
# MET AD571 Business Analytics Foundations

# Assignment 3

**Student name:** **Kunfei Chen**

**Student number:** **U15575304**

# Table of Contents

## 1. Executive Summary

This report aims to apply R-Studio to import the NYC Real Estate data from the Boston University SQL server, and do conduct the descriptive analytics in terms of different statistics, including the total number of sales, the mean sale price and gross square footage, the five-number summary for both sale price and gross square footage, the proportion of sales for different property types, the standard deviation of sale prices, and the correlation between sale price and gross square feet, and then perform K-mean clustering of all neighborhoods based on three selected Key Performance Indicators. Finally, a comparison of the average residential property costs between Bedford-Stuyvesant and Crown Heights was performed to test the alternative hypothesis that the average sale pries are different in both neighborhoods.

The total number of sales in Bedford-Stuyvesant since 2009 amount to 7711. The mean sales price and size in square feet are $968402.3 and 3366.474 respectively and the standard deviation of sale prices for residential properties fluctuated during the last 11 years. The Pearson correlation coefficient of 0.537 between sale price and gross square feet indicates there is a week correlation existing in the two variables. The K-mean clustering analysis categorized all neighborhoods into 3 clusters which is the best clustering number based on comparing the three KPIs' total squared errors. Bedford-Stuyvesant was assigned to cluster 2 which is the most common one. The result of t-test concluded that there is no statistically difference in average residential property costs between Bedford-Stuyvesant and Crown Heights.

## 2. Task 1 – Descriptive analytics for Bedford-Stuyvesant neighborhood

### 2.1 Question A

The total number of sales in Bedford-Stuyvesant achieves 7711 (See Figure 1) and fluctuates around average 700 per year since 2009 (See Figure 2, Figure 3). It can be seen that, compared with the figure in 2013 and 2014, the traction number is relatively low in recent years such as 2018 and 2019, which means less estates were sold recently.

### 2.2 Question B

Although the traction number in recent years is not high as discussed in Question A, the pattern of mean sale price for residential properties since 2009 seems to be diverse. Despite a certain decrease in sales between 2017 and 2018, the figure in other years show an upward trend especially a surge from $1176125.5 in 2018 to $1842784.6 in 2019 (See Figure 4, Figure 5). As for the pattern of mean gross square footage for residential properties, it experienced a stable decrease within the last 4 years from 3550.055 in 2016 to 2417.339 but an increase to 3208.658 in 2019 (See Figure 4, Figure 6). It is assumed that the economy environment was not good from 2016 to 2018 but in 2019 became better. Generally, the mean sale price and mean gross square footage within the past 11 years were $968402.3 and 3366.474 respectively (See Figure 7).

### 2.3 Question C

The five-number summary (minimum,1st quantile, median, mean, 3rd quantile, maximum), for both sale price and gross square footage for residential properties since was performed (See Figure 8), it is noticeable that the maximum in mean sale price per year is much higher than its 3$^{rd}$ quantile, which means the uncertainty of sale is significant than that of mean gross feet.

### 2.4 Question D

The proportion analysis indicated that residential sales comprise 90.1% of the number of sales, followed by commercial (3%) and mixed (6.2%) in Bedford-Stuyvesant (See Figure 8, Figure 9). This means the sales staff need to pay more attention to the residential market.

### 2.5 Question E

The standard deviation of sale prices for residential properties in Bedford-Stuyvesant since 2009 is shown in Figure 11 and Figure 12. Since the standard deviation always indicates the uncertainty level of a column of data, it is clearly that the sale price in this area became more and more unstable from 2009 to 2016 and from 2018 to 2019.

Whereas during the period between 2016 and 2018, the sale price is more balanced.

### 2.6 Question F

The Pearson correlation coefficient between sale price and gross square feet for residential properties in Bedford-Stuyvesant since 2009 achieves 0.537, which indicates there is a week correlation existing in the two variables (See Figure 13). According to Figure 14, this scatter plot shows that generally the sales price increases along with the increase of gross square feet.

