Effect of Neighborhoods on Housing Pricing in City of Los Angeles

Arvin Zarookian

June 2020

1- Introduction

1-1- Housing in Los Angeles

Los Angeles County is the most populous county in the United States, with more than ten million inhabitants as of 2018. It has 88 incorporated cities and many unincorporated areas at 4,083 square miles. Its county seat, Los Angeles, is also California's most populous city and the second most populous city in the U.S., with about four million residents [1].

Los Angeles County, like most metropolitan cities, has its own housing issue. At the census of 2000, there were 3,133,774 households, and 2,137,233 families in the county. The population density was 2,344 people per square mile. There were 3,270,909 housing units at an average density of 806 per square mile. The homeownership rate is 47.9%, and the median value for houses is $409,300. 42.2% of housing units are in multi-unit structures [1].

The 2010 United States Census reported Los Angeles had a population of 3,792,621. The population density was 8,092.3 people per square mile. There were 1,413,995 housing units at an average density of 2,812.8 households per square mile, of which 503,863 (38.2%) were owner-occupied, and 814,305 (61.8%) were occupied by renters. The homeowner vacancy rate was 2.1%; the rental vacancy rate was 6.1%. 1,535,444 people (40.5% of the population) lived in owner-occupied housing units and 2,172,576 people (57.3%) lived in rental housing units [2].

1-2- Problem

Due to high demand and low supply (available houses for sale), Los Angeles housing market is very competitive and real estate's average selling price is above nation average. Redfin, a real estate company, describes housing market in LA as follow giving a compete score of 71 (very competitive) [3]:

- Many homes get multiple offers, some with waived contingencies
- Homes sell for around list price and go pending in around 43 days
- Hot Homes can sell for about 3% above list price and go pending in around 21 days

In such a competitive market, evaluating housing price is very important. Different factors such as housing type (single family residential, condo or townhouse), neighborhood, property square footage, number of beds and baths, etc.… will effect on housing pricing. While evaluating each of mentioned factors requires a separate study, here we will focus on how housing price is related to its neighborhood. The goal is to answer this question:

"Do similar neighborhoods have same average selling price per square footage?"

1-3- Methodology:

This study is divided into two sections. First a dataset of Los Angeles neighborhoods will be created. Using Foursquare API searching tool, these neighborhoods will be clustered in similar clusters based on common venues in each neighborhood. In order to have a better visual understanding, using Follium library, we will create a map of clustered neighborhoods.

In second section, data for sold houses prices in each neighborhoods will be collected and sorted. The average selling price per square foot will be calculated for each neighborhood.

By comparing the data from these two sections, we will try to find an answer for our question.
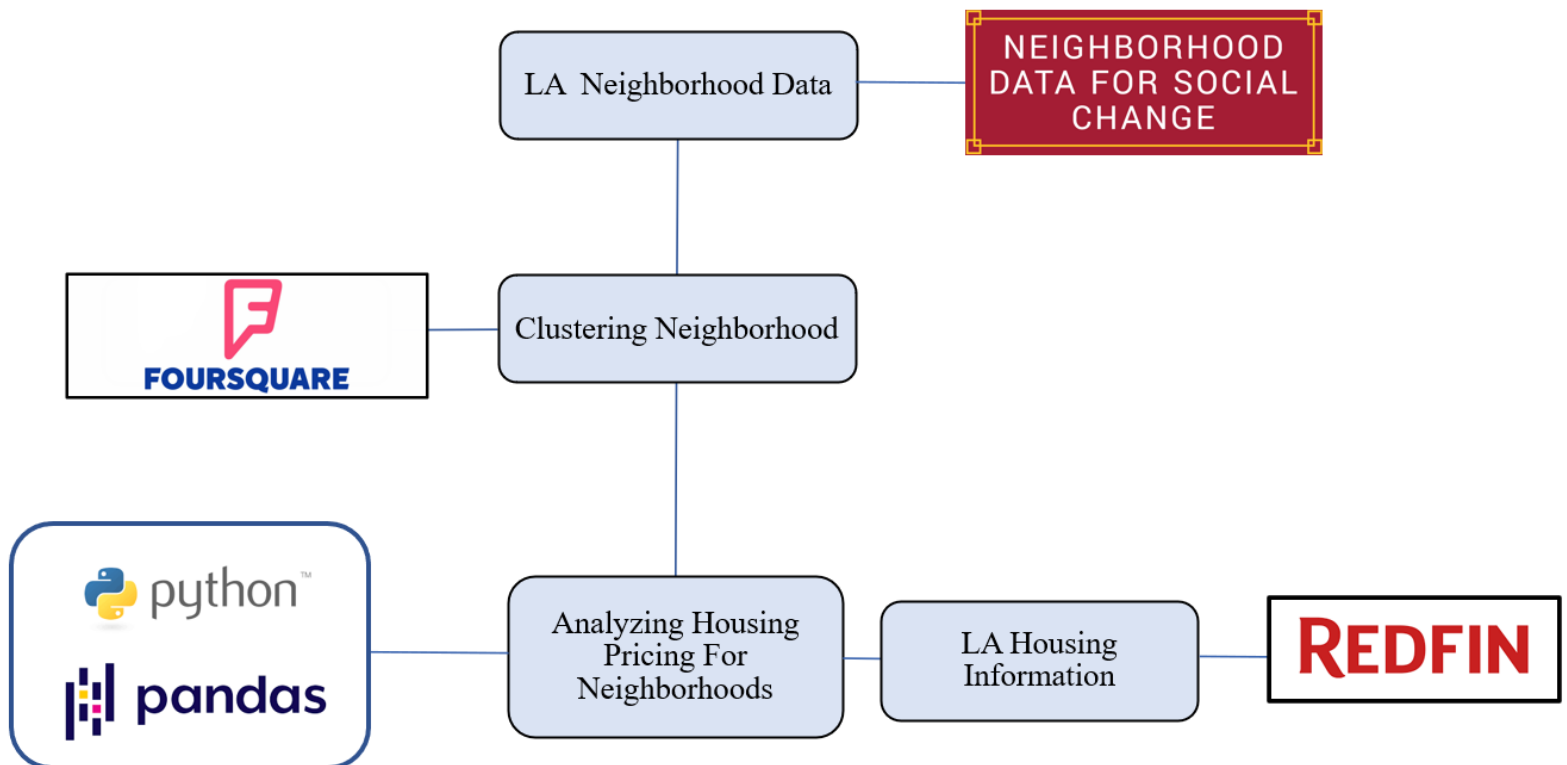

