

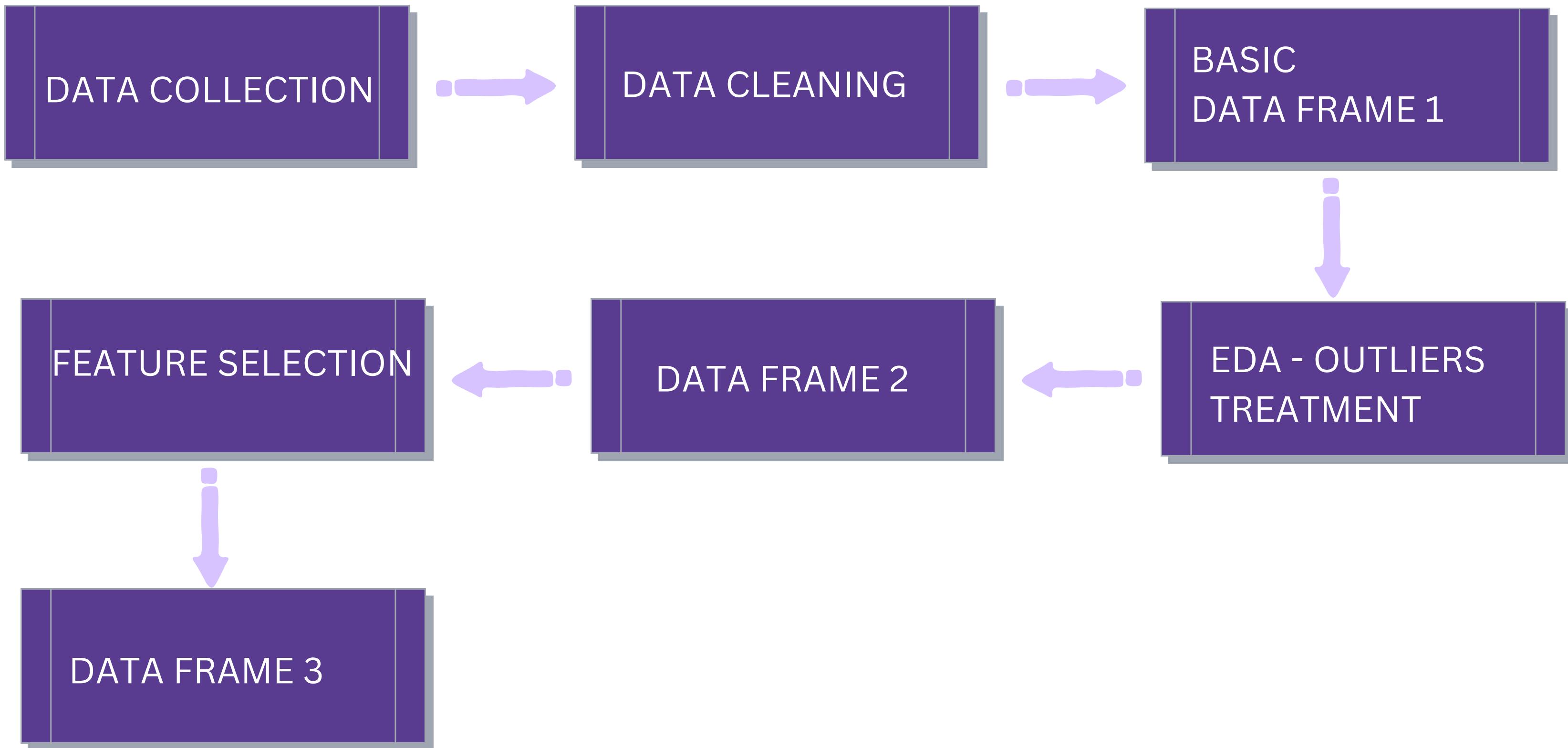
CALORIES BURNT PREDICTION

SAITEJA GURRAM

PROBLEM STATEMENT

Develop a machine learning model which can predict the number of calories a person has burnt during physical activities

FLOW CHART



DATASET INSIGHTS

DATA SHAPE

- This dataset contains 15000 rows and 8 columns

DATASET COLUMNS

- The target column of this data is “Calories” which has numerical data
- There are 7 independent columns such as Age, Height, Weight, Duration, Heart_Rate, Body_Temp

