Assignment 2
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 19th, 2024

Table of Contents

Executive Summary **3**

Introduction to Bayside: A Descriptive Analysis for Real Estate **4**

Annual Trends: Average Price per Square Foot for Residential Real Estate **5**

Filtering the Data for Accurate & Precise Insights **6**

Comparing The Bayside's Raw Vs Cleaned Data **7**

Analysing The Historical Market Trends **7**

Neighborhood Comparison: Bayside vs. Others **9**

Correlation Analysis between sale price, gross square feet, and year **10**

Recommendation and Conclusion **11**

References **12**

Appendix **13**

Figure 1 13

Figure 2 14

Figure 3 15

Figure 4 15

Figure 5 16

Figure 6 16

Figure 7 17

Executive Summary

The New York City real estate market is thriving, and opening a new office requires a strategic approach grounded in robust data analysis. This report employs a descriptive analytics approach to determine the optimal neighborhood for expansion, focusing on Bayside within the Queens Borough. Essential data was imported from the metsql server into the R environment, enabling detailed examination and comparative analyses. Metrics such as property values and sales trends were scrutinized to position Bayside against other neighborhoods in Queens, providing a clear picture of its real estate landscape and identifying key opportunities and potential challenges.

The analysis revealed several key insights. The correlation matrix highlighted strong positive relationships between sale prices and gross square footage, indicating that larger properties tend to command higher prices. Also, there was a moderate positive trend in sale prices over the years, reflecting a growing market. The data suggests that investing in a brokerage firm in NYC's Bayside neighborhood is promising, as the increasing average sale prices demonstrate positive market growth and significant potential for future investments.

Focusing on the Bayside neighborhood in Queens Borough, the data shows that Bayside outperforms other neighborhoods in Queens in terms of sales performance. This superior performance highlights Bayside as a promising location with better-than-average market dynamics. By setting up a brokerage firm in Bayside, the firm can capitalize on its robust market activity, leveraging the neighborhood's strong sales performance to attract clients and drive business growth. This strategic decision ensures that the firm's expansion is both profitable and sustainable, positioning Bayside as a competitive and attractive option within the Queens Borough.

Introduction to Bayside: A Descriptive Analysis for Real Estate

As the New York City real estate market continues to thrive and evolve, the decision to open a new office requires a strategic approach grounded in robust data analysis. Our real estate brokerage firm aims to make informed decisions by leveraging advanced business analytics to identify the optimal neighborhood for expansion. Given the diverse range of properties and fluctuating sales trends across NYC, a thorough examination of each potential location is imperative.

In this report, we delve into a comprehensive analysis of Bayside, a neighborhood within the Queens Borough, to determine its viability as the site for our new office. The journey began with importing essential data from the metsql server into the R environment. This foundational step ensured that all relevant information was accurately captured and ready for detailed examination. Utilizing R's powerful analytical capabilities, I conducted a series of comparative analyses to position Bayside against other neighborhoods in Queens. This involved scrutinizing various metrics, including property values and sales trends, to paint a clear picture of Bayside's real estate landscape. By systematically evaluating these factors, we can identify key opportunities and potential challenges unique to this neighborhood.

Through this descriptive analytics approach, the aim is to provide actionable insights and sound recommendations that will guide our strategic decision-making process. The objective is to assess the current state of the market in Bayside, ensuring that our firm's expansion is both profitable and sustainable.

Annual Trends: Average Price per Square Foot for Residential Real Estate

The table presented in Figure 1, provides a detailed overview of residential real estate sales in a specific neighborhood from the year 2003 to 2021. This data is crucial for understanding the trends and dynamics of the real estate market over time. The table comprises four key columns including, SaleYear, TotalSales, TotalSqft, and Avg.

The 'SaleYear' column lists the years, ranging from 2003 to 2021, indicating the period over which the data was collected. This chronological arrangement allows for a clear view of how the market has evolved over the years. The 'TotalSales' column shows the total sales value in dollars for each year. This figure reflects the cumulative amount of money generated from residential real estate transactions in the neighborhood for each respective year. Observing these values can help identify years with higher or lower market activity, indicating possible economic or market shifts. The 'TotalSqft' column indicates the total square footage of residential real estate sold each year. This data is essential for assessing the volume of real estate activity, providing insight into the size of properties being sold and the overall market demand. The 'Avg' column provides the average price per square foot, which is calculated by dividing the 'TotalSales' by the 'TotalSqft' for each year. This metric is particularly useful for evaluating the relative value of real estate in the neighborhood, giving a clear picture of how prices have fluctuated over time.

