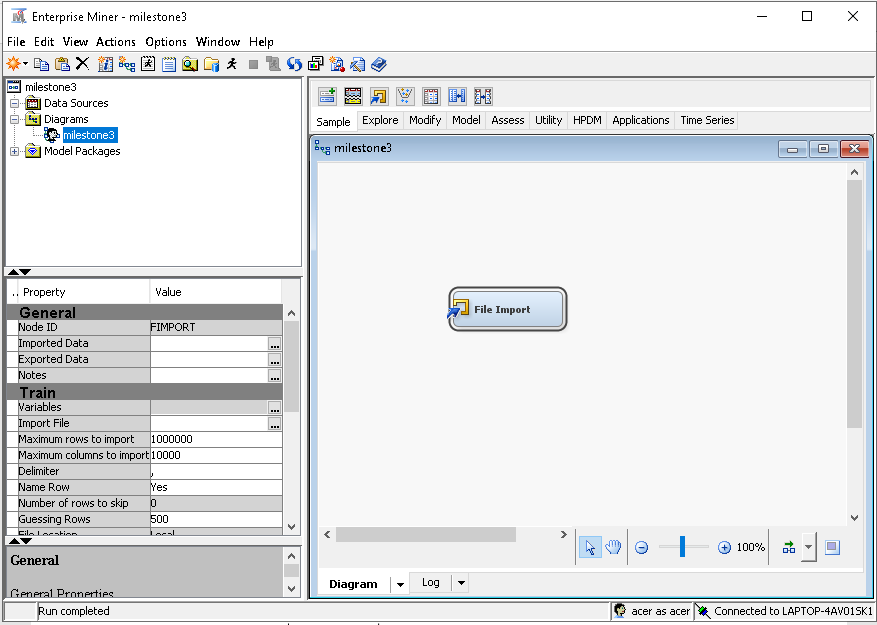
**Analysis Goal**

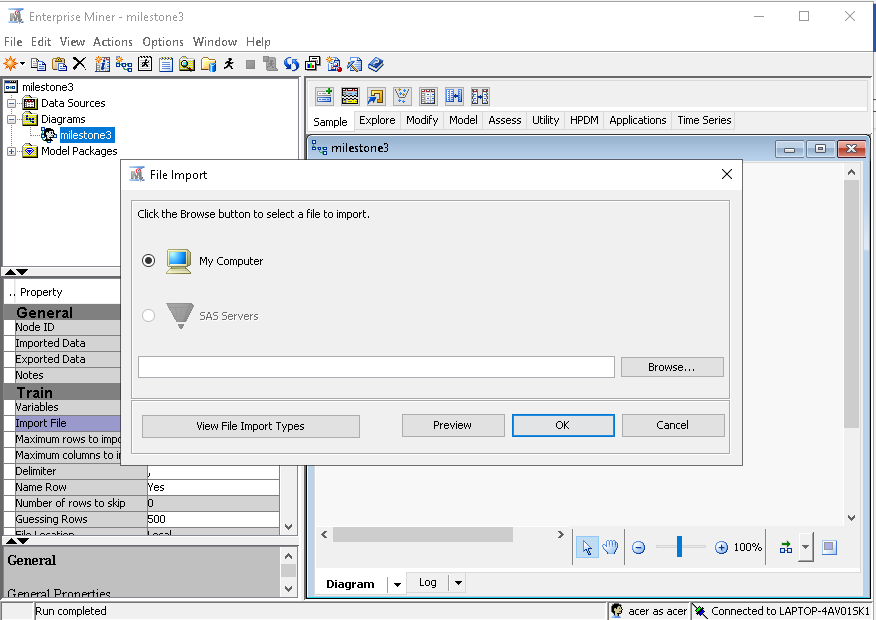
To apply machine learning methods in order to effectively predict future cryptocurrency price movements