Figure 1 – Methodology and Data Sources

2- Data Description and Cleaning:

Based on described methodology, two sets of data for Los Angeles Neighborhood and Housing Prices will be needed. In this section, we will review data sources and briefly present how data will be cleaned and wrangled for this study purpose.

2-1- Los Angeles Neighborhoods:

To acquire Los Angeles neighborhoods data, we will use information from Neighborhood Data for Social Change (NDSC) platform [4]. NDSC is a project of the University of Southern California Price Center for Social Innovation. It is a free, publicly available online resource for civic actors to learn about their neighborhoods.

This data set includes geographical information for about 272 city/neighborhoods within Los Angeles County. There are three main neighborhood categories in this dataset, namely:

- unincorporated-area (71 neighborhood)
- segment-of-a-city, referring to city of Los Angeles (114 neighborhood)
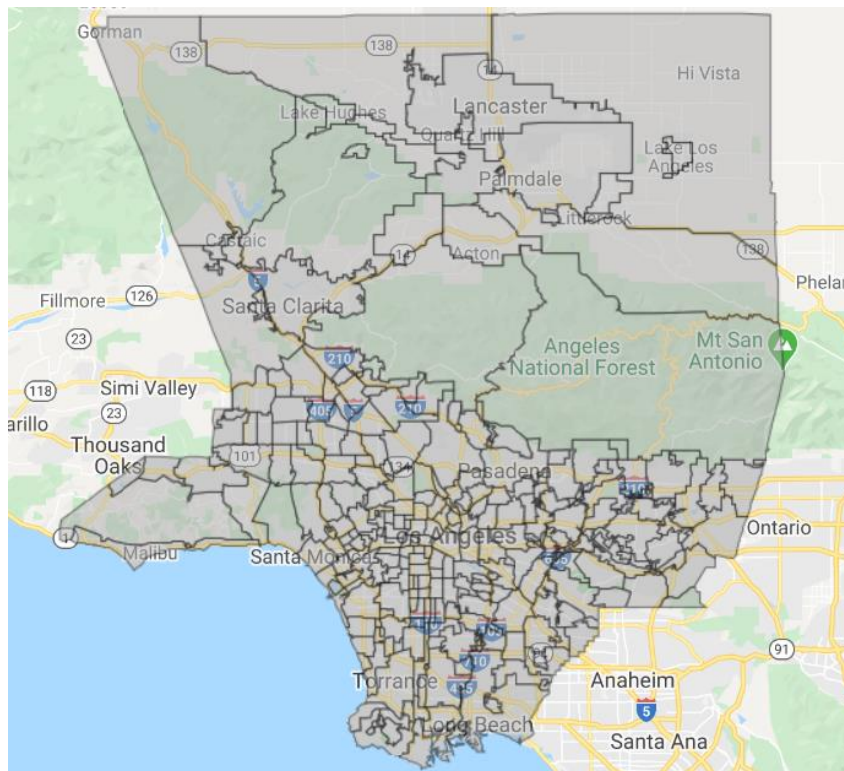- standalone-city (87 city/neighborhood)

Figure 2 – LA Neighborhoods

Unincorporated-areas are not populated and developed as other neighborhoods and are out of scope of this study. Also some standalone cities have their own pricing, sometimes double county average, which can result to a outlier in data set. Therefore we limit this study to neighborhoods within city of Los Angeles (114 neighborhoods).

