# Problem Statement PART 2

**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?

**Answer:**

Optimal value of **alpha** is **0.001** and **100.0** for **lasso** and **ridge** respectively.

```
Optimum alpha for ridge is 100.000000
ridge  Regression with  100
==================================
R2 score (train) :  0.9020609864160686
R2 score (test) :  0.8591688998758671
RMSE (train) :  0.1252687639453735
RMSE (test) :  0.14862291951821605


Optimum alpha for lasso is 0.001000
lasso  Regression with  0.001
==================================
R2 score (train) :  0.9031009713684458
R2 score (test) :  0.8618015366954971
RMSE (train) :  0.12460189321652129
RMSE (test) :  0.14722721919726536
```

If we change the alpha value for both ridge and lasso to double.

For Ridge (alpha=100) to (alpha=200)

R2 Score (variance on test data) which was 85.9% which can be explained by Model changed to 85.7% and RMSE (prediction made by model off by RMSE units) of test data changed from 0.1486 to 0.1497

```
Model Evaluation : Ridge Regression, alpha=200.0
R2 score (train) :  0.8995
R2 score (test) :  0.8571
RMSE (train) :  0.1269
RMSE (test) :  0.1497
```

For Lasso (alpha=0.001) to (alpha=0.002)

R2 Score (variance on test data) which was 86.1% which can be explained by Model changed to 86.3% and RMSE (prediction made by model off by RMSE units) of test data changed from 0.1472 to 0.1466

```
Model Evaluation : Lasso Regression, alpha=0.002
R2 score (train) :  0.9018
R2 score (test) :  0.863
RMSE (train) :  0.1254
RMSE (test) :  0.1466
```

**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?

**Answer:**

Lasso produced slightly high R2 Score than Ridge (86.1>85.9). Hence Choose Lasso as final 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?

**Answer:**

```
# Top 5 featues in Lasso final model

model_coefficients.sort_values(by='Lasso (alpha=0.0001)', ascending=False).head(5)
```

| | Ridge (alpha=9.0) | Lasso (alpha=0.0001) | Ridge (alpha = 18.0) | Lasso (alpha = 0.0002) |
|---|---|---|---|---|
| OverallQual | 0.082741 | 0.098495 | 0.092987 | 0.096424 |
| 1stFlrSF | 0.066591 | 0.074911 | 0.075069 | 0.077349 |
| LotArea | 0.041865 | 0.044386 | 0.044270 | 0.044772 |
| OverallCond | 0.039633 | 0.042852 | 0.043386 | 0.043935 |
| FullBath | 0.035606 | 0.037195 | 0.036807 | 0.037103 |

5 Most Important features are as fallows in final Model:

[ OverallQuall, 1stFlrSF, LotArea, OverallCond, FullBath]

| | Lasso |
|---|---|
| TotRmsAbvGrd | 0.087366 |
| GarageArea | 0.072636 |
| Fireplaces | 0.053252 |
| KitchenQual | 0.051756 |
| BsmtQual | 0.048789 |

after removing above features we will be getting features as follows:

[TotRmsAbvGrd, GarageArea, FirePlaces, KitcenQual, BsmtQual]

**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:**

By removing outliers in the data set and getting majority of the data without eliminating the majority of rows by making transformation to fit into the model. Trying to implement model with good data(train data) we get best output model and best result on test data by the model.