1. 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?

The final models for ridge and lasso regression yielded an optimal value of 50 for ridge and 0.01 for lasso. The top ten predicators for ridge and lasso with model's final alphas and new models with double the alphas as follow:

Ridge(alpha=50)	Ridge(alpha=100)	Lasso(alpha=0.01)	Lasso(alpha=0.02)
GrLivArea	GrLivArea	GrLivArea	GrLivArea
TotalBsmtSF	TotalBsmtSF	TotalBsmtSF	TotalBsmtSF
OverallQual_Very Good	OverallQual_Very Good	OverallQual_Very Good	OverallQual_Very Good
ExterQual_Others	ExterQual_Others	GarageCars	GarageCars
BsmtQual_Others	BsmtQual_Others	ExterQual_Others	ExterQual_Others
GarageCars	GarageCars	BsmtQual_Others	BsmtQual_Others
LotArea	GarageArea	LotArea	BsmtFinType1_GoodLivQtr
Foundation_PouredContrete	TotRmsAbvGrd	KitchenQual_Others	GarageArea
GarageArea	LotArea	Foundation_PouredContrete	LotArea
HouseStyle_2 Story	KitchenQual_Others	BsmtFinType1_GoodLivQtr	KitchenQual_Others

Increase in alpha results in pushing the coefficients of the predicators more towards zero in both ridge and lasso, however in case of lasso there is an increase of number of predictors with coefficient of zero thereby implying that regularization is high. This is because the penalty on the models has now increased and this results in decreasing variance and increasing bias. The model starts to become a much simpler model and may actually lead to underfitting.

Below are the metrics from before doubling the alphas and after doubling them.

	Model	R2	RSS	RMSE	Adj_R2
4	Ridge[Train]	0.8811	114.8585	0.3448	0.8746
5	Ridge[Test]	0.8696	54.1280	0.3611	0.8624
12	Ridge-x2[Train]	0.8776	118.2719	0.3499	0.8709
13	Ridge-x2[Test]	0.8681	54.7501	0.3632	0.8609

	Model	R2	RSS	RMSE	Adj_R2
8	Lasso[Train]	0.8759	119.8348	0.3522	0.8692
9	Lasso[Test]	0.8727	52.8241	0.3568	0.8658
14	Lasso-x2[Train]	0.8682	127.3176	0.3630	0.8610
15	Lasso-x2[Test]	0.8720	53.1351	0.3578	0.8650

It is noted that R2 has reduced slights, which both RSS and RMSE have slightly increased thus impacting adjusted R2.

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?

After applying regularization techniques on the model, it was found that lasso model with an alpha of 0.01 was best suited as an ideal model since it not only reduced the model complexity by reducing the coefficients of the few predicators to zero especially those that were not relevant to the dependent variable as per EDA performed earlier, but also yielded almost similar R2 scores for both train and test datasets with the most ideal RSS, RSME and adjusted R2 among all the model created during the assignment(see above screenshots).

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?

Alpha for Lasso model is still at 0.01 as the optimal value. Following are the five most important pred ictor variables after removing the five most important predictor variables from the initial lasso mode l: 'TotRmsAbvGrd', 'GarageArea', 'LotArea', 'BsmtQual_Others', 'KitchenQual_Others'

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?

An ideal model is simple enough but not too simple that it starts to underfit the data. This is achieved by trading off bias in favour of variance which is typically known as Bias-Variance trade-off. More the bias, lesser will be the variance resulting in an accuracy that is comparable between the train and test data set. The goal is to ensure that all patterns are picked from the data set provided, but not at the cost of a high variance which will reduce the accuracy of the unseen data.

However, this process also requires further steps to make the model generalisable by using cross validation such as k-fold on the train data set which assesses the model's performance on multiple subsets of the train data set while still not being able to peek at the unseen data thereby ensuring model integrity.

Feature engineering can also be used to engineer the features to capture the important information in the data. The model can focus on the most informative aspects and improves its generalization abilities.

Finally, applying regularization techniques like Ridge and Lasso prevent overfitting and helps to control the complexity of the model to reduce the impact of irrelevant and noisy predictors. Furthermore, Lasso helps in feature selection which aids in reducing the complexity of the model.

An ideal model is less likely to be affected by noise leading to more reliable predictions. It can also perform well on the unseen data but at the cost of bias and accuracy. However high accuracy does not necessarily mean robustness. A balance between accuracy and generalisability of the model is required.