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 alpha value are as listed below:

Scalar	StandardScaler	MinMax
ridge	600	0.09
Lasso	0.000005	0.000005

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: The optimal lambda value is 0.000005. I experimented with various values of polynomial degrees, alpha(lambda) values and combination of Linear regression models and found out that Lasso Regression (Regularization technique) with MinMaxScaler seems to be the best suit in above case(with said degrees and alpha value) as it is giving training R-Squared value of 87 and test R-Squared value of 86.

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: Original 5 most important predictor variables as per co-efficient values are as below. Note that I have used RFE and retain top 20 feature hence my list may vary from others who might have retained all the original columns

Feature_Name(Top 5)	Importance(Descending)
RoofStyle	0.032583
Fireplaces	0.018333
PavedDrive	0.014049
FullBath	0.011978
HalfBath	0.011315

The new top 5 features after dropping above 5 are

Feature_Name(Top 5)	Importance(Descending)
RoofMatl	0.036873
GarageCars	0.024683
KitchenAbvGr	0.010967
ExterQual	0.010634
BsmtFullBath	0.007858

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 condition for satisfactory model are as follows-

Both R2 scores for Train and Test should be high.

Difference between train and test performance should be low.