To Renovate or Not to Renovate

BY: Carlos Marin

Outline

Background

Business Problem

Data

Methods

Results

Conclusions

Background

- Every Year, Americans renovate their home so when it comes time to sell, they can get best return. I create two multiple linear regression models based on whether they were renovated or not.

Business Problem

- How much does renovating or not renovating your home affect pricing of the home? By creating a multiple linear regression models we can see how each coefficient works with pricing. From this we can create an actual formula that can be used to calculate how much your home will sell for, once renovated or not renovated

Data

- Using King County data set. The data set is separated by renovated and non-renovated and then normalization data processing is done.

- Renovated data set has 744 rows

- Non renovated data set has 3842 rows.

- Both have 5 columns

Methods

- First separate data a sets by whether they were renovated or not.
- Data Processing
- Perform Multiple Linear Regression identify formula
- Data trimmed to remove unnecessary outliers.
- Post-trimming Update formula.

Results

Pre-Trimming Multiple Linear Regression –Renovated and Non data set

Effect on Price	Area Renovated
-105,896	Bedrooms
14,605	Bathrooms
549, 565	Square footage of living area
-112,215	Square footage apart from basement

Effect on Price	Areas Not- Renovated
-53,475	Bedrooms
5,390	Bathrooms
276,346	Square footage of living area
-31,960	Square footage apart from basement

After running the Multiple linear regression model, first thing we notice is that bedrooms and square footage apart from basement have a negative correlation on price, after renovations We can see apart of number of bedrooms, and square foot above, there is a greater impact on price when it comes to bathrooms and square foot of living area. The negative values in bedrooms and squarefoot above area indicates as price increases, they share negative correlation ("as one goes up the other goes down").
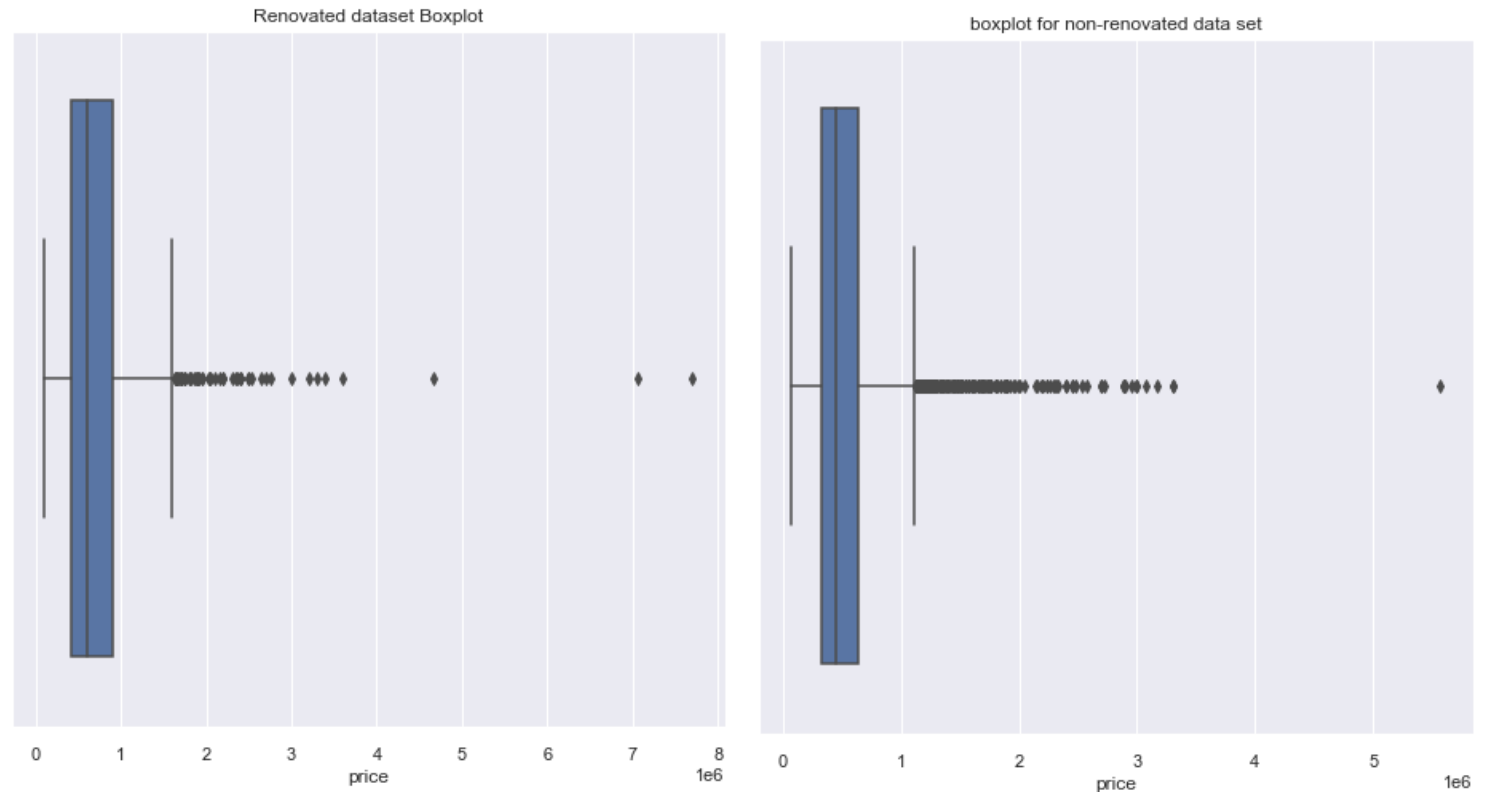
Results

BUT WAIT THERE IS MORE!