DATA CLEANING

- Data type conversion

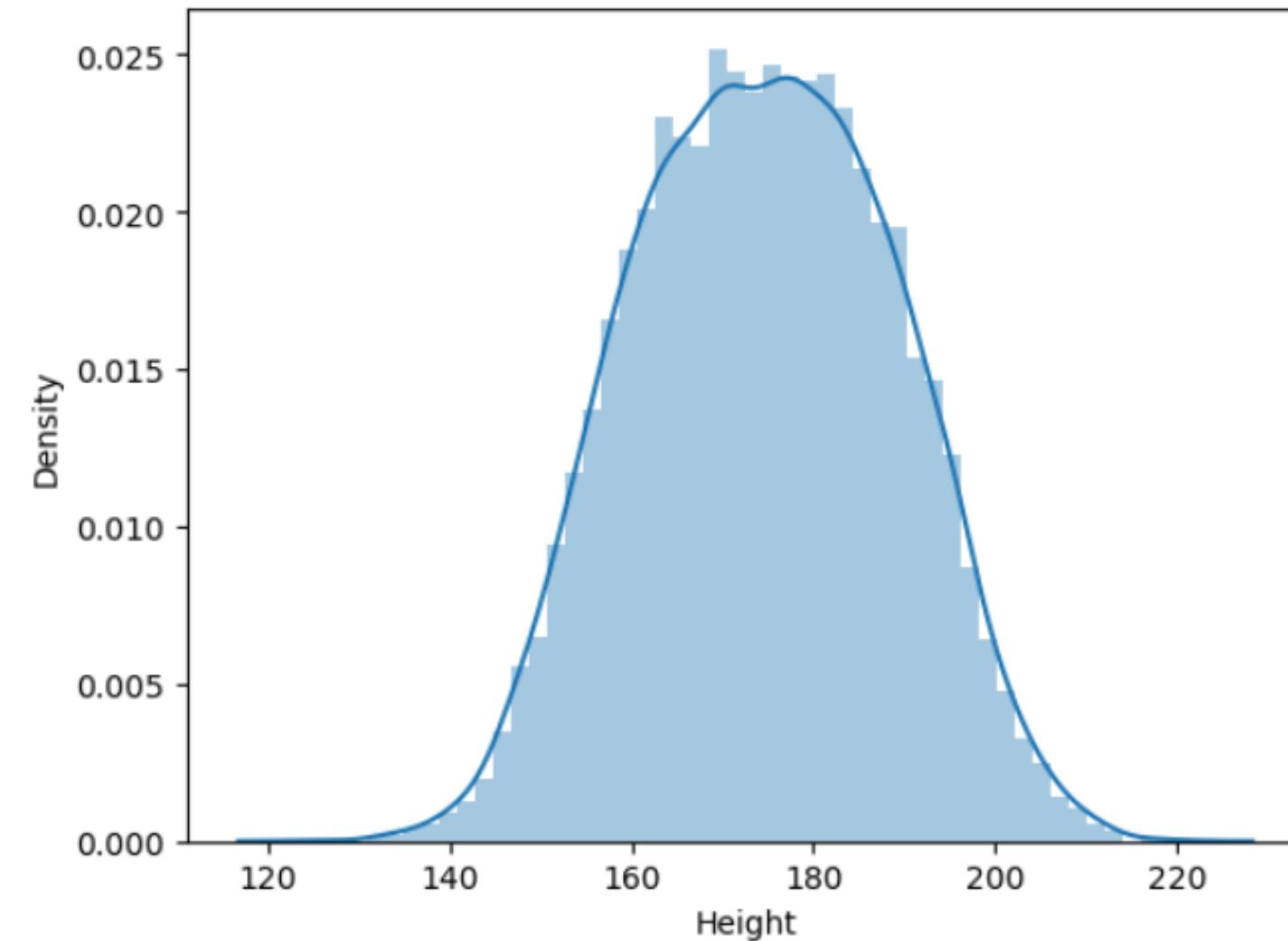
BASIC DATAFRAME

Index	Modelname	R-square	Adj-R2	MSE
0	Linear	0.967477	0.967458	132.783854
1	Lasso	0.964716	0.964696	144.254756
2	Ridge	0.967477	0.967458	132.785602
3	Adaboost Regressor	0.963617	0.963595	131.632058
4	Gradient Boost Regressor	0.997003	0.997001	13.740505
5	RF-Regressor	0.999698	0.999698	9.347478
6	DT Regressor	1.000000	1.000000	30.511000
7	XGB Regressor	0.999580	0.999580	4.634860

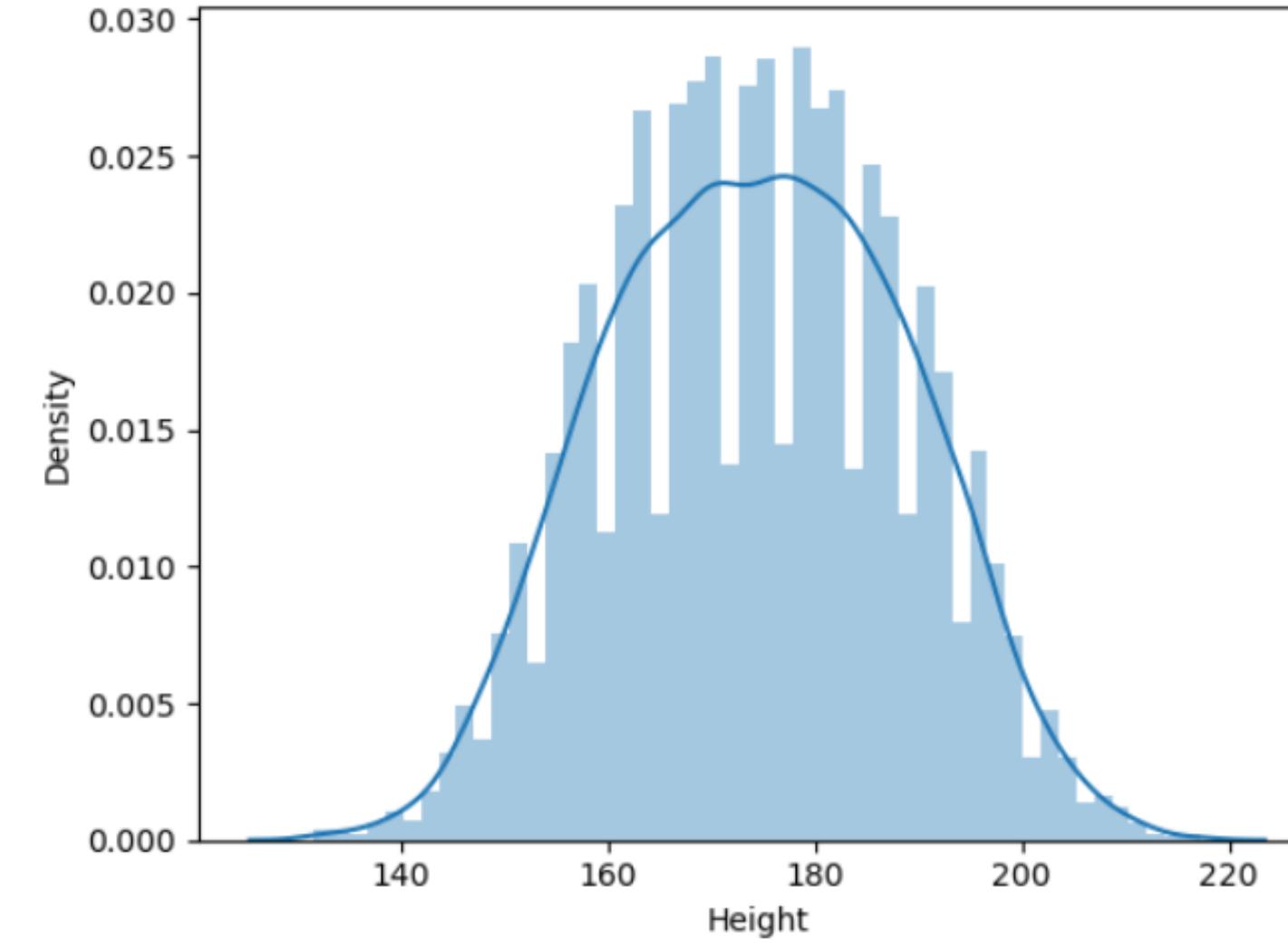
- The conclusion from the above data frame is XGB Regressor is best suited for my data .

