Question 1

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

- I applied GridSearchCV and use the metrics **neg_mean_absolute_error** and **fold is 5** to compare the outcome of alpha value
- Detail

Ridge

	param_alpha	mean_train_score	mean_test_score	rank_test_score
28	500	-15345.398726	-19083.715866	1
27	400	-15039.765571	-19147.708658	2
26	300	-14721.445746	-19303.388477	3
29	1000	-16775.984771	-19445.293906	4
25	200	-14366.603623	-19600.294503	5
24	100	-13906.315422	-20142.523317	6
13	2.0	-11284.275322	-20473.832315	7
14	3.0	-11447.678347	-20489.310573	8
12	1.0	-11131.600797	-20525.159630	9
11	0.9	-11119.599847	-20539.505772	10
15	4.0	-11598.866959	-20540.224620	11
10	0.8	-11107.874553	-20556.233673	12
9	0.7	-11096.994882	-20575.349988	13
16	5.0	-11737.292152	-20590.589668	14
8	0.6	-11087.154466	-20596.371941	15
23	50	-13499.891401	-20607.124885	16
7	0.5	-11078.250823	-20620.888472	17
17	6.0	-11869.151236	-20632.412678	18
6	0.4	-11070.107849	-20652.657876	19
18	7.0	-11992.323961	-20672.424020	20
5	0.3	-11062.488244	-20695.320991	21
19	8.0	-12102.561041	-20708.783799	22
20	9.0	-12200.423373	-20742.071403	23
4	0.2	-11055.358909	-20744.884619	24
21	10.0	-12288.477424	-20774.767592	25
3	0.1	-11048.334480	-20802.232401	26
2	0.01	-11043.046273	-20861.396002	27
1	0.001	-11042.580064	-20869.130670	28
0	0.0001	-11042.534071	-20869.930161	29
22	20	-12862.425227	-20882.890489	30

Lasso

	param_alpha	mean_train_score	mean_test_score	rank_test_score
28	500	-13493.582026	-19047.664107	1
27	400	-13074.719609	-19065.603881	2
26	300	-12646.520253	-19178.253666	3
29	1000	-15226.803114	-19193.537792	4
25	200	-12084.445761	-19301.535950	5
24	100	-11565.597877	-19752.311978	6
23	50	-11301.528025	-19992.210141	7
22	20	-11139.345757	-20302.939946	8
21	10.0	-11089.505895	-20556.631647	9
20	9.0	-11084.487007	-20588.047875	10
19	8.0	-11079.463775	-20620.776030	11
18	7.0	-11074.927874	-20655.868936	12
17	6.0	-11070.085085	-20690.704933	13
16	5.0	-11065.765193	-20723.628562	14
15	4.0	-11061.332598	-20756.514626	15
14	3.0	-11056.840369	-20790.759999	16
13	2.0	-11052.449634	-20831.092655	17
12	1.0	-11048.500158	-20876.238113	18
11	0.9	-11048.122263	-20881.005941	19
10	0.8	-11047.718933	-20885.708986	20
9	0.7	-11047.297137	-20890.512291	21
8	0.6	-11046.902681	-20895.640383	22
7	0.5	-11046.547031	-20899.988260	23
6	0.4	-11046.180970	-20903.626572	24
5	0.3	-11045.777770	-20908.770236	25
4	0.2	-11045.404899	-20914.904017	26
3	0.1	-11045.060430	-20921.479622	27
2	0.01	-11044.796780	-20925.414451	28
1	0.001	-11044.776009	-20926.056700	29
0	0.0001	-11044.774367	-20926.122646	30

