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:
i. Optimal Value:
Ridge: 2
Lasso: 100

ii.
Ridge:
Alpha 2:
R2score(train): 0.9277294745972668
R2score(test): 0.8915019137941209

Alpha 4: (post double the alpha value)
R2score(train): 0.9222097346399525
R2score(test): 0.8862435414650203

Lasso:
Alpha-100:
R2score(train): 0.9186327292619871
R2score(test): 0.8912257646336962

Alpha-200: (post double the alpha value)
R2score(train): 0.9048537576117922
R2score(test): 0.874717723422264

R2score on training and testing data has decreased post double the optimal value of Ridge and Lasso

iii. Below are the most important predictor variables.

- GrLivArea
- OverallQual
- TotalBsmtSF
- Neighborhood_NridgHt
- BsmtFinSF1

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?

Answer:
Ridge:
Alpha 2
R2score(train): 0.9277294745972668
R2score(test): 0.8915019137941209

Lasso:
Alpha-100
R2score(train): 0.9186327292619871
R2score(test): 0.8912257646336962

As per above R2 score lasso will look better compared to ridge, as lasso will help in feature elimination from given dataset. Also, R2 difference of test and train with respect lasso is ~2.74%, whereas for ridge it is ~3.62%

Question 3:
After building the model, you realized 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:
Earlier below mentioned predictor variables were important.

- GrLivArea
- OverallQual
- TotalBsmtSF
- Neighborhood_NridgHt
- OverallCond

Post dropping above variables, we are getting below mentioned predictor variables as important variables

- 1stFlrSF
- 2ndFlrSF
- BsmtFinSF1
- GarageArea
- MasVnrArea

Question 4:

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

Answer:

If there is any drastic change in input or predictor variable it shouldn't impact the accuracy of model then that model is considered as robust and due to addition of any new condition on basic dataset its performance is not impacted can be considered as generable model.

Model should be generalizable so that accuracy is not lesser on training data and outliers exist in dataset should be removed to get high accuracy.

If the model is not robust and generalizable then that model cannot be considered for predictive analysis.