OUTLIER TREATMENT

Before Treating the Outliers

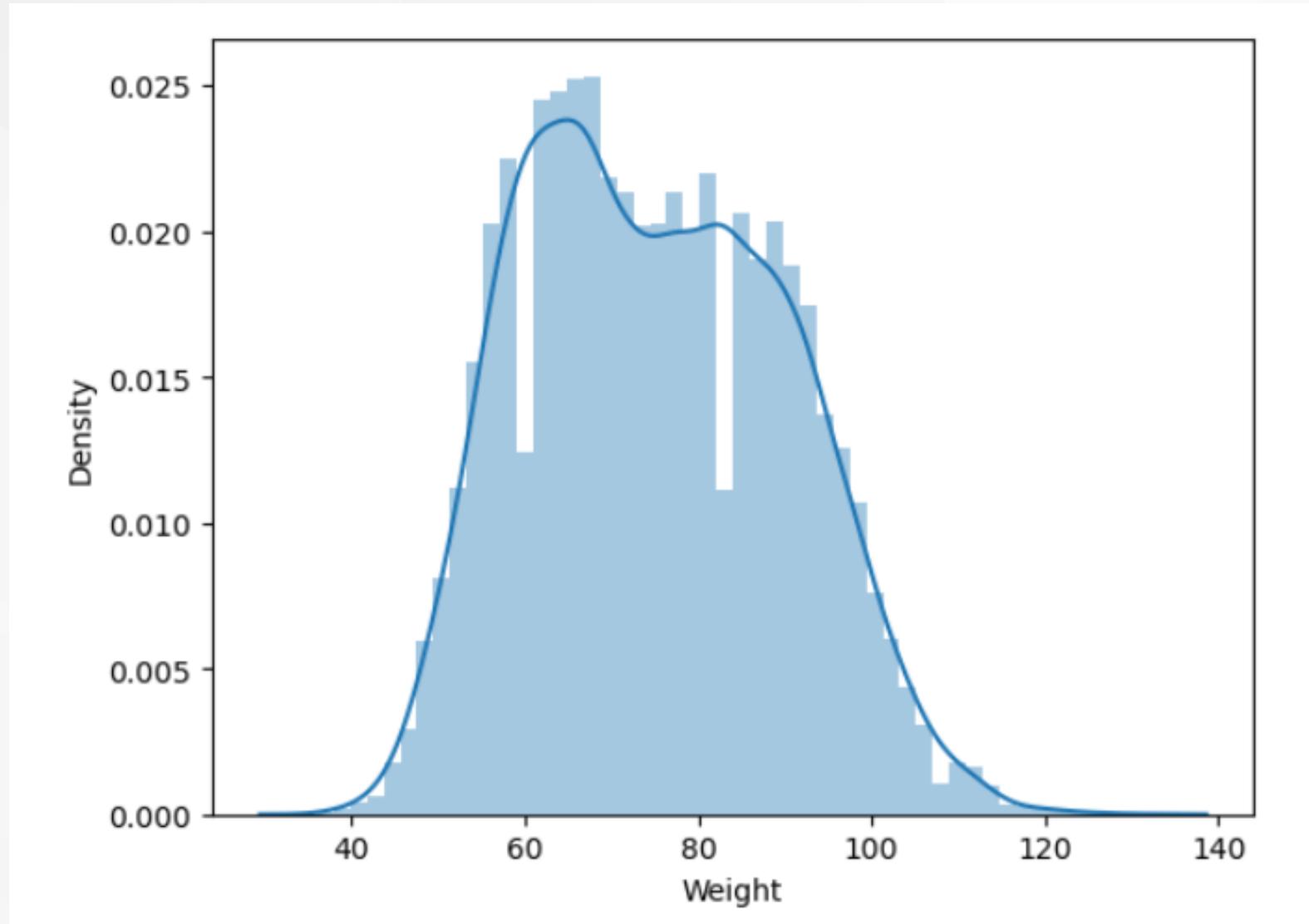


After Treating the Outliers

