PGPDSBA Capstone Project

- Siddhant Raj Singh

PGPDSBA.O. AUG22.A



-

|  |  |
| --- | --- |
| INDEX | |
| TABLE OF CONTENT | PAGE NUMBER |
| 1) Introduction of the business problem | 2-5 |
| 1.1 Defining problem statement | 3 |
| 1.2 Needs of the study/project | 4 |
| 1.3 Understanding business/social opportunities | 5 |
| 2)Data Report | 6-9 |
| 2.1 Visual inspection of data | 6-7 |
| 2.2 Understanding of attributes | 7-9 |
| 3) Exploratory data analysis | 10-21 |
| 3.1 Univariate analysis | 10-13 |
| 3.2 Bivariate analysis | 13-19 |
| 3.3 Removal of unwanted variables (if applicable) | 20 |
| 3.4 Missing Value treatment (if applicable) | 20 |
| 3.5 Outlier treatment (if required) | 20-21 |
| 3.6 Variable Transformation | 21 |
| 4) Business insights from EDA | 22-25 |
| 4.1 Data imbalance | 22 |
| 4.2 Clustering | 25 |
| 4.3 Business insights from EDA | 23 |

**Introduction**

Problem Statement

A house value is simply more than location and square footage. Like the features that make up a person, an educated party would want to know all aspects that give a house its value. For example, you want to sell a house and you don’t know the price which you may expect — it can’t be too low or too high. To find house price you usually try to find similar properties in your neighbourhood and based on gathered data you will try to assess your house price. The price of the car. This assignment should help the student in exploring the summary statistics, contingency tables, conditional probabilities & hypothesis testing.

Need of the Project

Real estate is one of the most popular domain amongst individuals all around the world. Also, real estate is the most popular choice of investment as well because of its nature of investment and return.

But it is the least transparent industry in our ecosystem. Housing prices keep changing day in and day out and sometimes are hyped rather than being based on valuation. Predicting housing prices with real factors will provide an estimate price of the house which will be highly accurate and will help those which are not familiar with this domain. We do so by making our evaluations based on every basic parameter that is considered while determining the price of the house about which an individual might not be aware of. Therefore, we can analyse the available data of the properties in the area and can predict the price.

Understanding Business/Social Opportunity

House Price prediction, is important to drive Real Estate efficiency. As earlier, House prices were determined by calculating the acquiring and selling price in a locality. Therefore, the House Price prediction model is very essential in filling the information gap and improve Real Estate efficiency.

As people are not aware of the features that helps to determine the property price, we can provide them a service through our house price prediction model.

**EDA and Business Implications**

# Data Description

1. cid: a notation for a house

2. dayhours: Date house was sold

3. price: Price is prediction target

4. room\_bed: Number of Bedrooms/House

5. room\_bath: Number of bathrooms/bedrooms

6. living\_measure: square footage of the home

7. lot\_measure: quare footage of the lot

8. ceil: Total floors (levels) in house

9. coast: House which has a view to a waterfront

10. sight: Has been viewed

11. condition: How good the condition is (Overall)

12. quality: grade given to the housing unit, based on grading system

13. ceil\_measure: square footage of house apart from basement

14. basement\_measure: square footage of the basement

15. yr\_built: Built Year

16. yr\_renovated: Year when house was renovated

17. zipcode: zip

18. lat: Latitude coordinate

19. long: Longitude coordinate

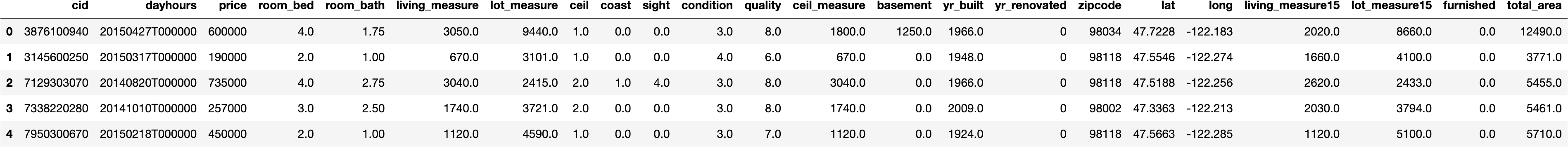
20. living\_measure15: Living room area in 2015(implies-- some renovations) This might or

might not have affected the lotsize area

21. lot\_measure15: lotSize area in 2015(implies-- some renovations) 22. furnished: Based on the quality of room

23. total\_area: Measure of both living and lot

# Sample of the dataset:



* Shape of the data is (21613, 23) i.e. the table has 21613 rows and 23 columns
* The sample of the dataset i.e. the first five rows appear to be perfect.

# Data Set Columns:

A picture containing text, font, screenshot

Description automatically generated

* Above are the columns presented in the data set. As discussed earlier, it seems all the columns are essential to predicting the HOUSE costs. We will perform the required statistical analysis further to find out the spread and distribution of the variables and other important features.

# Check for Duplicates:

**check for duplicated rows has been done. We found that there are no duplicate rows presented in the dataset. Hence the duplicate entries are 0.**

# Info of the Data Set:



* There are total 21613 rows and 23 columns in the dataset. All the columns are of float64 or object or int64 type.
* We have both Categorical and Continuous data.
* There are null values in the data set as observed from the above table.
* What we observe from the above table is that certain variables that are showing up as object data type even though the data in them is of numeric types. These columns have been identified and they are as follows:
  + ['ceil', 'coast', 'condition', 'yr\_built', 'long']
* above mentioned variables should be of type number but instead they are of type object because we can see after listing their unique values, variables have the erroneous value '$'. This is the reason their data type is object type.

