

# Analysis of Real Estate trends using rolling sales data of different regions in New York

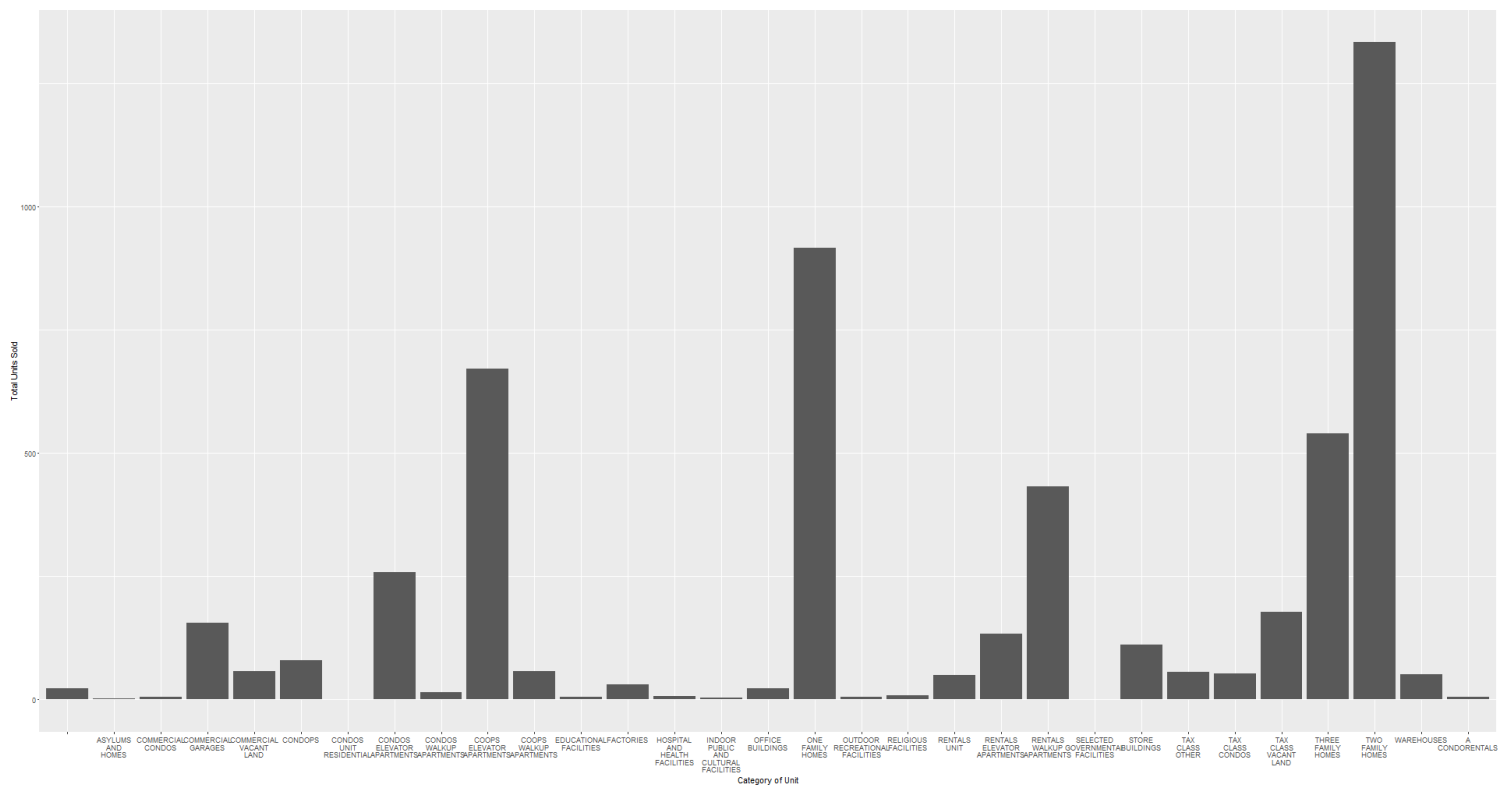
## Data Set

The data used for analysis was rolling sales data from different New York boroughs like Manhattan, Queens, Brooklyn, Staten Island etc. The data captured various variables like borough, building class, year sold, year built, neighborhood, total units, land area, gross area etc.

