New York Housing Report

by J. Deaton

Abstract I have been tasked by management board to find a new location for certain clients wishing to buy a house in New York close to the city. By presenting visualizations of different correlations and statistics we will present enough options for the client to proceed in their home buying process.
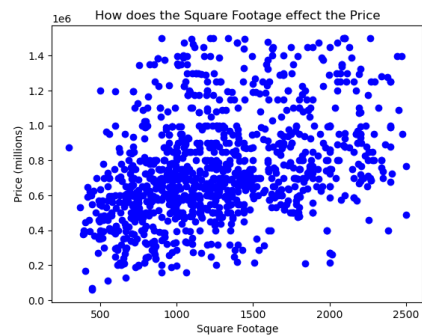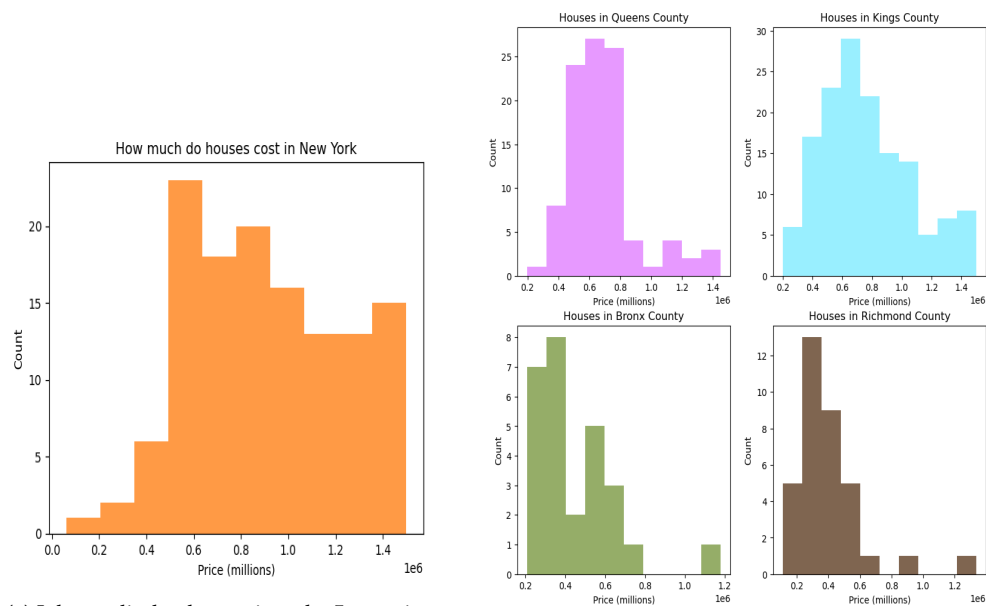
Visualizations

Figure 1: By using our cleaned dataset I wanted to see if the Square footage of a house was the main factor in the increase of price. As seen in the visualization there is little to no correlation between these variable amongst the entire dataset.


(a) I then split the dataset into the 5 counties in New York city including the city itself. This bar graph shows the spread of prices in houses in New York County.

(b) This bar graph shows the other four boroughs in New York city and their spread of price in houses.
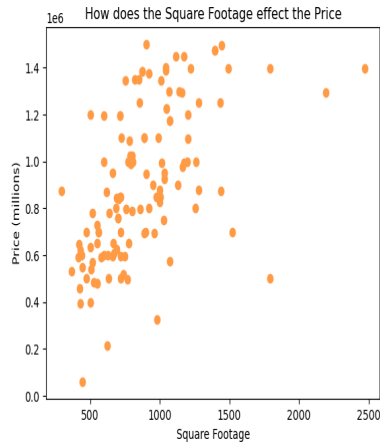
(a) Then from the split data I then tried to see if the size of the house had an effect on the price. As seen in the graph there is a better correlation then the original scatter plot. Although not perfectly positive there is a positive correlation.
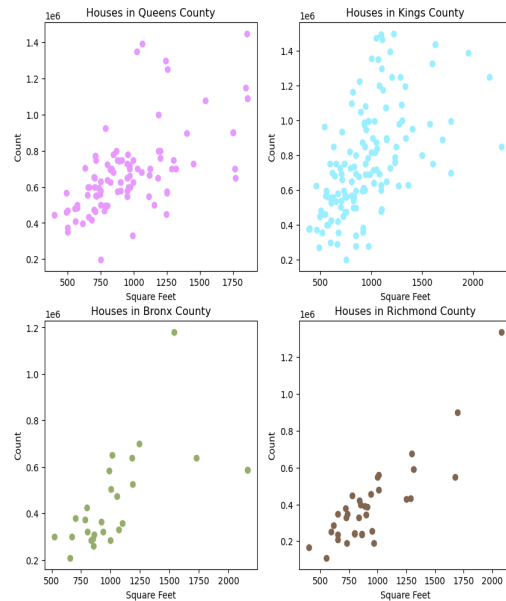
(b) Again I then looked at the four boroughs to see whether they also had a correlation of size of house and price. These houses also seemed to have a positive correlation even though it is not incredibly strong.

Summary

We have determined from the images above it would be wise to first decide on which area you would prefer to live in. This is because the price of the house you may buy is largely dependent on where you live without much correlation to how large the house is. Later after deciding where you would like to live then it would be easier to find a price value on the size of the house. Within each area there is a stronger correlation between size of the house and the price of the house.

From our visualizations we would prefer to live in Queens because of the larger range of prices in the house. Later in the visualizations we can see there is a positive correlation between size and price of the house which means Queens most likely has a larger range in size of the house as well. Now if you are looking for houses that are not as expensive then you would probably prefer the Bronx or Staten Island. If you would prefer more expensive houses then there is always the option to live in the city, but you could also choose Kings, or Queens counties.

In future data we would ask that it be required for the data to include which county you live in so there would be more data on each county. In the cleaning process we decided to cut all the data that did not include which county the house was in. We also zoomed in to look at where the majority of the data was. This took out some outliers which may have proven helpful. This left us with several data values but a much deeper study could be done with more data values in these specific counties.