After Replacing ‘$’ with nan we get the following info of the Data Set:



## Describing the Data:

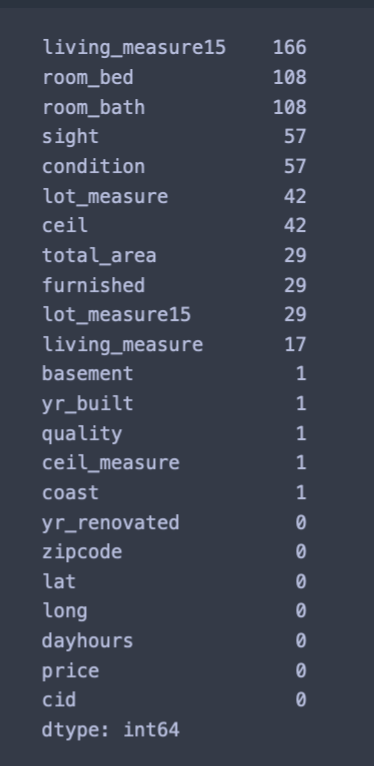
A picture containing text, screenshot, black and white, document

Description automatically generated

* The variable cid is a unique notation given to a house and therefore it provides no significant insight towards the dataset in terms of visualisation or describing data.
* The variable price tells us the price of the house. This is our target variable. The average price is 5.401e+05 with a standard deviation of 3.67362 e+05. The median for the same is 4.5e+05 which is less than the mean. Therefore we can say that the variable price is slightly skewed to the right.
* The variable room\_bed tells us the number of bedrooms in a house. The average bed rooms is 3.37 with a standard deviation of 9.3e-01. The median for the same is 3 which is less than the mean. Therefore we can say that the variable room\_bed is slightly skewed to the right.
* The variable room\_bath tells us the number of bathrooms in rooms. The average bath rooms is 2.1 with a standard deviation of 7.7e-01. The median for the same is 2.25 which is slightly more than the mean. Therefore we can say that the variable room\_bath is slightly skewed to the left.
* The variable living\_measure tells us the square footage of the house. The average living measure is 2.07e+03 with a standard deviation of 9.1e+02. The median for the same is 1.91e+03 which is less than that of mean. Therefore we can say that the variable living\_measure is slightly skewed to the right.
* The variable lot\_measure tells us the square footage of the lot. The average lot measure is 1.51e+04 with a standard deviation of 4.1e+04. The median for the same is 7.61e+03 which is less than the mean. Therefore we can say that the variable is slightly skewed to the right.
* The variable ceil tells us the total floors in the house. The average ceil is 1.492 with a standard deviation of 5.4e-01. The median for the same is 1.5 which is greater than the mean. Therefore we can say that the variable is slightly skewed to the left.
* The variable coast tell us the house which has a view to the waterfront. The coast variable will not be treated as continuous type variable but a categorical variable.
* The variable sight tells us the number of times a house has been viewed. The average sighting is 2.34e-01 with a standard deviation of 7.6e-01. The median for the same is 0.
* The variable ceil\_measure tells us the square footage of house apart from basement. The average ceil measure is 1.78e+03. The median for the same is 1.56e+03 which is less than the mean. Therefore we can say that the variable is skewed to the right.
* The variable basement will be treated as a categorical column indicating whether the house has basement or not.
* The variable yr\_built tells us the built year of the house. The average year in which houses were built is 1969. The maximum is 2015 i.e. the data also includes details of the houses that were built in 2015.
* The variable living\_measure15 and lot\_measure15 is same as living\_measure and lot\_measure but post renovation. The average is 1.98e+03 and 1.27e+04 respectively.
* The variable yr\_renovated tells us the year when the house was renovated. The latest year when the house was renovated is 2015 which is indicated by the column max.

# 

## Check for Missing Values:



* The total number of missing values is 689.
* If we decide to drop the missing values total Number of rows that we will be dropping if we drop null values is 226.
* The total percentage of the values that we will drop if we decide to drop them is 1.045%. Therefore we can treat the null values with mean, median or mode. We can choose mean if the data is normally distributed. But since the data is skewed either to right or left as observed from the description of the data in the above section, we will replace the null values with median. We will replace categorical column with mode.

After treating the null values with mean we get the following table:



As we can see we have replaced all the null values with median and categorical columns with mode.

