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:

- Best alpha value for Lasso : {'alpha': 0.001}
- Best alpha value for Ridge : {'alpha': 0.1}
- If we choose to double the value of alpha for both ridge and lasso: In case of ridge that will lower the coefficients and in case of Lasso there would be more less important features coefficients turning 0.
- The most important predictor variable after the change is implemented are those which are significant.

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:

- Best alpha value for Lasso : {'alpha': 0.001}
- Best alpha value for Ridge : {'alpha': 0.1}
- As we Got good score for both the models so we can go with Lasso regression as it results in model parameters such that lesser important features coefficients become zero.
- Ridge: Train :94.97 Test :87.08 and Lasso : Train :90.55 Test :87.17

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?

Answer:

On running the same notebook and removing the top 5 significant variables:

We found below variables as next 5 significant.

-Ridge

1. RoofMatl_WdShake
2. RoofMatl_Roll
3. RoofMatl_WdShngl

4. MSZoning_RH
5. Condition2_PosN

-Lasso

1. Neighborhood_StoneBr
2. Neighborhood_OldTown
3. Electrical_FuseF
4. BsmtExposure_Mn
5. Exterior1st_CBlock

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?

Answer:

- Changes for model:
    - Use a model that's resistant to outliers. Tree-based models are generally not as affected by outliers, while regression-based models are. If you're performing a statistical test, try a non-parametric test instead of a parametric one.
- Changes for data:
    - Transform your data. If your data has a very pronounced right tail, try a log transformation.
    - Remove the outliers. This works if there are very few of them and you're fairly certain they're anomalies and not worth predicting.