Assignment-based Subjective Questions	
Question	**Answer**
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?	Optimal value of lambda for Ridge Regression = 6 Optimal value of lambda for Lasso = 0.001 **After doubling the value of Alpha:** **Changes in Ridge Regression metrics:** R2 score of train set decreased from 0.94 to 0.937 R2 score of test set remained same at 0.92 **Changes in Lasso metrics:** R2 score of train set decreased from 0.92 to 0.91 R2 score of test set decreased from 0.92 to 0.91 **So, the most important predictor variables after we double the alpha values are (taking the commons from both model):-** • **GrLivArea**: Above grade (ground) living area square feet • **TotalBsmtSF** : Total square feet of basement area • **OverallQual_9** : Rates the overall material and finish of the house - 9 = Excellent • **OverallQual_8** : Rates the overall material and finish of the house - 8 = Very good • **LotArea** : Lot size in square feet • **BsmtFinSF1** : Type 1 finished square feet **The working code of this is part of the Jupyter notebook**
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?	The model we will choose is <u>not dependent on their values</u> but will depend <u>on the use case</u>. • If we have too many variables and one of our primary goals is feature selection, then we will use **Lasso**. • If we don't want to get too large coefficients and reduction of coefficient magnitude is one of our prime goals, then we will use **Ridge** Regression.
After building the model, you realised that the five most important predictor variables in the **lasso model** are not available in the incoming	• **Top 5 Lasso predictors were:** GrLivArea, TotalBsmtSF, OverallQual_9, OverallQual_8, Neighborhood_Crawfor

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?	**After dropping our top 5 lasso predictors, and recreating the lasso model again - we get the following new top 5 predictors:** • **1stFlrSF:** First Floor square feet • **2ndFlrSF:** Second floor square feet • **TotalBsmtSF:** Total square feet of basement area • **Exterior1st_BrkFace:** Exterior covering on house = Brick Face • **GarageCars:** Size of garage in-car capacity **The working code of this is part of the Jupyter notebook**
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?	• A model is robust when any variation in the data perform well with the unseen data and captures the underlying pattern sufficiently. o It is a generalizable model. o It should not overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. o the model should not be too complex o There is a trade-off between accuracy and generalizable. In order to be robust and generalized we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease. In general, we have to find strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.