The dataset from NDSC will be cleaned and wrangled to only include target neighborhoods. Unnecessary data will be deleted and at the end the dataset will include neighborhood name, latitude and longitude.

	neighborhood	latitude	longitude
1	Adams-Normandie	-118.300208	34.031461
10	Arleta	-118.430757	34.243100
11	Arlington Heights	-118.323408	34.044910
13	Atwater Village	-118.262373	34.131066
17	Vermont-Slauson	-118.290358	33.983914

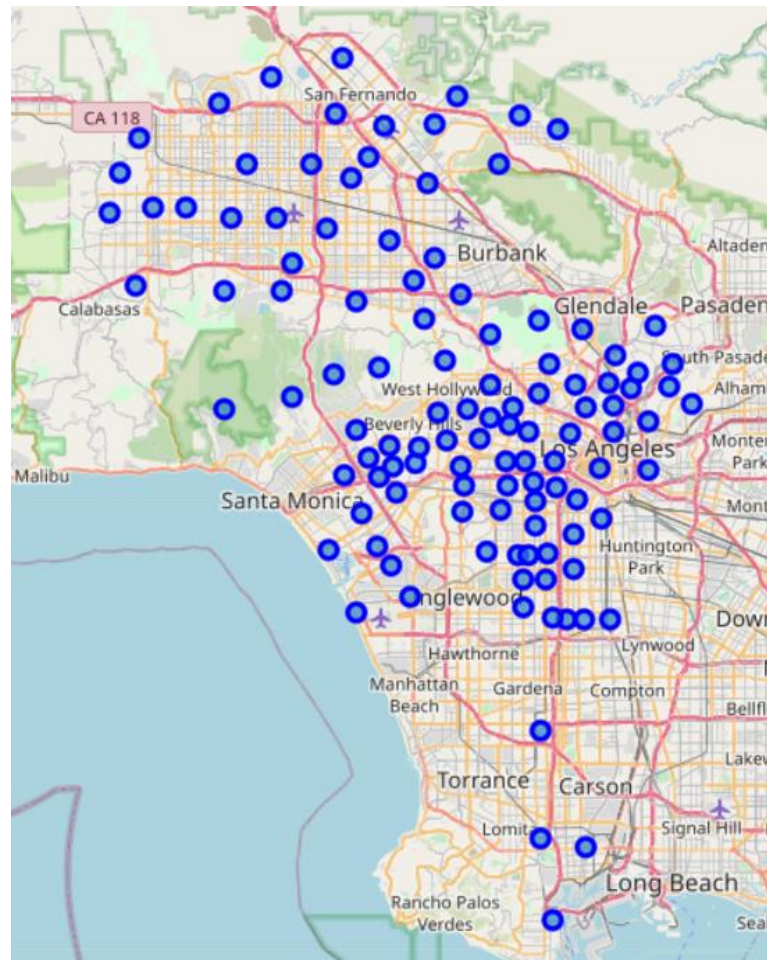**Table 1 – LA neighborhood data set after cleaning**

Figure 3 – LA neighborhood distribution

2-2- Los Angeles Housing Prices:

Redfin [5] website is the source for housing pricing in this study. This Seattle based company provides a comprehensive data for sold houses for each city, neighborhood or even zip code. Using Redfin search tool for city of Los Angeles, we can create a dataset of sold houses for different property types and given period of time (up to last three years). We will select our data set based on following condition