- Conclusion the best alpha
  - Ridge: 500
  - Lasso: 500

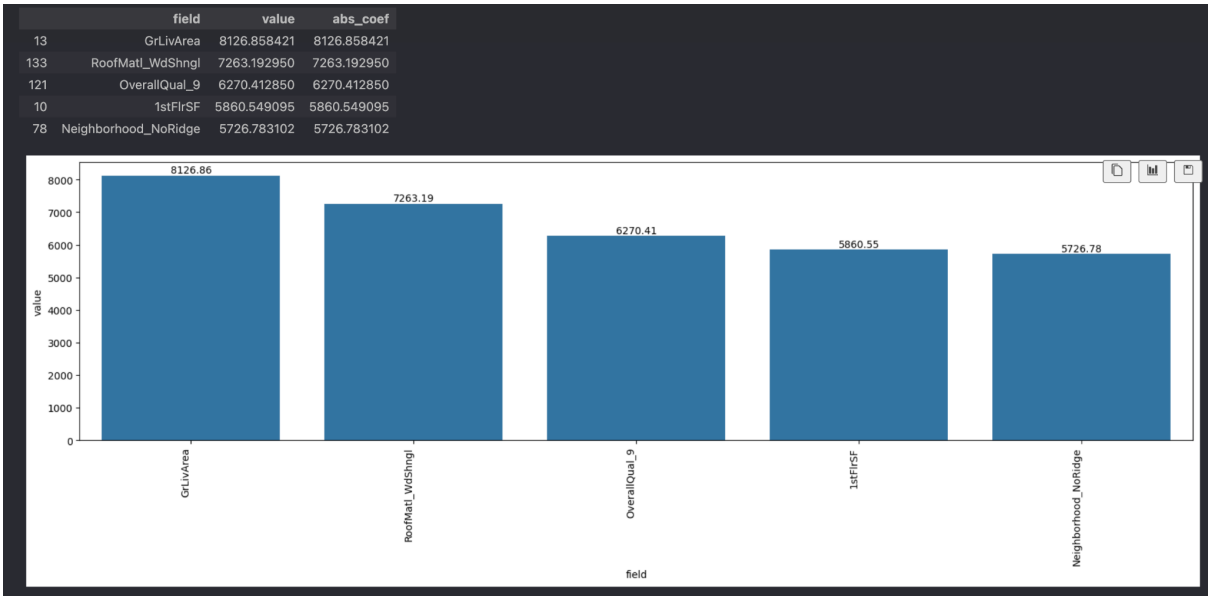
1.2 What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

	Before	**After**

Ridge Metrics	r2_train = 0.88 r2_test = 0.83 rss_train = 713428853122 rss_test = 414482032987.7721 mse_train = 739304511 mse_test = 1001164330	r2_train = 0.89 r2_test = 0.83 rss_train = 667779160766 rss_test = 424503240467 mse_train = 691999130 mse_test = 1025370146
Lasso Metrics	r2_train = 0.92 r2_test = 0.83 rss_train = 461885142905 rss_test = 429313320892 mse_train = 478637453 mse_test = 1036988697	r2_train = 0.89 r2_test = 0.82 rss_train = 667779160766 rss_test = 424503240467 mse_train = 691999130 mse_test = 1025370146

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

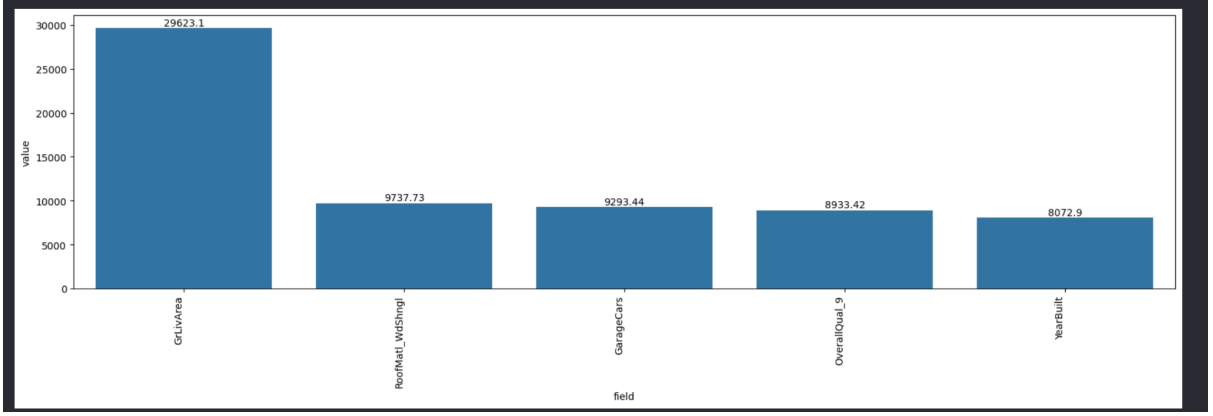
Ridge before:



	field	value	abs_coef
13	GrLivArea	8126.858421	8126.858421
133	RoofMatl_WdShngl	7263.192950	7263.192950
121	OverallQual_9	6270.412850	6270.412850
10	1stFlrSF	5860.549095	5860.549095
78	Neighborhood_NoRidge	5726.783102	5726.783102