The data was first cleaned to get the numeric values within the set and then observations made to find relationships within the various variables present. First a single borough was analyzed (Brooklyn) and then the analysis was expanded to all the boroughs to find both local and universal trends. The analysis was made with the objective of helping in understanding both local and regional trends in real estate for the New York region.

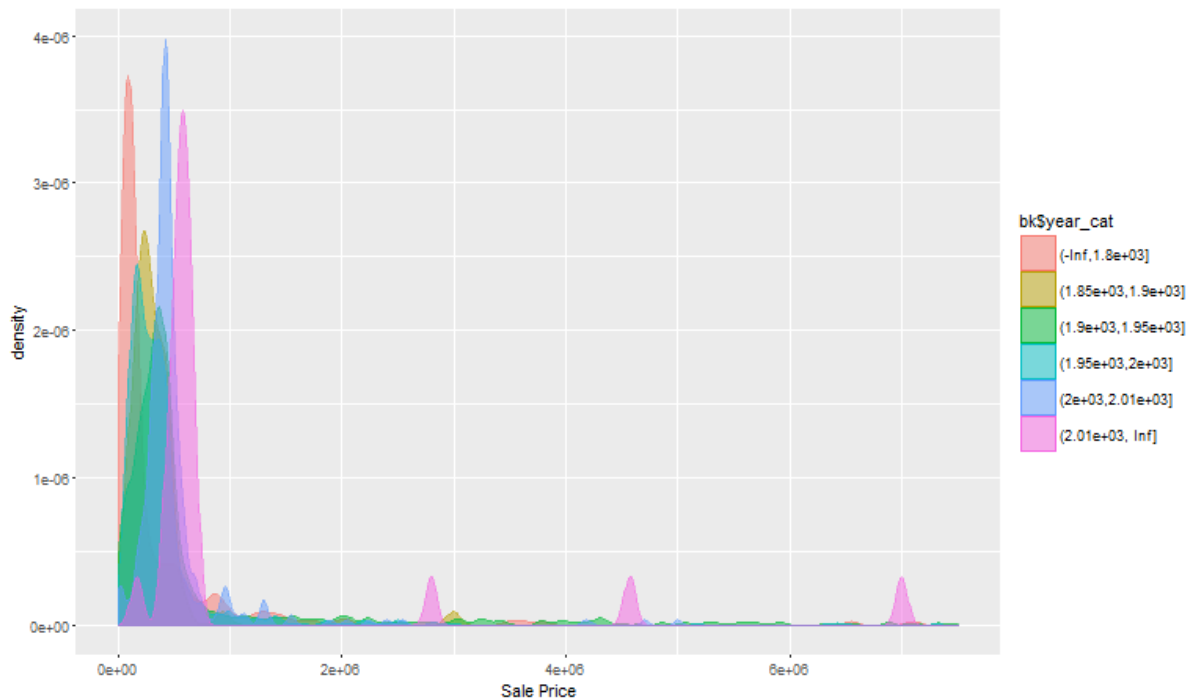
## [1] Analysis on Brooklyn Borough

First, the number of units sold within each category was observed.



It can be seen that the most number of units were sold for the property which were either one, two or three family homes, or elevator apartments. This gives a good insight into the most popular type of real estate in Brooklyn.

It was observed that the year in which the property was built showed a wide variation in data, which ranged from houses built in 1800 to more recent ones in 2013. The data was thus analyzed to understand how the year in which the house was built affected the Sale price or the market value of the property. The Data was divided into separate age groups of 50 years each and their price distribution density within each subgroup observed.

