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 for ridge is 5 and for lasso regression is 0.0005**.** With increase in λ, the model becomes progressively simpler as the coefficients are pushed down towards zero in ridge and some of the coefficients become zero in lasso.
The most important predictor variables are:

	Features
11	MSZoning_RL
5	GrLivArea
12	MSZoning_RM
1	OverallQual
9	MSZoning_FV
2	OverallCond
4	TotalBsmtSF
14	Foundation_PConc
7	GarageCars
3	BsmtFinSF1

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: The optimal lambda value in case of Ridge and Lasso is as below:

- Ridge - 5
- Lasso - 0.0005

The Mean Squared error in case of Ridge and Lasso are:

- Ridge - 0.0136779
- Lasso - 0.0135336

The Mean Squared Error of Lasso is slightly lower than that of Ridge

Also, since Lasso helps in feature reduction (as the coefficient value of one of the feature became 0), Lasso has a better edge over Ridge.

Hence based on Lasso, the factors that generally affect the price are the Zoning classification, Living area square feet, Overall quality and condition of the house, Foundation type of the house, Number of cars that can be accomodated in the garage, Total basement area in square feet and the Basement finished square feet area

Therefore, the variables predicted by Lasso in the above bar chart as significant variables for predicting the price of a house.

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 five most important predictor variables are :

Features
MSZoning_RL
GrLivArea
MSZoning_RM
OverallQual
MSZoning_FV

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?**

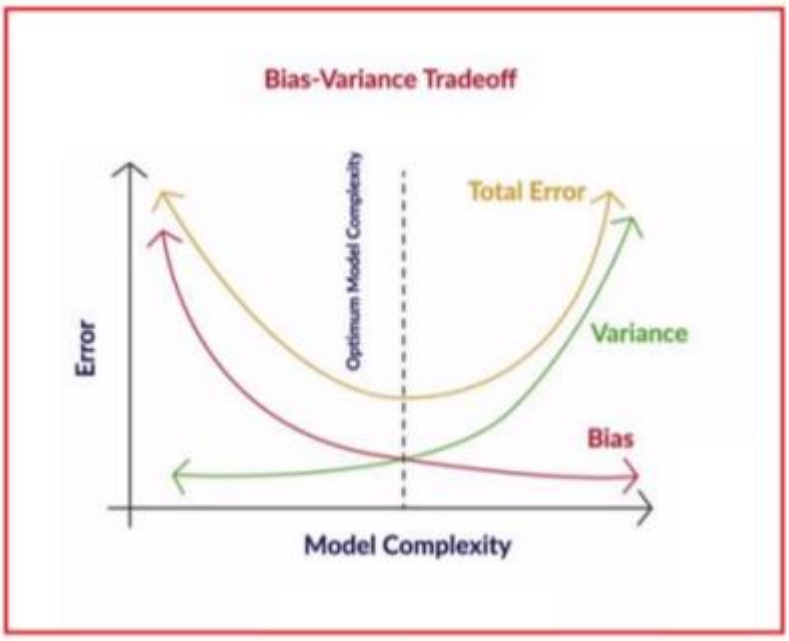
Ans: Per, Occam's Razor – given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to the following reasons :-

- Simpler models are usually more generic and are more widely applicable
- Simpler models require fewer samples for effective training than the more complex ones and hence easy to train the data
- Simpler models are more robust
    - Complex models tend to change widely with the changes in the training data set
    - Simple models have low variance , high bias and complex models have high variance and low bias
- Simpler models make more errors in the training set. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples

Therefore to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use.
Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.
Also, Making a model simple leads to Bias-Variance Trade-off:

- A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed. Bias quantifies how accurate is the model likely to be on test data.

A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.
Variance refers to the degree of changes in the model itself with respect to changes in the training data.
Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph


Bias-Variance Tradeoff
Optimum Model Complexity
Total Error
Variance
Bias
Error
Model Complexity