**SAS Project**

**Bitcoin Project Prediction**

**DESCRIPTION OF VARIABLES:**

coin\_avg (Outcome Variable) - It is the weekly average price of the bitcoin.

new\_wallets – Number of active wallets in a week.

coin\_day – Number of sold bitcoins aggregated to a weekly level.

buy\_bitcoin – Percentage of google searches for the bitcoin aggregated to a weekly level.

avg\_dxy\_value – Average dollar value aggregated to a weekly level.

gold – Price of gold aggregated to a weekly level.

Intel\_avg – Average price of intel stock aggregated to a weekly level.

wallets\_total – Total number of new wallets aggregated to a weekly level.

snp\_close – The value of snp at the end of the week or the closing value of snp at the end of the week.

The above variables were first-differenced or differenced once to reduce the autocorrelation among the variables for the regression.

**REGRESSION TABLES:**

1. **Regression table with all predictor variables**

|  |
| --- |
| **The SAS System** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: coin\_avg\_diff**

|  |  |
| --- | --- |
| **Number of Observations Read** | 206 |
| **Number of Observations Used** | 205 |
| **Number of Observations with Missing Values** | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** **Squares** | **Mean** **Square** | **F Value** | **Pr > F** |
| **Model** | 8 | 39534253 | 4941782 | 10.17 | <.0001 |
| **Error** | 196 | 95259732 | 486019 |  |  |
| **Corrected Total** | 204 | 134793985 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 697.15066 | **R-Square** | 0.2933 |
| **Dependent Mean** | 113.99637 | **Adj R-Sq** | 0.2644 |
| **Coeff Var** | 611.55515 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** **Estimate** | **Standard** **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | -302.12776 | 88.72618 | -3.41 | 0.0008 |
| **new\_wallet\_diff** | 1 | 0.00473 | 0.00108 | 4.36 | <.0001 |
| **coin\_day\_diff** | 1 | 0.00000448 | 0.00012133 | 0.04 | 0.9706 |
| **buy\_bitcoin\_diff** | 1 | 19.64082 | 9.42561 | 2.08 | 0.0385 |
| **avg\_dxy\_value\_diff** | 1 | 83.78310 | 96.25677 | 0.87 | 0.3851 |
| **gold\_diff** | 1 | 2.54638 | 1.65496 | 1.54 | 0.1255 |
| **Intel\_avg\_diff** | 1 | -46.74957 | 34.73712 | -1.35 | 0.1799 |
| **wallets\_total\_diff** | 1 | 0.00155 | 0.00029657 | 5.23 | <.0001 |
| **snp\_close\_diff** | 1 | 2.51185 | 1.17566 | 2.14 | 0.0339 |

**Explanation:**

The R-squared value of the regression model is 0.2933 which indicates that 29.33% variability can be explained by the predictor variables. The F-value is 10.17 with a corresponding value of < 0.0001 which is less than the significance level of 0.01. This means that the model is statistically significant.

The new\_wallet\_diff is the first-differenced value for the number of active wallets per week. The t-value (4.36) with the corresponding p-value of <.0001 which is less than significance level of 0.05 indicates that the new\_wallet\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicates that for every increase in 1000 active wallets, the value of bitcoin goes up by $4.73.

The coin\_day\_diff is the first-differenced value for the number of bitcoins sold per week. The t-value (0.04) with the corresponding p-value of 0.9706 which is more than significance level of 0.05 which indicates that the coin\_day\_diff is not statistically significant (not responsible for variability in bitcoin price).

The buy\_bitcoin\_diff is the first-differenced value for the percentage of google searches for buying bitcoins per week. The t-value (2.08) with the corresponding p-value of <.0.0385 which is less than significance level of 0.05 indicates that the buy\_bitcoin\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicates that for every 1% increase in google searches, the value of bitcoin goes up by $19.64.

The avg\_dxy\_value\_diff is the first-differenced value for the value of US dollar per week. The t-value (0.87) with the corresponding p-value of 0. 3851 which is more than significance level of 0.05 which indicates that the avg\_dxy\_value\_diff is not statistically significant (not responsible for variability in bitcoin price).

The gold\_diff is the first-differenced value for the price of gold per week. The t-value (1.54) with the corresponding p-value of 0.1255 which is more than significance level of 0.05 which indicates that the gold\_diff is not statistically significant (not responsible for variability in bitcoin price).