## 3. Test 2 - K-mean Clustering Analysis

Task 2 required to perform k-means clustering, comparing the assigned Bedford-Stuyvesant neighborhood to other neighborhoods. The selected Key Performance Indicators (KPISs) includes Median Sale Price, Standard deviation of sales, Price of 1 gross square foot (See Figure 15). This is because the median reflects the central tendency of the sales, the standard deviation reflects the uncertainty level of the price in a cluster, and the price per gross square foot ensure the estates' unit value in a cluster are similar.

To identify the best number of clusters to be classified, the total squared errors of the three KPIs are compared in terms of different clustering number (See Figure 16). By iterating the figure from 1 to 15, the number of 3 is identified as a proper clustering number.

After running the cluster analysis, all of the neighborhoods are assigned to 3 clusters (See Figure 17). The visualization result is displayed in a form of three-dimensional scatter plot (See Figure 18). The color of cluster 1 is black, and that of cluster 2 and cluster 3 are red and green. It can be seen that Bedford-Stuyvesant is assigned to Cluster 2, which is the most common type. Generally, the neighborhoods in cluster 1 have more transactions with characteristic of cheap price and high stability, whereas those in cluster 2 and cluster 3 seems to be more expensive and unstable.

## 4. Task 3 T-Test for Bedford-Stuyvesant and Crown Heights

Test 3 required choosing a second neighborhood to test the alternative hypothesis that, starting in 2009, the average residential property costs are statistically different from the Bedford-Stuyvesant. The neighborhood to be compared is selected from the areas surrounding Bedford-Stuyvesant (See Figure 19). Within the four nearby neighborhoods, their average price per square footage are compared (See Figure 20). It is clearly that the lines of Bedford-Stuyvesant and Crown Heights are really close, therefore the Crown Heights is selected for the comparison.

Since the data length of the two neighborhoods are different, the parameter 'paired' is set as FALSE. In addition, the alternative hypothesis is true difference in means is not equal to 0, and the confidence interval level is set as 0.95. The t-test result indicates the p-values equals 0.3393, which is much higher than the significance level of 0.05 (See Figure 21). Therefore, the null hypothesis is accepted, which means there is no statistically difference in average residential property costs between the two neighborhoods.

## 5. Appendix

***Figure 1***

| | TotalSales |
|---|---|
| 1 | 7711 |

***Figure 2***

| | Year | TotalSales |
|---|---|---|
| 1 | 2009 | 434 |
| 2 | 2010 | 575 |
| 3 | 2011 | 618 |
| 4 | 2012 | 782 |
| 5 | 2013 | 996 |
| 6 | 2014 | 924 |
| 7 | 2015 | 806 |
| 8 | 2016 | 650 |
| 9 | 2017 | 594 |
| 10 | 2018 | 802 |
| 11 | 2019 | 530 |