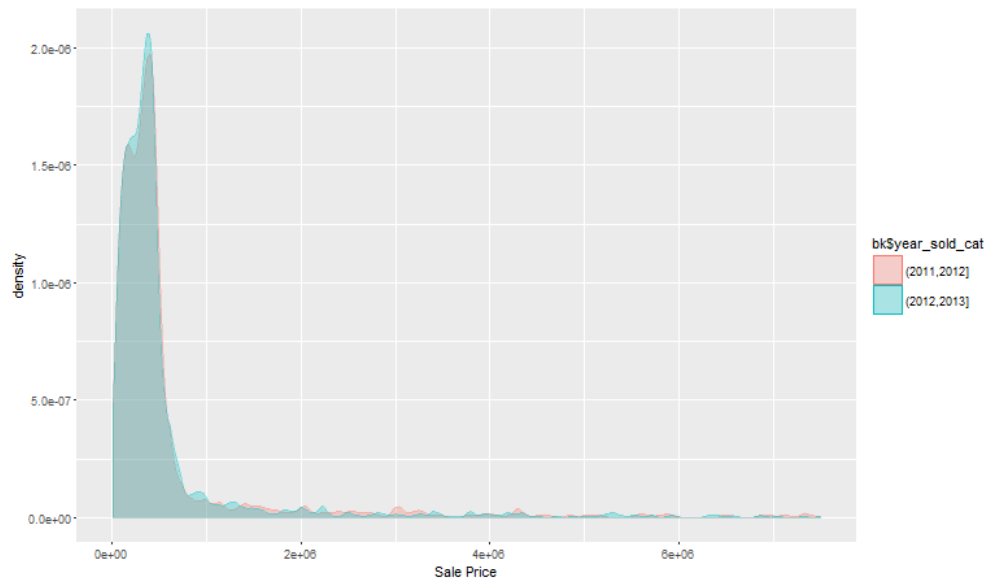


It can be seen that older houses are more peaked towards the start of the x axis indicating an overall lower price distribution, where as those made after 2013 have several peaks all along the x-axis indicating a higher value. Interestingly, houses made between 1850 and 1900 show a more flattened curve and have a small peak even at the higher price end indicating that houses from this period in Brooklyn may go for high prices. In context of real direct, this fact can help in determining a correct evaluation of a property based on the period in which it was built.

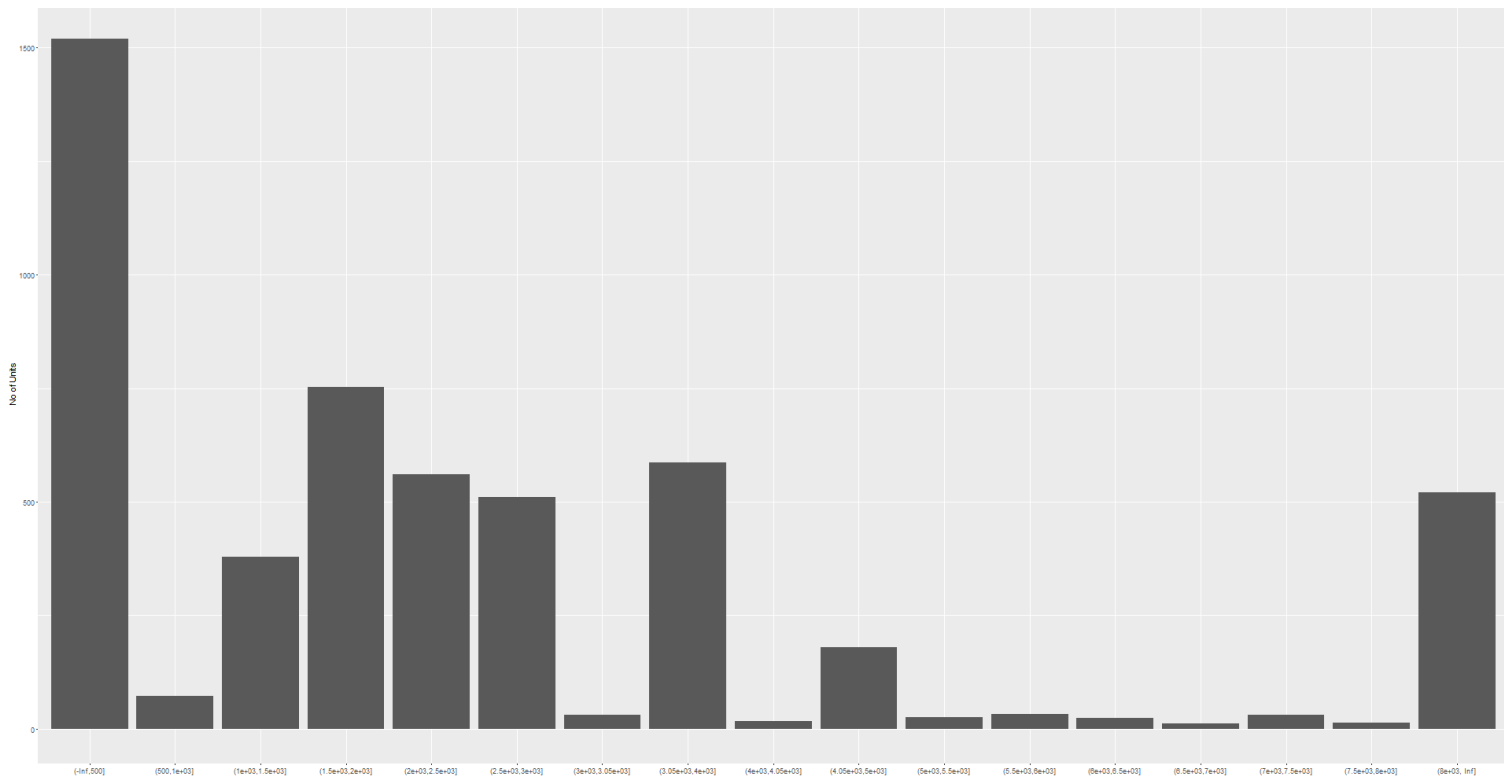
In order to determine how the yearly sales have been affected, an analysis of the sale price distribution for yearly sales revealed that although the number of units sold in 2013 was higher than the number of units sold in 2012, the total sale price was greater than for 2012, indicating that the average value of the property sold in 2012 was more than that for 2013.

