

Best location to open a gym in Toronto



CHOUVATTANAK ENG

Introduction



- Targeted audiences of this project would be the business people who want to open a new gym or expand their franchised
- Through this study, they will have a clear overview of the locations in Toronto and can confidently target their specific clients.
- The question in this study will be which location in Toronto is the optimum point?

Data acquisition and cleaning



- We mainly focus on 4 data sources in this instance
 - Wikipedia: Postal code, borough and neighborhoods in Toronto.
 - Toronto Geospatial Data
 - Foursquare API: Information of venues in Toronto
 - Toronto Census data: Total population and household income in the neighborhoods in Toronto
- Below procedures are summary the data cleaning and wrangling process:
 1. Pulling data from data sources.
 2. Drop row and column based on data quality
 3. Mapping all data into one table
 4. Prepare data by selecting and applying standard scaling to the features that will be used in the K-mean clustering model

Expected Data Table



- Below is the expected table that will be used in modeling:

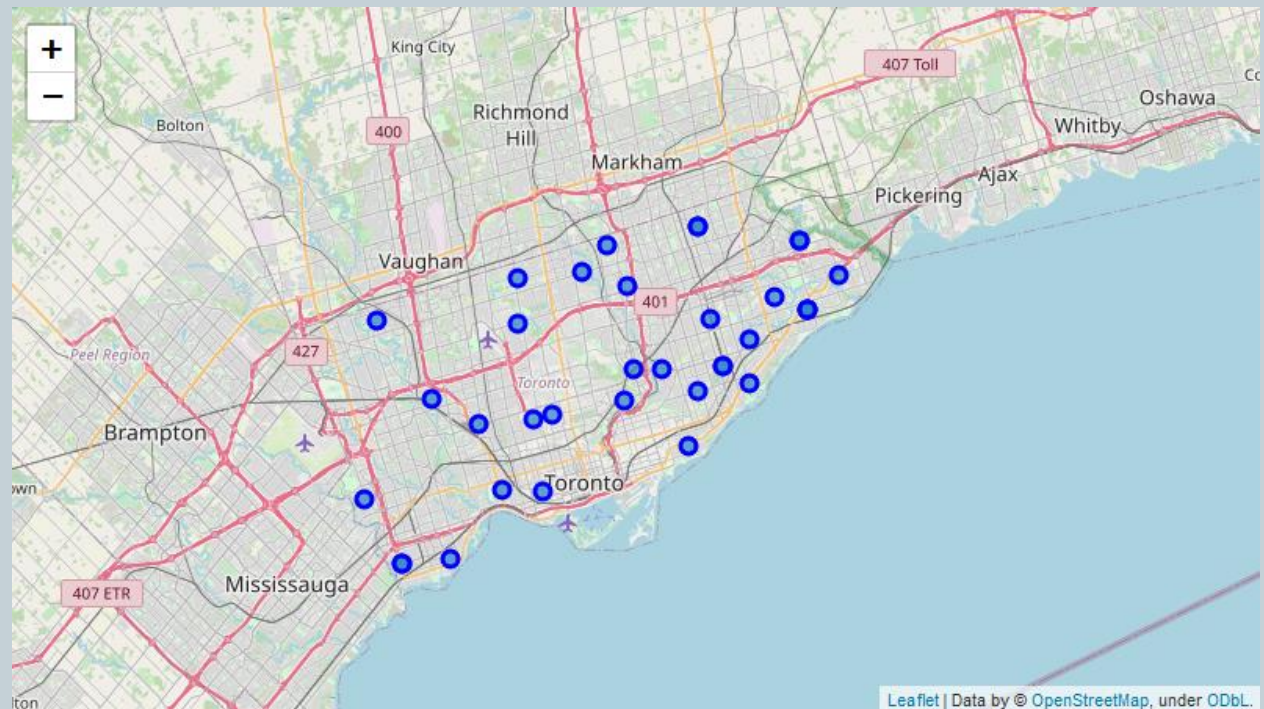
PostalCode	Borough	Neighborhood	Latitude	Longitude	Total Population	Average Family Income	No. of Gym Center
M4A	North York	Victoria Village	43.725882	-79.315572	17510.0	65104.0	11.0
M1B	Scarborough	Rouge	43.806686	-79.194353	46496.0	86997.0	2.0
M1B	Scarborough	Malvern	43.806686	-79.194353	43794.0	64497.0	2.0
M1C	Scarborough	Highland Creek	43.784535	-79.160497	12494.0	98857.0	2.0
M3C	North York	Flemington Park	43.725900	-79.340923	21933.0	55824.0	13.0

Methodology



- the samples are roughly 40 samples and 3 to 4 features will be selected based on quality and its relation to the study.
- K-mean clustering will be used to modeling in this study.
- Python libraries are required to facilitate the modeling.

