Problem Statement

Toronto is one of the biggest cities in Canada, and it is the most populous Canadian city, and offers many Economic Opportunities! Also If you're looking for a career in arts, culture or tech ... , this city may be a great choice for you.

However, the percent of adults that had an overweight/obese status increased significantly in Toronto from 2007 to 2014, and it is always increasing until now! In fact, it is due to many reasons including the lack of restaurants which provide healthy food!

Normally to open a restaurant you must guaranty accessibility and visibility, but it is hard to do that manually, whereas if it is done with ML it will requires just a dataset and some processing on it !

1.2 Target Audience

Who will be more interested in this project ?

• people suffering from obesity.

• Many business man, investors and entrepreneurs (because as we said before, Toronto offers many Economic Opportunities)

• because Toronto is characterized by Multicultural Diversity, this project will be so profitable for people used to consummate healthy food as part of their culture (people from Sweden as example).

2 Data acquisition and cleaning

2.1 Data Sources

To recommend a good location for a new healthy food restaurant,

- I will use a dataset of Toronto city which contains information about its Boroughs and Neighborhoods (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

- We will also utilize "https://cocl.us/Geospatial_data" csv file to get all the geographical coordinates of the neighborhoods.
- I will also use Foursquare API to get data location and other information in Toronto.

2.2 Data cleaning

- I scrapped data from the following Wikipedia page : "List of Postal code of Canada: M" to make a dataframe which consists of three columns: PostalCode, Borough, and Neighborhood.
- Download geographical coordinates to the neighborhoods and merged it with the dataframe above into one using the postal code.
- Gather venues data from Foursquare API for each neighborhood within a radius of 500 m.

3 Exploratory Data Analysis

- Extract the following venue categories : Vegetarian / Vegan Restaurant, Gym / Fitness Center, Gym, Yoga Studio, which appears more significant.
- I used Foluim library to draw an interactive map based on Neighborhood Latitude, Longitude and Venue Category .
- I used one-hot encoding to code Venue Category values, and then I insert into it a column which contains neighborhoods.
- Then I grouped the latest data by neighborhoods and turn our summarized values into percentages.

