**Utilizing Foursquare to Determine the Optimal Location for Apartment Development in Seattle**

Nicholas Wapstra

**Introduction**

There is a huge influx of people moving to the Seattle area with the growth of major tech companies and startups. There is a need for more apartment living to house the growing population. Therefore, prospective apartment developers need to know the most resident friendly neighborhoods to build their housing. Ideal apartment living would be close to restaurants, public transportation, grocery stores, and parks or beaches. In order to make this determination, it would be helpful to group neighborhoods into clusters based on what offerings are nearby. Foursquare data can be leveraged to provide this information and determine which neighborhoods would be the best suited for apartment living.

**Data**

The data required to solve the problem is the list of Seattle neighborhoods, the latitude and longitudes of these neighborhoods, and venues that are nearby these neighborhoods. A list of Seattle neighborhoods was collected from findwell.com1 and Google Maps2 provided latitude and longitude coordinates. Foursquare3 was used for collecting data on locations of nearby venues using geographic coordinates. In total, there were 54 Seattle neighborhoods examined for apartment feasibility.

**Methodology**

The neighborhood and geographic location data were combined manually in .csv file. The .csv was then imported in to a Jupyter Notebook. The latitude and longitude were gathered from the geopy.geolocater tool. A map of Seattle was generated using Folium and all of the neighborhoods were represented by points on the map. A function called “getNearbyVenues” was defined to take in a list of neighborhood names, latitudes, and longitudes and output a list of nearby venues. The category types of venues were checked for accuracy. One hot encoding was used to generate a list of categories for nearby venues and dummy values were inserted to stand-in for whether or not a venue type was present in that neighborhood. Frequency tables were generated for each neighborhood to see the top 5 most common venues. Another function called “return\_most\_common\_venues” was implemented to return a table containing the list of neighborhoods and top 5 venues. K-means clustering was used to divide the neighborhoods into groups. K-means clustering was a good model fit for this situation because it has the ability to divide the data into non-overlapping subsets of similar objects without internal structure. A k of 5 was selected because it gave the most similar findings for the neighborhood groupings. For this use case, the neighborhoods need to be partitioned by what their nearby venues are, but the structure of these clusters is irrelevant. The cluster label was then added to the table containing top 5 venues and a map was generated to visualize the clusters.

**Results**

The neighborhoods were sorted in to 5 clusters. Cluster 1 had 47 neighborhoods with the majority of close venues being a mix of restaurants, parks, and transportation. Cluster 2 contained 1 neighborhood with specialty stores nearby. Cluster 3 contained 1 neighborhood with a mountain and park nearby. Cluster 4 contained 4 neighborhoods with parks and recreation close by. Cluster 5 contained 1 neighborhood with only transportation nearby.

**Discussion**

Based on the results of the analysis, my advice to apartment developers would be to build their complex in one of the 47 neighborhoods presented in cluster 1. These neighborhoods generally presented a nice mix of venues, such as restaurants, bars, grocery stores, and parks, that would make them excellent candidates for apartment living. The other clusters seemed to represent strictly parks, specialty stores, or transportation so may not be as wide in their appeal for potential renters. This report should not be the only evidence taken into account by apartment developers. Proximity to large tech companies, availability of space, and traffic among other factors were not factored in when constructing this model. K-means clustering also is not necessarily the most accurate model for this case since it picks it’s starting points for clustering at random.

**Conclusions**

Apartment developers looking to build their units should consider development in one of the neighborhoods in cluster 1 based on their proximity to consumer-friendly nearby venues. Future research can be done to determine which of these neighborhoods are more livable based on space availability, proximity to work, and levels of traffic.