Makao Limited Company

Project Overview:

Using regression modeling to analyze house sales in a northwestern county.

Business Problem:

Makao Ltd Company is interested in the need to provide advice to homeowners about how different home features might increase the estimated value of their homes, and by what amount.

Business Understanding:

Makao ltd is a Real Estate Agency based in King County. Makao Ltd company was formed and registered in the year 2010 to help home owners and residents of King County to buy, sell and rent their homes. The company has been doing well but not well enough in helping their customers buy and or sell their homes. They have received a number of negative feeedbacks and complains from customers they helped purchase homes. They also lost a number of customers and would like to do better in terms of serving their customers' and meeting their customers' needs. They would also like to regain the trust of the customers they lost and improve their services.

Company Mission:

Dedicated in transforming lives, homes and in customer satisfaction.

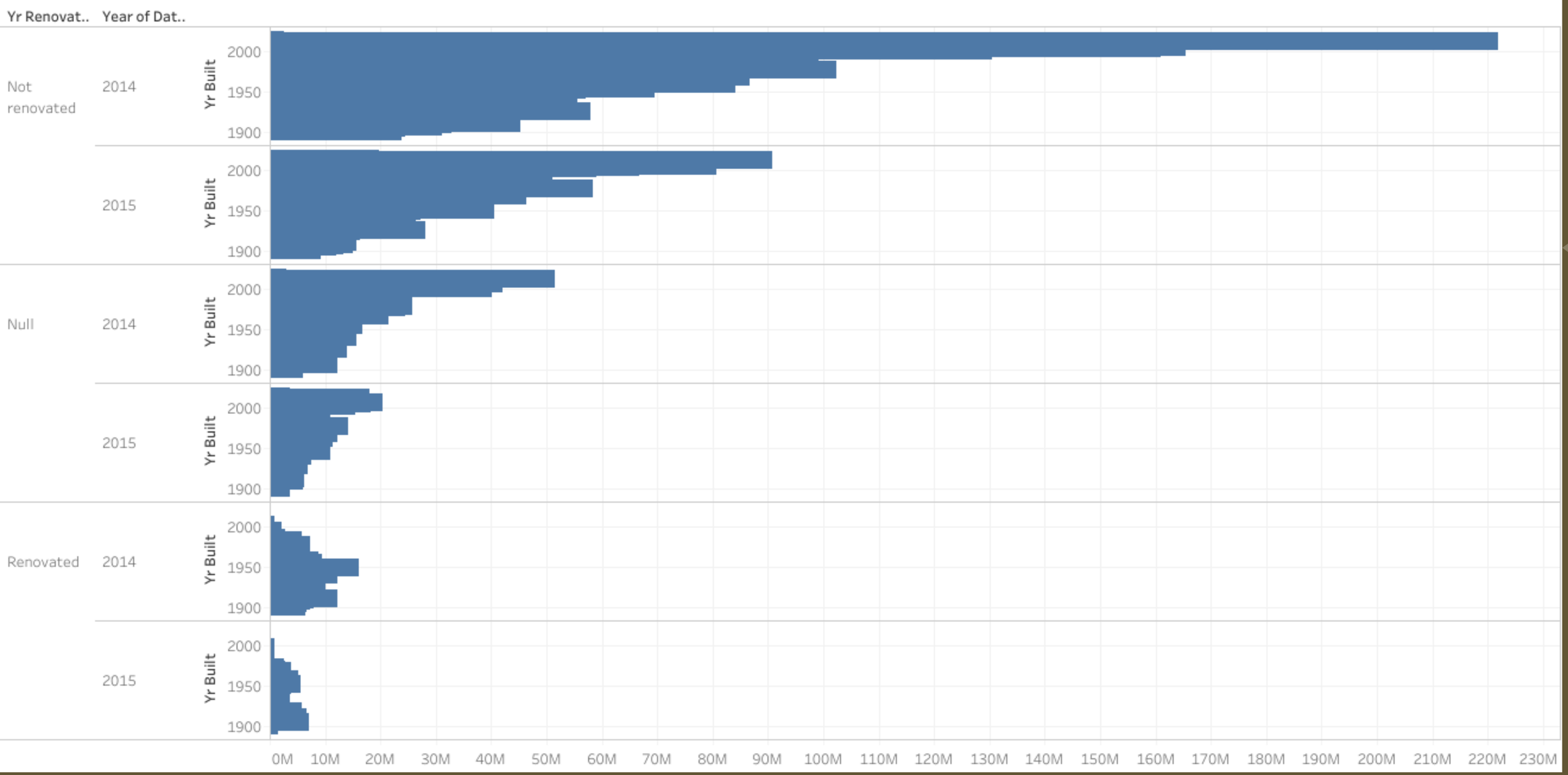
Company Goals:

