Neighborhood Recommender System: Comparing Neighborhoods Between Cities

Hasmik Grigoryan

August 2019

This project was completed for IBM Data Science Professional Certificate specialization on Coursera.[2] This serves as my final project for the Capstone. I will explore neighborhoods in different cities and create a recommender system for neighborhoods in a city of interest based on preferences in a different city. For the start, the project will support neighborhood comparisons between New York, Chicago and LA. However, I will make the code easy to scale and generalize and add more cities over time.

Introduction / Business Insight

Most big enough cities have similar sets of neighborhoods; the homey ones, the ones great for night life and entertainment, the ones with the most museums, etc. In my experience, it takes some time before you figure out what your favorite ones are when moving into a new city. For this reason, I will try to make a recommender system for neighborhoods. I will be exploring a few different methods for this: content-based recommendation and clustering.

While this tool may be particularly useful to individuals who are looking to relocate, it can be of interest to realtors and property management companies. For instance, a realtor can use this tool to provide better recommendations for their clients. Property management companies may use the content for better targeted advertising and for analyzing their clientele's preferences.

Data

For this project I will be mainly using Foursquare API [1], the list of neighborhoods scraped from Wikipedia [3] and Geopy for address search. Foursquare provides an easy way for exploring venues in proximity of given locations. These venues come with a lot of accompanying data, including categories, ratings, and names. I will be using this data for feature engineering per neighborhood. For instance, I can design categorical features that indicate most common category

of venue within the neighborhood. Similarly, we can look at the categories with highest cumulative ratings. The Wikipedia data will be scraped for neighborhood list using Pandas.

My attempt is to make this tool as general as possible, so that it can work for any city that Foursquare data is available for. However, a list of neighborhoods of a city will be necessary. For this reason, I will be making the tool work for LA, Chicago and New York. But the tool will work for any city when the data on neighborhoods in appropriate format is provided.

References

[1] Foursquare. https://foursquare.com/.

[2] Ibm data science professional certificate. https://www.coursera.org/professionalcertificates/ibmdatascience.

[3] Lists of neighborhoods by city. https://en.wikipedia.org/wiki/Lists_of_neighborhoods_by_city.