Consequently, a distribution analysis for price revealed that there were several small peaks for sales in 2012 which would have led to an increase in total sale for the year by having some units sold at higher prices.

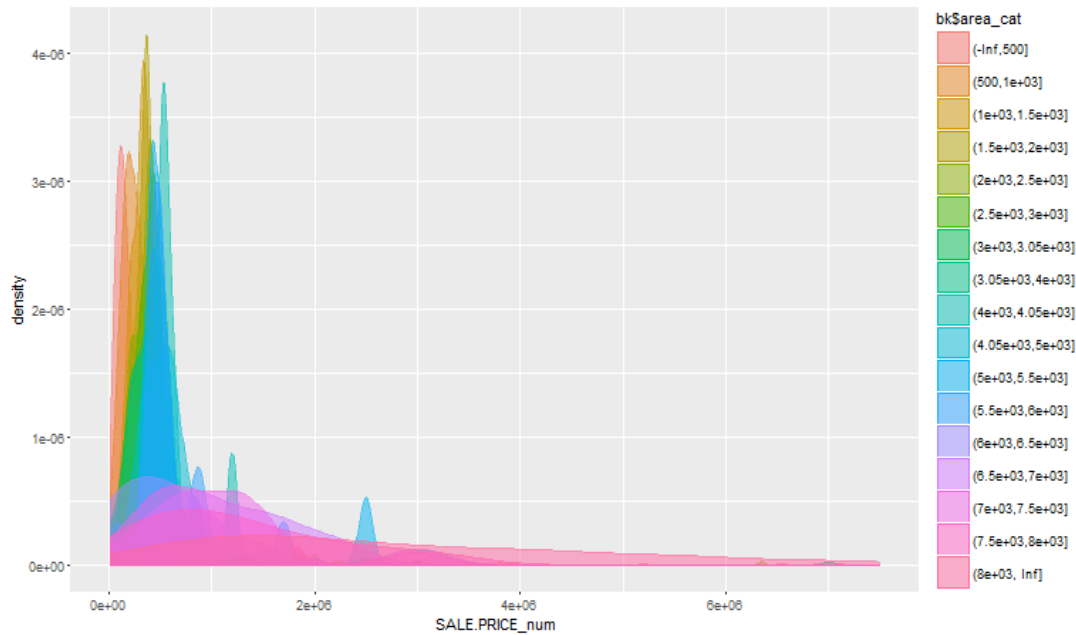
Year Sold	Units Sold	Max Unit Sale Price
2012	2227	650988.2
2013	3041	462727



