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:

IN CASE OF RIDGE REGRESSION: IF THE ALPHA VALUE IS HIGH IT LEAD TO UNDERFITTING AND IF THE ALPHA VALUE IS TO LOW IT LEADS TO OVERFITTING SO WE SHOOULD FIND THE OPTIMAL VALUE. RIDGE REGRESSION ALSO REDUCES THE MAGNITUDE OF COEFFICIENTS

IN CASE OF LASSO REGRESSION: FOR LASSO REGRESSION THE ALPHA VALUE IS 1. THE OUTPUT IS THE BEST CROSS-VALIDATED LAMBDA, WHICH COMES OUT TO BE 0.001. ONCE WE HAVE THE OPTIMAL VALUE OF LAMBDA. WE TRAIN THE LASSO MODEL.

THE VALUE OF ALPHA IN LASSO REGRESSION IS A HYPERPARAMETER THAT DETERMINES THE STRENGTH OF THE REGULARIZATION TERM IN THE MODEL. A HIGHER VALUE OF ALPHA MEANS A STRONGER REGULARIZATION TERM, WHICH CAN LEAD TO MORE PARAMETERS BEING SET TO ZERO AND A SIMPLER MODEL. A LOWER VALUE OF ALPHA MEANS A WEAKER REGULARIZATION TERM, WHICH CAN ALLOW MORE PARAMETERS TO BE NON-ZERO AND RESULT IN A MORE COMPLEX MODEL.

THERE IS NO "OPTIMAL" VALUE OF ALPHA THAT WORKS BEST IN ALL CASES, AND THE APPROPRIATE VALUE OF ALPHA WILL DEPEND ON THE SPECIFIC CHARACTERISTICS OF YOUR DATASET AND THE GOALS OF YOUR MODELING. IN GENERAL, IT IS RECOMMENDED TO USE CROSS-VALIDATION TO TUNE THE VALUE OF ALPHA AND FIND THE VALUE THAT LEADS TO THE BEST MODEL PERFORMANCE ON YOUR DATASET

IT IS ALSO IMPORTANT TO NOTE THAT LASSO REGRESSION IS SENSITIVE TO THE SCALE OF THE FEATURES, SO IT IS OFTEN RECOMMENDED TO STANDARDIZE THE FEATURES BEFORE APPLYING LASSO REGRESSION. THIS CAN HELP TO ENSURE THAT THE REGULARIZATION TERM IS APPLIED UNIFORMLY TO ALL FEATURES, RATHER THAN BEING DOMINATED BY THE SCALE OF ANY INDIVIDUAL FEATURE.

- ➔ **IF YOU CHOOSE DOUBLE THE VALUE OF ALPHA FOR BOTH RIDGE AND LASSO?**

λ GIVEN HERE, IS ACTUALLY DENOTED BY ALPHA PARAMETER IN THE RIDGE FUNCTION. SO BY CHANGING THE VALUES OF ALPHA, WE ARE BASICALLY CONTROLLING THE PENALTY TERM. HIGHER THE VALUES OF ALPHA, BIGGER IS THE PENALTY AND THEREFORE THE MAGNITUDE OF COEFFICIENTS ARE REDUCED

A HIGHER ALPHA VALUE RESULTS IN A STRONGER PENALTY, AND THEREFORE FEWER FEATURES BEING USED IN THE MODEL
WHEN WE DOUBLE THE VALUE OF ALPHA FOR OUR RIDGE REGRESSION NO WE WILL TAKE THE VALUE OF ALPHA EQUAL TO 10 THE MODEL WILL APPLY MORE PENALTY ON THE CURVE AND TRY TO MAKE THE MODEL MORE GENERALIZED THAT IS MAKING MODEL MORE SIMPLER AND NO THINKING TO FIT EVERY DATA OF THE DATA SET.

