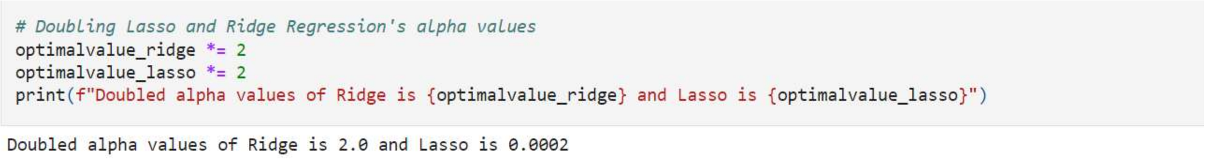
1. What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Based on the house Python code implementation, these are the optimal values of alpha

The optimal value of the Ridge is 0.2

The optimal value of the Lasso is 0.0001

Below are the results after doubling the alpha values for Ridge and Lasso in the Python code implementation



```
# Doubling Lasso and Ridge Regression's alpha values
optimalvalue_ridge *= 2
optimalvalue_lasso *= 2
print(f"Doubled alpha values of Ridge is {optimalvalue_ridge} and Lasso is {optimalvalue_lasso}")
```

```
Doubled alpha values of Ridge is 2.0 and Lasso is 0.0002
```

```
# Doubling Lasso and Ridge Regression's alpha values
optimalvalue_ridge *= 2
optimalvalue_lasso *= 2
print(f"Doubled alpha values of Ridge is {optimal-
value_ridge} and Lasso is {optimalvalue_lasso}")
```

```
Doubled alpha values of Ridge is 0.4 and Lasso is 0.0002
```

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?

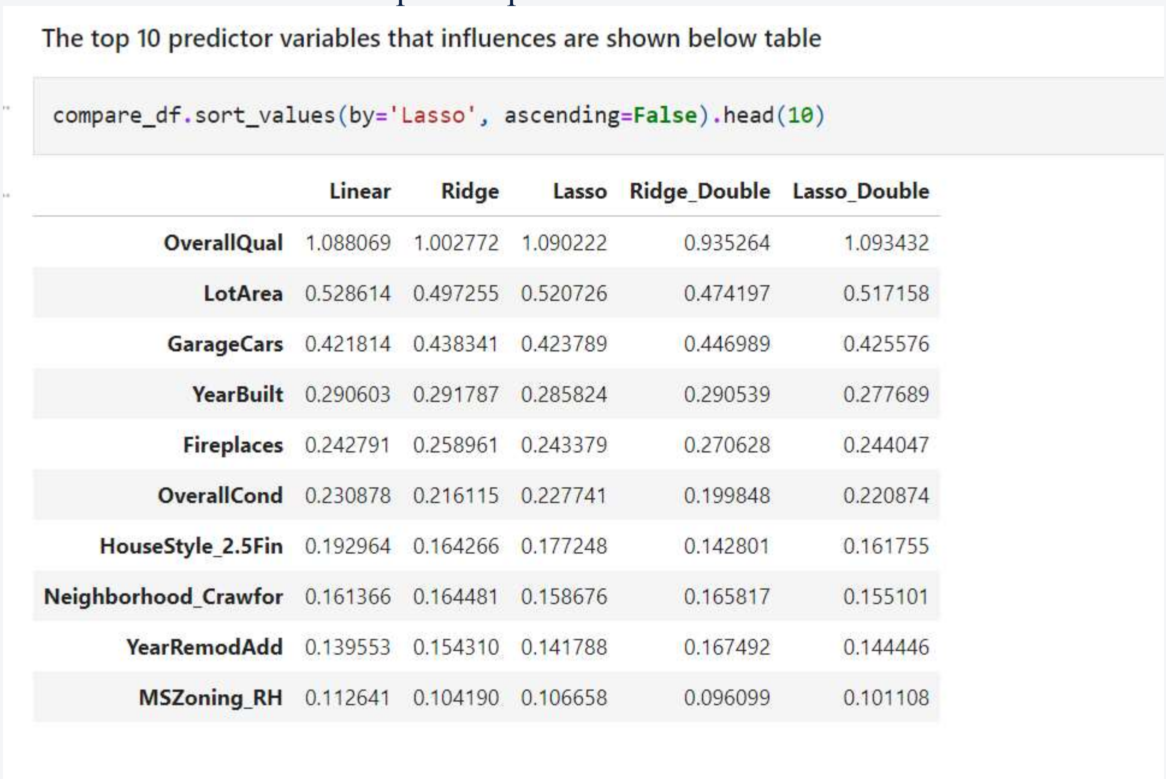
Ans: According to my Python code implementation, the final metrics for Ridge, Lasso, Double Ridge, and Double Lasso R2 Scores, RSS, and MSE Train & Test values are as follows:

Metric	Linear Regression	Ridge Regression	Lasso Regression	Double Ridge Regression	Double Lasso Regression	
0	R2 Score (Train)	0.840025	0.838955	0.839560	0.836908	0.838862
1	R2 Score (Test)	0.803322	0.806939	0.807843	0.806255	0.811618
2	RSS (Train)	22.315495	22.464738	22.380375	22.750261	22.477741
3	RSS (Test)	24.558512	10.818806	10.768116	10.857134	10.556612
4	MSE (Train)	0.158169	0.158697	0.158399	0.159702	0.158743
5	MSE (Test)	0.169637	0.168070	0.167676	0.168367	0.166021

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?



The top 10 predictor variables that influences are shown below table

```
compare_df.sort_values(by='Lasso', ascending=False).head(10)
```

	Linear	Ridge	Lasso	Ridge_Double	Lasso_Double
OverallQual	1.088069	1.002772	1.090222	0.935264	1.093432
LotArea	0.528614	0.497255	0.520726	0.474197	0.517158
GarageCars	0.421814	0.438341	0.423789	0.446989	0.425576
YearBuilt	0.290603	0.291787	0.285824	0.290539	0.277689
Fireplaces	0.242791	0.258961	0.243379	0.270628	0.244047
OverallCond	0.230878	0.216115	0.227741	0.199848	0.220874
HouseStyle_2.5Fin	0.192964	0.164266	0.177248	0.142801	0.161755
Neighborhood_Crawfor	0.161366	0.164481	0.158676	0.165817	0.155101
YearRemodAdd	0.139553	0.154310	0.141788	0.167492	0.144446
MSZoning_RH	0.112641	0.104190	0.106658	0.096099	0.101108

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?

- Make your model resistant to outliers by making the following changes:

A regression-based model is less affected by outliers than a tree-based model. Non-parametric tests are better than parametric tests when performing a statistical test.

- The error metric should be more robust.: Using mean absolute difference (or something like Huber Loss) reduces the influence of outliers. The median is a measure of central tendency for a variety of reasons, which I explain here Why is the median a measure of central tendency? There is no relationship between it and any other values of the data set, so how does it "describe" the data set?

The following changes can be made to your data:

- Winsorize it. You can artificially cap your data at a certain level. How can winsorization be applied? Data transformation.
- Try a log transformation if your data has a very pronounced right tail. Identify the outliers and remove them. Obviously, this works if there are a very small number of them and you are fairly sure they are anomalies and not worth predicting.