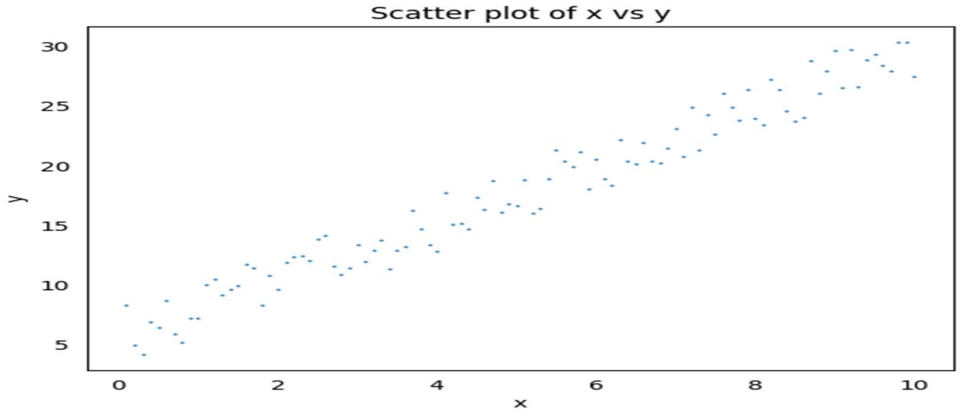
Linear Regression Analysis Report on Different Datasets

Data	MSE	RMSE	R_2	Cost by Gradient	Cost by Numerical
Data1	2.0785	1.4417	0.9579	[[5.680758659438731], [2.3840643453748847]]	[[5.6807871267612455], [2.384060066057181]]
Data2	0.0764	0.2764	0.9904	[[3.6821037544687716], [0.9730002947444361]]	[[3.682122666293189], [0.9729974518460569]]
Data3	0.1617	0.4021	0.3136	[[1.1770558584302782], [0.09419114914478582]]	[[1.1770620783119963], [0.09419021414817919]]
Data4	34.620	5.8839	0.9841	[[13.239475793250877], [6.132434333809515], [2.3922682969119413], [7.746810935958502]]	[[13.239477824421101], [6.132437632710662], [2.3922655421880563], [7.746810379932413]]

Analysis:

