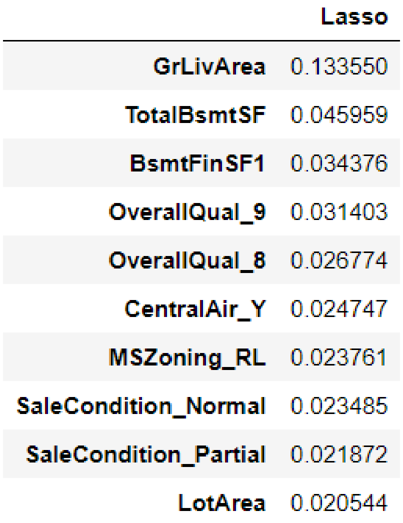
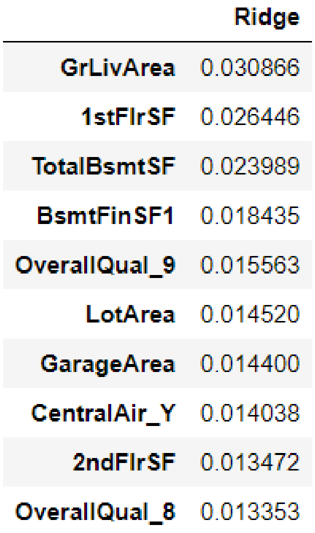
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

The optimal value of alpha for lasso regression is 0.001 and for Ridge regression is 500.If the alpha value is doubled in both the Lasso and regression models then below are the most important features:

Lasso **Ridge**



	Lasso
GrLivArea	0.133550
TotalBsmtSF	0.045959
BsmtFinSF1	0.034376
OverallQual_9	0.031403
OverallQual_8	0.026774
CentralAir_Y	0.024747
MSZoning_RL	0.023761
SaleCondition_Normal	0.023485
SaleCondition_Partial	0.021872
LotArea	0.020544



	Ridge
GrLivArea	0.030866
1stFlrSF	0.026446
TotalBsmtSF	0.023989
BsmtFinSF1	0.018435
OverallQual_9	0.015563
LotArea	0.014520
GarageArea	0.014400
CentralAir_Y	0.014038
2ndFlrSF	0.013472
OverallQual_8	0.013353

Below are the metrics of Lasso when Alpha is 0.001.

```
r2 value on training dataset using lasso Regression :0.9553046427897006
r2 value on test dataset using lasso Regression :0.8519191099529513
rss value on train dataset using lasso Regression :7.144577736823862
rss value on test dataset using lasso Regression :9.657120989988545
mse value on train dataset using lasso Regression :0.007130317102618625
mse value on test dataset using lasso Regression :0.022458420906950105
```

Below are the metrics of Lasso when Alpha is 0.002. the value of r2 is decreased in both test and training dataset for the model

```
r2 value on training dataset using lasso Regression :0.965077910001262
r2 value on test dataset using lasso Regression :0.8432507406333101
rss value on train dataset using lasso Regression :5.582315531216929
rss value on test dataset using lasso Regression :10.222430202264913
mse value on train dataset using lasso Regression :0.005571173184847234
mse value on test dataset using lasso Regression :0.02377309349363933
```

Below are the metrics of Ridge when Alpha is 500.

```
r2 value on training dataset using Ridge Regression :0.9482706289629425
r2 value on test dataset using Ridge Regression :0.8612665004550237
rss value on train dataset using Ridge Regression :8.26896876363027
rss value on test dataset using Ridge Regression :9.04752929324427
mse value on train dataset using Ridge Regression :0.008252463835958353
mse value on test dataset using Ridge Regression :0.02104076579824249
```

Below are the metrics of Ridge when Alpha is 1000.decrease in the R2 value in both datasets and other error values have been increased.

```
r2 value on training dataset using Ridge Regression :0.9326080097352037
r2 value on test dataset using Ridge Regression :0.8610796827602023
rss value on train dataset using Ridge Regression :10.772647168264761
rss value on test dataset using Ridge Regression :9.059712641692458
mse value on train dataset using Ridge Regression :0.010751144878507746
mse value on test dataset using Ridge Regression :0.021069099166726647
```

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 Regression is suitable for this model as the model provides better r2 value and other metric values also the model has many features and most of them are poorly corelated. lasso regression will help the feature selection to have efficient model.

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?

1stFlrSF, 2ndFlrSF, GarageArea, CentralAir_Y, OverallQual_9 are the five most important predictor variables when Lasso regression model has been re-built with missing features.

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 ensure the model is robust and generalisable for any dataset regularisation is performed. Regularisation helps in reducing the error terms and ensures the model is fitted appropriately and prevents overfitting and makes the model simpler.

Regularisation helps in tuning the regression function by adding the penalty term(hyperparameter) to the function to prevent the function turning into a complex one.

If the value of hyperparameter increases, the error term in the model constantly increases and the accuracy of the model decreases, model becomes too simpler and falling into the risk of underfitting.

If the value of hyperparameter decreases, the error term will decrease with increases the accuracy value, but the model becomes too complex and can lead to overfitting the data.