SUMMARY STATISTICS OF THE NUMERIC VARIABLES

> summary(final\_table$Volume)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 0.000e+00 | 3.267e+07 | 9.545e+08 | 1.134e+10 | 1.665e+10 | 3.510e+11 |

> summary(final\_table$Marketcap)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 7.784e+08 | 6.299e+09 | 3.755e+10 | 1.205e+11 | 1.500e+11 | 1.186e+12 |

> summary(final\_table$avg\_price)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 74.21 | 429.48 | 2286.58 | 6683.85 | 8486.63 | 63316.73 |

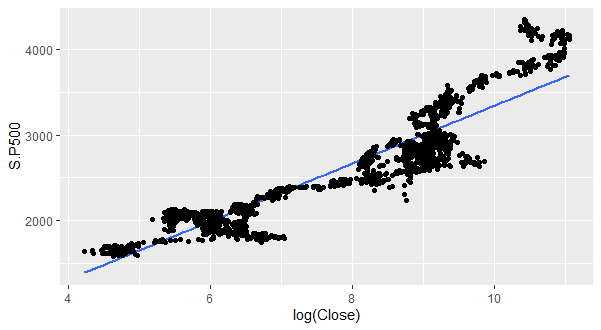
> summary(final\_table$S.P500)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 1573 | 2038 | 2403 | 2517 | 2884 | 4352 |

**CORRELATION BETWEEN DEPENDENT VARIABLE 'CLOSE' AND THE INDEPENDENT 'S.P500'**

> cor(final\_table[c("Close", "S.P500")])

|  |  |  |
| --- | --- | --- |
|  | Close | S.P500 |
| Close | 1.0000000 | 0.8292034 |
| S.P500 | 0.8292034 | 1.00000 |

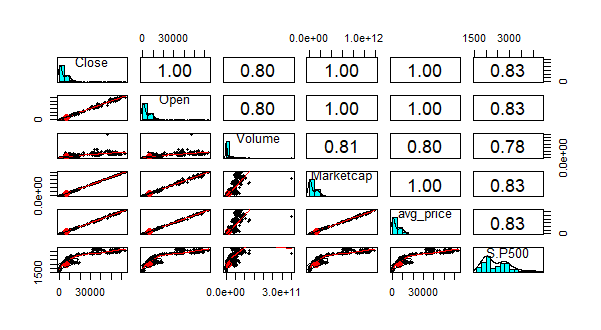


**Here, 0.829 is a strong positive correlation between the Closing price of BTC and stock price. More so it is a positive correlation meaning, and upward rise in stock price causes a rise in the price of BTC as well.**

**CORRELATION MATRIX TO KNOW THE VARIABLES TO PUT IN THE MODEL**

> cor(final\_table[c("Close", "Open", "Volume", "Marketcap", "avg\_price", "S.P500")])

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Close | Open | | Volume | Marketcap | avg\_price | S.P500 |
| Close | | 1.0000000 | 0.9983747 | | 0.8040466 | 0.9997132 | 0.9995942 | 0.8292034 |
| Open | | 0.9983747 | 1.0000000 | 0.8048494 | | 0.9981193 | 0.9995929 | 0.8290565 |
| Volume | | 0.8040466 | 0.8048494 | 1.0000000 | | 0.8057855 | 0.8047747 | 0.7837008 |
| Marketcap | | 0.9997132 | 0.9981193 | 0.8057855 | | 1.0000000 | 0.9993230 | 0.8263721 |
| avg\_price | 0.9995942 | | 0.9995929 | 0.8047747 | | 0.9993230 | 1.0000000 | 0.8294671 |
| S.P500 | 0.8292034 | | 0.8290565 | 0.7837008 | | 0.8263721 | 0.8294671 | 1.0000000 |

**This is a correlation matrix of all the variables. All the variables are strongly correlated with each other, so we need not include them in our model. This is one way to avoid data leakage**The above matrix shows all the variables are strongly correlated. Also, the histogram shows they are heavily skewed. Therefore, we had to use logarithmic transform for our dependent variable ‘close’.

The essence of the logarithmic transformation is to normalize the heavily skewed variable.

**MODEL SUMMARY**

Residuals:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | 1Q | Median | 3Q | Max |
| -1.78108 | -0.49692 | -0.07967 | 0.48249 | 1.92422 |

Coefficients:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | | 1.141e+00 | 7.153e-02 | 15.95 | <2e-16 \*\*\* |
| **S.P500** | 2.551e-03 | | 2.758e-05 | 92.51 | <2e-16 \*\*\* |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Residual standard error**: 0.658 on 1441 degrees of freedom

**Multiple R-squared**: 0.8559, **Adjusted R-squared**: 0.8558

F-statistic: 8559 on 1 and 1441 DF, p-value: < 2.2e-16

**The t value observed is statistically significant. What this means is that our model fits.**

**Multiple R squared: is the amount of variation in the dependent variable that can be explained by the independent variable. Here, 85.6% of the variation in the closing price of BTC can be explained by the variation in stock price.**

**Adjusted R squared has the same rendition as multiple R squared however in a case of multiple independent variable.**

**EVALUATION METRICS FOR PREDICTION ON THE TEST\_SET, TRAIN\_SET, RIDGE REGRESSION AND LASSO REGRESSION**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | train\_set | test\_set | Ridge regression | Lasso regression |
| MAE | 5194.971 | 3794.871 | 842.356 | 265.4229 |
| MSE | 303821174 | 172182335 | 1810835 | 124457.8 |
| RMSE | 17430.47 | 13121.83 | 1345.673 | 352.7858 |

**MSE: The lower the MSE, the better the model. The best of the model is the Lasso regression model.**

**RMSE: A lower RMSE is better than a higher RMSE, the model performed better on the testing set than on the training set. The best model is the Lasso regression model.**

**CONCLUSION**

**Bitcoin closing price and the SP500 price are highly correlated, that is investors can monitor the price of the SP500 to make decisions as regards to bitcoin buying and selling and vice versa.**

**The correlation is a positive one, meaning an upward move in the price of the SP500 will result in an upward move in the price of bitcoin while a downward move in the price of SP500 will result in a downward move in the price of bitcoin as well. It is important however, to note that correlation is not causation, this means that the fluctuations in the price of the SP500 is not what is causing the changes seen in the price of bitcoin, rather it simply imply that there is a relationship between the two.**

**We could use this conclusion to predict Bitcoin prices to help investors understand the potential investment opportunity for Bitcoin as the S&P 500 has a more detailed history and Bitcoin is relatively new as an investment.**

**The best model to make predictions to base investing decision on is the Lasso regression model evidenced by the fact that it has the least RMSE. In building the Lasso regression model, Market capital was added to the independent variable. What this means is that to get a better prediction value, in addition to the stock price, including the market cap can help us make better prediction and by extension better investment decisions.**