Clustering Neighborhoods of Portland, Oregon

Skyler Schilke

June 16, 2019

1. Introduction

Portland, OR is a very popular city among many different demographics. It has friendly people, great food, good shops, and the list goes on.

Within the city of Portland, there are 94 neighborhoods. Choosing which neighborhood to live in can be a daunting decision to someone who would like to move to Portland or to a different neighborhood in Portland.

For this project, I will be grouping and ranking Portland's neighborhoods based on categories that are important to a remote employee who works from home. These categories will be walking score, biking score, number of bars, and number of coffee shops

2. Data acquisition and cleaning

In order to group all these neighborhoods, I will need the following data from the following sources:

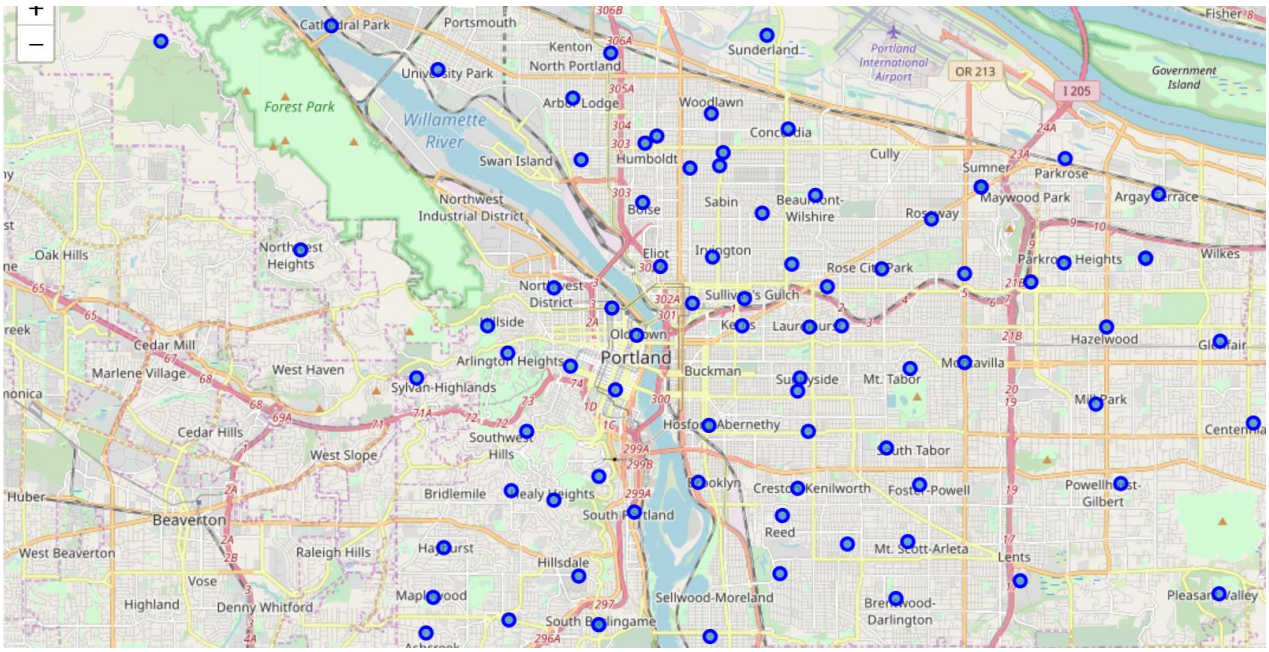
1. List of neighborhoods from Portland Oregon Gov website
2. Latitude and Longitude of the center of each neighborhood from Google Maps Geocode API
3. Walking and Biking score from Walk Score API
4. Number of Bars and Coffee shops from Foursquare API

In order to clean this data, it will be normalized and equally weighted to return a reliable out.