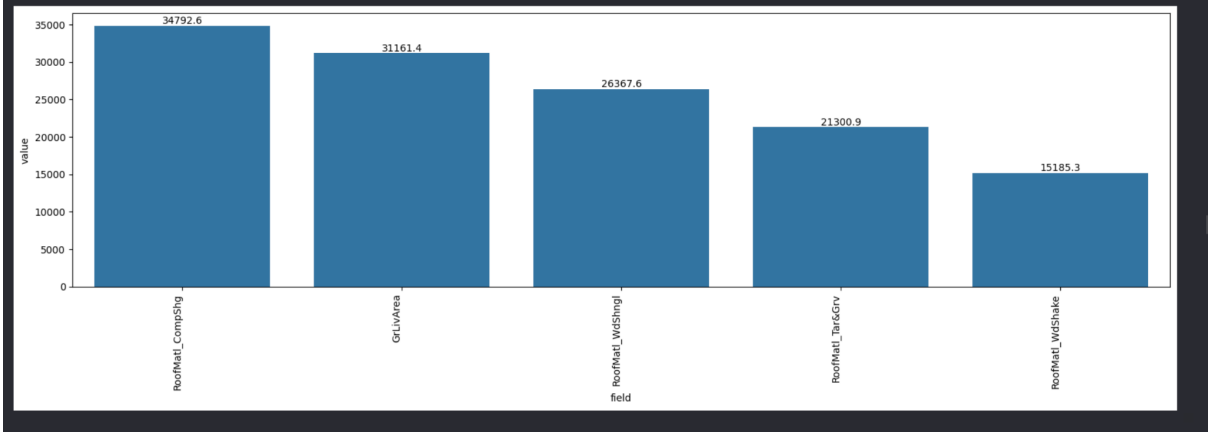
Ridge after:

	field	value	abs_coef
13	GrLivArea	29623.055167	29623.055167
133	RoofMatl_WdShngl	9737.734789	9737.734789
21	GarageCars	9293.442556	9293.442556
121	OverallQual_9	8933.421353	8933.421353
3	YearBuilt	8072.895015	8072.895015



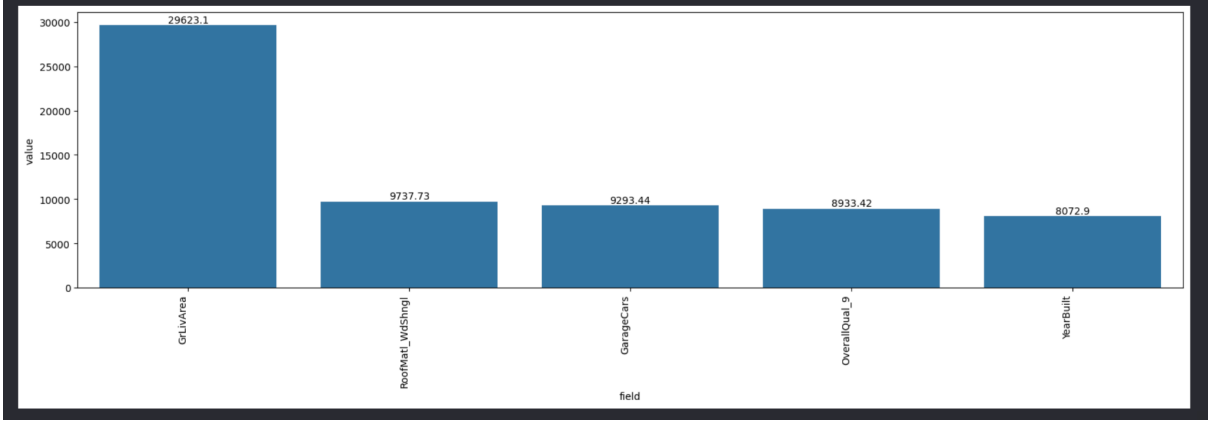
Lasso Before:

	field	value	abs_coef
127	RoofMatl_CompShg	34792.631011	34792.631011
13	GrLivArea	31161.381288	31161.381288
133	RoofMatl_WdShngl	26367.558380	26367.558380
131	RoofMatl_Tar&Grv	21300.871945	21300.871945
132	RoofMatl_WdShake	15185.305618	15185.305618



