Battle of the Neighborhoods - Seattle Metropolis

NK

ABSTRACT

Analysis of the Seattle Metropolitan area has been conducted to evaluate the home values in each neighborhood, which shows the thirty most expensive and thirty least expensive neighborhoods in the Seattle Metropolis. Along with this data analysis and machine learning is used to cluster neighborhoods into **six** different clusters based on the the most popular locations found in ***FourSquare***. About 71% of the neighborhoods fall into a single cluster where coffee shop, restaurants, and retail are among the top venues. The neighborhood is also classified based on the Walk and Bike score. Results show that about 30% of neighborhoods are either bike or walk friendly or both.

Contents

1 Introduction

The population of Seattle is growing rapidly, and most of the concentration have expanded to new neighborhoods outside of the Seattle area. In the city of around **four million people**, the city of Seattle has grown outward to accommodate its residents. People There are several existing neighborhoods in the Seattle Metropolis, but the new population is moving away from the Seattle city center. The population is moving towards the east in cities like Bellevue, Redmond, Issaquah and as far south Puyallup and close to Tacoma.

Here we will estimate average housing price in each neighborhood and also local attraction so that customer can make choice with their housing locations in the metro area the Seattle Metropolis. The aim will be create clusters of the most popular attractions and the cost of housing various neighborhoods.

In this project, we will try to find an optimal location for a house based on the average home value in different neighborhoods, and a location where might benefit the public based on the walk and bike scores. The solution will help to inform customers on different locations and neighborhood where they might want to settle into.

The project, in this initial stage, will use k-means clustering machine learning tool to identify common popular local attractions in different neighborhoods. Most and least expensive house value neighborhood in the past five years.

	Neighborhood	State	City	County Name	Latitude	Longitude
0	Lea Hill	WA	Auburn	King County	47.325909	-122.180028
1	Meydenbauer	WA	Bellevue	King County	47.609155	-122.206071
2	Enatai	WA	Bellevue	King County	47.587412	-122.197571
3	Bellevue Downtown	WA	Bellevue	King County	47.615241	-122.192841
4	Newport Shores	WA	Bellevue	King County	47.570655	-122.192345

Figure 1. List of neighborhoods with geographical information

2 Methodology

2.1 Home Value

At first we will need to identify the the neighborhoods in Seattle Metro, obtain the geo-location (latitude and longitude). The neighborhood information was obtained from the Zillow Research, where the home value for all the neighborhoods in the United States has been listed by month. The data for the Seattle Metropolis is parsed and the value the last five year is collected.

Next step is to identify the geographical location for the identified neighborhoods, using python library geocoders, a Gecoding API. Most of the neighborhoods Latitude and Longitude were identified with an exception of a few, shown in Fig. 1.

About 10% of the listed neighborhood (29 to be precise) did not have geographical location information. The neighborhood information in Zillow Research is very extensive and many not be quite necessary for our analysis. Lets drop and plot the remaining neighborhood in a geographical map. If the missing neighborhood significantly distort the neighborhood information in the map; then we might consider adding those in. The Seattle metro map with different neighborhoods is presented in Fig. 2.