3. Exploratory Data Analysis

3.1 Calculation of target variable

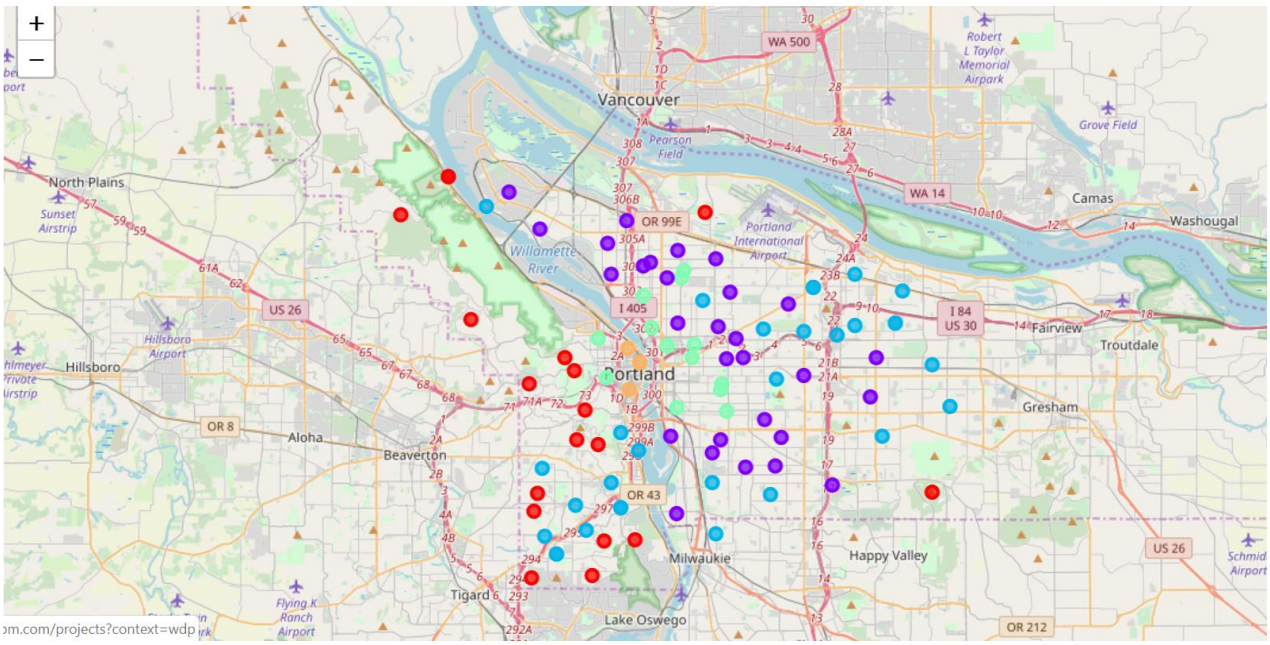
First, the original data was plotted over Portland. This was done to make sure that all of the datapoints were on the correct map. The following is the plotted map:



Next, we normalize the data from Foursquare and Walkscore API which gives us a ranking of all the neighborhoods in Portland. After normalizing the data, we will run a k-means clustering with a k-value of 5 to group all of the neighborhoods into 5 different groups based on Walkscore, Bikescore, number of bars, and number of coffee shops. The following is a map of the resulting cluster

4. Predictive Modeling

Next, we normalize the data from Foursquare and Walkscore API which gives us a ranking of all the neighborhoods in Portland. After normalizing the data, we will run a k-means clustering with a k-value of 5 to group all of the neighborhoods into 5 different groups based on Walkscore, Bikescore, number of bars, and number of coffee shops. The following is a map of the resulting cluster:



Then, we want to see the top 20 neighborhoods based on our selected attributes. The following are the top 20 neighborhoods in Portland, OR:

Mt. Tabor
Glenfair
Woodland Park
Lloyd District Community Association
Portland Downtown
Portsmouth
South Portland
North Tabor
Hosford-Abernethy Neighborhood District Assn.
Healy Heights
Northwest District Association
Vernon
Richmond
St. Johns
Goose Hollow Foothills League
Forest Park
Boise
Humboldt
King
Hayhurst

5. Conclusions

From the model, we can conclude that the best place for a remote employee is in the area of downtown or in a purple on the map.

6. Future directions

The main flaw with this project is the accuracy of the neighborhoods and the flaws in not creating equal areas in the map. For the future, the model will be updated to include user input of a city or address and the model will return an equally spaced 10 by 10 grid of locations, each being ranked from best to worst and clustered based on these same attributes. This will be much more fluid and adaptable of a model than the Portland one presented in this model.