Vuesol

ANALYSIS OF THE NORTHEAST OHIO HOUSING MARKET VISUALIZATION & REGRESSION ANALYSIS

by Rikka Vivekanand Reddy
Operation Research Analyst

Executive Summary

Northeast Ohio offers a wide range of homes of different prices, sizes, and lot sizes, ranging from urban, modern homes, to large country estates, to grand historic homes in older neighborhoods.

Average sizes: The average listing price is $176,040, but this mean is driven up by some very expensive homes in the more exclusive neighborhoods of Gates Mills, Pepper Pike, Solon and Aurora. There are many other options that are well below this average price. The average lot size is 19,404 square feet, which again is driven up by more exclusive neighborhoods like Gates Mills, where the average lot size is 168,263 square feet, and Pepper Pike, where the average lot size is 55,723 square feet. The average square footage of homes follows a similar pattern. The average size is 1,860 square feet, with the larger homes located in the more exclusive neighborhoods, but there are many sizes available. The number of bathrooms will also follow this trend: larger homes have more bathrooms. In spite of the vast range of home sizes, lot sizes and listing prices, the numbers of bedrooms are surprisingly consistent. The majority of homes have three bedrooms, though there many homes with four or more bedrooms.

Cost drivers: List price will be driven significantly by where the house is located. Exclusive neighborhoods like Gates Mills and Pepper Pike have greater property values and taxes, so the cost of the home will reflect this. There is a strong relationship between the list price of the house with the number of baths and the square footage of the house. This means that the more baths a house has, the higher the average list price. The square footage of the house will also impact the average list price of the house. The number of beds, the lot size and the number of parking spots will also increase the price, but not as much as these first two factors. Interestingly, the age of the home has two effects on price: the older the house is, the lower the listing price becomes until after age 100. At that point, the prices begin to rise for older, more historic homes. The majority of the homes available are aged 50-75 years, but like the other factors, there are a wide range of ages.

In fact, we determined that the general trend is for every additional bathroom, the price will increase by $34,611. For every additional parking spot, the price will increase by $12,014. The price per square foot of the house is $94.33. Finally, each year that the house is older, the price will decrease by $739.

Conclusions:

John Doe is seeking a home that will be large enough for himself and his wife, a new baby and visiting parents, which will require at least three bedrooms. He would also like a large yard for his child and dogs to play in. The data has demonstrated that there are numerous homes in the area that fulfill these criteria. John has also assumed that size and location influence price, which the data demonstrates is correct. Other factors that he will need to consider will be the number of baths, parking spots and age, as these factors will also have an effect on price.

Given his salary of $120,000, we are confident that John and his family will be able to find a home that meets their home and budgetary requirements.

Author:
Rikka Vivekanand Reddy
Technology Practice Lead
Vuesol Technologies Inc

Index

Data Cleansing

Missing values:

When cleansing the data, we immediately noticed that there were several pieces of data missing:

- 15 missing in zip codes: Because this is closely correlated with city, we determined this would not affect results
- 3 missing number of bathrooms: This is a very small number. We determined this would not af-fect results.
- 337 missing square footage data: We felt this would not significantly skew results, as there were other pieces of data such as lot size, bedrooms and baths that would provide a complete picture of the housing market.
- 242 missing lot size data: Again, we felt the missing values would be complemented by other variables such as square footage, bedrooms and baths.
- 75 missing the year the house was built: Because this was not a large percentage, we did not feel this would skew data.

Duplicates:

We discovered and removed one duplicate entry.

Errors:

Year built was miscategorized as continuous data. We changed this to nominal.

Summary statistics & Potential errors

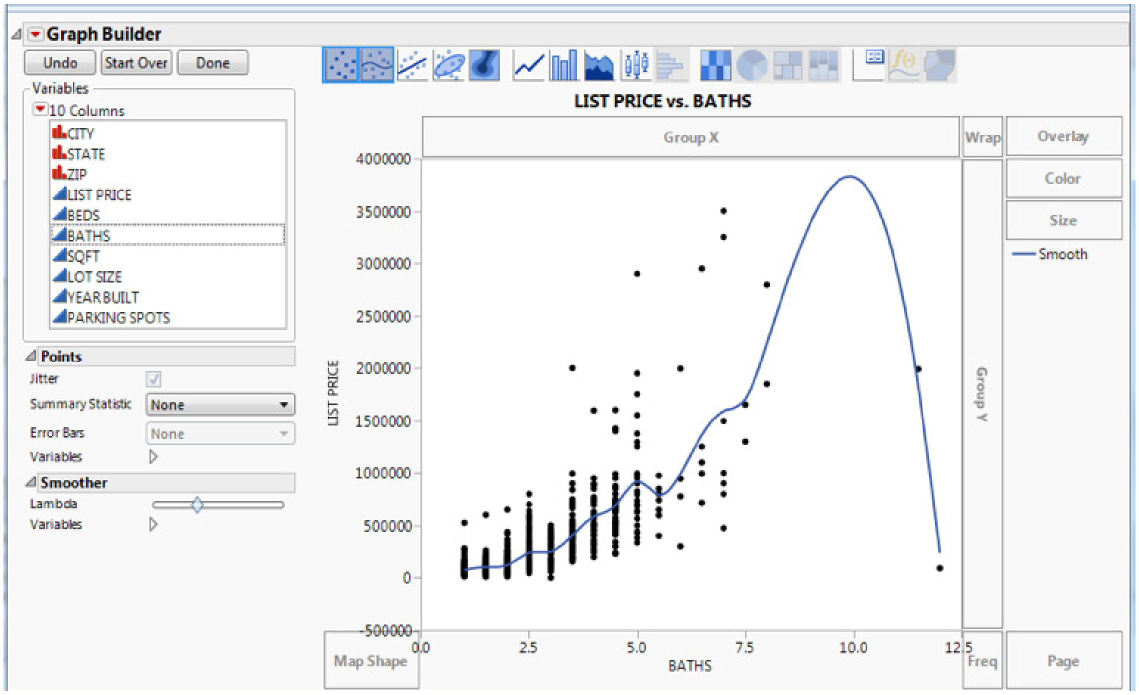
We did an immediate scan for potential errors in data entry by looking at the Columns View summary data (see Figure 1 in the Appendix). We noticed several potential errors:

- Maximum value for number of bathrooms: 12 bathrooms. This could be possible with a large house, but it appeared strange.
- Maximum value for parking spots: 30 parking spots. This seemed unlikely.
- Minimum for list price: 100. This value seemed too low.
- Maximum value for square footage: 15,233. This value may be too high

Outliers:

After making note of these potential errors found in the Summary Statistics, we also looked at the box plots to identify outliers (See Figure 2 in Appendix). For listing price, we immediately saw that there was an outlier for $2,950,000 (Row 77) in Gates Mills, which is a very nice area; therefore, we included this data. We also noticed three outliers in lot size. Again, after looking more closely at the data, we determined that these values potentially made sense; the lots were located in Solon and Gates Mills, which are known for spacious properties.

We also noticed that Row 1312 had 12 bathrooms. We decided to exclude this data because the list price was only $90,000, so this value did not make sense. This outlier also skewed our visualization to help us understand how the number of baths affected price. We saw a general positive correlation with price and number of baths, but the outlier skewed the model:



Summary Statistic: None

Error Bars: None

Outlier skewing data and price v. bath model

Given the large sample size of almost 2,000 values and the uncertainty of the validity of these outlier values, we decided to only include data that fell within four standard deviations of the mean for the categories with the greatest outliers, to avoid these values skewing the data:

- **Number of baths:** Excluded 13 rows of data due to the number of baths lying outside of four standard deviations outside the mean. The mean was 3.349, with standard deviation of .82662.
- **Number of bedrooms**: Excluded 12 rows of data that fell outside four standard deviations, or seven beds and greater. The mean was 2.01248, and the standard deviation was 1.09966.
- **Number of parking spots:** Excluded two of data. This was after finding row 86, which had a value of 30 parking spots, as an outlier. The mean was 1.813084, the standard deviation was 1.0658. We excluded houses with more than 7 parking spots, which were 4 standard deviations beyond the mean.
- **Lot size:** Excluded 6 additional rows for lot size, also within 4 standard deviations. We saw a range in values that seemed too large to make sense, such as 4,467,949 square feet, and others, such as a minimum of 109 square feet, which seemed too low. The mean was 29,969 sq. ft. and the standard deviation was 165,081 sq. ft. Within 4 standard deviations we excluded houses with a lot size of more than 690,295 sq. ft.

We determined that all other data was acceptable. By doing this we were able to eliminate our outliers while maintaining the integrity of our data with 99.99% of the points remaining within 4 standard deviations of the true mean.

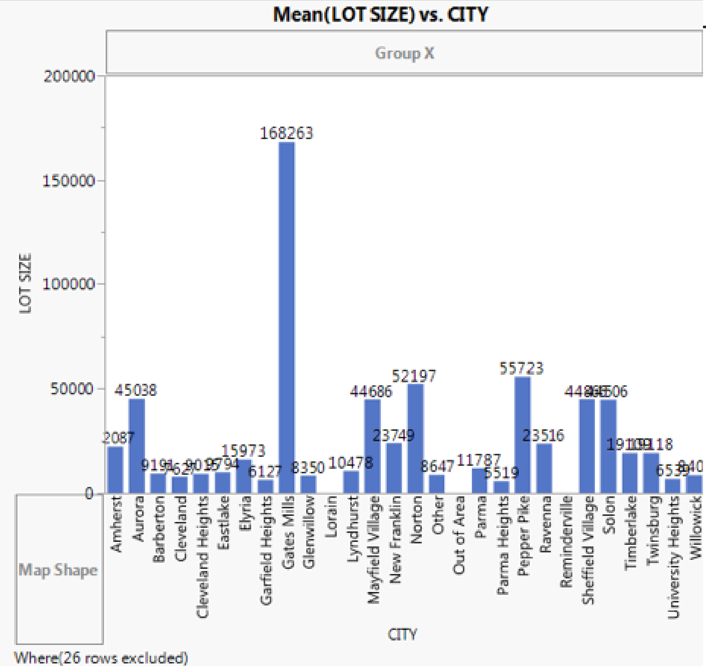
We created an age column, which was built from the year built, using the formula “2016-Year Built.”