*Figure 3*

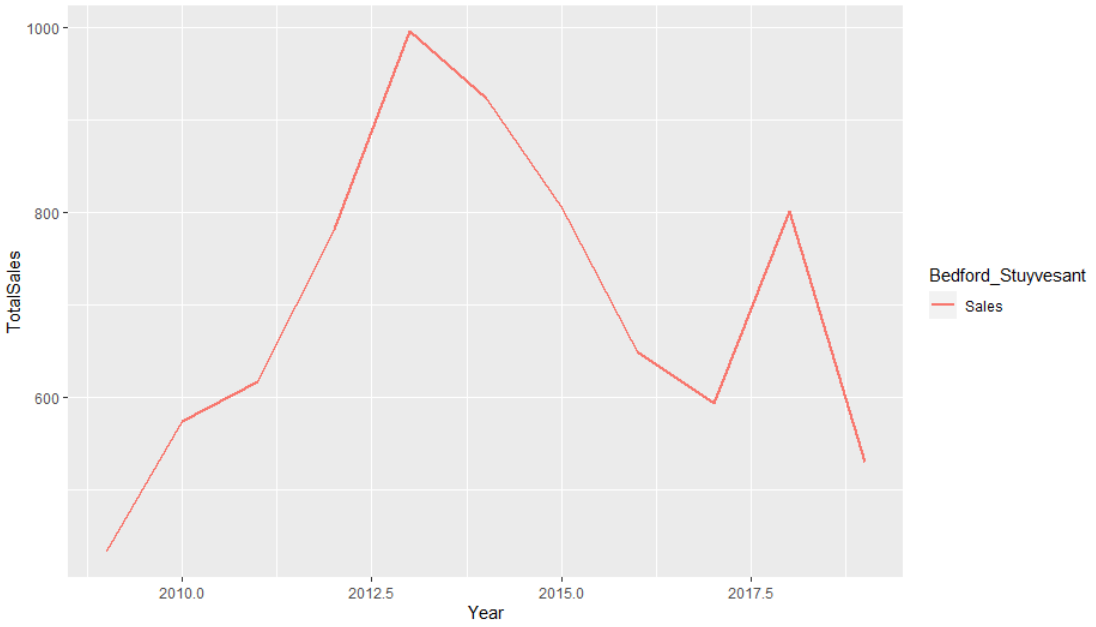


*Figure 4*

| | Year | MeanSalePrice | MeanGrossFeet |
|---|---|---|---|
| 1 | 2009 | 453115.5 | 3217.854 |
| 2 | 2010 | 423181.1 | 3161.783 |
| 3 | 2011 | 473768.0 | 3346.299 |
| 4 | 2012 | 609013.1 | 4321.907 |
| 5 | 2013 | 698078.5 | 3351.289 |
| 6 | 2014 | 951364.0 | 3574.417 |
| 7 | 2015 | 1188471.9 | 3515.146 |
| 8 | 2016 | 1441039.4 | 3550.055 |
| 9 | 2017 | 1395483.6 | 3366.467 |
| 10 | 2018 | 1176125.5 | 2417.339 |
| 11 | 2019 | 1842784.6 | 3208.658 |

