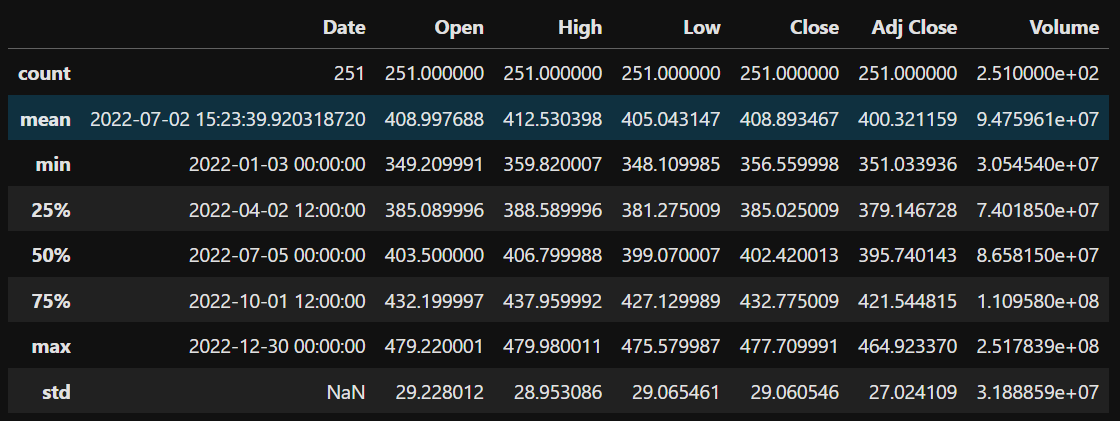
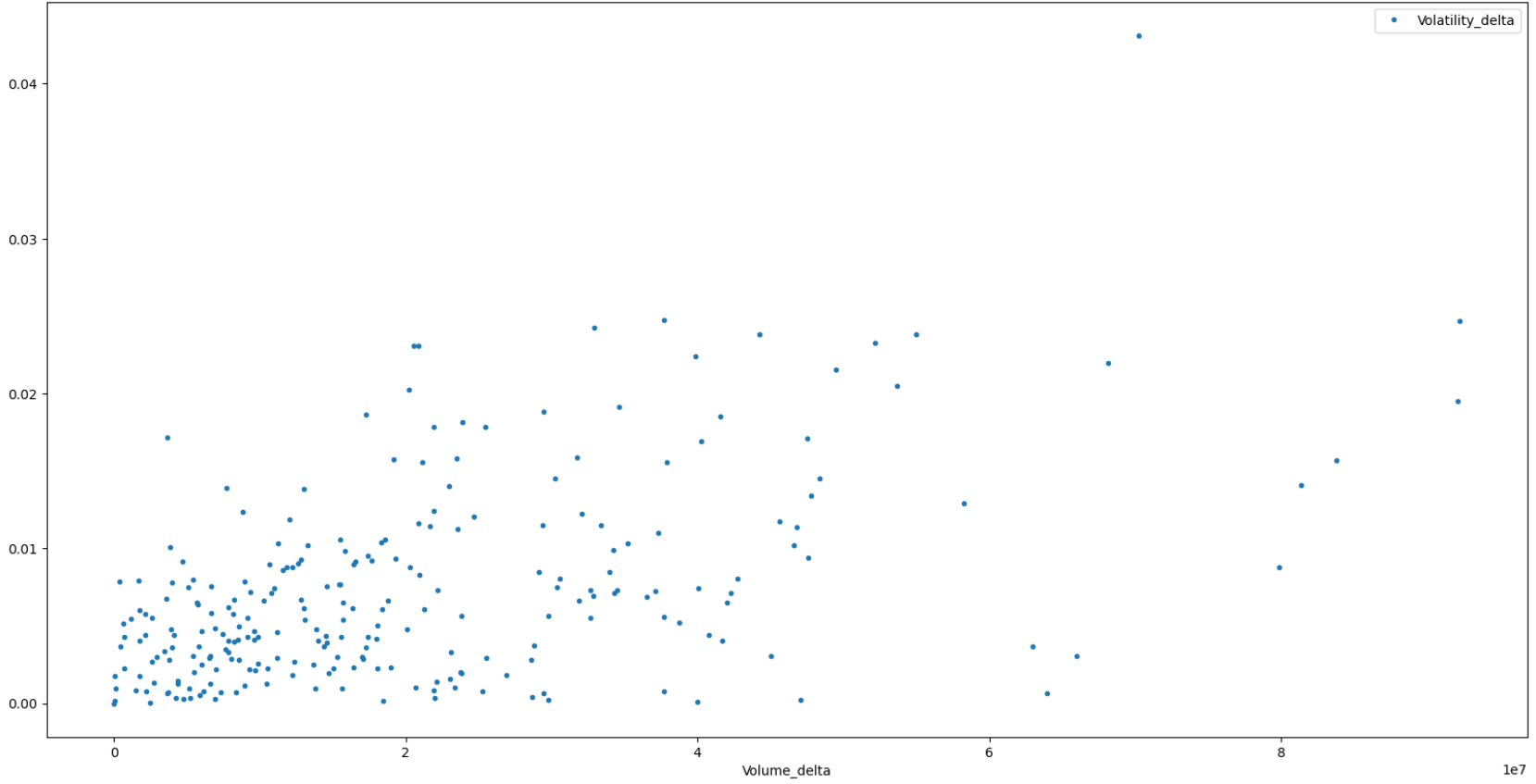
# **Summary**

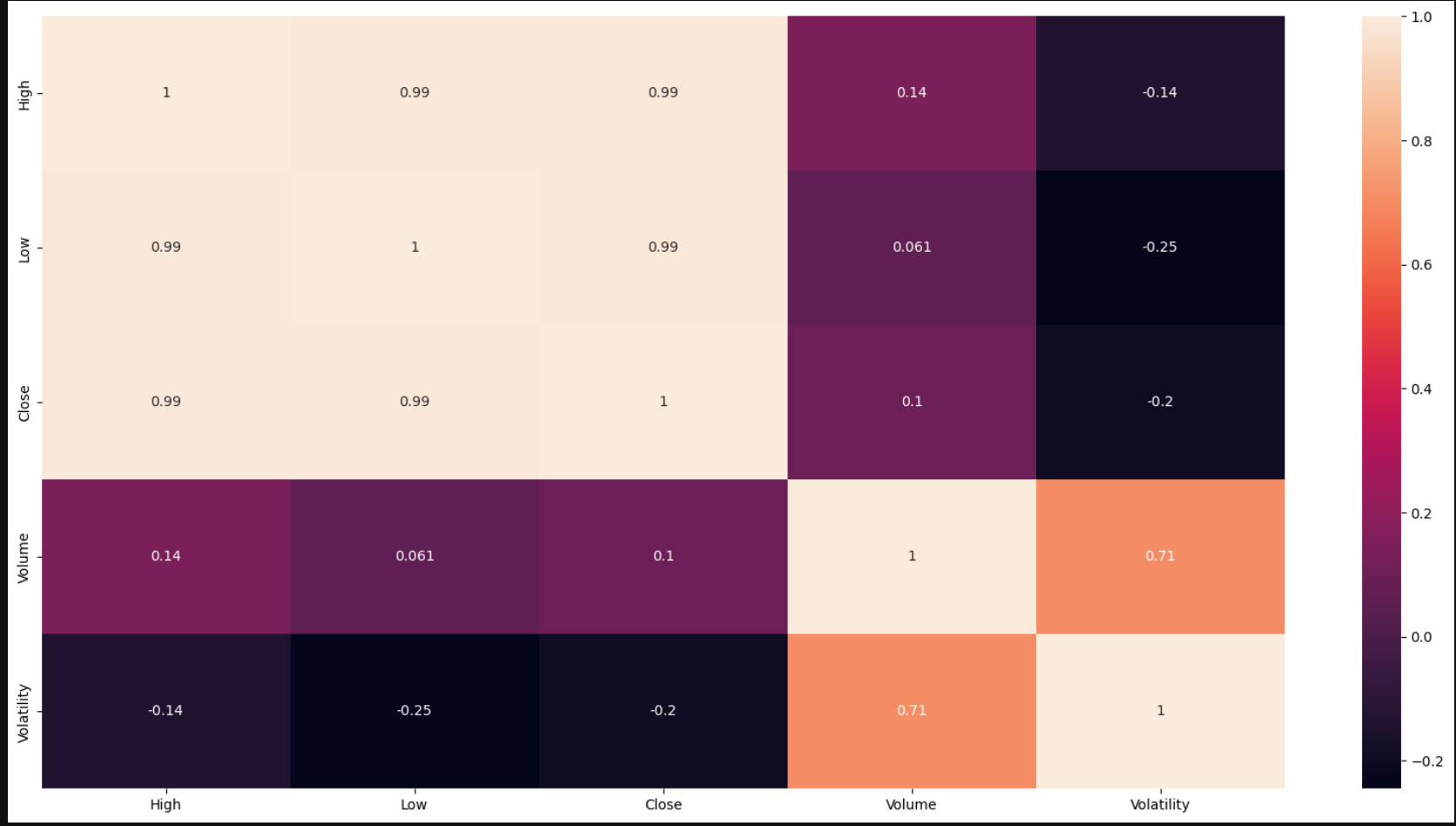
The historical data of SPDR S&P 500 ETF Trust (SPY) for the year of 2022 was used to develop the equation. Initially the data was loaded, the date column was converted from Object class to DateTime class, and the summary statistics were checked.



