SSC 442 Lab 3

Team Awesome (19) - Sayem Lincoln, Joshua Schwimmer, John Townshend.

Ex2 Part 3-

Task -

In a PDF write-up, describe the resulting model. Discuss how you arrived at this model, what interactions you're using (if any) and how confident you are that your group's prediction will perform well, relative to other groups.

Answer –

Our resulting model values are Train RMSE: 44072.28 and Test RMSE: 37741.19.

We received these outcomes from our model by running selected variables through our model. We selected 12 variables from the Ames dataset making them our predictor variables as we believe these variables can present us with the best correlation causation scenario when regressed against our response variable SalePrice.

The model we are using is model 12:

lm(data2$SalePrice ~ MSSubClass+LotFrontage+LotArea+YearBuilt+YearRemodAdd+MasVnrArea+BsmtFinSF1+ BsmtFinSF2+BsmtUnfSF+TotalBsmtSF+X1stFlrSF+X2ndFlrSF, data = data2).

We interacted the predictor variables against the response variable through a linear model regression function, then to get a predicted RMSE value we ran our final resulting model from Ex 1 Part 2 (the one with 15 variables and complexities) through our Test and Train function, we did this in order for the outcome to get tested and trained through our model until we received an outcome that has the lowest RMSE possible. Additionally, it can be intuitively implied that these variables play a part on a property's Sale Price, so regressing SalePrice on these variables can present us with a smaller RMSE value for both Train and Test components.

We are confident with our outputs and we believe that our model has regressed properly on the variables that we chose, and an optimal RMSE values for both Train and Test component was outputted.