*Figure 5*

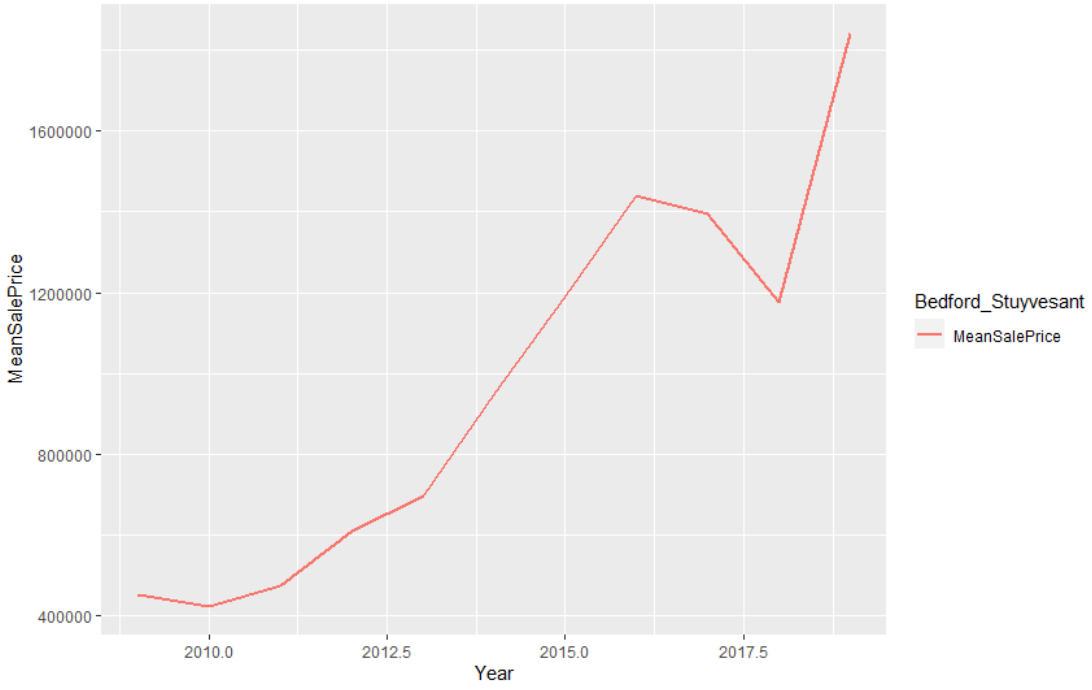


*Figure 6*

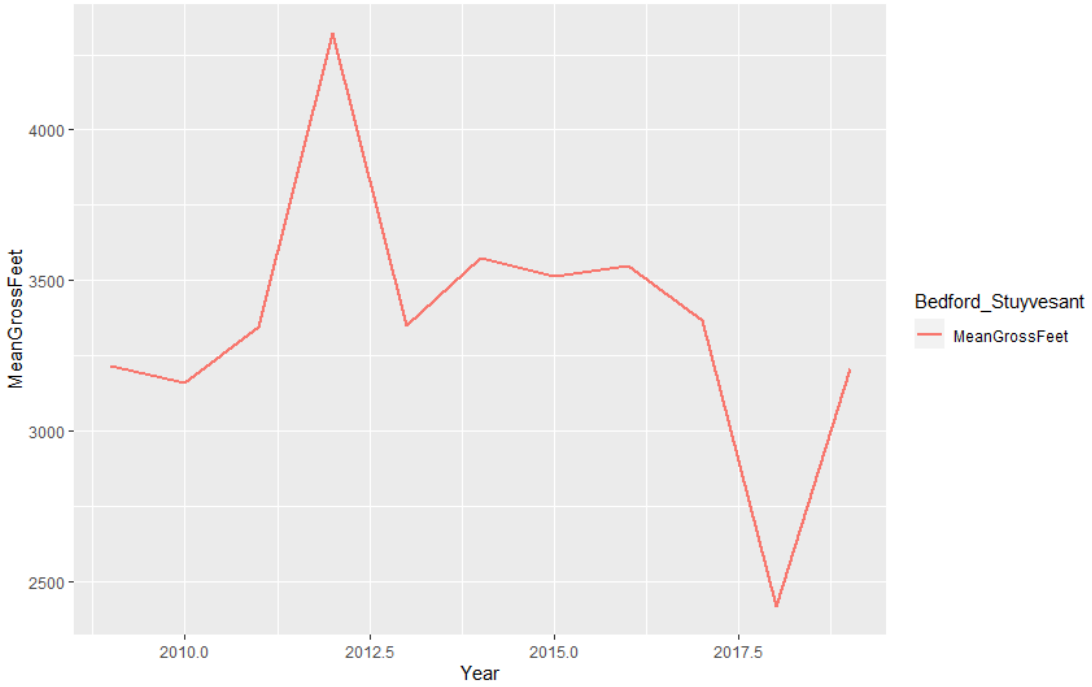

*Figure 7*

| | MeanSalePrice | MeanGrossFeet |
|---|---|---|
| 1 | 968402.3 | 3366.474 |

*Figure 8*

```
MeanSalePrice     MeanGrossFeet
Min.    : 423181   Min.    :2417
1st Qu.: 541391   1st Qu.:3213
Median : 951364   Median :3351
Mean    : 968402   Mean    :3366
3rd Qu.:1291978   3rd Qu.:3533
Max.    :1842785   Max.    :4322
```

*Figure 9*

| | TYPE | TotalUnits | proportion |
|---|---|---|---|
| 1 | COMMERCIAL | 235 | 0.030475943 |
| 2 | MIXED | 477 | 0.061859681 |
| 3 | OTHER | 50 | 0.006484243 |
| 4 | RESIDENTIAL | 6949 | 0.901180132 |