However, it is important to note that in some years, specifically 2018, 2020, and 2021, the Avg column contains "NA" values. This indicates that the data for these years is incomplete or not filtered properly, resulting in missing average price per square foot calculations. The presence of NA values suggests that further data cleaning and filtering are necessary to obtain accurate and comprehensive insights.

Filtering the Data for Accurate & Precise Insights

After an initial analysis of the residential real estate data, it became evident that some years contained incomplete or unfiltered data, resulting in "NA" values in the average price per square foot column. To ensure the accuracy and reliability of our findings, we undertook a thorough data cleaning process. This process involved filtering out erroneous entries, correcting inconsistencies, and ensuring that all data points were complete and accurate.
As seen in Figure 2, the cleaned data is now reflected in the updated table, which provides a clearer and more precise picture of the average price per square foot of residential real estate in our neighborhood from 2003 to 2021.

The Avg column now reliably provides the average price per square foot, calculated by dividing TotalSales by TotalSqft for each year (See Figure 2).
The summary statistics of the average price per square foot (Avg) reveal further insights:

- The minimum average price per square foot recorded is $159.1.
- The first quartile (1st Qu.) is $259.8, indicating that 25% of the values lie below this price.
- The median value is $350.4, representing the midpoint of the dataset.
- The mean (average) value across all years is $352.7.
- The third quartile (3rd Qu.) is $427.2, showing that 75% of the values are below this price.
- The maximum average price per square foot observed is $564.9.

This cleaned and validated data provides a robust foundation for further analysis and decision-making. It allows us to accurately assess trends, identify key opportunities, and make well-informed recommendations regarding the potential for opening a new office in the neighborhood. By ensuring the integrity of our data, we can confidently rely on these insights to guide our strategic planning and real estate investment decisions.

Comparing The Bayside's Raw Vs Cleaned Data

The graph in Figure 3 provides a visual comparison between the cleaned and uncleaned data for Bayside's residential properties, spanning from 2003 to 2021. The x-axis represents the years, while the y-axis indicates the average price per square foot (Avg). The two lines on the graph correspond to the cleaned data (CLEAN_BAYSIDE) and the uncleaned data (BAYSIDE), allowing for a clear side-by-side comparison of the datasets.

From the graph in figure 3, it is evident that the uncleaned data (represented by the orange line in Figure 3) exhibits several anomalies and inconsistencies. There are noticeable fluctuations and erratic patterns, particularly in the years where NA values were initially present. These irregularities underscore the importance of data cleaning, as they can lead to misleading interpretations and conclusions about market trends.

In contrast, the cleaned data (represented by the blue line in Figure 3) presents a more stable and coherent trend over the years. While there are still fluctuations, they are more consistent with typical market behavior and provide a reliable basis for analysis. The cleaned data ensures that the average price per square foot is accurately represented, eliminating the distortions caused by incomplete or erroneous data entries.

Analysing The Historical Market Trends

The Figure 4 presents the past performance of the residential market in Bayside, showcasing the average price per square foot from 2003 to 2021. This graph highlights several key trends and notable dips, reflecting broader economic events and local market conditions that have significantly impacted the real estate landscape in this neighborhood.

The 2006 Market Correction

One of the most striking features of the graph is the sharp dip in 2006. This decline can be attributed to the housing market correction that began in 2005 (Boykin, 2023).

The previous decade had seen an unusual surge in property values, particularly in downstate areas, driven by the wide availability of credit. This rapid increase in property values led to a housing price bubble that ultimately proved unsustainable. The collapse of the subprime mortgage industry, fueled by bad lending practices, triggered this correction. As credit markets tightened and the economy slowed, there was a sharp decline in market values, which is evident in the significant drop in average price per square foot in Bayside during this period (Boykin, 2023).

