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?

Optimal Values

- Ridge optimal α = 184.21
- Lasso optimal α = 0.001

Doubling Alpha

- Ridge: When α is doubled, the average coefficient magnitude decreases, confirming stronger shrinkage.
- Lasso: Reduces the number of active features. Produces a simpler model with fewer predictors at the same time, test R^2 increases, meaning the model generalises better after removing more weak/noisy features.

Most Important Predictor:

- Ridge: GrLivArea, TotalBsmtSF, 1stFlrSF, BsmtFinSF1, OverallQual_Very Good, YearBuilt, OverallQual_Excellent, 2ndFlrSF, OverallCond_Fair, LotArea
- Lasso: GrLivArea, YearBuilt, TotalBsmtSF, OverallQual_Very Good, OverallQual_Excellent, SaleCondition_Normal, LotArea, Neighborhood_Crawfor, BsmtFinSF1, SaleType_New

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?

Model Preference: Lasso Regression from the results.
Reason: As Lasso gives the highest test R^2 and the lowest RMSE, which means it predicts house prices more accurately on unseen data.

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?

Initial Top 5 predictor:
GrLivArea, YearBuilt, TotalBsmtSF, OverallQual_Very Good, OverallQual_Excellent

New Top 5 predictor: (after old 5 removal):
1stFlrSF, 2ndFlrSF, BsmtFinSF1, BsmtUnfSF, YearRemodAdd

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?

A model can be made robust and generalisable by using techniques cross-validation, regularisation (Ridge/Lasso), and proper hyperparameter tuning. These methods prevent overfitting and ensure the model performs consistently on unseen data.

The implication is that training accuracy may reduce slightly due to added bias, but test accuracy improves. This bias - variance trade-off results in a more reliable model that generalises better to new data.