**DATA VISULAIZATION**

**CRYPTOCURRENCIES**

**SUBMITTED BY**

**MALAV DANI**

**AKSHARA KUMAR**

**SHUBHANGI GARG**

**LIKITHA CHAGANTI**

**CRYPTOCURRENCIES**

1. **SOURCE DATA FILES:**

* The Primary Data set is from: <https://www.kaggle.com/jessevent/all-crypto-currencies>
* The Associate Dataset is from: <https://www.investing.com/commodities/gold-historical-data>
* The Associate Data set is from: (Paste the Link)
  1. **DESCRIPTIONS**
* The primary data source is “Cryptocurrencies.xls” which has 583,174 rows.
* There are Two secondary original data sources, listed as follows:
* USD\_Historical\_Data.xls with 2097 rows
* Gold \_Futures\_Historical\_Data.xls with 2274 rows

1. **DATA** **CLEANING/MANIPULATION & JOINS:**

Paste the Tableau Joins Here

We have used **INNER JOIN** on the dataset. At first, we performed the OUTER JOIN for combining all the sheets and to preserve all the values. Since there were few missing values we preferred using INNER JOIN on the datasets.

1. **DATA ANALYSIS:**

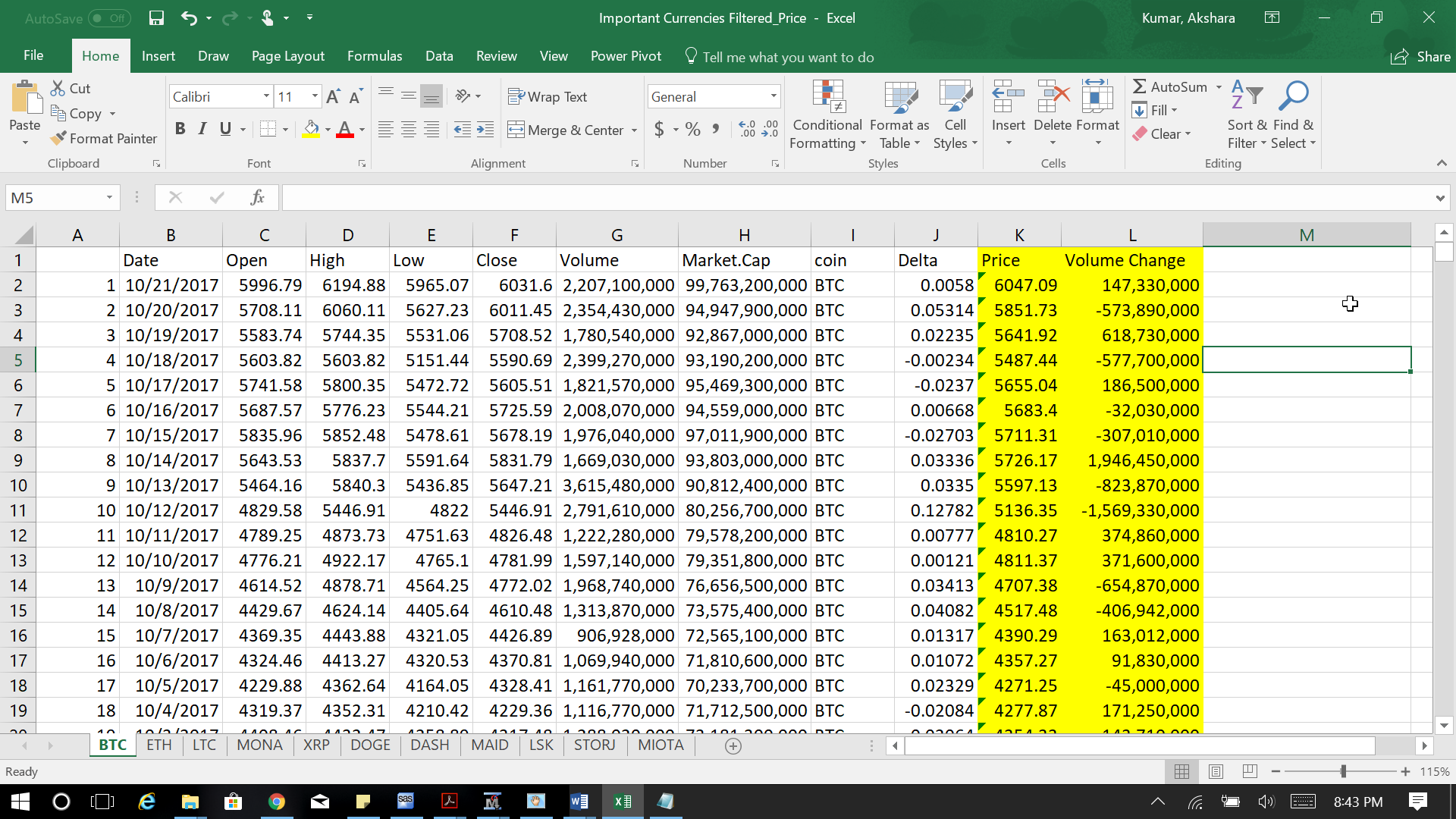
We have performed the following Ratios:

* **Price:**

Price was calculated by taking the Average of the Variables (Low, high, Open, Close)

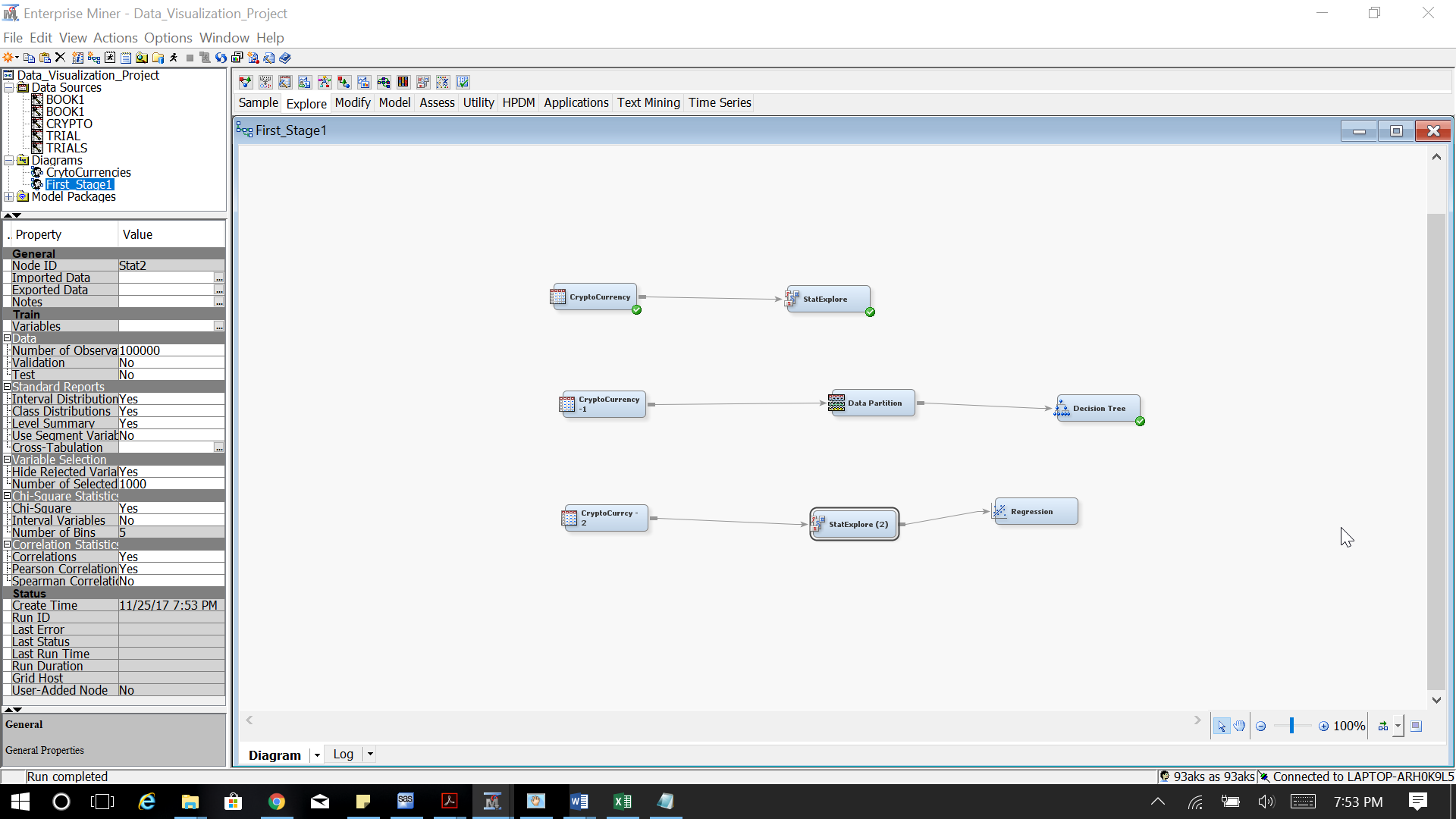
* **Volume\_Change:**

Volume Change can be calculated by taking the difference between the Volume of the Current day to the Volume of the previous day



1. **STATISTICAL ANALYSIS:**

We gave used the SAS Enterprise Miner to perform the Statistical Analysis. We have exported the Crypto currencies file and found the important variables for the analysis and, we have used the Decision Tree Model for the Analysis.



