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?*

Alpha for Ridge Regression: 8
Alpha for Lasso Regression: 0.0001

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.077704e-01	8.756955e-01	8.778517e-01
1	R2 Score (Test)	-7.058080e+22	7.974415e-01	8.081834e-01
2	RSS (Train)	5.745834e+11	7.744074e+11	7.609746e+11
3	RSS (Test)	1.542938e+35	4.428048e+11	4.193225e+11
4	MSE (Train)	2.467122e+04	2.864170e+04	2.839220e+04
5	MSE (Test)	1.951851e+16	3.306577e+04	3.217707e+04

Alpha for Ridge Regression: 16
Alpha for Lasso Regression: 0.0002

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.077704e-01	8.574591e-01	8.633980e-01
1	R2 Score (Test)	-7.058080e+22	7.985206e-01	8.183197e-01
2	RSS (Train)	5.745834e+11	8.880193e+11	8.510199e+11
3	RSS (Test)	1.542938e+35	4.404460e+11	3.971640e+11
4	MSE (Train)	2.467122e+04	3.067081e+04	3.002506e+04
5	MSE (Test)	1.951851e+16	3.297758e+04	3.131536e+04

Top 5 Features - Lasso:

Parame	Linear	Ridge	Lasso
GrLivArea	2.14E-02	0.055178	0.231142
OverallQua	1.11E-01	0.089148	0.143089
Neighborh	1.21E-01	0.063375	0.084769
GarageAre	7.37E-02	0.05614	0.063732
Neighborh	8.17E-02	0.043043	0.059043

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 using Lasso regression. There is a better match between Test and train Data. The top reasons for selecting reasons are:

- There are a large number of features, and only a few will be relevant
- Perform feature selection and identify the most relevant variables.
- We can sacrifice some bias for the sake of a simpler and more interpretable 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?*

1. **Train a New Lasso Model:** Exclude the five variables that were identified earlier and retrain the model

2. **Retrieve and Sort Coefficients:** Extract the coefficients from the newly trained Lasso model. These coefficients indicate the importance of each predictor variable in the new model.
3. **Select the Top Five Variables:** Choose the five variables with the largest absolute coefficients in the new Lasso model

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 a model is robust and generalizable:

a. **Data Splitting:** Use train-test splits and cross-validation to assess performance on unseen data.
b. **Feature Engineering:** Selecting only relevant features and scaling them
c. **Regularization:** Using Ridge or Lasso Regression
d. **Hyperparameter Tuning:** Optimize model hyperparameters.
e. **Model Complexity:** Choosing a model that's neither too easy or complicated
f. **Managing the Bias-Variance Trade-off:** Achieve a balance between bias and variance to enhance generalization
g. **Assessing with Varied Test Data:** Test the model's performance using datasets that encompass diverse scenarios
h. **Ongoing Oversight and Refinement:** Continuously oversee and refine the model during its deployment in production.

Implications for Accuracy:

Prioritizing robustness and generalization may lead to reduced training data accuracy, but it improves the model's real-world performance by reducing overfitting and ensuring adaptability to new data patterns.