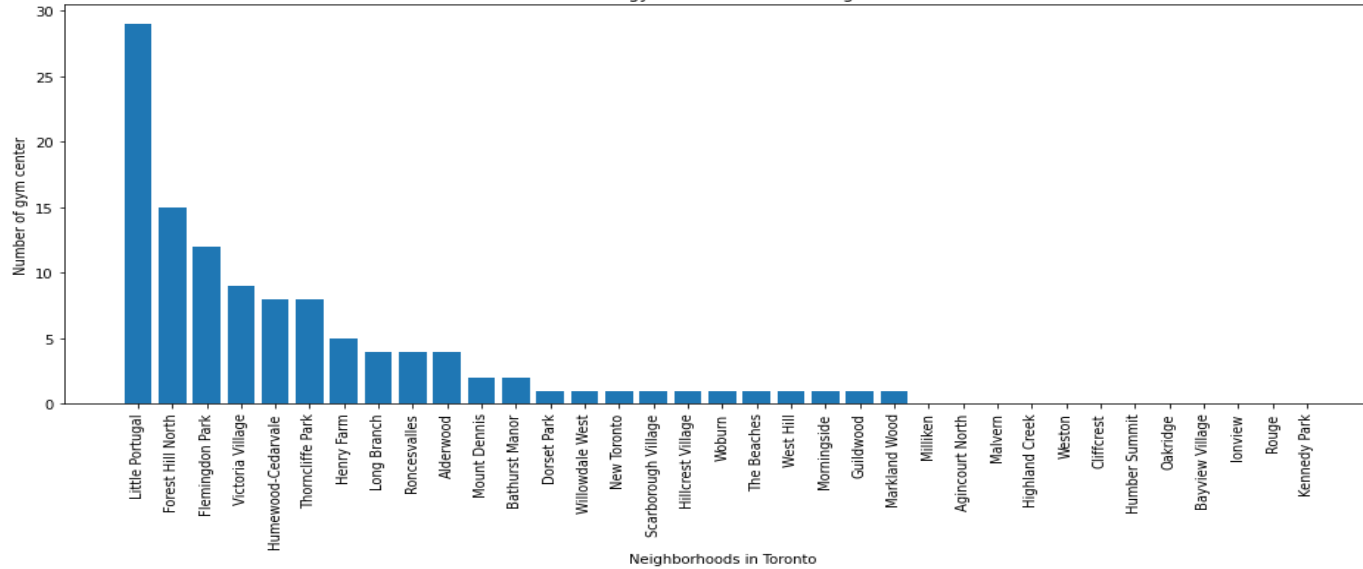
Toronto's
neighborhoods



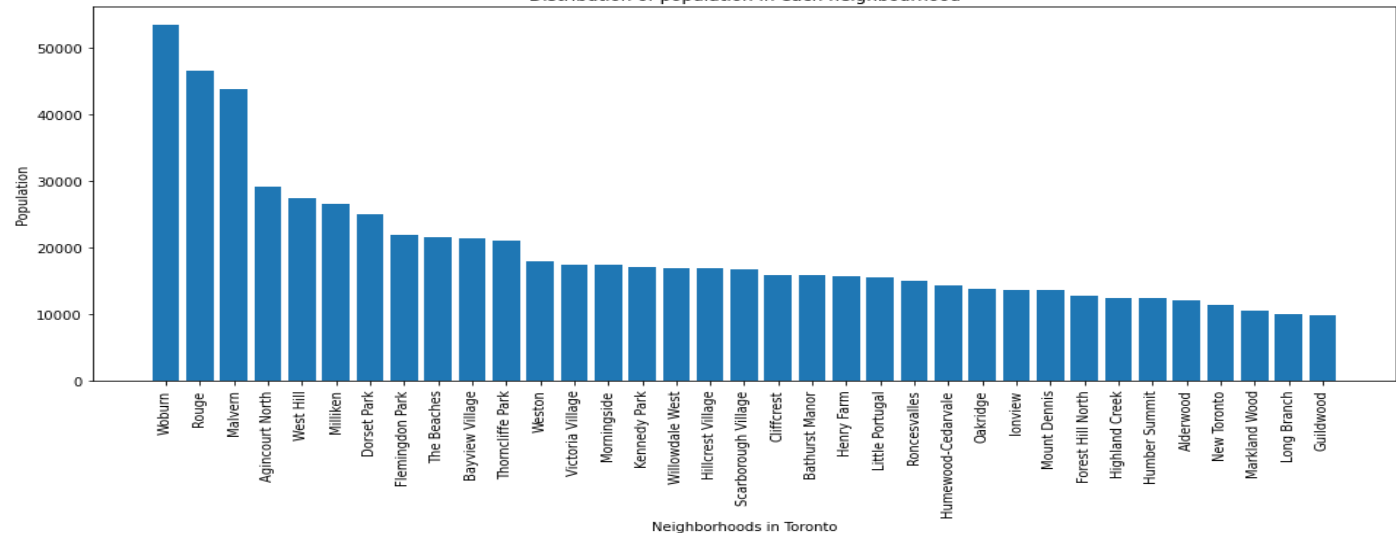
Exploratory Data Analysis