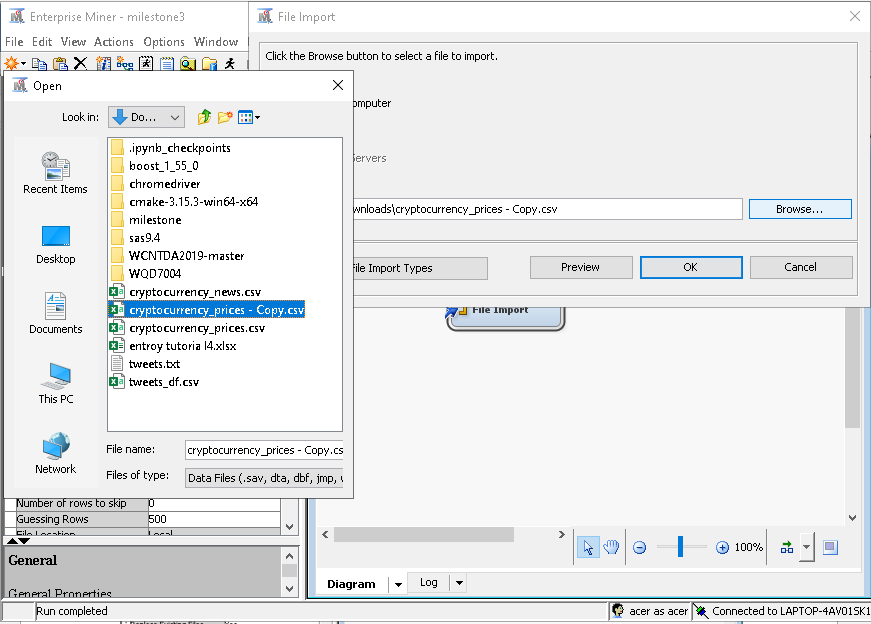
**Analysis Data**

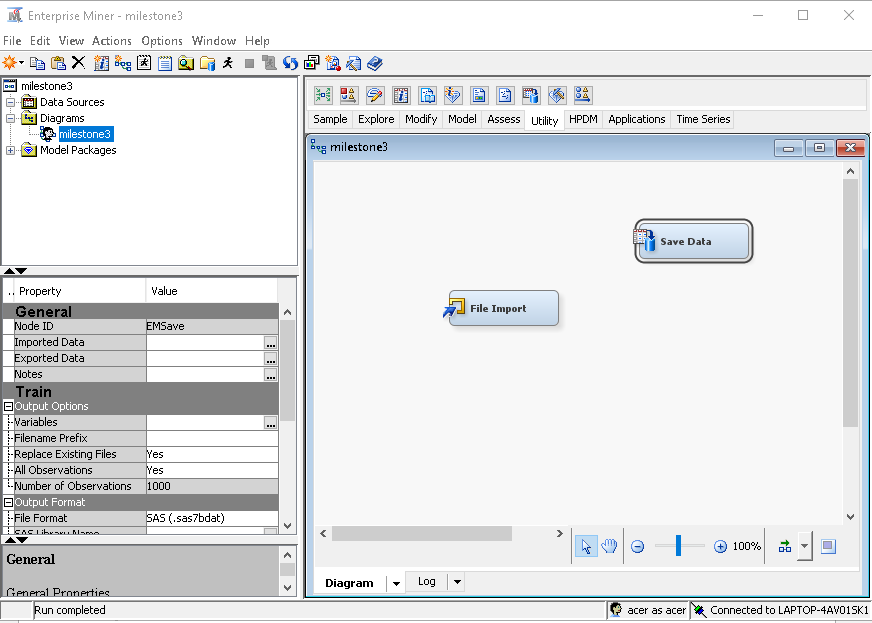
To gain better understanding how cryptocurrency prices are affected by market sentiments, political and economic climates.

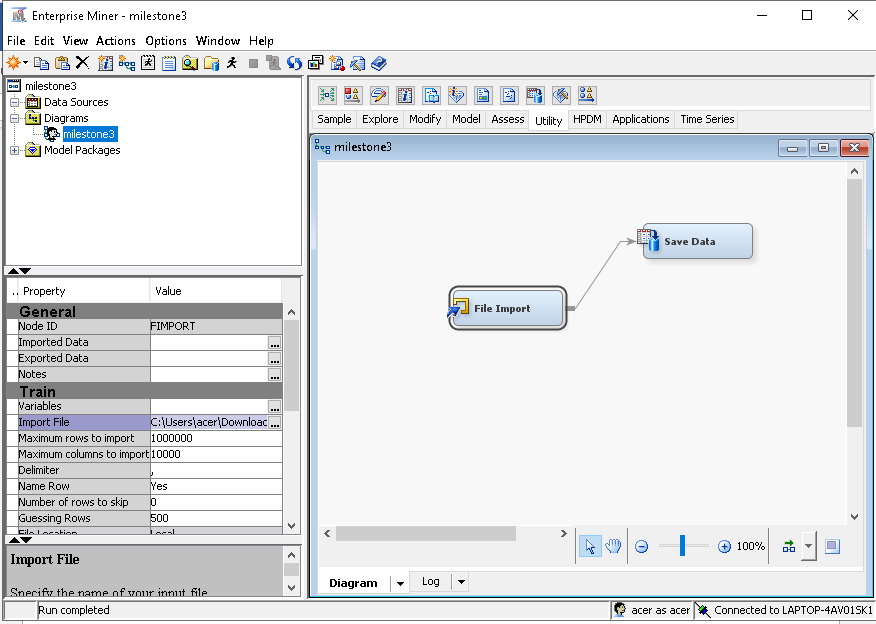
**Import CSV file into SAS Enterprise Miner**

