Comparison of Selection Techniques

* We can see in the table below that performance of forward selection and stepwise selection was the same in terms of Adjusted R Squared, but stepwise selection edged out forward selection in terms of pressed cross validation and Kaggle score.
* The backward selection model had a higher Adjusted R squared score, but lower CV press and Kaggle scores, so we went with the stepwise model.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Adjusted R Squared | CV Press | Kaggle Score |
| Forward Selection | 0.8776 | 25.7512 | 0.13944 |
| Backward Selection | 0.9361 | 50.0041 | 0.28573 |
| Stepwise Selection | 0.8776 | 25.622 | 0.13933 |
| Custom (Stepwise Selection) | 0.8776 | 25.622 | 0.13933 |

Conclusion

* Once we turned our attention to the full gamut of explanatory variables available, our ability to predict house prices improved markedly, but there become more possibilities to choose combinations of explanatory variables. Different models chose different features of importance, but there was consensus around some important features such as year, overall condition, overall condition square feet, neighborhood of a house that play an important role in predicting house price. Our best model used only 7 of the available features and had a Kaggle score of 0.13933 with and Adjusted of 0.8776.