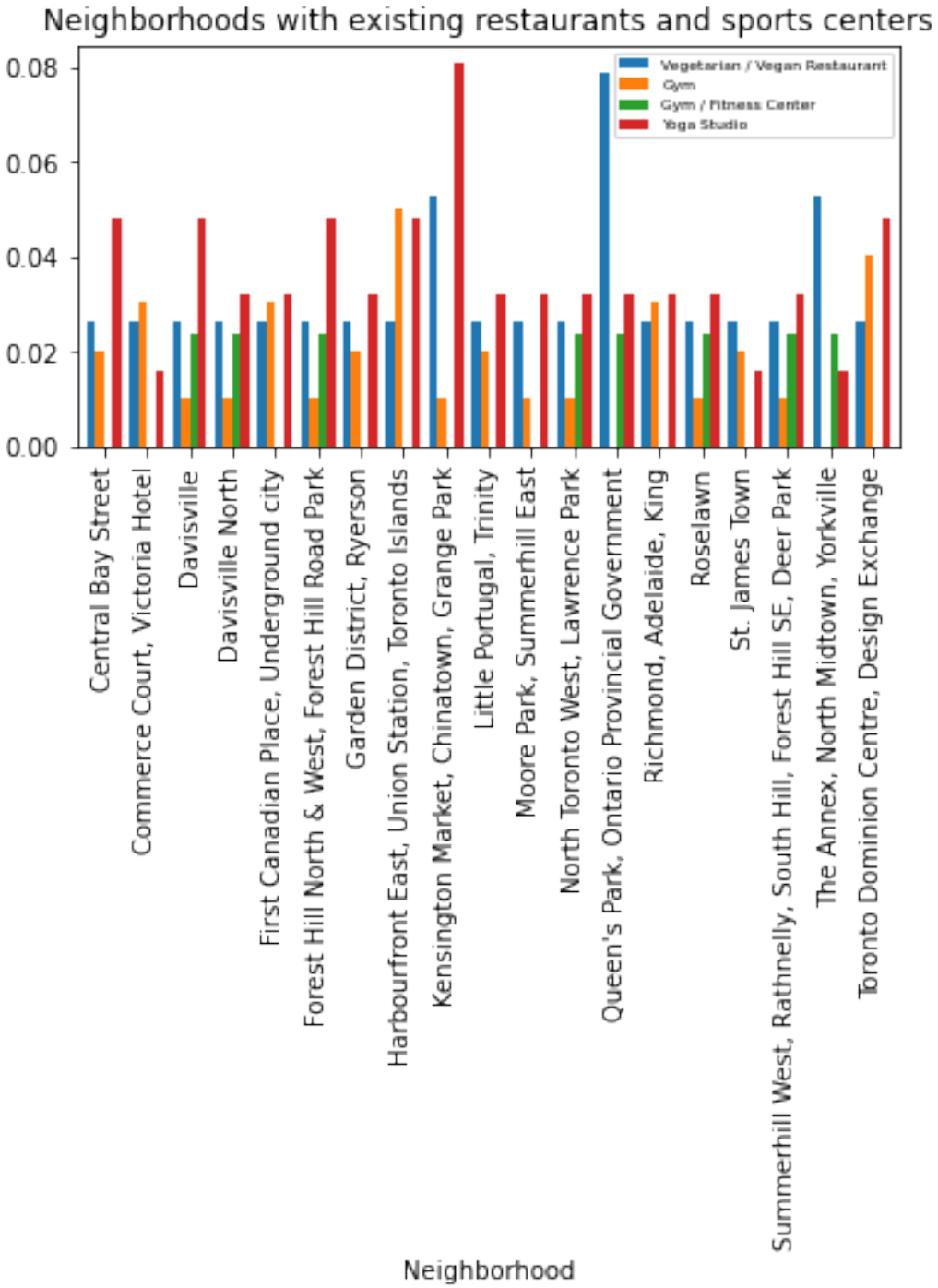

Figure 1: Neighborhoods with existing restaurants and sports centers

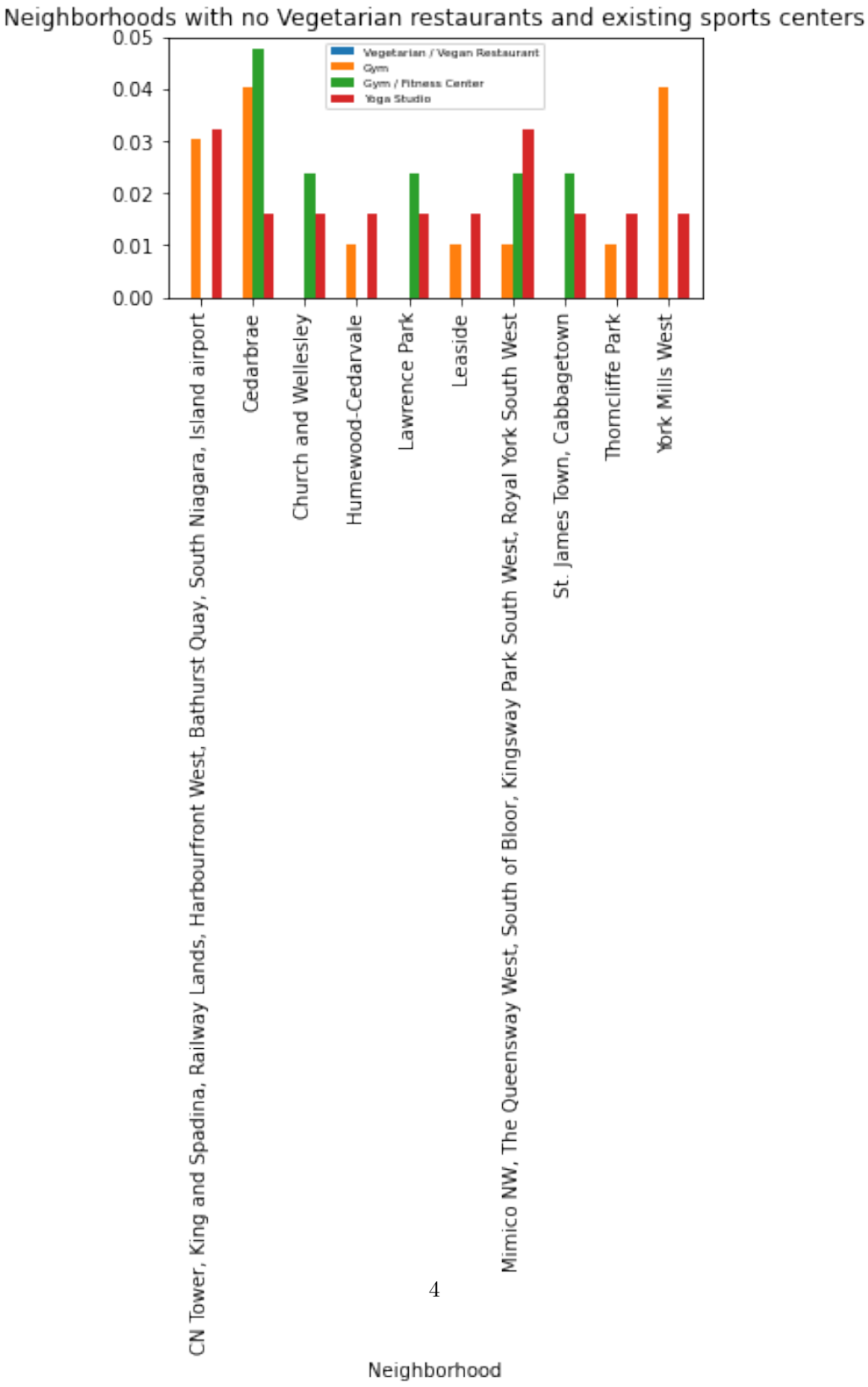

Figure 2: Neighborhoods with no Vegetarian restaurants and existing sports centers

4 Predictive Modeling

For segmenting neighborhood in Toronto based on existed vegan restaurants, gym, fitness and yoga, and finding recommendation for a good start of restaurant business, clustering algorithm will be used especially K-mean method.