We also created our validation column using a stratified random method because we excluded outlier data and wanted to make sure that the groups were not skewed by the excluded data. The purely random approach included the data that we excluded, which could cause some of the groups to become skewed.

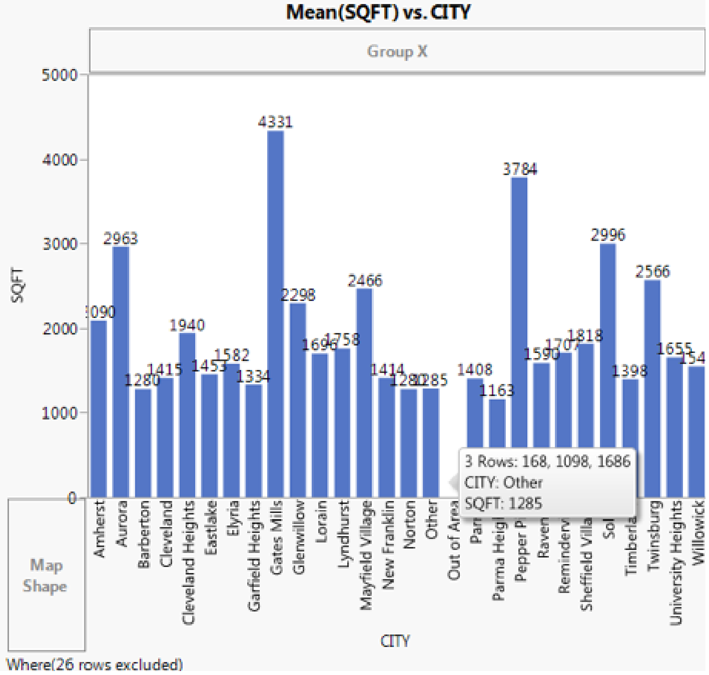
Data Visualization & Descriptive Statistics:

Descriptive Statistics:

After eliminating our outliers, we had a cleaner set of Summary Statistics and box plots. The mean list price for Northeast Ohio is $176,040. This data was skewed to the right, meaning that the mean is greater than the median, and there are many larger values on the right side of the scale. Lot size (mean of 19,404 sq. ft.), and square footage (mean of 1,860 sq. ft.) were also quite skewed to the right, suggesting that there is a wide range of sizes available, from smaller, urban lots to vast properties further outside of the city. As the two bar charts demonstrate, these averages are driven up by more exclusive neighborhoods such as Gates Mills and Pepper Pike, both of which have significantly higher mean lot and home square footage values :

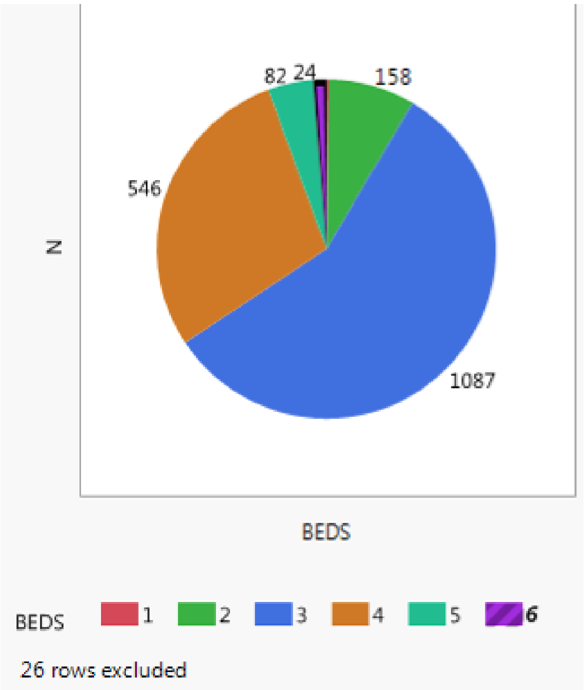


Average lot size by city



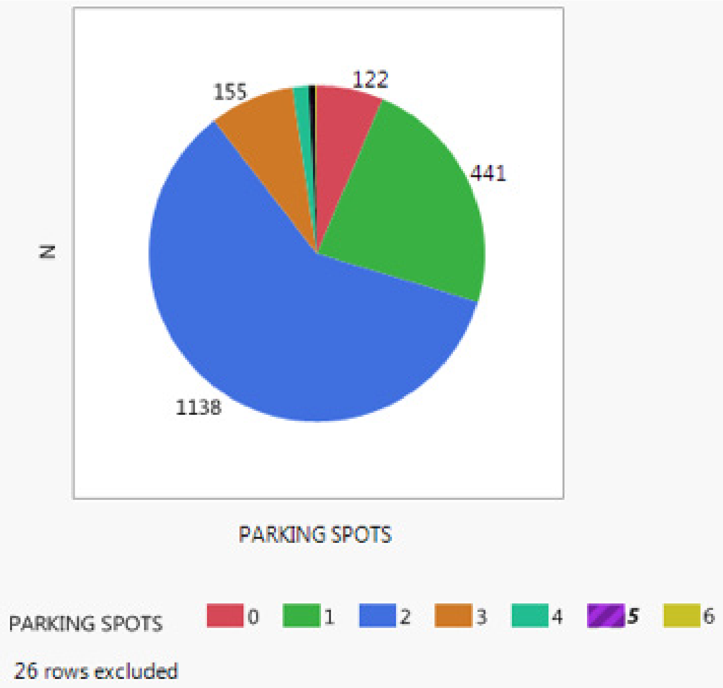
Average square footage by city

The data for number of bathrooms (mean of 1.96, which is just below the median of 2 baths) was also slightly skewed to the right, coinciding with the wide range of home sizes. Surprisingly, the number of bedrooms was normally distributed. The average home has 3 bedrooms, regardless of the size of the house or lot, though there are a number of homes with four or more bedrooms, as shown in this pie chart:



Number of bedrooms

Parking spots were skewed to the left. Almost sixty percent of the houses in the data set had two parking spots. Nearly 90% of the homes have two spots or fewer. However, almost two hundred homes have three parking spots or greater.



Number of parking spots

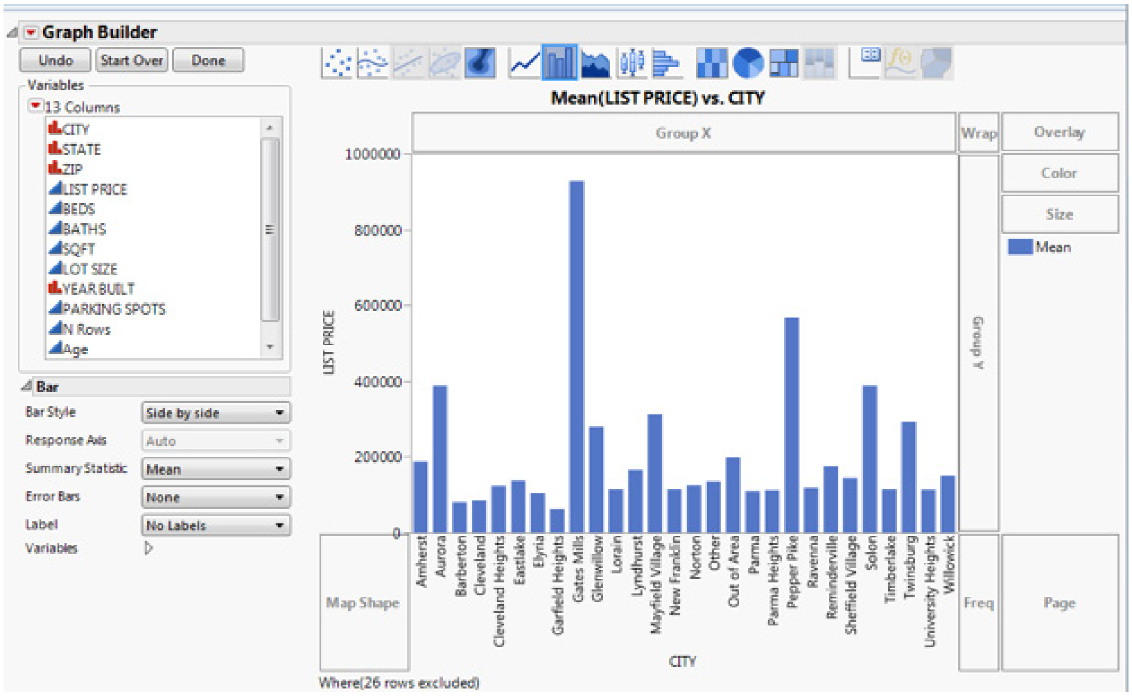
The age of the home was fairly normally distributed. The average age was 61.2 years, and the median age was 62 years. There are more homes to the left of the mean and median than the right, meaning that John will have more new homes to choose from than very old (more than 100 years, for example).

See Figures 3 and 4 in the Appendix for the jmp outputs on data distribution for all variables.

Visualizations:

The variable that we were most interested in understanding was price, and how the other variable values influence listing price. The visualizations we selected will help "John" understand the type of house he will be able to purchase while staying within his budget.