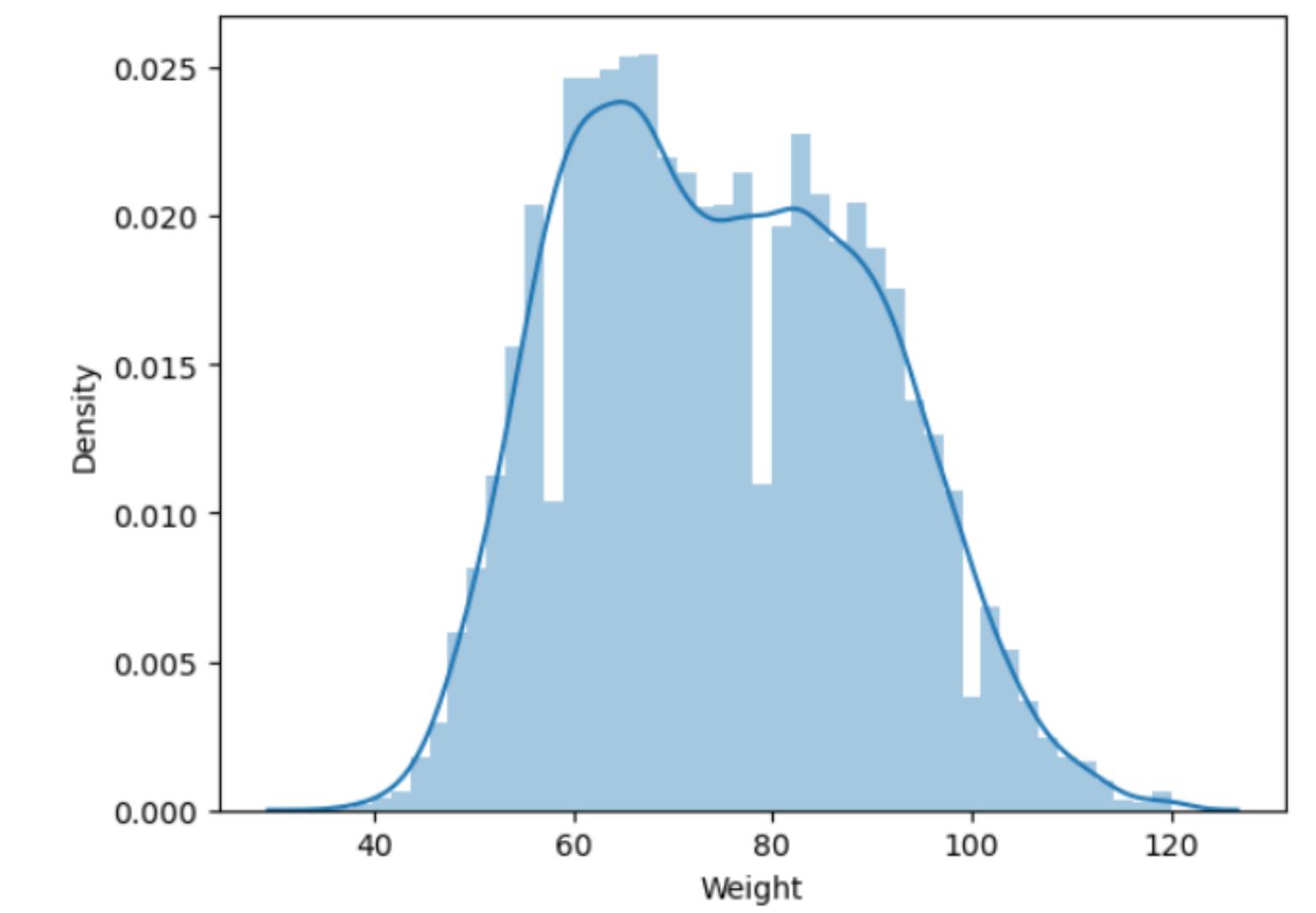


OUTLIER TREATMENT

Before Treating the Outliers

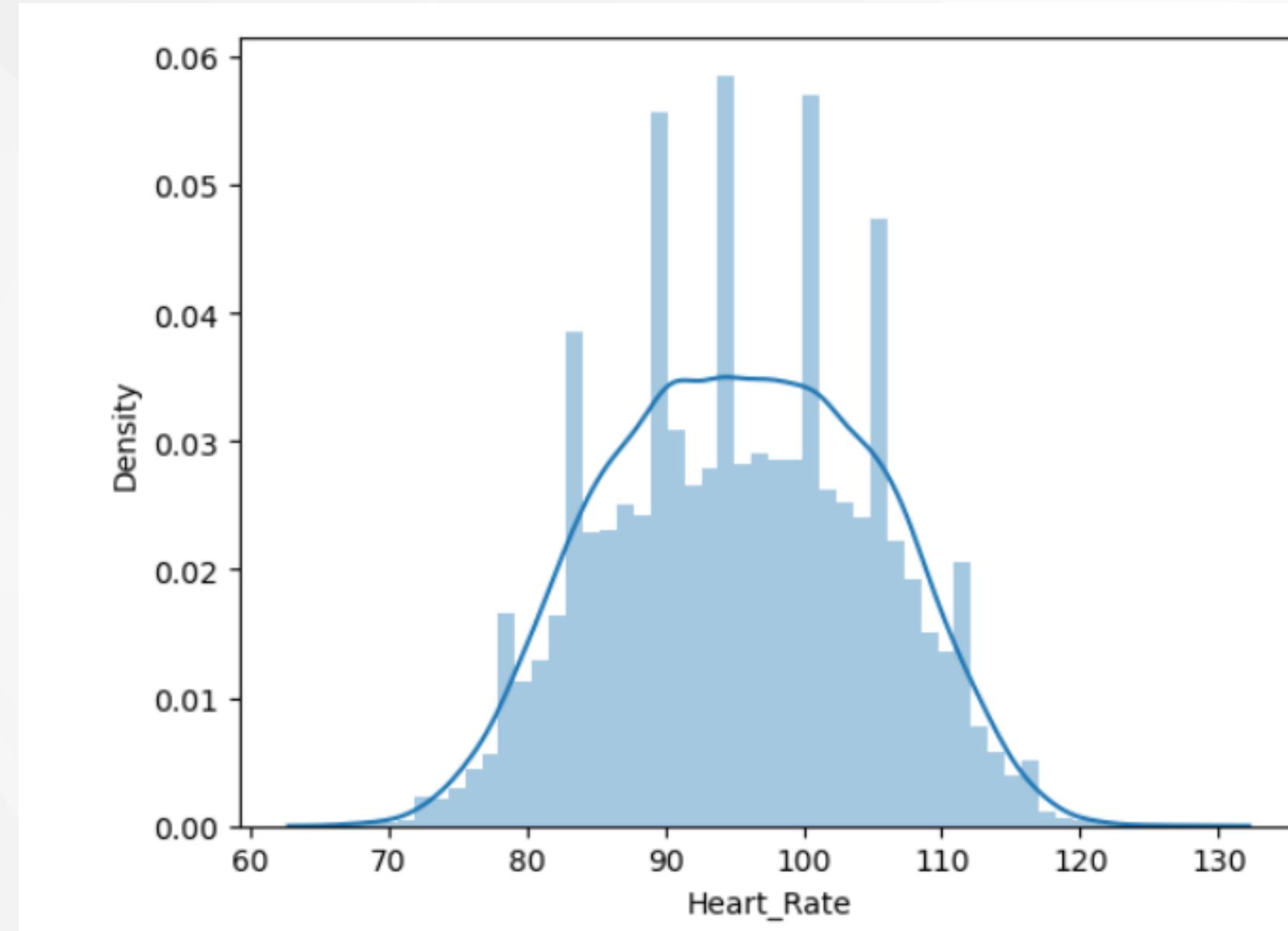