Distribution of gym centers in each neighbourhood



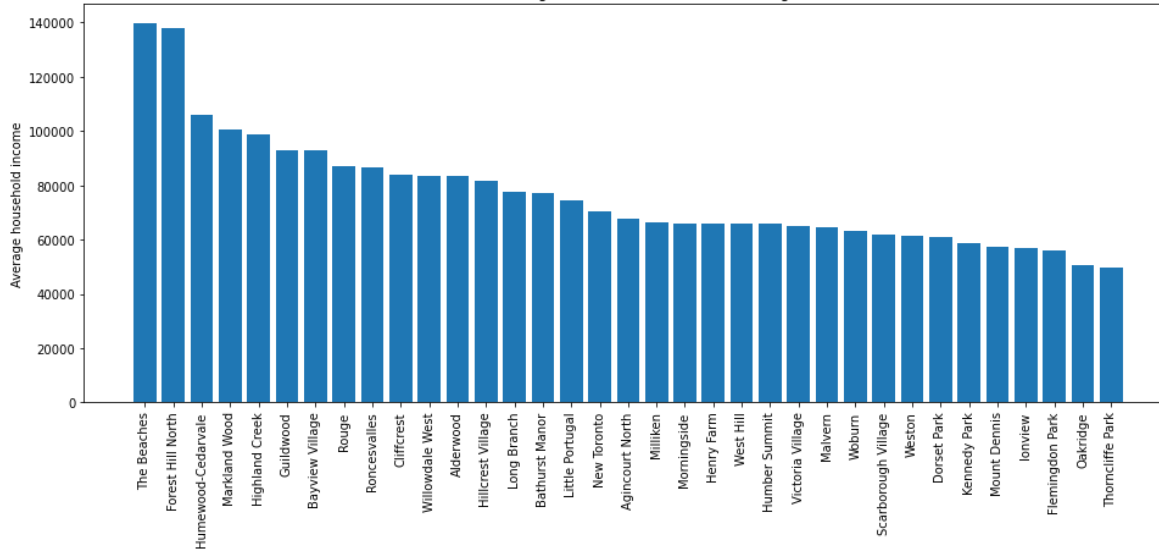
Distribution of population in each neighbourhood