Lasso After:

	field	value	abs_coef
13	GrLivArea	29623.055167	29623.055167
133	RoofMatl_WdShngl	9737.734789	9737.734789
21	GarageCars	9293.442556	9293.442556
121	OverallQual_9	8933.421353	8933.421353
3	YearBuilt	8072.895015	8072.895015



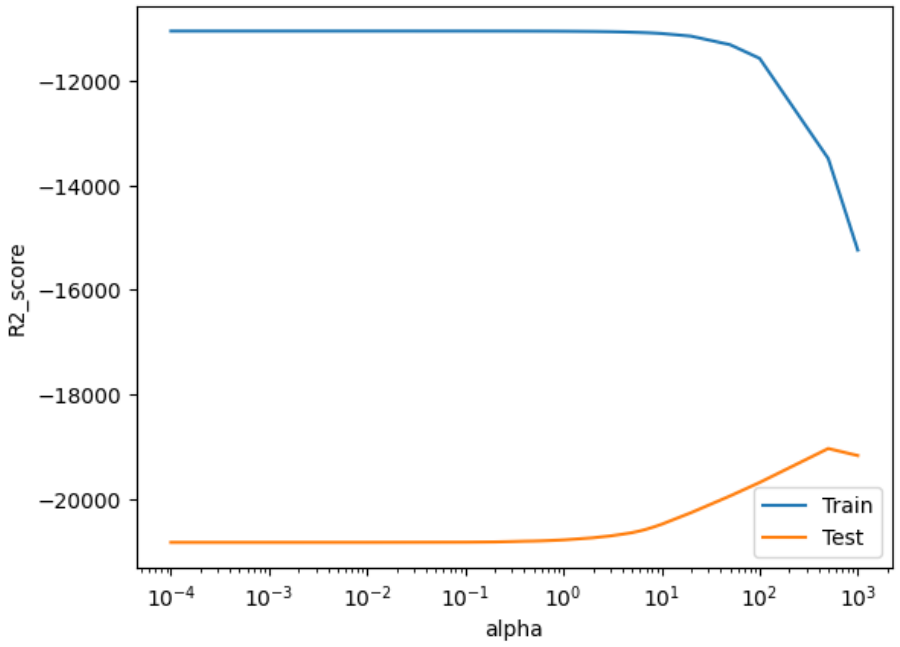
Question 2

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?

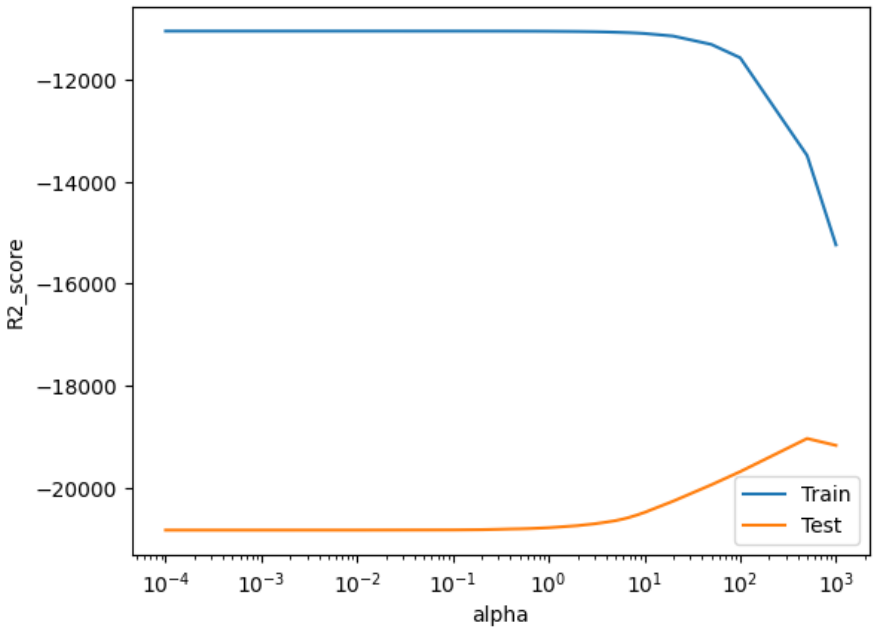
- We apply GridSearchCV to find out the alpha has the best mean_test_score and mean_train_score.
- When the score continues improving we will increase the alpha until the score decreases. Grid search is a hyperparameter tuning technique that performs an exhaustive search over a specified hyperparameter space to find the combination of hyperparameters that yields the best model performance.
- Grid search then systematically explores every possible combination of hyperparameters from the parameter grid. It fits and evaluates the model for each combination using cross-validation and selects the combination that yields the best performance.