The 2009 Recession

The graph also highlights another steep decline in 2009, coinciding with the global financial crisis and the recession that severely impacted the U.S. real estate market. The recession caused widespread economic instability, leading to a significant drop in real estate values across the country (Wikipedia, 2024). Bayside, like many other neighborhoods in New York City, was not immune to these effects. The 2009 recession resulted in a substantial reduction in property values, as indicated by the notable decrease in the average price per square foot in Figure 4. This period marked a challenging time for the real estate market, with recovery taking several years.

The Dip in 2016

Additionally, the graph shows a sharp dip in 2016, which can be linked to specific local market dynamics. In 2016, there was a considerable decrease in the number of building permits issued for new housing units citywide, following a spike in 2015 as developers rushed to begin projects before the expiration of the 421-a property tax exemption (Furman, 2017). This reduction in new housing permits led to a decline in property sales, with Queens experiencing one of the largest drops in sales volume at -17.2% (Furman, 2017). This decrease in market activity is reflected in the significant dip in average price per square foot for Bayside in 2017, as seen in Figure 4.

Neighborhood Comparison: Bayside vs. Others

The Figure 5 compares the average sale price per square foot across several neighborhoods in the Queens Borough, including Astoria, College Point, Flushing-South, Flushing-North, and Bayside, represented by CLEAN_BAYSIDE. This comparative analysis provides a comprehensive view of how these neighborhoods have performed over the years, highlighting both the volatility and growth trends within the residential real estate market.

From the Figure 5, it is evident that Bayside, as shown by the CLEAN_BAYSIDE line, has exhibited a relatively stable performance with a notable dip in 2016. Despite this downturn, Bayside has demonstrated resilience and recovery, maintaining a consistent upward trend in average sale prices over the subsequent years. This recovery and growth suggest that Bayside has managed to rebound from market fluctuations more effectively compared to some of the other neighborhoods. This stability is particularly impressive given the broader economic challenges faced by the real estate market during this period.

In comparison, neighborhoods like Astoria, College Point, and the Flushing areas have shown more variability in their average sale prices. While some of these areas experienced significant peaks and troughs, Bayside's performance indicates a more reliable and steady increase in property values.

As seen in Figure 7, when comparing the median average sale prices, Bayside's median of $350.4 per square foot is higher than those of Flushing-South ($341.4), Flushing-North ($300.8), College Point ($279.1), and Astoria ($264.7).

The mean average sale price per square foot in Bayside Neighborhood (CLEAN_BAYSIDE) is $352.7, which is higher than the mean values for Flushing-South ($366.39), Flushing-North ($399.54), College Point ($321.13), and Astoria ($311.94). This indicates that Bayside generally commands higher property values, showcasing its relative stability and potential as a lucrative real estate market (See Figure 7).

Also, Bayside was not greatly impacted by the COVID-19 pandemic's effect on real estate, unlike other neighborhoods in Queens that experienced more pronounced volatility. The consistent growth trend observed in Bayside underscores its potential as a future market opportunity, making it an attractive option for real estate investment. This steady rise in average sale prices signals a healthy demand and a robust market, positioning Bayside as a competitive neighborhood within the Queens Borough.

Correlation Analysis between sale price, gross square feet, and year

The correlation matrix in Figure 6 visualizes the relationships between three variables: SALE_PRICE, GROSS_SQUARE_FEET, and SaleYear. This graphical representation helps to understand the degree and direction of the linear relationships among these variables. The correlation matrix uses circles of varying sizes and colors to convey this information effectively.

The correlation between SALE_PRICE and GROSS_SQUARE_FEET is depicted by a large, dark blue circle. This indicates a strong positive correlation, meaning that as the gross square footage of a property increases, the sale price tends to increase as well. The strong correlation suggests that property size is a significant factor in determining sale price. This relationship is expected in real estate markets, where larger properties generally command higher prices.

In contrast, the correlation between SALE_PRICE and SaleYear is represented by a smaller, lighter blue circle, indicating a weak to moderate positive correlation. This suggests that while there is a trend of increasing sale prices over the years, the relationship is not as strong as that between SALE_PRICE and GROSS_SQUARE_FEET. The weak to moderate correlation may reflect market conditions, economic factors, or other variables influencing property prices over time.

