Assignment 3
NYC Real Estate Data Analysis using R

Moiz Deshmukh
BU ID: U63698143

Metropolitan College, Boston University
AD517: Applied Business Analytics (Summer 2)
Professor Prasad Kothapalli
July 23rd, 2024

Table of Contents

Executive Summary **3**

Descriptive Statistics **4**

K-means Clustering **5**

T-Test Analysis **7**

Conclusion **7**

Appendix **8**

Figure 1 8

Figure 2 8

Figure 3 8

Figure 4 9

Figure 5 9

Figure 6 9

Figure 7 9

Figure 8 10

Figure 9 10

Figure 10 11

Figure 11 11

Executive Summary

This report presents a comprehensive analysis of Bayside's residential real estate market since 2009, highlighting substantial sales activity and high property values. Bayside's market is diverse, offering both affordable and luxury housing options with a wide range of property sizes. This diversity underscores the neighborhood's attractiveness for buyers seeking spacious homes.

The K-means clustering analysis compared Bayside with other New York City neighborhoods, focusing on median sale price, standard deviation of sale prices, and the number of units sold. Bayside is grouped in a cluster with moderate property values and significant variability, indicating a dynamic market with varied investment opportunities. This cluster highlights Bayside's active and diverse market characteristics.

A two-sample t-test comparing sale prices between Bayside and Astoria reveals a statistically significant difference, with Bayside being more moderately priced. This distinction emphasizes the unique market dynamics within Queens Borough. Overall, Bayside's robust and versatile real estate market offers valuable insights and opportunities for investors and developers, making it an attractive area for a wide range of buyers.

Descriptive Statistics

This report presents the results of a comprehensive descriptive analysis of residential real estate data for the Bayside neighborhood since 2009. The analysis was conducted using R, and each question has been addressed with corresponding data insights. The total number of sales in the Bayside neighborhood since 2009 is represented by an aggregate total sales value of $428,973,997 (as seen in Figure1). This substantial figure underscores the volume of real estate activity in the neighborhood over the past decade, reflecting a thriving market.

The mean sale price for residential properties in Bayside stands at $833,282.8 (Figure 2), while the mean gross square footage is 2,444.605 square feet (Figure 3). These averages suggest that Bayside's real estate market is characterized by relatively high property values and sizeable residential properties, making it an attractive area for buyers looking for spacious homes.

A five-number summary for sale prices, as seen in Figure 4, reveals a wide range of property values. The minimum sale price recorded is $78,594, while the 1st quartile is $630,000, the median is $795,000, the 3rd quartile is $970,000, and the maximum sale price reaches $6,100,000. This summary indicates a diverse market with properties ranging from more affordable options to high-end luxury homes.

Similarly, the five-number summary for gross square footage in Figure 5, shows a range of property sizes. The minimum recorded square footage is 504 sq ft, the 1st quartile is 1,228 sq ft, the median is 1,632 sq ft, the 3rd quartile is 2,268 sq ft, and the maximum is a significant 489,440 sq ft. This data reflects the availability of both compact and expansive residential properties within Bayside.

The proportion of sales in Figure 6, by property type further highlights Bayside's market characteristics. Residential properties dominate the market, comprising 90.319% of all sales. Commercial properties account for 7.517%, mixed-use properties for 1.595%, and

other types for 0.564%. This overwhelming majority of residential properties underscores Bayside's primary focus as a residential area, with a small but notable presence of commercial real estate.

The standard deviation of sale prices for residential properties is $348,649.8 (Figure 7), indicating significant variability in property values. This high standard deviation suggests that while the average property value is high, there is substantial diversity in sale prices, accommodating a wide range of buyer preferences and financial capabilities.

The correlation between sale price and gross square footage, as indicated in Figure 7, is 0.180964, indicating a positive but relatively low correlation. This suggests that while larger properties tend to be more expensive, other factors such as location, amenities, and property condition also significantly influence property values in Bayside.

K-means Clustering

The K-means clustering analysis was employed to determine the relationships between various New York City neighborhoods, focusing on how Bayside compares with these diverse areas. The clustering analysis considered three key indicators: median sale price, standard deviation of sale prices (SDSalePrice), and the number of units sold. These indicators were chosen to capture the central tendency, variability, and volume of real estate transactions, providing a comprehensive view of the market dynamics in each neighborhood.