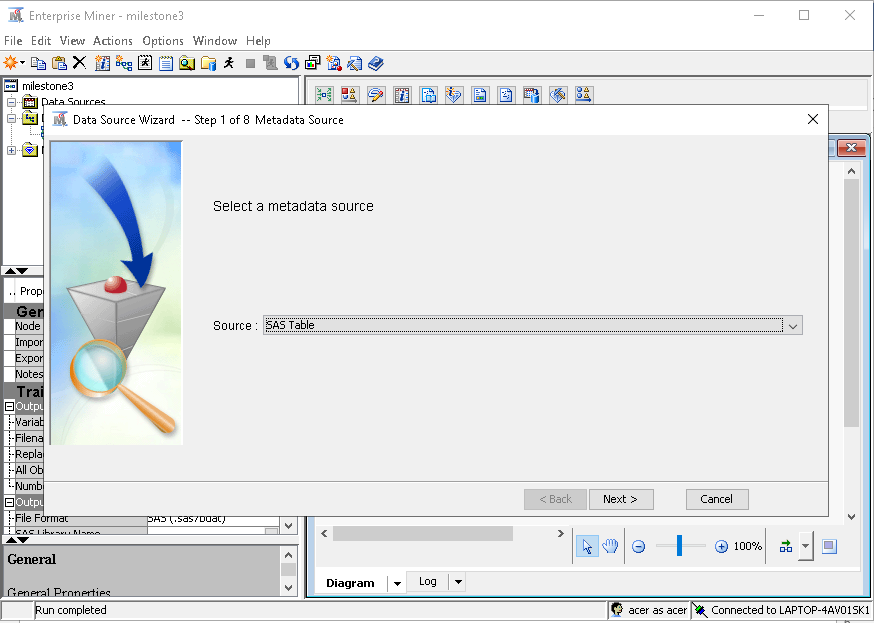
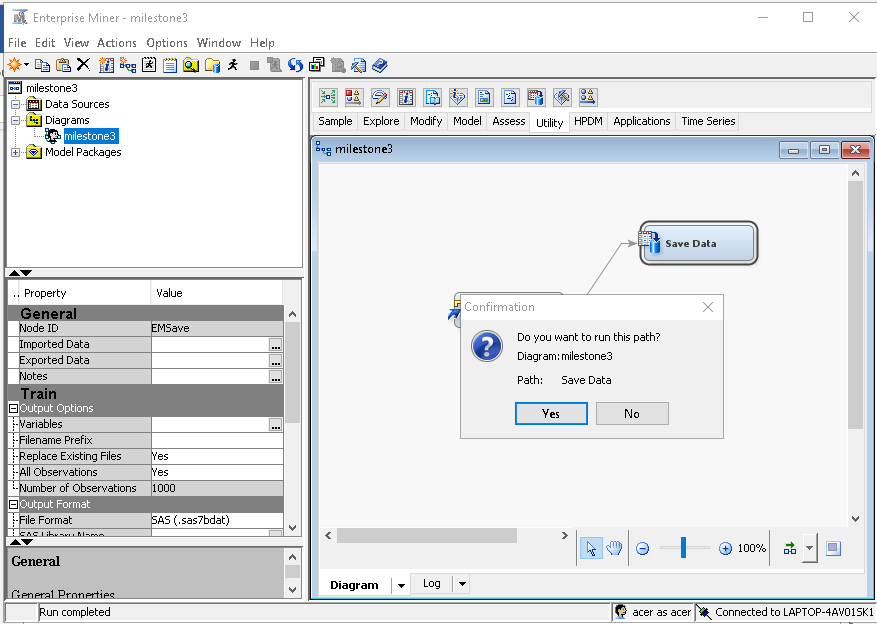
Drag the **File Import** node from the **Sample** Tab. Then go to the **Import File** to import the csv file