The relationship between GROSS_SQUARE_FEET and SaleYear is shown by a light blue circle, indicating a weak positive correlation. This implies that the gross square footage of properties sold has slightly increased over the years. While the increase in property size over time is minimal, it suggests a trend towards larger properties being sold in the market, albeit not as pronounced.

Recommendation and Conclusion

After employing a descriptive analytics approach and analyzing the real estate data through R, several key insights have emerged. The correlation matrix highlighted strong positive relationships between sale prices and gross square footage, indicating that larger properties tend to command higher prices. Also, the data suggested a moderate positive trend in sale prices over the years, reflecting a growing market. These findings underscore the relevance of property size in determining sale prices and reveal subtle trends of increasing property sizes over time.

Based on this analysis, I recommend investing in a brokerage firm in the Bayside neighborhood. The increasing average sale prices demonstrate positive market growth, indicating significant potential for future investments. The data analysis reveals that Bayside outperforms other neighborhoods in Queens in terms of sales performance. This superior performance highlights Bayside as a promising location with better-than-average market dynamics. By setting up a brokerage firm in Bayside, one can capitalize on its robust market activity, leveraging the neighborhood's strong sales performance to attract clients and drive business growth.

References

- Boykin, R. (2023, August 31). *The Great Recession's Impact on the Housing Market*. Investopedia. https://www.investopedia.com/investing/great-recessions-impact-housing-market/#:~:text=In%202006%2C%20the%20housing%20market,securities%20backing%20up%20subprime%20mortgages

- Wikipedia contributors. (2024, July 11). *Subprime mortgage crisis*. Wikipedia. https://en.wikipedia.org/wiki/Subprime_mortgage_crisis#:~:text=Two%20proximate%20causes%20were%20the,non%2Dinvestors%20when%20prices%20fell.

- Furman Center Blog. (2017). *Citywide sales volume down in 2016 as housing prices continue to rise*. NY Furman Center. https://furmancenter.org/thestoop/entry/citywide-sales-volume-down-in-2016-as-housing-prices-continue-to-rise

Appendix

Figure 1

Bayside's Residential Properties' Average Price Per Square Footage (Uncleaned Data)

	SaleYear	TotalSales	TotalSqft	Avg
1	2003	407395469	2383347	170.9342
2	2004	519608747	2492941	208.4320
3	2005	525094852	1927997	272.3525
4	2006	461128025	2497922	184.6047
5	2007	558765332	1658384	336.9336
6	2008	392535122	1211058	324.1258
7	2009	319881237	1258131	254.2511
8	2010	312026844	2127604	146.6564
9	2011	291578954	1101462	264.7199
10	2012	317141577	1282119	247.3574
11	2013	420393524	1339577	313.8256
12	2014	443276303	1209664	366.4458
13	2015	473675361	1573188	301.0927
14	2016	492577854	1195455	412.0422
15	2017	592582127	3462473	171.1442
16	2018	579548039	1386750	417.9182
17	2019	500662059	*NA*	*NA*
18	2020	432535339	*NA*	*NA*
19	2021	672888479	*NA*	*NA*

Figure 2

Cleaned Bayside's Residential Properties' Average Price Per Square Footage

	SaleYear	TotalSales	TotalSqft	Avg
1	2003	300924626	1634850	184.0686
2	2004	388345348	1842456	210.7759
3	2005	386992339	1217053	317.9749
4	2006	348548559	1827394	190.7353
5	2007	361397280	1031323	350.4210
6	2008	247915977	695351	356.5336
7	2009	214340931	629515	340.4858
8	2010	248585584	1562067	159.1389
9	2011	214818996	616981	348.1777
10	2012	235499783	663983	354.6774
11	2013	302585752	825824	366.4046
12	2014	307241360	764072	402.1105
13	2015	348476381	1128308	308.8486
14	2016	346715751	766666	452.2383
15	2017	424342151	2611526	162.4882
16	2018	463735818	830725	558.2302
17	2019	354975674	672252	528.0396
18	2020	318353048	584765	544.4119
19	2021	508745883	900547	564.9299