- *Make clients life easier when it comes to buying and or selling their homes.*
- *Deliver excellent services to clients by putting client's interest first.*
- *Built trust relations with clients.*
- *Building a reputation when it comes to buying and selling homes*

E.D.A

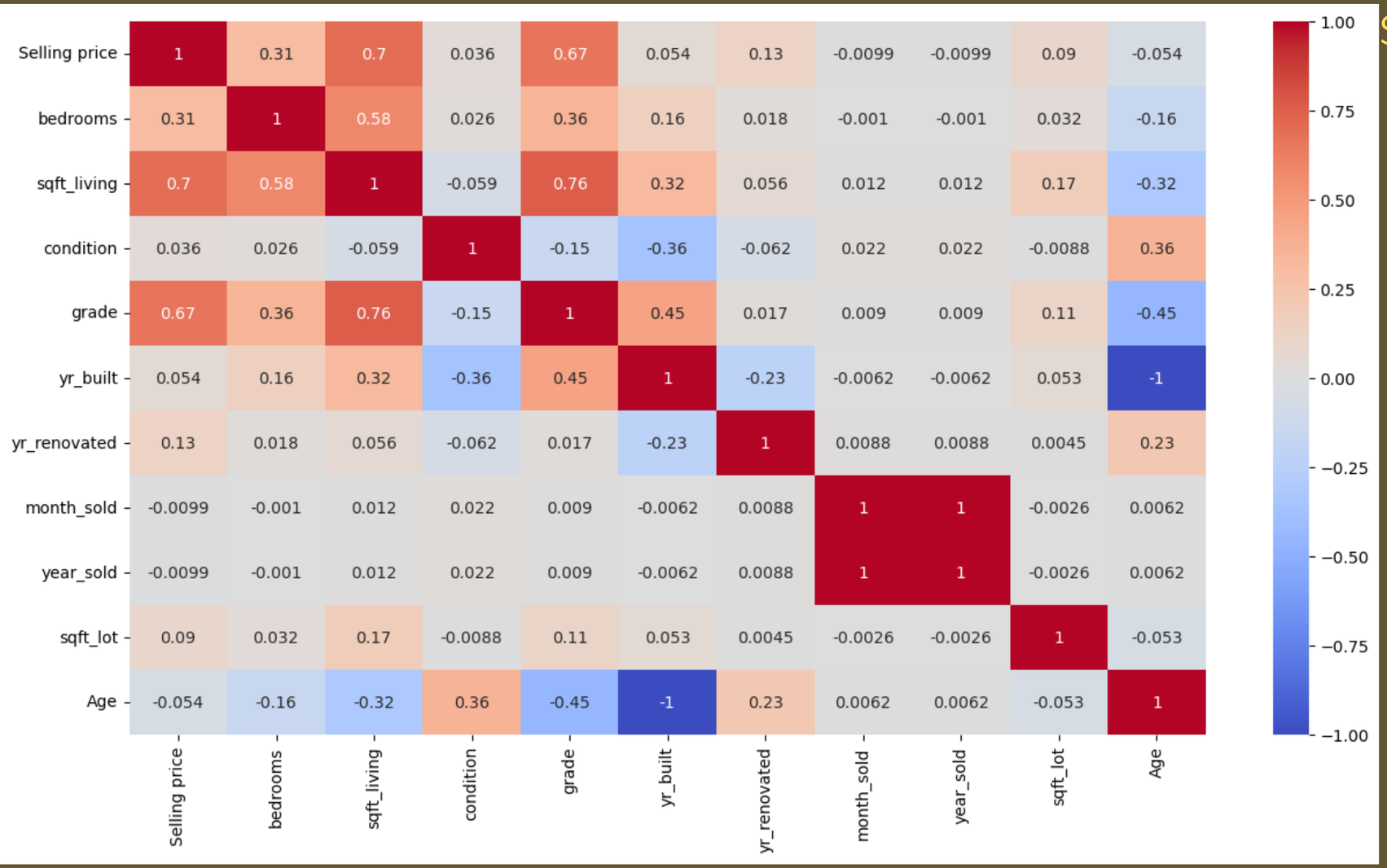
Exploratory Data Analysis

In the following sllides we'll look at graphs between different features and make conclusions on the graphs



From the above picture we can see that people prefer houses were not renovated, which had 0 value in yr_renovated column of the data compared to renovated houses.

