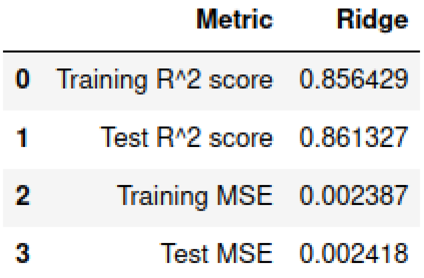
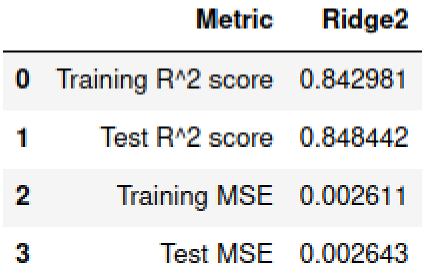
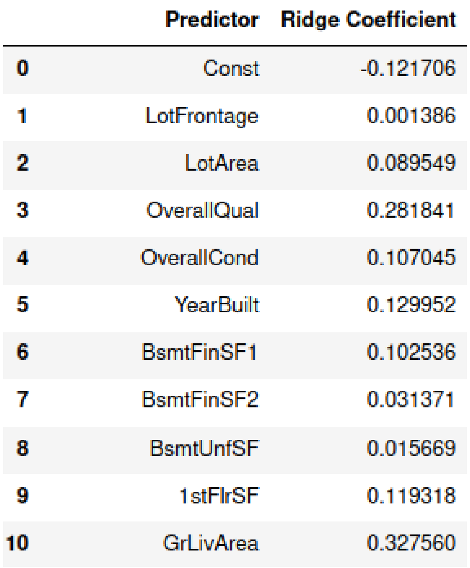
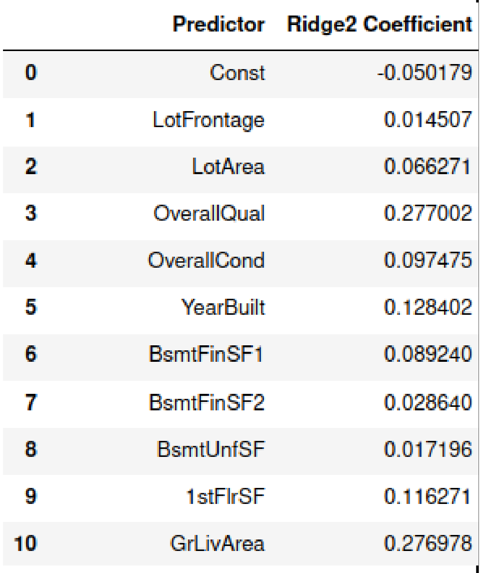
Question 1:
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

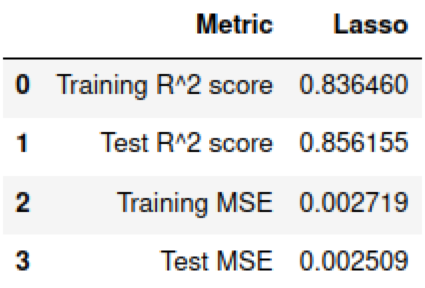
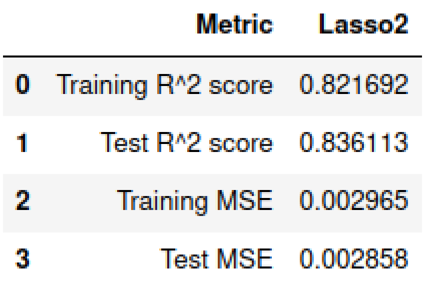
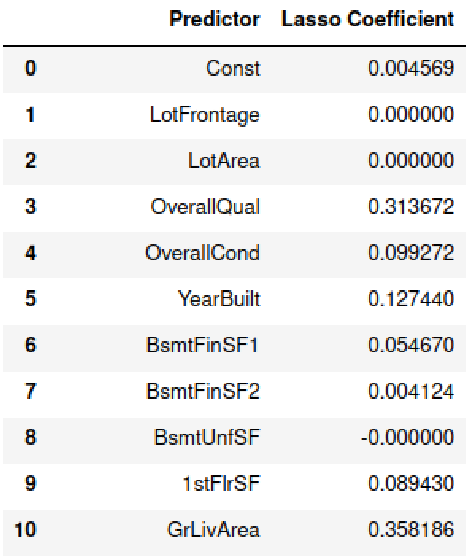
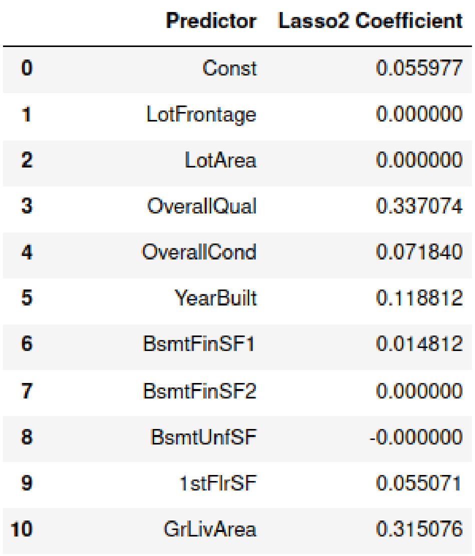
Answer:
Optimal value of alpha
The optimal value of alpha for **ridge regression is 2** and for **lasso regression is 0.0004**.

