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?

Ans:
The optimal value of alpha comes out to be:

- Ridge:
  - Alpha: 2.5
  - Training score: 92.13%
  - Test score: 90.13%
- Lasso:
  - Alpha: 0.0003
  - Training score: 92.03%
  - Test score: 89.94%

To achieve these values of alphas, for the respective model, we followed the process of GridSearchCV. We provided various values for alpha, and trained models on the. The model which performed the best was selected.

Doubling the value of alpha:
Ridge:

- Alpha: 5
- Training score: 91.85
- Test score: 90.07

Lasso:

- Alpha: 0.0006
- Training score: 91.39
- Test score: 90.16

As observed from the above results, the training and test score have not changed much after doubling the value of alpha. But the top 5 predictors have changed for both the model, as shown below:

Top predictors (Ridge)	
Alpha = 2.5	Alpha = 5
OverallQual_Excellent (+)	OverallQual_Excellent (+)
OverallCond_Excellent (+)	OverallCond_Fair (-)
OverallCond_Fair (-)	GrLivArea (+)

Exterior1st_BrkComm (-)	OverallCond_Excellent (+)
MSZoning_FV (+)	OverallCond_Very_Good (+)

Top predictors (Lasso)	
Alpha = 0.0003	Alpha = 0.0006
Exterior1st_BrkComm (-)	OverallQual_Excellent (+)
OverallQual_Poor (-)	OverallCond_Fair (-)
OverallQual_Excellent (+)	OverallCond_Excellent (+)
OverallQual_Very_Excellent (+)	GrLivArea (+)
OverallCond_Excellent (+)	OverallQual_Very_Excellent (+)

This suggests that the models have adjusted the values of coefficients for its predictors to compensate with the change in value of alpha. Most of the predictors (4 out 5) now have a positive coefficient.

The top predictor in both of the models is the same:- OverallQual_Excellent with coefficient 0.150875 and 0.186742 for Ridge and Lasso respectively.

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?

Ans:
During modelling for RIdge and Lasso we observed that the training and test scores for both these models was similar:

Ridge:

Training score: 92.13%
Test score: 90.13%

Lasso:

Training score: 92.03%
Test score: 89.94%

Since, both of the models perform equally well, we can choose any one of these. Now, Lasso has a property of feature selection. It makes the coefficient of predictors 0 for which it realises is redundant, whereas, Ridge, reduces the value of coefficient but does not make them 0. Hence, a model made by Lasso would be simpler as compared to the RIdge model.
As calculated in the notebook shared along, Ridge has coefficient value for all 70 columns selected in RFE. Whereas, Lasso has only 50 coefficients for which coefficient is not 0.

So, choosing the Lasso model makes more sense, since it is simpler and more general as compared to the Ridge 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?

Ans:
The most important variables and performance of the initial Lasso model are as follows:

Predictors:

1. Exterior1st_BrkComm (-)
2. OverallQual_Poor (-)
3. OverallQual_Excellent (+)
4. OverallQual_Very_Excellent (+)
5. OverallCond_Excellent (+)

Training score: 92.03%

Test score: 89.94%

After removing the top 5 predictors from the Lasso model, and training again. Below are the results:

Predictors:

1. ExterQual_Fa (-)
2. OverallCond_Fair (-)
3. Neighborhood_StoneBr (+)
4. GrLivArea (+)
5. ExterQual_TA (+)

Training score: 91.11%

Test score: 89.71%

We observe that there is no significant change in the model performance, the model still behaves good on both, train set and test set

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?