<u>How city affects prices::</u>



Graph Builder

Variables

13 Columns

CITY
STATE
ZIP
LIST PRICE
BEDS
BATHS
SQFT
LOT SIZE
YEAR BUILT
PARKING SPOTS
N Rows
Age

Bar

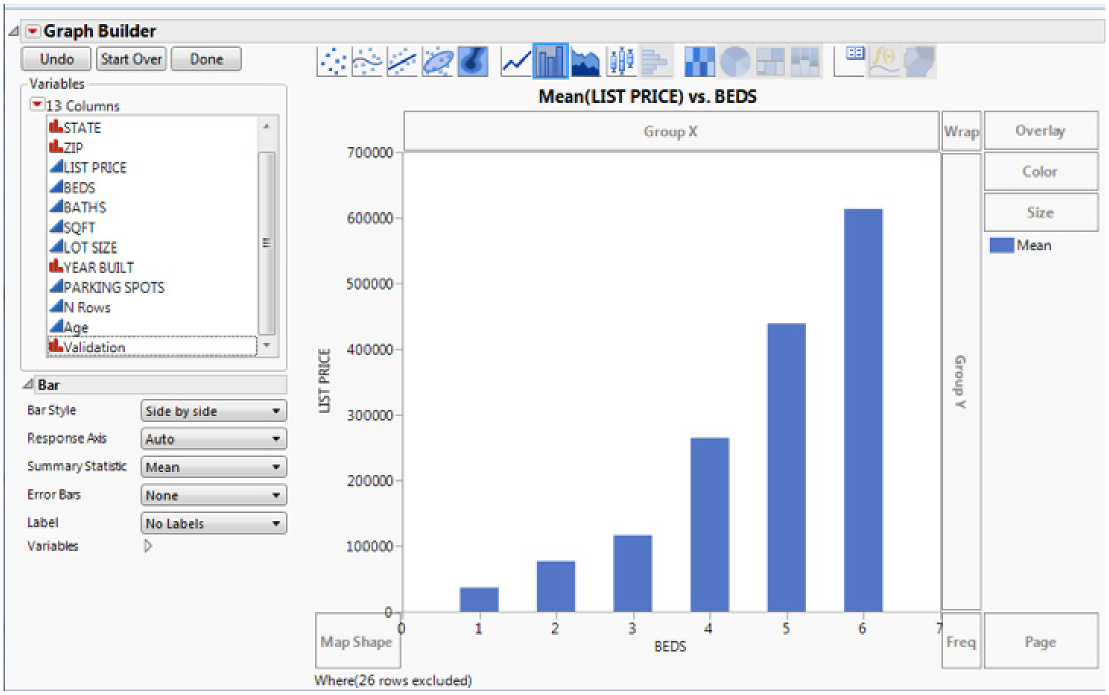
Bar Style: Side by side
Response Axis: Auto
Summary Statistic: Mean
Error Bars: None
Label: No Labels
Variables

Where(26 rows excluded)

This chart shows Price by City, which shows that some cities are significantly more expensive than other cities. From having knowledge of the area, we know that this is a realistic depiction of the areas. For example, Gate Mills and Pepper Pike are historically wealthier communities. This graph would allow us to show "John" the most expensive areas and the areas that are within his budget. We can then focus on those communities when searching for houses.

<u>How number of bedrooms affect price:</u>

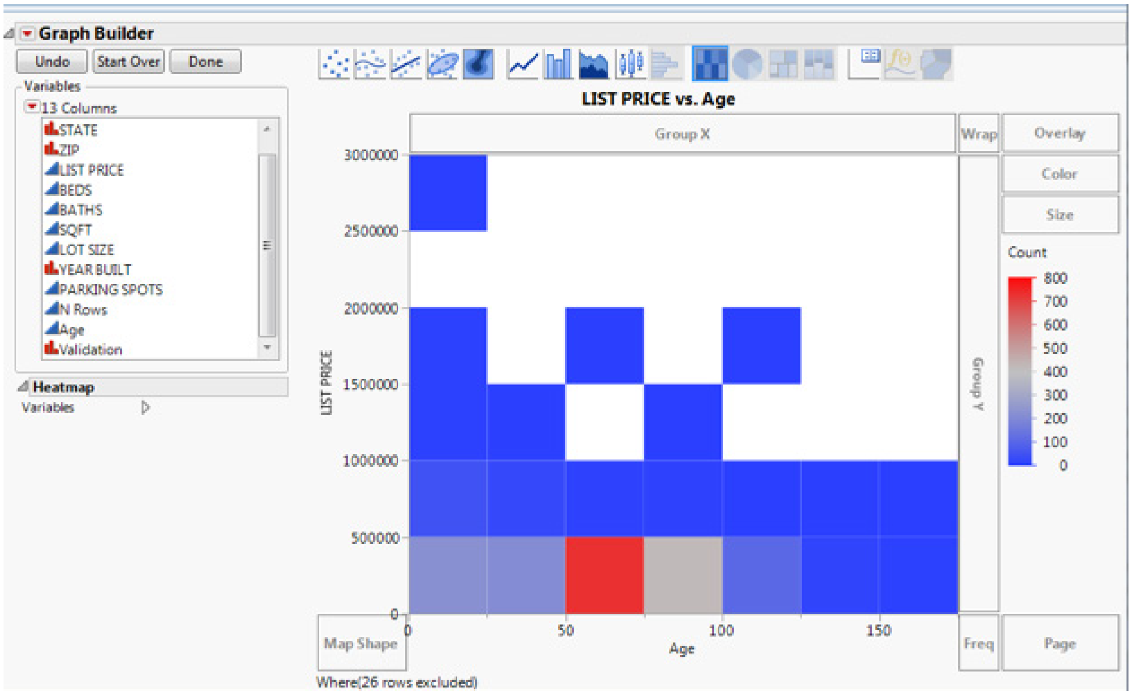
In this graph, we can see that on average the more expensive houses have more bedrooms.



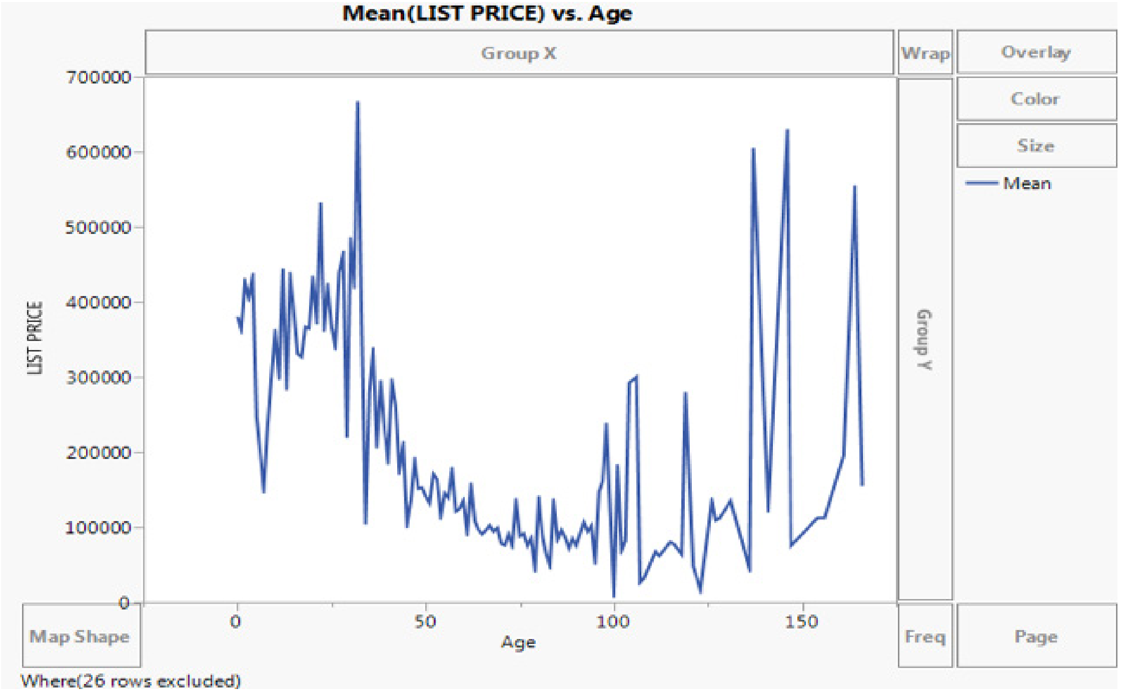
Bar Style: Side by side
Response Axis: Auto
Summary Statistic: Mean
Error Bars: None
Label: No Labels

How age of the house affects price:

This first chart shows where the greatest concentration of houses lies: age 50-75, at price range from $0 - $500,000

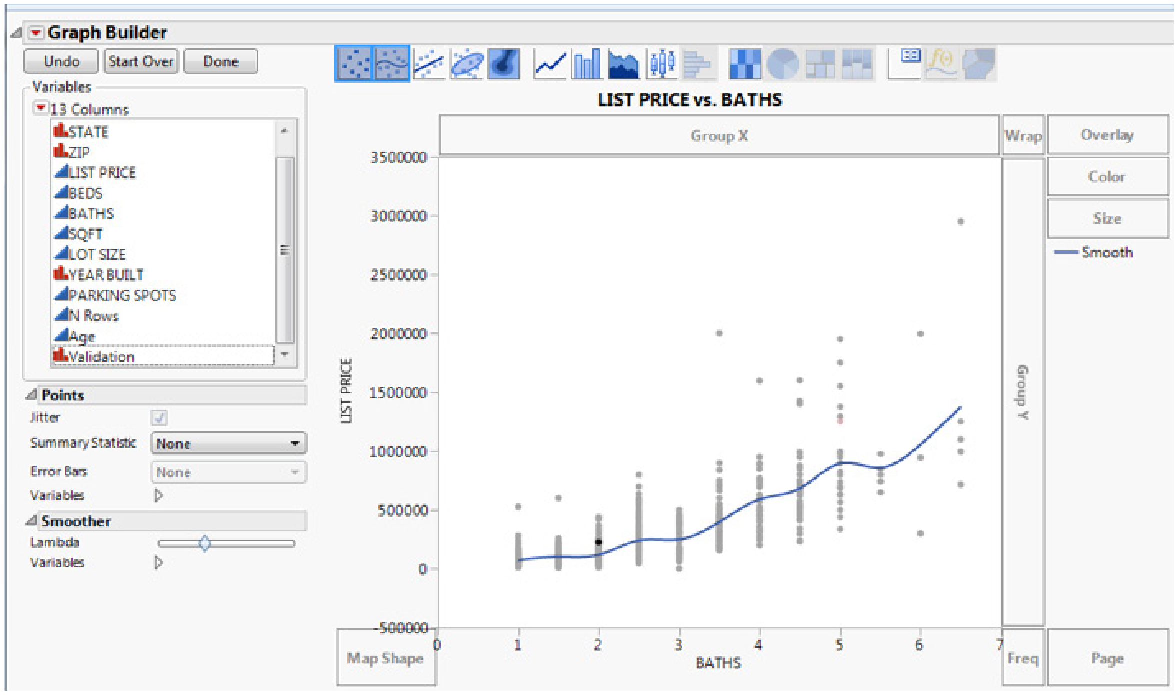


This second visualization of price by age of house shows that there is a negative correlation between list price and age (i.e. the older the house is, the lower the list price). However, homes that are 100 years old or greater begin to see increases in list price.