Change in the model if alpha is doubled for ridge regression
If alpha is doubled for ridge regression, the training and test R^2 score and MSE decrease slightly, and the coefficients of the predictors mostly shrink towards 0 as shown below.

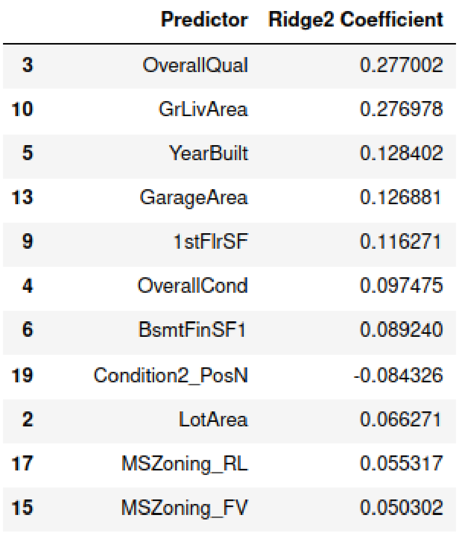
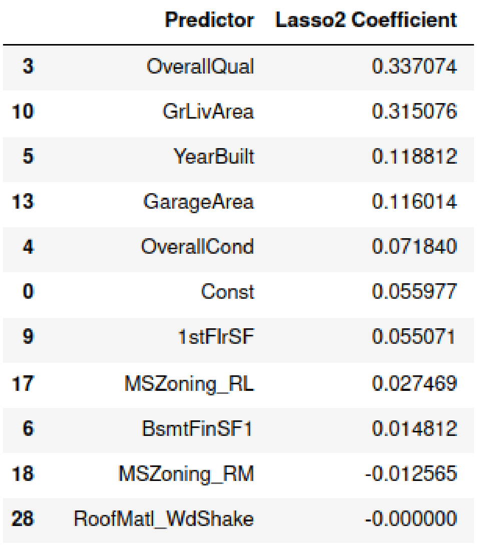
<table>
<tr><th>Model</th><th>alpha = 2</th><th>alpha = 4</th></tr>
<tr><td>Performance Metrics</td><td>
<table>
<tr><th></th><th>Metric</th><th>Ridge</th></tr>
<tr><td>0</td><td>Training R^2 score</td><td>0.856429</td></tr>
<tr><td>1</td><td>Test R^2 score</td><td>0.861327</td></tr>
<tr><td>2</td><td>Training MSE</td><td>0.002387</td></tr>
<tr><td>3</td><td>Test MSE</td><td>0.002418</td></tr>
</table></td><td>
<table>
<tr><th></th><th>Metric</th><th>Ridge2</th></tr>
<tr><td>0</td><td>Training R^2 score</td><td>0.842981</td></tr>
<tr><td>1</td><td>Test R^2 score</td><td>0.848442</td></tr>
<tr><td>2</td><td>Training MSE</td><td>0.002611</td></tr>
<tr><td>3</td><td>Test MSE</td><td>0.002643</td></tr>
</table></td></tr>
<tr><td>Coefficients</td><td>
<table>
<tr><th></th><th>Predictor</th><th>Ridge Coefficient</th></tr>
<tr><td>0</td><td>Const</td><td>-0.121706</td></tr>
<tr><td>1</td><td>LotFrontage</td><td>0.001386</td></tr>
<tr><td>2</td><td>LotArea</td><td>0.089549</td></tr>
<tr><td>3</td><td>OverallQual</td><td>0.281841</td></tr>
<tr><td>4</td><td>OverallCond</td><td>0.107045</td></tr>
<tr><td>5</td><td>YearBuilt</td><td>0.129952</td></tr>
<tr><td>6</td><td>BsmtFinSF1</td><td>0.102536</td></tr>
<tr><td>7</td><td>BsmtFinSF2</td><td>0.031371</td></tr>
<tr><td>8</td><td>BsmtUnfSF</td><td>0.015669</td></tr>
<tr><td>9</td><td>1stFlrSF</td><td>0.119318</td></tr>
<tr><td>10</td><td>GrLivArea</td><td>0.327560</td></tr>
</table></td><td>
<table>
<tr><th></th><th>Predictor</th><th>Ridge2 Coefficient</th></tr>
<tr><td>0</td><td>Const</td><td>-0.050179</td></tr>
<tr><td>1</td><td>LotFrontage</td><td>0.014507</td></tr>
<tr><td>2</td><td>LotArea</td><td>0.066271</td></tr>
<tr><td>3</td><td>OverallQual</td><td>0.277002</td></tr>
<tr><td>4</td><td>OverallCond</td><td>0.097475</td></tr>
<tr><td>5</td><td>YearBuilt</td><td>0.128402</td></tr>
<tr><td>6</td><td>BsmtFinSF1</td><td>0.089240</td></tr>
<tr><td>7</td><td>BsmtFinSF2</td><td>0.028640</td></tr>
<tr><td>8</td><td>BsmtUnfSF</td><td>0.017196</td></tr>
<tr><td>9</td><td>1stFlrSF</td><td>0.116271</td></tr>
<tr><td>10</td><td>GrLivArea</td><td>0.276978</td></tr>
</table></td></tr>
</table>