As all columns have significant standard deviation greater than 0, we can use all the columns for equation. Then Intraday Volatility which is equal to (High-Low)/Close along with daily volume change and daily volatility change were calculated and plotted.



We can see that, The volatility delta value ranges from 0 to 0.05. Before building the equations, correlation plot was plotted to check the influence of variables.

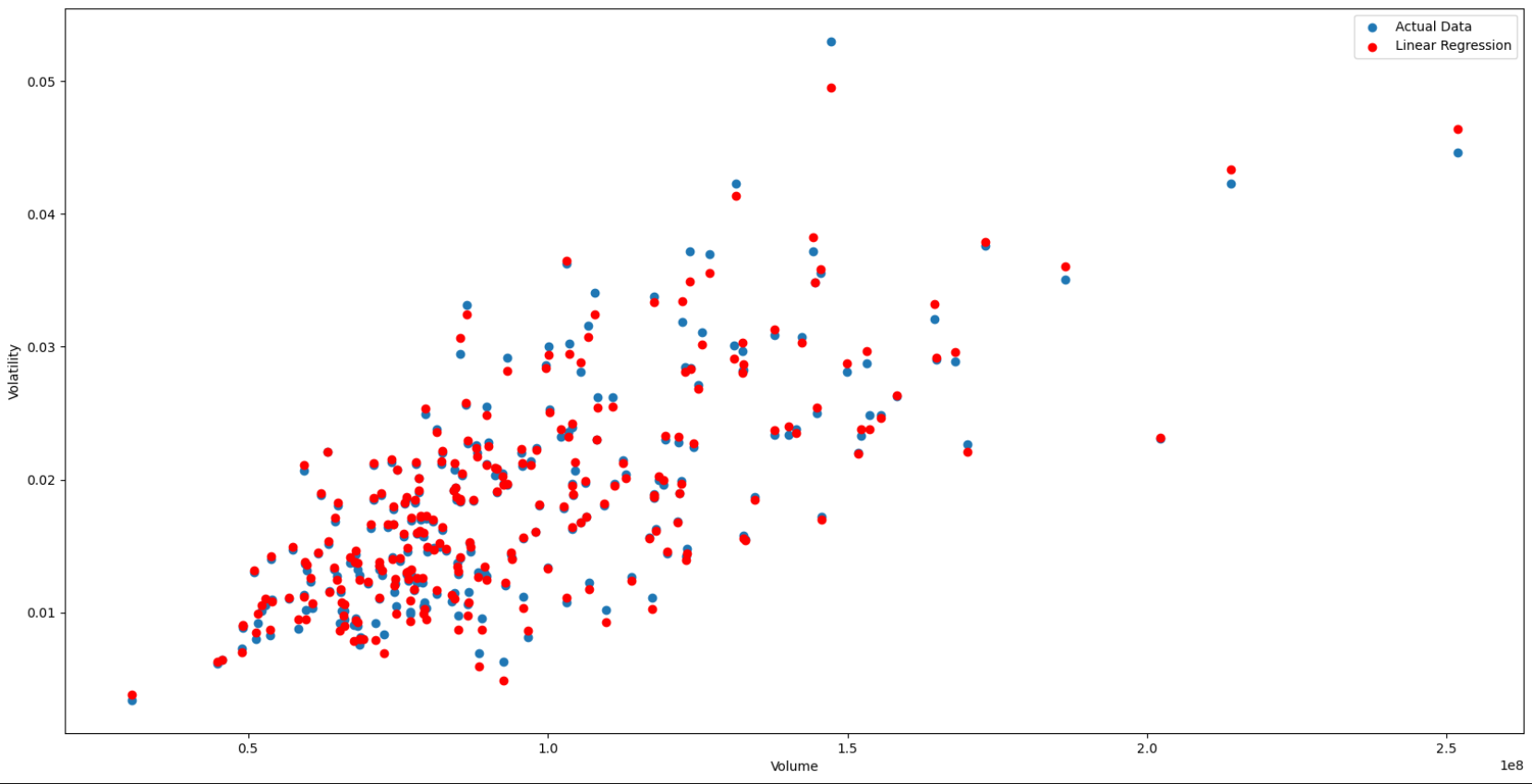


We can see that volume has the highest influence of 0.71 on volatility which means if volume increases the volatility increases. Other variables have mild negative correlation meaning if Volatility increases the other decreases. Linear Regression was used to build the multi variate linear equations using Volume, High, Low and Close.

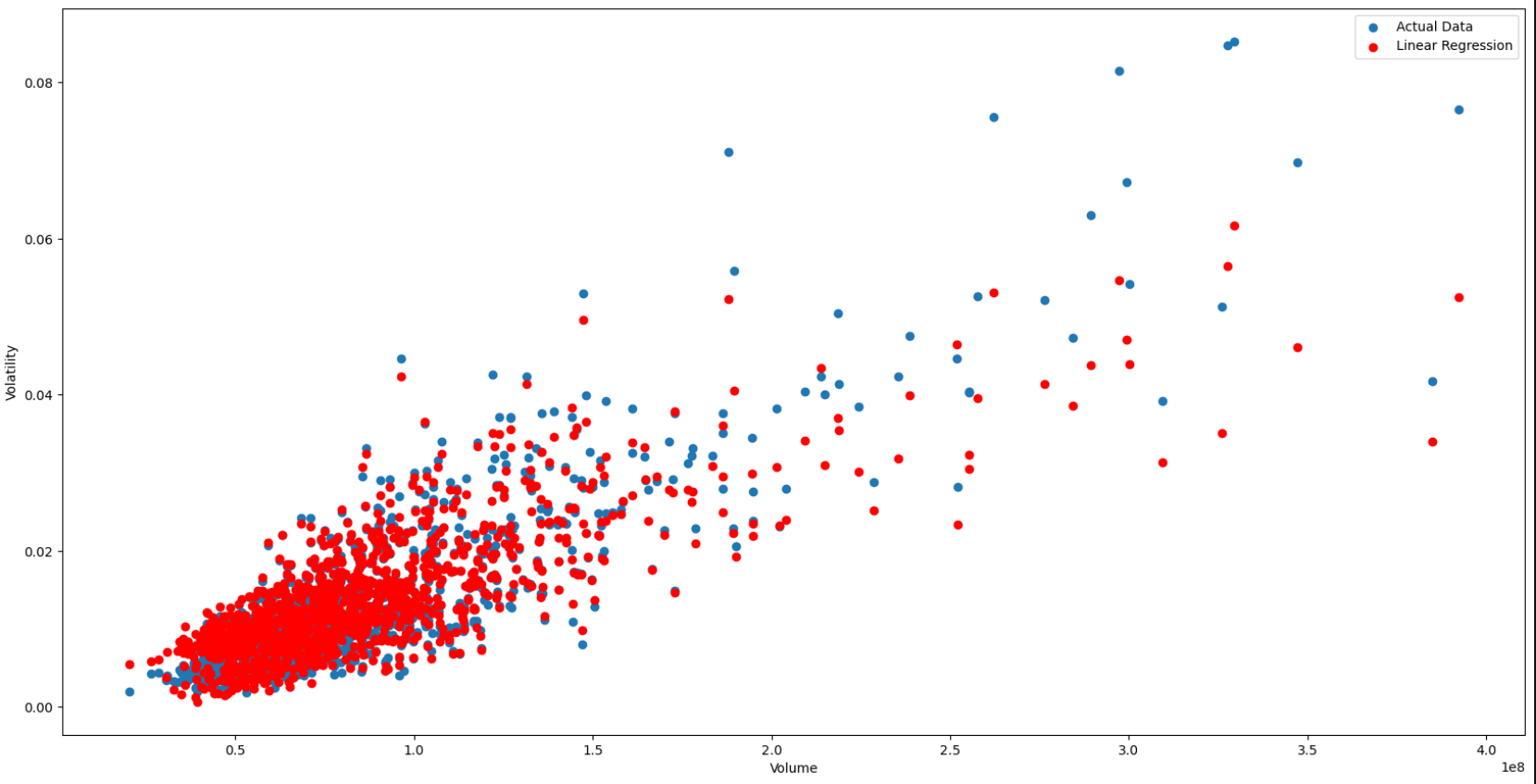
## Case 1: Using Equation built using SPY on other ETFs and Stocks:



