Capstone Project

Shirley

January 2020

1 Introduction

In the past few years, as the increasing number of workers in high-tech industries in Seattle, the housing market faces sharply rising demand. There are sixty neighborhoods in Seattle, separated by their zip codes. These neighborhoods vary a lot in both housing prices and types of venues. When new workers move into Seattle, they usually consider both the affordability of the house and amenities around the house. The recent surge in newcomers in Seattle and the lack of real estate inventory drives the average housing value to a historically high level. The goal of this project is to provide neighborhood-level information to potential buyers in the Seattle housing market.

2 Data Description

Ideally, the data should house-level data of both price and nearby venues. The limitation of data leads to the neiborhood-level study in this project. I first identify the list of neighborhoods in Seattle. For each neighborhood, I collect information about existing venues in that area. A venue is categorized into a specific group that describes the nature of the venue. I also collect the average housing price in each neighborhood. Specifically, I use the following data in my analysis:

- The zip code list of Seattle is from zip-codes.com. The link is https://www.zip-codes.com/city/wa-seattle.asp.

- To locate each neighborhood, I find coordinates of Seattle neighborhoods with Opencage Geocoder API.

- The numbers and types of venues in each neighborhood are collected from the Foursquare API.

- I find the average housing prices from www.zillow.com. The limitation of this dataset is that there are only 34 neighborhoods available.

- To draw the choropleth map of housing prices, I use the boundaries of zip code areas from ArcGIS Hub.

3   Methodology

From the venue data, there are more than three hundred categories of venues, including different types of restaurants, bars, parks, gym, etc. The first step of the analysis is to find the structure of the data. I use the K-means algorithm to find the typical amenities in a neighborhood. The neighborhoods are separated into different clusters by the venue category distribution. I then mark the clustering of neighborhoods on the map of Seattle.

I illustrate the distribution of housing prices in Seattle with a Choropleth map. For the set of neighborhoods with both venue and housing price information, I provide both information on the same map.

4   Results

4.1   Neighborhood Clustering

I separate the sixty neighborhoods into four clusters. There are 17 neighborhoods in cluster 0, 9 neighborhoods in cluster 1, 33 in cluster 2, and 1 in cluster

Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Coffee Shop	Hotel	Cocktail Bar	Bakery	Breakfast Spot	Italian Restaurant	Sandwich Place	Seafood Restaurant	Sushi Restaurant	French Restaurant
1	Park	Coffee Shop	Pizza Place	Vietnamese Restaurant	Brewery	Grocery Store	Gas Station	Playground	Pub	Bakery
2	Coffee Shop	Pizza Place	Sandwich Place	Bar	Mexican Restaurant	Burger Joint	Park	Bakery	Ice Cream Shop	Pet Store
3	Park	Convenience Store	Coffee Shop	Pizza Place	Gym	Motel	Bus Station	Storage Facility	Baseball Field	Organic Grocery

Figure 1: Top 10 Most Common Venues In Each Cluster

3.

I use the characteristics of the centroid of a cluster to illustrate the frequency of venues in each group. Both cluster 0 and cluster 2 are attractive for city dwellers, featuring cafes, restaurants, and bars. Venues in cluster 0 may cater to customers who prefer a more luxurious lifestyle. With the park as the most frequent venue category, neighborhoods in cluster 1 provide a good living area for city dwellers who enjoy the beauty of nature. Neighborhoods in cluster 3 seem to have fewer amenities for a city lifestyle.

Now let us see the locations of neighborhoods in each cluster on the map.

In figure 2, red marks are neighborhoods in cluster 0, purple marks are neighborhoods in cluster 1, green dots are neighborhoods in cluster 2, and the yellow mark is the neighborhood in cluster 3. Not surprisingly, neighborhoods in cluster 0 are in downtown Seattle.

Figure 2: Locations of Neighborhoods in Each Cluster

Cluster Labels	price
0	787416.666667
1	688837.500000
2	658410.526316
3	411600.000000

Figure 3: Average Housing Price

4.2 Housing Price

From figure 3, neighborhoods in cluster 0 have the highest average price, maybe because of the land value in downtown Seattle. The average housing price of cluster 1 is higher than that of cluster 2, illustrating the preference of amenities like parks.