Change in the model if alpha is doubled for lasso regression

If alpha is doubled for lasso regression, the training and test R^2 score and MSE decrease slightly, and the coefficients of the predictors mostly shrink towards 0 as shown below.

<table>
<tr><th>Model</th><th>alpha = 0.0004</th><th>alpha = 0.0008</th></tr>
<tr><td>Performance Metrics</td>
<td>
<table>
<tr><th></th><th>Metric</th><th>Lasso</th></tr>
<tr><td>0</td><td>Training R^2 score</td><td>0.836460</td></tr>
<tr><td>1</td><td>Test R^2 score</td><td>0.856155</td></tr>
<tr><td>2</td><td>Training MSE</td><td>0.002719</td></tr>
<tr><td>3</td><td>Test MSE</td><td>0.002509</td></tr>
</table></td>
<td>
<table>
<tr><th></th><th>Metric</th><th>Lasso2</th></tr>
<tr><td>0</td><td>Training R^2 score</td><td>0.821692</td></tr>
<tr><td>1</td><td>Test R^2 score</td><td>0.836113</td></tr>
<tr><td>2</td><td>Training MSE</td><td>0.002965</td></tr>
<tr><td>3</td><td>Test MSE</td><td>0.002858</td></tr>
</table></td></tr>
<tr><td>Most important Predictors</td>
<td>
<table>
<tr><th></th><th>Predictor</th><th>Lasso Coefficient</th></tr>
<tr><td>0</td><td>Const</td><td>0.004569</td></tr>
<tr><td>1</td><td>LotFrontage</td><td>0.000000</td></tr>
<tr><td>2</td><td>LotArea</td><td>0.000000</td></tr>
<tr><td>3</td><td>OverallQual</td><td>0.313672</td></tr>
<tr><td>4</td><td>OverallCond</td><td>0.099272</td></tr>
<tr><td>5</td><td>YearBuilt</td><td>0.127440</td></tr>
<tr><td>6</td><td>BsmtFinSF1</td><td>0.054670</td></tr>
<tr><td>7</td><td>BsmtFinSF2</td><td>0.004124</td></tr>
<tr><td>8</td><td>BsmtUnfSF</td><td>-0.000000</td></tr>
<tr><td>9</td><td>1stFlrSF</td><td>0.089430</td></tr>
<tr><td>10</td><td>GrLivArea</td><td>0.358186</td></tr>
</table></td>
<td>
<table>
<tr><th></th><th>Predictor</th><th>Lasso2 Coefficient</th></tr>
<tr><td>0</td><td>Const</td><td>0.055977</td></tr>
<tr><td>1</td><td>LotFrontage</td><td>0.000000</td></tr>
<tr><td>2</td><td>LotArea</td><td>0.000000</td></tr>
<tr><td>3</td><td>OverallQual</td><td>0.337074</td></tr>
<tr><td>4</td><td>OverallCond</td><td>0.071840</td></tr>
<tr><td>5</td><td>YearBuilt</td><td>0.118812</td></tr>
<tr><td>6</td><td>BsmtFinSF1</td><td>0.014812</td></tr>
<tr><td>7</td><td>BsmtFinSF2</td><td>0.000000</td></tr>
<tr><td>8</td><td>BsmtUnfSF</td><td>-0.000000</td></tr>
<tr><td>9</td><td>1stFlrSF</td><td>0.055071</td></tr>
<tr><td>10</td><td>GrLivArea</td><td>0.315076</td></tr>
</table></td></tr>
</table>