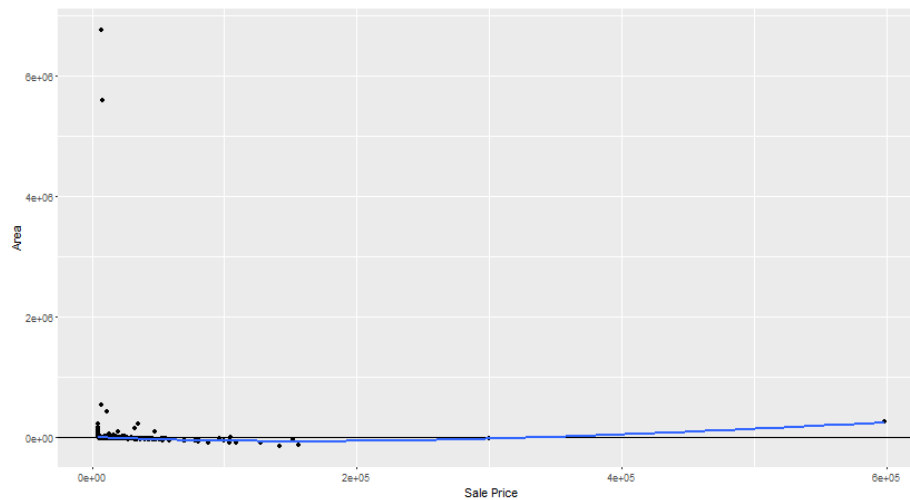
In order to determine how the gross total area affects the price of a unit and how the data was represented by units of different areas, the data was categorized based on the gross area of size 500 sq ft. The total number of units within each group were as shown.



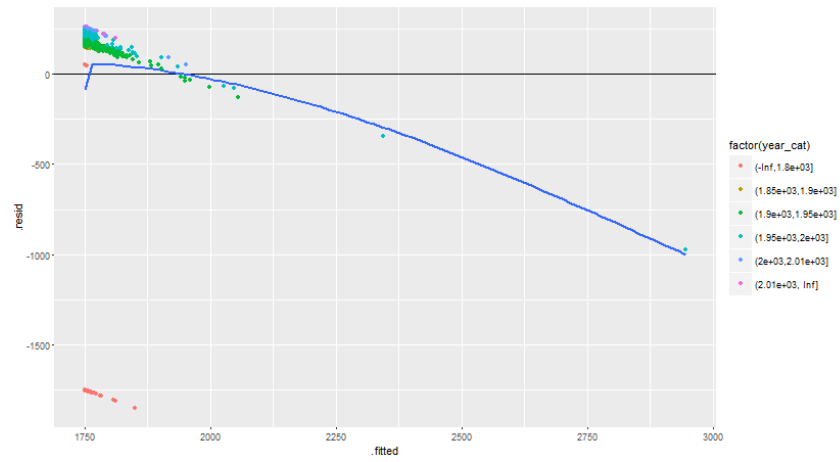
As can be seen, most of the units were less than 500 sq. However, there was a significant number of houses which had total area more than 8000 sq feet. The individual sale price distribution within each of the area sub category were further. This can help realdirect.com in targeting the area categories that have the most unit sales within them.



It can be seen that the unites with areas between 2-2.5 thousand square feet have a higher peak than the units belonging to other area groups. It can be inferred that houses in this range of gross area show little variation in their prices. Making them reliable for price projections and estimates. In order to further investigate this, a linear model between price and the total area was made. It can be seen that the model is slightly linear. It should be noted that this is a plot of the model itself and not the actual values.



Similarly linear model between Year built and price is found to be non linear which corresponds to the earlier observation that the price may be higher for certain houses that belong to a particular year category



## [2] Analysis on all the boroughs combined

In order to expand the dataset, similar analysis like [1] was done on all the borough datasets for New York region to find relationship between variables and find trends consistent among all the boroughs.

The total sales among all the boroughs were as shown. The labels used are:

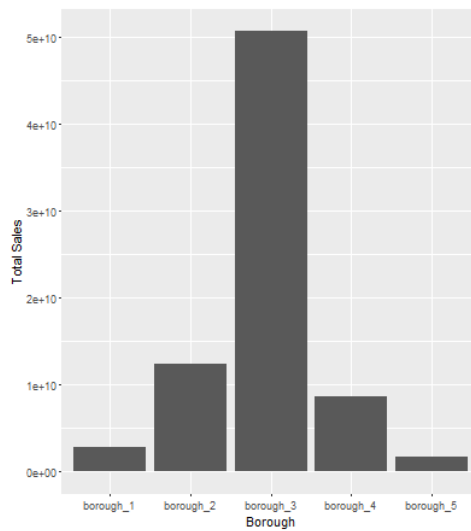
*Borough\_1: Bronx*

*Borough\_2: Brooklyn*

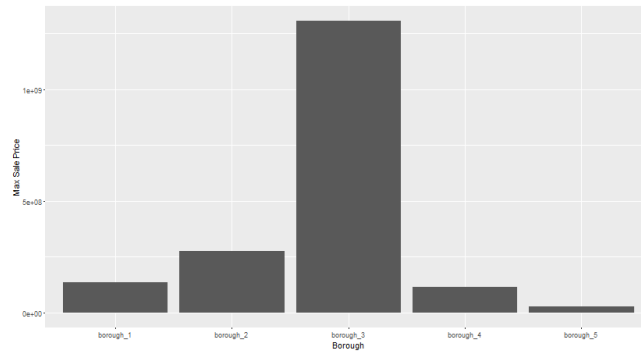
*Borough\_3: Manhattan*

*Borough\_4: Queens*

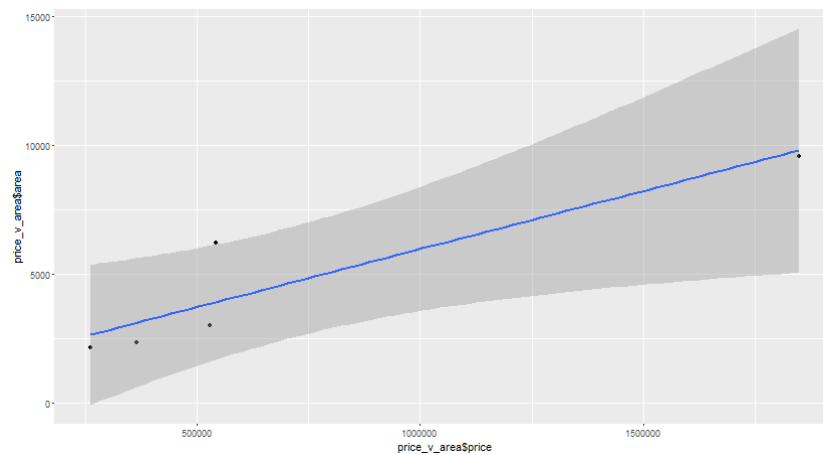
*Borough\_5: Staten Island*



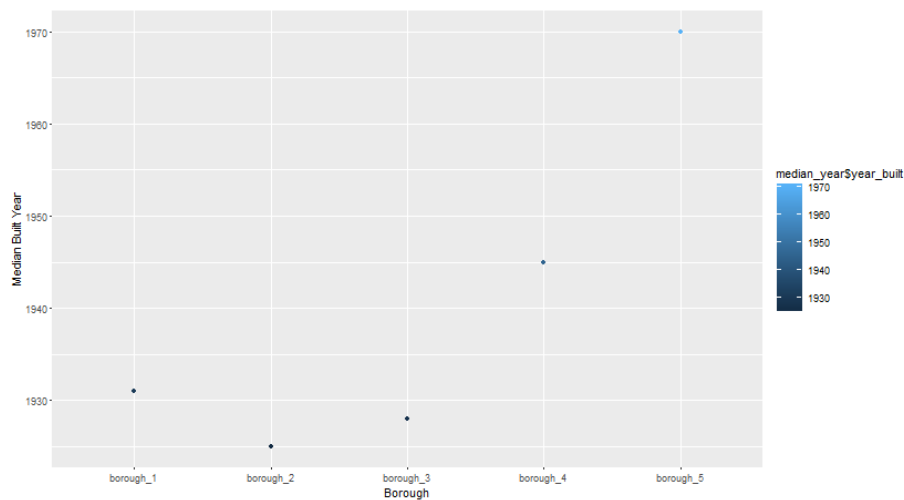
It can be seen that among all the borough, Manhattan has the highest total sale price, and it is considerably higher than all the other boroughs. In order to further investigate this, max sale price within each borough were analyzed. Again sale prices for Manhattan were more than that for any other borough.



As in [1], relationship between sale price and gross area for each borough was analyzed. It was found that even across boroughs, this relationship can be described using a linear model.



The median year of built within each borough was observed, and it was seen that for manhattan the median year was 1928, while for Staten Island it was 1970. Thus the year in which a unit was built has less effect on the price of the unit as compared to the borough in which the unit is. This can be corroborated by seeing the total number of units sold within each borough.



<b>Borough Name</b>	<b>No of Units</b>
Bronx	3589
Brooklyn	14582
Manhattan	19802
Queens	15587
Staten Island	3777

Although Brooklyn, Manhattan and Queens have had same no of units sold, the total sale price for manhattan is considerably larger than the other two.