* 1. **VARIABLE WORTH**

We Calculated the Variable Worth. It is the method which is based on binning the input variable to identify their maximum log worth which is known as the Variable Worth.

A screenshot of a computer

Description generated with very high confidence

Hence, from the output. We can identify that the Variables (Low, high, Open, Close, Market Cap) with correlations of above 99 Percentage and more.

The all 5 Variables have equal worth for identifying the Target ‘Price’ hence we can use them for our analysis.

Keeping ‘Price’ as Target Variable, Correlations are shown for the input variables

A screenshot of a cell phone

Description generated with high confidence

A screenshot of a social media post

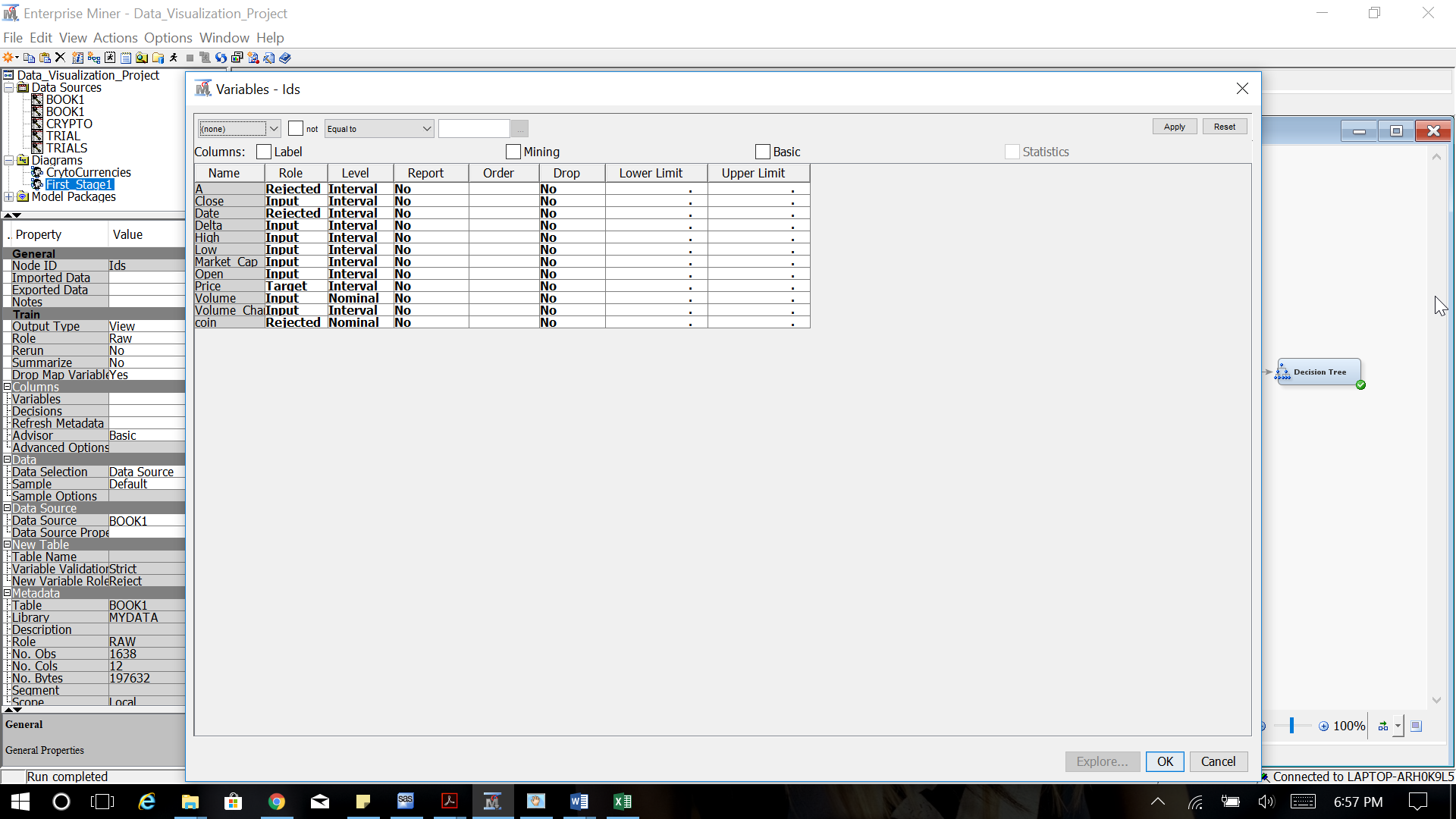
Description generated with very high confidence

* 1. **DECISION TREE**

We Performed the Decision tree analysis. Variable Importance is calculated using the SAS decision tree methodology. It attempts to evaluate the overall value or importance of the variable over the fitted tree. The variable used to split the root node impacts every observation while those which split nodes lower in the tree impact a smaller number of observations. Variable Importance is impacted by both the number of observations impacted and the purity of the resulting split.