The plot generated from the K-means clustering analysis, as seen in Figure 10, illustrates these relationships. The x-axis represents the median sale price, the y-axis denotes the number of units sold, and the size of the points corresponds to the standard deviation of the sale prices, with larger points indicating higher variability. The color gradient represents the cluster assignment, with Bayside being assigned to Cluster 1.

The analysis of the optimal number of clusters was conducted using both the silhouette and WSS (within-cluster sum of squared errors) methods, as observed in figure 8 and 9 respectively. These methods help to determine the most appropriate number of clusters by evaluating the compactness and separation of the clusters. The results from both methods indicated that there are two major clusters within the data, suggesting a natural grouping of the neighborhoods based on the chosen indicators.

In Cluster 1 (Figure 10), which includes Bayside, the data points generally show lower median sale prices and a wide range in the number of units sold. This cluster is characterized by neighborhoods with more moderate property values and substantial variability in the number of units sold. The presence of large points within this cluster indicates neighborhoods with higher standard deviations in sale prices, reflecting significant variability in property values. Bayside's inclusion in this cluster suggests that it shares market characteristics with other neighborhoods that have moderate sale prices and varied transaction volumes.

In contrast, Cluster 2 (Figure 10), which comprises neighborhoods with higher median sale prices, typically shows a different market dynamic. These neighborhoods often have fewer units sold but at higher prices, indicating a more exclusive market segment. The smaller point sizes in this cluster suggest lower standard deviations, meaning that property values in these neighborhoods are more consistent and less variable.

The clustering analysis provides valuable insights into Bayside's position within the broader New York City real estate market. Bayside's inclusion in Cluster 1 highlights its role as a neighborhood with diverse and active market activity, characterized by moderate property values and significant variability.

T-Test Analysis

As a final measurement, the sale prices in the neighborhoods of Bayside and Astoria, both located in the Queens Borough of New York City, were compared using a two-sample t-test. This statistical test was employed to determine whether the neighborhoods represent similar samples from the same population (the null hypothesis) or if they differ from each other in a statistically significant way. The results of this t-test provide important insights into the real estate market dynamics of these two neighborhoods, as observed in Figure 11.

The two-sample t-test conducted compared the sale prices between Bayside and Astoria. The test yielded a t-value of -8.8412 with degrees of freedom (df) of 3411.3, and a p-value of almost zero (< 0.0000000000000022). This extremely low p-value indicates that the null hypothesis should be rejected, signifying a statistically significant difference in the sale prices between the two neighborhoods.

Additionally, the 95% confidence interval for the difference in means does not include zero, ranging from -642560.3 to -409295.3. This further supports the conclusion that there is a significant difference in the average sale prices between Bayside and Astoria. The sample estimates show that the mean sale price in Bayside is $833,282.8, whereas the mean sale price in Astoria is substantially higher at $1,359,210.6.

Conclusion

The comprehensive descriptive analysis of Bayside's residential real estate market reveals a thriving and diverse market with substantial sales activity and high property values. Bayside showcases a dynamic market with a wide range of property values and sizes. The K-means clustering analysis highlights Bayside's unique position, grouping it with neighborhoods that have moderate property values and substantial variability in sales. This underscores Bayside's role as an active and varied market, attractive for diverse investment opportunities. Additionally, the T-test comparison with Astoria indicates Bayside's more

moderately priced market compared to Astoria's higher-end market. Overall, these analyses emphasize Bayside's robust and versatile real estate market within the Queens Borough, offering valuable insights for investors and developers.

