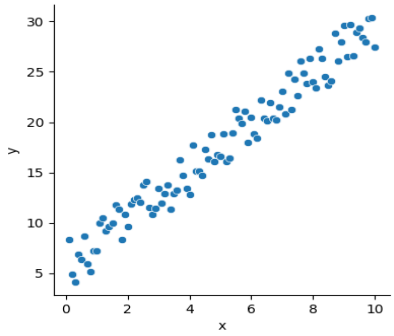
Linear Regression Analysis
Sahil Kumar

Data1:

Positive linear relationship: There is a positive linear relationship between the X and Y values. As the X-values increase, the Y-values also increase. This indicates that we can use linear regression to model this data effectively.



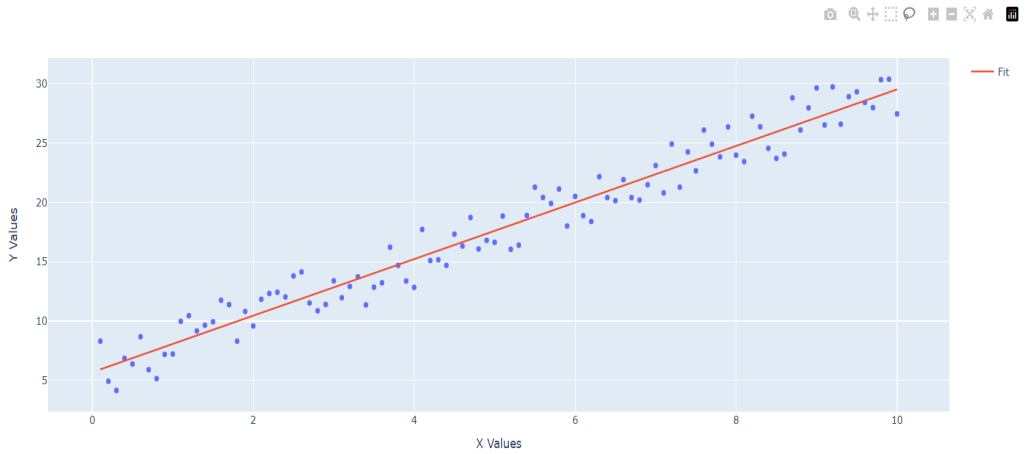
No outliers: There are no outliers or extreme values in your data that may affect the model performance or accuracy. All the data points are close to the best fit line, indicating a low error or variance.

Algo	MAE	MSE	RMSE	R^2
Simple linear regression	1.28055978429146	2.0785254017773265	1.4417091945941547	0.9579571905586358
Gradient descent	1.2842293730562 4	2.1150369286293262	1.4543166534937728	0.9572186635410976

In the above table error are shown using both the algorithms both the algorithms giving almost same error

coefficient of determination:0.9579… which is tending to one which is a signal that we have a good predictor.

Regression model : Linear as we can see from the following figure.

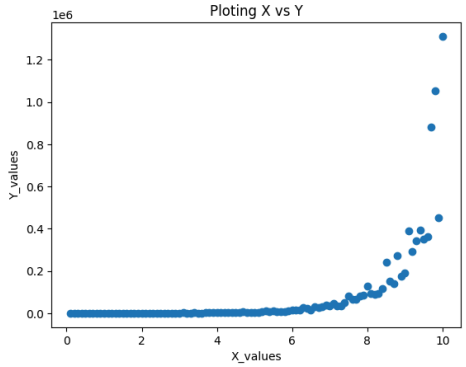


Best hyperplane : Y=5.67068712+2.38306007*X

Data2:

Non-Linear relationship: I scattered the data points and I found that there is a non-linear relationship between the x and the y variables.The data points are clustered near the origin and then sharply increase after an X-value of around 8.

Outliers: there are four distinct outliers where y value sharply increases and x values are around 9 or 10

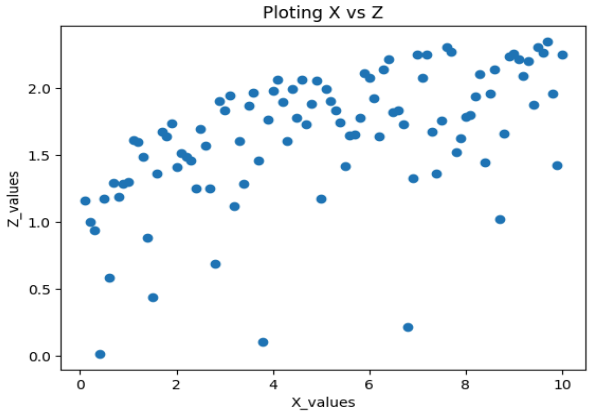


algo	MAE	MSE	RMSE	R^2
Simple linear regression	99929.7832971416	27577785853.164074	166065.6070749271	0.3433890964390125
Using gradient descent	99929.76847886939	27577785853.168034	166065.607074939	0.3433890964438069
Logarithmic	25883.6365244	8892037637.21	94297.6014393	0.78828580001

regression	30727	335	4388	5314

For this data values of all the errors are large as we can see from the above plot that points are shown like they are following the linear trend rather they are something like exponential.
coefficient of determination:0.3433… which tends to 0 is indicating that linear regression model will not give the best fit.
In such types of data we will do the non-linear transformation,by taking a log of actual y value in the dataset. After that we will calculate the regression model to evaluate the weight.This model gives reasonable values of errors and also gives the good value of R^2 indicating the model is good.
Hyperplane Y = exp(2.01637719+0.15128513*X)

DATA3:
Non-linear pattern: The scatter plot of our data shows a non-linear pattern. It suggests that a linear regression model might not be the best fit for this data as it does not follow a straight line trend. We might consider exploring polynomial regression to better fit and predict the data.



algo	MAE	MSE	RMSE	R^2
Linear regression	0.2946779330131036 3	0.1617304414308855 2	0.4021572347116057	0.3136973226728079
LR using Gradient descent	0.3076562835325742	0.1656296623508913 5	0.4069762429809526	0.2971509895693162

Polynomial regression	o.27284952969 17847	0.14466752744 736108	0.38035184690 93598	0.38610375059 25682

We chose different polynomials with different degrees and tried to fit into the regression model but we are not able to generate the value of good R^2.

DATA4:

algo	MAE	MSE	RMSE	R^2
Linear regression	5.15550563037 7769	34.6204971753 4355	5.88391713310 47435	0.98417449058 943147
LR using GD	5.16004993538 41425	34.6298971753 5018	5.88471713310 474	0.98417060166 28395

In this data simple linear regression gives the best fit with R^2 value very close to 1.

Best fit Hyperplane y=6.13243*X1+2.39336*X2+7.7468102+13.239577