Figure 3

Comparing Clean (CLEAN_BAYSIDE) vs Uncleaned (BAYSIDE) data

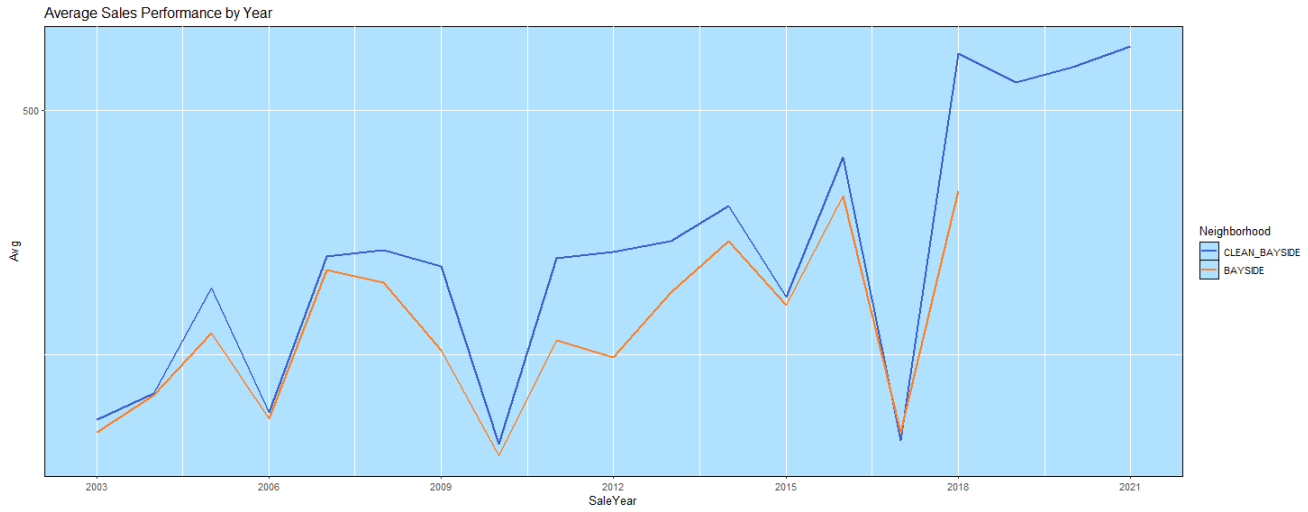

Figure 4

Analysing Bayside (CLEAN_BAYSIDE)'s performance by each year

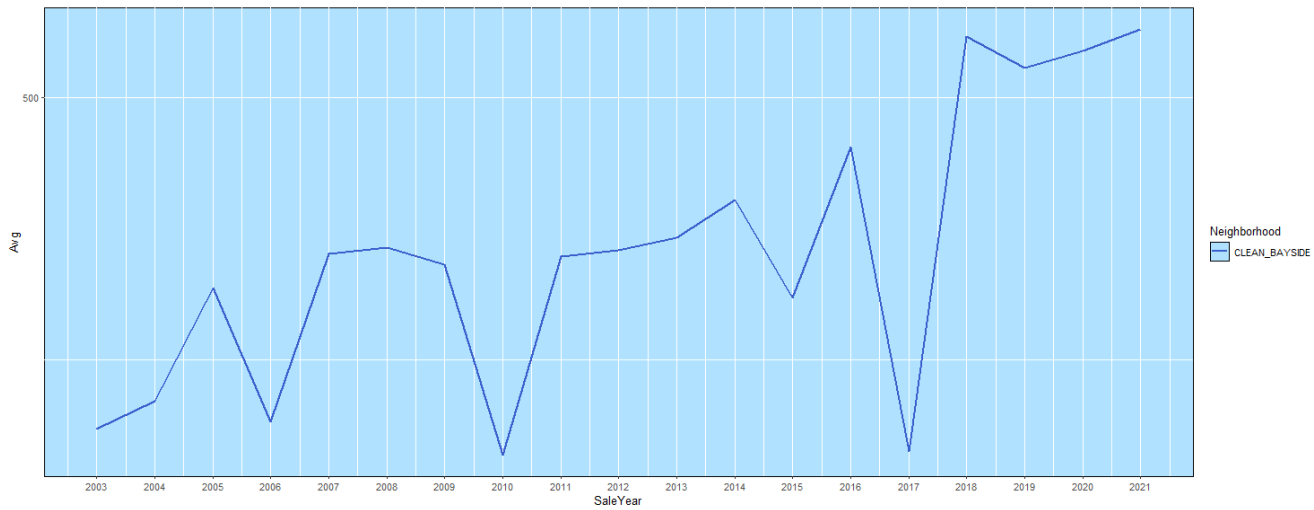

Figure 5

Bayside Neighborhood Comparison with Other Neighborhoods

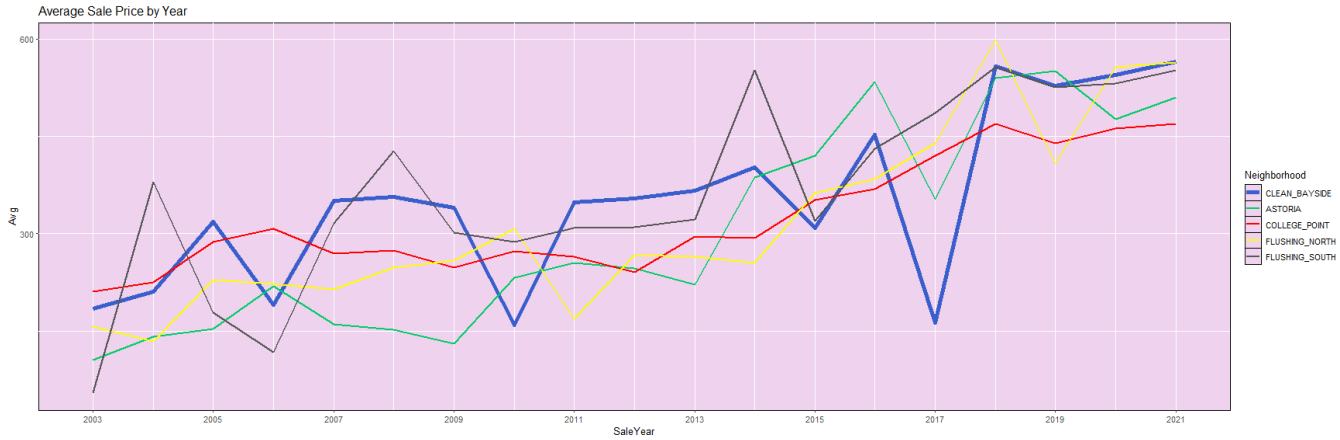

Figure 6

Correlation Analysis between sale price, gross square feet, and year

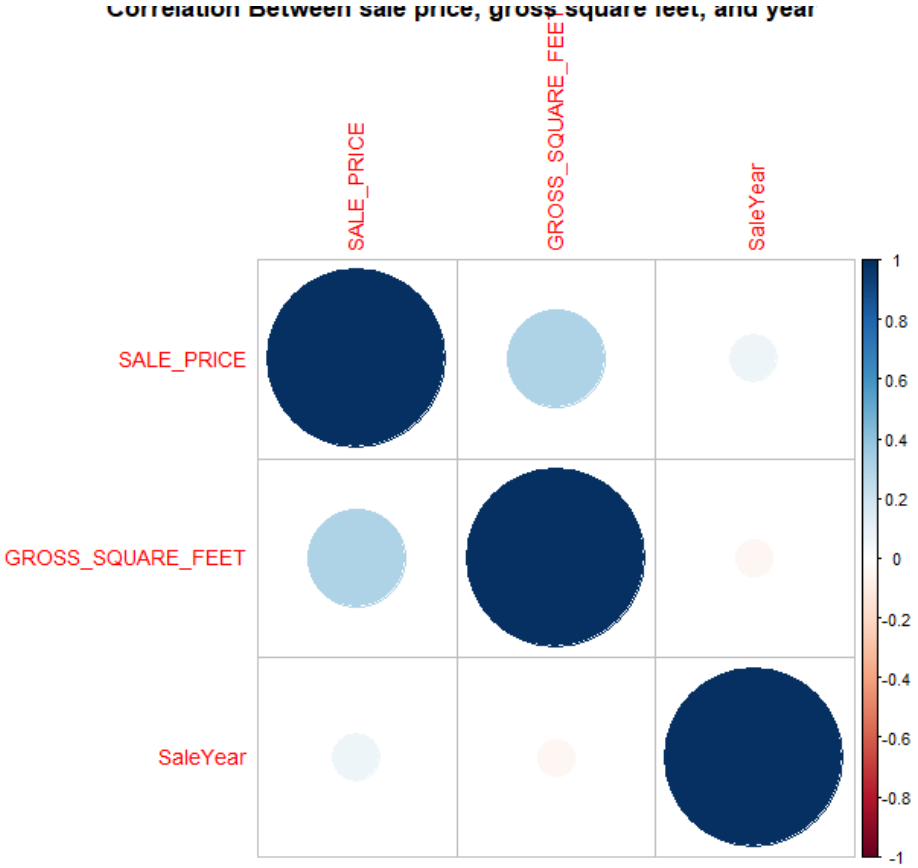

Figure 7

Summary of all the Neighborhoods taken into Comparison

```
> summary(FLUSHING_SOUTH)
    SaleYear      TotalSales          TotalSqft           Avg        
 Min.   :2003   Min.   :123748292   Min.   : 369782   Min.   : 54.94  
 1st Qu.:2008   1st Qu.:196084394   1st Qu.: 464910   1st Qu.:305.59  
 Median :2012   Median :222392208   Median : 553472   Median :321.73  
 Mean   :2012   Mean   :248958128   Mean   : 907766   Mean   :366.39  
 3rd Qu.:2016   3rd Qu.:294382592   3rd Qu.: 908814   3rd Qu.:506.04  
 Max.   :2021   Max.   :543683901   Max.   :4015864   Max.   :557.29  