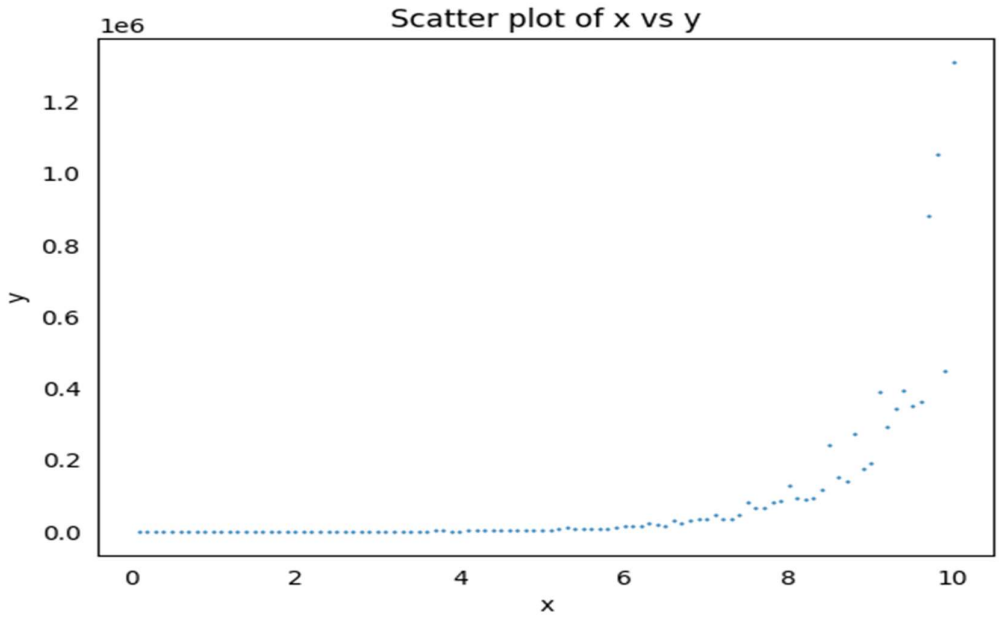
Dataset 1

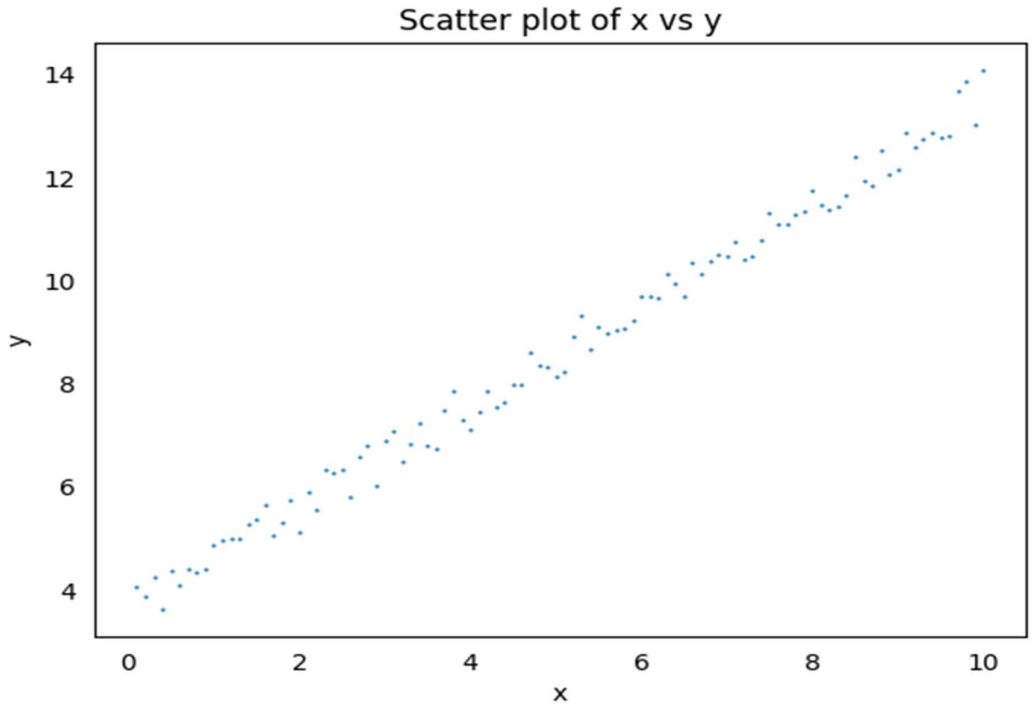
The data points in this dataset appeared linear when plotted on the x and y axis. The R-squared value is 0.9579571905586357, which is close to 1, indicating a strong linear relationship between the variables. Therefore, this dataset is suitable for the standard Linear Regression algorithm.



Dataset 2

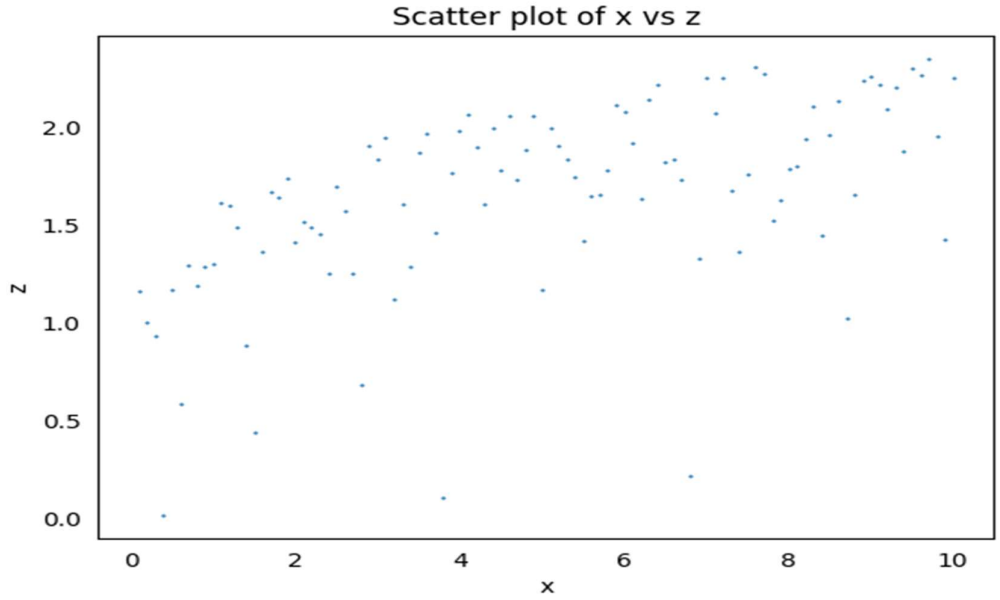
Initially, the data points in this dataset appeared exponential. However, after applying a logarithmic transformation to the input features, the transformed data appeared linear. The R-squared value is 0.9904038522690993, which is very close to 1, indicating a strong linear relationship between the transformed variables. Therefore, this dataset requires a non-linear transformation of the input features to be suitable for the standard Linear Regression algorithm. First plot of x vs y before transformation and second plot is after transformation.