By taking the ‘Price’ column as the Target variable. We have also rejected the Columns such as:

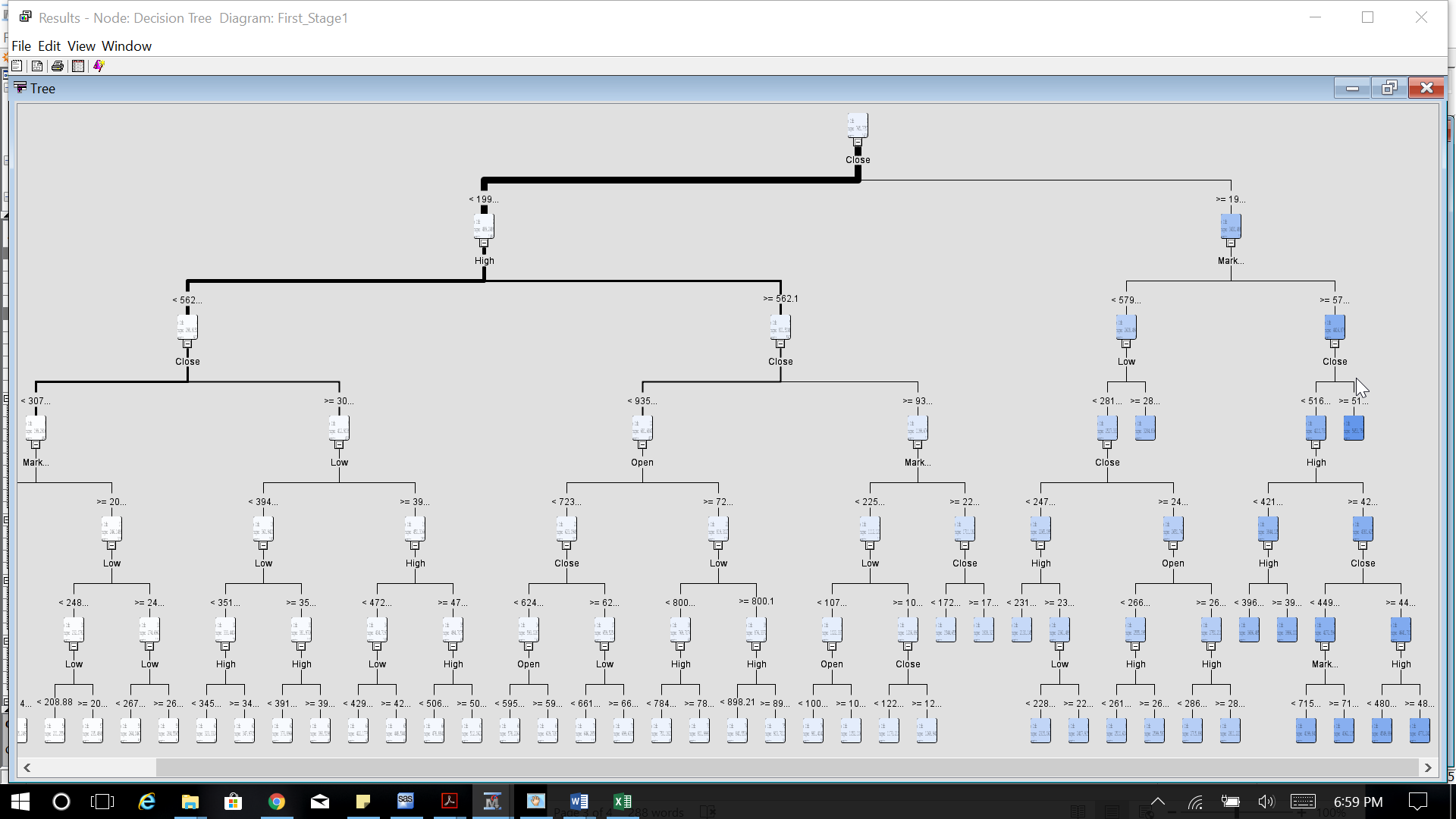
* **Date** (Every day’s record from the year 3013-1017)
* **A - ID** (Which represents the count of the entries)
* **Coin** (Represents the Name of the Currencies)
* **Volume\_Change** (Newly Created column to find the change in the volume between the current and previous day)



We then partition our date by taking 70% of the data for Training, 30% of the data for the Validation.