Correlation heatmap of the housing data

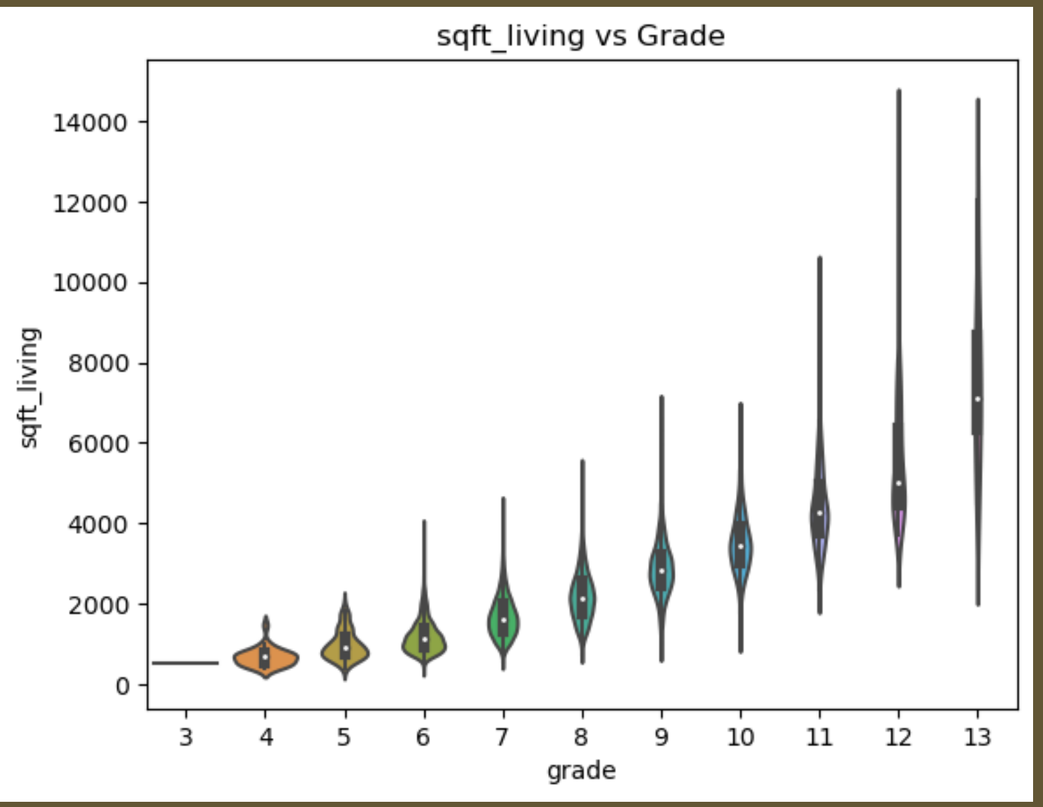


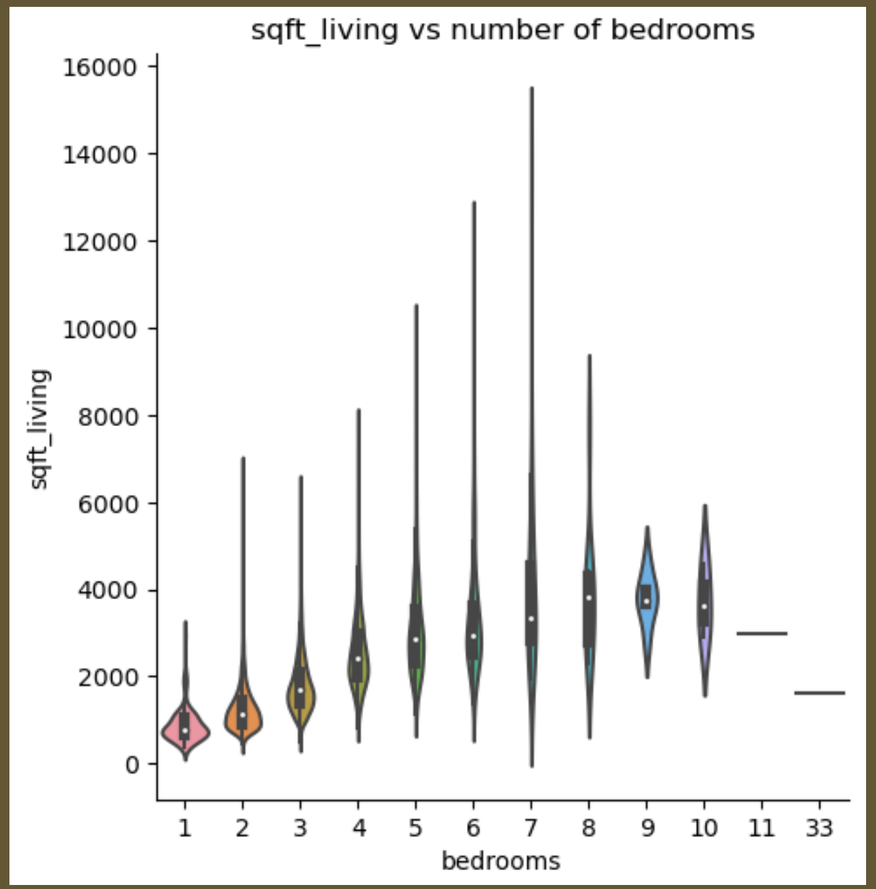
Shows the relationship of various data feature with each other