*Figure 10*

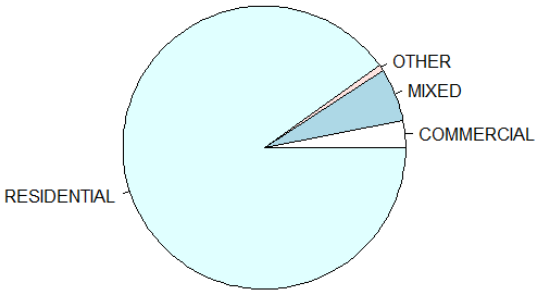

*Figure 11*

| | Year | sdSalePrice |
|---|---|---|
| 1 | 2009 | 220651.8 |
| 2 | 2010 | 250612.2 |
| 3 | 2011 | 502043.7 |
| 4 | 2012 | 1083127.4 |
| 5 | 2013 | 1126361.7 |
| 6 | 2014 | 1553690.6 |
| 7 | 2015 | 1398389.8 |
| 8 | 2016 | 2913349.1 |
| 9 | 2017 | 1326170.0 |
| 10 | 2018 | 978619.4 |
| 11 | 2019 | 2451147.3 |

*Figure 12*

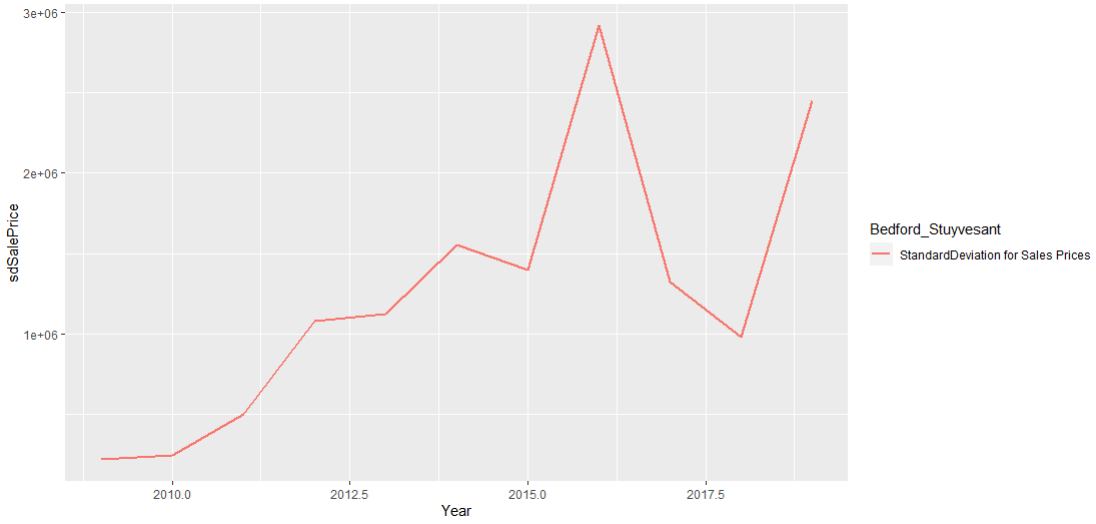


*Figure 13*

| | SALE_PRICE | GROSS_SQUARE_FEET |
|---|---|---|
| SALE_PRICE | 1.0000000 | 0.5365923 |
| GROSS_SQUARE_FEET | 0.5365923 | 1.0000000 |

