Introduction

Many IT companies landed their headquarters in the pacific northwest. To name a few: Amazon, Tableau, Redfin from Seattle; Microsoft from Redmond and Expedia from Bellevue. This not only caused a growth in population but also a great increase in housing price in those cities. Because houses are becoming unaffordable to a lot of people working in those big cities, the outer cities have become popular for newcomers to move into. This notebook will explore the most popular cities with IT companies: Seattle, Bellevue and Redmond grouped and 5 nearby cities and observe the differences.

Data

Zillow Housing Data

Home value forecast data can be downloaded from the Zillow research website. By using this data, I will find five nearby cities from the popular cities (Seattle, Bellevue and Redmond) with the highest forecast percentage increase in home value.

Latitude and Longitude Data

For each neighborhood, I will use the geocoder library to get the latitude and longitude.

Foursquare Location Data

By using the /venues/explore endpoint, I will download venues for Seattle, Bellevue and Redmond, then put the data into one dataframe. Another dataframe will also be created to store venues for 5 nearby cities. For each city, I will follow the steps from a previous assignment and group the venues to show the top 10 venue categories. Finally, apply clustering to both group venues and explore the differences.

Methodology

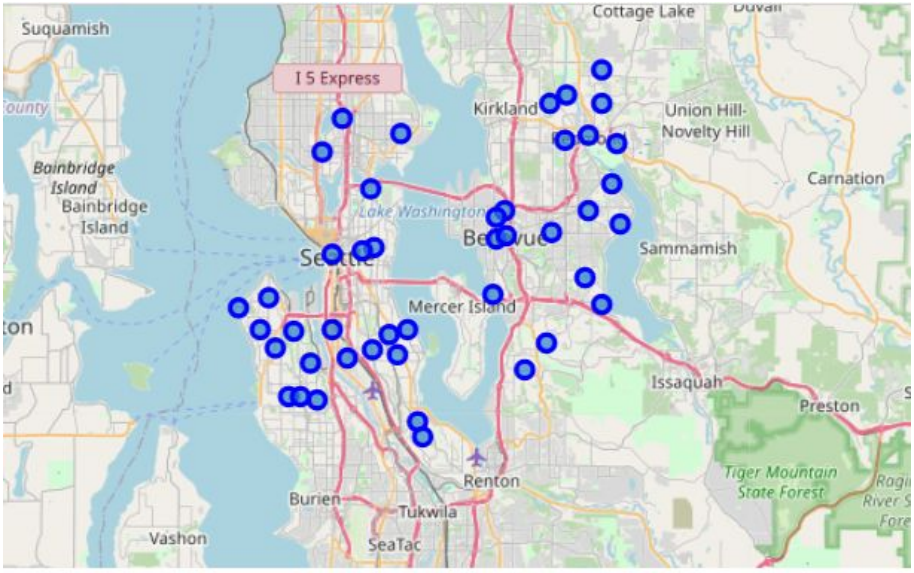
The first step is to find the five nearby cities of Seattle, Bellevue and Redmond so they can be used as comparisons. The method that is used to find the nearby cities is by finding the top five cities from the same county as the popular cities (King County) with ForecastYoYPctChange > 0.5 in Zillow Housing data. The five cities are Kent, Federal Way, Auburn, Des Moines and Burien.

CityName	ForecastYoYPctChange
Kent	1.857143
Federal Way	1.621429
Auburn	1.600000
Des Moines	1.355556
Burien	1.242857

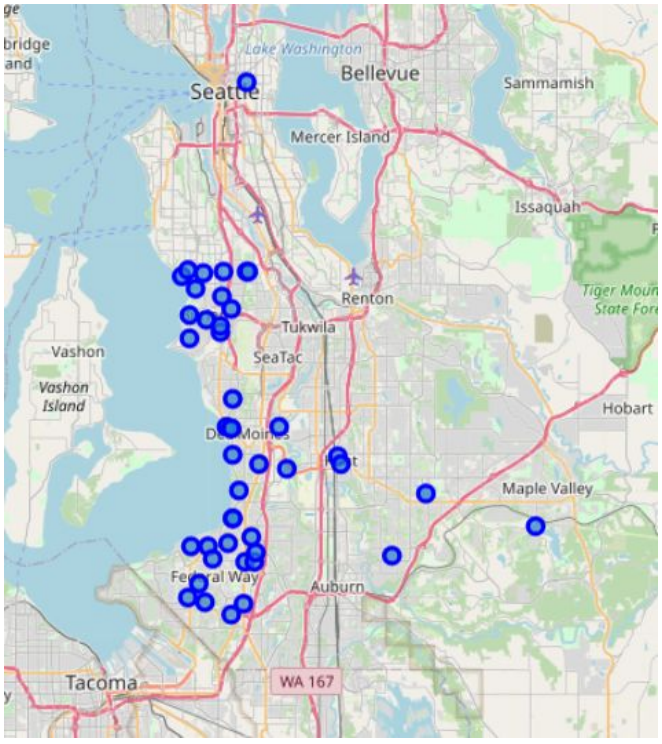
Next, also by using the Zillow Housing data, each city can be broken down into smaller neighborhoods. There are 129 neighborhoods in Seattle, Bellevue Redmond, and there are 45 neighborhoods in Kent, Federal Way, Auburn, Des Moines and Burien. Because 129 is too many from the popular cities, we will eliminate the neighborhoods that have ForecastYoYPctChange <= 0.5. This gives us 44 neighborhoods from the popular cities.

To show the neighborhoods in the map, we use the geocode library to retrieve the latitude and longitudes for each neighborhood.

