Portland Neighborhood Selection

1. Introduction

1.1 Background

Portland is the largest city in Oregon both by total area and population (Wikipedia, 2019). It was founded in 1845 and was named after Portland, Maine by Francis W. Pettygrove who won the right to name the city in a coin toss (Wikipedia, 2019). The city was estimated to house 653,115 residents in 2018 (Wikipedia, 2019). Both the Willamette and Columbia rivers run through the city and it has access to many outdoor activities, including hiking, mountain biking, wind surfing, skiing, and surfing. It is the third most affordable city on the West Coast of the United States, with a median home price of $375,425, and was ranked number 8 on the "Overall Best Places to Live List" (Thorsby, 2019). Therefore, it is an attractive place to both live and/or invest in a home.

1.2 Problem

I currently live in Southern California where Real Estate is simply out of my price range. I want to own a home so I can A) begin putting my money towards equity instead of rent and B) secure a home to live for myself and family. However, in order to solve this problem what neighborhoods would be in my price range? Furthermore, of those neighborhoods within my price range which neighborhoods would be suitable to live based on crime statistics and nearby venues?

1.3 Interest

This problem and analysis presented below is potentially of interest to those seeking to relocate to the city of Portland. It is also of interest to potential real estate investors looking to purchase property in Portland. Furthermore, the analysis can be duplicated for any city for those looking to invest or move to a particular location.

2. Data

2.1 Data Description

The following data will be required to determine which neighborhoods are acceptable:
- Median home price data for each of the neighborhoods
- Crime statistics by neighborhood and the offense category
- Neighborhood data such as neighborhood area, longitude, latitude, and neighborhood names

- Nearby venues using the longitude and latitude generated for each neighborhood

The Median home price data will be used to assess which neighborhoods are within my price range and which are not. I will set my acceptable price threshold to $500,000 and below, but I hope to purchase a house in the low $400,000 range. I am setting it to this threshold because there will likely be some homes within my target price range even if the median price is $500,000 and because housing prices are negotiable. After setting my threshold, I will then create a function to convert the median home price data into binary data. Meaning, those houses above my threshold will receive a zero and those within it will receive a one.

The crime data statistics for the current year will be collected and weighted according to the offense category. For example, Homicides will receive a weighted score of five, while petty theft will receive a weighted score of one. After they have been weighted, they will be multiplied by the number of counts associated with the particular incidents. The weighted counts will then be grouped by neighborhood, divided by neighborhood area in order to factor in neighborhood size to the data, then normalized to a score of 100, where 100 is the highest weighted incident count by area. Similar to the real estate data, a threshold for weighted incident counts by area will be set and any neighborhoods with a weighted incident count by area higher than the threshold will be assigned a zero and any neighborhood below the threshold will be assigned a one.

The neighborhood data will also need to be gathered and generated. This includes longitude and latitude values, neighborhood names, neighborhood area, and a neighborhood json file. The neighborhood area will be used to divide the weighted crime statistics as described above, and the json file will be used the visualize the data on a map.

The binary data for both crime and median price will then be multiplied together in order to generate a list of neighborhoods that meet both thresholds. Neighborhoods that do not meet both thresholds will be dropped from the dataframe. Afterwards, the top five venues for each of these neighborhoods will be collected using Foursquare API, then the K-Means machine learning algorithm will be applied to these neighborhoods using venue data, median price data, and weighted crime counts by area to generate clusters. These clusters will then be analyzed based on their suitability and the cluster will be used to narrow the home search to these particular neighborhoods.

References

Thorsby, D. (2019, October 2). *The Best Affordable Places to Live on the West Coast*. Retrieved from US News: https://realestate.usnews.com/real-estate/slideshows/best-affordable-places-to-live-on-the-west-coast?slide=9

Wikipedia. (2019, November 5). *Portland, Oregon*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Portland%2C_Oregon