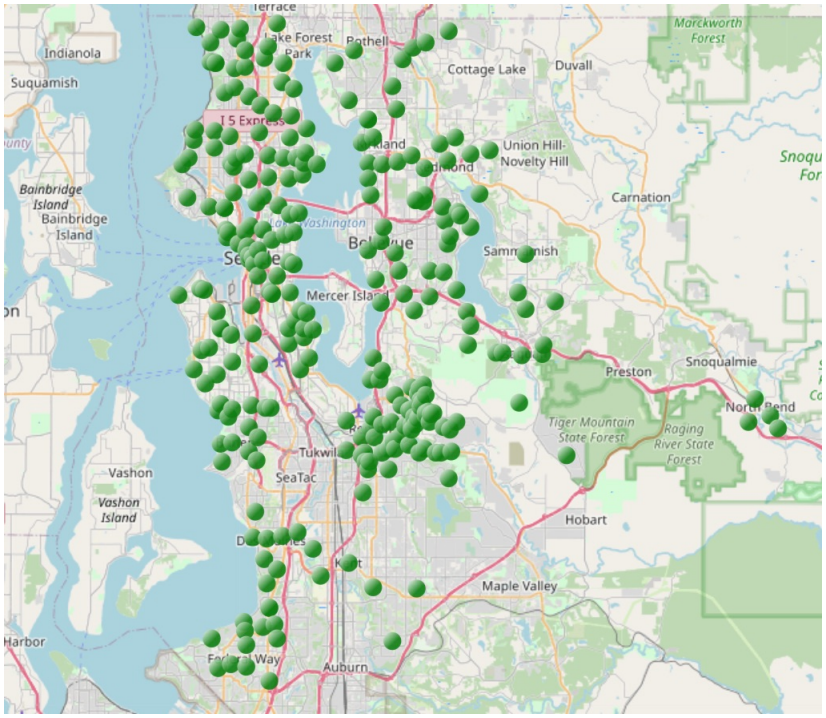

Figure 2. Map of Seattle metropolis with different neighborhood in the surrounding cities

The data shown here shows good representation of the neighborhood in the Seattle metropolitan area; therefore the dropped neighborhoods are not added back into the list. Next step is to shows the plot

of the home values for each year for different neighborhoods. The distribution can be better represented using a box plot, as shown in Fig. 3.

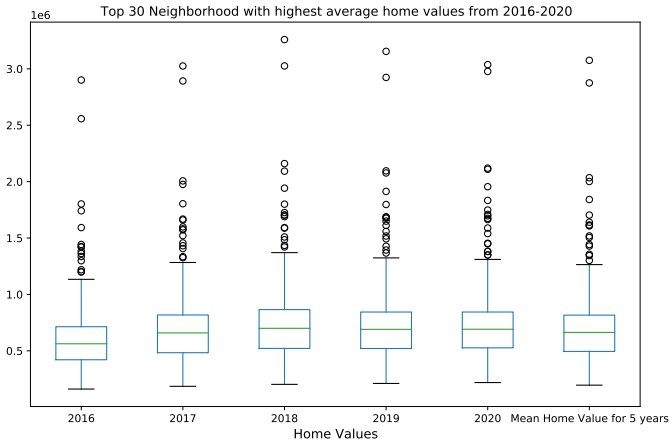

Figure 3. Box plot showing the distribution of the home value for fives years from 2016-2020. The last column shows the distribution of the mean of the home values from those same five years.

For those five years (2016-2020) the largest value (excluding the outliers) are nominally over $ 1 million, and the smallest home values are much lower than $ 500k. Generally, the estimated outlier only exist in the higher end, which reach on the upwards of $ 3.5 millions on average. The spread is also seen to be consistent with the mean value of the housing prices. Next, we will use the average of the last five years home values to identify the top and bottom 30 neighborhoods with highest and lowest home values. The highest and lowest average home values for the top and bottom neighborhoods are shown in Fig. 4. Analysis of the results and discussion on the home values will be done in the following sections.

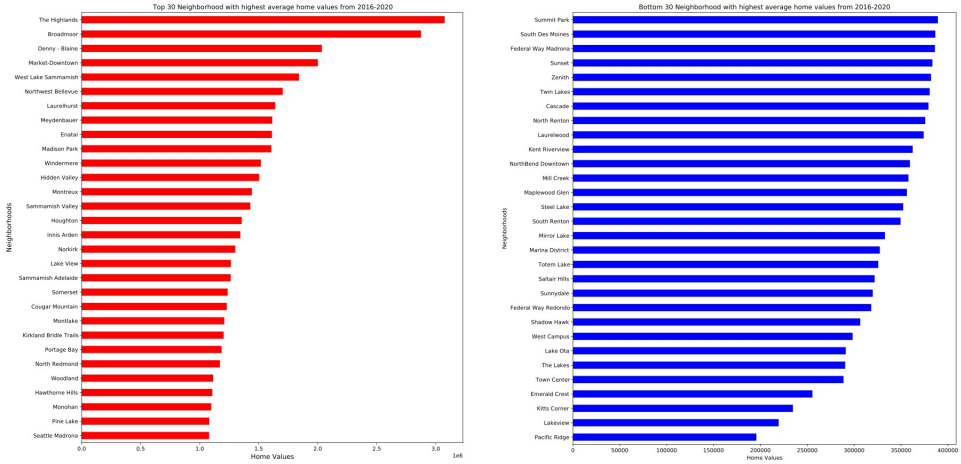

Figure 4. Bar plots showing the home prices for top thirty (left) and bottom thirty (left) neighborhoods.