Appendix

Figure 1

The total number of sales in your Bayside since 2009

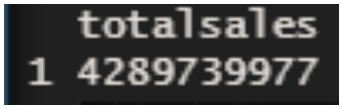

Figure 2

The mean sale price for residential properties in Bayside since 2009

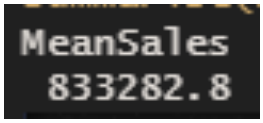

Figure 3

The mean gross square footage for residential properties in Bayside since 2009

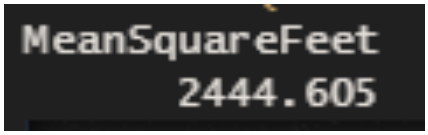

Figure 4

A five-number summary for sale price

```
> fivenum(BAYSIDE_RES$SALE_PRICE)
[1]    78594   630000   795000   970000 6100000
```

Figure 5

A five-number summary for gross square footage

```
> fivenum(BAYSIDE_RES$GROSS_SQUARE_FEET)
[1]     504    1228    1632    2268 489440
```

Figure 6

The proportion of sales of residential, commercial, mixed, and other properties in Bayside since 2009

```
> PROPORTION*100
 COMMERCIAL       MIXED       OTHER RESIDENTIAL
  7.5173993   1.5952458   0.5684854  90.3188694
>
```

Figure 7

The standard deviation of sale prices and correlation between sale price and gross square feet for residential properties in Bayside neighborhood

	MEAN SALE PRICE	MEAN GROSS SQ FEET	MEDIA SALE PRICE	MEDIAN GROSS SQ FEET	SD SALE PRICE	SD GROSS SQ FEET	Cor Sale Price	Cor Gross Sq Feet
BAYSIDE	833282.8	2444.605	795000	1632	348649.8	14034.33	0.1809694	833282.8