The Intel\_avg\_diff is the first-differenced value for the average price of intel stock per week. The t-value (-1.35) with the corresponding p-value of 0.1799 which is more than significance level of 0.05 which indicates that the Intel\_avg\_diff is not statistically significant (not responsible for variability in bitcoin price).

The wallets\_total\_diff is the first-differenced value for the number of new wallets per week. The t-value (5.23) with the corresponding p-value of <.0001 which is less than significance level of 0.05 indicates that the wallets\_total\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicate that for every 1000 new wallets, the value of the bitcoin goes up by $1.55.

The snp\_close\_diff is the first-differenced value for the closing value of snp per week. The t-value (2.14) with the corresponding p-value of <.0.0339 which is less than significance level of 0.05 indicates that the snp\_close\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicate that for every 1% increase in closing value of snp, the value of the bitcoin goes up by $2.51.

Out of all the 8 predictor variables, only new\_wallet\_diff, buy\_bitcoin\_diff, wallets\_total\_diff and snp\_close are statistically significant whereas the other variables, coin\_day\_diff, avg\_dxy\_value\_diff, gold\_diff and Intel\_avg\_diff are not statistically significant. Hence, the non-significant predictor variables have been dropped in the final (parsimonious) model which can be found below.

1. **FINAL MODEL: Regression table with only the significant predictor variables**

|  |
| --- |
| **The SAS System** |

**The REG Procedure**

**Model: MODEL1**

**Dependent Variable: coin\_avg\_diff**

|  |  |
| --- | --- |
| **Number of Observations Read** | 206 |
| **Number of Observations Used** | 205 |
| **Number of Observations with Missing Values** | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of** **Squares** | **Mean** **Square** | **F Value** | **Pr > F** |
| **Model** | 4 | 37311774 | 9327943 | 19.14 | <.0001 |
| **Error** | 200 | 97482211 | 487411 |  |  |
| **Corrected Total** | 204 | 134793985 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 698.14830 | **R-Square** | 0.2768 |
| **Dependent Mean** | 113.99637 | **Adj R-Sq** | 0.2623 |
| **Coeff Var** | 612.43030 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter** **Estimate** | **Standard** **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 1 | -297.69648 | 86.82029 | -3.43 | 0.0007 |
| **new\_wallet\_diff** | 1 | 0.00496 | 0.00108 | 4.61 | <.0001 |
| **buy\_bitcoin\_diff** | 1 | 19.83409 | 9.32004 | 2.13 | 0.0346 |
| **wallets\_total\_diff** | 1 | 0.00155 | 0.00028805 | 5.40 | <.0001 |
| **snp\_close\_diff** | 1 | 1.77398 | 0.84937 | 2.09 | 0.0380 |

**Explanation:**

The R-squared value of the regression model is 0.2768 which indicates that 27.68% variability can be explained by the predictor variables. The F-value is 19.14 with a corresponding value of < 0.0001 which is less than the significance level of 0.01. This means that the model is statistically significant. This model only contains significant predictor variables; hence it is the final regression model and can also be called as “parsimonious model”.

The new\_wallet\_diff is the first-differenced value for the number of active wallets per week. The t-value (4.61) with the corresponding p-value of <.0001 which is less than significance level of 0.05 indicates that the new\_wallet\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicates that for every increase in 1000 active wallets, the value of bitcoin goes up by $4.96.

The buy\_bitcoin\_diff is the first-differenced value for the percentage of google searches for buying bitcoins per week. The t-value (2.13) with the corresponding p-value of <.0.0346 which is less than significance level of 0.05 indicates that the buy\_bitcoin\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicates that for every 1% increase in google searches, the value of bitcoin goes up by $19.83.

The wallets\_total\_diff is the first-differenced value for the number of new wallets per week. The t-value (5.40) with the corresponding p-value of <.0001 which is less than significance level of 0.05 indicates that the wallets\_total\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicate that for every 1000 new wallets, the value of the bitcoin goes up by $1.55.

The snp\_close\_diff is the first-differenced value for the closing value of snp per week. The t-value (2.09) with the corresponding p-value of <.0.0380 which is less than significance level of 0.05 indicates that the snp\_close\_diff is statistically significant (responsible for variability in bitcoin price). The parameter estimates indicate that for every 1% increase in closing value of snp, the value of the bitcoin goes up by $1.77.

Out of all the significant predictor variables, wallets\_total\_diff has the highest t-value. This could indicate that wallets\_total or the total number of new wallets is the strongest predictor among the significant predictor variables.