2.2 Top Attraction using FourSquare API

Each neighborhood location data is used to obtain the top 10 attractions in the given location within 500 m of the geo-coordinates provided by the Geocoder API. The information is then used to create a cluster of neighborhoods using k-means clustering to group neighborhoods with similar top attractions. In the

first step during this process, FourSquare API is used to identify the venues located in the neighborhood, then the k-means clustering is used as shown by the screenshot from the Jupyter notebook below

```
# set number of clusters
kclusters = 6

seattle_grouped_clustering = seattle_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(seattle_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

Figure 5. Using k-means clustering to cluster the neighborhood into 6 different clusters

2.3 Walk and Bike score using Walk Score API from Redfin

The walk and bike data from WalkScore is used to obtain the ease to walk or bike in various neighborhoods. The geolocation data again is primarily used to obtain the walk and bike scores and walk and bike descriptions. The descriptions gives succinct knowlege on where the neighborhood is convenient for the residents to walk to work, places and events or will they have to be dependent on a car or other transportation. Similarly, bike description provides it is extremely convenient to bike or not.

3 Results

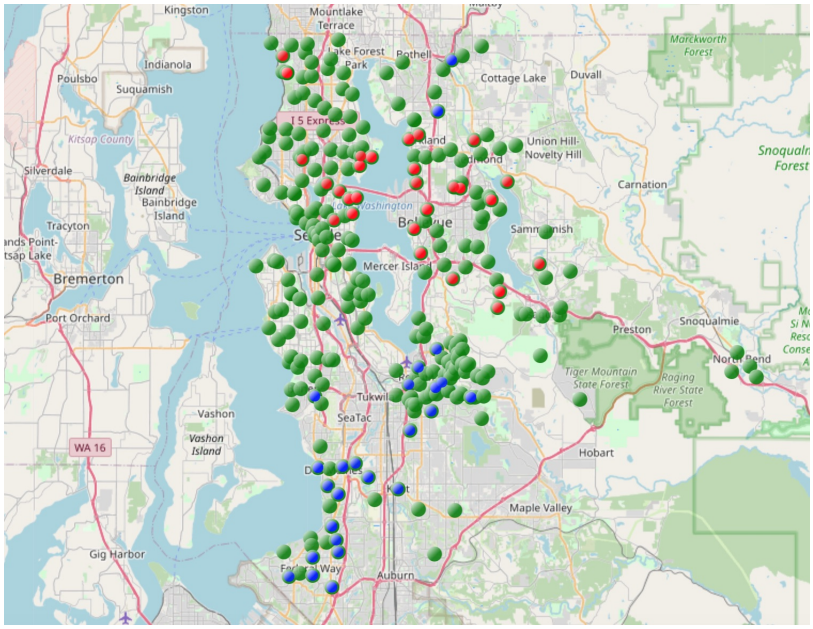

Figure 6. Map of Seattle metropolis with different neighborhood in the surrounding cities along with top and bottom neighborhoods with high and low home values

The home value analysis show that the there is large disparity between the high home value and low value neighborhoods in the Seattle Metro. The overlay of the high and low home value neighborhood (Fig. 6) will help us identity the location with respect to the Seattle City. Interestingly we can be observe that neighborhood with high home values are located in the north-east Seattle Metro, and the low home values are located in south Seattle Metro. In the next iteration of this study each neighborhood with

geographical boundary and gradient color scale would provide a better information. As we move north from the bottom of the map in Fig. 6, one could find the average home value to rise from mid $ 500k towards $ 1 million.

The next of identifying top attractions in different neighborhood is useful.

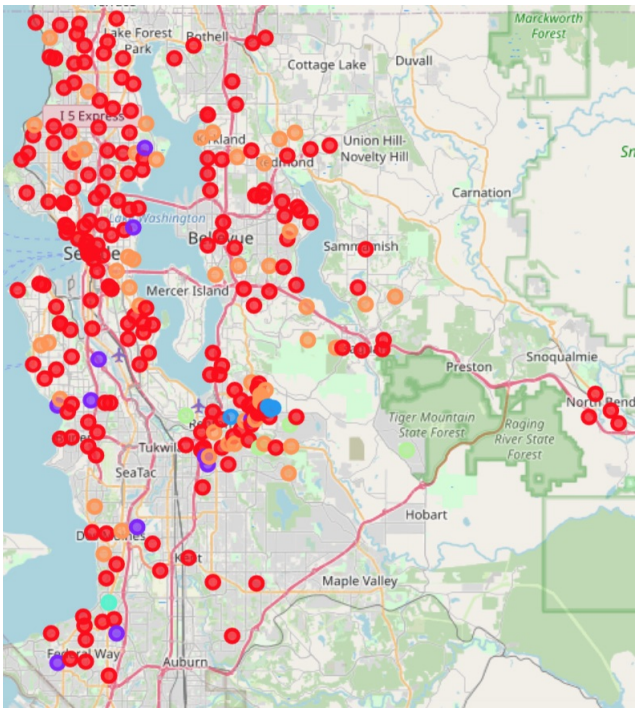

Figure 7. Using k-means clustering to cluster the neighborhood into 6 different clusters. A majority of the neighborhoods seem to clustered into one category

The map in Fig. 7 shows the clustered map of each neighborhood. The neighborhoods are clustered into 6 different categories. Analysis show that out of 281 neighborhood used for data analysis the largest cluster consists of 201 neighborhoods, while the second largest cluster consists of 48 neighborhood, and the remaining 32 neighborhood are divided into 4 other clusters. Most of the neighborhoods are not very different from the top attraction The largest cluster shows that most popular item in those area are coffee shop, food places, and retail stores.