List price vs baths:

The data suggests that the list price will rise as the number of bathrooms increase.



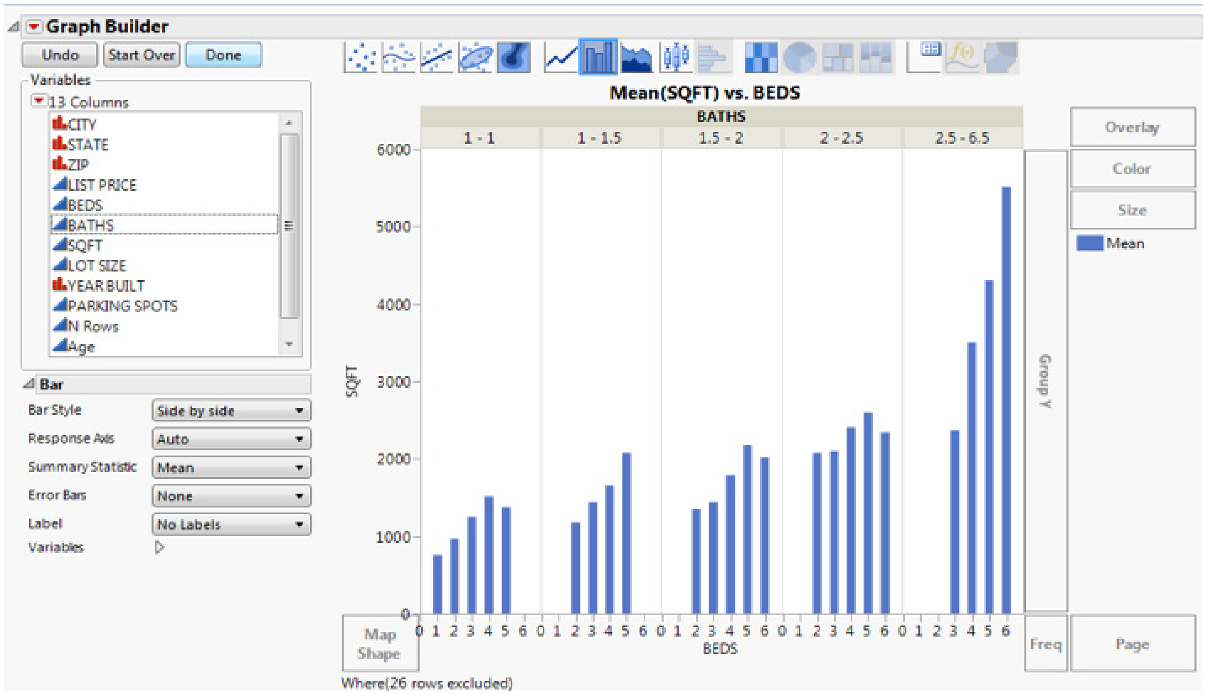
13 Columns
- STATE
- ZIP
- LIST PRICE
- BEDS
- BATHS
- SQFT
- LOT SIZE
- YEAR BUILT
- PARKING SPOTS
- N Rows
- Age
- Validation

Points
- Jitter ☑
- Summary Statistic: None
- Error Bars: None
- Variables

Smoother
- Lambda
- Variables

Square footage and number of bedrooms:

We can see from this graph, that if John prefers to have a bathroom for himself, his parents and a guest bathroom then he would need to purchase a house with three or more bedrooms. This graph shows that there are no houses available with more than three baths that has less than three beds.

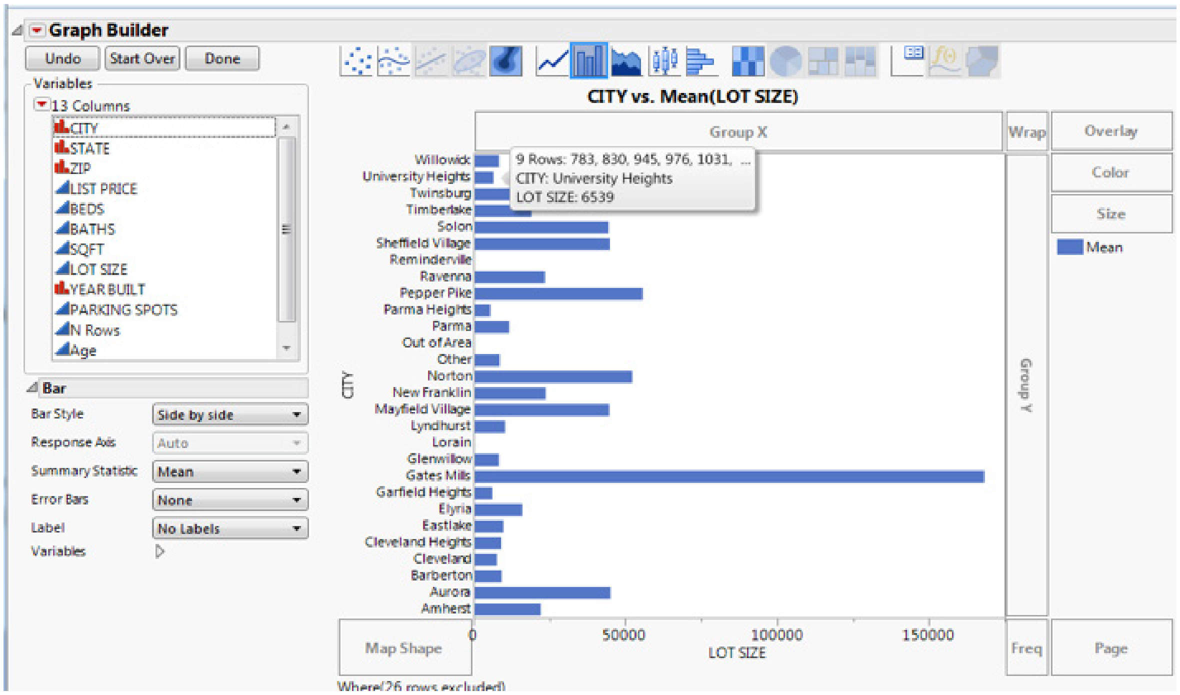


Bar Style: Side by side
Response Axis: Auto
Summary Statistic: Mean
Error Bars: None
Label: No Labels
Variables

Where(26 rows excluded)

<u>Average lot size by city:</u>

John would like to purchase a house that has a yard for his dogs and future children. This graph tells him where he is more likely to find larger lots, and what the mean lot size is in each neighborhood.



9 Rows: 783, 830, 945, 976, 1031, ...
CITY: University Heights
LOT SIZE: 6539

Bar Style: Side by side
Response Axis: Auto
Summary Statistic: Mean
Error Bars: None
Label: No Labels
Variables

Where(26 rows excluded)

All of these charts help John to get a better understanding of the communities in which he will most likely find the type of home he wants, what the average price in those areas are, and the kinds of things that affect listing price, such as the age of the home, number of bedrooms and number of bathrooms.

Pairwise Correlation Analysis:

Our analyses show that the list price of the house is highly positively correlated with the number of baths and the square footage of the house. With a correlation of 0.8846 and 0.7617, respectively, the square footage and number of baths will increase the list price as the variables increase. This means that the more baths a house has, the higher the average list price. The square footage of the house will also impact the average list price of the house. The number of beds, the lot size and the number of parking spots is also correlated with list price, though the correlation is not strong. Age is the only variable that is negatively correlated with list price at a correlation of -0.3975. Again, this variable does not have a strong correlation which means that there may not be a large effect on price when the age differs among houses. This negative relationship indicates that as the age of a house increases, the list price of that house will decrease. However, as our visualization demonstrated, houses age 100 years and greater did increase in price.

There are several variables that are correlated with each other, which can cause a problem of multicollinearity. The dependent variables that are correlated are baths and square footage and beds and square footage. This makes sense because larger houses have greater square footage and more room for additional beds and baths. We may take this into consideration when creating our regression analysis because we would want to make sure that our dependent variables are not correlated. We may find that excluding square footage from the analysis would give us a more explanatory model.

See Figures 5 & 6 of the Appendix for our Pairwise Correlation matrix and outputs.

Simple Regression Analysis

Our regression analyses looked at the influence that one variable had on another.

List price and age:

We looked at the relationship between list price and age. Our null hypothesis was that the beta of the age variable will be equal to zero meaning that the age variable has no effect on the list price. Our alternative hypothesis is that the beta of the age variable is not equal to zero meaning that the age variable has an effect on list price. We would reject the null hypothesis because the p-value is significant and less than our alpha of 0.05. The R2 of this regression is 0.160. This small R2 indicates that this model is not a good indicator of the variables that affect list price. We would need to add additional variables to make this model more comprehensive. As our pairwise regression analysis suggests, the beta for the age variable is negative meaning that as there is a one year increase in age, there is a $2,863.59 decrease in price.