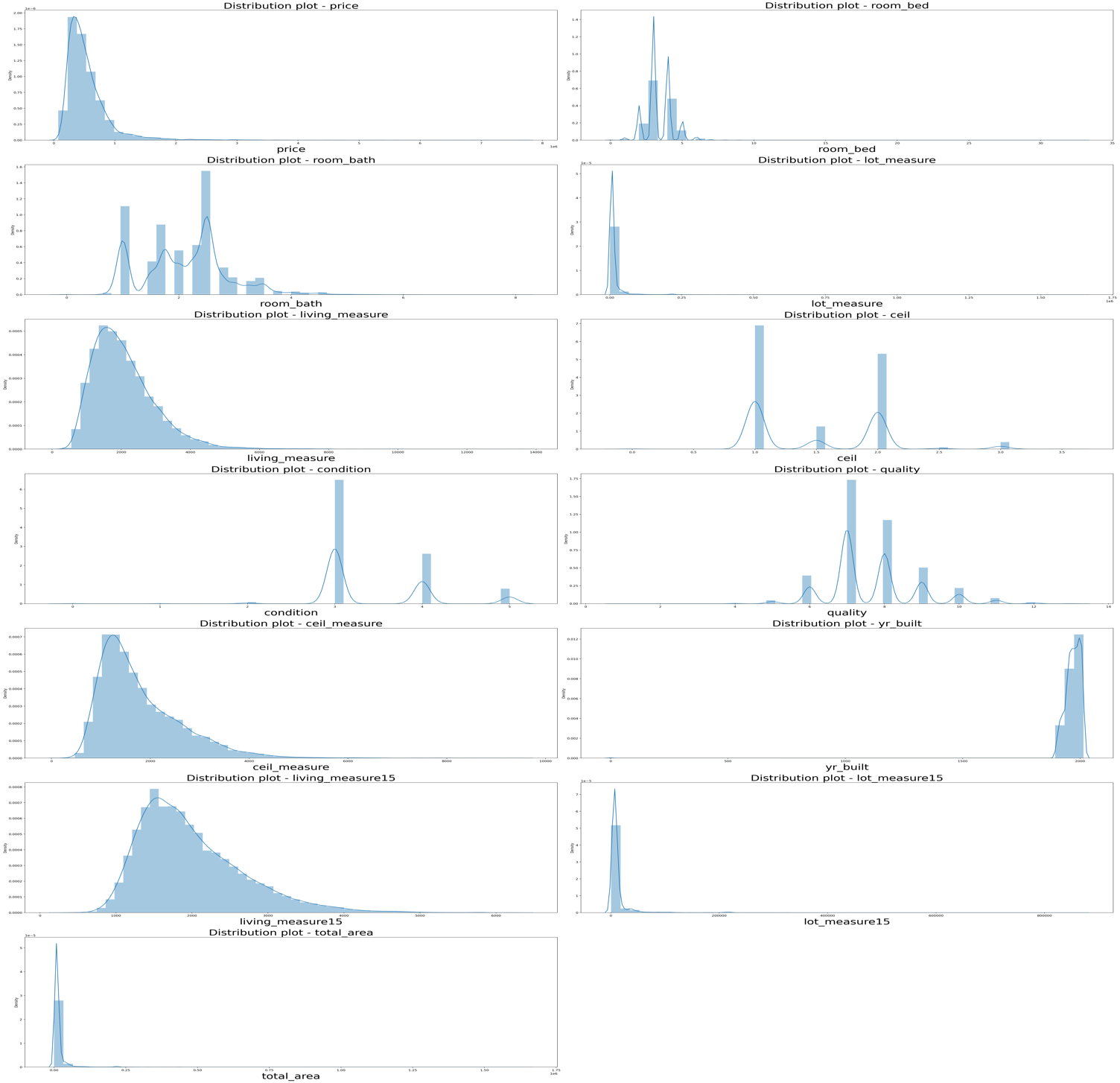
Univariate Analysis

Box Plot

# 

# We can see, there are lot of features which have outliers. So we might need to treat those before building model. OUTLIER TREATMENT DONE IN PYTHON NOTEBOOK.

Distribution Plot



* The variable price tells us the price of the house. This is our target variable. The average price is 5.401e+05 with a standard deviation of 3.67362 e+05. The median for the same is 4.5e+05 which is less than the mean. Therefore we can say that the variable price is slightly skewed to the right.
* The variable room\_bed tells us the number of bedrooms in a house. The average bed rooms is 3.37 with a standard deviation of 9.3e-01. The median for the same is 3 which is less than the mean. Therefore we can say that the variable room\_bed is slightly skewed to the right.
* The variable room\_bath tells us the number of bathrooms in rooms. The average bath rooms is 2.1 with a standard deviation of 7.7e-01. The median for the same is 2.25 which is slightly more than the mean. Therefore we can say that the variable room\_bath is slightly skewed to the left.
* The variable living\_measure tells us the square footage of the house. The average living measure is 2.07e+03 with a standard deviation of 9.1e+02. The median for the same is 1.91e+03 which is less than that of mean. Therefore we can say that the variable living\_measure is slightly skewed to the right.
* The variable lot\_measure tells us the square footage of the lot. The average lot measure is 1.51e+04 with a standard deviation of 4.1e+04. The median for the same is 7.61e+03 which is less than the mean. Therefore we can say that the variable is slightly skewed to the right.
* The variable ceil tells us the total floors in the house. The average ceil is 1.492 with a standard deviation of 5.4e-01. The median for the same is 1.5 which is greater than the mean. Therefore we can say that the variable is slightly skewed to the left.
* The variable coast tell us the house which has a view to the waterfront. The coast variable will not be treated as continuous type variable but a categorical variable.
* The variable sight tells us the number of times a house has been viewed. The average sighting is 2.34e-01 with a standard deviation of 7.6e-01. The median for the same is 0.
* The variable ceil\_measure tells us the square footage of house apart from basement. The average ceil measure is 1.78e+03. The median for the same is 1.56e+03 which is less than the mean. Therefore we can say that the variable is skewed to the right.
* The variable basement will be treated as a categorical column indicating whether the house has basement or not.
* The variable yr\_built tells us the built year of the house. The average year in which houses were built is 1969. The maximum is 2015 i.e. the data also includes details of the houses that were built in 2015.
* The variable living\_measure15 and lot\_measure15 is same as living\_measure and lot\_measure but post renovation. The average is 1.98e+03 and 1.27e+04 respectively.
* The variable yr\_renovated tells us the year when the house was renovated. The latest year when the house was renovated is 2015 which is indicated by the column max.

Countplot Room\_Bed

A picture containing screenshot, rectangle, colorfulness, square

Description automatically generated

Majority of the house has 3 bedrooms followed by 4 bedrooms, 2 bedrooms,5 bedrooms,6 bedrooms,1 bedroom,7,8 and 0 bedrooms. The 33 bedroom what we got in the above analysis is definitely an outlier it needs to be deleted.

Countplot Room\_Bath

A picture containing screenshot, design

Description automatically generated

# Majority of the properties have bathroom in the range of 1.0 to 2.5

# 

Countplot Condition

A picture containing screenshot, rectangle, colorfulness, square

Description automatically generated

Most of the house condition is 3, followed by 4 and then 5. Very less number of houses has a condition rating of 2.

Countplot Quality

A picture containing screenshot, rectangle, colorfulness, square

Description automatically generated

Majority of the houses has been rated in terms of quality from the range of 7-9 and quality 13 seems to be an outlier which calls for further interpretation might be required. There are only 13 building which has the highest quality rating.

Countplot Basement

A blue and orange rectangles

Description automatically generated with low confidence

Most houses do not have basement. But a substantial number of houses comes with basement.

Countplot Furnished

A blue and orange rectangles

Description automatically generated with low confidence

Most properties are not furnished. Furnish column need to be converted into categorical column

Living Measure VS Price

A picture containing screenshot, print

Description automatically generated

There is clear increment in price of the property with increment in the living measure But there seems to be one outlier to this trend. Need to evaluate the same.