Figure 8

The silhouette method

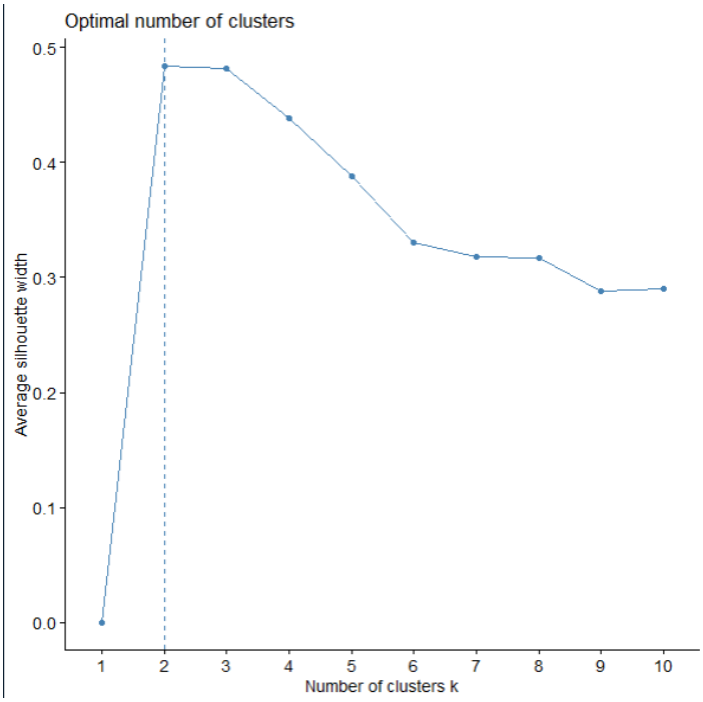

Figure 9

The WSS method

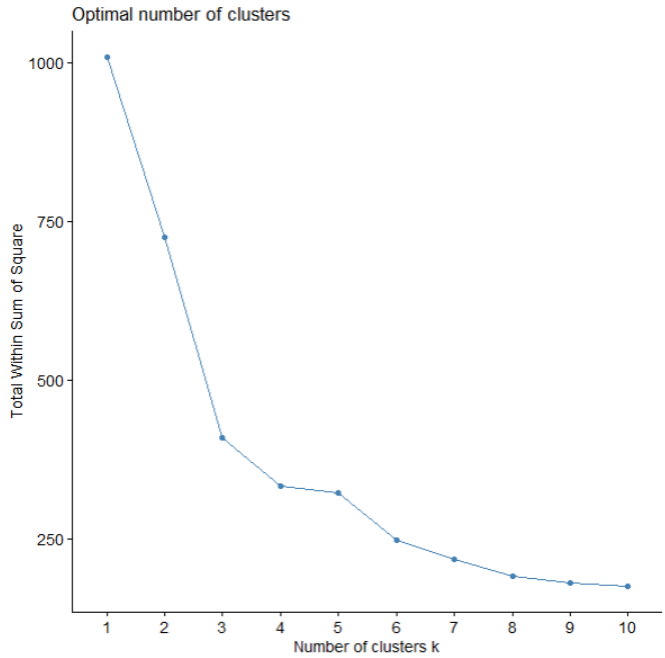

Figure 10

The K-means Clustering

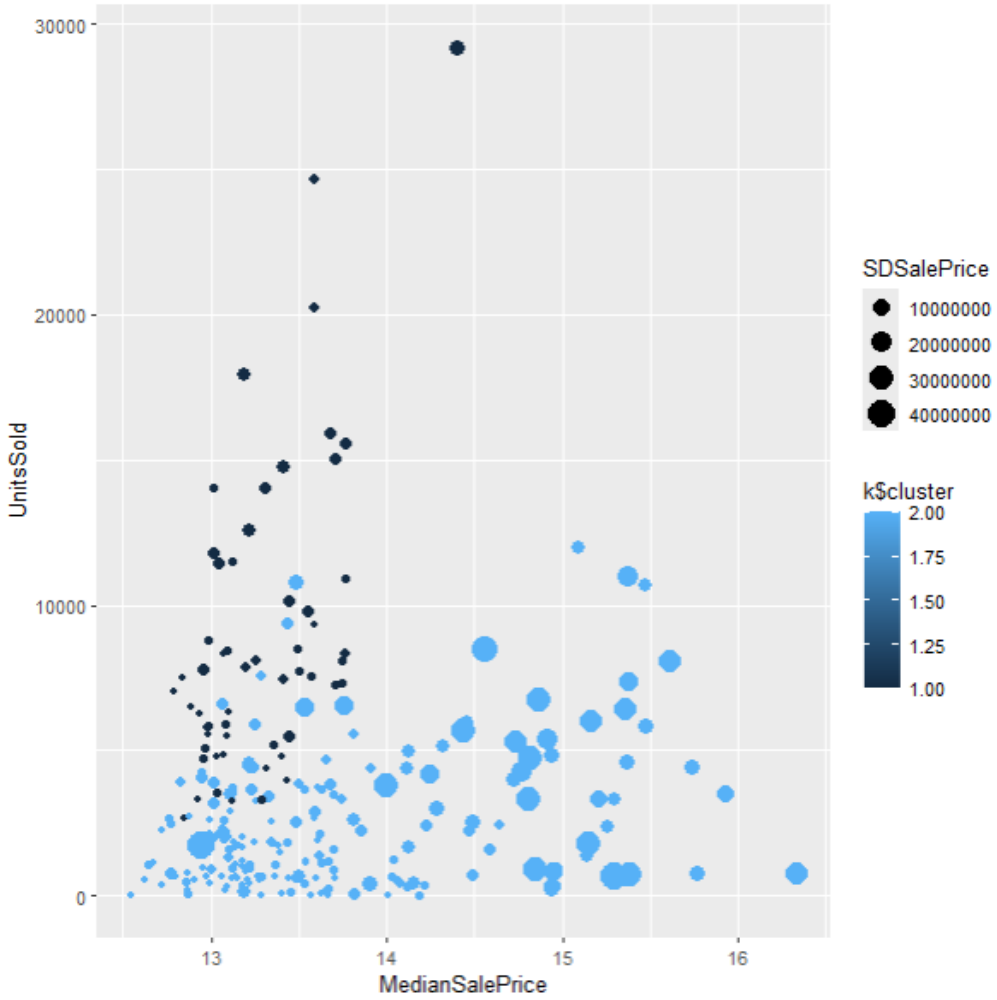

Figure 11

The T-test Analysis

```
        Welch Two Sample t-test

data:  BaysideData$SALE_PRICE and AstoriaData$SALE_PRICE
t = -8.8412, df = 3411.3, p-value < 0.00000000000000022
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -642560.3 -409295.3
sample estimates:
mean of x mean of y
 833282.8 1359210.6
```