List price and square footage:

We looked at the relationship between list price and the square footage of the house. Our null hypothesis was that the beta of the square footage variable will be equal to zero meaning that the square footage of the house has no effect on the list price. Our alternative hypothesis is that the beta of the square footage variable is not equal to zero meaning that the square footage variable has an effect on list price. We would reject the null hypothesis because the p-value is significant and less than our alpha of 0.05. The R2 of this regression is 0.751916. This R2 is higher and indicates that this model can explain 75% of the variables that affect list price. With this higher R2, we can see that this model is starting to explain how list price is determined. We still need to add additional variables to see if R2 increases, showing that the variables in the model explain the list price. As our pairwise regression analysis suggests, the beta for the square footage variable is positive meaning that as there is a one square foot increase in the size of the house, there is a $146.59 increase in price.

List price and number of beds:

We looked at the relationship between list price and number of beds. Our null hypothesis was that the beta of the beds variable will be equal to zero meaning that the bed variable has no effect on the list price. Our alternative hypothesis is that the beta of the beds variable is not equal to zero meaning that the bed variable has an effect on list price. We would reject the null hypothesis because the p-value is significant and less than our alpha of 0.05. The R2 of this regression is 0.21941. This small R2 indicates that this model is not a good indicator on its own of the variables that affect list price. We would need to add additional variables to make this model more comprehensive. As our pairwise regression analysis suggests, the beta for the bed variable is positive meaning that for each additional bedroom in a house, there is a $131,580 increase in price. This is unreasonable and another reason why this model is not reliable. We will now create a multiple regression to examine the effect of multiple variables on the list price.

In summary here are the results of our regression. See Figure 7 of the Appendix for our jmp outputs

Model	Regression equation	p-value	r2
Y by age	LIST PRICE = 346805.23 - 2863.5896*Age	<.0001	.160
Y by beds	LIST PRICE = -261498.1 + 131580.03*BEDS	<.0001	.219
Y by square footage	LIST PRICE = -100955.4 + 146.58881*SQFT	<.0001	.752

Multiple Regression Analysis:

We analyzed the effect that the number of beds, number of baths, the square footage of the house, the lot size, the number of parking spots and the age of the house has on the list price of the house. We excluded city, zip and state from our analysis because they are correlated with each other and we wanted to avoid multicollinearity. Our null hypothesis is that the betas of each of the variables are equal to zero, meaning that the variables in the model have no effect on the list price. Our alternative hypothesis is that at least one of the betas of the variables will not be equal to zero, meaning that at least one of the variables has an effect on the list price.

We ran a stepwise regression analysis. We found that the variables we selected were all significant in this model as their p-values were less than an alpha of 0.05. We had an R2 of 0.8257, meaning that 83% of the list price can be explained by the model. The equation that we developed is as follows:

List Price = -35175.34 - 15953.38*Beds + 38069.56*Baths + 99.471376*SQFT + 0.7845009*Lot Size + 13176.636*Parking Spots - 687.5151*Age

We found it interesting that the beta for beds was negative, as our pairwise regression analysis indicated that the number of beds in a house has a strong positive correlation with list price. This model would suggest that for every additional bedroom, the price of the house would decrease by $15,953. We know that this is not reasonable. Because of this, we analyze the Variance Inflation Factor of the model. We found that the VIF for all of the variables was fewer than 5, meaning that there was no multicollinearity. We still found this curious, so we decided to run the model again. This time, we excluded the beds variable on the basis of our pairwise regression analysis indicating a correlation between beds and square footage.

With this new model, our results were as follows:

- The beta for beds is now positive aligning with our assumption that the more bedrooms a house has, the higher the price
- The p-values of all of the variables are statistically significant and we would reject the null hy-pothesis because p is less than alpha.
- Our R2 value has decreased very slightly to 0.82 indicating that this model still does a good job ex-plaining the list price of a house.

The new equation is:
List Price = -67,817.29 + 34,611.428*Baths + 94.330562*SqFt + 0.8296131*Lot Size + 12,014.661*Parking Spots - 739.0884*Age

This model indicates the effect of these variables on list price. For every additional bathroom, the price will increase by $34,611. The price per square foot of the house is $94.33. For every additional parking spot, the price will increase by $12,014. Finally, each year that the house is older, the price will decrease by $739. This model explains 82% of the variables that make up list price.

The R2 for the three different types of data differs by the test. For the validation tests, a larger sample size will yield a better the R2 and results that explain that larger sample.

<u>Difference between simple regressions and multiple regression analyses:</u>

There is a difference between the 3 simple regressions and the multiple regression we ran. First, the R2 of the simple regressions are lower for the most part in the simple regressions. By adding more variables, we were able to explain more of the model. We also noticed that the betas were different between the simple regressions and the multiple regression. This indicates that the magnitude of the effect of each variable differed depending on how many other variables were taken into account in the model.

Recommendations and Conclusions:

The selling price of the house per square foot is $94.33. The price per additional bathroom is $34,611. The average list price of the house is calculated by inputting the conditions into the regression equation.

The average list price of the house is calculated as follows:

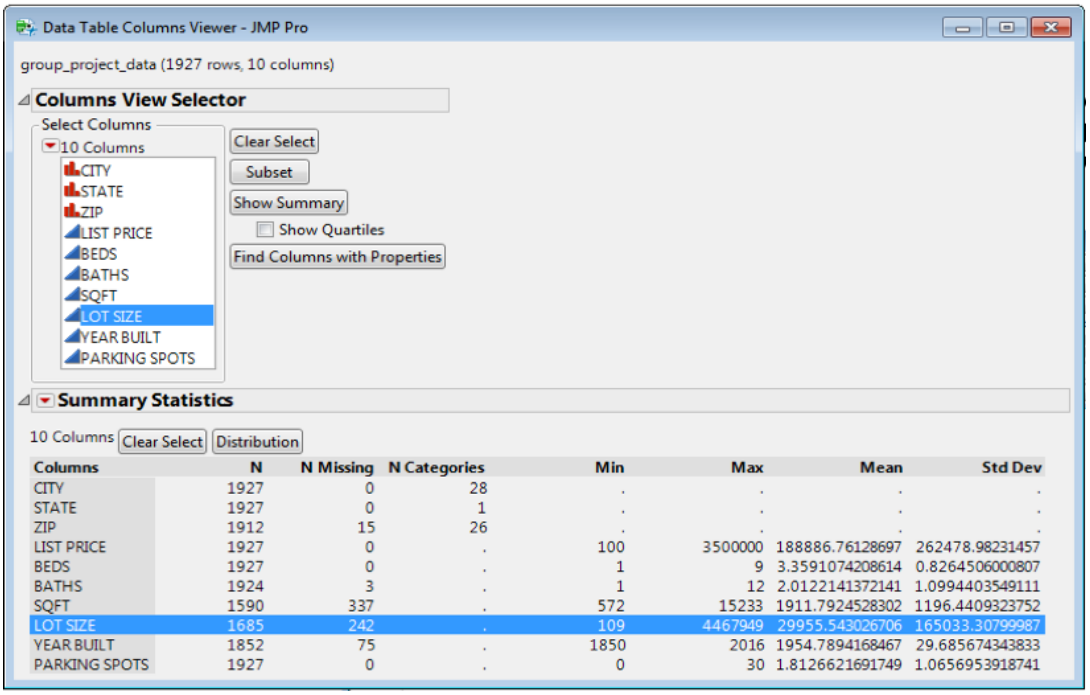
List Price = -67817.29 + 34611.428*Baths + 94.330562*SqFt + 0.8296131*Lot Size + 12014.661*Parking Spots - 739.0884*Age

Based on our model, the most important variable in predicting price is the number of bathrooms because of the large beta.

Given the parameters of 3,000 square feet, four bedrooms, three baths, lot size of 15,000 square feet, and age of not more than 20 years, we can expect the price to cost $316,671. While it is difficult to determine if he can afford such a house without knowing his credit history or other debt, with a salary of $120,000, it is likely that such a house would be in his budget.

Appendix

Figure 1: Summary Statistics before removing outliers



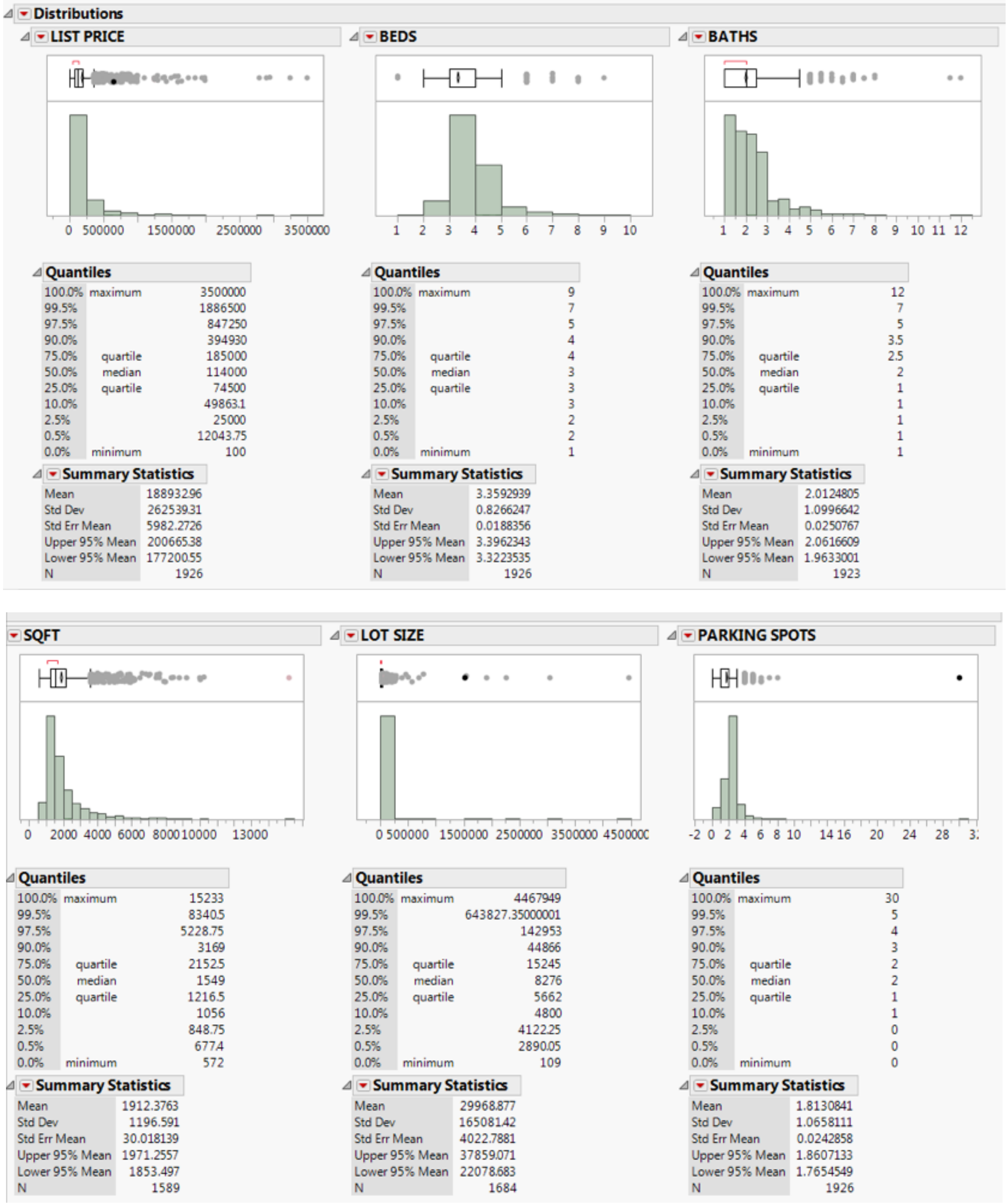
group_project_data (1927 rows, 10 columns)