Room Bed VS Price

A picture containing line, diagram, plot

Description automatically generated

There is clear increasing trend in price with room\_bed

Lot Measure VS Price

A picture containing screenshot

Description automatically generated

There doesn’t seem to be no relation between lot\_measure and price trend.

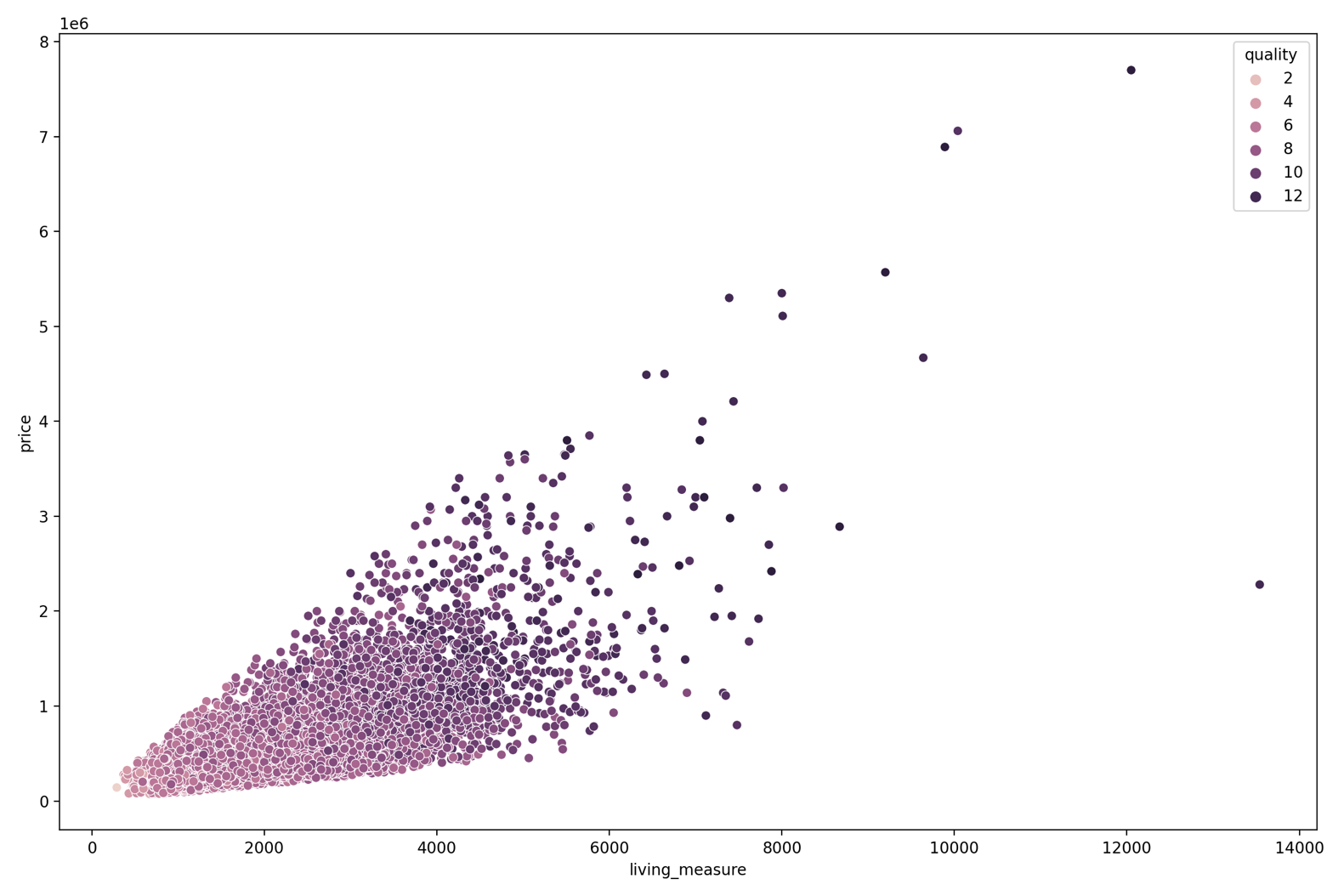
Living Measure VS Price with Sight

A picture containing screenshot

Description automatically generated

Properties with higher price have more number of sights compared to that of houses with lower price

Living Measure VS Price with Quality



Most houses are graded as 6 or more. We can see some outliers as well

Ceil Measure VS Price

A picture containing screenshot, print

Description automatically generated

There is upward trend in price with ceil\_measure

Living Measure VS Price with Renovation

A picture containing screenshot, print

Description automatically generated

Renovated properties have higher price than others with same living measure space.