Correlation ranges from -1 to 1

A correlation of -1 means the variables are negatively correlated, 0 means no correlation and 1 to mean positively correlated, to mean, increase of one variable leads to the increase of the other

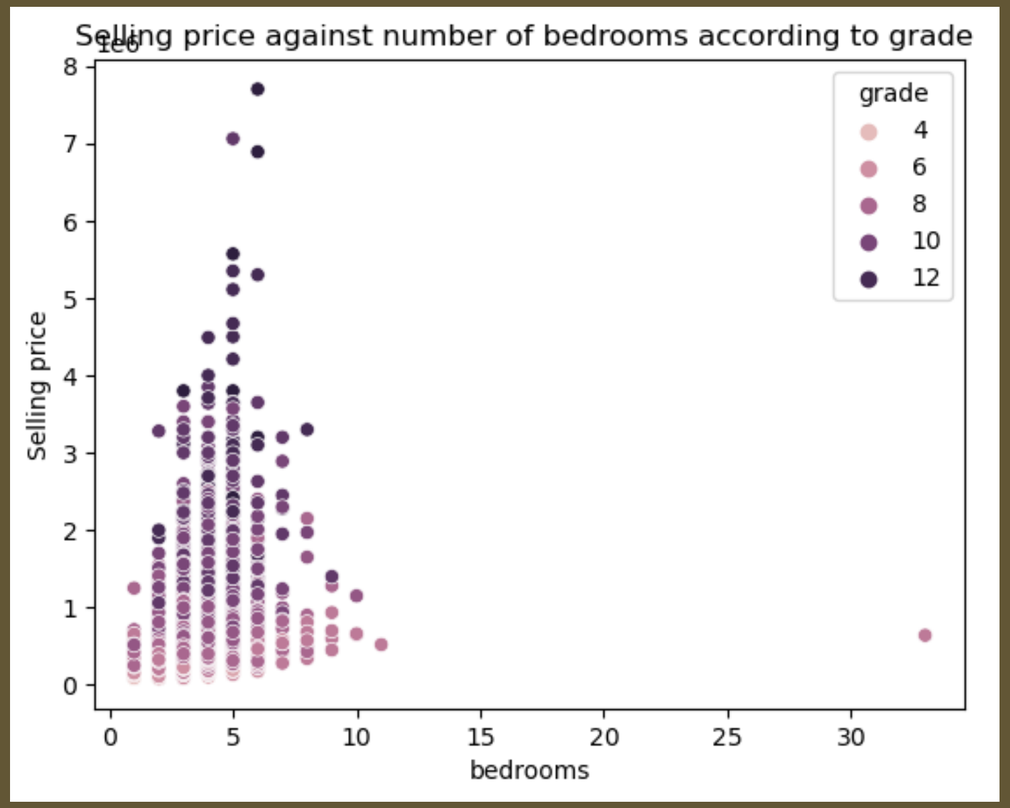
Violin plots on how sqft_living affects grade and bedrooms