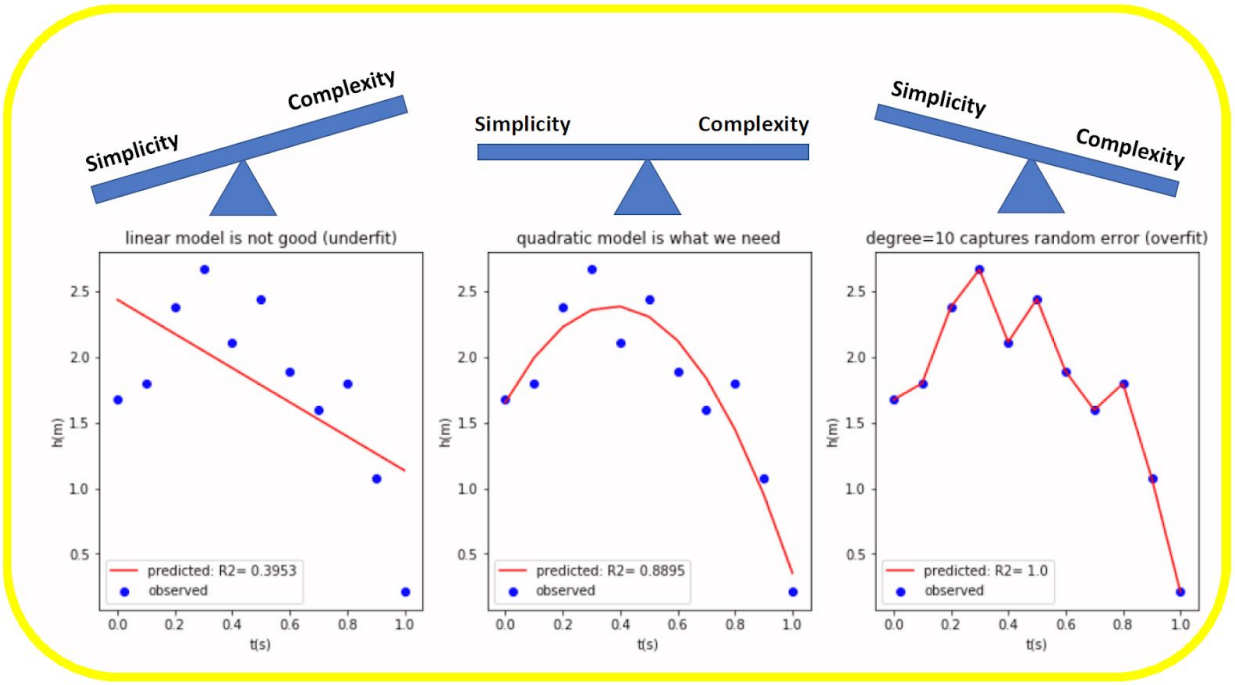
Answer:
A few highly contaminated predictors, i.e., predictors having minimum information and many outliers can cause the model to learn incorrectly and perform poorly since the Lasso model is based on Least-squares estimator. But during our process of model building, we took care of outliers in the independent variables, hence, the model is stable.

Also, since we did not drop/cap outliers in the dependent variable and rather transformed it, the model should work well in surprisingly high valued houses as well.

Since we used RFE and Lasso to build our model, both of the techniques are known to perform well in feature selection. In RFE, we selected 70 features and after training the Lasso model, we found that the learning algorithm had made coefficients for 20 variables as 0. Since our model performed well on training as well as test data with 92.03% and 89.94% accuracy, we can say that our model is general as compared to the Ridge model which uses all 70 predictors to achieve similar performance.

The accuracy of the model is comparable with the Ridge model, i.e., there is no significant hit on accuracy. This is because the Ridge model reduces the value of coefficients for redundant predictors so that their implication on prediction is negligible and Lasso makes the coefficient 0 for redundant variables. Hence, there is not much difference in accuracy. But the lasso model is simpler and more general as compared to the Ridge model because it uses less number of predictors for achieving the same performance.

Another implication of making a model simpler is that it can cause underfitting because the model fails to learn some feature that might impact the prediction. So, there's a tradeoff between simplicity and complexity of the model (see image below). But in our case, since our model is performing well in both training and test data, we can say that the model is accurate.


Simplicity
Complexity
linear model is not good (underfit)
predicted: R2= 0.3953
observed
h(m)
t(s)
Simplicity
Complexity
quadratic model is what we need
predicted: R2= 0.8895
observed
h(m)
t(s)
Simplicity
Complexity
degree=10 captures random error (overfit)
predicted: R2= 1.0
observed
h(m)
t(s)