SUBJECTIVE QUESTIONS:

1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

- The optimal value of alpha for ridge is 6.0 and for lasso is 0.001.
- When we choose to double the alpha value,
    - R-Squared value for both lasso and ridge model is reduced in training set.
    - MSE has reduced in ridge model and R-Squared test score.
- Most important predictor variables are,
    - GrLivArea
    - OverallQual_Excellent
    - Functional_Typ
    - Neighborhood_Crawfor
    - OverallCond_Excellent
    - Neighborhood_StoneBr
    - Exterior1st_BrkFace
    - OverallQual_Very_Good
    - SalesCondition_Alloca
    - RoofMatl_WdShngl

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?

Lasso Model is chosen to be applied and the reason is,

- The difference between the R-Square for train and test is least for lasso comparatively.
- Comparatively, the R-Squared score is better and the Mean Square Error (MSE) is least for lasso model which a sign of a good model.

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?

- The model is re-trained by removing the top five feature that we have got earlier. As we already know the lasso model performs better in this use case, we re-train the model with alpha = 0.001.
- After re-training the model, the top predictors are,
    - RoofMatl_WdShngl
    - Exterior1st_BrkFace
    - 2ndFlrSF
    - Neighborhood_StoneBr
    - SalesCondition_Alloca

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 make sure the model is robust and generalizable, it has to be a simple model which has low variance and high bias and they have to be generic in nature.
- To have a good accuracy for the above discussed simple model, the bias-variance trade off has to be considered as key factor since the simple model can also be defined by high bias and lower variance.
- So, it is a key factor to have bias-variance trade off for accuracy of the model.