Living Measure VS Price with Furnished

A picture containing screenshot, design

Description automatically generated

Furnished houses have higher price than that of the Non-furnished houses

Multivariate Analysis

Pair Plot

A picture containing symmetry, screenshot, square, pattern

Description automatically generated

From above pair plot, we observed/deduced below

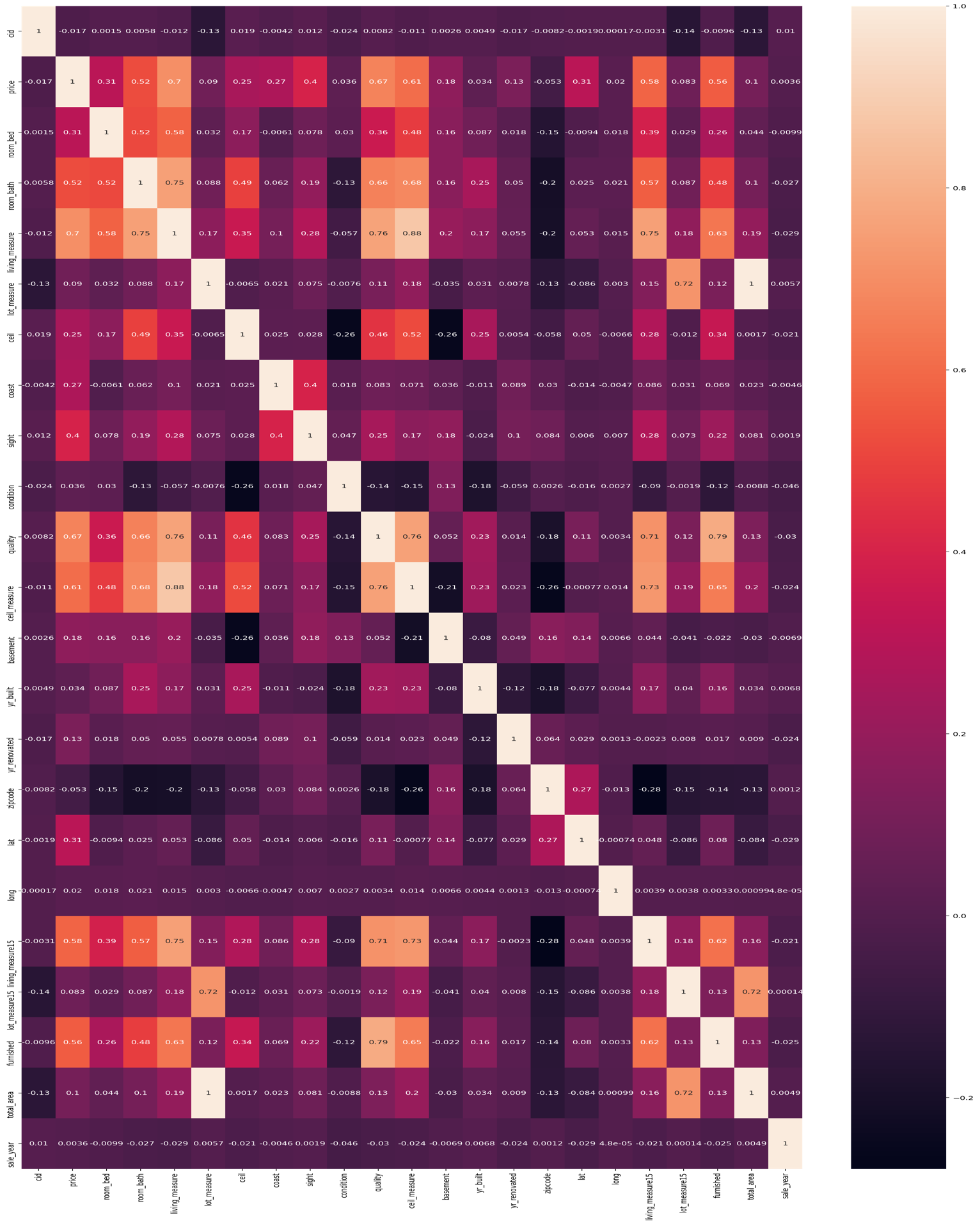
* price: price distribution is Right-Skewed
* room\_bed: our target variable (price) and room\_bed plot is not linear. It's distribution have lot of gaussians
* room\_bath: It's plot with price has somewhat linear relationship. Distribution has number of gaussians.
* living\_measure: Plot against price has strong linear relationship. It also have linear relationship with room\_bath variable. So might remove one of these 2. Distribution is Right-Skewed.
* lot\_measure: No clear relationship with price.
* ceil: No clear relationship with price. We can see, it's have 6 unique values only. Therefore, we can convert this column into categorical column for values.
* coast: No clear relationship with price. Clearly it's categorical variable with 2 unique values.
* sight: No clear relationship with price. This has 5 unique values. Can be converted to Categorical variable.
* condition: No clear relationship with price. This has 5 unique values. Can be converted to Categorical variable.
* quality: Somewhat linear relationship with price. Has discrete values from 1 - 13. Can be converted to Categorical variable.
* ceil\_measure: Strong linear relationship with price. Also with room\_bath and living\_measure features. Distribution is Right-Skewed.
* basement: No clear relationship with price.
* yr\_built: No clear relationship with price.
* yr\_renovated: No clear relationship with price. Have 2 unique values. Can be converted to

Categorical Variable which tells whether house is renovated or not.

* zipcode, lat, long: No clear relationship with price or any other feature.
* living\_measure15: Somewhat linear relationship with target feature. It's same as living\_measure. Therefore we can drop this variable.
* lot\_measure15: No clear relationship with price or any other feature.
* furnished: No clear relationship with price or any other feature. 2 unique values so can be converted to Categorical Variable
* total\_area: No clear relationship with price. But it has Very Strong linear relationship with

lot\_measure. So one of it can be dropped.

Heat Map



We have linear relationships in below features as we got to know from above matrix

* price: room\_bath, living\_measure, quality, living\_measure15, furnished
* living\_measure: price, room\_bath. So we can consider dropping 'room\_bath' variable.
* quality: price, room\_bath, living\_measure
* ceil\_measure: price, room\_bath, living\_measure, quality
* living\_measure15: price, living\_measure, quality. So we can consider dropping living\_measure15 as well. As it's giving same info as living\_measure.
* lot\_measure15: lot\_measure. Therefore, we can consider dropping lot\_measure15, as it's giving same info.
* furnished: quality
* total\_area: lot\_measure, lot\_measure15. Therefore, we can consider dropping total\_area feature as well. As it's giving same info as lot\_measure.

BUSINESS INSIGHTS

* The data doesn’t seem to be unbalanced however there are certain outliers in the data which makes the data meaningless, for example the data given suggests that a house with 1 bedroom is sold for more than 10000000 which is practically not possible, it would be better to drop such entries during model building.
* The missing data in the data set is already imputed based on the data type.
* Living measure is the most significant variable in our analysis and since living measure, lot measure and ceil measure are proportional we need not spend a lot of time in analysing these variables, analysis based on living measure would provide much greater insights.
* It is evident from EDA that an ideal house would be the one with 2-3 bedrooms and 3 bathrooms, even though houses with 8 and >8 bedrooms and bathrooms have sold for a higher price a lot of people doesn’t seem to be buying them, higher number of records are sold with three-bedroom houses hence an equal or even more revenue could be obtained by selling more houses with three bedrooms and bathrooms.
* Although majority of houses are not furnished, it is seen in bivariate analysis that furnished houses produce more revenue compared to unfurnished ones.