- To avoid complication in house pricing over the years such as annual inflation rate or market surge in a specific time, we will limit our search for last on year.
- single family houses have a different pricing standard comparing to condos and town homes, therefore our data set will only include sold condos and town homes.
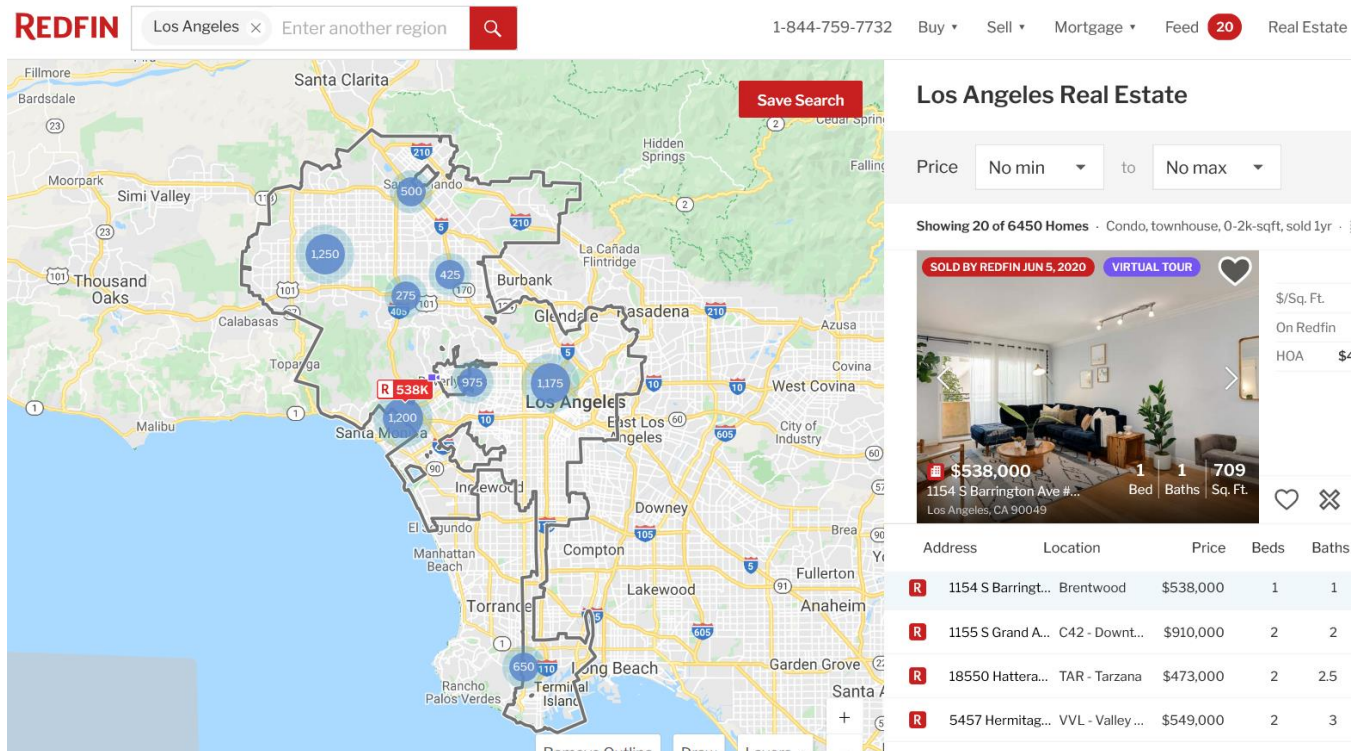- Property size will be limited to 2,000 square feet.

Figure 4 – Redfin Website Information

Applying above criteria in Redfin search tool will result in about 6,500 cases (it's should be noted that this number is being updated on daily bases and in any given date the total number could be different). This dataset includes information such as property type, address, city, number of beds and baths, sold price, square feet, $/square feet, latitude and longitude.

For purpose of our study, columns containing information for city where property is located, average price per square feet, latitude and longitude is required. Dataset will be cleaned to include only this information.

This dataset is based on city name, different from LA neighborhood data bases (there are some common neighborhoods). First we have to update neighborhoods to match with LA neighborhoods. Distance of each sold property will be calculated from each neighborhood (using Geopy distance function) and closest neighborhood will be assigned to row

	neighborhood	$/SQUARE FEET	LATITUDE	LONGITUDE
0	Sherman Oaks	463	34.153650	-118.454706
1	Echo Park	627	34.082351	-118.254832
2	Sawtelle	600	34.049958	-118.469722
3	Panorama City	333	34.239891	-118.448450
4	Sherman Oaks	384	34.165849	-118.465634

Table 2 – LA sold properties

At the end, housing information for only 63 neighborhood was retrieved.

3- Results

Before getting ready for clustering, neighborhoods' venues information was extracted using Foursquare API. The inquiry was limited to 100 venue per neighborhood in 2,000 m radius. Information for about 9,000 were returned by API. Some neighborhoods reached to 100 venue limit, while for some other only information for 10 venue was available.

