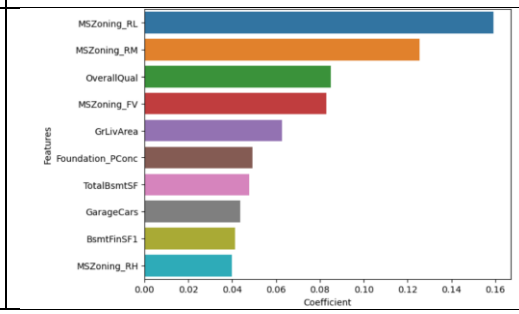
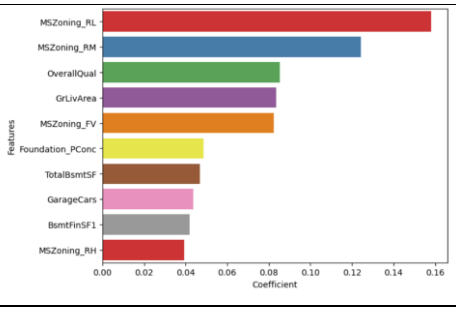
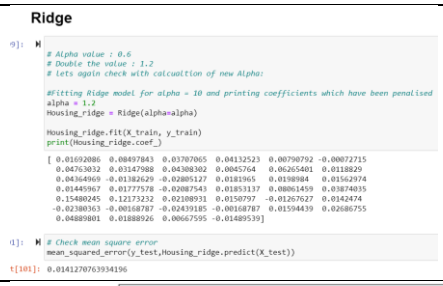
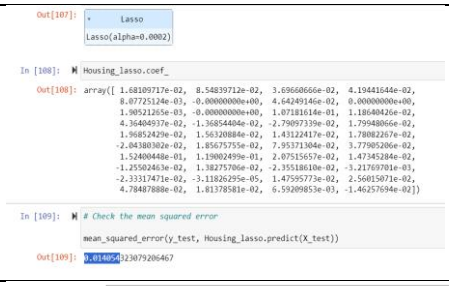
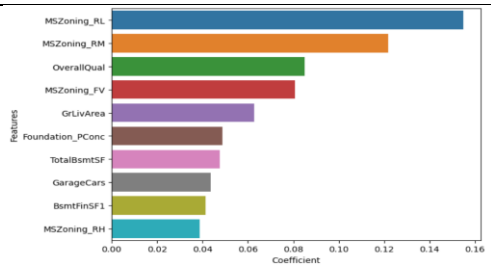
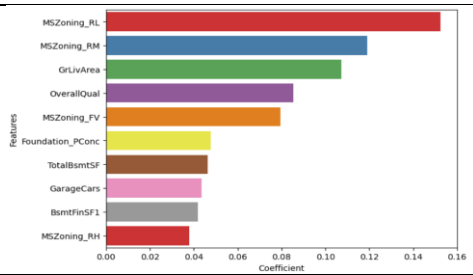
Question 1

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 : Please check below table for the observation as per question asked:

Observations	Ridge	Lasso
Optimal value of Alpha	0.6	0.0001
Mean square error	0.01415	0.01411
Bar Charts		
Double the value of Alpha – Refer Notebook for details		
Alpha	0.6 * 2 = 1.2	0.0001 * 2 = 0.0002
Mean square error	0.01412	0.01405
Evidence		
Bar Charts of important variable		
Effect	Coefficient of some variables are changed	Coefficient of some variables are changed

Conclusion : On changing alpha value, predictors are same but coefficient of some variable is changed.

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:

The Mean Squared error in case of Ridge and Lasso are:

Ridge - 0.01415

Lasso - 0.01411

1) The Mean Squared Error of Ridge is higher then Lasso.
2) Also, in case of Lasso some features get reduced (due to some of coefficients of feature became 0)
3) R2 score of Lasso is greater that Ridge on test data set

Hence we will choose Lasso.

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 : Looking at below coefficients below are 5 most important variables:-> Refer Jupitor notebook for deatils

1. **GrLivArea**
2. **MSZoning**
3. **GarageCars**
4. **Foundation**
5. **TotalBsmtSF**

	Lasso_Housing
TotalBsmtSF	0.074108
2ndFlrSF	0.013773
GrLivArea	0.113868
HalfBath	0.013784
GarageCars	0.050693
IsRemodelled	-0.013010
Built_Remodel_Age	-0.045123
b_BsmtExposure	0.030866
b_HeatingQC	0.022040
b_KitchenQual	0.038370

b_FireplaceQu	0.025710
b_GarageFinish	0.025754
b_BldgType	-0.028729
b_SaleCondition	0.021580
MSZoning_FV	0.108356
MSZoning_RH	0.052498
MSZoning_RL	0.200750
MSZoning_RM	0.162973
Neighborhood_Crawfor	0.031414
Neighborhood_NridgHt	0.020373
Neighborhood_OldTown	-0.011062
Neighborhood_Somerst	0.012928
Exterior1st_BrkComm	-0.028683
Exterior1st_CBlock	-0.004854
Exterior1st_Wd Sdng	-0.023429
Exterior2nd_Wd Sdng	0.018424
Foundation_CBlock	0.030769
Foundation_PConc	0.041584
Foundation_Slab	0.016054
GarageType_Attchd	0.019582
SaleType_WD	-0.010501

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 th e model and why?

Answer:

To have robust and generalised model, the test accuracy should not be lesser than the training score. The model should give accurate result for datasets other than the training dataset. Outliers should be avoided to have better accuracy. To get thi s addressed, outliers should be analysed and data which is rel evant for modelling should only be kept. Unnecessary outliers should be removed from dataset. If the model is not robust, pr edictive analysis done by this can not be trusted.