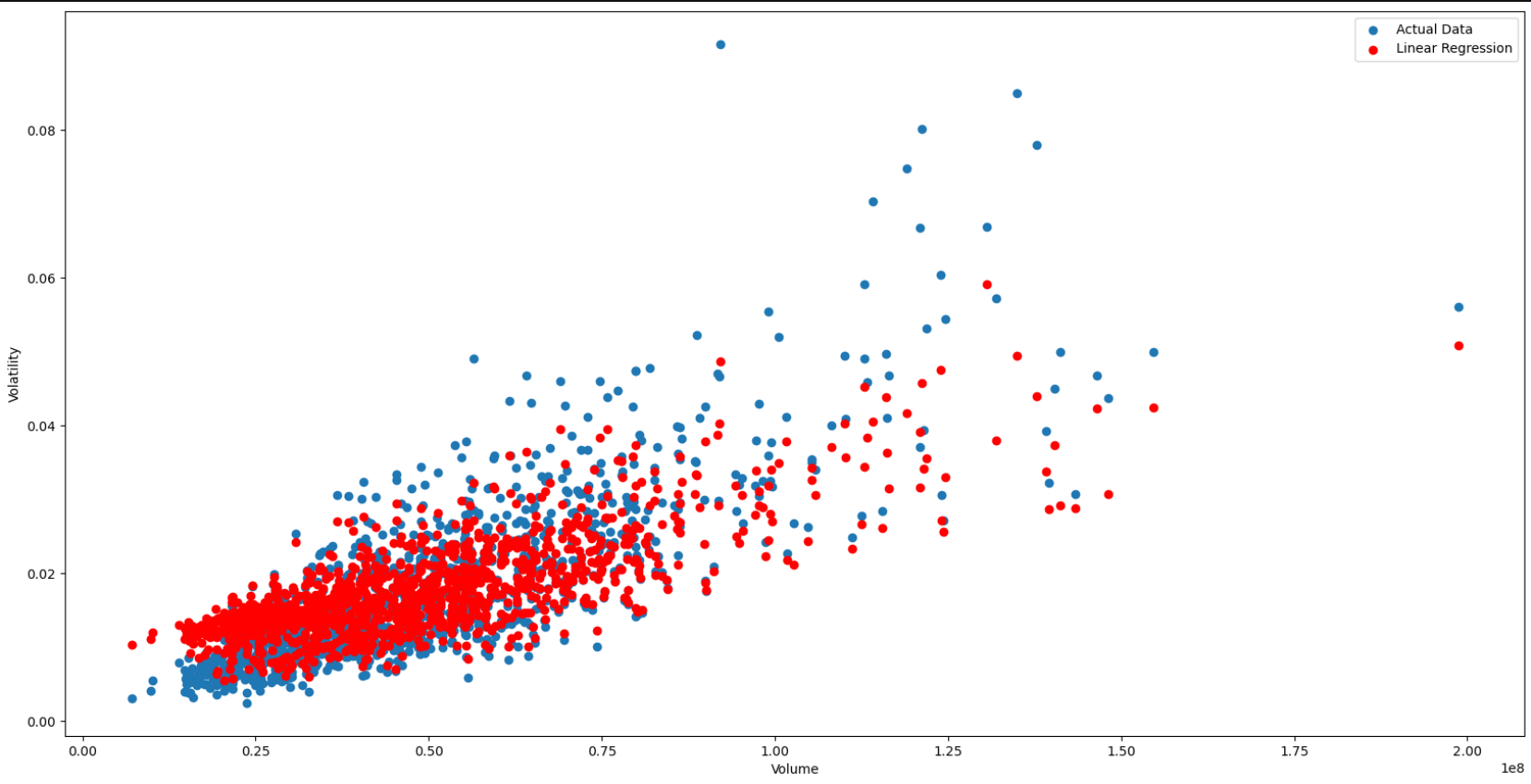
With the above built equation, the Volatility values predicted, and percentage of error was calculated which is the absolute difference between predicted and actual value divided by the actual times 100. Which gives the change between actual and predicted value. The actual and predicted value were plotted.



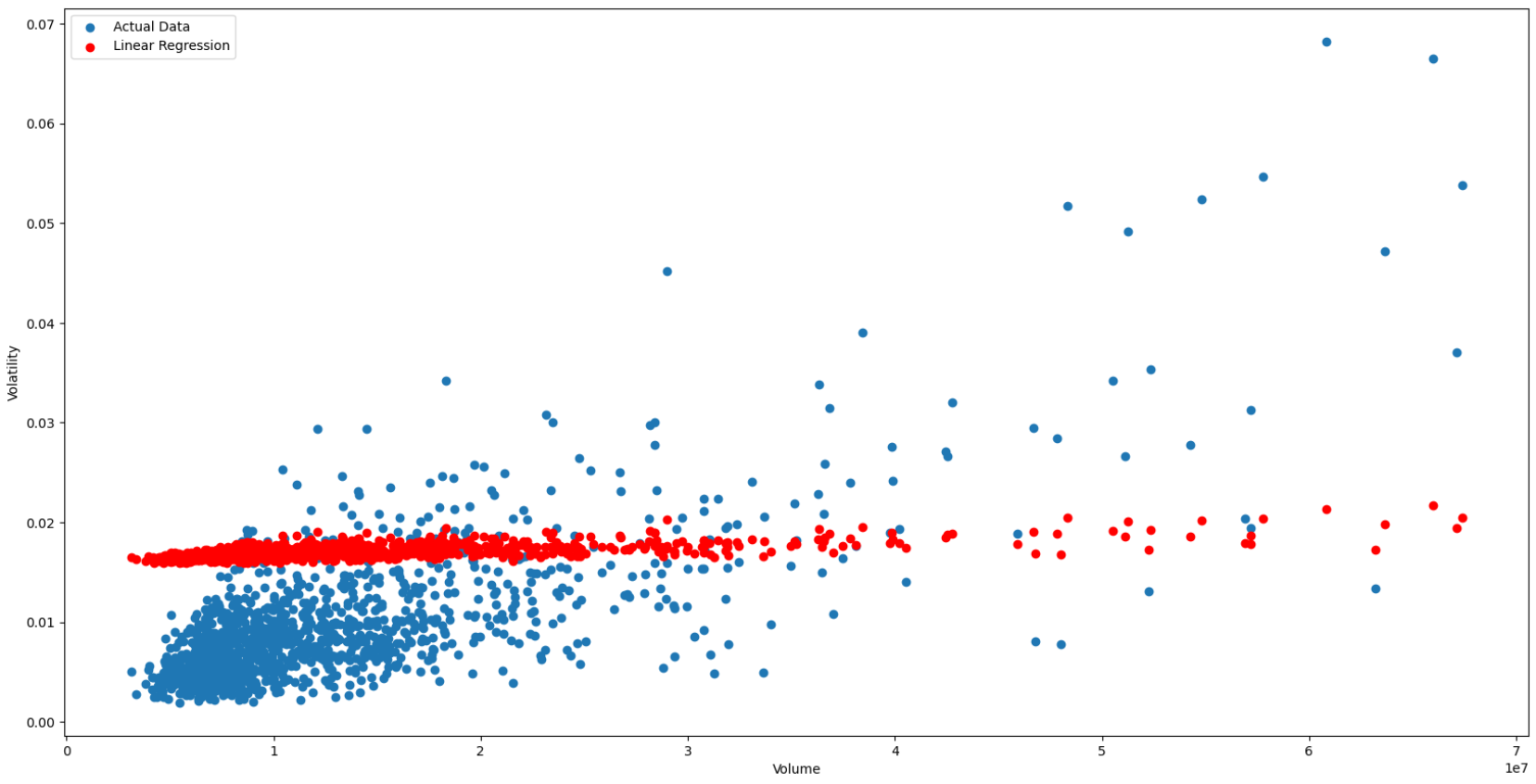
From the plot we can say that the equations fit good for given data. The coefficients and intercept are of small value because the variables ranges are big whereas volatility is always less than 1. This equation is then used on the same data but for 5 Year interval and its Percentage Error and scatter plot were calculated and plotted.



