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 was found to be **100**.
If we double the value of alpha in Ridge and Lasso Regression, the model will be penalized more heavily and thus will result in the shrinkage of the coefficients of the features.
In Lasso Regression, it is likely that features will be eliminated as a result of doubling the alpha, however, it also depends on the dataset at hand.

Top 5 features in Ridge Regression:

Alpha = 100		Alpha = 200	
Feature	**Coef**	**Feature**	**Coef**
TotRmsAbvGrd	**12.891**	**TotRmsAbvGrd**	**11.62**
OverallQualCondMean	**11.122**	**OverallQualCondMean**	**10.75**
FullBath	**9.213**	**FullBath**	**8.48**
GarageArea	**8.184**	**GarageArea**	**8.05**
Neighborhood_NridgHt	**8.12**	**TotalBsmtSF**	**7.86**

Top 5 features in Lasso Regression:

Alpha = 1		Alpha = 2	
Feature	**Coef**	**Feature**	**Coef**
TotRmsAbvGrd	**13.55**	**TotRmsAbvGrd**	**12.89**
OverallQualCondMean	**11.76**	**OverallQualCondMean**	**12.14**
Neighborhood_NridgHt	**10.25**	**KitchenQual**	**8.97**
FullBath	**9.11**	**Neighborhood_NridgHt**	**8.93**
Neighborhood_NoRidge	**8.66**	**TotalBsmtSF**	**8.58**

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: After performing regularization, the performance of both Ridge and Lasso is fairly similar when compared on R2 Scores and RMSE for both train and test data. However, in Lasso many of the feature coefficients are reduced to zero, thereby performing feature selection and therefore Lasso is selected as an appropriate choice to predict the SalePrice.
More specifically, in Ridge Regression Model the final features are **87** whereas in Lasso the final features are **70**, both having very similar scores of the error metrics. Therefore, it is more appropriate to select the Lasso Regression Model.

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: The top 5 important predictor variables after removing the top 5 predictors found initially are:

Feature	**Coef**
KitchenQual	**15.34**
BsmtQualCondMean	**7.53**
BsmtExposure	**6.32**
Functional	**5.70**
Condition1_Norm	**4.66**

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: To make sure that a model is robust and generalisable we can take care of the following things:

- **Data Quality:** We must ensure that we have perform extensive EDA on the data to remove or address any inconsistencies in the data which can come in various forms such as incorrect or extreme values, improper format. A high quality model with bad data will result in bad predictions.

- **Bias v/s Variance Tradeoff:** We have to choose the appropriate complexity of the model so that the model generalizes well on the unseen data. An overly complex model is likely to perform poorly on test data whereas an overly simple model will not learn the patterns significantly enough to provide acceptable predictions. Therefore, it is imperative that we address any overfitting or underfitting in the model (if any).

- **Hyperparameter Tuning:** It is necessary that we perform the hyperparameter tuning so that we find the appropriate hyperparameter (alpha in case of regularized linear regression), so that we obtain the best predictions from the model (model with the best error metrics).

- **Cross Validation:** Cross Validation helps us to evaluate the model's performance on different subsets of data, this technique is essential because it helps us to obtain less optimistic and more appropriate estimate of the model's performance.