# Results

|  |  |  |
| --- | --- | --- |
| KCHPM RMSE | | |
| Approaches | Training RMSE | Testing RMSE |
| Model made by selecting Attributes by doing analysis | 172774 | 168872 |
| stepAIC |  |  |
| Boruta | 139800 | 140324 |

# Conclusions

.

In our KCHPM (King County House Pricing Model) There are 21 attributes. I made a linear regression model to predict the price of the house taking account of attributes which influence it to great extent.

I used following three different approaches and made three different models to predict the house prices

* Featured selection based on basic statistical analysis:

I did the analysis on each and every attributes. I used boxplot, plot, histogram,correlation etc. I also checked the correlation between independent variable. because if they are highly correlated with each other ,they may mislead the result of the model. I removed all the outliers of if it is there in any of the attributes. Based on the analysis I selected some attributes which is best suitable for our model to predict price.

I made linear regression model with the help of all selected attributes.

Then I calculated its training and testing error.

* Boruta algorithm :

Boruta is an algorithm of finding important attributes in information systems by iterative learning of the randomForest classifier.

Boruta selected 19 attributes out of 21 to be added in the model

I made linear regression model with the help of all selected 19 attributes.

Then I calculated its training and testing error.

* stepAIC algorithm:

stepAIC Performs stepwise model selection by AIC. The set of models searched is determined by the scope argument.

stepAIC selected 20 attributes out 0f 21 for our model.

I made linear regression model with the help of all selected 20 attributes.

Then I calculated its training and testing error.

I think the model having least RMSE will be the best Model.