5 Discussion

Figure 4 shows both the housing prices and the clustering of neighborhoods. For people who enjoy metropolis life and people prefer the beauty of nature, there are expensive houses around central Seattle and more affordable choices in other neighborhoods.
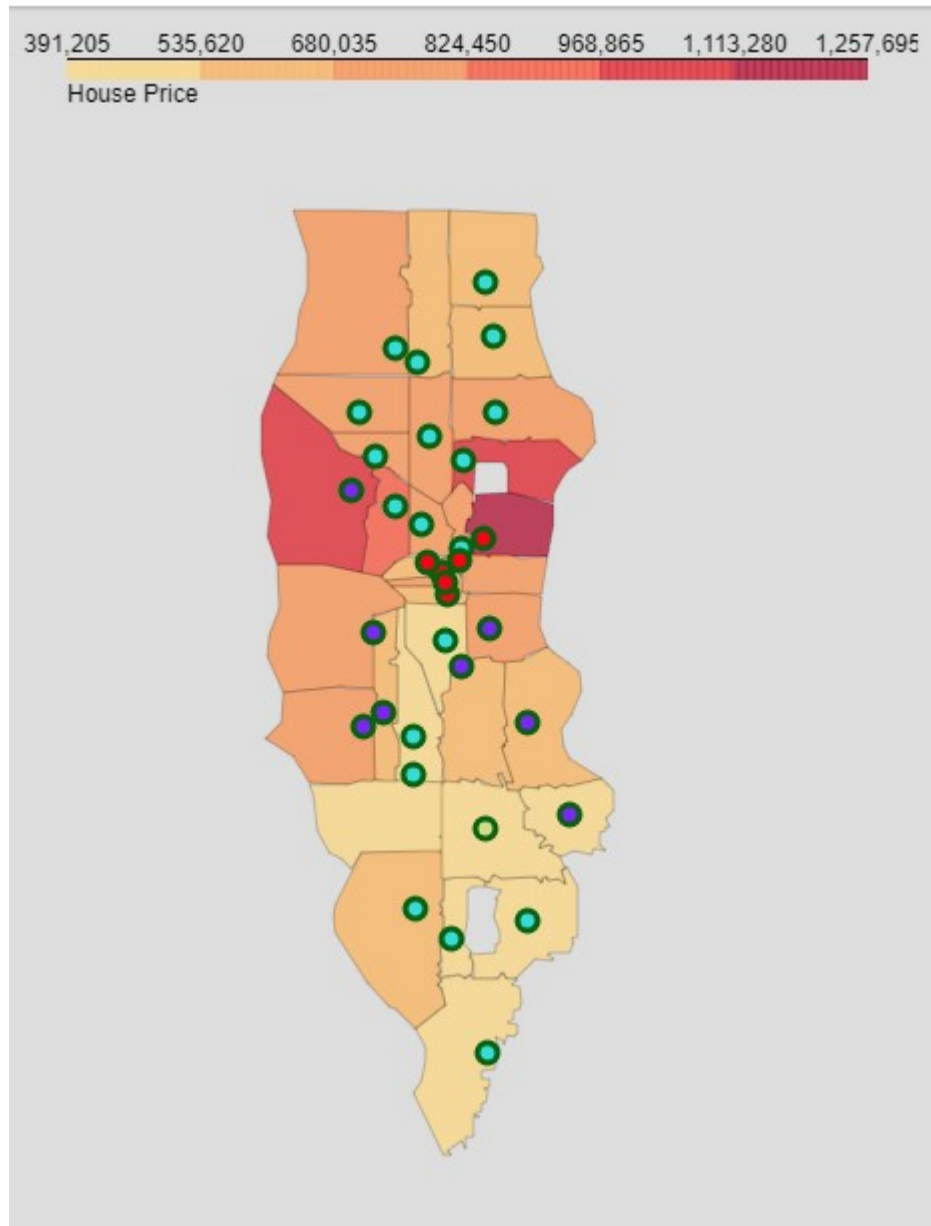
Figure 4: Housing Prices and Amenities

6 Conclusion

Seattle provides neighborhoods of various characteristics that could satisfy the needs of different people. In this project, I create a map to illustrate both categories of information that could help the city dwellers pick the neighborhood that suits their preferences.

There are two limitations to our data. First, there are only 34 neighborhoods with housing price information. Second, an ideal dataset of this analysis should use amenity and housing price information for each housing transaction. My study thus only provides limited guidance for people to search for a new home in the Seattle area.