Question 1

What is the optimal value of alpha for ridge and lasso regression?

The optimum alpha values are:

- Ridge: 5.0
- Lasso: 0.01

```
ridge_alpha = ridge_model_cv.best_params_.get('alpha')
# ridge_alpha = 10.0

ridge = Ridge(alpha=ridge_alpha)
ridge.fit(X_train, y_train)
```

```
Ridge(alpha=5.0)
```

```
lasso_alpha = lasso_model_cv.best_params_.get('alpha')
lasso = Lasso(alpha=lasso_alpha)
lasso.fit(X_train, y_train)
```

```
Lasso(alpha=0.01)
```

What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

Ridge: the coefficient changes but not significantly enough to change the order of Features. Higher alpha causes the coefficient to lower for positive value and the opposite for negative. Ridge uses Lamba as penalty.

Alpha: 5.0

	Coefficient	Features
0	-0.254	MSSubClass_1-1/2 STORY FINISHED ALL AGES
1	-0.692	MSSubClass_1-STORY 1945 & OLDER
2	-0.137	MSSubClass_2 FAMILY CONVERSION - ALL STYLES AN...
3	0.022	MSSubClass_2-STORY 1945 & OLDER
4	0.417	MSSubClass_2-STORY 1946 & NEWER
5	-0.153	MSSubClass_2-STORY PUD - 1946 & NEWER
6	-0.036	MSSubClass_DUPLEX - ALL STYLES AND AGES
7	-0.295	MSSubClass_PUD - MULTILEVEL - INCL SPLIT LEV/F...
8	-0.202	MSSubClass_SPLIT FOYER
9	-0.008	LandSlope_Sev
10	-0.137	BldgType_2fmCon

Alpha: 10.0

	Coefficient	Features
0	-0.211	MSSubClass_1-1/2 STORY FINISHED ALL AGES
1	-0.566	MSSubClass_1-STORY 1945 & OLDER
2	-0.097	MSSubClass_2 FAMILY CONVERSION - ALL STYLES AN...
3	0.017	MSSubClass_2-STORY 1945 & OLDER
4	0.381	MSSubClass_2-STORY 1946 & NEWER
5	-0.123	MSSubClass_2-STORY PUD - 1946 & NEWER
6	-0.021	MSSubClass_DUPLEX - ALL STYLES AND AGES
7	-0.180	MSSubClass_PUD - MULTILEVEL - INCL SPLIT LEV/F...
8	-0.141	MSSubClass_SPLIT FOYER
9	-0.014	LandSlope_Sev
10	-0.097	BldgType_2fmCon

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \sum_{j=0}^{p} w_j^2$$

Ridge puts constraint on the coefficient (w) in the formula above. Large w will get penalized.

Lasso penalizes the sum of their absolute values. This also known as L1 regularization; where as Ridge is L2. Higher alpha can lead to zero coefficient; hence auto feature selection.

<table>
<tr><th>Alpha = 0.01</th><th>Alpha = 0.02</th></tr>
<tr><td>

	Coefficient	Features
0	-0.122	MSSubClass_1-1/2 STORY FINISHED ALL AGES
1	-0.646	MSSubClass_1-STORY 1945 & OLDER
2	0.438	MSSubClass_2-STORY 1946 & NEWER
3	0.165	HouseStyle_2Story
4	0.427	Exterior1st_CemntBd
5	0.108	Exterior1st_VinylSd
6	0.887	Exterior2nd_CmentBd
7	0.627	Exterior2nd_VinylSd
8	-0.664	GarageType_NA

</td><td>

	Coefficient	Features
0	-0.016	MSSubClass_1-1/2 STORY FINISHED ALL AGES
1	-0.481	MSSubClass_1-STORY 1945 & OLDER
2	0.423	MSSubClass_2-STORY 1946 & NEWER
3	0.158	HouseStyle_2Story
4	0.421	Exterior1st_CemntBd
5	0.107	Exterior1st_VinylSd
6	0.465	Exterior2nd_CmentBd
7	0.604	Exterior2nd_VinylSd
8	-0.553	GarageType_NA

</td></tr>
</table>

$$\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{M}\left(y_i - \sum_{j=0}^{p} w_j \times x_{ij}\right)^2 + \lambda \sum_{j=0}^{p} |w_j|$$

Note: in the formular above, the absolute on the fight side.