- ➔ **WHAT WILL BE THE MOST IMPORTANT PREDICTOR VARIABLES AFTER THE CHANGE IS IMPLEMENTED?**

MOST IMPORTANT VARIABLE AFTER CHANGES HS BEEN IMPLEMENTED IN THE RIDGE REGRESSION:

1. OverallQual	9. MSZoning_FV
2. GrLivArea	10. MSZoning_RM
3. 1stFlrSF	11. Neighborhood_NridgHt
4. GarageArea	12. BsmtFullBath
5. MSZoning_RL	13. MSZoning_RH
6. BedroomAbvGr	14. LotArea
7. FullBath	15. Neighborhood_Crawfor
8. 2ndFlrSF	16. ScreenPorch

MOST IMPORTANT VARIABLE AFTER CHANGES HS BEEN IMPLEMENTED IN THE LASSO REGRESSION:

1. GrLivArea	7.MSZoning_RM
2. OverallQual	8. Neighborhood_NridgHt
3. GarageArea	9. BsmtFullBath
4. MSZoning_RL	10. ScreenPorch
5. MSZoning_FV	11. Neighborhood_Crawfor
6. MSZoning_RH	12. FullBath

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?

IN RIDGE REGRESSION I WILL CHOOSE THE VALUE OF ALPHA IS 3.0 BECAUSE THE DATA WHICH I HAVE PERFORM CERTAIN OPERATIONS ON IT LIKE , DROPPING THE COLUMNS AND PERFORMING THE EDA AND APPLYING THE LINEAR REGRESSION THEN FROM THAT WE ARE APPLYING THE RIDGE AND LASSO REGRESSION. I DON'T THAT WHICH I HAVE PERFORMED IS WELL OR I MAY BE IMPROVE.

WHEN I PERFORMED THE LINEAR REGRESSION THE R2 FOR TRAIN WAS GOOD THAT IS 90% BUT THE TEST SCORE WAS VERY LOW NOT ONLY LOW IT WAS IN NEGATIVE .
THEN I PERFORMED THE RIDGE REGRESSION AND PASS THE ALPHA VALUE ACCORDING TO THE RIDGE IT GIVE ALPHA VALUE 2.0 THEN IF I CALCULATE THE R2 SQUARED VALUE FOR TRAIN ,THE TRAIN SCORE WAS DECREASED 90 TO 89 % WHERE HAS R2 SQUARE FOR TEST CASE WAS INCRESED UP TO 83% . IF WE INCREASED OR DECREASE THE VALUE OF ALPHA LIKE 0.1,2.0,3.0 THERE IS NO CHANGE IN R2 SQURE VALUE.

WHERE AS IN LASSO REGRESSION THE ALPHA VALUE IS 0.0001

- **IF INCREASE THE VALUE OF ALPHA FROM 0.0001 TO ALPHA VALUES LIKE 2.0,3.0,4.0 THERE IS DECRES IN R2 SQUARED VALUE. THE R2 SQUARED VALUE FOR TRAIN TEST WILL BE ZERO.**
- **IF WE DECRESE THE ALPHA FROM 2,3,4 TO 0.01 THEN**
  THE R2 SQUARE VALUE FOR TRAIN WILL INCRESE TO 71% AND FOR TEST IT WILL BE THE 64 %
- **IF WE INCREASE THE VALUE OF ALPHA VALUE FROM 0.01 TO 0.001**
  THEN THE R2 SQUARE VALUE FOR TRAIN WILL INCREASE TO 86 % AND FOR TEST IT WILL BE 83 %
- **IF WE INCREASE THE VALUE OF ALPHA VALUE FROM 0.001 TO 0.0001**
  THEN THE R2 SQUARE VALUE FOR TRAIN WILL INCREASE TO 89 % AND FOR TEST IT WILL BE DECREASE 83 TO 80 %

	METRICS	LINEAR_REGREESION	RIDGE_REGRESSION	LASSO_REGRESSION
0	R2 Score (Train)	8.953741e-01	0.896807	0.000000
1	R2 Score (Test)	-2.378887e+22	0.837925	-0.029860
2	RSS(Train)	1.290044e+01	12.723708	123.300658
3	RSS(Test)	1.249191e+24	8.510815	54.079566
4	MSE(Train)	1.317717e-02	0.012997	0.125946
5	MSE(Test)	2.974264e+21	0.020264	0.128761
6	RMSE(Train)	1.147918e-01	0.114003	0.354888
7	RMSE(Test)	5.453682e+10	0.142351	0.358833

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:

Those 5 most important predictor variables that will be excluded are :-

1. GrLivArea **4. TotalBsmtSF**

2. OverallQual **5. GarageArea**

3. OverallCond

QUESTION 4:

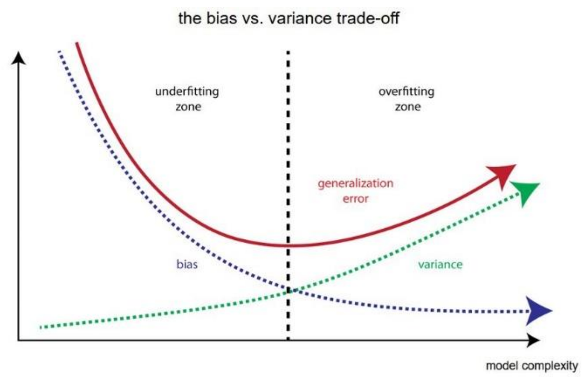
How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

ROBUST: A MODEL IS CONSIDERED TO BE ROBUST IF ITS OUTPUT AND FORECASTS ARE CONSISTENTLY ACCURATE EVEN IF ONE OR MORE OF THE INPUT VARIABLES OR ASSUMPTIONS ARE DRASTICALLY CHANGED DUE TO UNFORESEEN CIRCUMSTANCES.

GENERALIZATION IS A TERM USED TO DESCRIBE A MODEL'S ABILITY TO REACT TO NEW DATA. THAT IS, AFTER BEING TRAINED ON A TRAINING SET, A MODEL CAN DIGEST NEW DATA AND MAKE ACCURATE PREDICTIONS. A MODEL'S ABILITY TO GENERALIZE IS CENTRAL TO THE SUCCESS OF A MODEL. IF A MODEL HAS BEEN TRAINED TOO WELL ON TRAINING DATA, IT WILL BE UNABLE TO GENERALIZE. IT WILL MAKE INACCURATE PREDICTIONS WHEN GIVEN NEW DATA, MAKING THE MODEL USELESS EVEN THOUGH IT IS ABLE TO MAKE ACCURATE PREDICTIONS FOR THE TRAINING DATA. THIS IS CALLED OVERFITTING. THE INVERSE IS ALSO TRUE. UNDERFITTING HAPPENS WHEN A MODEL HAS NOT BEEN TRAINED ENOUGH ON THE DATA. IN THE CASE OF UNDERFITTING, IT MAKES THE MODEL JUST AS USELESS AND IT IS NOT CAPABLE OF MAKING ACCURATE PREDICTIONS, EVEN WITH THE TRAINING DATA.
THE MODEL SHOULD BE AS SIMPLE AS POSSIBLE, THOUGH ITS ACCURACY WILL DECREASE BUT IT WILL BE MORE ROBUST AND GENERALISABLE. IT CAN BE ALSO UNDERSTOOD USING THE BIAS-VARIANCE TRADE-OFF. THE SIMPLER THE MODEL THE MORE THE BIAS BUT LESS VARIANCE AND MORE GENERALIZABLE. ITS IMPLICATION IN TERMS OF ACCURACY IS THAT A ROBUST AND GENERALISABLE MODEL WILL PERFORM EQUALLY WELL ON BOTH TRAINING AND TEST DATA I.E. THE ACCURACY DOES NOT CHANGE MUCH FOR TRAINING AND TEST DATA.

BIAS VARIENCE TRADE OFF:

- ➔ **BIAS MEANS THE DIFFERNECE BETWEEN THE ACTUAL AND PREDICTED**
- ➔ **WHERE AS VARIENCE MEANS HOW PRIDICTED VALUES ARE SCATTERED**
- ➔ **THE IDEAL MODEL SHOULD HAVE LOW BIAS AND LOW VARIENCE**
- ➔ **WHEN WE ARE INCRESING THE COMPLEXITY OF THE MODEL, BIAS IS REDUCED VARIENCE GETS INCRESED**



- ➔ **BIAS HELPS YOU QUANTIFY, HOW ACCURATE IS THE MODEL LIKELY TO BE ON TEST DATA.**
- ➔ **VARIANCE IS THE DEGREE OF CHANGES IN THE MODEL ITSELF WITH RESPECT TO CHANGES IN THE TRAINING DATA.**
- ➔ **ACCURACY OF THE MODEL CAN BE MAINTAINED BY KEEPING THE BALANCE BETWEEN BIAS AND VARIANCE AS IT MINIMIZES THE TOTAL ERROR**

IF THE MODEL HAS LOW BIAS AND LOW VARIENCE THEN THAT MODEL IDENTIFIES ALL THE PATTERNS THAT SHOULD BE PERFORMED WELL ON THE UNSEEN DATA