Columns View Selector

Summary Statistics

Columns	N	N Missing	N Categories	Min	Max	Mean	Std Dev
CITY	1927	0	28	.	.	.	.
STATE	1927	0	1	.	.	.	.
ZIP	1912	15	26	.	.	.	.
LIST PRICE	1927	0	.	100	3500000	188886.76128697	262478.98231457
BEDS	1927	0	.	1	9	3.3591074208614	0.8264506000807
BATHS	1924	3	.	1	12	2.0122141372141	1.0994403549111
SQFT	1590	337	.	572	15233	1911.7924528302	1196.4409323752
LOT SIZE	1685	242	.	109	4467949	29955.543026706	165033.30799987
YEAR BUILT	1852	75	.	1850	2016	1954.7894168467	29.685674343833
PARKING SPOTS	1927	0	.	0	30	1.8126621691749	1.0656953918741

Figure 2: Box plots before removing outliers



Distributions

LIST PRICE

Quantiles

100.0%	maximum	3500000
99.5%		1886500
97.5%		847250
90.0%		394930
75.0%	quartile	185000
50.0%	median	114000
25.0%	quartile	74500
10.0%		49863.1
2.5%		25000
0.5%		12043.75
0.0%	minimum	100

Summary Statistics

Mean	188932.96
Std Dev	262539.31
Std Err Mean	5982.2726
Upper 95% Mean	200665.38
Lower 95% Mean	177200.55
N	1926

BEDS

Quantiles

100.0%	maximum	9
99.5%		7
97.5%		5
90.0%		4
75.0%	quartile	4
50.0%	median	3
25.0%	quartile	3
10.0%		3
2.5%		2
0.5%		2
0.0%	minimum	1

Summary Statistics

Mean	3.3592939
Std Dev	0.8266247
Std Err Mean	0.0188356
Upper 95% Mean	3.3962343
Lower 95% Mean	3.3223535
N	1926

BATHS

Quantiles

100.0%	maximum	12
99.5%		7
97.5%		5
90.0%		3.5
75.0%	quartile	2.5
50.0%	median	2
25.0%	quartile	1
10.0%		1
2.5%		1
0.5%		1
0.0%	minimum	1

Summary Statistics

Mean	2.0124805
Std Dev	1.0996642
Std Err Mean	0.0250767
Upper 95% Mean	2.0616609
Lower 95% Mean	1.9633001
N	1923

SQFT

Quantiles

100.0%	maximum	15233
99.5%		8340.5
97.5%		5228.75
90.0%		3169
75.0%	quartile	2152.5
50.0%	median	1549
25.0%	quartile	1216.5
10.0%		1056
2.5%		848.75
0.5%		677.4
0.0%	minimum	572

Summary Statistics

Mean	1912.3763
Std Dev	1196.591
Std Err Mean	30.018139
Upper 95% Mean	1971.2557
Lower 95% Mean	1853.497
N	1589

LOT SIZE

Quantiles

100.0%	maximum	4467949
99.5%		643827.35000001
97.5%		142953
90.0%		44866
75.0%	quartile	15245
50.0%	median	8276
25.0%	quartile	5662
10.0%		4800
2.5%		4122.25
0.5%		2890.05
0.0%	minimum	109

Summary Statistics

Mean	29968.877
Std Dev	165081.42
Std Err Mean	4022.7881
Upper 95% Mean	37859.071
Lower 95% Mean	22078.683
N	1684

PARKING SPOTS

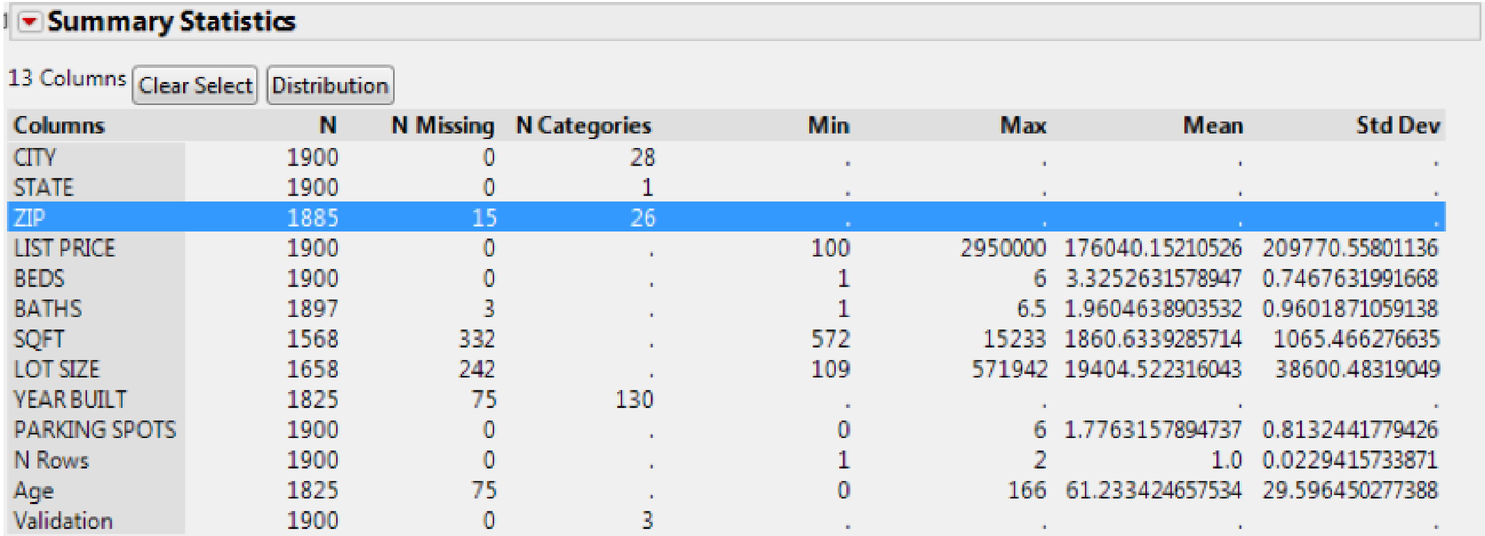
Quantiles

100.0%	maximum	30
99.5%		5
97.5%		4
90.0%		3
75.0%	quartile	2
50.0%	median	2
25.0%	quartile	1
10.0%		1
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics

Mean	1.8130841
Std Dev	1.0658111
Std Err Mean	0.0242858
Upper 95% Mean	1.8607133
Lower 95% Mean	1.7654549
N	1926