What will be the most important predictor variables after the change is implemented?

<table>
<tr><th>Ridge</th><th>Lasso</th></tr>
<tr><td>Important variables have rfe_ranking of 1:

Exterior1st_BrkComm

Exterior2nd_CmentBd

Exterior2nd_VinylSd
GarageType_NA

GarageQual_Po</td><td>Important variables have rfe_ranking of 1:

MSSubClass_1-STORY 1945 & OLDER
MSSubClass_2-STORY 1946 & NEWER
Exterior2nd_CmentBd
Exterior2nd_VinylSd
GarageType_NA</td></tr>
</table>

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?

In this use case, I'd go with Ridge. Out of the bat and without performing hyper parameter turning, Ridge 0.511 R-squared and 0.472 Adj. R-squared. Also, the VIF value for Ridge is a lot smaller. I have not removed any variables and retrain the model. The result is subjected to change.

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.511
Model:	OLS	**Adj. R-squared:**	0.472
Method:	Least Squares	**F-statistic:**	13.26
Date:	Wed, 01 Feb 2023	**Prob (F-statistic):**	4.27e-48
Time:	02:00:36	**Log-Likelihood:**	-501.28
No. Observations:	467	**AIC:**	1073.
Df Residuals:	432	**BIC:**	1218.

On the other hand, Lasso gives 0.480 R-squared and 0.470 Adj. R-squared.

OLS Regression Results

Dep. Variable:	SalePrice	**R-squared:**	0.480
Model:	OLS	**Adj. R-squared:**	0.470
Method:	Least Squares	**F-statistic:**	46.95
Date:	Wed, 01 Feb 2023	**Prob (F-statistic):**	1.34e-59
Time:	02:00:02	**Log-Likelihood:**	-515.25
No. Observations:	467	**AIC:**	1050.
Df Residuals:	457	**BIC:**	1092.

Question 3

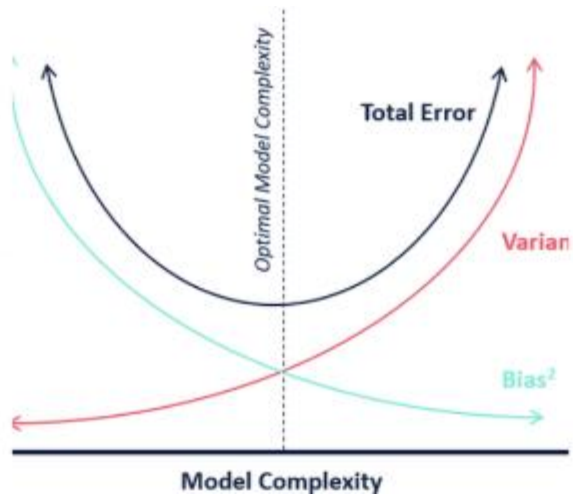
After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Comparing the Coeff, VIF and p-value, I would say the five important predictor variables are:

- MSSubClass_2-STORY 1946 & NEWER
- MSSubClass_1-1/2 STORY FINISHED ALL AGES
- HouseStyle_2Story
- Exterior2nd_VinylSd
- Exterior1st_CemntBd

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?



To generalize the model, we have to discuss the diagram above. The middle dotted line is the optimal where the model is not so complex and both bias and variance meet at an equilibrium middle section. Highly complex model will have low bias and high variance. The opposite is true where not so complex model will have low variance and high bias. On the left side (high bias and low variance), the model is considered as underfitting. Contrary, low bias high variance is overfitting. That means the model is doing very good with training data, but do poorly with real data.