* **Q:** Which variables are significant in predicting the price of a house, and

**Ans:** Below variables are significant variables

* 1. **LotFrontage**
  2. **MasVnrArea**
  3. **BsmtFinSF1**
  4. **BsmtFinSF2**
  5. **BsmtUnfSF**
  6. **GrLivArea**
  7. **Fireplaces**
  8. **GarageArea**
  9. **WoodDeckSF**
  10. **BsmtQual\_encoded**
  11. **KitchenQual\_encoded**
* Q: How well those variables describe the price of a house.

**Ans:** **Below variables are most important variables to describe price of a house as they have comparatively higher coefficient than other variables.**

* 1. **BsmtFinSF1 --** 0.1421
  2. **BsmtUnfSF --** 0.0915
  3. **GrLivArea --** 0.3431
  4. **GarageArea --** 0.1040

**Problem Statement - Part II :**

**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:** Below are the optimal values of alpha:

**Ridge** 🡪 1000 🡪 below are the r2 score

**# alpha train Test**

1000 0.81 0.85

After double the value of alpha=2000

**# alpha train Test**

2000 0.78 0.83

Test r2Score is higher than train data set, hence 2000 is not correct value.

**Lasso** 🡪 2500 🡪 Below are r2 score

**# alpha train Test**

2500 0.90 0.88

After double the value of alpha =5000

**# alpha train Test**

5000 0.87 0.86

10000 0.82 0.83

**Changing value of alpha is not impacting that much in Lasso, Hence it seems it is not reliable model.**

**Q: 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: Below are standard deviation of both model**

**Ridge:** 0.097

**Lasso:** 0.57

**Lasso has very high standard deviation hence it is not a good model, on the other hand Ridge is having low standard deviation Hence I would chose Ridge Model.**

**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: we can check Cross Validation score using K-fold and hyper parameter tuning method.**

**If standard deviation is higher than we can say that model is not robust.**