Most important predictors after the change is implemented

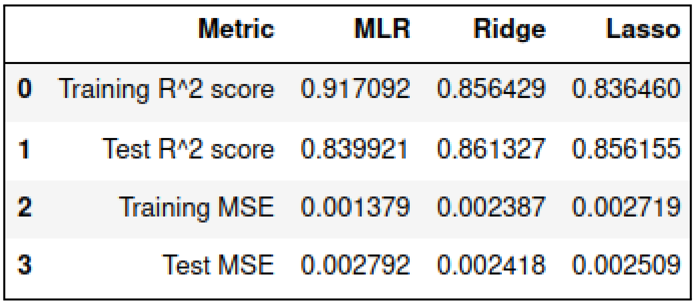
Model	Ridge	Lasso
Most Important Predictors		

	Predictor	Ridge2 Coefficient
3	OverallQual	0.277002
10	GrLivArea	0.276978
5	YearBuilt	0.128402
13	GarageArea	0.126881
9	1stFlrSF	0.116271
4	OverallCond	0.097475
6	BsmtFinSF1	0.089240
19	Condition2_PosN	-0.084326
2	LotArea	0.066271
17	MSZoning_RL	0.055317
15	MSZoning_FV	0.050302

	Predictor	Lasso2 Coefficient
3	OverallQual	0.337074
10	GrLivArea	0.315076
5	YearBuilt	0.118812
13	GarageArea	0.116014
4	OverallCond	0.071840
0	Const	0.055977
9	1stFlrSF	0.055071
17	MSZoning_RL	0.027469
6	BsmtFinSF1	0.014812
18	MSZoning_RM	-0.012565
28	RoofMatl_WdShake	-0.000000

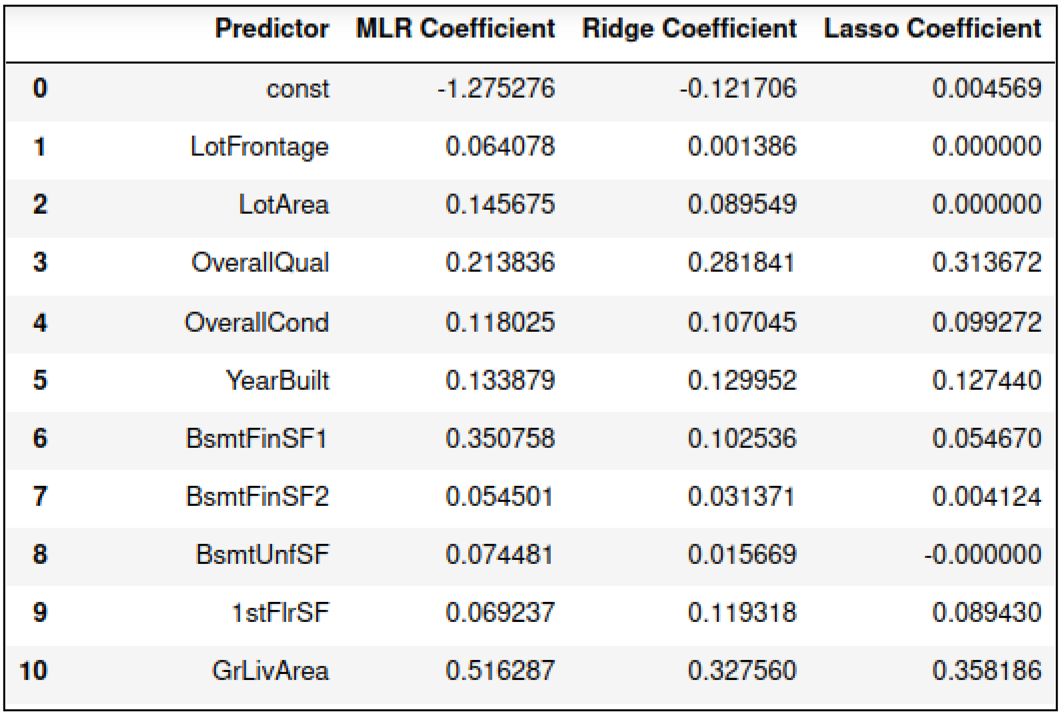
Question 2:
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer:
1. The optimal value of lambda for **ridge regression is 2** and for **lasso regression is 0.0004**. For these values of lambda, ridge regression has better values of training and test R^2 score and MSE.



