

Toronto Best Neighborhood for sports lovers and potential investors

Capstone Project of IBM Data Science

Lishan Yu

1. Introduction

In this project, the best neighborhood will be found in Toronto for sports lovers and potential investors. The selected neighborhood will include gym, yoga club, soccer field, hockey arena, etc.

2. Data

Based on the purpose, the following data will be needed:

1. Toronto neighborhoods map
2. Kinds of sports places in each neighborhood
3. Total numbers of sports places in each neighborhood

FourSquare will be used for the geo service.

Methodology and Discussion

Part One: Find out all the neighborhood in Toronto.

The data is coming from:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (Wikipedia Site)

Part Two: Collecting data from “FourSquare” and reform the data

In this part, FourSquare API will be used to collect the information about all the types of venues in each neighborhood. Besides, the foursquare will gather all the venues required within 1km from the center of the neighborhood.

Part Three: Analyze the reformed data and set up a new data frame

After the data has been fully collected and reformed, the types of the sports places and similar places will be sorted out in details. After this step, the dataset is ready to process the final step.

Part Four: Machine Learning with K-means clustering.

Under this stage, the neighborhood were dividend to 5 clusters. And the best neighborhoods for sports lovers will be those with highest scores, while the best neighborhoods for potential investors will be those with relatively low scores since those place have more chance to build new sports places. The decision equalization used is $SUM\ SCORE = NUMBER\ OF\ SPORTS\ PLACES + RATING$

Result and Conclusion

To conclude, G3, G4 and G5 are best places for sports lovers, while G1 and G2 are the best neighborhoods for potential investors.