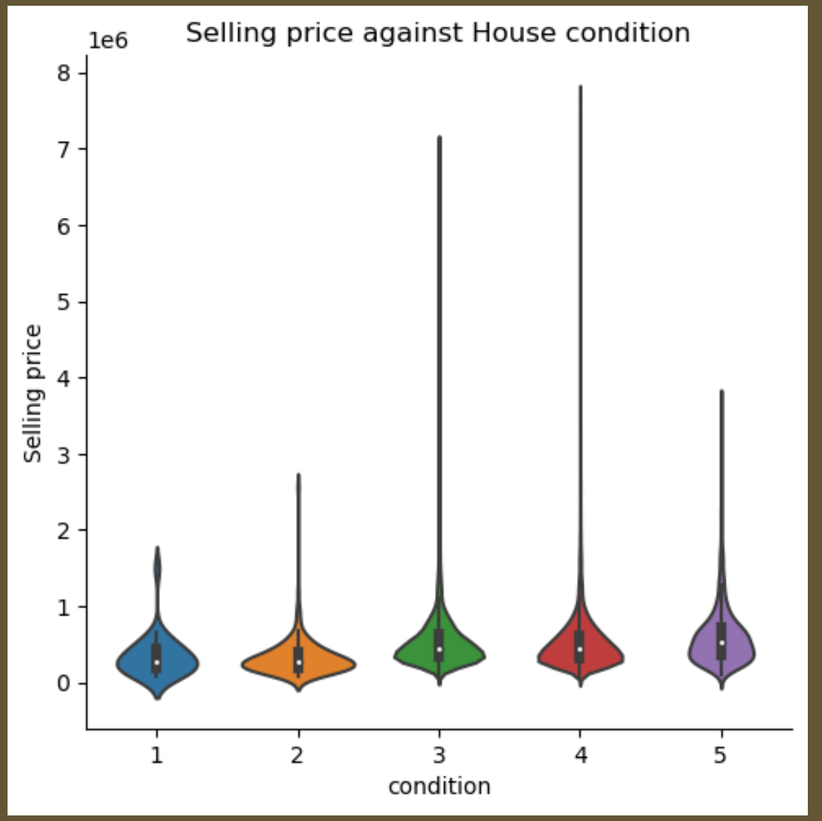




The grade and number of bedrooms increase with increase in the size of the living room

How Selling price is affected by various features

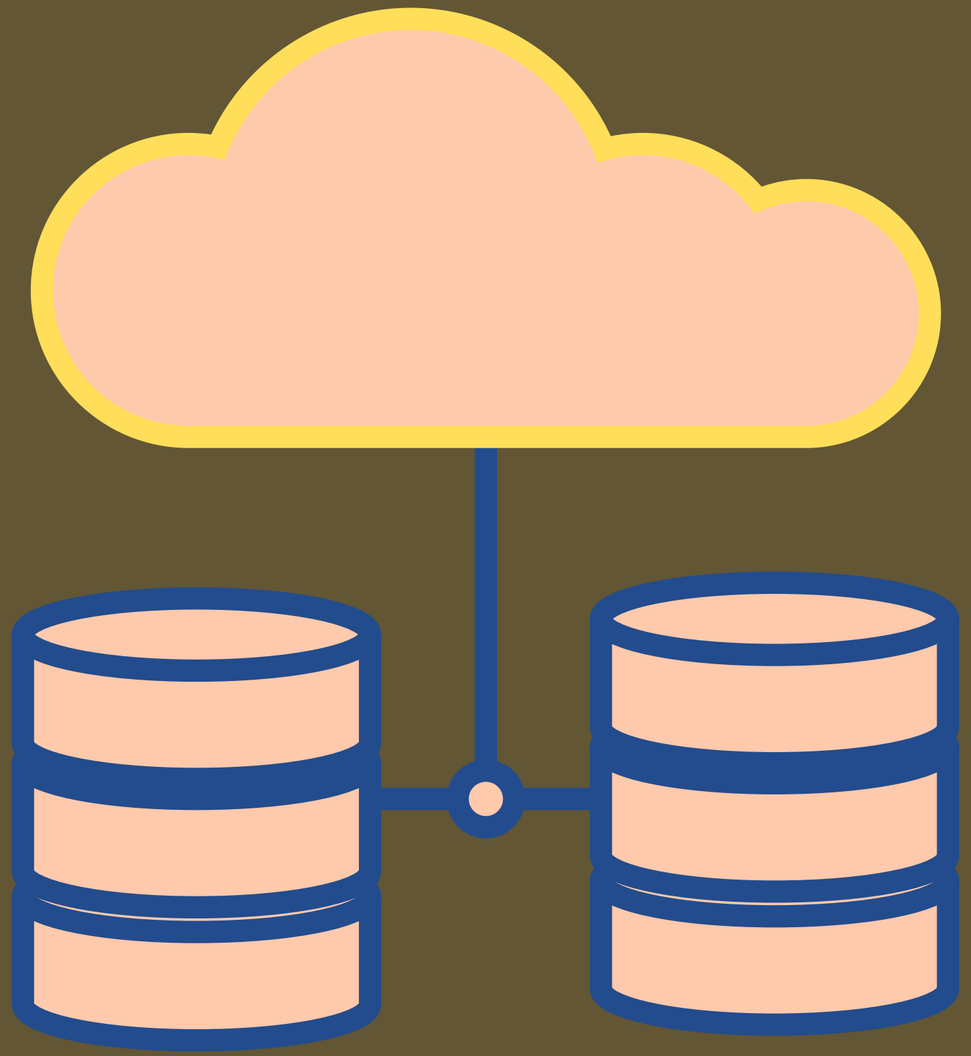




Selling Price increases with increase in number of bedrooms and grade.

As the condition of the house gets better, the selling price increases

Data split



The housing data was then split into training and testing data sets.

