Question 1: 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?

Answer: The optimal value of alpha for Ridge and Lasso Regression we got is below:

Lasso: {'alpha': 0.001**}**
Ridge: {'alpha': 7.0**}**

For Ridge: {Alpha = 7.0**}**

R2 Score on Train data	**0.9314**
R2 Score on Test data	**0.9134**
MSE on Train data	**0.0102**
MSE on Test data	**0.0123**

For Lasso: {Alpha = 0.001**}**

R2 Score on Train data	**0.9198**
R2 Score on Test data	**0.9130**
MSE on Train data	**0.0119**
MSE on Test data	**0.0124**

Now if we double the Alpha value for both Ridge and Lasso then we got:

For Ridge: {Alpha = 14.0**}**

R2 Score on Train data	**0.9282**
R2 Score on Test data	**0.9145**
MSE on Train data	**0.0107**
MSE on Test data	**0.0122**

For Lasso: {Alpha = 0.002}

R2 Score on Train data	**0.9118**
R2 Score on Test data	**0.9102**
MSE on Train data	**0.0132**
MSE on Test data	**0.0128**

From above table when we double the value of Alpha for both Ridge and Lasso then R2 score is decreasing slightly. The top 5 coefficient are given below, when we multiplying alpha by 2 for both Ridge and Lasso:

Feature Name	Ridge_Coeff_Double	Lasso_Coeff_Double
GrLivArea	**0.074944**	**0.128101**
OverallQual	**0.074685**	**0.093489**
SaleCondition_Partial	**0.056171**	**0.082568**
Neighborhood_Crawfor	**0.088234**	**0.070798**
OverallCond	**0.055847**	**0.056892**

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?

Answer: Ridge and Lasso are two Regularisation techniques which is used to tune the coefficients also maintain balance between bias-variance trade-off to reduce model complexity and prevent overfitting.

Ridge Regression: It performs regularisation by adding penalty equivalent to square of magnitude of Coefficients.
Ridge Regression Cost = Least Square Object + α * (sum of square of coefficients)

<u>Lasso Regression</u>: It performs regularisation by adding penalty equivalent to absolute value of magnitude of Coefficients.
Lasso Regression Cost = Least Square Object + α * (sum of absolute value of coefficients)

Now except the regularisation, Lasso regression shrinks the coefficient value towards 0, means the penalty term pushes some of insignificant coefficients to be exactly to 0 when we have large features. Lasso helps in feature selection as well along with regularisation. So, we can say that model generated by Lasso makes model simpler as compared to model generated by 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?

Answer: According to Lasso regression our top 5 features are as below:

Feature Name	Lasso_Coeff
GrLivArea	0.129
OverallQual	0.085
SaleCondition_Partial	0.112
Neighborhood_Crawfor	0.097
OverallCond	0.058

And when we remove these top 5 features and again fit the Lasso regression, then we have below top 5 features:

Feature Name	Lasso_Coeff
MSZoning_RH	0.3204
MSZoning_FV	0.3191
MSZoning_RL	0.2987
MSZoning_RM	0.2949
2ndFlrSF	0.1323

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?

Answer: The model should not be unnecessarily complex; it should be as simple as possible though it may impact the accuracy of model but it will be more robust and generalisable. Also, by using Bias-Variance trade-off we can understand it, Simpler the model more bias but less variance result in more generic. Its implication in terms of accuracy is that model will perform uniformly on both training and test data i.e., accuracy does not change much.

Simpler model having some advantage over Complex model, due to which we prefer this.

1. Simpler models are usually more generic.
2. Simpler models require fewer training samples for train the model as compare to complex one.
3. Simpler models are more robust.
4. Simpler models make more errors in the training set.

Bias: Bias is basically the correctness of the model; it occurs when our model fails to learn pattern from the data. Model with high bias pays little attention on training data. It always leads to high error on training and test data.

Variance: Variance is the inconsistency of model prediction for given data points. Model with high variance pays a lot of attention to training data and does not generalise on the unseen data. So as a result, such models perform very well on training data but has high error rates on test data.

So, it is very important to keep balance between Bias and Variance to avoid model complexity and overfitting.

We should try to model have low bias and low variance or we can say find an optimal point where total error is low.