Subjective Questions

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?

Ans: Alpha are 10 and 100 for ridge and lasso regression respectively. The r2 score goes down if we double the value of alpha.

Important predictor for Lasso are,
1stFlrSF,2ndFlrSF,Alley,BldgType,BsmtCond,BsmtExposure,BsmtFinSF1,BsmtFinType1,BsmtFinType2,BsmtQual,BsmtUnfSF,CentralAir,Condition1,Electrical,ExterCond,ExterQual,Exterior1st,Exterior2nd,Fence,FireplaceQu,Foundation,Functional,GarageArea,GarageCond,GarageFinish,GarageQual,GarageType,GarageYrBlt,GrLivArea,HeatingQC,HouseStyle,KitchenQual,LandContour,LotArea,LotConfig,LotFrontage,LotShape,MSZoning,MasVnrArea,MasVnrType,Neighborhood,OverallCond,OverallQual,PavedDrive,RoofStyle,SaleCondition,TotalBsmtSF,WoodDeckSF

Important predictor for Ridge are,

Alley,BldgType,BsmtCond,BsmtExposure,BsmtFinType1,BsmtFinType2,BsmtQual,CentralAir,Condition1,Electrical,ExterCond,ExterQual,Exterior1st,Exterior2nd,FireplaceQu,Foundation,Functional,GarageFinish,GarageType,GarageYrBlt,Heating,HeatingQC,HouseStyle,KitchenQual,LandContour,LandSlope,LotConfig,MSZoning,MasVnrType,Neighborhood,OverallCond,OverallQual,PavedDrive,PoolQC,RoofStyle,SaleCondition,SaleType

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?

Ans: Both regressions can be taken into consideration as both have r2 of around 86% . But in Ridge regression there 37 important features and lasso has 48 features. So, I would go for Ridge regression on which we can focus to drive the business decision.

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?

Ans: Important features are OverallQual, SaleCondition, Neighborhood, Functional, RoofStyle

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?

Ans: The model is performing well on unseen data with the r2 score of 86% which is a good indicator and that implicates the robustness and generalisation capability of the model.