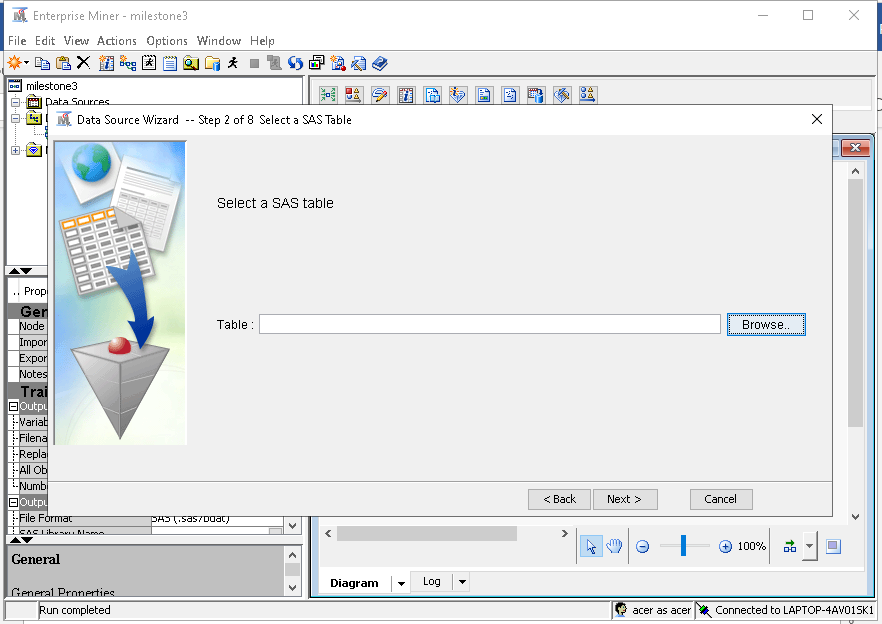
Browse the local directory and then click **OK**