Train data was used when training the model and test data was used to test the model built using the train data

Building, Evaluating and validating a model

1. Linear Regression Model

From the Linear Regression modlel we calculated the Mean Square Error for the train and test data and found the difference between the two. The difference was too large for MSE.

We then got the R squared and found that the model explains only 55% of the data.

2. Decision Tree Regression

We calculate the R-Squared and found that the model explains only 31% of the data using this model

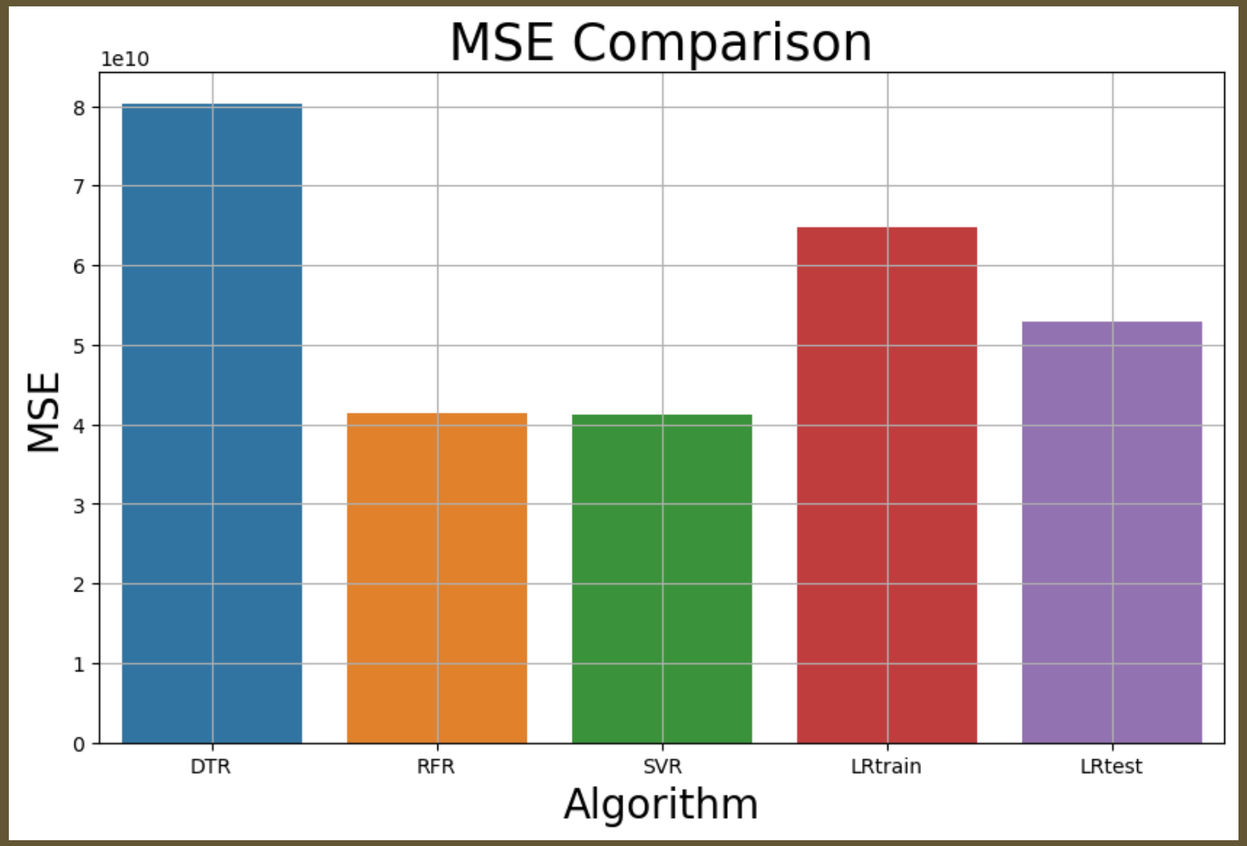
3. Random Forest Regression

Calculating the R- Squared using this model, we got that the model explains 100% of the data.