Ridge:

	param_alpha	mean_train_score	mean_test_score	rank_test_score
25	500	-15362.095904	-19071.852779	1
26	1000	-16798.276076	-19454.338394	2
24	100	-13908.741846	-20149.319648	3
13	2.0	-11289.767516	-20442.125180	4
14	3.0	-11452.505449	-20466.620546	5
12	1.0	-11139.289406	-20470.172155	6
11	0.9	-11127.528259	-20480.565438	7
10	0.8	-11116.218180	-20492.887605	8
9	0.7	-11105.269934	-20508.960391	9
15	4.0	-11599.710139	-20519.799907	10
8	0.6	-11094.898968	-20527.999709	11
7	0.5	-11085.867079	-20550.647632	12
16	5.0	-11735.846419	-20571.196428	13
6	0.4	-11077.565786	-20580.092162	14
5	0.3	-11069.941823	-20617.119161	15
17	6.0	-11864.758218	-20618.629703	16
23	50	-13500.575094	-20629.944436	17
4	0.2	-11063.067774	-20659.003115	18
18	7.0	-11985.350599	-20663.158122	19
19	8.0	-12094.699141	-20702.916409	20
3	0.1	-11057.407252	-20707.747173	21
20	9.0	-12193.172209	-20736.658976	22
2	0.01	-11052.977818	-20760.646239	23
21	10.0	-12281.983409	-20766.003717	24
1	0.001	-11052.536649	-20766.580197	25
0	0.0001	-11052.494444	-20767.180896	26
22	20	-12859.306155	-20883.894830	27



Lasso



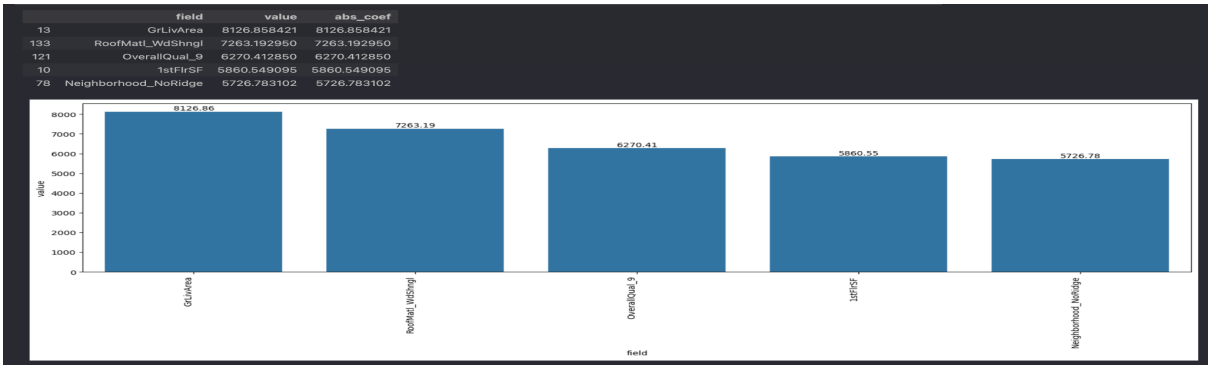
	param_alpha	mean_train_score	mean_test_score	rank_test_score
25	500	-13492.152795	-19036.503260	1
26	1000	-15243.188318	-19171.220276	2
24	100	-11576.918963	-19685.407220	3
23	50	-11315.228553	-19942.976567	4
22	20	-11151.867740	-20263.148581	5
21	10.0	-11100.966241	-20485.539988	6
20	9.0	-11095.812870	-20514.905129	7
19	8.0	-11090.730362	-20546.694864	8
18	7.0	-11085.958706	-20579.897983	9
17	6.0	-11081.550472	-20613.548052	10
16	5.0	-11076.850917	-20645.116960	11
15	4.0	-11071.943279	-20674.907048	12
14	3.0	-11066.995489	-20707.702456	13
13	2.0	-11062.319153	-20743.830019	14
12	1.0	-11057.884548	-20784.018614	15
11	0.9	-11057.463621	-20788.159586	16
10	0.8	-11057.013948	-20792.284209	17
9	0.7	-11056.563067	-20796.939260	18
8	0.6	-11056.230091	-20801.756427	19
7	0.5	-11055.882388	-20805.891954	20
6	0.4	-11055.540176	-20809.086760	21
5	0.3	-11055.189617	-20815.407172	22
4	0.2	-11054.834609	-20822.514856	23
3	0.1	-11054.508941	-20827.151660	24
2	0.01	-11054.272039	-20830.256372	25
1	0.001	-11054.254306	-20830.575472	26
0	0.0001	-11054.252996	-20830.621933	27

Question 3

After building the model, you realized 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.

