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?

The optimal value of alpha for ridge and lasso regression is 3.0 and 0.0001 respectively.

Ridge	Lasso
R2 dropped from 0.926996578389455 To 0.9190614239696848	R2 dropped from 0.927183179766154 To 0.9178523439123412
There is small change observed with Features like FullBath taking prominence	There is a small change observed in predictor features like SaleCondition_Partial taking prominance
There is shrinkage in coefficients observed	There is shrinkage in coefficients observed
The most important predictors now are • GrLivArea , 1stFlrSF, OverallQual, OverallCond, BsmtFinSF1, GarageArea, 2ndFlrSF, BsmtUnfSF, LotArea, FullBath	The most important predictors now are • GrLivArea , OverallQual, OverallCond, BsmtFinSF1, GarageArea, YearBuilt, BsmtUnfSF, LotArea, Neighborhood_Crawfor, SaleCondition_Partial

Question 2
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?

- I will be choosing the lasso primarily due to following factors
  - R2 is better for Lasso. Though they are very close but the deciding factor is variables
  - The number of non-zero coefficients in lasso is 59 vs 131 in ridge. This means model is much simpler in lasso vs ridge.

Question 3
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?

- 1stFlrSF
- 2ndFlrSF
- GarageArea
- MSZoning_RL
- MSZoning_FV

Question 4
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 a model is robust and generalizable, a number of best practices can be followed:
  - Use a large and diverse training dataset: This helps to reduce overfitting and improve the model's ability to generalize to new data.
  - Use regularization techniques: Regularization helps to reduce overfitting by adding a penalty term to the loss function. Common regularization techniques are lasso and ridge.
  - Cross-validation: This involves dividing the data into several folds and training the model multiple times on different folds, with each fold used as the validation set once.
  - Bias and Variance tradeoff is highly required to minimize the error and to avoid overfitting and underfitting

The implications of having a robust and generalizable model for accuracy are two-fold:

- Improved accuracy on new and unseen data: A robust and generalizable model will perform better on new and unseen data compared to a model that is overfitted to the training data.
- Increased stability: A robust and generalizable model will be less sensitive to small changes in the data compared to a model that is overfitted to the training data.