Help mover to find similar area

Peter Boktor

Problem Overview

Many people like to stay in places with certain features and when they have to move they struggle with finding similar place
This project will use Machine learning clustering algorithm techniques along with real data from Foursquare API to quantify and provide guidance to movers from New York city, USA to the city of Toronto, Canada

Data Overview

For this project the Foursquare API will be used along with A list of neighborhoods in New York and Toronto is downloaded with location in longitude and latitude coordinates
New York neighborhoods: https://ibm.box.com/shared/static/fbpwbovar7lf8p5sgddm06cgipa2rxpe.json
Toronto neighborhoods: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
The data downloaded are the neighborhoods located in New York and Toronto and will be determined based on the frequency of the categories found in the neighborhoods.

Methodology

feature extraction

Each feature becomes binary, this means that 1 means this category is found in the venue and. Then, all the venues are grouped by the neighborhoods, computing at the same time the mean. This will give us a venue for each row and each column will contain the frequency of occurrence of that particular category.

Unsupervised Learning

K-Means is a clustering algorithm.

TRONTO MAP
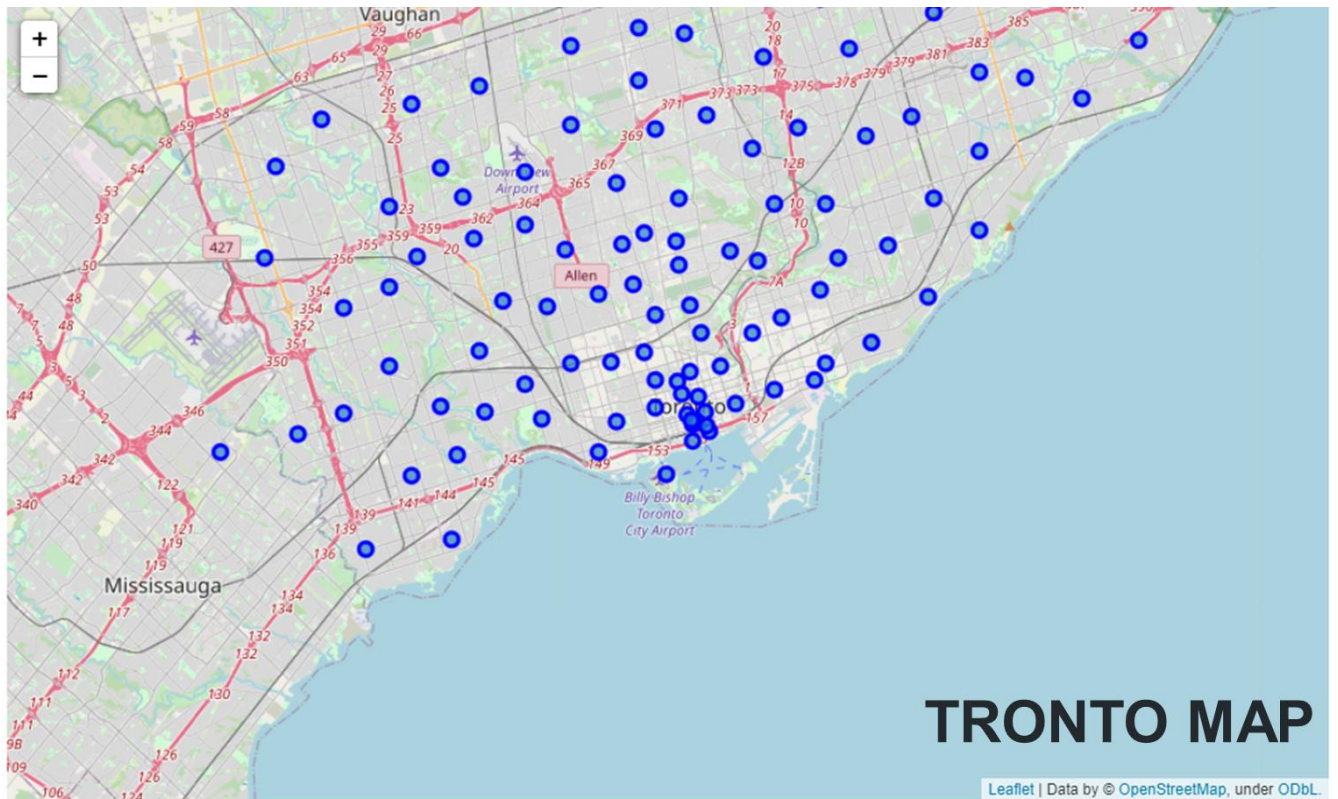
Conclusion

The K-Means clustering algorithm is used for finding similarities between the neighborhoods.
the K selected is 5.
Results show as sown in the figure
Neighborhoods that have around parks, bus lines and sandwich places.

Neighborhoods that have around parks, playgrounds and trails.

Neighborhoods that have around coffee shops, pubs and Italian restaurants.

Neighborhood that have around gardens.

Neighborhoods that have around coffee shops, parks and bakeries.