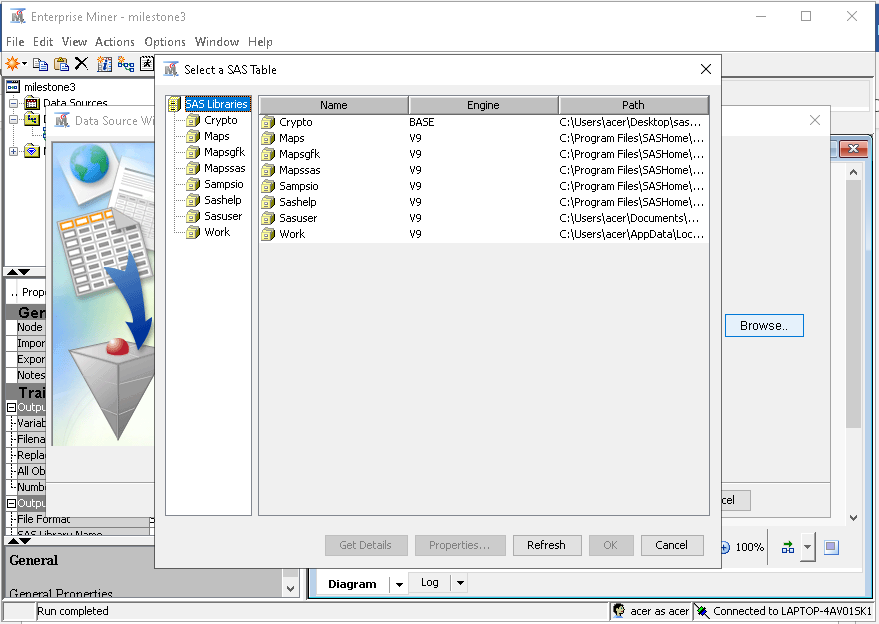


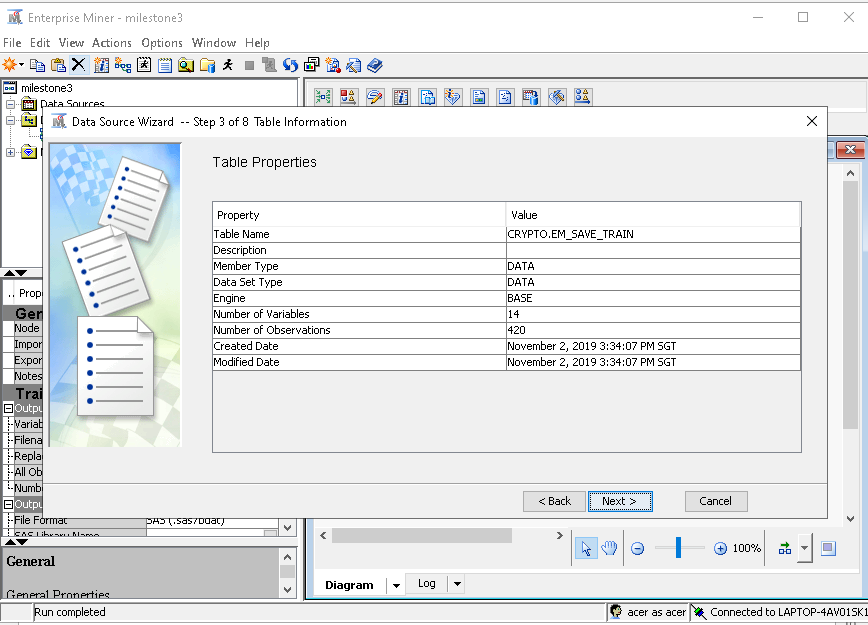
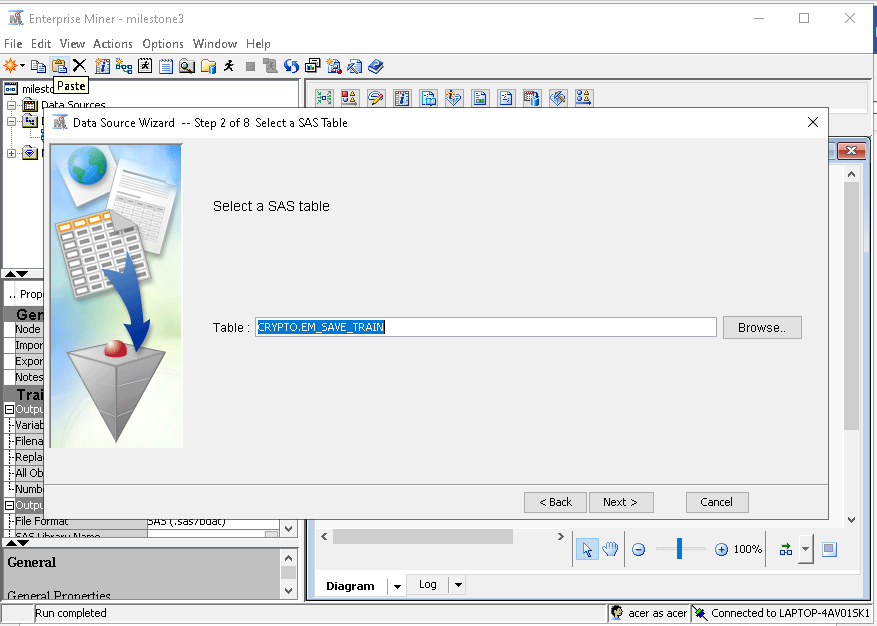
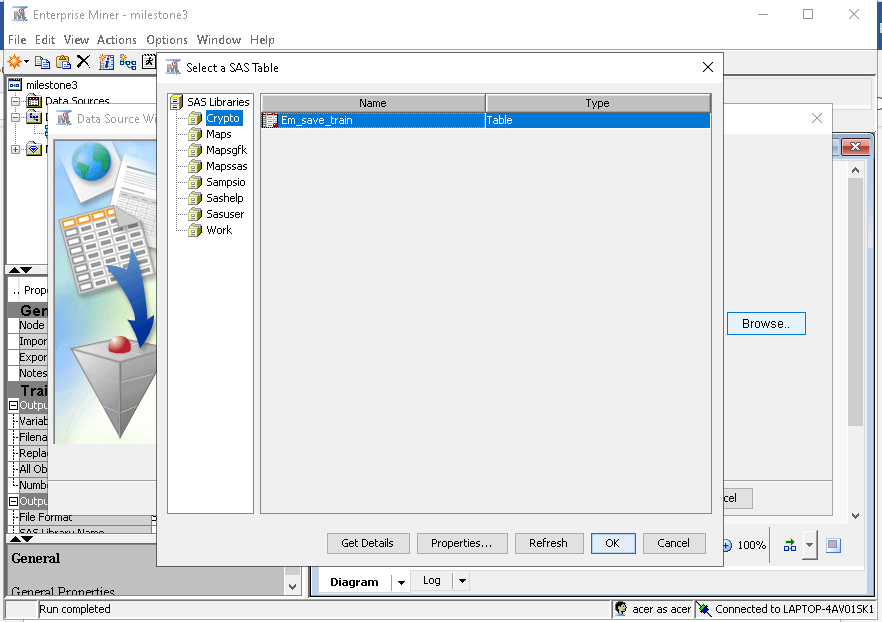
Next, pull the **Save Data** node from the **Utility** Tab

