**Predicting the Perfect Neighborhood Fit**

Mearl

August 19, 2019

1. **Introduction**

Moving to a new apartment or home can be an extremely stressful and time consuming experience. There are many steps in the process and it is easy to feel overwhelmed. As a transplant moving across the country or even the world to begin university, start their career, or perhaps looking for a change of scenery it is difficult to know which neighborhood would feel the most like home. Alleviating the stress and worry of not sure where within a new city would be a good fit is beneficial to those moving, but it doesn’t stop there. It could also help realtors locate locations for their clients, better understand the market, provide local shops with some insights into their new residents, or even city planners when designing new or updating locations.

For purposes of this assignment I am planning to start small and help solve the problem of someone leaving their current home and moving to Brooklyn NY, a location with as much diversity among neighborhoods as its residents. The goal is to cluster each neighborhood and then predict the top 3 neighborhoods that best fit with the user’s preferences.

Being someone extremely new to machine learning as well as python in general I wanted to create a simple project. Please keep this in my when reviewing the work as I have tried my best!

1. **Data**

The main data source used in this analysis is Foursquare to pull the information about each neighborhood and their venues. Pulling out the types of venues and their frequency within a neighborhood paints a fair picture of the type of experience living there. Another source of data which for purposes of this assignment was created is the user specific information/preferences. We are going to compare the user preferences to the clustered neighborhoods. Please see below for some ideas on expanding the analysis for future iterations.

1. **Methodology**

In this assignment I used the Foursquare API to explore neighborhoods in Brooklyn, New York. I used the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters. Using the k-means clustering algorithm to complete this task and visualized Folium library to visualize the neighborhoods in Brooklyn and their emerging clusters. Once that was completed I then created a dummy user profile based on a listing of four priority preferences the user cared about. After visualizing the top 10 priority venues count based on neighborhood using matplot lib there was additional insight provided to help the user select the top 3 neighborhoods.

It was important to do some exploratory analysis and data preprocessing to take the format of the data from foursquare and manipulate it into some more useable for modeling purposes. Throughout the process it was critical to visualize the manipulations made either through viewing the data frame first few rows or by using visualization such as folium and matplot lib. Utilizing machine learnings for clustering was really helpful as well. The previous labs set the foundation for being able to replicate a similar analysis. I do not think I would have been able to result in the same outcome without using the machine learning techniques taught in the course. There was no use of inferential statistics such as ANOVA or Regression in this analysis, but on future iteration of this there should be some.

This study’s scope was intentionally small to prove it was possible to create a POC of what I had in mind. I am excited to expand and enrich what was already built in the future. I hope as my skills increase I can continue to solve complex and interesting problems using the learnings from this specialization.

1. **Results**

Upon final analysis it was evident based on the user listing of priorities/preferences the following 3 neighborhoods were a good start in selecting a potential apartment or home to move into:

* 1. Bushwick
  2. Gerritsen Beach
  3. Bedford-Stuyvesant

1. **Discussion**

As mentioned before, Brooklyn, NY is a diverse and large borough of NYC and it can be very stressful trying to determine which neighborhood can provide the best living experience for its residents.

I used the Kmean algorithm as part of this clustering study as well as dummy user data to use predictive modeling to create a content-based recommendation system. In future studies real user data should be used and maintained.

1. **Conclusion**

As result of using this recommendation system to help generate a starting point for users to select neighborhoods in which they want to live created a better moving experience and hopefully will expand to provide more accurate predictions as users begin to onboard.

This initial iteration was purposefully small scoped. In future iterations I believe it will be relatively easy to expand on certain aspects which are necessary to create a more accurate and complete recommendation model. Below is a non-exhaustive list of ideas to take forward:

* Residential demographic information
* Expanding model to account for more cities and neighborhoods
* Enriching the neighborhood information (e.g. venue ratings, public services, crime data etc.)
* Pooling all user information to provide collaborative based filtering on top of the initial content based approach and comparing results.