Neighborhoods in Seattle, Bellevue and Redmond:



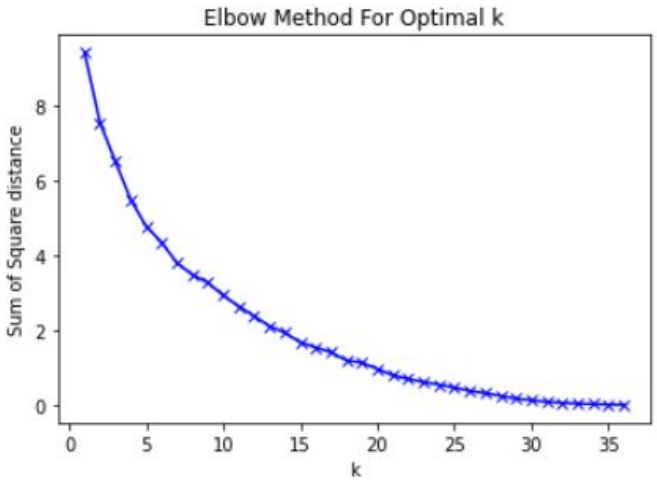
Neighborhoods in Kent, Federal Way, Auburn, Des Moines and Burien:



Kent, Federal Way, Auburn, Des Moines and Burien are cities south of Seattle, Bellevue and Redmond.

For each neighborhood, a request is sent to the Foursquare API to get the venues. There are a total of 201 unique venue categories for the popular cities and 151 unique venue categories for the nearby cities. Because there are too many venue categories, a grouping was performed to group similar venues together. The results are there are 89 venue categories for the popular cities and 74 venue categories for the nearby cities.

The venues from the two city groups are clustered using the k-means method so we can explore the similarities of the cities. The elbow method is using to find the optimal number of clusters but for both city groups, the method doesn't give us much information:



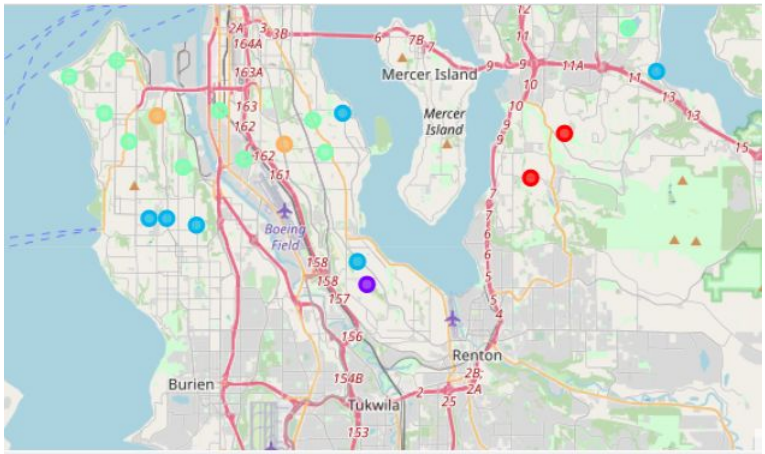
There is no ‘elbow” in the graph above, so for the purpose of the analysis, 5 is picked to be the number clusters to be explored.

Results

We categorized 5 clusters for both the popular cities and nearby cities.

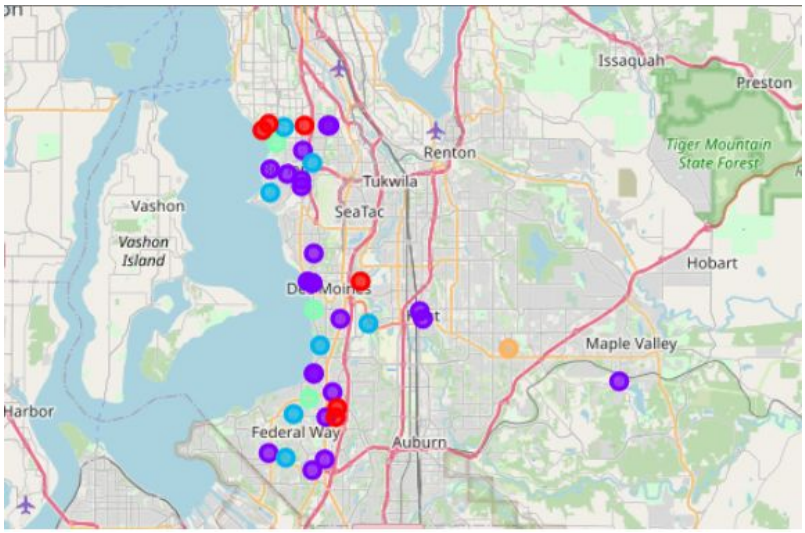
The five clusters for Seattle, Bellevue and Redmond are:

- parks and fields
- workout and exercise studios or gyms
- parks and playground
- restaurants
- transportation services



The five cluster for Kent, Federal Way, Auburn, Des Moines and Burien:

- parks and fields
- restaurants
- transportation services
- home services
- lake



Discussion

Both the popular and nearby cities have a lot of restaurants. An extended analysis that can be done is to explore the differences in the restaurants. The popular cities have more workout places while the nearby cities might have more bigger lands to build services such as Home Depot, Lowes, Ikea, or other stores for home improvement resources.

Because Seattle, Bellevue and Redmond have more tech companies and people tend to live closer to shopping, restaurants and even a variety of personal enhancement services such as yoga studios, cycling studios, etc.

The elbow method doesn't seem to work very well in the two models we explored. So the *k* value that is used in k-means models might not be the optimal value.

Conclusion

This assignment is a good start for exploring two distinct areas due to their professional differences. We can further clean up the data and perhaps look at other data that is not

venue related. Other data that can be considered could be housing value, store or service locations, traffic data, crime data, etc. This assignment is a good start for exploring two distinct areas due to their professional differences.