The results from the Walk Score API from Redfin shows that about 30% of the neighborhoods are walk friendly and for the remaining one would need to depend of motor vehicle. Similarly, again about 30% of the neighborhoods are extremely bike friendly and for the remaining one would need to depend of motor vehicle. Figs. 9 and 8 shows the table of the number of neighborhoods with the information.

Walk Description	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Walk Score	Bike Score	Bike Description
Car-Dependent	139	139	139	139	139	139
Somewhat Walkable	63	63	63	63	63	63
Very Walkable	49	49	49	49	49	49
Walker's Paradise	29	29	29	29	29	29

Figure 8. Shows the number of walker's friendly neighborhoods. About 30% of then are walker friendly about 281 neighborhoods being studied.

Bike Description	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Walk Score	Walk Description	Bike Score
Bikeable	72	72	72	72	72	72
Biker's Paradise	13	13	13	13	13	13
Somewhat Bikeable	136	136	136	136	136	136
Very Bikeable	59	59	59	59	59	59

Figure 9. Shows the number of biker's friendly neighborhoods. About 30% of then are biker friendly about 281 neighborhoods being studied.

4 Discussion and Conclusion

From the three different analysis view point there is distinction between the home value between the north, north-east metro and south metro in the Seattle area. The customer looking to purchase or settle in the Seattle area should consider this depending on their budget. In terms of the top attraction there is not much difference among most of the neighborhood. If there is something in particular a customer requires then one of four outlying cluster can be considered; otherwise there should not be a particular preference based on the venues. Seattle metro does not seem to be quite bike or walk friendly with only 30% of the total neighborhood suitable for those customers. Customers need to consider owning a vehicle or look for local transit system to visit various places. In the next step if the study, we should consider adding transit information to our analysis. Moreover, the school district and crime rate in a particular neighborhood should ago be considered to better inform the customers.

The analysis at hand shows that area around Renton, WA is suitable for most customers with less than outlier home values, with many cluster of local attraction. The walk score analysis needs to improve to identify if Renton, WA is walk and bike friendly. Using the location, and local attraction and the cost of housing, and walking a customer could decide which area they would prefer to reside on in the Metropolis.

Acknowledgements

I would like to thank Zillow Research for the details neighborhood list and the home values provides for data analysis and also Walk Score for their database and API to get the walk and bike score of the neighborhoods in this study.