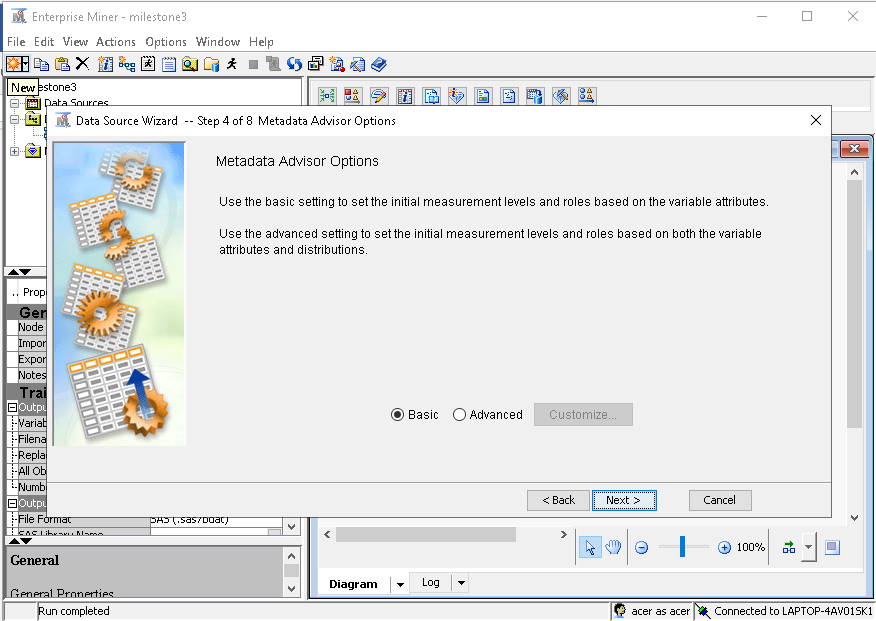
Connect the two nodes by pulling an arrow across. Then proceed the run the nodes.

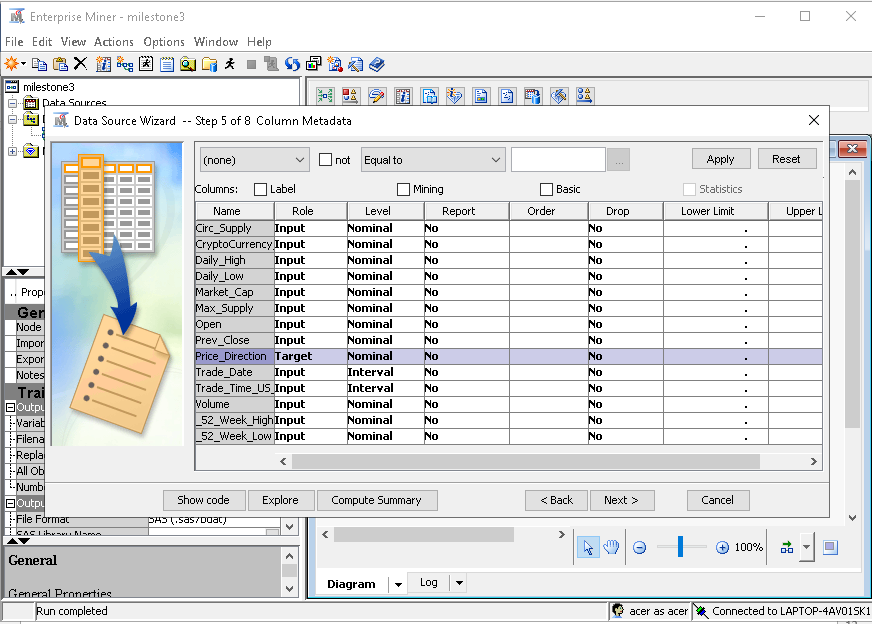
Proceed the create a new **Data Source**, go to **File**, **New** and click **Data Source**. It will prompt you to the SAS Table, click **Next.**



We have previously saved the csv file in the Library called Crypto. So click that SAS Table and click the table in it.