Figure 3: Summary after removing outliers

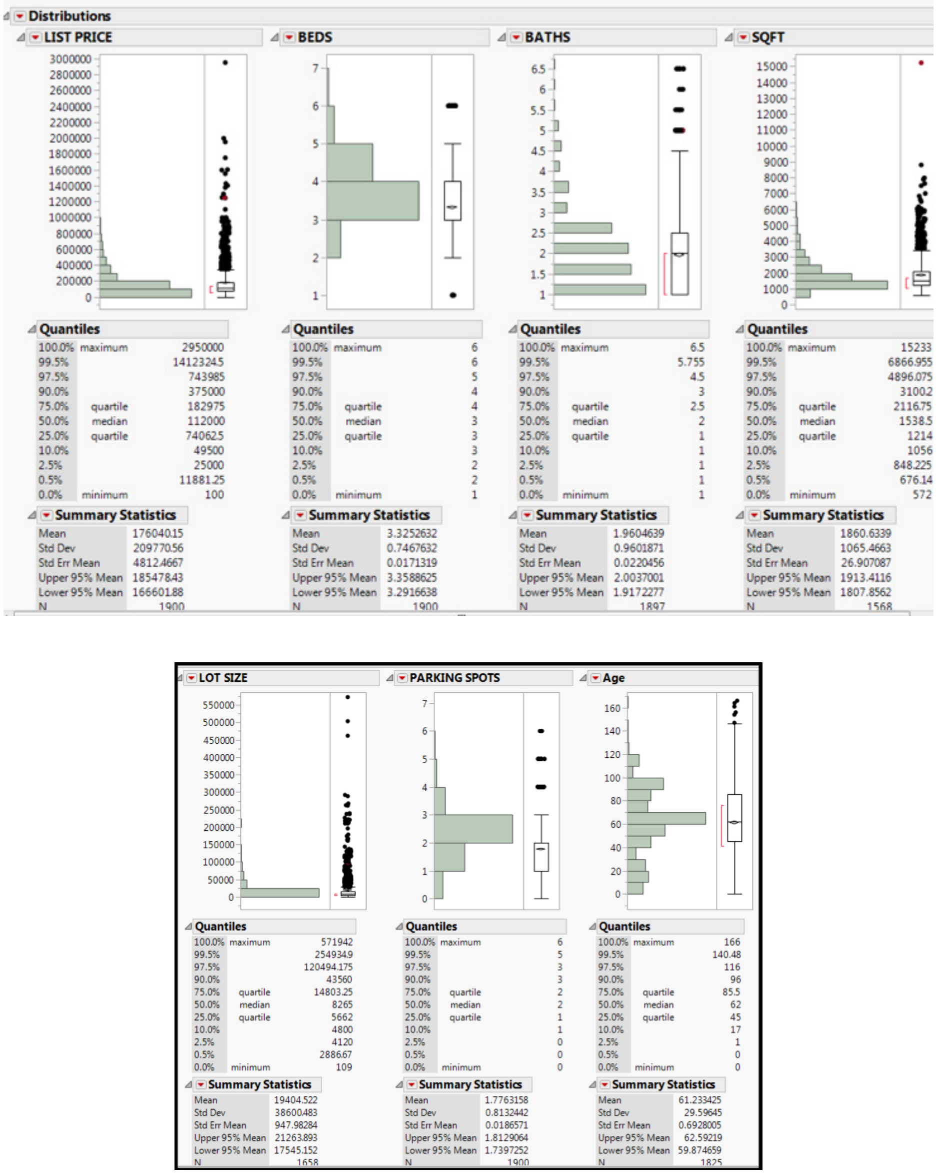


Summary Statistics

13 Columns Clear Select Distribution

Columns	N	N Missing	N Categories	Min	Max	Mean	Std Dev
CITY	1900	0	28	.	.	.	.
STATE	1900	0	1	.	.	.	.
ZIP	1885	15	26	.	.	.	.
LIST PRICE	1900	0	.	100	2950000	176040.15210526	209770.55801136
BEDS	1900	0	.	1	6	3.3252631578947	0.7467631991668
BATHS	1897	3	.	1	6.5	1.9604638903532	0.9601871059138
SQFT	1568	332	.	572	15233	1860.6339285714	1065.466276635
LOT SIZE	1658	242	.	109	571942	19404.522316043	38600.48319049
YEAR BUILT	1825	75	130	.	.	.	.
PARKING SPOTS	1900	0	.	0	6	1.7763157894737	0.8132441779426
N Rows	1900	0	.	1	2	1.0	0.0229415733871
Age	1825	75	.	0	166	61.233424657534	29.596450277388
Validation	1900	0	3	.	.	.	.

Figure 4: Box Plots after removing outliers



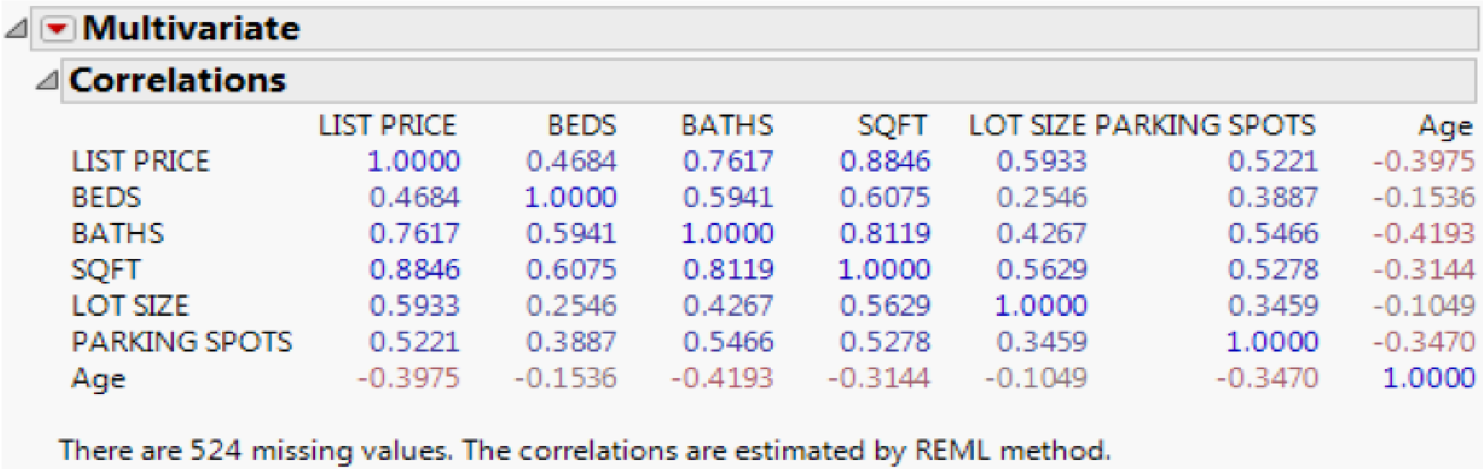
Quantiles

		LIST PRICE	BEDS	BATHS	SQFT	LOT SIZE	PARKING SPOTS	Age
100.0%	maximum	2950000	6	6.5	15233	571942	6	166
99.5%		1412324.5	6	5.755	6866.955	254934.9	5	140.48
97.5%		743985	5	4.5	4896.075	120494.175	3	116
90.0%		375000	4	3	3100.2	43560	3	96
75.0%	quartile	182975	4	2.5	2116.75	14803.25	2	85.5
50.0%	median	112000	3	2	1538.5	8265	2	62
25.0%	quartile	74062.5	3	1	1214	5662	1	45
10.0%		49500	3	1	1056	4800	1	17
2.5%		25000	2	1	848.225	4120	0	1
0.5%		11881.25	2	1	676.14	2886.67	0	0
0.0%	minimum	100	1	1	572	109	0	0

Summary Statistics

	LIST PRICE	BEDS	BATHS	SQFT	LOT SIZE	PARKING SPOTS	Age
Mean	176040.15	3.3252632	1.9604639	1860.6339	19404.522	1.7763158	61.233425
Std Dev	209770.56	0.7467632	0.9601871	1065.4663	38600.483	0.8132442	29.59645
Std Err Mean	4812.4667	0.0171319	0.0220456	26.907087	947.98284	0.0186571	0.6928005
Upper 95% Mean	185478.43	3.3588625	2.0037001	1913.4116	21263.893	1.8129064	62.59219
Lower 95% Mean	166601.88	3.2916638	1.9172277	1807.8562	17545.152	1.7397252	59.874659
N	1900	1900	1897	1568	1658	1900	1825

Figure 5: Co-relations



Multivariate

Correlations

	LIST PRICE	BEDS	BATHS	SQFT	LOT SIZE	PARKING SPOTS	Age
LIST PRICE	1.0000	0.4684	0.7617	0.8846	0.5933	0.5221	-0.3975
BEDS	0.4684	1.0000	0.5941	0.6075	0.2546	0.3887	-0.1536
BATHS	0.7617	0.5941	1.0000	0.8119	0.4267	0.5466	-0.4193
SQFT	0.8846	0.6075	0.8119	1.0000	0.5629	0.5278	-0.3144
LOT SIZE	0.5933	0.2546	0.4267	0.5629	1.0000	0.3459	-0.1049
PARKING SPOTS	0.5221	0.3887	0.5466	0.5278	0.3459	1.0000	-0.3470
Age	-0.3975	-0.1536	-0.4193	-0.3144	-0.1049	-0.3470	1.0000

There are 524 missing values. The correlations are estimated by REML method.

Figure 6: Scatterplot Matrix

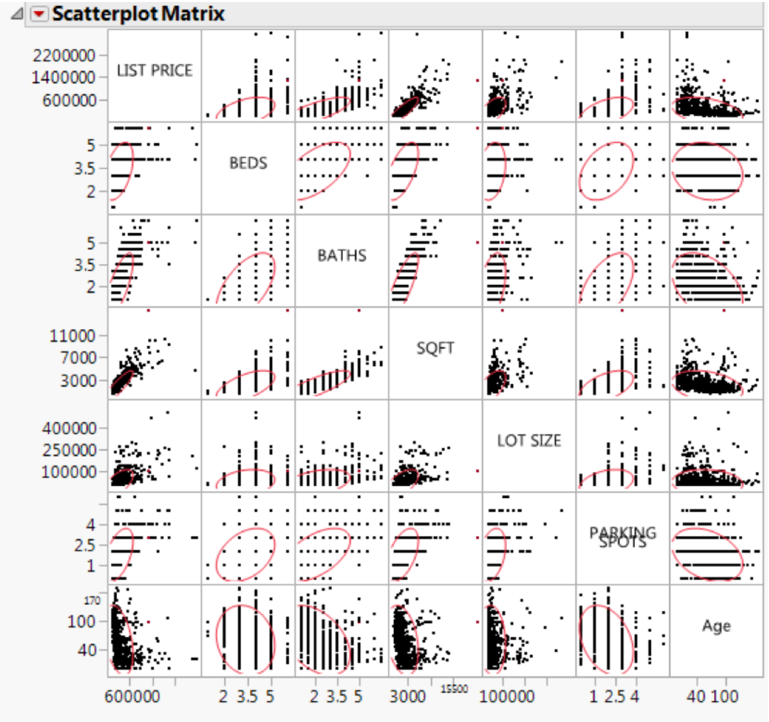
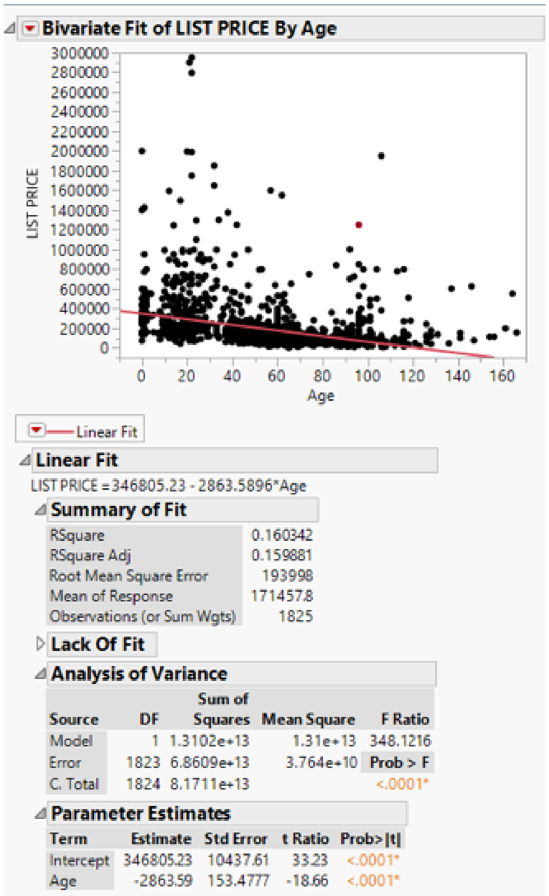