After Treating the Outliers

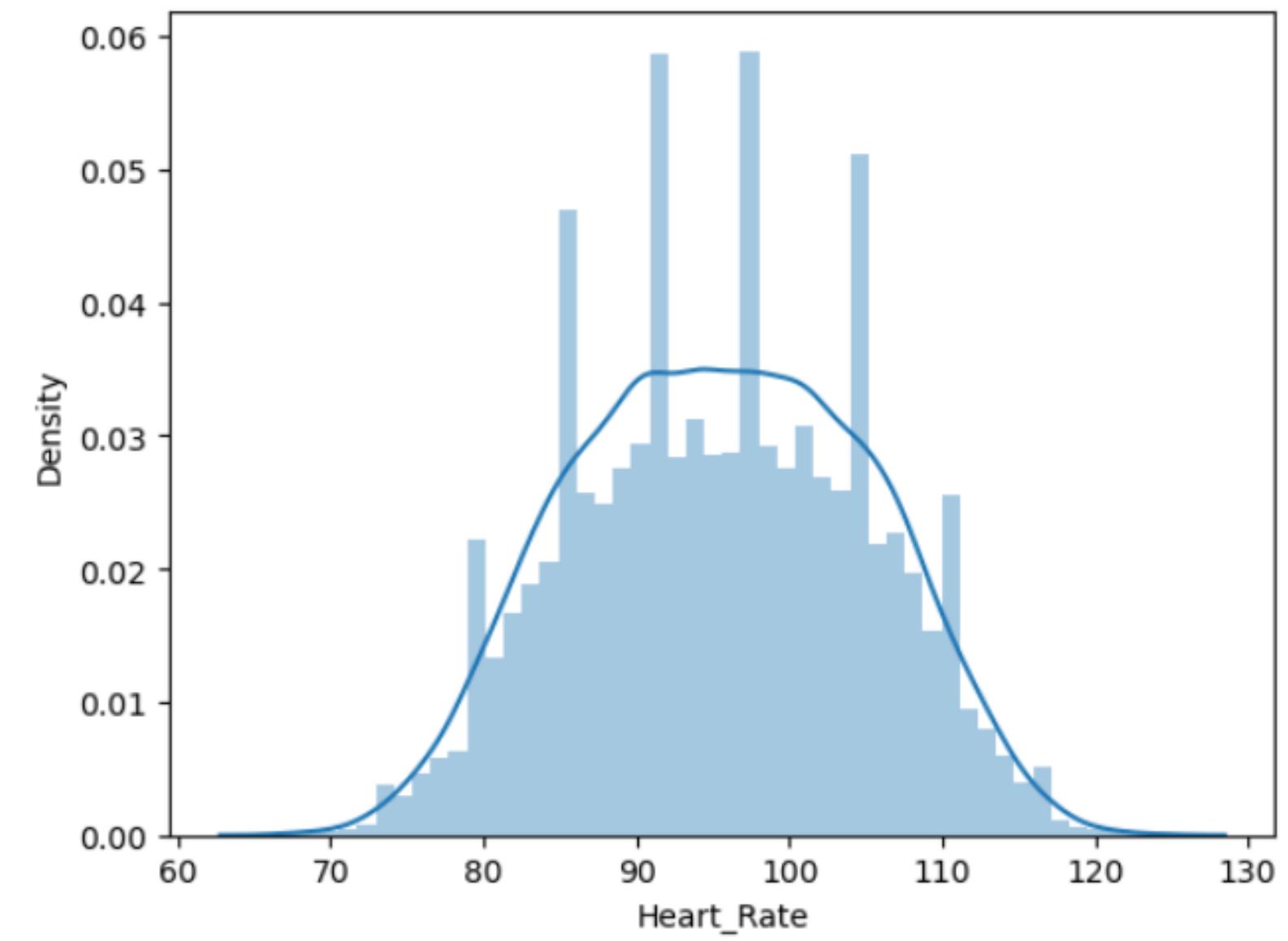


OUTLIER TREATMENT

Before Treating the Outliers