4. Support Vector Regression

Similarly, this regression technique explains 100% of the data

Comparing the Mean Squared Errors of different Model;

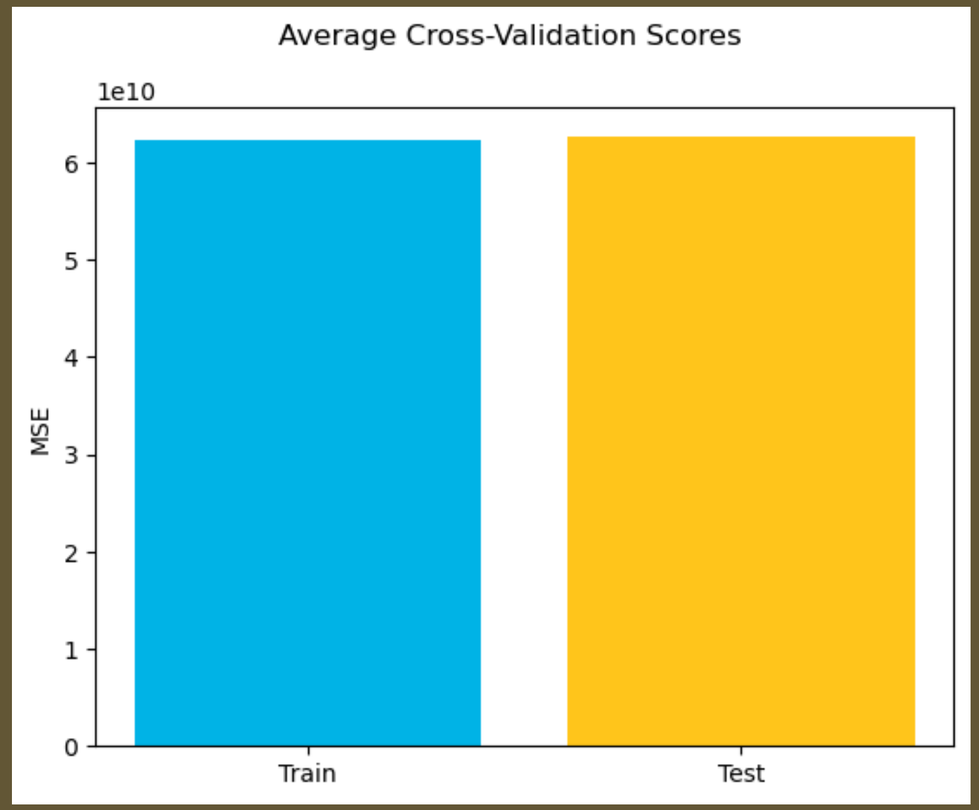


RFR and SVR have close MSE and are the best models since they have a low MSE.

Model Validation

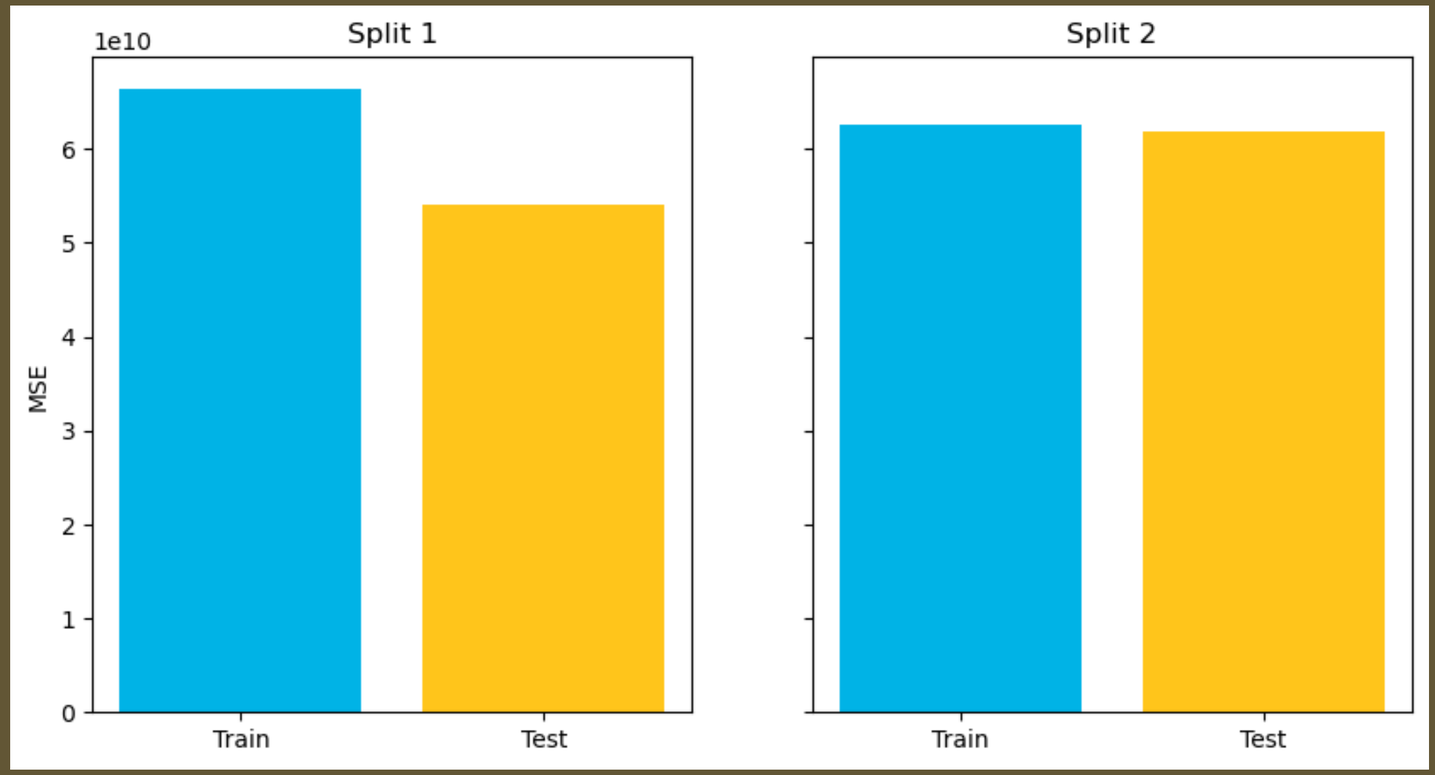
We then validated the models using a number of validation methods;