Because I did my predictions with KMeans clustering, the first step was to determine k : the number of clusters in our dataset with elbow method.

I got k = 4 .

5 Results and Discussion

5.1 Results

We obtained the following four clusters :

- Cluster 0 contains neighborhoods with high number of Vegetarian / Vegan Restaurants and less number of Gym and yoga centers (red).
- Cluster 1 contains neighborhoods with high number of Gym/ Fitness Centers and fewer Vegetarian / Vegan Restaurants (purple).
- Cluster 2 contains neighborhoods which are densely populated with Gym and yoga centers, the same as the number of Vegetarian / Vegan Restaurants (blue).
- Cluster 3 contains neighborhoods with a good number of Gym and yoga centers, and has least number of Vegetarian / Vegan Restaurants (yellow).

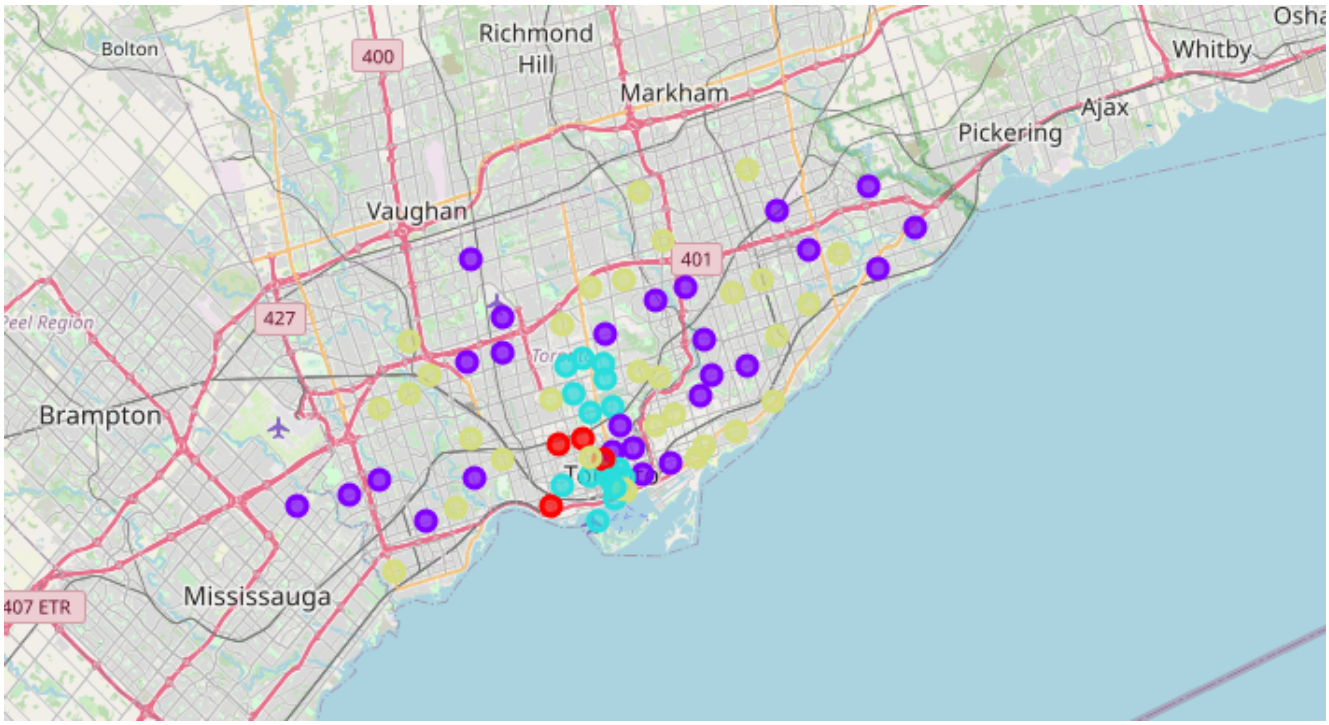

Figure 3: the four resulting clusters

5.2 Discussion

According to these results I concluded that adding more features such as the population and number of parking and the existence of a high foot or car traffic will help us to make more perfect results, also other methods of unsupervised learning can be used such as recommendation systems.

5.3 Conclusion

As the statistics of our clusters shows, from 80 neighborhood, Downsview, Cedarbrae and York Mills West from cluster 1 and 3 can be a good location for our new restaurant, because there is significant number of Gym/ Fitness and yoga centers, and we know that people who are registries there have often tendency to follow a diet and eat healthier, also there is less number of Vegetarian / Vegan Restaurants (a good example of healthy food restaurants) which gives lower competition.

Thus, this region could potentially be a perfect place for starting quality healthy food restaurants.