k-means algorithm was used to cluster neighborhoods. After a few iteration, it was determined that best results can be achieved with k=5. Not all clusters have equal neighborhoods, Cluster 2 with 51 and cluster 3 with two neighborhoods have the most and least neighborhood respectively. Rest of neighborhoods have been divided between cluster 4 with 45, cluster 1 with 11 and cluster 0 with 5.
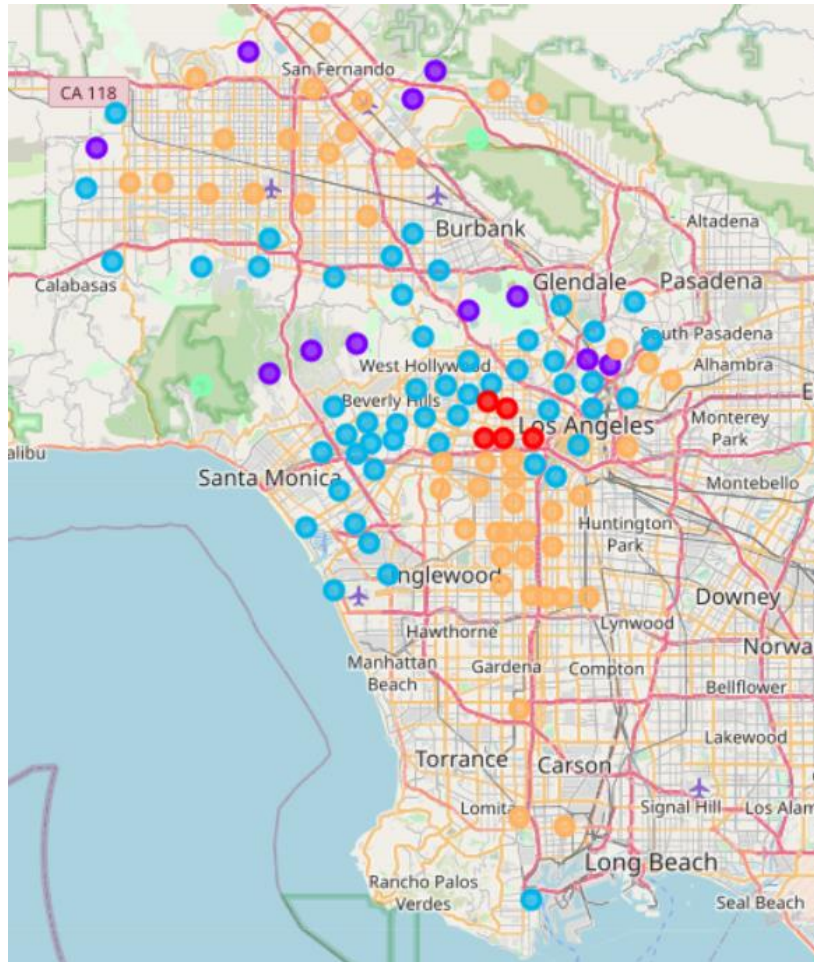
Figure 5 – LA neighborhood clusters