*Figure 14*

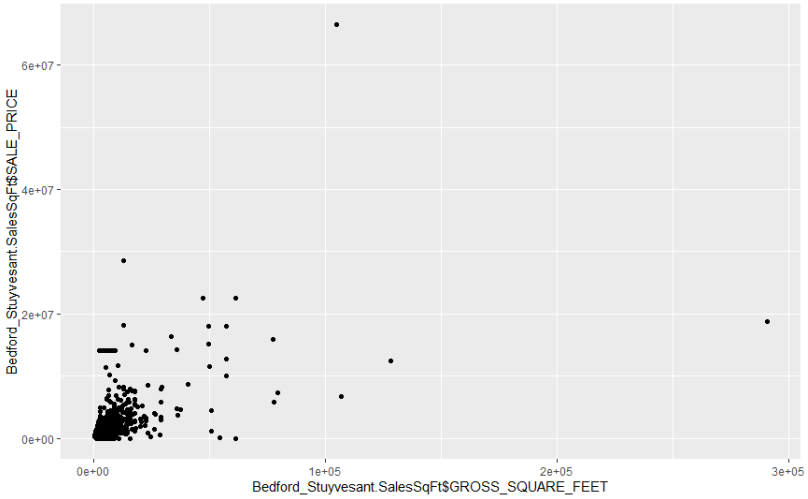


*Figure 15*

| | NEIGHBORHOOD_NAME | PricePerFeet | MedSales | sdSalePrice |
|---|---|---|---|---|
| 1 | AIRPORT LA GUARDIA | 397.55455 | 620000.0 | 233913.4 |
| 2 | ALPHABET CITY | 528.74215 | 3500000.0 | 9959156.7 |
| 3 | ANNADALE | 270.16728 | 625000.0 | 326633.4 |
| 4 | ARDEN HEIGHTS | 239.35516 | 350000.0 | 165104.8 |
| 5 | ARROCHAR | 255.72744 | 456999.5 | 208447.2 |
| 6 | ARROCHAR-SHORE ACRES | 265.00952 | 431500.0 | 223840.0 |
| 7 | ARVERNE | 177.99465 | 370045.5 | 2950295.6 |
| 8 | ASTORIA | 330.07098 | 840000.0 | 3607671.9 |
| 9 | BATH BEACH | 327.84127 | 771000.0 | 888847.4 |
| 10 | BATHGATE | 147.88803 | 410000.0 | 1364276.6 |
| 11 | BAY RIDGE | 348.20272 | 875000.0 | 1049086.6 |
| 12 | BAYCHESTER | 193.46262 | 390000.0 | 585837.2 |
| 13 | BAYSIDE | 310.76918 | 750250.0 | 348402.1 |
| 14 | BEDFORD PARK/NORWOOD | 120.54310 | 550000.0 | 2714942.3 |
| 15 | BEDFORD STUYVESANT | 285.36390 | 700000.0 | 1525509.4 |
| 16 | BEECHHURST | 428.64579 | 858000.0 | 390372.3 |
| 17 | BELLE HARBOR | 318.50221 | 700000.0 | 326250.8 |
| 18 | BELLEROSE | 345.38926 | 505000.0 | 145479.8 |
| 19 | BELMONT | 177.97921 | 499000.0 | 2781260.0 |
| 20 | BENSONHURST | 341.23827 | 860000.0 | 578096.1 |
| 21 | BERGEN BEACH | 303.03369 | 640000.0 | 1184005.2 |
| 22 | BLOOMFIELD | 269.63323 | 499900.0 | NA |
| 23 | BOERUM HILL | 726.53874 | 2000000.0 | 3133401.2 |
| 24 | BOROUGH PARK | 331.31723 | 900000.0 | 910852.6 |
| 25 | BRIARWOOD | 165.93322 | 570000.0 | 1500490.5 |
| 26 | BRIGHTON BEACH | 216.86699 | 585000.0 | 3340441.0 |
| 27 | BROAD CHANNEL | 235.64544 | 318453.5 | 146894.8 |
| 28 | BRONX PARK | 181.95615 | 515000.0 | 6285308.7 |
| 29 | BRONXDALE | 174.02215 | 408900.0 | 1851926.1 |

