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?

Choosing double the value of alpha for Ridge and Lasso regression will generally increase the amount of regularization applied to the models. Here's how the models and the most important predictor variables may be affected
Ridge Regression:

- Increasing alpha will strengthen the regularization effect in Ridge regression
- The overall magnitude of the coefficients will decrease
- The most important predictor variables will depend on the original dataset and the relationships between the predictors and the target variable
- The training R2 (coefficient of determination) is 0.9109, indicating that the linear regression model explains approximately 91.1% of the variance in the training data.
- The testing R2 is 0.4885, suggesting that the linear regression model performs less well on unseen or testing data, explaining only around 48.9% of the variance.

Lasso Regression:

- Doubling the value of alpha in Lasso regression will increase the strength of regularization even more
- More coefficients will be pushed towards zero, and some may become exactly zero
- The training R2 is 0.9001, indicating that the lasso regression model explains approximately 90.0% of the variance in the training data.
- The testing R2 is 0.6753, suggesting that the lasso regression model performs better on unseen or testing data compared to linear regression, explaining around 67.5% of the variance.

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?

Choosing between Ridge and Lasso regression depends on the specific requirements and characteristics of the problem at hand. However, based on the information provided, it seems that Lasso regression may be a better choice.

Improved Testing R2: In the provided regression comparison, the Lasso regression model has a higher testing R2 (0.6753) compared to the linear regression model (0.4885). This indicates that the Lasso model has better generalization performance and better at predicting unseen data.

Feature Selection: Lasso regression has the advantage of performing feature selection by shrinking the coefficients of less important variables to zero.

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?

The below given variables have higher coefficients indicating that they have the most significant impact on the predicted SalePrice. If these five variables are not available in the incoming data, they cannot be used as predictors in the new model that excludes them.

GrLivArea
OverallQual
2ndFlrSF
TotalBsmtSF
BsmtFinSF1

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?

Cross Validation
Train/Test Data Split
Feature Selection
Regularization
Imputation
Outlier Handling