Later cluster information was transferred to Los Angeles housing dataset. As mentioned earlier, housing information only for 63 neighborhood is available. It should be noted that no information was available for two neighborhoods in cluster 3, so the results is presented for other 4 clusters. The square per feet price for each cluster has been shown in following figures.
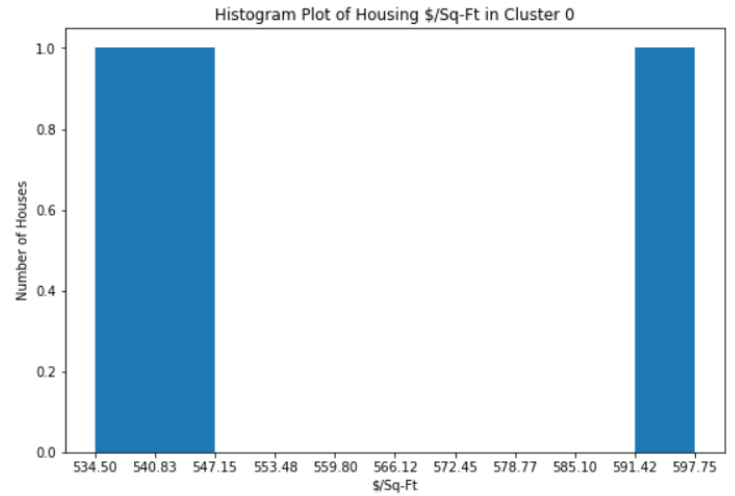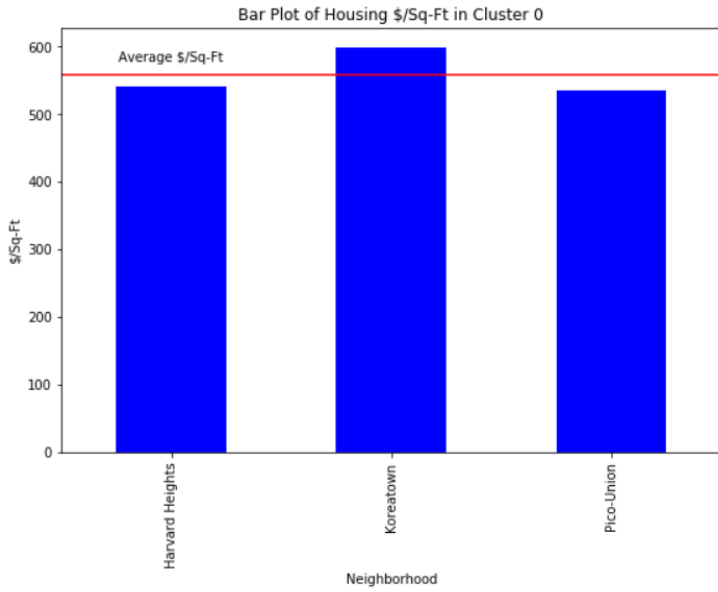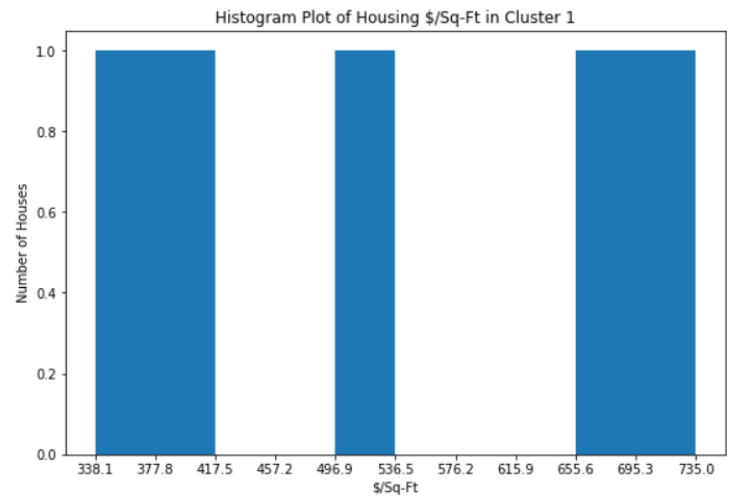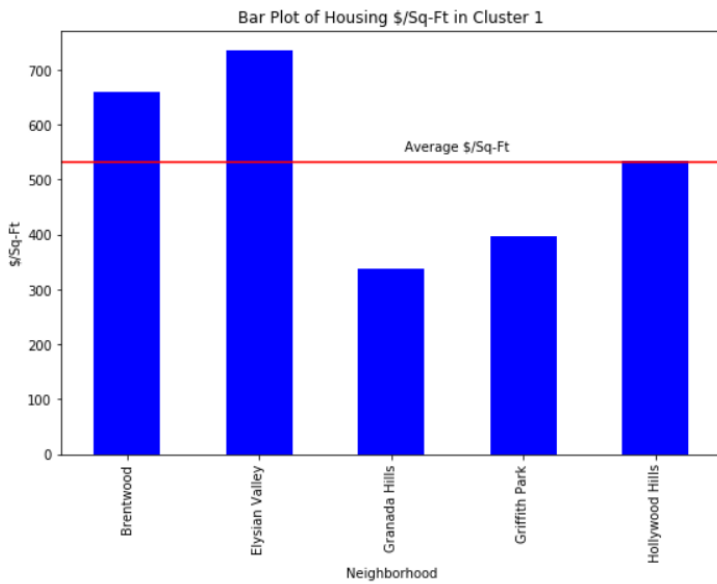
Figure 6 – Pricing Information for Cluster 0
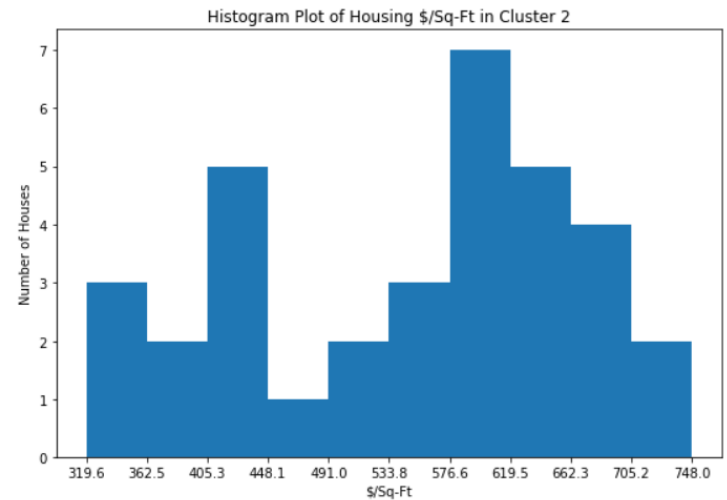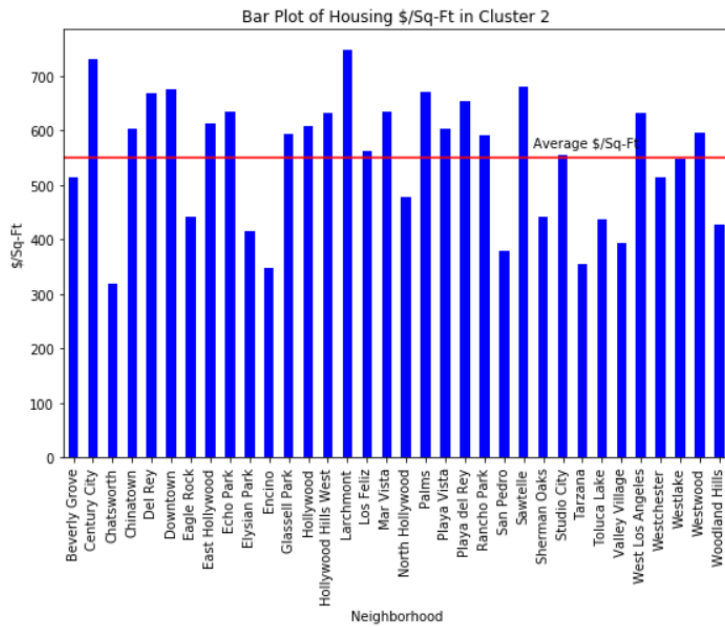