> summary(FLUSHING_NORTH)
    SaleYear      TotalSales          TotalSqft           Avg        
 Min.   :2003   Min.   :3.714e+08   Min.   : 862714   Min.   :133.7  
 1st Qu.:2008   1st Qu.:4.944e+08   1st Qu.:1744555   1st Qu.:226.0  
 Median :2012   Median :6.661e+08   Median :2193363   Median :264.2  
 Mean   :2012   Mean   :6.581e+08   Mean   :2351253   Mean   :317.9  
 3rd Qu.:2016   3rd Qu.:7.610e+08   3rd Qu.:2761032   3rd Qu.:396.5  
 Max.   :2021   Max.   :1.163e+09   Max.   :5555150   Max.   :598.4  
> summary(COLLEGE_POINT)
    SaleYear      TotalSales          TotalSqft          Avg        
 Min.   :2003   Min.   : 74530853   Min.   :224776   Min.   :211.2  
 1st Qu.:2008   1st Qu.:103321642   1st Qu.:359169   1st Qu.:267.1  
 Median :2012   Median :132306708   Median :397154   Median :292.9  
 Mean   :2012   Mean   :136612083   Mean   :440450   Mean   :324.9  
 3rd Qu.:2016   3rd Qu.:170487922   3rd Qu.:435100   3rd Qu.:394.9  
 Max.   :2021   Max.   :206218632   Max.   :798093   Max.   :469.9  