Post model evaluation, we see that some coefficients are not really relevant. The most possible cause is due to outliers (or abnormal values).

One fix is the removal of outliers. Well, how many outliers are there in the data set? Well let's look…
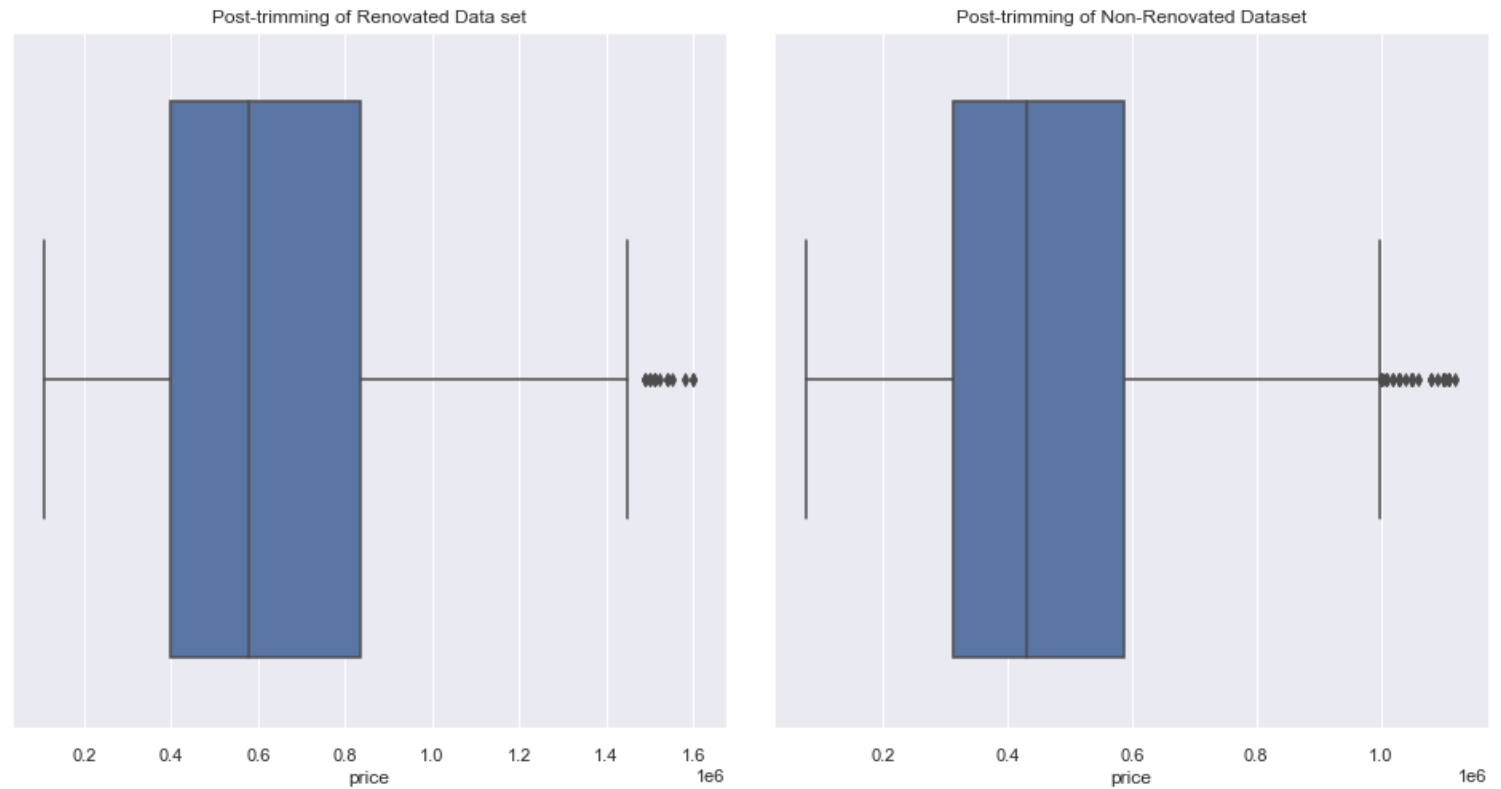
Results

Visualization of Outliers in Renovated and non-Renovated

To see outliers use what's known as boxplots, where everything outside of the lines and boxes, are considered outliers. They affect our model because they are considered abnormal values.

Results

Visualization of Outliers POST-TRIMMING in Renovated and non-Renovated

This is what post-removal of outliers looks like.

Results

Post-Trimming Multiple Linear Regression –Renovated and Non data set

Effect on Price	Area Renovated
-56,480	Bedrooms
99,200	Bathrooms
281,300	Square footage of living area
-47,790	Square footage apart from basement

Effect on Price	Areas Not- Renovated
-29,340	Bedrooms
11,030	Bathrooms
181,400	Square footage of living area
-16,270	Square footage apart from basement

Post Trimming the new shape for the renovated data set is 644 rows (down from 744), 5 columns.
Post trimming new shape for non-renovated data set is 3648 rows (down from 3842), 5 columns

Conclusions/ Findings

- After modeling the multiple linear regression model. We see both negative coefficients (areas renovated and areas not renovated). While both do share negative coefficients with bedrooms and square footage apart from basement. In the renovated data set, the positive coefficients (bathrooms and square footage of living area), quickly negate the negative affects and greatly impact price in comparison on non-renovated houses.

Next Steps…

- 1) Evaluate renovated homes by decade renovated

- 2) Evaluate homes based on what year they were built (by decade built) and comparing whether renovation is necessary.

- 3) Evaluate what other areas of home can greatly impact price for renovated and non-renovated homes (e.g., kitchen, garage, etc.)

Thank you!