Figure 7 – Pricing Information for Cluster 1

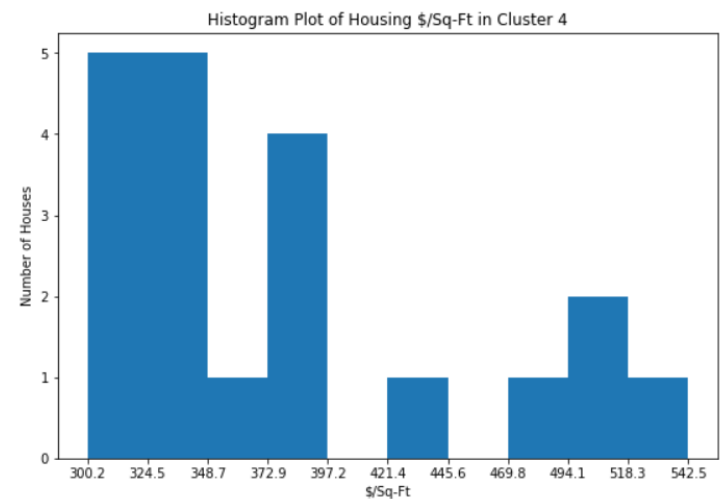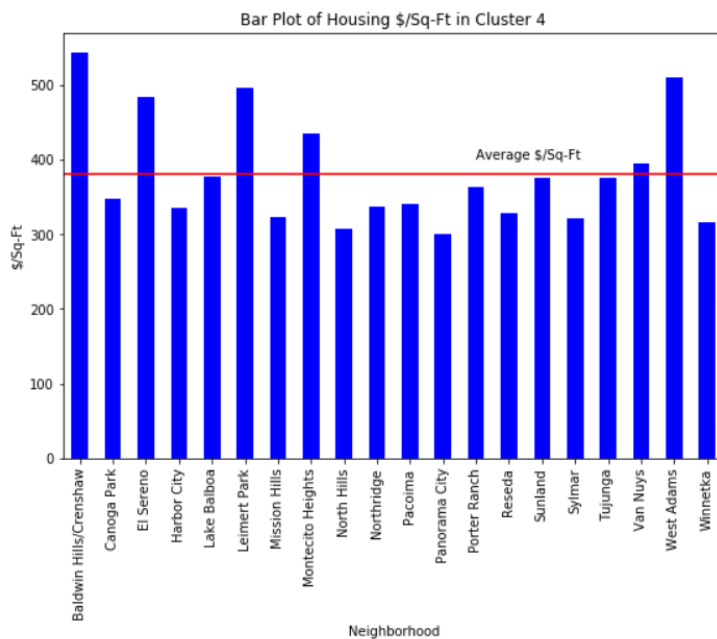**Figure 8 – Pricing Information for Cluster 2**

Figure 9 – Pricing Information for Cluster 4

4- Discussion:

In order to analyze the results, the average and standard deviation for each cluster has been shown in Table 3 and Figure 10.

Cluster 0 and 4 have relatively small standard variation which means neighborhoods prices are close to cluster average. On the other hand cluster 1 and 2 have a larger standard variation. From respective figures it's clear that in these clusters neighborhood pricings are spread over a wider