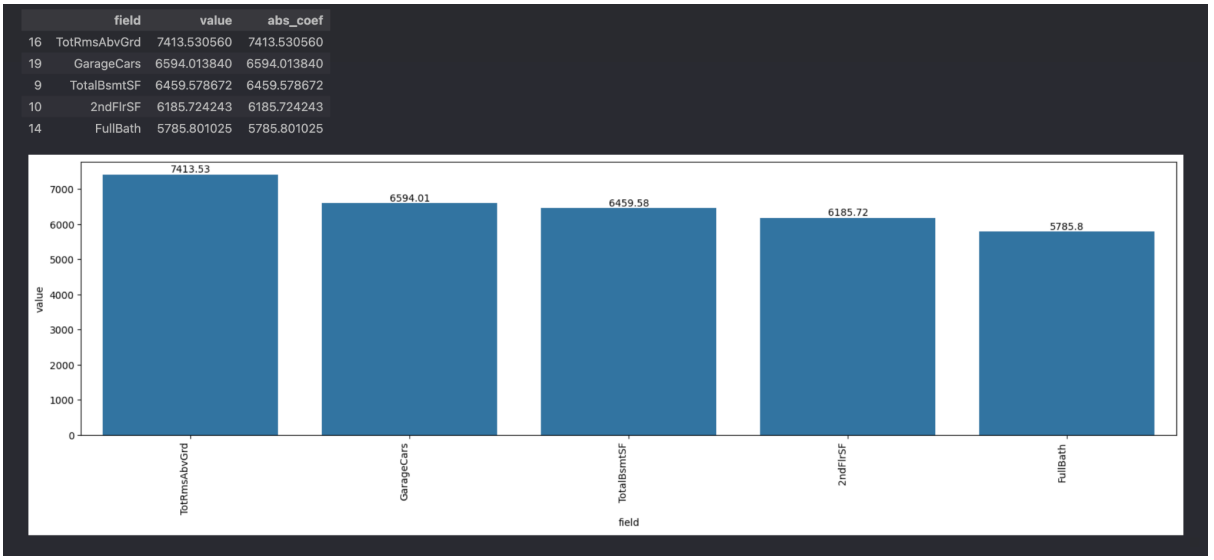
Answer:

1. Ridge top 5 features in the first training with the best alpha (500) include:



	field	value	abs_coef
13	GrLivArea	8126.858421	8126.858421
133	RoofMatl_WdShngl	7263.192950	7263.192950
121	OverallQual_9	6270.412850	6270.412850
10	1stFlrSF	5860.549095	5860.549095
78	Neighborhood_NoRidge	5726.783102	5726.783102

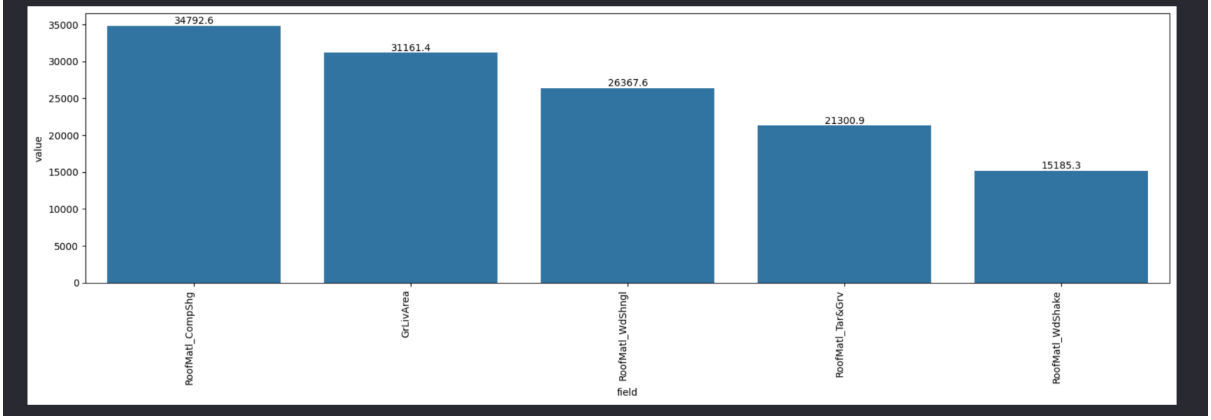
After excluding the top 5 features we have the new top 5 features