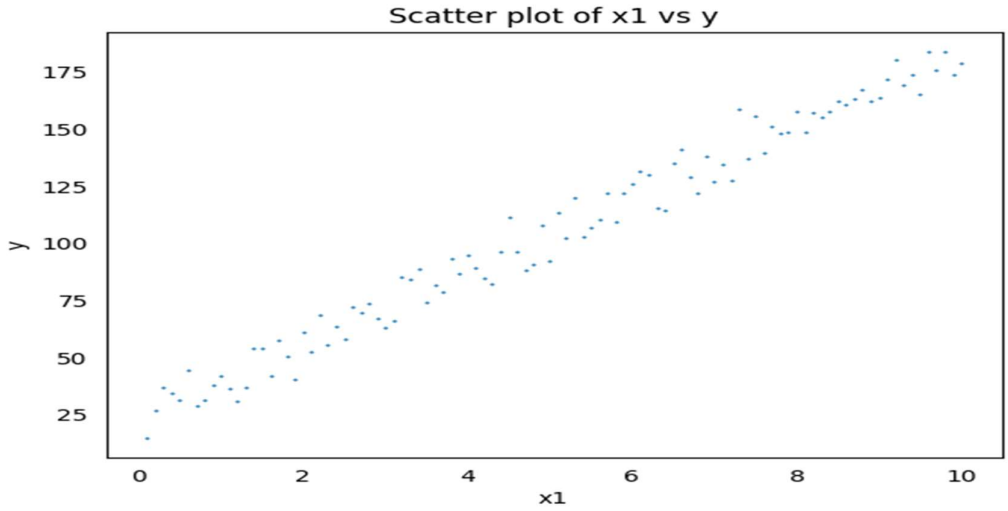
Dataset 3

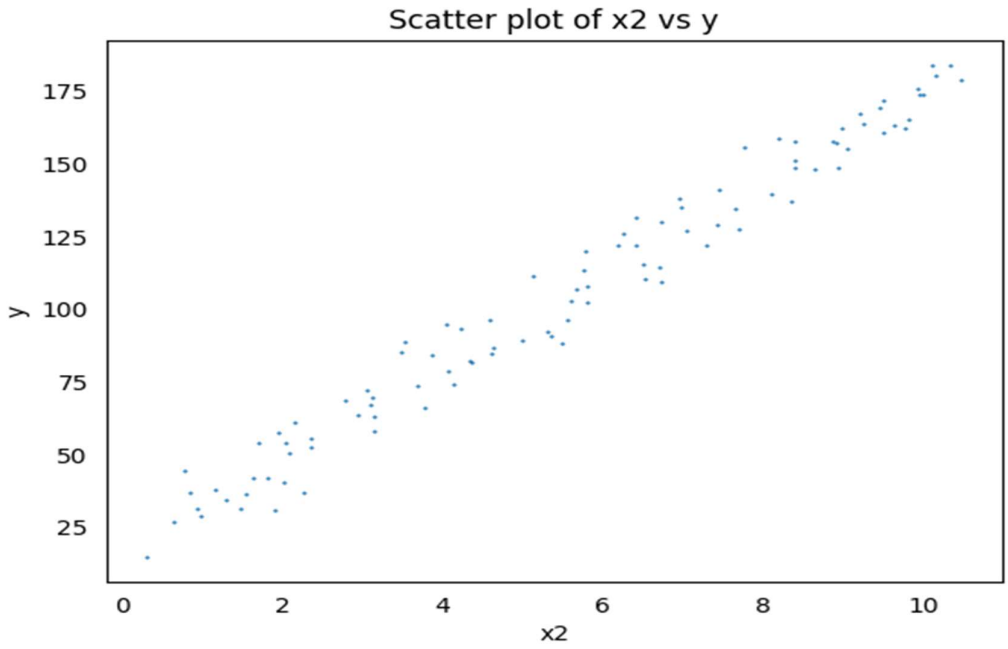
The data points in this dataset were scattered across the graph, making it difficult to decide a clear pattern or relationship. The R-squared value is 0.31369732267280803, which is far from 1, indicating a weak linear relationship between the variables. Therefore, the standard Linear Regression algorithm may not be suitable for this dataset. Other regression methods or algorithms that can handle non-linear relationships may be more appropriate.

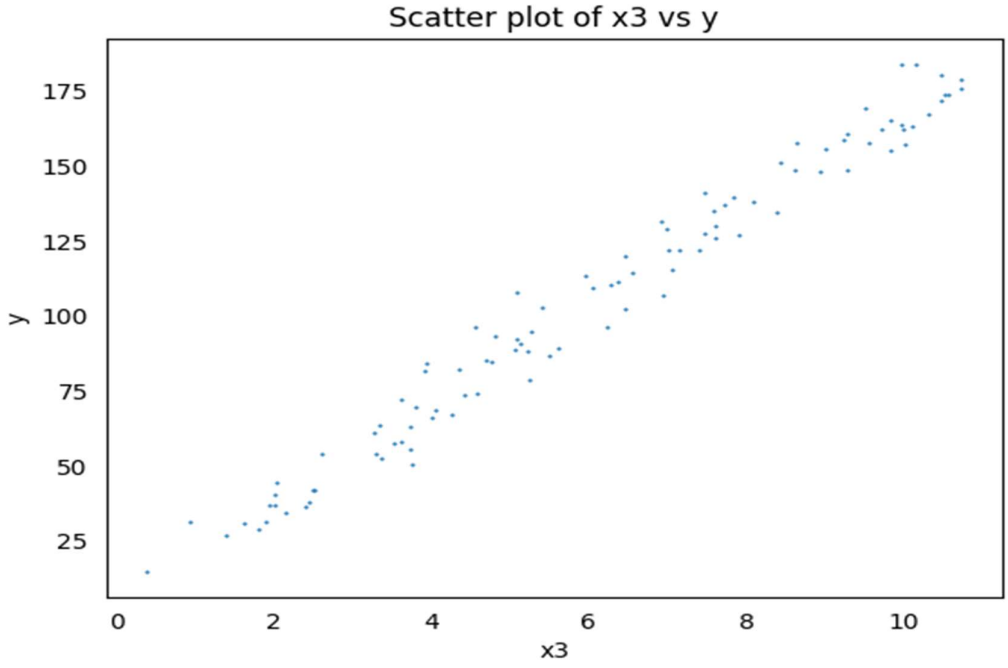


Dataset 4

This is a multivariable dataset. When plotting x1, y and x2, y and x3, y, all appeared linear. The R-squared value is 0.9841749058943147, which is close to 1, indicating a strong linear relationship between the variables. Therefore, this dataset is suitable for the standard Linear Regression algorithm.







Conclusion

Datasets 1 and 4 are suitable for the standard Linear Regression algorithm, Dataset 2 requires a non-linear transformation of the input features, and Dataset 3 may require a different regression method due to the weak linear relationship between the variables. These conclusions are based on the visual inspection of the plots and the R-squared values.