range. These neighborhoods needs more refining by adding more detail information for neighborhoods.
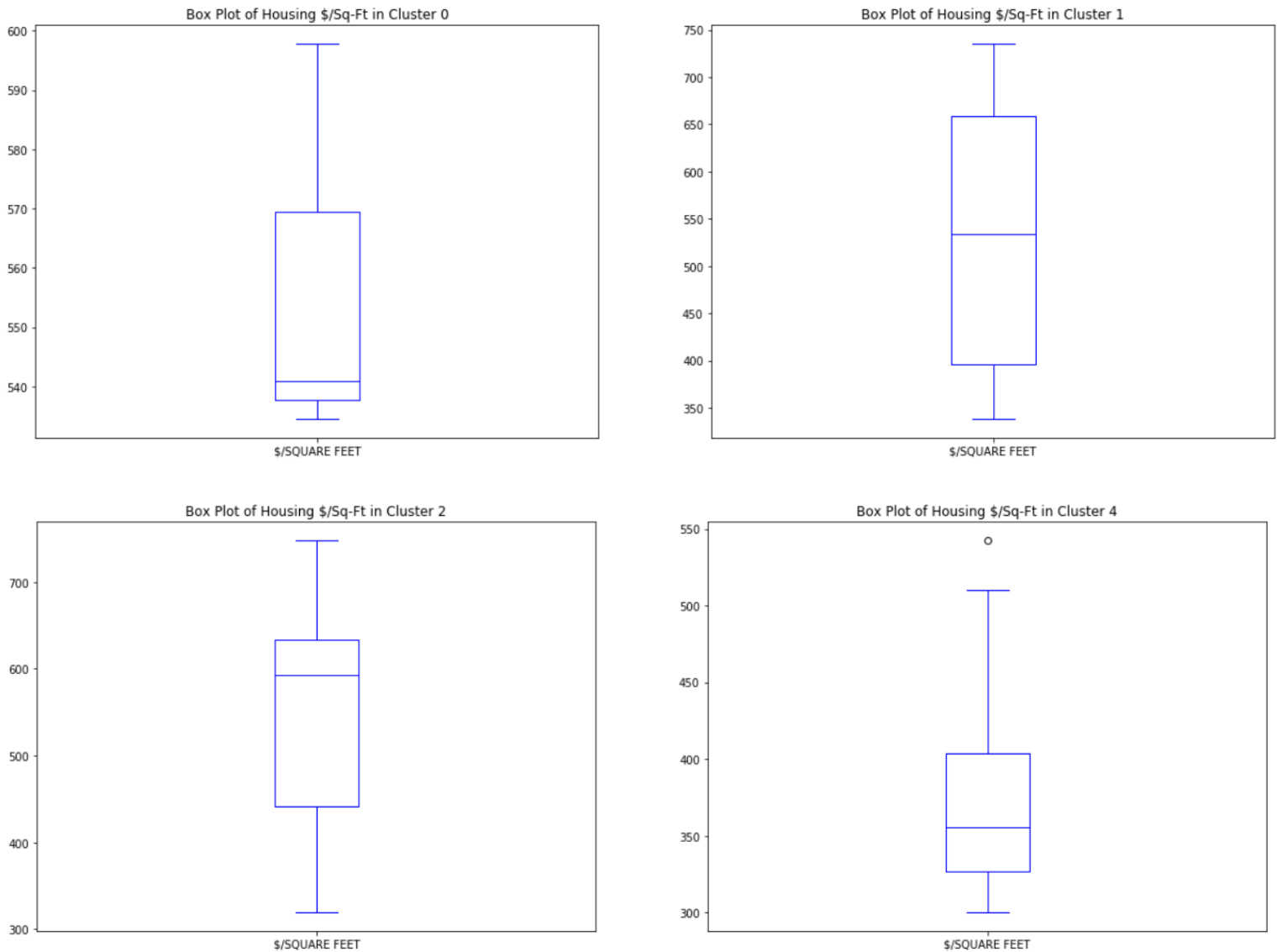

Figure 10 – Box Plot for Clusters Pricing

Cluster	Number of Neighborhoods	Average $/Sq-Ft	Standard Deviation
0	3	557.75	34.79
1	5	532.27	168.26
2	35	549.67	117.32
4	20	380.29	73.59

Table 3 – Cluster Average Pricing and Standard Deviation

5- Conclusion:

City of Los Angeles is a populated area at hearth of LA County in California. As any populated urban area housing is a major issue and pricing is very competitive. Predicting housing pricing and have a model to predict and determine pricing has a significant role in city planning, investors, sellers and buyers decision making.

In this study, It was tried to analyze effect of neighborhood on housing pricing. The neighborhood were clustered based on venues within each neighborhood and pricing for each neighborhood was checked. Also this study was only limited to condos and town homes, excluding single family houses.

As explained in previous section, this study reached to some of it's goals, but there is room for improvement and more detail investigation. Some suggestion for future research are:

- Improving clustering by adding more variables, i.e. school districts
- Expanding neighborhood list to other cities in LA County
- Adding Single Family Residentials to housing database

References

1- Los Angeles County, California, Wikipedia, en.wikipedia.org/wiki/Los_Angeles_County,_California

2- City of Los Angeles, Wikipedia, https://en.wikipedia.org/wiki/Los_Angeles

3- Los Angeles Housing Market Insight, Redfin, www.redfin.com/city/11203/CA/Los-Angeles/housing-market

4- Los Angeles Neighborhoods, Neighborhood Data for Social Change (NDSC), usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr

5- Sold Properties in City of Los Angeles, Redfin