Metropolitan Living in the Pacific Northwest

by MARTINE CARTIER, budding data scientist

INTRODUCTION

I examined the similarity between neighborhoods in Seattle and Portland. This analysis will be used to show which neighborhoods between the two cities are most similar based on local venues. People moving from city to city can use this information to determine which neighborhoods they would like to move to in the new city.

DATA

For this project I will be using a couple distinct datasets. I will use Foursquare location data to evaluate venue types in the neighborhoods of two Pacific Northwest metropoli, Portland and Seattle. I also had to obtain data regarding the neighborhoods and geographic location of the neighborhoods. For Seattle, I found data of the neighborhoods in King County that included latitude and longitude. For Portland, I had to construct a data frame using a list of the city's neighborhoods from Wikipedia and location data generated by the geolocator.

After obtaining my preliminary data, the Seattle and Portland datasets were combined. This is so that when the neighborhoods are clustered based on venue type, the different cities' areas will be evaluated on the same criteria and can therefore be compared. Following the merging, I obtained venue information from Foursquare for each neighborhood, and made a new dataframe that contained this information.

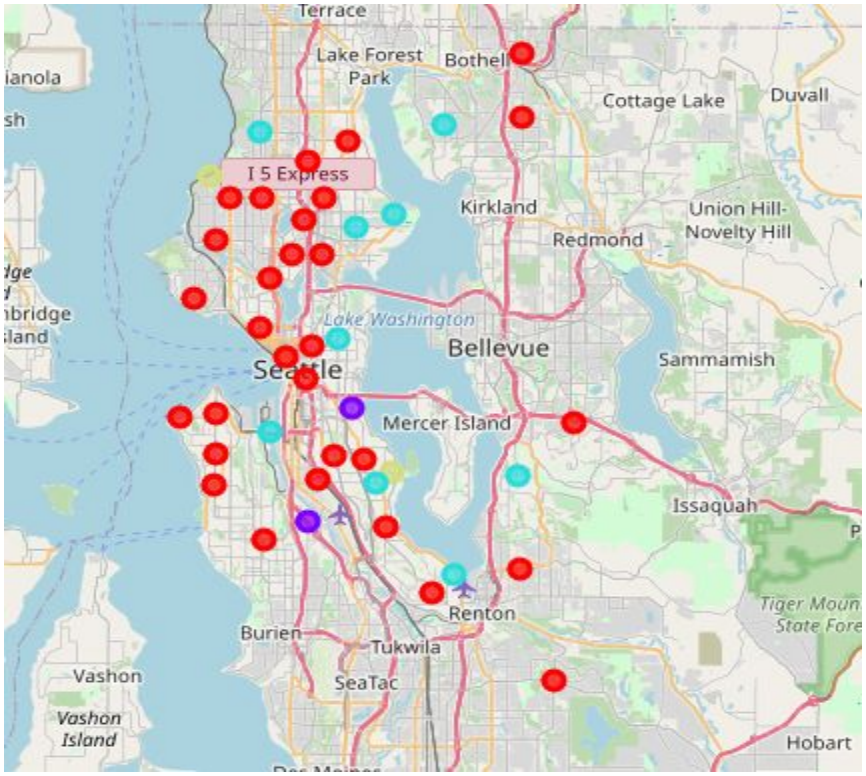
METHODOLOGY

The majority of the work I did was obtaining and preparing the data. When that was complete, the analysis was relatively straightforward. I used the unsupervised machine learning technique k-clustering to cluster the neighborhoods based on popular venue type and analyze the data. I grouped the neighborhoods and found the mean of the frequency for each venue type. I then created the data frame mentioned above, which included the neighborhood names and the top venue types for each neighborhood, in order to gain a better understanding for what the data was showing.