The Decision Tree’s Output is shown below:

‘Close’ variable is chosen for the first split



**TARGET VARIABLE: ‘PRICE’**

We can find that the Most Significant variables are:

* **High**: The highest Price for the given day
* **Open**: The opening price is the price at which a security first trades upon the opening of an exchange on a given day
* **Close:** the price of a security at the end of the day's business in a financial market.
* **Low**: The Lowest price for the given Day

**SAMPLE NODE RULES FOR THE DECISION TREE:**

**NODE RULES:**

**Node = 13**

if Market.Cap < 5.8E10 or MISSING

AND Low >= 2818.08

AND Close >= 1993.29

then

Tree Node Identifier = 13

Number of Observations = 9

Predicted: Price = 3284.8363889

**Node = 15**

if Market.Cap >= 5.8E10

AND Close >= 5169.46

then

Tree Node Identifier = 15

Number of Observations = 10

Predicted: Price = 5653.7565

**Node = 42**

if Market.Cap >= 2.26E10

AND High >= 562.1

AND Close < 1723.8 AND Close >= 935.525

then

Tree Node Identifier = 42

Number of Observations = 7

Predicted: Price = 1544.6557143

**Node = 43**

if Market.Cap >= 2.26E10

AND High >= 562.1

AND Close < 1993.29 AND Close >= 1723.8 or MISSING

then

Tree Node Identifier = 43

Number of Observations = 12

Predicted: Price = 1808.32375

**Node = 44**

if Market.Cap < 5.8E10 or MISSING

AND Low < 2818.08 or MISSING

AND High < 2315.95

AND Close < 2471.52 AND Close >= 1993.29

then

Tree Node Identifier = 44

Number of Observations = 10

Predicted: Price = 2130.145

**Node = 48**

if Market.Cap >= 5.8E10

AND High < 3960.07

AND Close < 5169.46 AND Close >= 1993.29 or MISSING

then

Tree Node Identifier = 48

Number of Observations = 9

Predicted: Price = 3686.4855556

1. **INSIGHTS:**

**Question 1:**

**Insight 1:**

**Question 2:**

**Insight 2:**

**Question 3:**

**Insight 3:**

1. **VISUALIZATION FRAMES:**

**Visualization 1:**

**Visualization 2:**

**Visualization 3:**

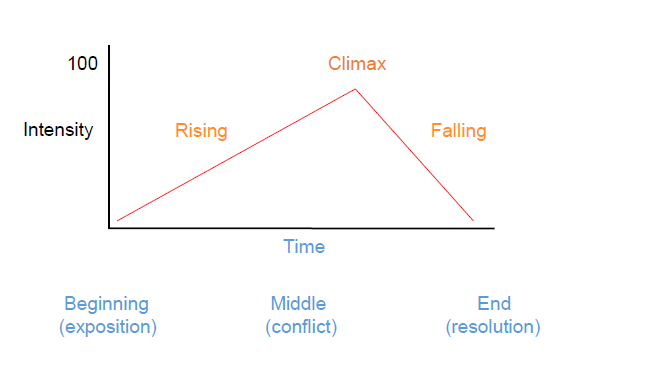
**Visualization 4:**

**Visualization 5:**

**Visualization 6:**

1. **VISULAIZATION STORY:**

We use Freytag’s model as our type of the story line.



**STORYLINE TYPE**: **Freytag’s Model**

According to Frey, the story is divided into 5 various stages:

* Beginning (exposition)
* Rising
* Middle (conflict)
* Falling
* End (resolution)

**CHARACTERS:**

* Gold historical Data
* US Dollar Historical Data
* Various Important Crypto Currencies

**RHETORIC TYPE USED: Logical**

1. **Beginning (exposition)**
2. **Rising**
3. **Middle (conflict)**
4. **Falling**
5. **End (resolution)**
6. **VISUALIZATION & ENCODING METHODS**

The visual encoding is the way in which data is mapped into visual structures, upon which we build the images on a screen.

|  |  |
| --- | --- |
| **VISUALIZATION** | **DATA ENCODING MODEL & EXPLANATIONS** |
| **Visualization 1** |  |
| **Visualization 2** |  |
| **Visualization 3** |  |
| **Visualization 4** |  |
| **Visualization 5** |  |
| **Visualization 6** |  |
|  |  |