	Metric	MLR	Ridge	Lasso
0	Training R^2 score	0.917092	0.856429	0.836460
1	Test R^2 score	0.839921	0.861327	0.856155
2	Training MSE	0.001379	0.002387	0.002719
3	Test MSE	0.002792	0.002418	0.002509

2. However, as shown below the lasso regression model performs automatic variable selection (by setting some coefficients to zero) and gives more insight into the variables that influence SalePrice, which is the goal of the business.

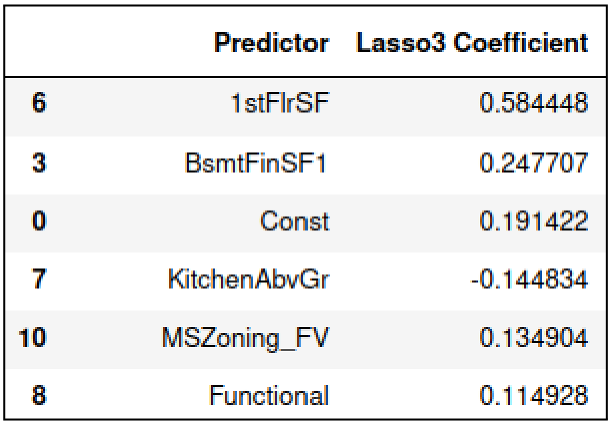


	Predictor	MLR Coefficient	Ridge Coefficient	Lasso Coefficient
0	const	-1.275276	-0.121706	0.004569
1	LotFrontage	0.064078	0.001386	0.000000
2	LotArea	0.145675	0.089549	0.000000
3	OverallQual	0.213836	0.281841	0.313672
4	OverallCond	0.118025	0.107045	0.099272
5	YearBuilt	0.133879	0.129952	0.127440
6	BsmtFinSF1	0.350758	0.102536	0.054670
7	BsmtFinSF2	0.054501	0.031371	0.004124
8	BsmtUnfSF	0.074481	0.015669	-0.000000
9	1stFlrSF	0.069237	0.119318	0.089430
10	GrLivArea	0.516287	0.327560	0.358186

3. Therefore, I have selected the Lasso regression model for the analysis.

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:
I've created the lasso model after excluding the five most important predictor variables of the original model in section 9.2 of the Jupyter Notebook and the 5 most important predictor variables in the new model are given below.



	Predictor	Lasso3 Coefficient
6	1stFlrSF	0.584448
3	BsmtFinSF1	0.247707
0	Const	0.191422
7	KitchenAbvGr	-0.144834
10	MSZoning_FV	0.134904
8	Functional	0.114928

- 1stFlrSF: Sale price increases when 1stFlrSF (first floor square feet) increases
- BsmtFinSF1: Sale price increases when BsmtFinSF1 (Type 1 finished square feet) increases
- KitchenAbvGr: Sale price decreases when KitchenAbvGr increases
- MSZoning_FV: Sale price is higher when MSZoning takes the value FV (Floating Village Residential)
- Functional: Sale price increases when functionality increases

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:
Robust and generalisable models

- A model is said to be robust and generalisable when it is able to make accurate predictions on unseen data after training on training data.
- In order to make sure that a model is robust and generalisable, we need to ensure that it learns the correct patterns from training data and does not underfit or overfit the training data.
- An underfitting model will have high bias and an overfitting model will have high variance. Thus, ensuring that the model is robust and generalisable involves balancing the bias and variance to maximise performance on unseen data.

Making linear regression models robust and generalisable

- Linear regression models in which the number of observations (n) is not much greater than the number of predictors (p) tend to have high variance after training and generally overfit the training data
- We can improve such models using regularisation, which decreases variance at the cost of a small increase in bias.
- Regularisation is of main types:
    - l1 regularisation (Lasso) - adds the l1 norm of the model as a penalty to the cost function, and
    - l2 regularisation (Ridge) - adds the l2 norm of the model as a penalty to the cost function.
- Both types of regularisation techniques tend to shrink the model coefficients towards zero and improve performance on unseen data.

Implications on accuracy
Since regularisation reduces overfitting, the accuracy of the model on training data decreases and its accuracy on unseen data increases.