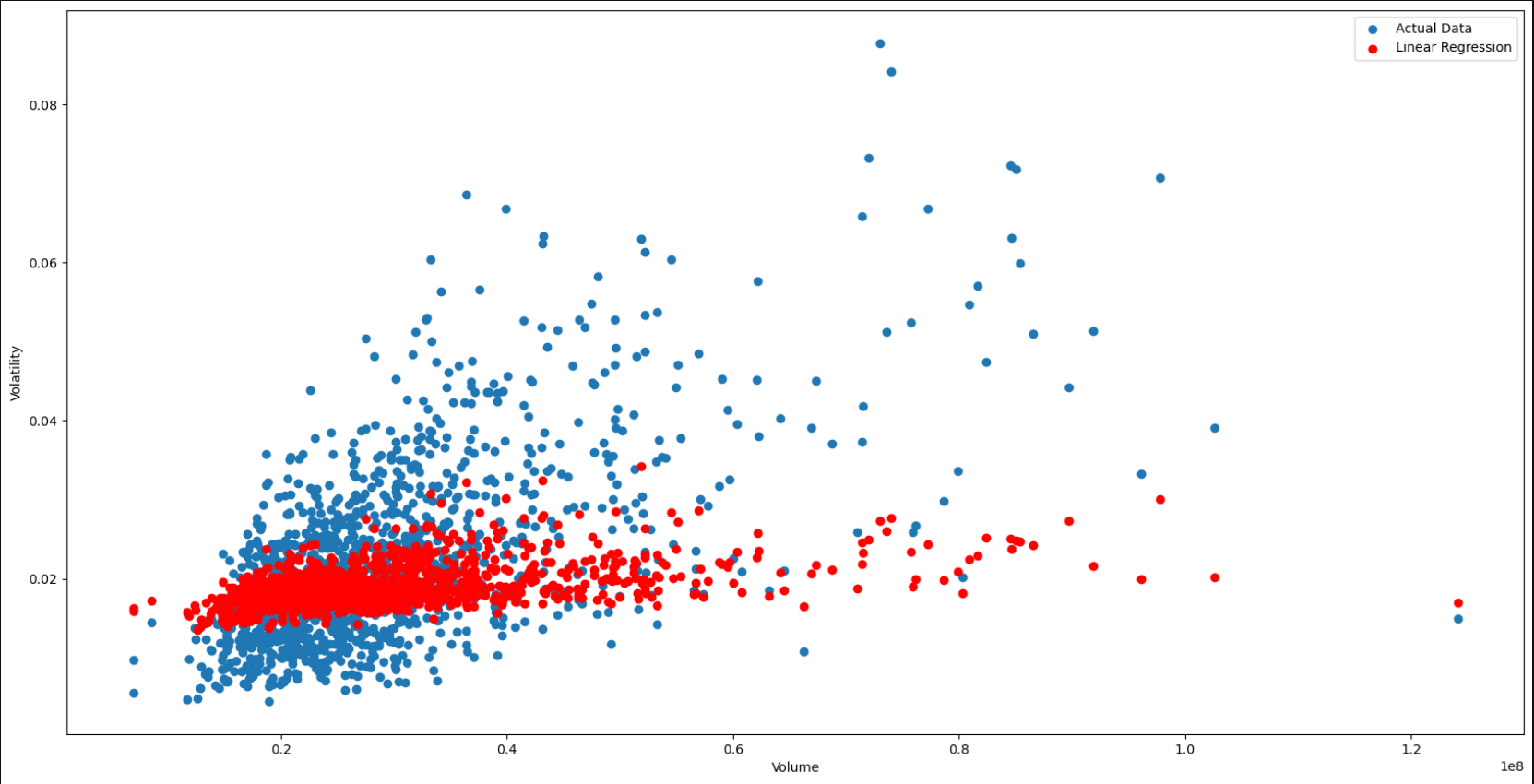
The same equation is plotted for other ETFs QQQ Invesco QQQ Trust Series I and VEA Vanguard FTSE Developed Markets ETF and Stocks Google and Microsoft respectively and the results were noted down.



