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 to 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 for alpha in Ridge is 3.0 and in Lasso is 0.00019 for the given data set after cleaning, pre-processing, and train-test split with 70-30 ratio. The value of alpha in both Ridge and Lasso Regression is a non-negative float which controls the strength of Regularization. Higher the value of alpha, higher will be the Regularization in the model to avoid overfitting. However, beyond an optimum value increase in the value of alpha, meaning even more regularization, will cause the model to not learn enough from the underlying patterns, which leads to underfitting of the model.

Hence, if we use double the value of optimum alpha, the increased regularization leading to underfitting will cause the model to perform poorer than for the optimum value. Thus, there will be a reduction in the R2 score and increase in Root Mean Squared Error (RMSE) of the models with doubled alpha values. The same can be observed from the values in the below table.

Model	Ridge (alpha = 3)	Ridge (alpha = 6)	Lasso (alpha = 0.00019)	Lasso (alpha = 0.0002)
R2 Score	0.8689	0.8629	0.8740	0.8688
RMSE	0.0542	0.0555	0.0532	0.0543

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?

The following table illustrates the optimal alpha value and the corresponding performance metrics for Ridge and Lasso.

Model	Optimum alpha	R2 Score (Train) in %	RMSE (Train)	R2 Score (Test) in %	RMSE (Test)
Ridge Regression	3.0	91.5964	0.0453524	86.8918	0.0542962
Lasso Regression	0.00019	91.4384	0.0457767	87.4013	0.0532303

As can be seen from the above table the Root Mean Squared Error is least and R2 Score is the maximum for Lasso. Moreover, lasso also helps in feature reduction, and have more edge over Ridge in that aspect well.

Thus, the best model built is the optimized Lasso Regression and the variables chosen by Lasso can be used for Sale Price prediction of the house, which are:

Priority	Variable	Coefficient	Description
1	OverallQual	0.2417	The Overall Quality of Materials inside house and finish of Construction
2	BsmtFinSF1	0.1707	The size in square feet of the finished Basement (Type 1)
3	OverallCond	0.1091	The overall condition of the house
4	2ndFlrSF	0.0902	Size of the Second Floor in square feet
5	MasVnrArea	-0.0684	Size of the Mansory Veneer in square feet
6	Neighborhood_Crawfor	-0.0432	If the house is located in the neighbrhood of Crawford
7	TotalBsmtSF	0.043	Total Basement Area in square feet
8	WoodDeckSF	0.0395	Size of the Wood Deck in square feet
9	Fireplaces	0.0351	Number of Fireplaces
10	LotConfig_CulDSac	0.0348	If Lot Configeration is Cul-de-sac

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?

The top five predictor variables in the lasso models are:

1	OverallQual
2	BsmtFinSF1
3	OverallCond
4	2ndFlrSF
5	MasVnrArea

If data of those variables are not available, then they have to removed and a new model needs to be created. And in the new lasso model, the five most important predictor variables are:

Priority	Variable	Coefficient	Description
1	GrLivArea	0.2556	Living Area above ground in sqaure feet
2	GarageArea	0.178	Area of Garage in square feet
3	Fireplaces	0.098	Number of Fireplaces
4	HouseStyle_2.5Fin	0.0492	If the house has two stories and an unfinished level
5	KitchenQual_TA	-0.0397	If Kitchen is of typical quality

Note: Code for dropping variables and creation of new model is included in the python notebook.

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?

When it comes to the question of generalisability and robustness of a model, one can safely depend on the Principle of Parsimony, otherwise known as Occam's Razor, which clearly states that, "other things equal, explanations that posit fewer entities, or fewer kinds of entities, are to be preferred to explanations that posit more", which can be interpreted as follows, in the scenario of Machine Learning models.

If two machine learning models show similar performance on a finite training and/or test data set, then one should always choose the model, which has made fewer assumptions on the test data. That is, out of two models with similar performance, always choose the simpler model.

This statement can be supported using the following facts regarding ML models.

- Simpler models have lesser variance, and thus will be more generic and robust compared to a complex model.
- Simpler models are easier to train, as they require fewer training compared to complex ones to give similar performance.
- Simpler models have lesser tendency to overfit compared to complex models lead, and thus will give better results on unseen data, making them more generalizable.

By ensuring that my model follows all the above criteria, and I follow the Occam's Razor during model selection, generalisability and robustness of the model can be ensured. Use of regularization to avoid overfitting and choosing best Bias-Variance Trade off scenario will reassure the generic nature of model.

On the other hand, accuracy of the model depends upon the bias-variance trade off. A model can memorize the whole training data and can have great accuracy on the training set, which is the perfect example of an overfit complex model. But such a model will perform poorly on unseen test data. Thus, choosing a Bias-Variance trade-off that results in almost similar accuracies in both train and testing set will be ideal, as such model will be simpler and will not vary much based on change data. To conclude, by selectin a simple model according to Occam's Razor, to ensure generalisability and robustness, it also affirms the choice of an accurate model that will perform well on unseen data.