

DA5401 Assignment 3 – MM21B051

Task 1:

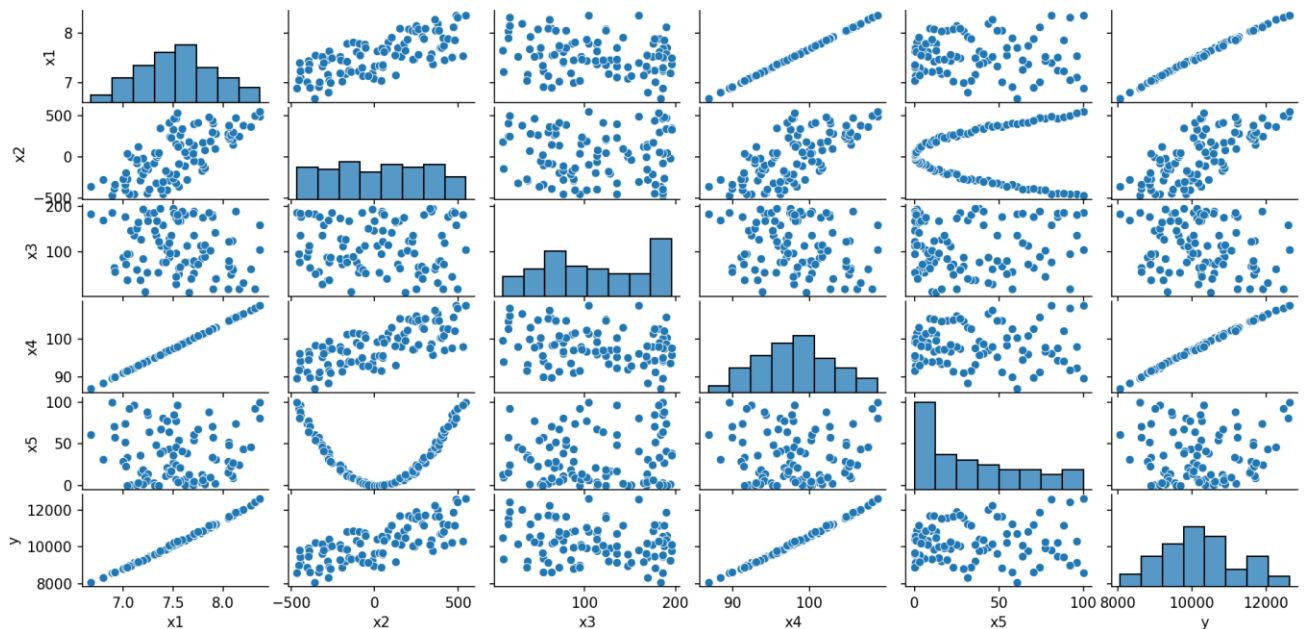
I split the data into 80% training and 20% test. Upon training the model directly using OLS and testing it on the test data set we get a Loss (SSE): 42022.69

Task 2:

Correlation of y with the features:

```
[[ True True False True False True]
 [ True True False True False True]
 [False False True False False False]
 [ True True False True False True]
 [False False False False True False]
 [ True True False True False True]]
```

Pairplot showing correlation between features and y



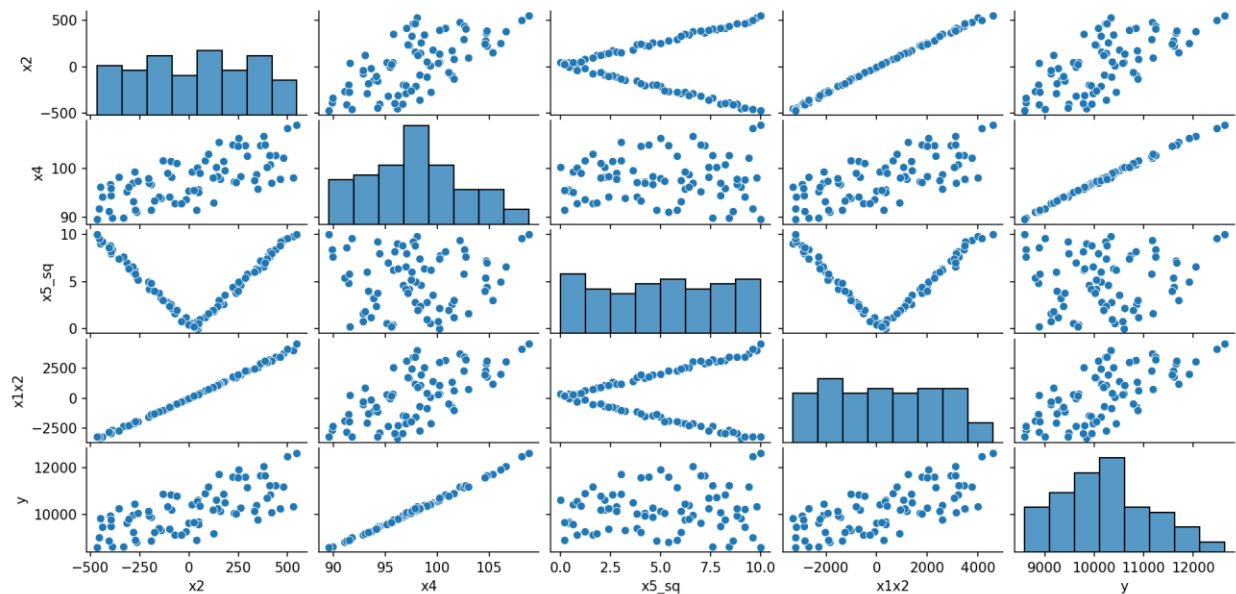
We notice that x_1 and x_4 are linearly correlated: x_4 is 13 times x_1 , hence including both x_1 and x_4 would be redundant. Therefore we remove x_1 though it is linearly correlated with y .