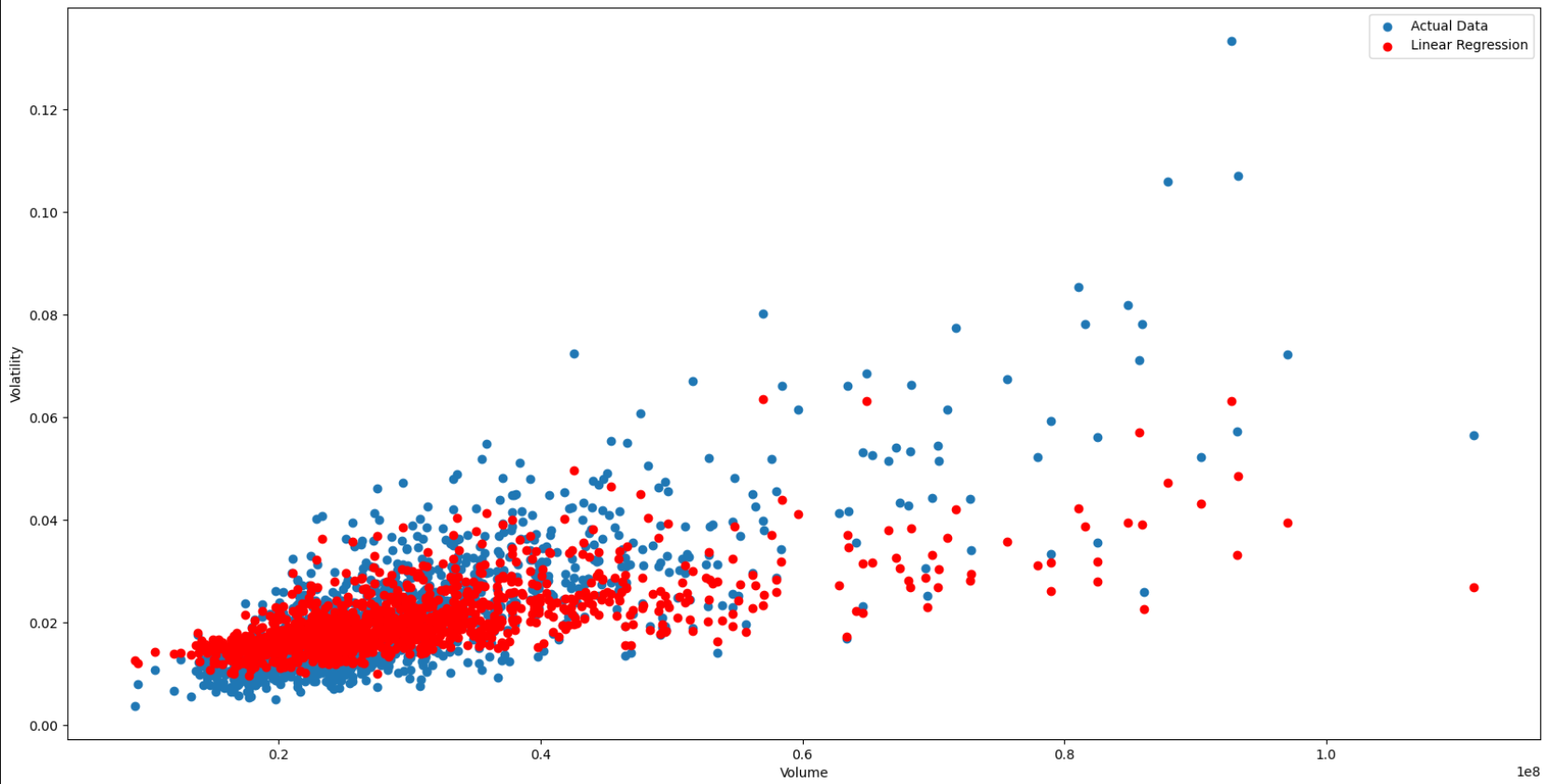
QQQ



VEA



GOOG



MSFT

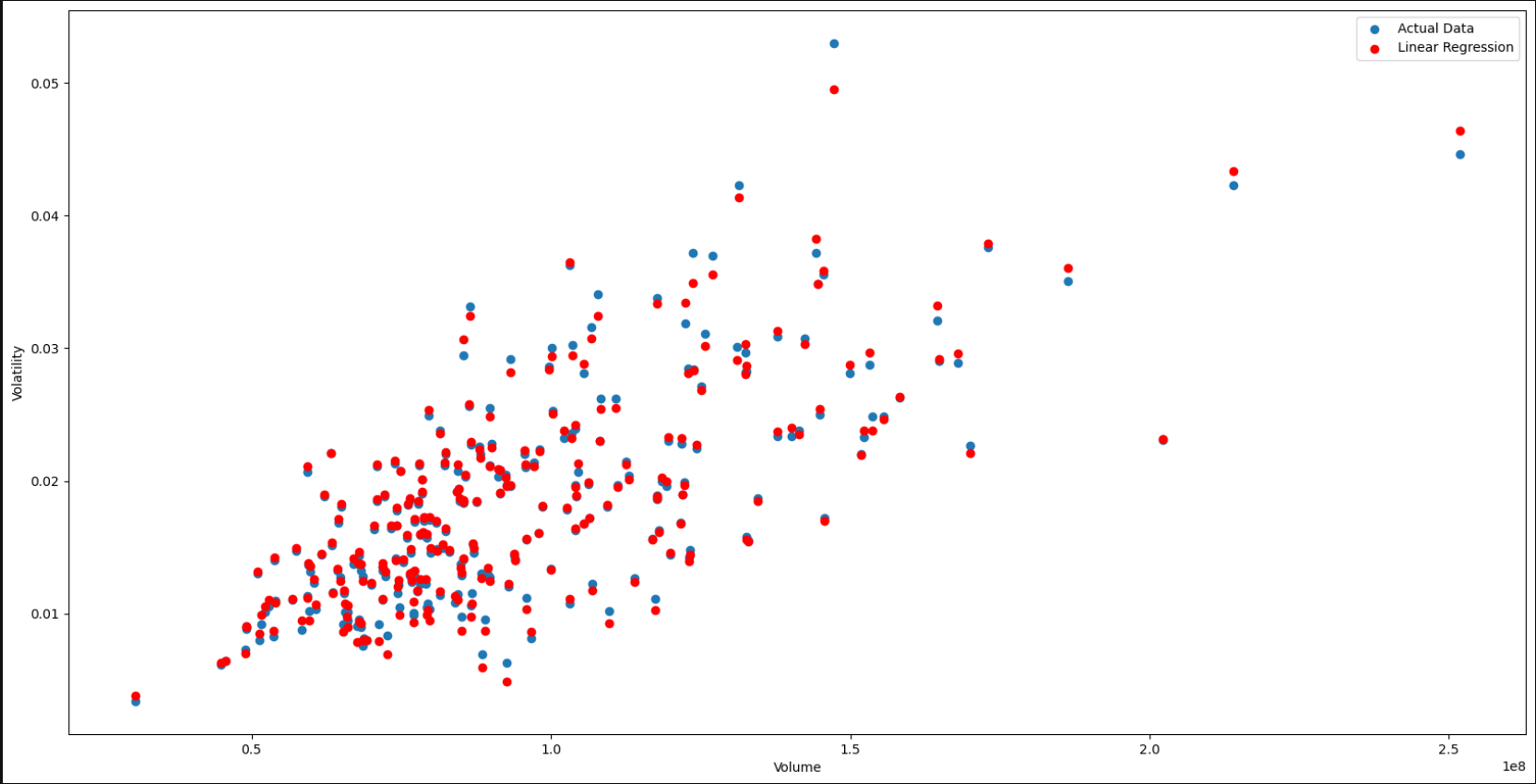


Results

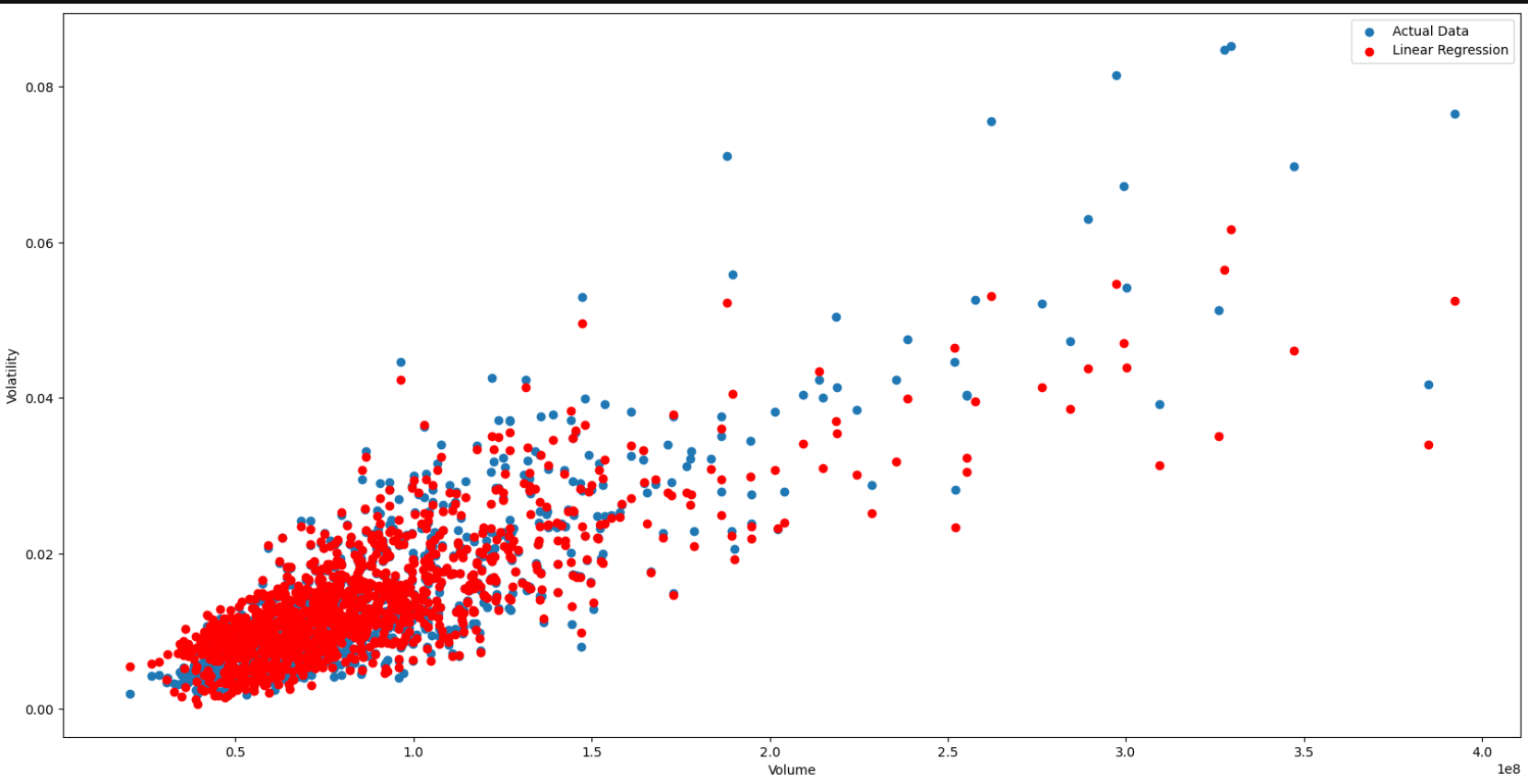
We can see the equation could not represent the actual value well and the mean percentage error increases with different data. This might be since the equation was largely influenced by variable Volume, if it was not as same SPY data then variations in representation arises and since the volatility in the dataset with which the equation was built had values ranging from 0 to 0.06 values with volatility above them were represented accurately and if the data had small values, then also the representation produced by the equation is wrong. The mean percentage error is small if the data has similar values compared to SPY data. Another Shortcoming of this equation is using dependent variables such as High, Low, Close which had high dependencies within them.

### Case 2: Building Equation for each data and Testing for 5-year data:

With SPY data using linear regression the equation was built and tested with its 5-year data. 