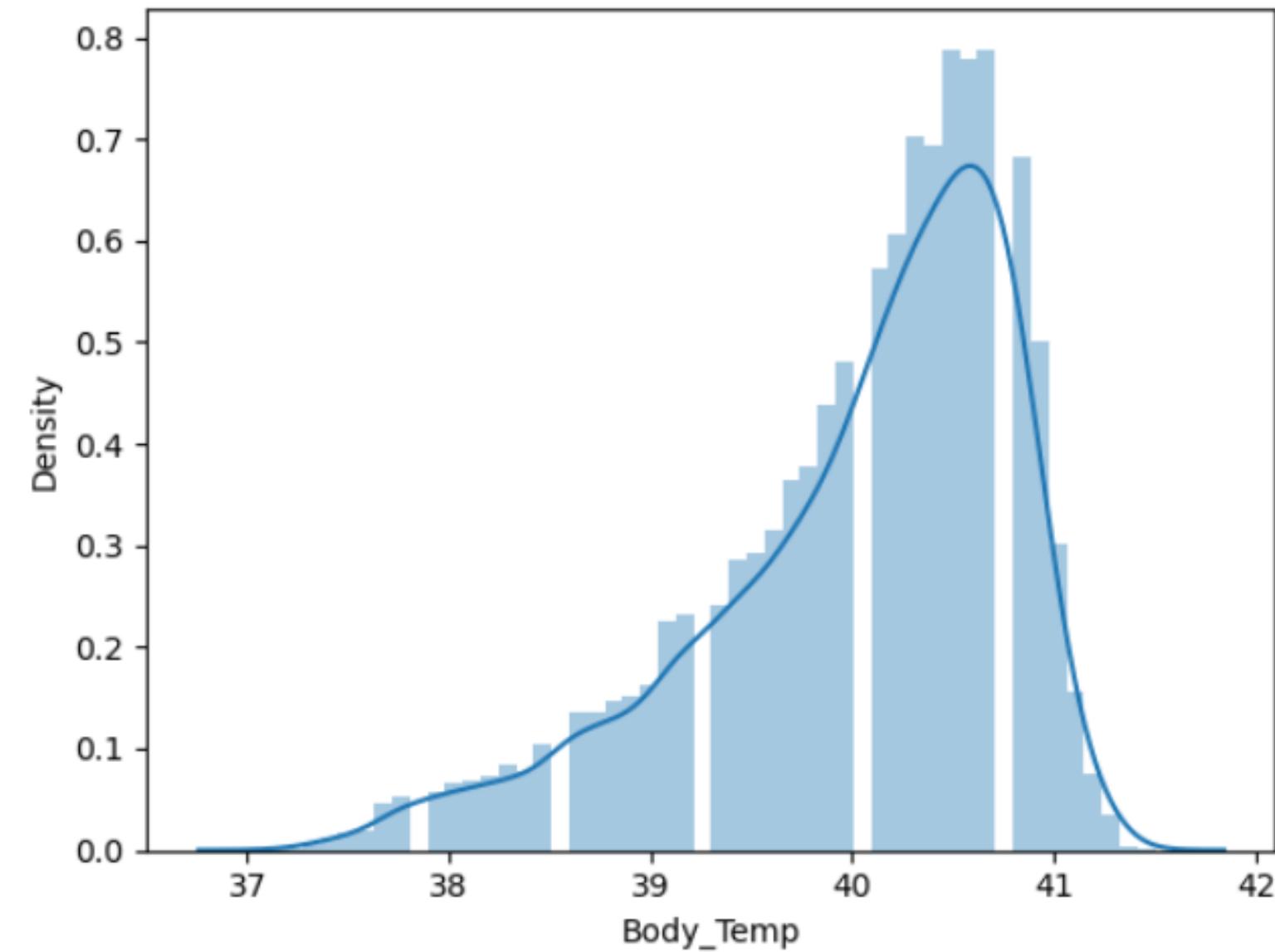


After Treating the Outliers

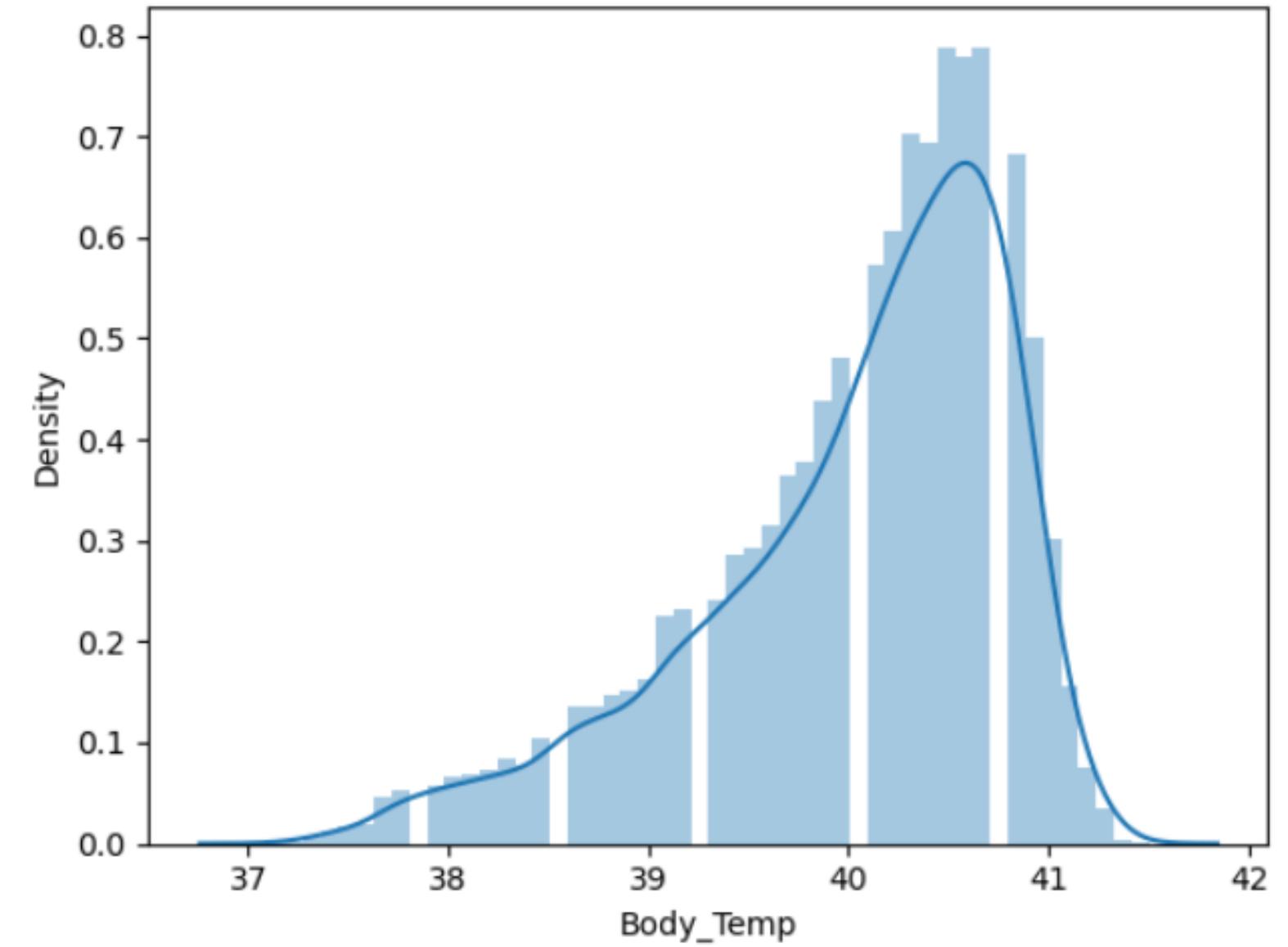


