**3.0 RESULTS AND DISCUSSIONS**

* Now, we will conclude the **Accuracy** of our models on the 20 % **Test data** through various measures but firstly we will verify it by Predict and post resample function.

**Model 1**

Graphical user interface, text, application

Description automatically generated

**Model 2**

Text

Description automatically generated

**Model 3**

Text, application

Description automatically generated

**Model 4**

Text

Description automatically generated

**Model 5**

Text

Description automatically generated

**Final Model**

Text

Description automatically generated with medium confidence

We can see that the R squared for the final model is 3.68 which is higher than the usual. We will further discuss the implications of the same.

* Further we have created a **Prediction column** to see the actual **Sale Price** and values predicted.

Table

Description automatically generated

Here we can see there is some difference between the actual sale price and the predicted prices.

* Now we will check the multicollinearity in regression models through **Caret** or **rms** library by selecting **vif** function of the final model.

A screenshot of a computer

Description automatically generated with medium confidence

Here we can see the independent analysis of vectors and can verify that condition.1feedr, condition.1Norm, Exter.QualityGd and Exter.QualityTA has a multicollinearity of more than 5. So, these variables can be problematic and non-reliable.

* **To Check Outliers Through Cooks Distance**

Chart, scatter chart

Description automatically generated

With the analysis we checked outliers and found there is 1, which has a cook’s distance greater than 1.

* **Final Plot Model**

Chart, line chart

Description automatically generated

As we can see that it’s **Not** **Normally Distributed** as it has deviated from the linear line.

* **Check Durbin-Watson Test**

Text

Description automatically generated

The DW test gave a result of 1.73 which lies between 1.5 to 2.5 as its has detected true auto correlation.

**4.0 CONCLUSION**

Hence through this overall analysis we have concluded that predicting the house prices involves various uncertainties. Starting from descriptive analysis of the data, further solving the data quality issues through various functions in Rstuio, then analysis done through pairing various correlation variables and then visualization of these variables. In addition to that the creating of regression analysis and testing various models and further matching it through predictions involves a lot of precision and even a data quality issues in one variable affects the overall quality of predictions and Analysis.

These findings can help and organization in analysis of various independent variables like Full Bathroom or Basement which can help them identify the important factors involved in a price hike of the house and they can further work on strengthening these variables.

The only limitations involved here is **few data quality** issues and most of the variables are **categorical** in nature which is always hard to quantify and analyse.