> summary(ASTORIA)
    SaleYear      TotalSales          TotalSqft           Avg        
 Min.   :2003   Min.   :142384410   Min.   : 558288   Min.   :104.9  
 1st Qu.:2008   1st Qu.:255070072   1st Qu.: 896556   1st Qu.:156.4  
 Median :2012   Median :299232918   Median :1018427   Median :246.9  
 Mean   :2012   Mean   :331176372   Mean   :1278790   Mean   :304.8  
 3rd Qu.:2016   3rd Qu.:423984766   3rd Qu.:1701918   3rd Qu.:448.7  
 Max.   :2021   Max.   :469685764   Max.   :2257507   Max.   :551.4  
> summary(CLEAN_BAYSIDE)
    SaleYear      TotalSales          TotalSqft           Avg        
 Min.   :2003   Min.   :214340931   Min.   : 584765   Min.   :159.1  
 1st Qu.:2008   1st Qu.:274755105   1st Qu.: 683802   1st Qu.:259.8  
 Median :2012   Median :346715751   Median : 830725   Median :350.4  
 Mean   :2012   Mean   :332765328   Mean   :1095035   Mean   :352.7  
 3rd Qu.:2016   3rd Qu.:374194810   3rd Qu.:1389560   3rd Qu.:427.2  
 Max.   :2021   Max.   :508745883   Max.   :2611526   Max.   :564.9  
```