OUTLIER TREATMENT

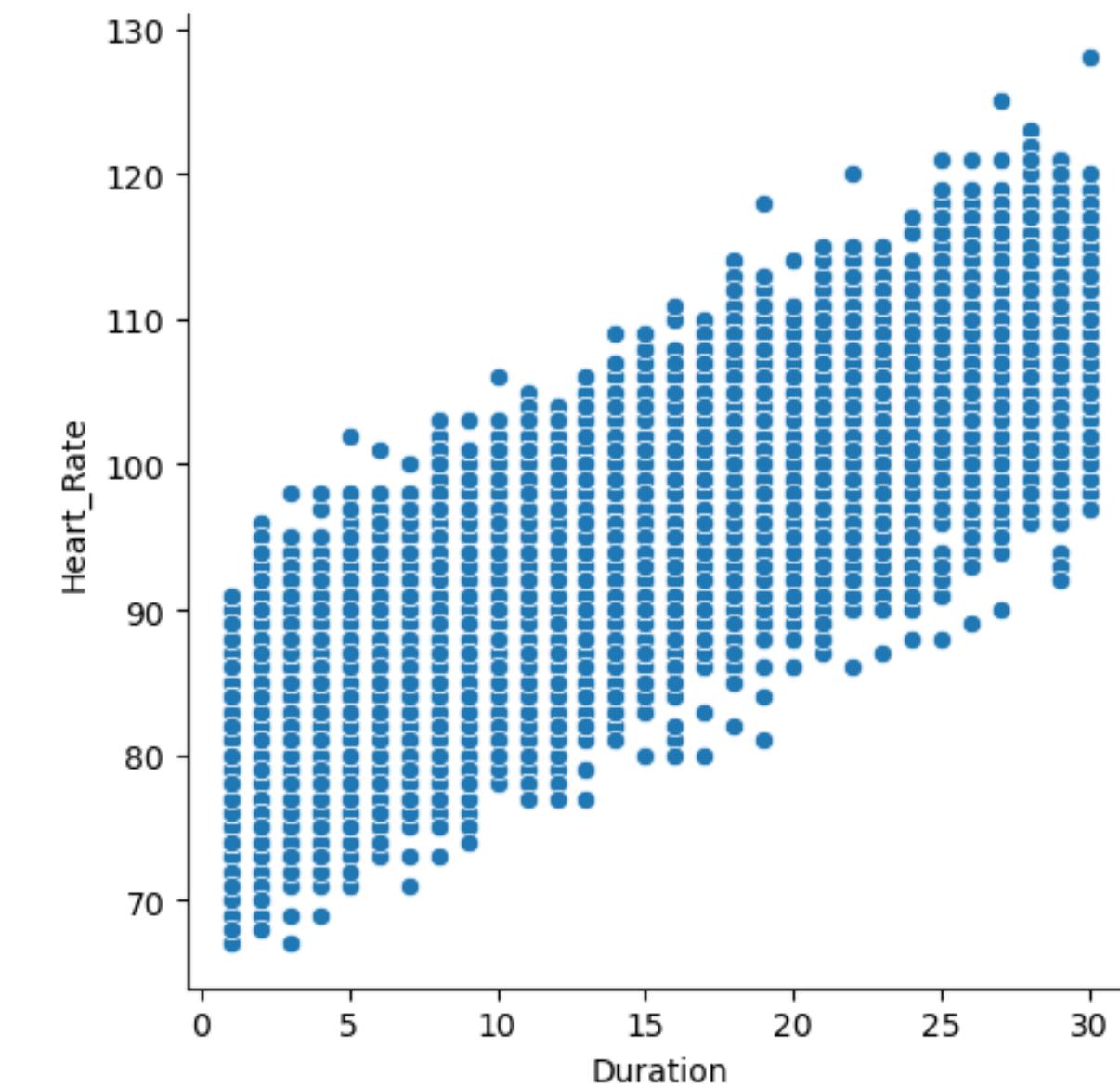
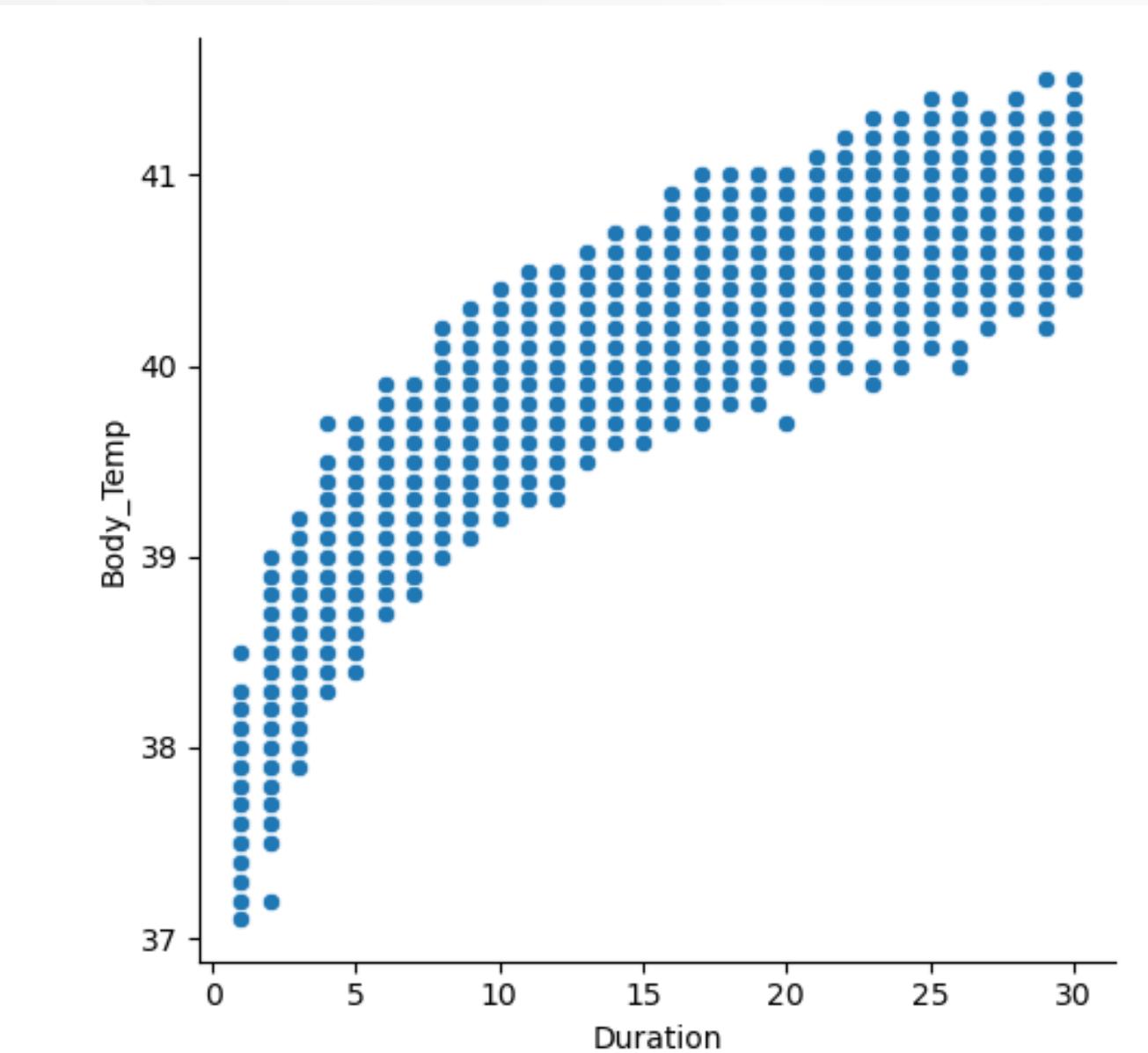
Before Treating the Outliers