Is the data unbalanced? If so, what can be done? Please explain in the context of the business

As the variable is continuous in nature we do not need to worry about the unbalancing of data. But we can look from the independent features as well to see if the data is balanced or unbalanced. For Example – 90% of the data has basement whereas 10% of the data do not have basement. To check if the data Is balanced or unbalanced, we need to look at the target variable with respect to the independent features.

Note: CLUSTERING DONE IN PYTHON NOTEBOOK

Variable considered for Model Building-

A screen shot of a computer code

Description automatically generated

We are dropping the following columns from the dataset as these information are repeated in multiple columns:

cid, dayhours , yr\_renovated, zipcode, lat, long

Data Cleaning and Pre-processing

* Missing data in a few columns. As we can see we have replaced all the null values with median and categorical columns with mode.
* Incorrect data type for some variables. What we observe from the above table is that certain variables that are showing up as object data type even though the data in them is of numeric types. These columns have been identified and they are as follows:
  + ['ceil', 'coast', 'condition', 'yr\_built', 'long']
* above mentioned variables should be of type number but instead they are of type object because we can see after listing their unique values, variables have the erroneous value '$'. This is the reason their data type is object type. Garbage Values in some columns of the dataset.
* Outliers in the dataset as seen from the BOXPLOT.

Model Building:

Machine Learning models can be understood as a program that has been trained to find patterns within new data and make predictions. These models are represented as a mathematical function that takes requests in the form of input data, makes predictions on input data, and then provides an output in response. First, these models are trained over a set of data, and then they are provided an algorithm to reason over data, extract the pattern from feed data and learn from those data. Once these models get trained, they can be used to predict the unseen dataset.

There are two important metrics which we use to measure the performance of the regression model-

1. Root Mean Square- Which measures the model prediction error. It corresponds to the average difference between the observed and predicted values by the model. The lower RMSE is the better the model.

2. R-Square- Representing the squared correlation between the observed known outcome values and the predicted values by the model . The higher R2 better the model.

Train Test Split:

* We will first create ‘X’ and ‘y’ which will contain Features/predictable variables and Target. ‘X’ will contain Features of the dataset and ‘y’ will contain the Target variable.
* To fairly evaluate our model’s performance, we don’t want to evaluate it on the same data it was trained at. Therefore, we will separate out a training set and a test set. There are four components to this – X\_train, X\_test, y\_train, y\_test. We will use Sklearn library in which we have sub-package model selection.

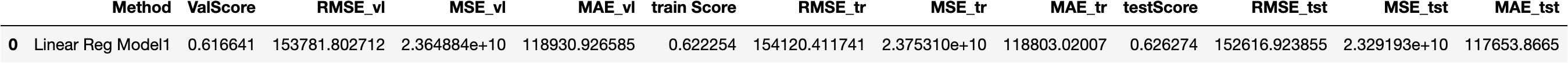
From model selection we will import ‘train\_test\_split’. We will keep the size of train and test in the ration of 70:30.

* We now invoke the linear regression function and find the best fit model on training data. Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.
* We create an instance of Linear Regression model. We have various parameters for the same. These parameters are called the hyperparameters. After creating the model we will train the model on the training dataset using the fit method. We will use this model which has been fit using the training dataset to make predictions of the dataset.

Linear Regression Model(with Lasso and Ridge) :

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable i.e. Linear Regression is used to predict a quantitative outcome variable (y) based on one or multiple predictor variables (x)

Linear Regression Model :



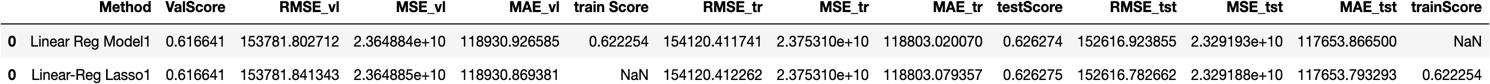
Linear Reg-Actual and Predicted joint plot:

A blue and black diagram

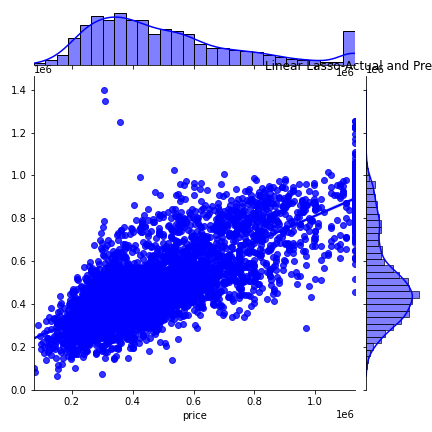
Description automatically generated

The linear regression model performed with scores 0.62 in training data, testing data set and 0.61 in validation data set respectively.

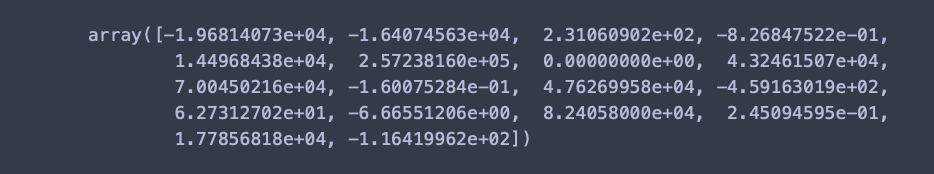
Lasso Model:



Linear Lasso-Actual and Predicted joint plot:

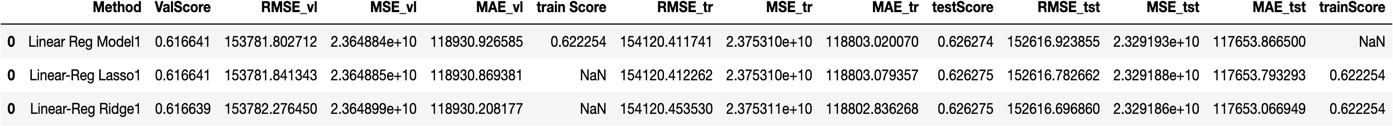


The lasso linear regression model performed with scores 0.622 in training data set and 0.61 in validation data set respectively.



