What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Optimal value of alpha for ridge and lasso is 3.3 and 0.0001 respectively. On doubling alpha below are the observations on train set

Ridge	Lasso
R2 dropped from 0.9292830376274619 To 0.9220415149504599	R2 dropped from 0.9300893238186959 to 0.9243791343970081
There is small change observed with Features like 1stFlrSF and 2ndFlrSF taking prominence	There is no change in predictor features
There is shrinkage in coefficients observed	There is shrinkage in coefficients observed
The most important predictors now are • GrLivArea • OverallQual • 1stFlrSF • GarageArea • BsmtFinSF1 • 2ndFlrSF • BsmtUnfSF • LotArea • OverallCond • KitchenQual	The most important predictors now are • GrLivArea • OverallQual • BsmtFinSF1 • OverallCond • GarageArea • YearBuilt • LotArea • BsmtUnfSF • Neighborhood_NridgHt • Neighborhood_StoneBr

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

We will be choosing the lasso primarily due to following factors

1. R2 is better for Lasso. Though they are very close but the deciding factor is variables
2. The number of non-zero coefficients in lasso is 69 vs 124 in ridge. This means model is much simpler in lasso vs ridge.

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

After we drop top 5 Features GrLivArea,OverallQual,BsmtFinSF1,OverallCond,GarageArea we see a drop in R2 for new model by 1%. The new most 5 important features are

- 1stFlrSF
- 2ndFlrSF
- Neighborhood_StoneBr
- YearBuilt
- LotArea

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

To ensure that model is robust and generalizable we need to ensure that the number of features that model learns is just right. Learning influencing features is good but when model learns unwanted features which is noise is not good. It causes model to overfit. If less important features are learnt then underfit occurs causing more errors in real world or test dataset. Overfit give better performance in train set but fares poorly on test set.