1. Using K fold cross Validation



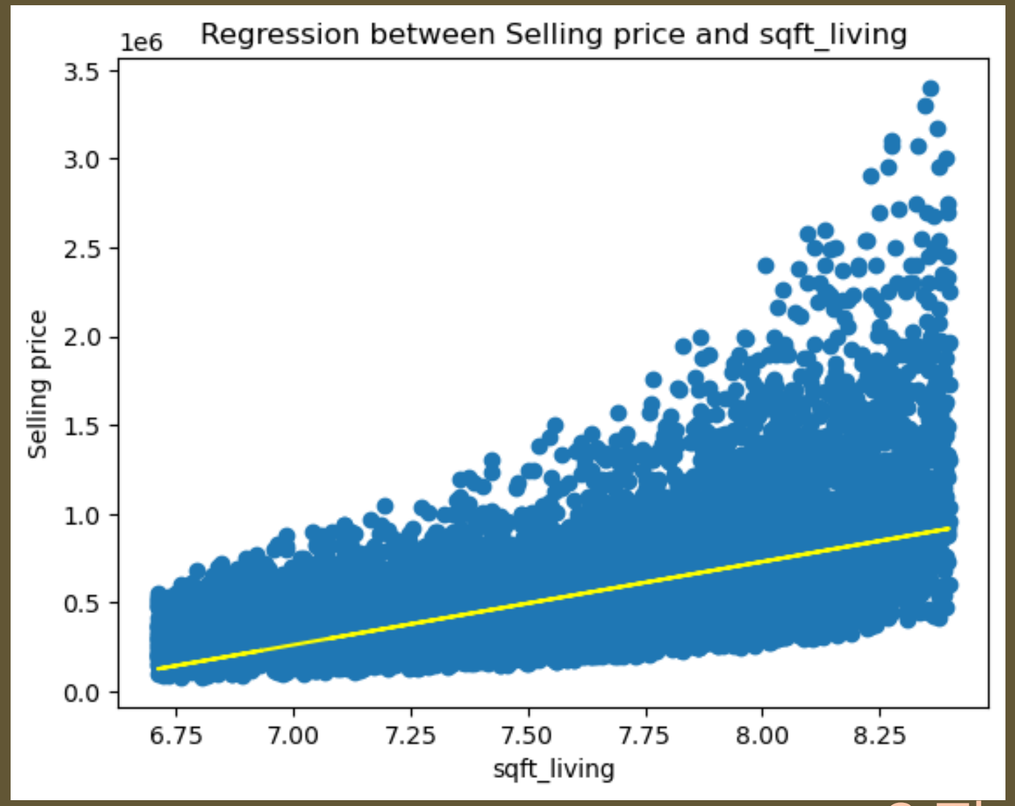
We found that the test and train data had the same Average cross validation Scores which means the model is overfitting.

2.Using Train test split



From this we got split 1 and split 2.
In split 1 the test and train data were appropriate but in split 2 we had an overfitting model

Answers to analytical questions



1. Renovating the house does not increase Selling price

2. House features like number of bedrooms, and the size of living room affect the selling price. As the number of bedroom and size of living room increase, the Selling price increases.

3. The grade of the house affect the Selling price as seen, higher house grades, cause more. Most house with condition 3 and above have a higher selling price than houses in condition 1 and 2.

Conclusion:

The Company should consider the customers opinion on how they would like the houses renovated.

The condition and grade of the house should be improved before selling because they are independent of how old the house is.

Sqft_living is the main feature in determining the Selling price.
A larger living area results to more selling price.