Exploratory Data Analysis

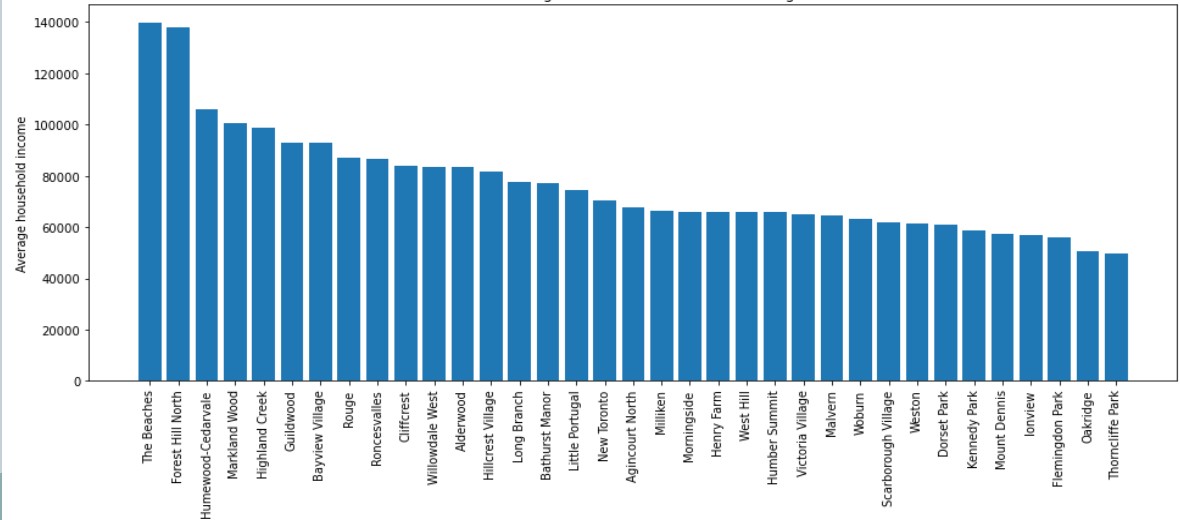


Distribution of average household income in each neighbourhood



Neighborhoods in Toronto

Distribution of average household income in each neighbourhood



Neighborhoods in Toronto

Data Preprocessing



- Features selection and apply standard scaling for K-mean modeling

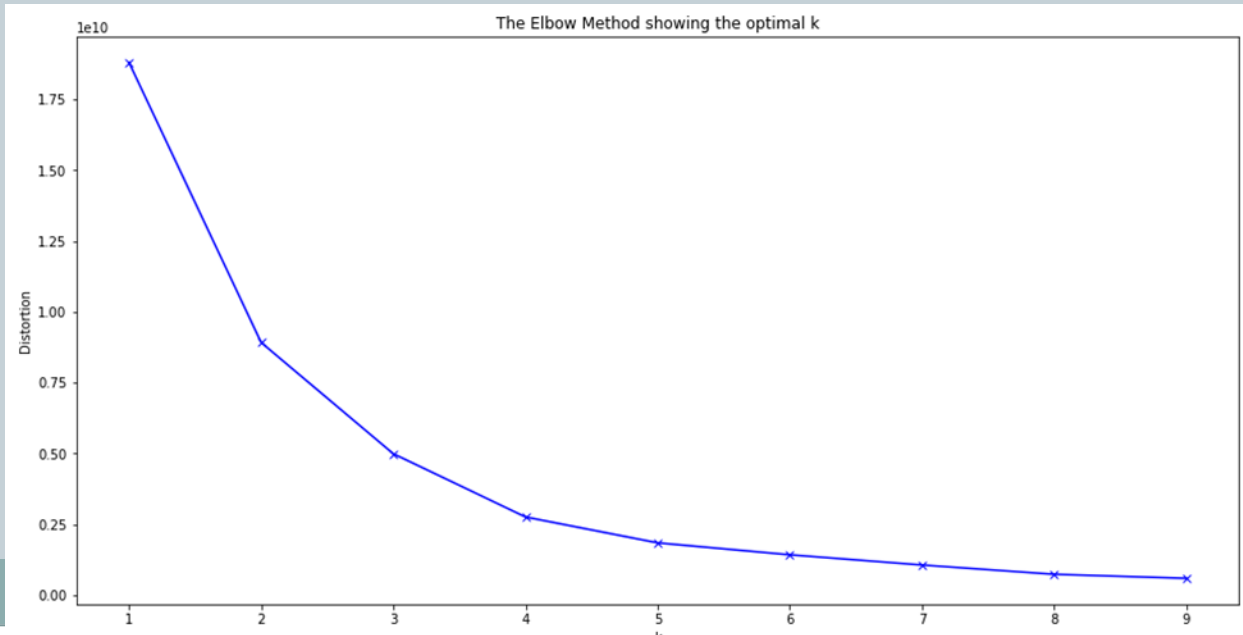
```
array([[ -0.20307124, -0.54406583,  1.00853868],  
       [ 2.69944643,  0.5026181 , -0.56418253],  
       [ 2.42888124, -0.57308594, -0.56418253],  
       [-0.70534919,  1.06963365, -0.56418253],  
       [ 0.23982656, -0.98773398,  1.53277908]])
```