*Figure 16*

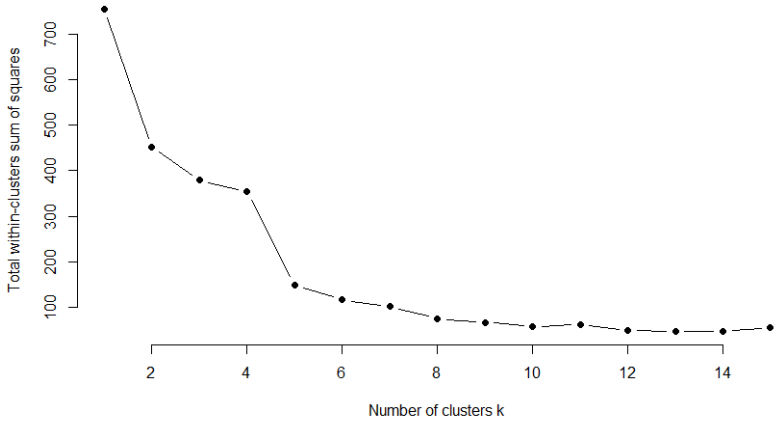


*Figure 17*

| | NEIGHBORHOOD_NAME | PricePerFeet | MedSales | sdSalePrice | V4 |
|---|---|---|---|---|---|
| 1 | AIRPORT LA GUARDIA | 0.288608218 | -0.369821539 | -0.222963893 | 2 |
| 2 | ALPHABET CITY | 0.896091928 | 1.736217686 | 0.196085465 | 1 |
| 3 | ANNADALE | -0.301277516 | -0.366165220 | -0.218968697 | 2 |
| 4 | ARDEN HEIGHTS | -0.443957633 | -0.567262716 | -0.225928778 | 2 |
| 5 | ARROCHAR | -0.368143346 | -0.489017874 | -0.224061203 | 2 |
| 6 | ARROCHAR-SHORE ACRES | -0.325161293 | -0.507664731 | -0.223397948 | 2 |
| 7 | ARVERNE | -0.728096610 | -0.552604171 | -0.105918164 | 2 |
| 8 | ASTORIA | -0.023884533 | -0.208943542 | -0.077592586 | 2 |
| 9 | BATH BEACH | -0.034209519 | -0.259400732 | -0.194743553 | 2 |
| 10 | BATHGATE | -0.867509794 | -0.523386899 | -0.174257866 | 2 |
| 11 | BAY RIDGE | 0.060077211 | -0.183349315 | -0.187839037 | 2 |
| 12 | BAYCHESTER | -0.656469866 | -0.538012171 | -0.207799911 | 2 |
| 13 | BAYSIDE | -0.113264379 | -0.274574452 | -0.218030712 | 2 |
| 14 | BEDFORD PARK/NORWOOD | -0.994134551 | -0.421009992 | -0.116059260 | 2 |
| 15 | BEDFORD STUYVESANT | -0.230907291 | -0.311320449 | -0.167310532 | 2 |
| 16 | BEECHHURST | 0.432580837 | -0.195780797 | -0.216222263 | 2 |
| 17 | BELLE HARBOR | -0.077455411 | -0.311320449 | -0.218985185 | 2 |
| 18 | BELLEROSE | 0.047049083 | -0.453916855 | -0.226774394 | 2 |
| 19 | BELMONT | -0.728168100 | -0.458304436 | -0.113201708 | 2 |
| 20 | BENSONHURST | 0.027827282 | -0.194318270 | -0.208133463 | 2 |
| 21 | BERGEN BEACH | -0.149084720 | -0.355196266 | -0.182025552 | 2 |
| 22 | BOERUM HILL | 1.812018509 | 0.639322257 | -0.098028355 | 1 |
| 23 | BOROUGH PARK | -0.018113554 | -0.165067725 | -0.193795377 | 2 |
| 24 | BRIARWOOD | -0.783948851 | -0.406384719 | -0.168388569 | 2 |
| 25 | BRIGHTON BEACH | -0.548092430 | -0.395415765 | -0.089107254 | 2 |
| 26 | BROAD CHANNEL | -0.461136039 | -0.590331524 | -0.226713426 | 2 |
| 27 | BRONX PARK | -0.709752294 | -0.446604219 | 0.037783650 | 2 |
| 28 | BRONXDALE | -0.746491844 | -0.524191289 | -0.153245621 | 2 |
| 29 | BROOKLYN-UNKNOWN | 0.525959509 | -0.091942094 | -0.089825571 | 2 |

