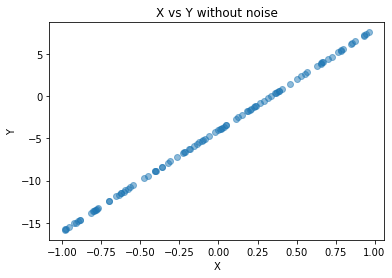
Assignment 3

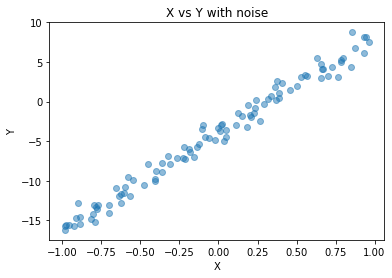
Analysis Report

Exercise 1:

Graph 1: Y=12X-4



Here in this graph, we can see all the points lies on the regression line Y=12X – 4. As we have selected the random values for X we have random points on the line.  
  
Graph 2: Y=12X-4 with noise



In this graph, we have added noise to the Y coordinate and hence we can see the points are not following a hard line (Y=12X-4) not the points are slightly above or below the line depending on the noise. But it shows the trend of the line (Y=12X-4).

Exercise 2:

Data columns (total 8 columns): Monthly Income, Transaction Time, Gender\_Female, Gender\_Male, City Tier\_Tier 1, City Tier\_Tier 2, City Tier\_Tier 3

R squared value for 1st model: 0.18408064631154986

Here for this run we can say that our model’s accuracy is 0.18 i.e. 18%. That means our models correctly map 18% of data on the linear regression line.

After adding “Record” column

Data columns (total 7 columns): Monthly Income, Transaction Time, Record, Gender\_Female, Gender\_Male, City Tier\_Tier 1, City Tier\_Tier 2, City Tier\_Tier 3

R squared value for 2nd model: 0.9187862767496578

The new accuracy of our model is 0.91 or 91%. That means our model correctly map 91% of data on the linear regression line. We can see this significant increase in accuracy because “Record” column is a strong feature to determine the “Total spend”