SPY

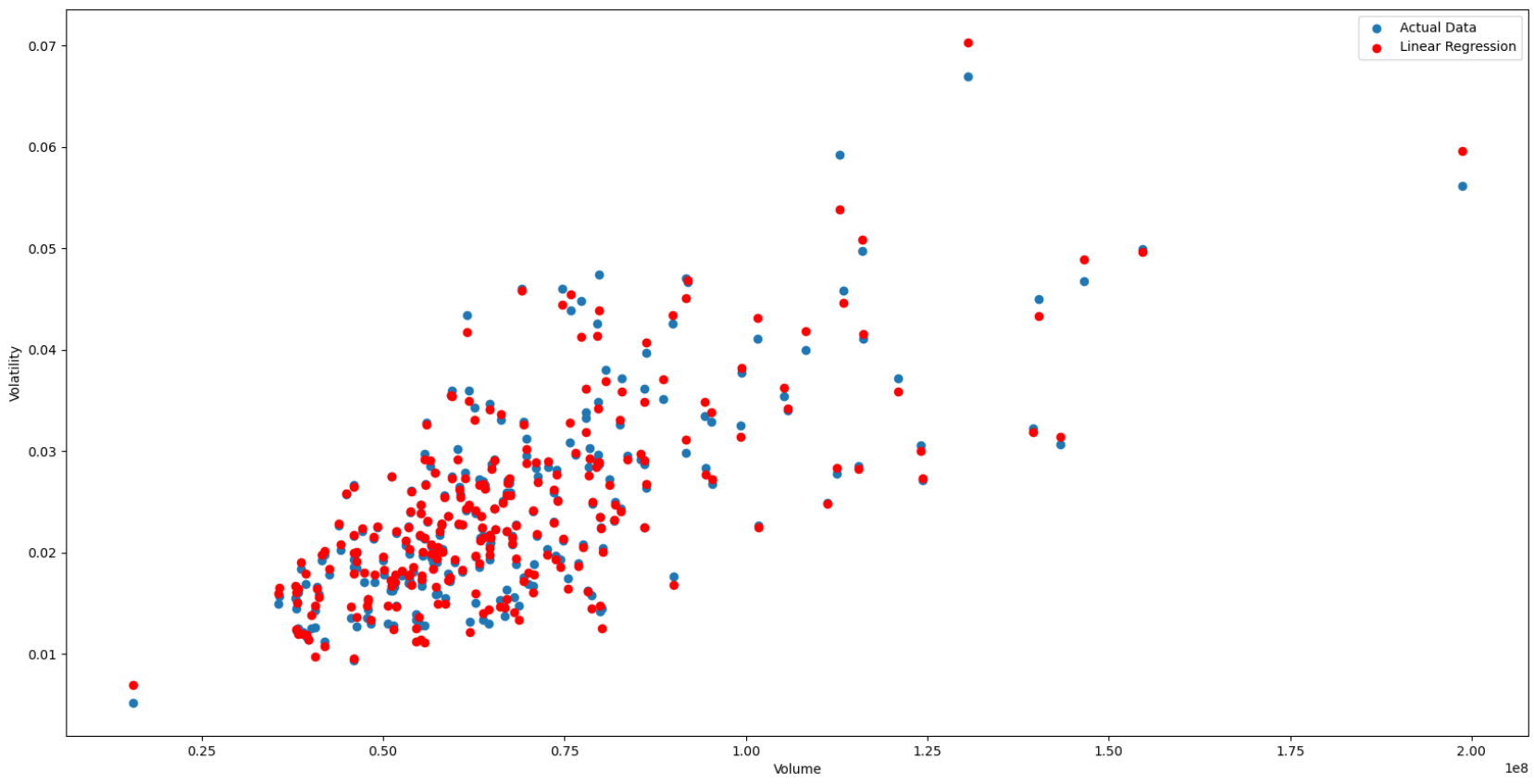


SPY-5Y

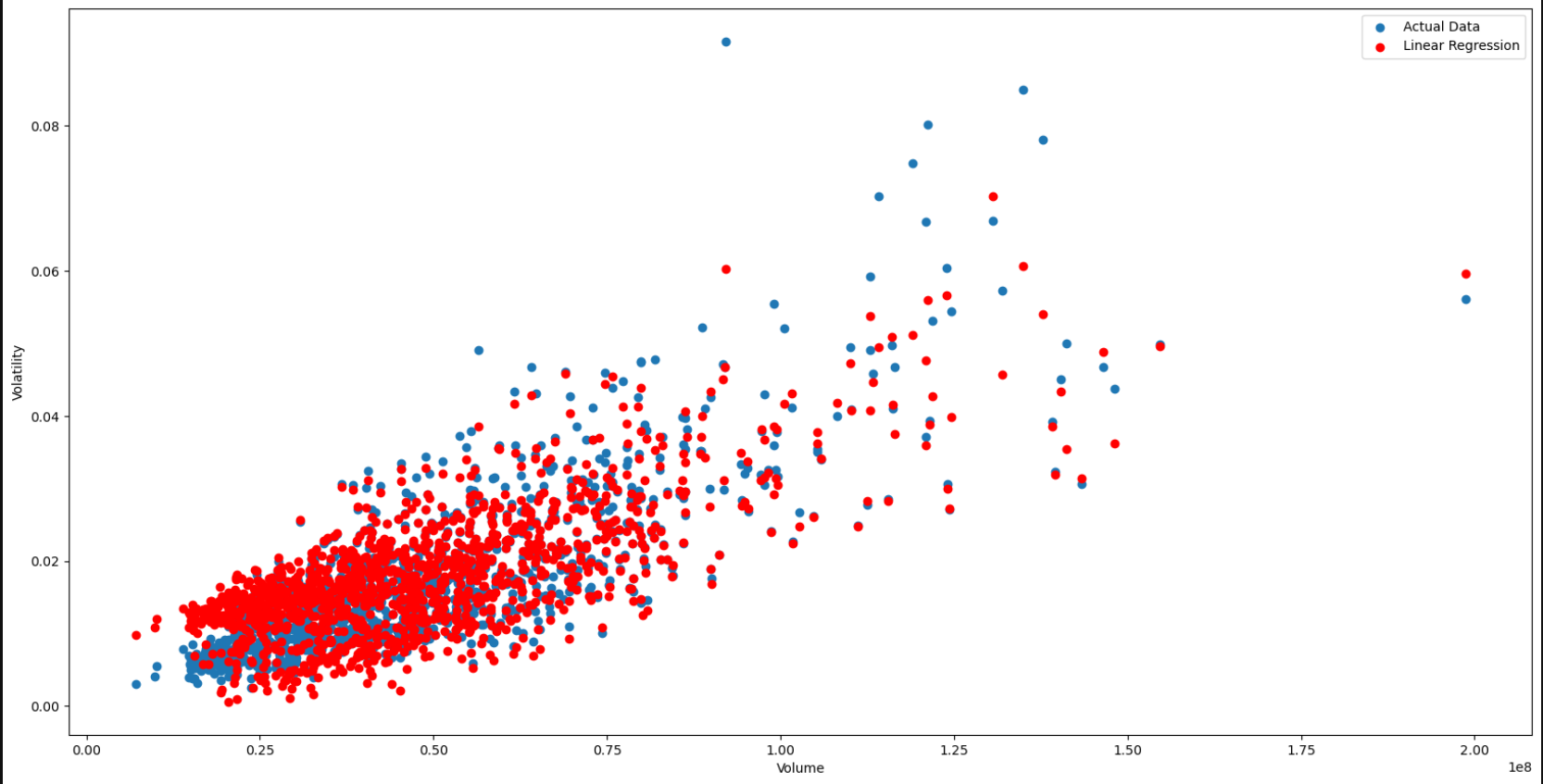
Similarly, 4 equations for each ETF and Stock were built and the values were plotted.

