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 is:

Ridge regression: 100
Lasso regression: 0.01

If we double the alpha:

```
            Most Important Features in Adjusted Ridge Model:
                  Feature  Coefficient
15              GrLivArea     0.055879
3             OverallQual     0.050410
12               1stFlrSF     0.040389
4             OverallCond     0.030429
2                 LotArea     0.028082
236             PoolQC_Gd    -0.024550
5               YearBuilt     0.023204
25             GarageCars     0.022231
71   Neighborhood_NridgHt     0.021403
165           BsmtQual_TA    -0.020293
```

```
Most Important Features in Adjusted Lasso Model:
          Feature  Coefficient
3     OverallQual     0.120612
15      GrLivArea     0.102899
25     GarageCars     0.047058
5       YearBuilt     0.042119
2         LotArea     0.036247
6    YearRemodAdd     0.027632
12       1stFlrSF     0.025903
8      BsmtFinSF1     0.023001
195  CentralAir_Y     0.015275
4     OverallCond     0.014255
```

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: We will be using lasso regression because it performed better than Ridge in both RMSE and Rsquare.
Ridge Regression RMSE: 0.15
Ridge Regression R^2: 0.88
Lasso Regression RMSE: 0.14
Lasso Regression R^2: 0.89

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:
Earlier:
GrLivArea: 0.117245
OverallQual: 0.100235
YearBuilt: 0.0494655
GarageCars: 0.0405937
LotArea: 0.0390393
If above variables dont come then:
OverallCond: 0.0310488
1stFlrSF: 0.0269366
BsmtFinSF1: 0.0236197
YearRemodAdd: 0.0222982
BsmtFullBath: 0.0161822

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?
Ans:

1. Data processing: we should process the data in best way possible to make sure it makes sense and match real world
2. Data scaling: We should make sure the to scale data and remove outliers
3. We should use validations like RSME and K-fold
4. Use regularization which increase the bais to reduce variance
5. Use good feature selection: we should properly select the features that matter to reduce the noice
6. Monitoring: We should monitor the performance and make sure it improves with time