We notice that y is linearly correlated with x_2 also and hence retain it. x_5 seems to be the square of x_2 and x_5 is not linearly correlated with y , therefore we use $\sqrt{x_5}$ as a feature.

x_3 is neither correlated to y nor to any other feature, therefore we remove it as it could be a feature with no relation to y .

Upon further exploration of combinations and exponents of the features we find that $x_1 \cdot x_2$ is linearly correlated to y as can be seen from the correlation matrix.

Below is a plot illustrating the relationships of the chosen features and y



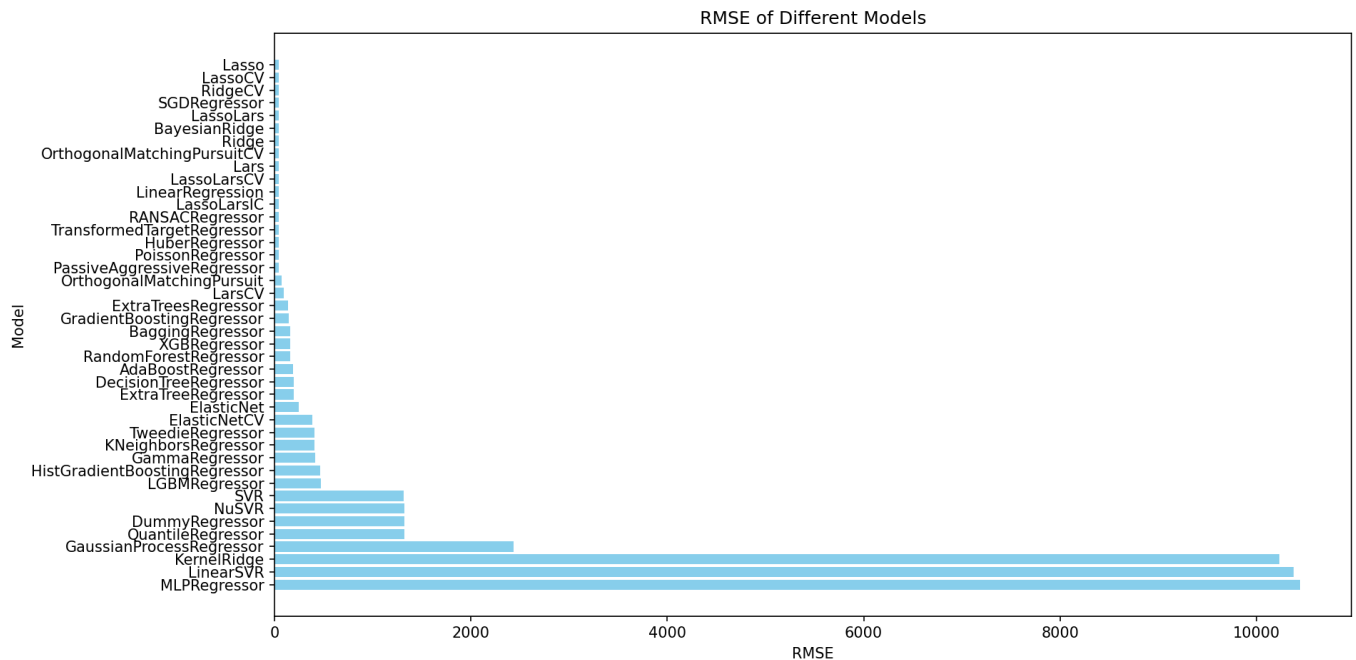
Task 3

The loss obtained upon training the model using the above features [x_2 , x_4 , $\sqrt{x_5}$, $x_1 \cdot x_2$] using OLS, and testing it on the test data set is:

Loss : 3987.3819494052077

The loss has significantly decreased, by 90%.

Task 4



Upon performing LazyRegressor, we find that the RMSE for LASSO is 43 and for OLS without any feature transformation it was 44.

RMSE after performing OLS on the new dataset after feature transformation we get the RMSE to be 13 which is lower. While LASSO only tries to achieve a sparse solution, we have eliminated all the irrelevant and redundant features and added features of higher order which are correlated to y , owing to the lower RMSE.