*Figure 18*


Retrieved (2020) from https://www.nyctourist.com/brooklyn-map

*Figure 19*

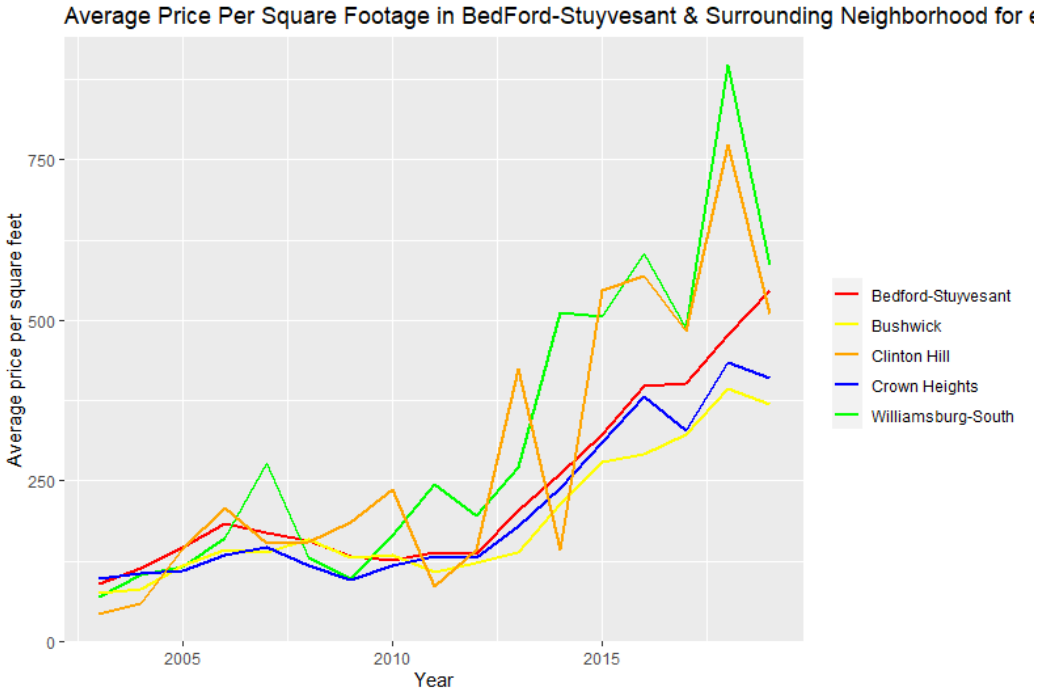

*Figure 20*

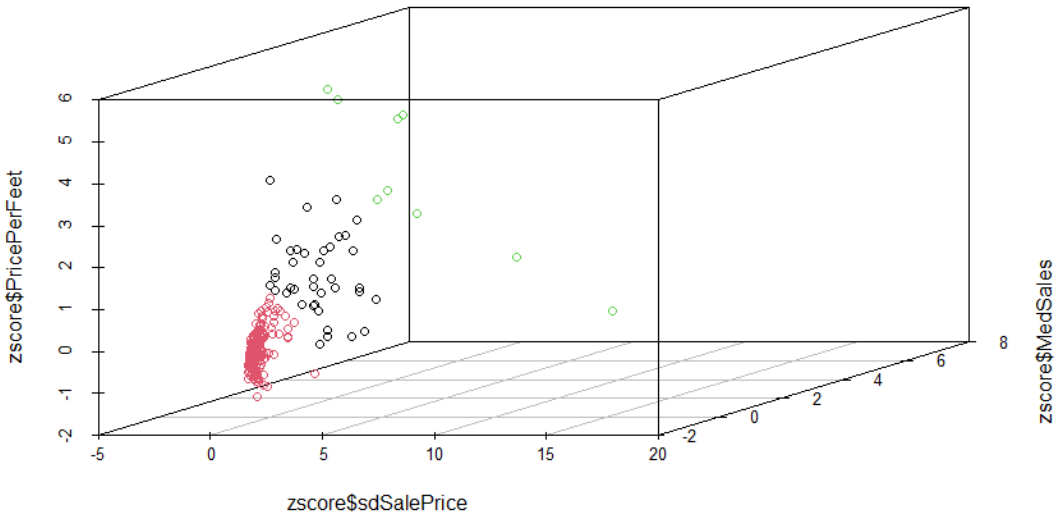


*Figure 21*

```
        Welch Two Sample t-test

data:  y1 and y2
t = 0.95566, df = 9550.2, p-value = 0.3393
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -7.496822 21.760699
sample estimates:
mean of x mean of y
 327.5340  320.4021
```