The coefficient of 1 variable in lasso model is almost '0', signifying that the variable with '0' coefficient can be dropped.

Ridge model:



Linear Ridge-Actual and Predicted joint plot:

A blue and white diagram

Description automatically generated

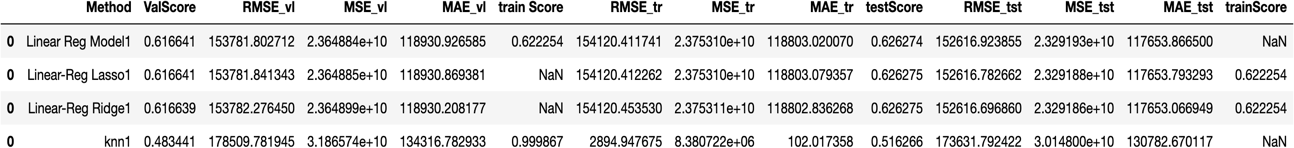
The Ridge linear regression model performed with scores 0.6222 in training data set , test data set and 0.616 in validation data set respectively.

A screenshot of a computer

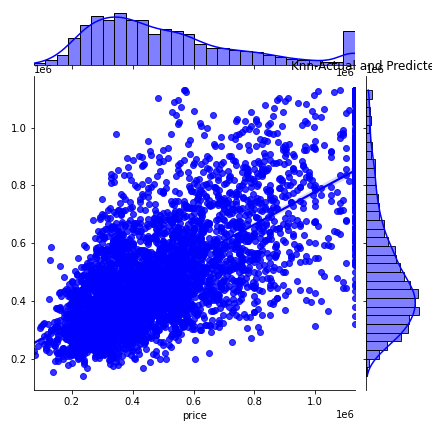
Description automatically generated

The coefficients of variables in ridge model are all non-zero, indicating that none of the variables can be dropped.

KNN Model:

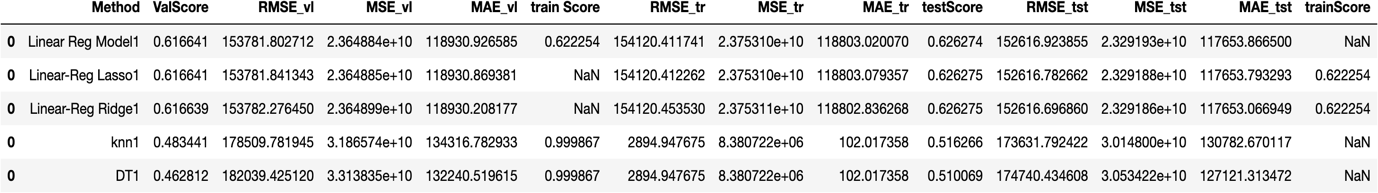


KNN-Actual and Predicted joint plot:

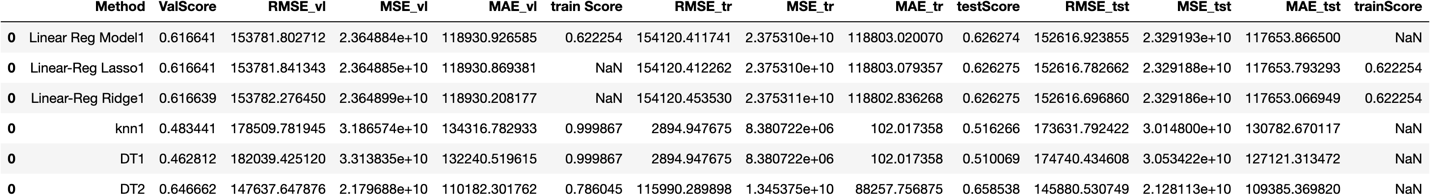


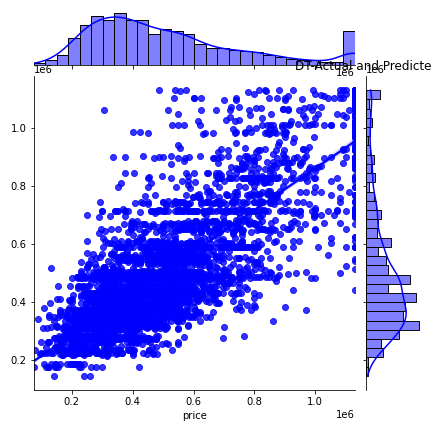
Though KNN regressor performed well in training set, the performance score in validation and test set is very less. This shows that the model is overfitted in training set

Decision Tree:



Above performance of initial Decision tree model shows overfit in training set with 1.00 score and low performance in validation and test set.





Above decision tree model with modified parameter has better performed on the training ,testing set, and validation set compared to initial decision tree model. But overall decision tree has not performed well than linear regression models.

Ensemble techniques:

Boosting and Bagging:

A screenshot of a graph

Description automatically generated

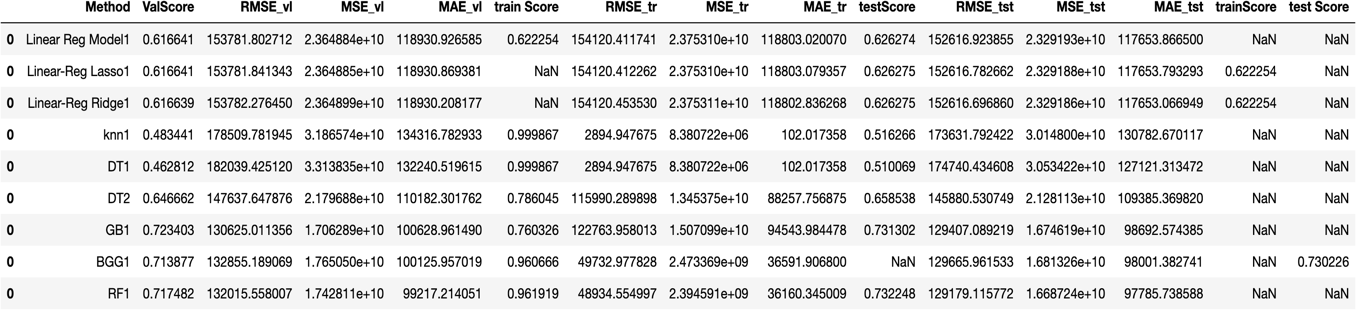
Gradient boosting model has provided good scores in both training, testing and validation sets .