Data Preprocessing

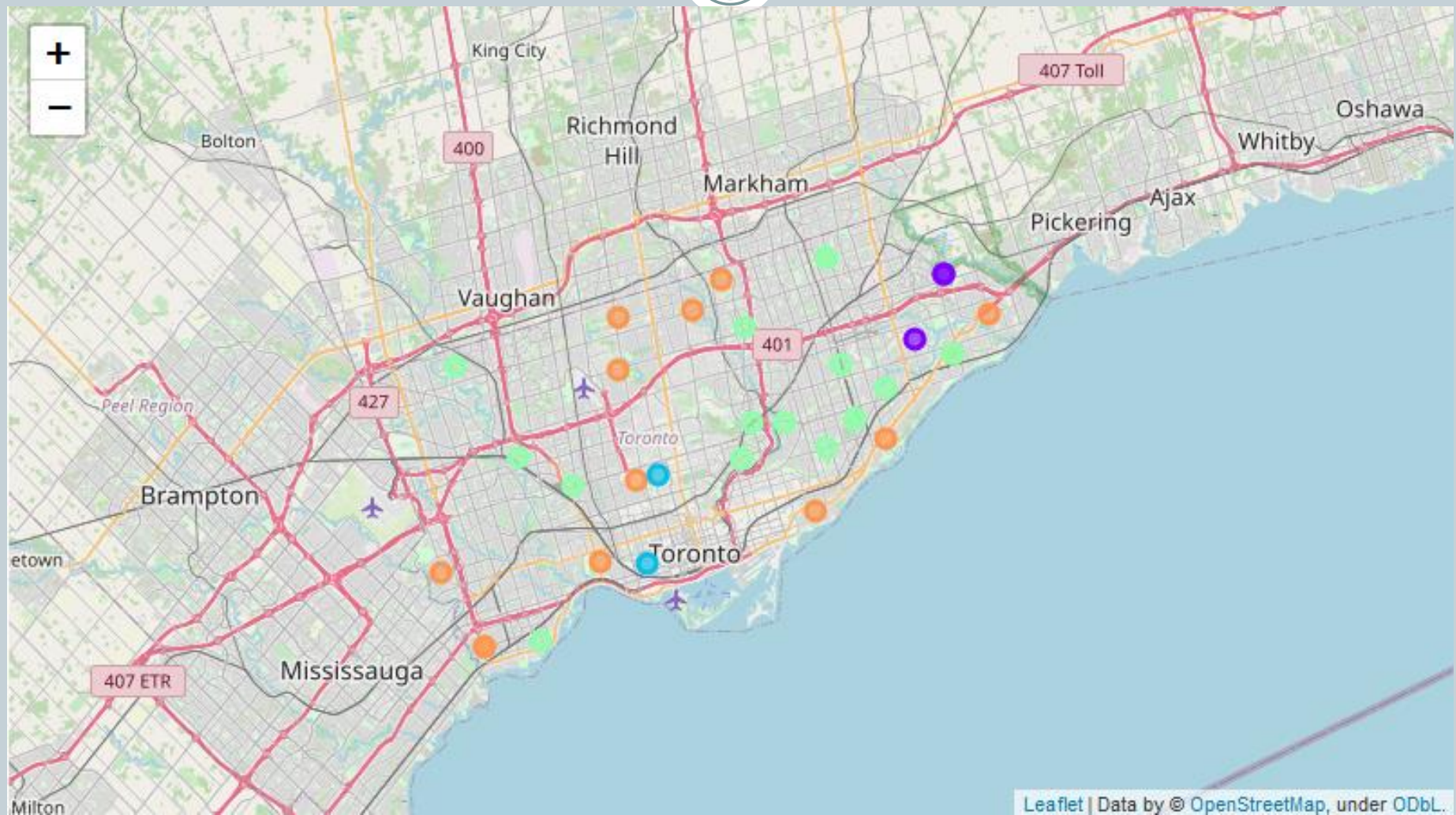


- Features selection and apply standard scaling for K-mean modeling

```
array([[ -0.20307124, -0.54406583,  1.00853868],  
       [ 2.69944643,  0.5026181 , -0.56418253],  
       [ 2.42888124, -0.57308594, -0.56418253],  
       [-0.70534919,  1.06963365, -0.56418253],  
       [ 0.23982656, -0.98773398,  1.53277908]])
```



Cluster Labels



Summary Clusters



- In summary:
 - Cluster 1: With high to mid population and spending power, with few competitors in the area signifies the lack of interest from population. But with good marketing campaign of healthy lifestyle, there could be a lot of potential customers in these neighborhoods.
 - Cluster 2: Low population in the area could be the concern of return of investment. Thus, investment in these neighborhoods wouldn't be a good choice.
 - Cluster 3: High competitors and spending power shows that these neighborhoods are premium customers even if the population is relatively low, because of the spending power, 1 customer could be a member of multiple gym centers.
 - Cluster 4: poses high barrier of entry with high competitors in the neighborhood and low population. Thus, customers in these neighborhoods would be cautious when choosing the gym of their choice.

Conclusion



- the most promising label would be in Cluster 1, as it has high population, mid spending power and fewer competitions give us opportunity to penetrate these neighborhoods.
- Another interest cluster would be cluster 3, as it has a lot of premium customers. if we could franchise the best brand, these high spending customer would surely choose our gym from competitors.
- In conclusion, this study is relatively useful as location selection is very important in the success of the business.
- We can also use these study as the blueprint for select the best location for other business as well.