Qtn. 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.

- Optimal Value of Alpha for Ridge Regression : 10
- Optimal Value of Alpha for Lasso Regression : 0.003
- For Ridge Regression, when we double the optimal alpha value,
  - the value of R^2 value decreases then the model with optimal alpha.
  - the Beta Coefficients values decreases.
  - The Top 5 features remains the same although the values of Beta Coefficient reduces a bit
  - The Top 5 feature for Ridge Model are:
    - GrLivArea (c)
    - TotRmsAbvGrd (Total rooms above grade (does not include bathrooms))
    - FullBath (Full bathrooms above grade)
    - 1stFlrSF (First Floor square feet)
    - 2ndFlrSF (Second floor square feet)
- For Lasso Regression, when we double the optimal alpha value,
  - the value of R^2 decreases compared to the model created with optimal alpha.
  - the Beta Coefficients values decreases.
  - The Top 5 features change as well the values of Beta Coefficient decreases
  - The Top 5 feature for Lasso Model for Optimal Alpha(0.003) are:
    - GrLivArea (Above grade (ground) living area square feet)
    - GarageCars (Size of garage in car capacity)
    - OverallQual (Rates the overall material and finish of the house) with Excellent Quality
    - MSZoning (Identifies the general zoning classification of the sale) with Residential High Density & Residential Low Density zone
  - The Top 5 feature for Lasso Model after doubling Alpha(0.006) are:
    - GrLivArea (Above grade (ground) living area square feet)
    - GarageCars (Size of garage in car capacity)
    - TotRmsAbvGrd (Total rooms above grade (does not include bathrooms))
    - Fireplaces (Number of fireplaces)
    - BsmtQual (Evaluates the height of the basement) with Excellent quality

Qtn.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.

Optimal Values of Alpha for Lasso is 0.003 and for Ridge is 10. I will choose the Lasso Regression because it has higher R^2 value on train and test data as well as it has selected optimal features for the best fit of regression

Qtn.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.

After excluding the top 5 features in Lasso Regression, we get the following top 5 features

1. TotRmsAbvGrd (Total rooms above grade)
2. GarageArea (Size of garage in square feet)
3. 1stFlrSF (First Floor square feet)
4. Fireplaces (Number of Fireplaces)
5. BsmtQual_Ex (Basement Quality with Excellent Quality)

Qtn.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.

To make the model robust and generalisable, we need to check that:

- the model must not have too many features and overfit the training data.
- It should be having optimal features, which makes it the best fit to predict the nearest values to the actual values on the unseen test data.
- The implications and reasons of the model being robust and generalizable are:
    o It can predict the values well on the unseen data
    o It is easy to adapt new features in the model and retrain the model
    o It gives correct results more often than wrong results i.e. it has the best accuracy
    o It can be reused for the similar data sets