A screenshot of a computer screen

Description automatically generated

Bagging model also performed well in training, testing and validation sets. There seems to be overfitting in training set. We need to analyse further by hyper tuning.

Random Forest Regressor:



Random forest model has performed well in training, testing and validation set. There is scope of further analysis on this model.

Model Selection/Validation:

Ensemble models: in summary ensemble models have performed well on training , Testing and validation sets. These models will be selected for further analysis with hyper tuning and feature selection. Out of the two ensemble models, Boosting is considered for fine tuning as it performed well on train data set and score of Train and test data set are nearly equal. We can clearly see gradient boosting is giving better result in comparison with other ensemble methods. Also, the score of 0.76 on training set indicates no overfitting of the model.

Feature Importance:

RF1-Feature\_Importancce:

A screenshot of a computer

Description automatically generated

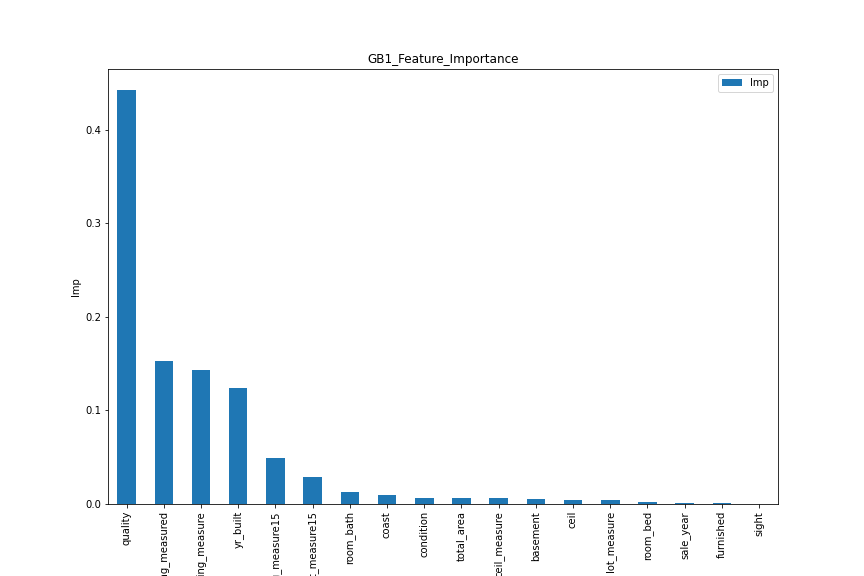
A screen shot of a graph

Description automatically generated

GB1-Feature\_Importance:

A screenshot of a computer

Description automatically generated



* The ensemble models have performed well compared to that of linear model, KNN models.
* The best performance is given by Gradient boosting model with training score 0.76,RMSE-122763, Validation (score-0.72,RSME-130625), Testing(score 0.73,RMSE- 129407).
* For further improvisation, the datasets can be made by treating outliers in different ways and hyper tuning the ensemble models.
* The best performance is given by Random Forest model with training (score 0.96,RMSE- 48934), Validation (score-0.71,RSME-132015), Testing(score 0.73,RMSE-129179)

Insights from Analysis:

* The top key features that drive the price of the property are as follows:
  + Quality
  + living\_measure15
  + lot\_measure15
  + ceil\_measure
  + Room\_bath
  + coast
  + Condition
* There is increasing trend in price with bed\_room increases.
* We have maximum of house prices 77,00,000 and min 75000.
* Majority of houses are not furnished , furnished houses produce more revenue as compared to unfurnished.
* From above analysis it indicates the fact , customers do not compromise on the purchase as against the quality of property.
* Higher bedrooms may signify bigger property. But also that signifies lower area for common amenities, which may not be preferable for customers.
* There is increasing Living measure is the most significant variable in our analysis and since living measure.
* There is increasing trend in price with bedroom increases .An ideal house would be the one with 2-3 bedrooms and 3 bathrooms even though houses with 7-8 or greater than 8 bedrooms and bathrooms have sold for a higher price a lot of people doesn’t seem to be buying them, higher number of records are sold with three-bedroom houses hence an equal or even more revenue could be obtained by selling more houses with three bedrooms and bathrooms.

Interpretation / recommendation:

* In this project we have focused only on factors which affects the price of house directly like its type, age, area. Through our research we have found that there are various hedonic factors which can influence buyers, like
  + Income
  + Crime rate
  + Amenities
  + Mortgage availability
  + Interest rates
  + Inflation
* A real estate developer must lay focus on the construction as it is well understood that a good quality construction ,shall fetch a higher price.
* An individual must carefully analyse all the significant variables before arranging necessary funds for purchase of a property. Similarly, a broker may also focus on attributes of built year, quality etc. while estimating the tentative brokerage.
* Factors such as proximity to coast, or property views do not determine the price and should not be considered by a buyer to estimate the prices.
* The parameter of number of bedrooms may be misleading and should be understood in relevance.
* A properties value is important in real estate transaction. Housing price trends are not only the concern of buyers and sellers, but it also indicates the current economic situation
* Property should be renovated time to time because customer will attract to renovated property
* Property should be furnished or semi furnished, to get higher price.
* if the all the facilities like school, hospital, stores, garden, airport, Railway station etc. are available nearby property then it will get more preference from customer and will get higher price. Properties with higher price have more no of viewed compared to that of houses with lower price.
* There is clear increase in price of the house with higher rating on quality.