Figure 7: Simple Regression analysis



Bivariate Fit of LIST PRICE By Age

Linear Fit

LIST PRICE = 346805.23 - 2863.5896*Age

Summary of Fit

RSquare	0.160342
RSquare Adj	0.159881
Root Mean Square Error	193998
Mean of Response	171457.8
Observations (or Sum Wgts)	1825

Lack Of Fit

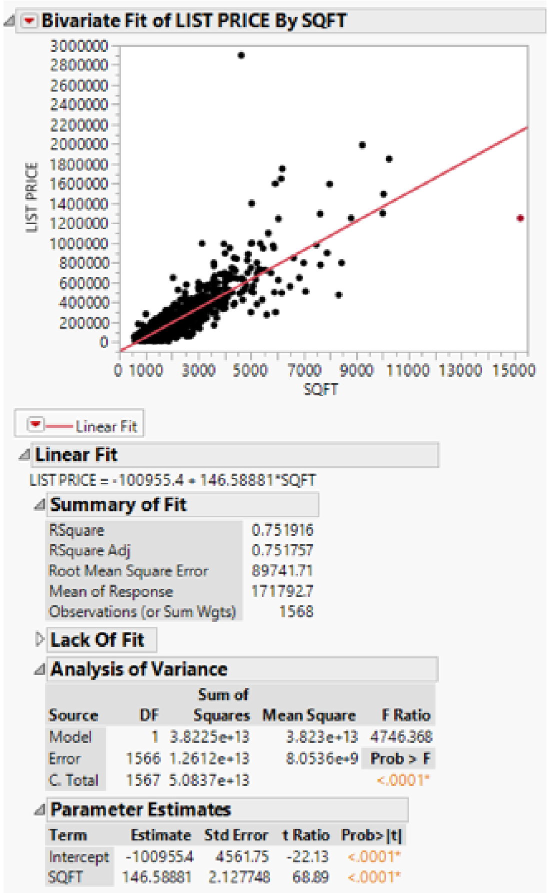
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.3102e+13	1.31e+13	348.1216
Error	1823	6.8609e+13	3.764e+10	Prob > F
C. Total	1824	8.1711e+13		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	346805.23	10437.61	33.23	<.0001*
Age	-2863.59	153.4777	-18.66	<.0001*

List Price by Age



Bivariate Fit of LIST PRICE By SQFT

Linear Fit

LIST PRICE = -100955.4 + 146.58881*SQFT

Summary of Fit

RSquare	0.751916
RSquare Adj	0.751757
Root Mean Square Error	89741.71
Mean of Response	171792.7
Observations (or Sum Wgts)	1568

Lack Of Fit

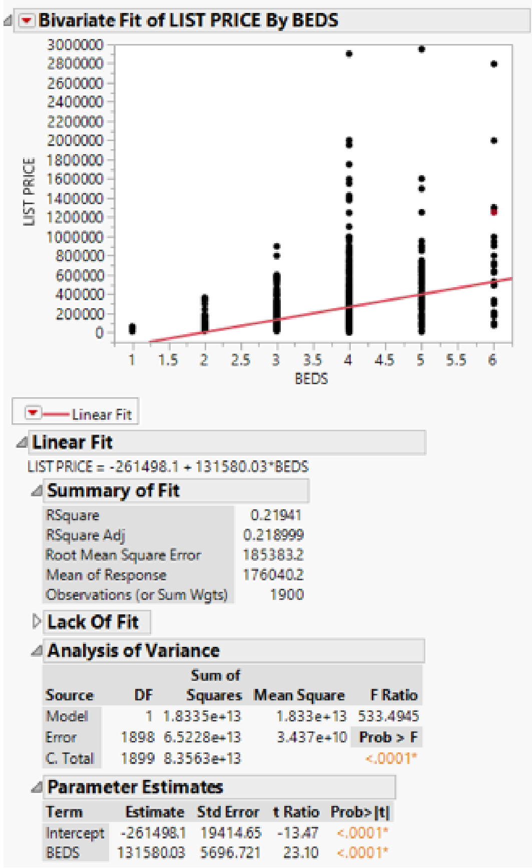
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.8225e+13	3.823e+13	4746.368
Error	1566	1.2612e+13	8.0536e+9	Prob > F
C. Total	1567	5.0837e+13		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-100955.4	4561.75	-22.13	<.0001*
SQFT	146.58881	2.127748	68.89	<.0001*

List Price by SQFT



Bivariate Fit of LIST PRICE By BEDS

Linear Fit

LIST PRICE = -261498.1 + 131580.03*BEDS

Summary of Fit

RSquare	0.21941
RSquare Adj	0.218999
Root Mean Square Error	185383.2
Mean of Response	176040.2
Observations (or Sum Wgts)	1900

Lack Of Fit

Analysis of Variance

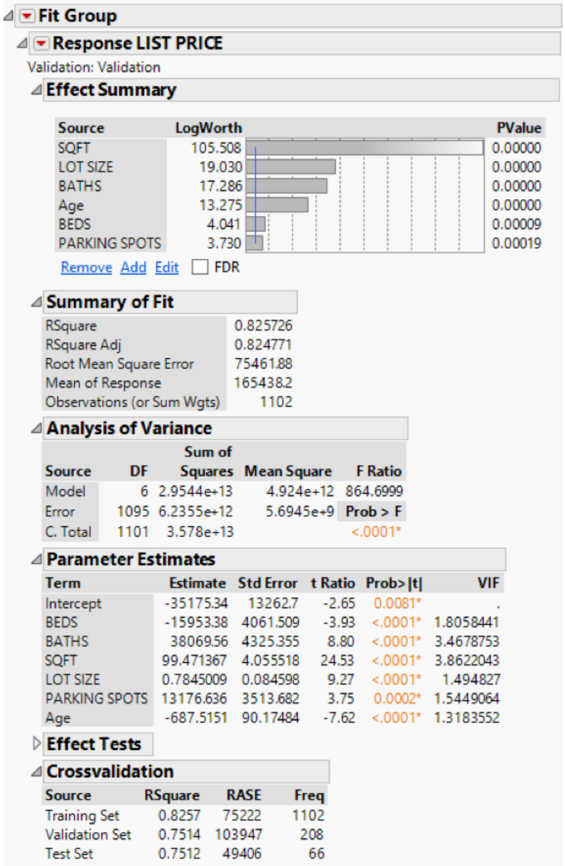
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1.8335e+13	1.833e+13	533.4945
Error	1898	6.5228e+13	3.437e+10	Prob > F
C. Total	1899	8.3563e+13		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-261498.1	19414.65	-13.47	<.0001*
BEDS	131580.03	5696.721	23.10	<.0001*

List Price by beds

Figure 8: Multiple Regression Analysis



Fit Group

Response LIST PRICE

Validation: Validation

Effect Summary

Source	LogWorth	PValue
SQFT	105.508	0.00000
LOT SIZE	19.030	0.00000
BATHS	17.286	0.00000
Age	13.275	0.00000
BEDS	4.041	0.00009
PARKING SPOTS	3.730	0.00019

Summary of Fit

RSquare	0.825726
RSquare Adj	0.824771
Root Mean Square Error	75461.88
Mean of Response	165438.2
Observations (or Sum Wgts)	1102

Analysis of Variance

		Sum of		
Source	DF	Squares	Mean Square	F Ratio
Model	6	2.9544e+13	4.924e+12	864.6999
Error	1095	6.2355e+12	5.6945e+9	Prob > F
C. Total	1101	3.578e+13		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	-35175.34	13262.7	-2.65	0.0081*	.
BEDS	-15953.38	4061.509	-3.93	<.0001*	1.8058441
BATHS	38069.56	4325.355	8.80	<.0001*	3.4678753
SQFT	99.471367	4.055518	24.53	<.0001*	3.8622043
LOT SIZE	0.7845009	0.084598	9.27	<.0001*	1.494827
PARKING SPOTS	13176.636	3513.682	3.75	0.0002*	1.5449064
Age	-687.5151	90.17484	-7.62	<.0001*	1.3183552

Effect Tests

Crossvalidation

Source	RSquare	RASE	Freq
Training Set	0.8257	75222	1102
Validation Set	0.7514	103947	208
Test Set	0.7512	49406	66

OTHER IMPORTANT CHARTS:

PUT OTHER CHARTS HERE

NOTE: All charts used for this analysis, including those posted in this report, can be found in the JMP file.

Vuesol

Headquarters

4625 Alexander Drive Suite
#245,Alpharetta, GA-30022, USA
+1-678-862-0550

Sales office

2033 Gateway Place,
Suite 605, San Jose,
CA 95110
+1-678-862-0550

Development Centre

Thinkspace Building, Plot no 82/11, Patrika
Nagar, HITEC City,
Hyderabad, Telangana 500081
040-29884477

E-mail

info@vuesol.com

To see how we empower your business
visit **www.vuesol.com**

/vuesol @vuesol /company/vuesol