**Task Report**

For the dataset which was given to us, the value of correlation coefficient came out to be

**R = 0.147818467**

Correlation coefficient really tells us about the strength of linear relationship between two variables x and y. If R is nearer to 1 means you can form a linear relationship which almost all points will satisfy whereas closer to 0 means such kind of linear relation which all points can satisfy can't be stated much clearly.

It is a small correlation.

As we can see, the value of R is between 0 and 1 and also it is close to 0 which indicates that the linear relationship between adj close and volume is not that strong.

Also, the value of R is positive which means that the slope of the tentative relation that exists between x and y is positive.

Now, we need to perform a regression analysis of the data given to us. Correlation coefficient told us about the strength whereas regression analysis will express the relationship between both.

Trading volume = independent var = x

Adj. Close = dependent var = y

Regression of y on x will give a regression equation i.e. the average value of y as a function of x.

|  |  |
| --- | --- |
| **SUMMARY OUTPUT** |  |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.151958169 |
| R Square | 0.023091285 |
| Adjusted R Square | 0.019152137 |
| Standard Error | 24.4314755 |
| Observations | 250 |

From the above summary output, we can infer that only 2.309% of the variation in adj. close is explained by the trading volume. More closer the value of R2 is to 1, better the regression equation fits the data.

We got the value of coefficients from the excel table and from there we can get the regression equation that is:

**Adj. Close = (3.258 \* 10-7)(trading volume) +(256.44389811472)**

The residual table that we have got from the analysis in excel tells us about how much the values of Adj. Close given in data are far away from the values of Adj. close that we have got after substituting the values of trading volume in the above equation.

Residual Val = (Value from eqn) – (Data Value)

Also, the value of significance F is 0.01619<=0.05 so all the variables are important.

Outlier = A point that is extreme in some way and falls far from other data points.

Now, if the absolute value of any residual is >2\*standard error then it is considered to be an outlier. So, after applying this in my excel sheet, I got to know that 4 observations are outliers that is 98.4% of the data points are within two standard errors. Observations **1 ,206 ,211 ,212** are outliers.

After removing all outliers,

|  |  |
| --- | --- |
| **SUMMARY OUTPUT** |  |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.160366603 |
| R Square | 0.025717447 |
| Adjusted R Square | 0.021724486 |
| Standard Error | 23.97474458 |
| Observations | 246 |

From the above summary output, we can infer that only 2.571% of the variation in adj. close is explained by the trading volume**. So, the percentage change in the value of R2 after removing all outliers is = 11.373%**

Now, regression equation is –

**Adj. Close = (3.354 \* 10-7)(trading volume) +(** **256.345)**

The basic purpose of removing outliers or the points which were influencing the linear relationship was to make the data fit the linear relationship much more perfectly.

**WEEKLY ANALYSIS-**

The value of correlation coefficient is = **R = - 0.050479091(NEGLIGIBLE CORRELATION)**

Negative correlation coefficient implies that as x increases, y decreases in the linear relationship or the slope is negative.

|  |  |
| --- | --- |
| **SUMMARY OUTPUT** |  |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.050479091 |
| R Square | 0.002548139 |
| Adjusted R Square | 0.000150418 |
| Standard Error | 89.63659677 |
| Observations | 418 |

From the above summary output, we can infer that only 0.254% of the variation in adj. close is explained by the trading volume. Closer the value of R2 is to 1, better the regression equation fits the data.

Regression equation-

**Adj. Close = (-8.18909812301665\*10-8)(Trading Volume) + (150.5097)**

Now, if the absolute value of any residual is >2\*standard error then it is considered to be an outlier. So, after applying this in my excel sheet, I got to know that 9 observations are outliers that is 98.4% of the data points are within two standard errors. Outliers are [357,365].

Now, removing all outliers. The R square decreases to 0.001518779. Removing outliers can caused the regression line to shift, and since the removed outliers would have been explaining a significant amount of variation in the dependent variable (Adj. Close), so the R square value decreased.

**MONTHLY ANALYSIS-**

The value of correlation coefficient is = **R = - 0.459997306**

Negative correlation coefficient implies that as x increases, y decreases in the linear relationship or the slope is negative.

**And since the value of R is quite large as compared to previous 2 analysis. So, we can say that in this case, linear plot will be much better. (LOW NEGATIVE CORRELATION)**

|  |  |
| --- | --- |
| **SUMMARY OUTPUT** |  |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.459997306 |
| R Square | 0.211597521 |
| Adjusted R Square | 0.206478025 |
| Standard Error | 79.99288073 |
| Observations | 156 |

From the above summary output, we can infer that 21.16% of the variation in adj. close is explained by the trading volume. Closer the value of R2 is to 1, better the regression equation fits the data.

Regression equation-

**(Adj. Close) = (-1.336\*10-7)(Volume)+( 202.21167972424)**

After removing all the outlies, the value of R2 = 0.240202499

|  |  |
| --- | --- |
| **SUMMARY OUTPUT** |  |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.49010458 |
| R Square | 0.240202499 |
| Adjusted R Square | 0.235033808 |
| Standard Error | 68.71581961 |
| Observations | 149 |

From the above summary output, we can infer that only 24.02% of the variation in adj. close is explained by the trading volume**. So, the percentage change in the value of R2 after removing all outliers is = 14.21%**

Now, regression equation is –

**Adj. Close = (-1.227\*10-7)(Volume)+ (184.1402)**

Now, calculating the RMSE value through excel, RMSE = **21.29122254**

Lower the RMSE, better the regression model fits the linear expression and predicts the actual price.

One way is to increase the number of data points we are considering for calculation but it is not always true. More effective method is to remove outliers and then calculate RMSE.

Calculating the RMSE of full years dataset gives RMSE = **24.62879772**

**CONCLUSION**

1. The stock prices used in analysis is Adj. Close which means Adjustment close which includes the value of close of stock price + the dividends given and splitting of face value. The volume is the number of trades happened during that time period. We need to draw a correlation between these two.
2. The basic need to find the correlation and do the regression analysis is to help us in the stock market in future. Getting a linear equation would help us know that will the we profit from the market in future or go in loss. The regression model provided an equation that can be used to predict the stock price based on the trading volume which will tell us when to buy and when to sell.
3. Also, we can conclude that some data points (called outliers) tend to disturb the dependency of y and x. These observations can be ignored while doing the analysis in order to get better relations.
4. Just like we did between Adj. Close and volume, we can develop relations between other quantities.

**Volatility** means the fluctuation the stock market of a particular company or index. It also has its own index which measures and plots it known as VIX.

**Liquidity** as the word suggests is the easy flow of any asset. It means that people are readily available to buy that asset. For e.g., people are readily available to trade stocks with others in exchange of money. It ensures you can easily buy and sell stocks in the market.

Higher the volume, more the liquidity.

Lower the volume, lesser the liquidity.

Higher the volatility, higher the volume.

Lower the volatility => volume can be both high and low.