### QQQ:



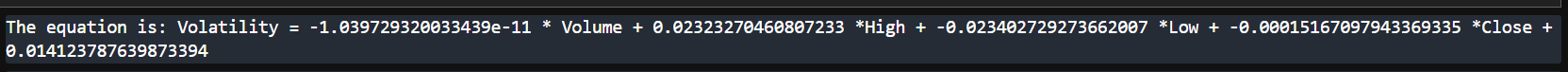


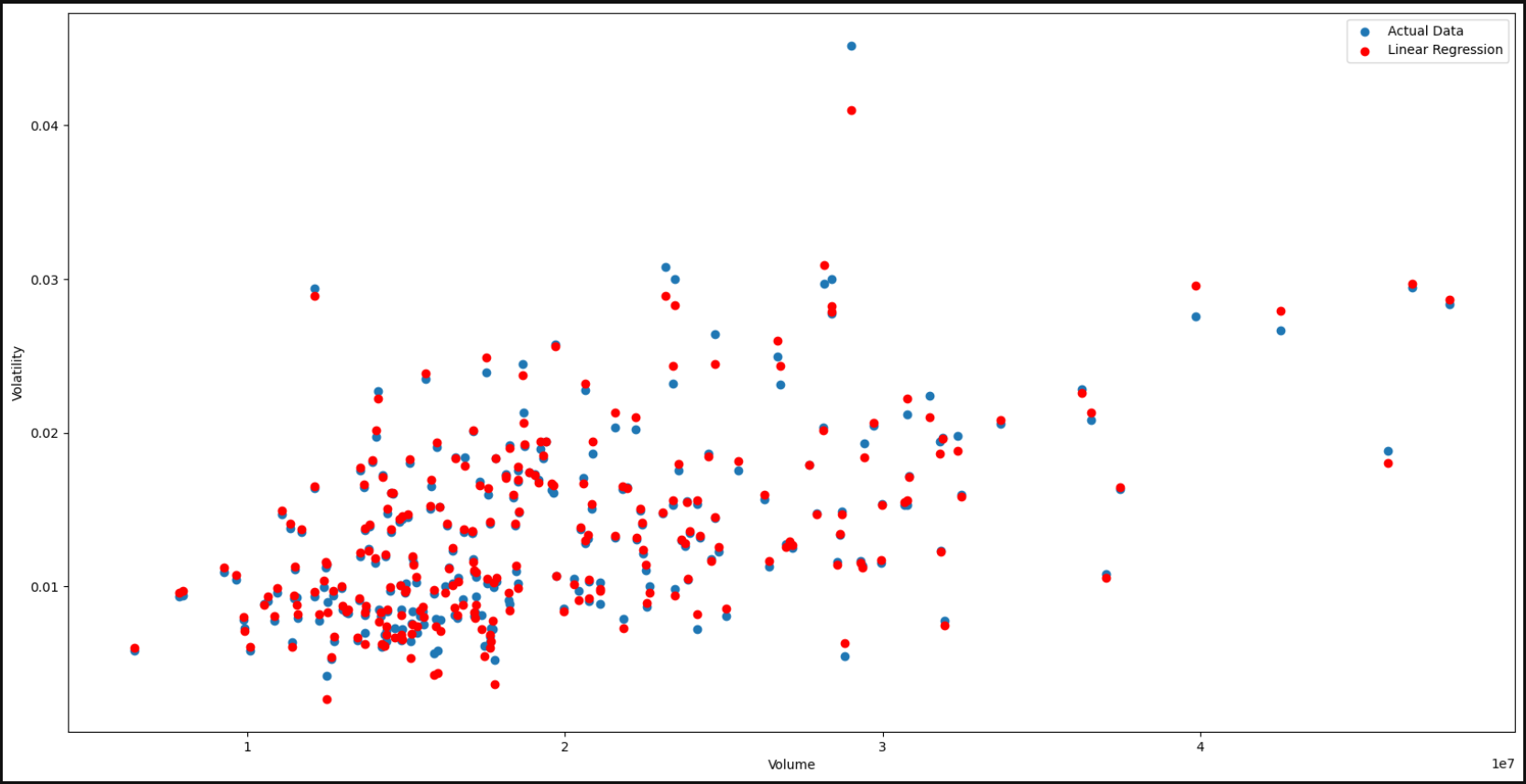
QQQ



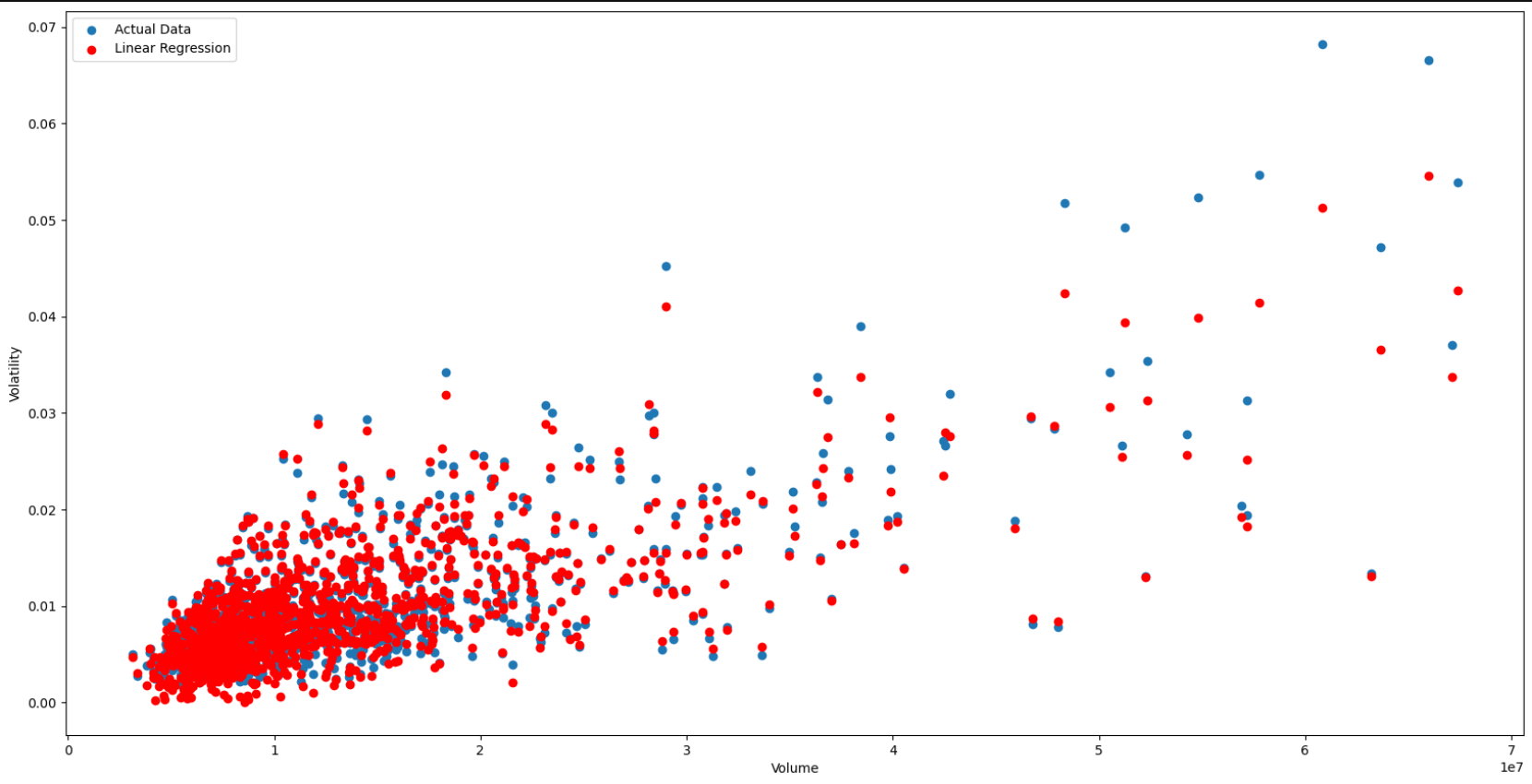
QQQ-5Y

### VEA:





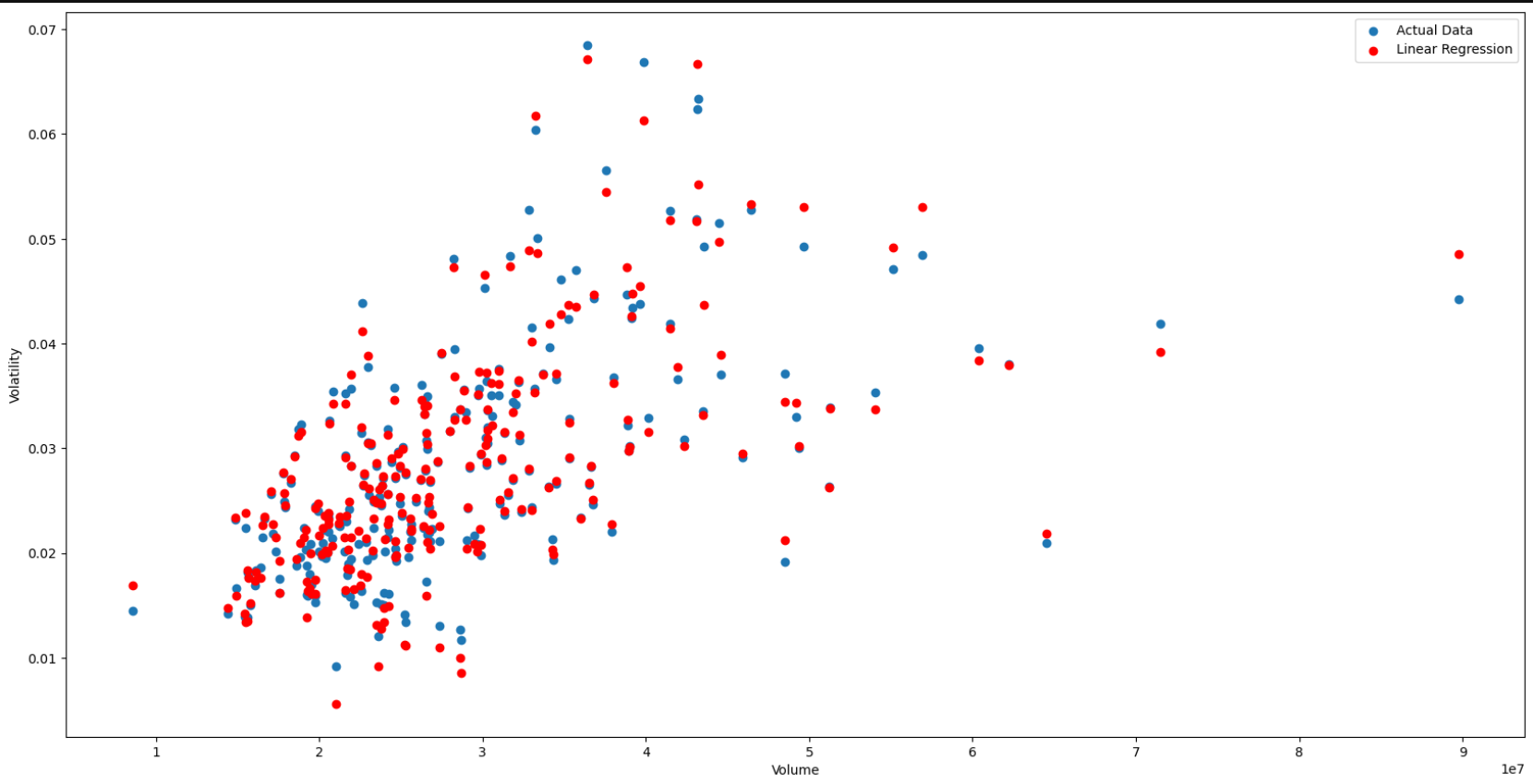
VEA



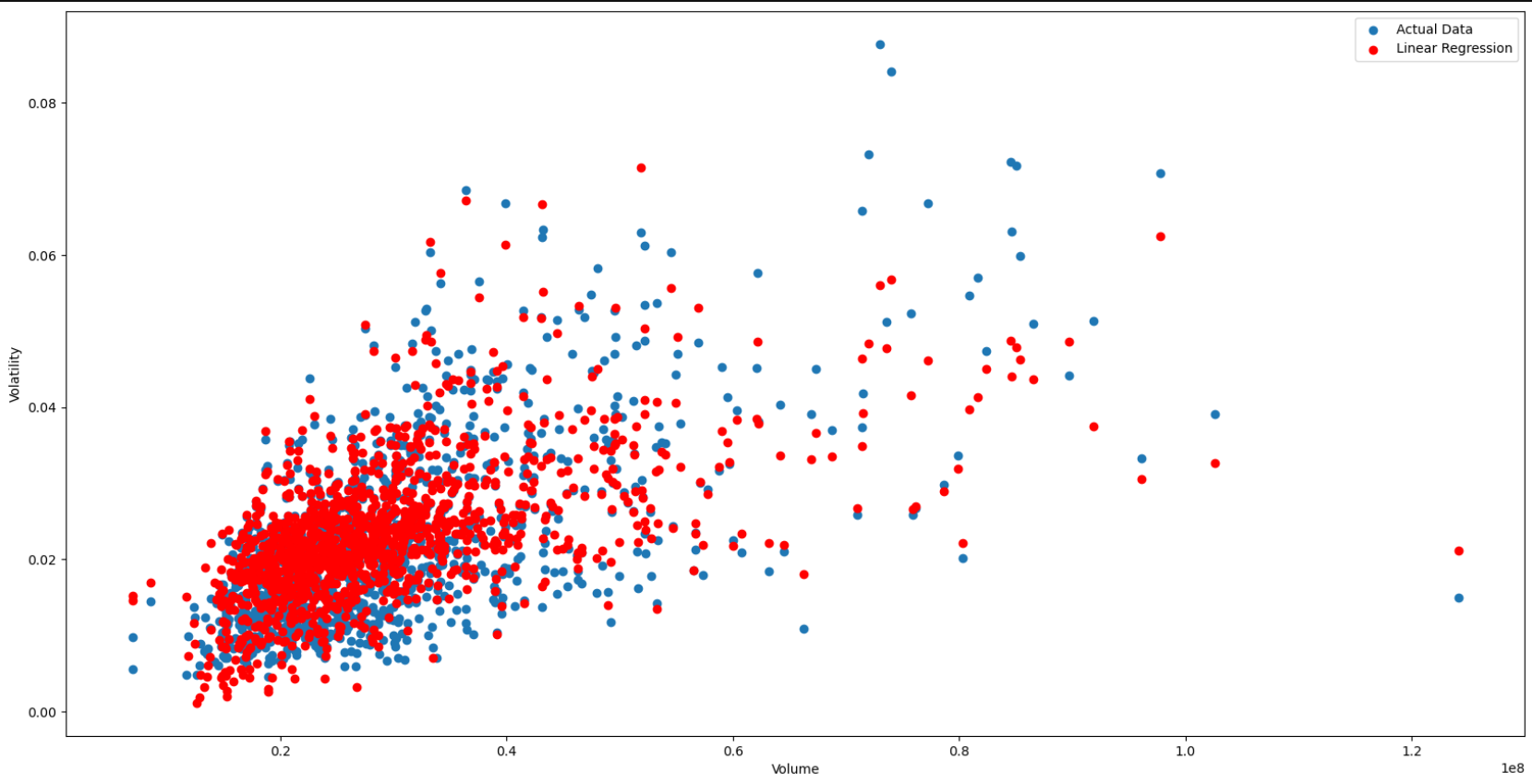
VEA-5Y

### Google:



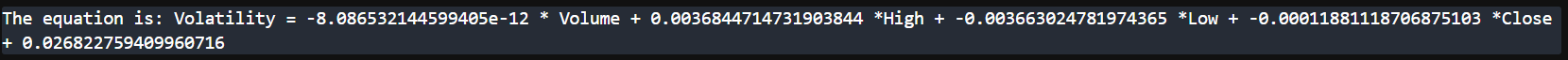


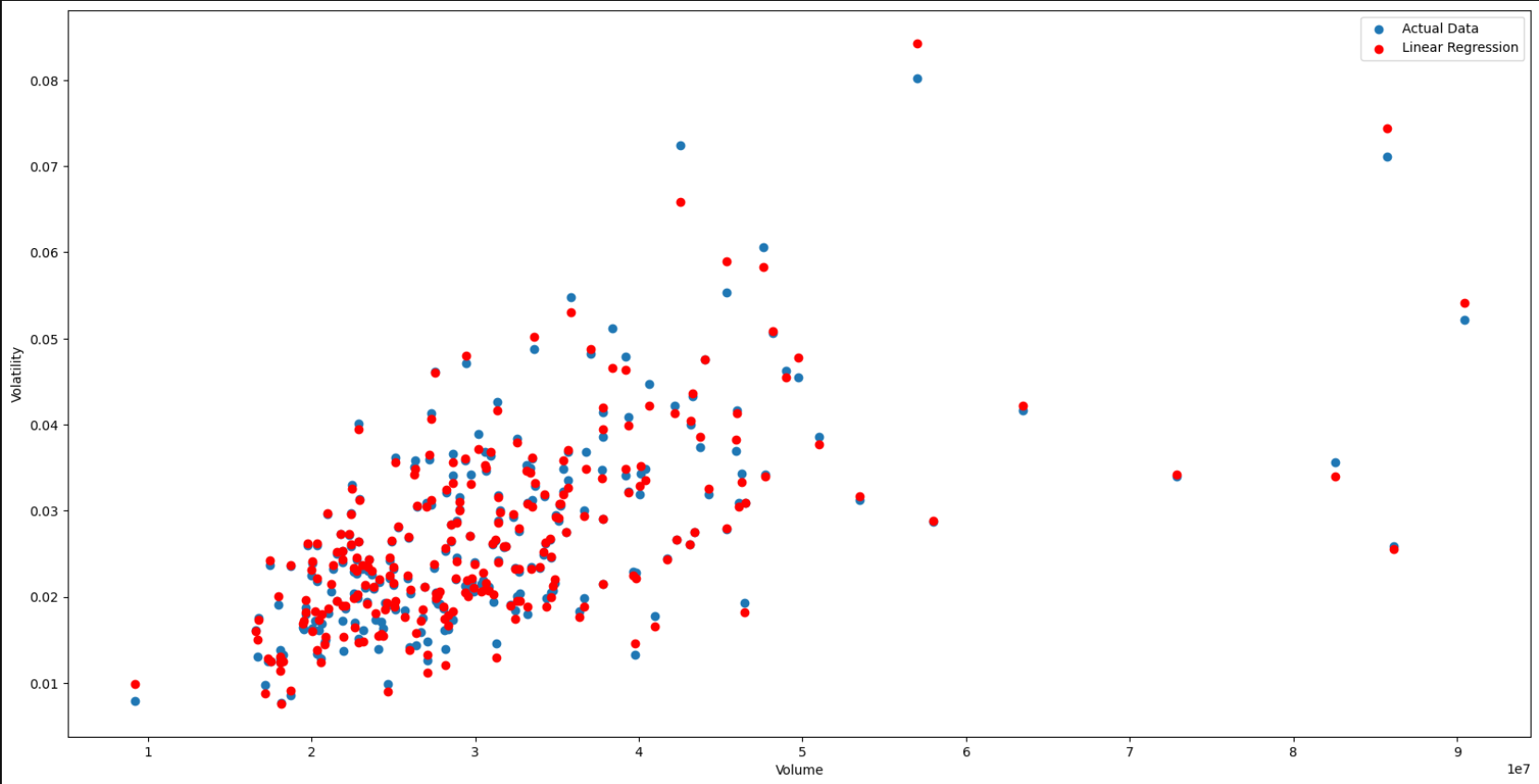
GOOG



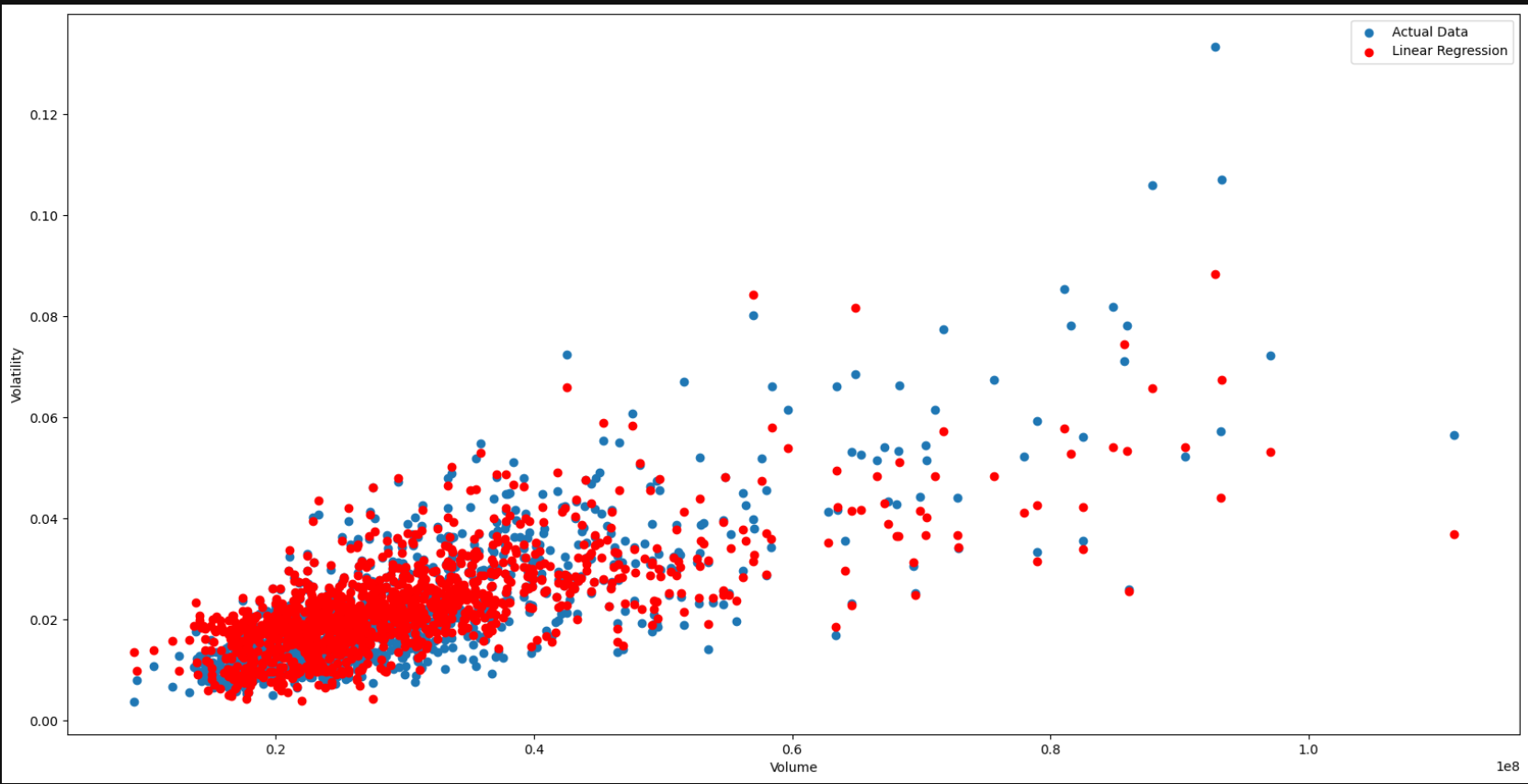
GOOG-5Y

### Microsoft:





MSFT



MSFT-5Y



Results

From the above results, the equations perform well with less mean percentage error and with 5Y data they have error because of the variation in Variables that was taken into consideration, and they cannot represent volatility values more than 0.08. As mentioned in the previous case the equation has dependent variables which also contribute to the error.