After Treating the Outliers



VISUALIZATION



SECOND DATFRAME

Index	Modelname	R-square	Adj-R2	MSE
0	Linear	0.967468	0.967449	133.094162
1	Lasso	0.964707	0.964696	144.581303
2	Ridge	0.967468	0.967458	133.095924
3	Adaboost Regressor	0.965797	0.965777	125.465376
4	Gradient Boost Regressor	0.997003	0.997001	13.740505
5	RF-Regressor	0.999685	0.999685	9.569155
6	DT Regressor	1.000000	1.000000	29.789000
7	XGB Regressor	0.999580	0.999580	4.895917

- The conclusion from the above data frame is that after EDA, MSE has been decreased slightly but still most of the models are performing overfitting.

FEATURE SELECTION

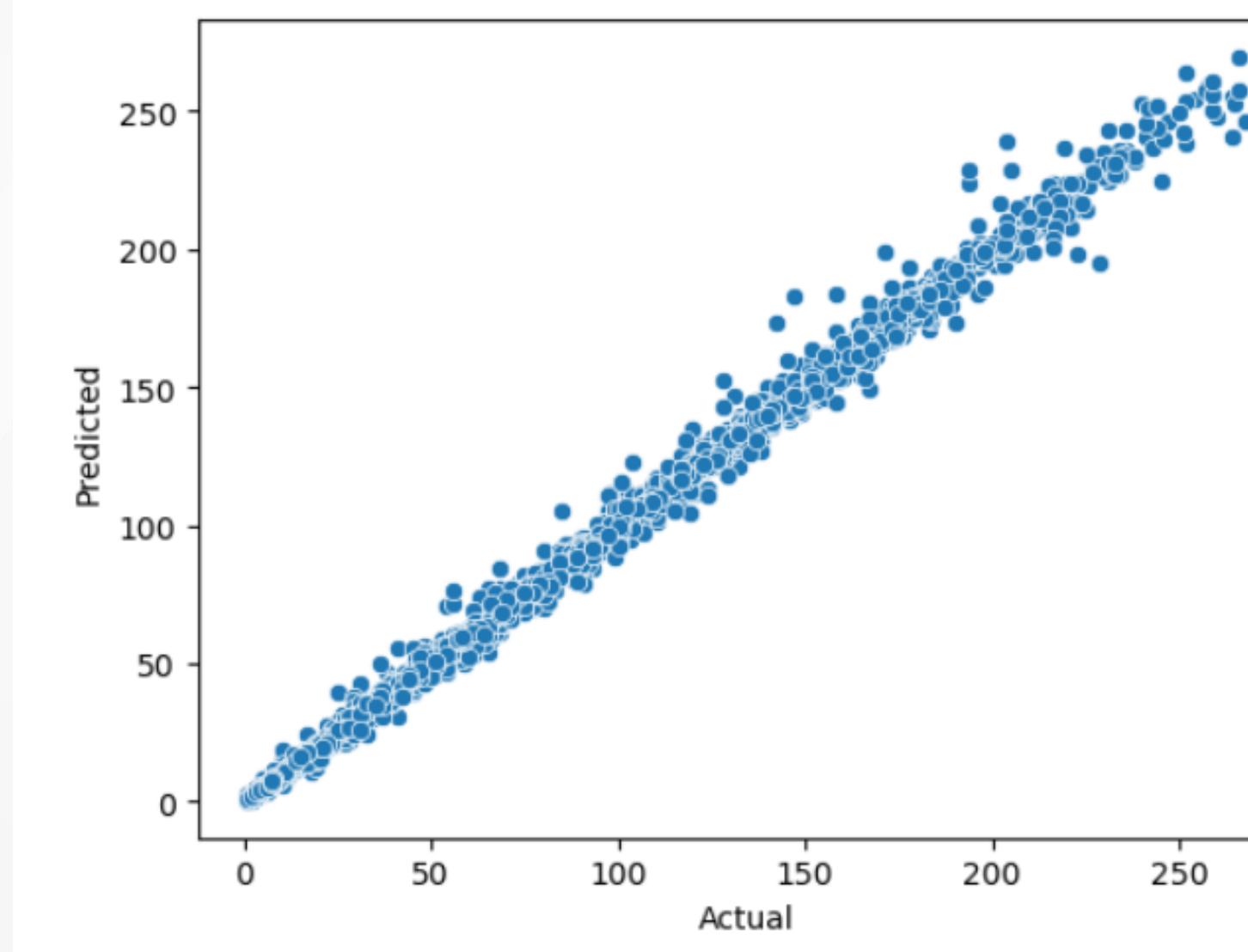
- **Lasso Regression**

THIRD DATAFRAME

Index	Modelname	R-square	Adj-R2	MSE
0	Linear	0.967210	0.967194	127.920212
1	Lasso	0.964448	0.964430	137.537756
2	Ridge	0.967210	0.967194	127.917925
3	Adaboost Regressor	0.958818	0.958797	161.183638
4	Gradient Boost Regressor	0.994847	0.994844	20.552283
5	RF-Regressor	0.999221	0.999221	18.311574
6	XGB Regressor	0.998661	0.998660	15.005201

CONCLUSION

- The conclusion from the above data frame is out of all the above models, XGB Regressor is best suited for my data .



MERITS

- Knowing how many calories one is burning during exercise can be highly motivating. It provides tangible feedback on the effort being put in and can help individuals stay committed to their fitness goals.
- With calorie predictions, users can set specific calorie-burning goals. Whether it's weight loss or improving cardiovascular fitness, having a target to work towards can enhance focus and determination.

DEMERITS

- Predicting calories burnt accurately can be difficult due to individual variations in metabolism, fitness levels, and even the type of exercise. Models may struggle to account for these differences.
- Even with accurate data, two individuals performing the same exercise may burn different amounts of calories due to differences in physiology, muscle mass, and other factors.

THANK YOU