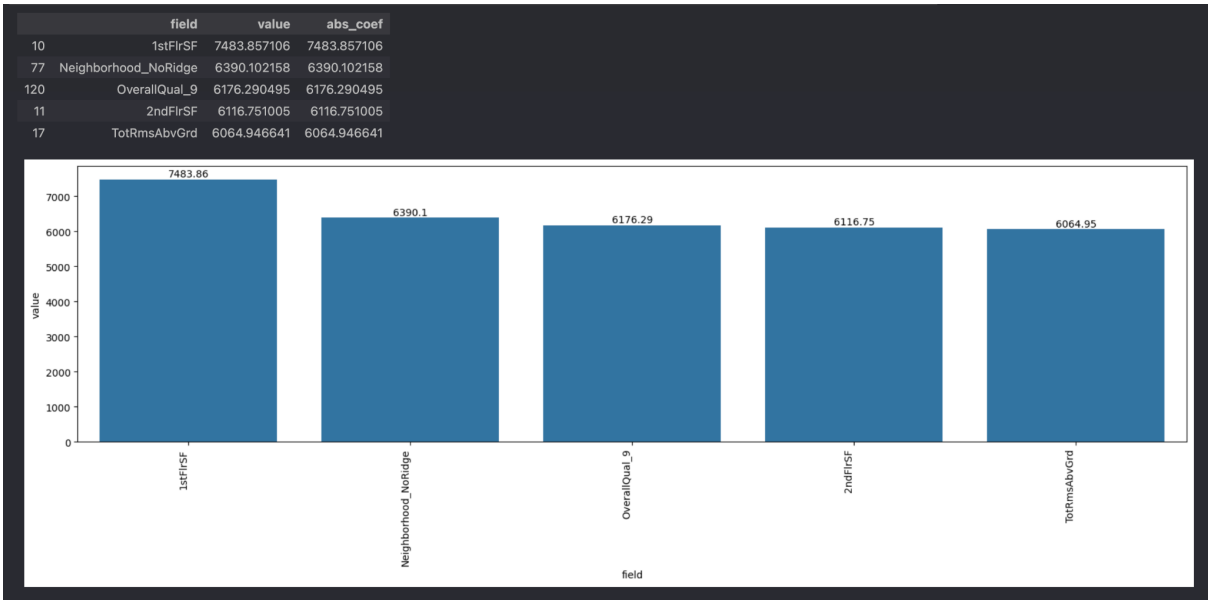
	field	value	abs_coef
16	TotRmsAbvGrd	7413.530560	7413.530560
19	GarageCars	6594.013840	6594.013840
9	TotalBsmtSF	6459.578672	6459.578672
10	2ndFlrSF	6185.724243	6185.724243
14	FullBath	5785.801025	5785.801025

2. Lasso top 5 features in the first training with the best alpha (500) include:

	field	value	abs_coef
127	RoofMatl_CompShg	34792.631011	34792.631011
13	GrLivArea	31161.381288	31161.381288
133	RoofMatl_WdShngl	26367.558380	26367.558380
131	RoofMatl_Tar&Grv	21300.871945	21300.871945
132	RoofMatl_WdShake	15185.305618	15185.305618



	field	value	abs_coef
10	1stFlrSF	7483.857106	7483.857106
77	Neighborhood_NoRidge	6390.102158	6390.102158
120	OverallQual_9	6176.290495	6176.290495
11	2ndFlrSF	6116.751005	6116.751005
17	TotRmsAbvGrd	6064.946641	6064.946641



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?

I think we can consider a model robust when they are not impacted by outliers in the training data and any variations in the test dataset will not affect its performance too much. To achieve this, we have to **make the model not overfit or underfit.** This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data. On the other hand, we have to find a way to balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.