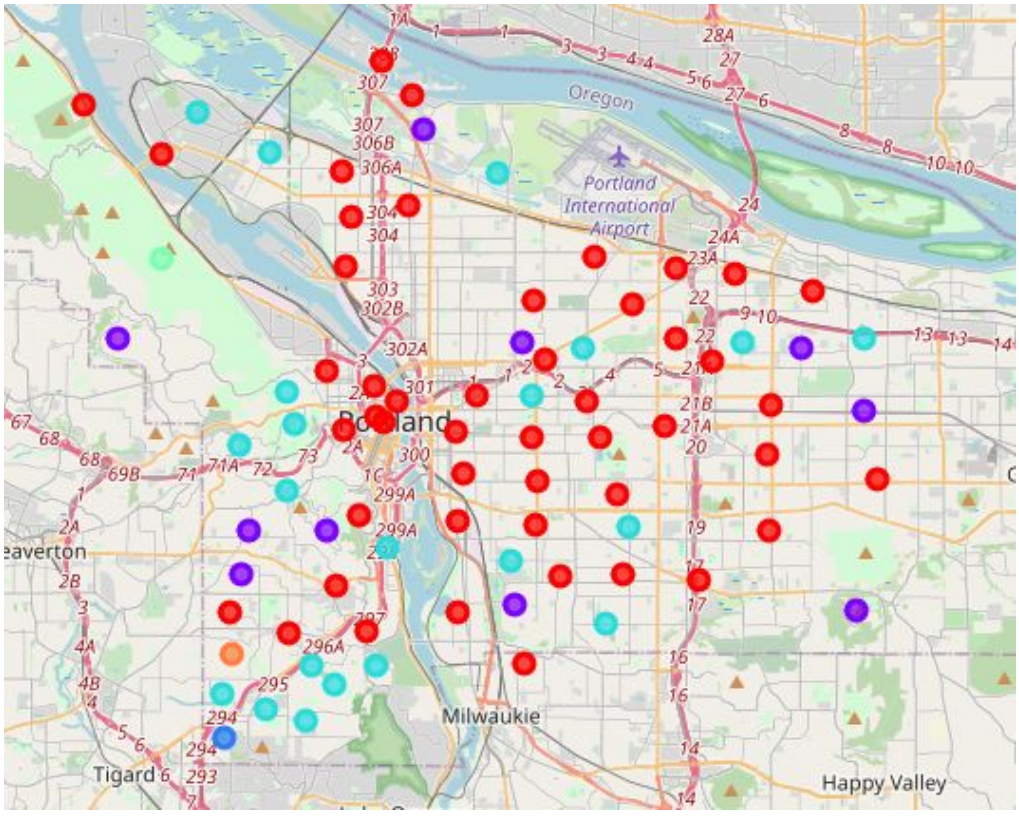
For the analysis, I applied the k-means clustering technique to the neighborhoods based on the mean frequency for each venue type. I used 7 clusters. Finally, I was ready to create a final data frame with all the information I had acquired. This data frame included neighborhood name, location, cluster label, and top venues. I mapped this information using colors for the clusters to easily compare the two cities.

RESULTS

Below are the two maps of Seattle and Portland with their neighborhoods clustered based on venue type.



Seattle



Portland

We can see that there is a lot of similarity between the neighborhoods of the two cities since the majority of each is the purple cluster.

DISCUSSION

This is an interesting preliminary investigation into the similarities of neighborhoods between Portland and Seattle. Based on this initial exploration, folks moving between cities will have ample options to maintain their lifestyle if they are coming from red-dot neighborhoods.

There are definitely opportunities to expand this analysis. One idea for future research would be to include housing cost analysis, as this is a major consideration when selecting a place to live. Additionally, this analysis is primarily for people who are already familiar with the neighborhoods of either Seattle or Portland. Further explanation of the clusters would assist people new to the region in understanding the neighborhoods.

CONCLUSION

In conclusion, this was an exciting foray into data science. This practical analysis will help inform Pacific Northwest movers and facilitate their moving process to an unfamiliar city. The initial observation shows that many neighborhoods in Seattle and Portland have red dots, meaning they have similar venue types. Therefore, people moving from those neighborhoods have numerous options in their destination city.