Assignment-based 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 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 value of alpha for ridge regression is 500 and for lasso regression is 0.01

After doubling the alpha for lasso reg, the r2_score dropped by 0.015 and RMSE increased by 0.002

After doubling the alpha for ridge reg, the r2_score dropped by 0.015 and RMSE increased by 0.002

Top 10 most important predictors as per Lasso after implementing the change:

```
'OverallQual',
 'GrLivArea',
 'GarageCars',
 'Fireplaces',
 'BsmtFullBath',
 'TotalBsmtSF',
 'CentralAir',
 'MSZoning',
 '1stFlrSF',
 'OverallCond'
```

Top 10 most important predictors as per Ridge after implementing the change:

```
'OverallQual',
 'GrLivArea',
 'GarageCars',
 'Fireplaces',
 '1stFlrSF',
 'GarageArea',
 'TotRmsAbvGrd',
 'FullBath',
 'TotalBsmtSF',
 'CentralAir'
```

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: Both ridge and lasso reg model are giving similar results but I will choose lasso regression model as lasso helped to shrink the beta coefficients to zero for some predictors i.e. it removed some of the features and make the model simpler and interpretable.

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: Fireplaces, CentralAir, TotalBsmtSF, MSZoning, LotArea will be the next five most important predictor variables.

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: Our model is not overfitting and the model is simple and interpretable which means the model is robust and generalisable. From model evaluation, we can see r2_score of both train and test data are almost nearly equal (r2_train_lasso: 0.869, r2_test_lasso: 0.865). So the model has enough accuracy maintaining the bias variance tradeoff.