As can be seen above, the data consists of 420 rows with 14 columns.





Description of Dataset

**Circ\_Supply**: Number of a particular cryptocurrency in market circulation(billion)

**Cryptocurrency**: Type of cryptocurrency

**Daily\_High**: Highest value reached for the particular cryptocurrency

**Daily\_Low**: Lowest value reached for the particular cryptocurrency

**Market\_Cap**: Share volume of cryptocurrency traded in the market(billion)

**Max.Supply:** Maximum volume of cryptocurrencies traded in the market(billion)

**Open**: Price of the cryptocurrency today

**Prev.Close**: Price of cryptocurrency the day before

**Price\_Direction**: Cryptocurrency price movements

**Trade\_Date**: Date of Traded cryptocurrency

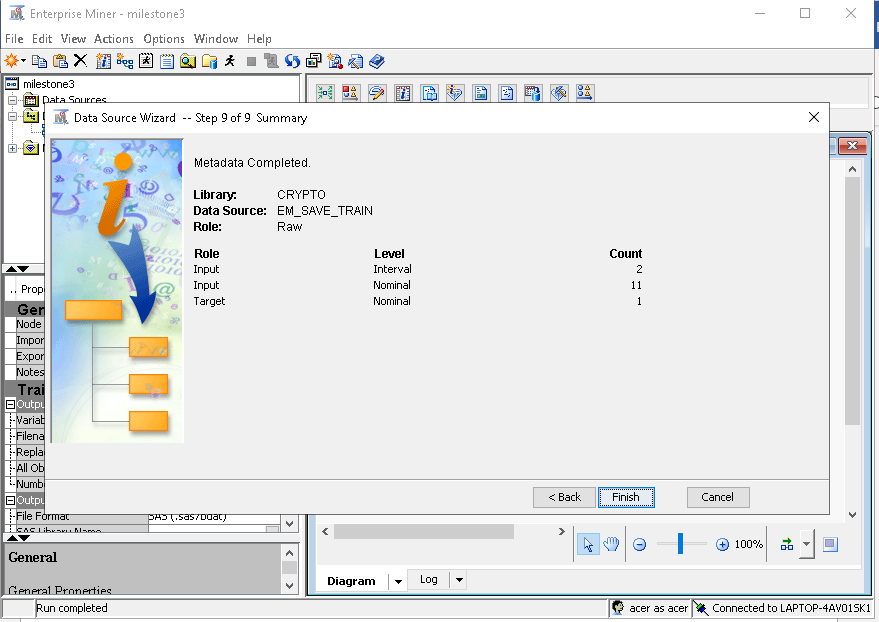
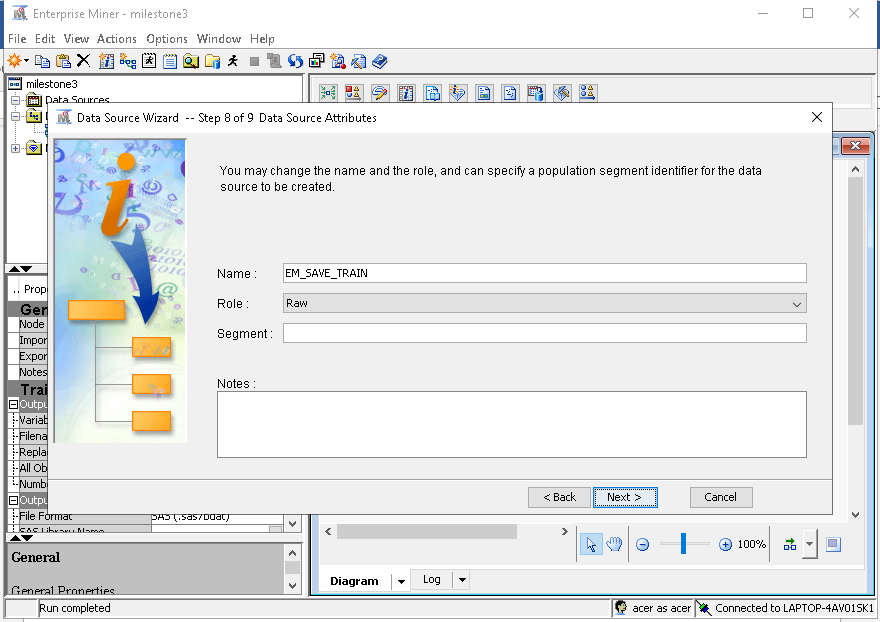
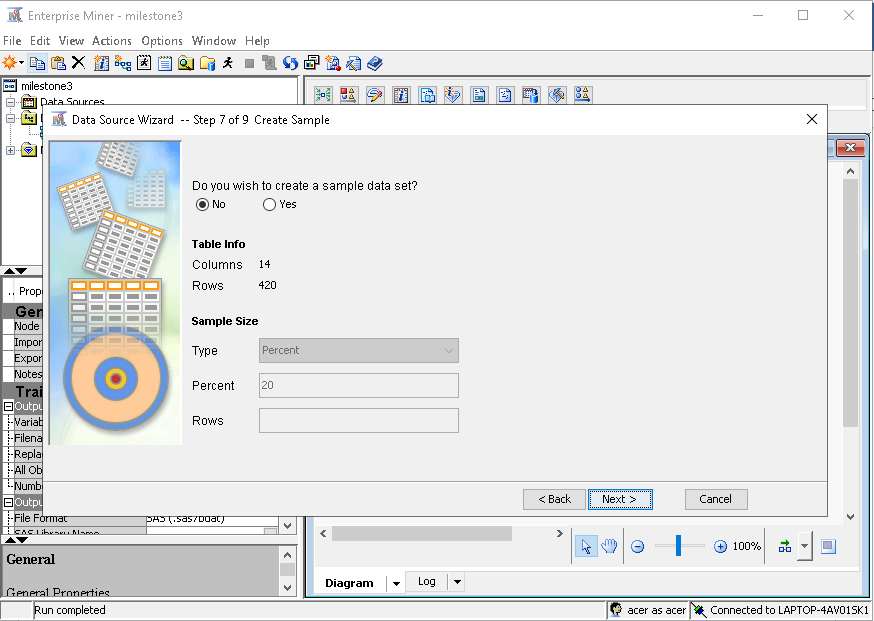
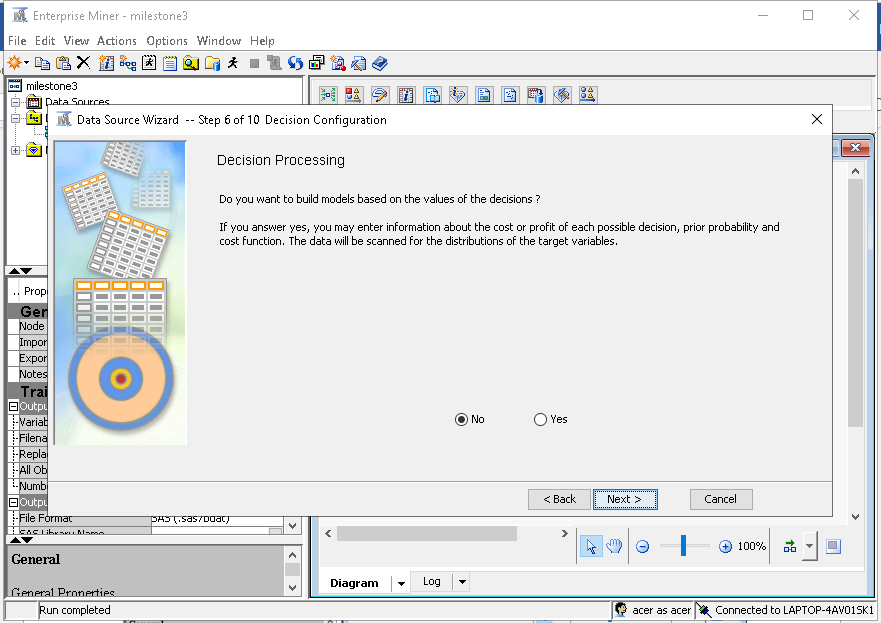
**Trade\_Time**: Time of Traded cryptocurrency

**Volume**: Volume of cryptocurrency(billion)

**52\_Week\_High**: Highest price in a year

**52\_Week\_Low**: Lowest price in a year

Metadata of the data, make sure that all the variables are in the **Role** of **Input.** We make the P**rice Direction** as the **Target Variable**

**Exploratory Data Analysis**

Plot histogram for **Market\_Cap**

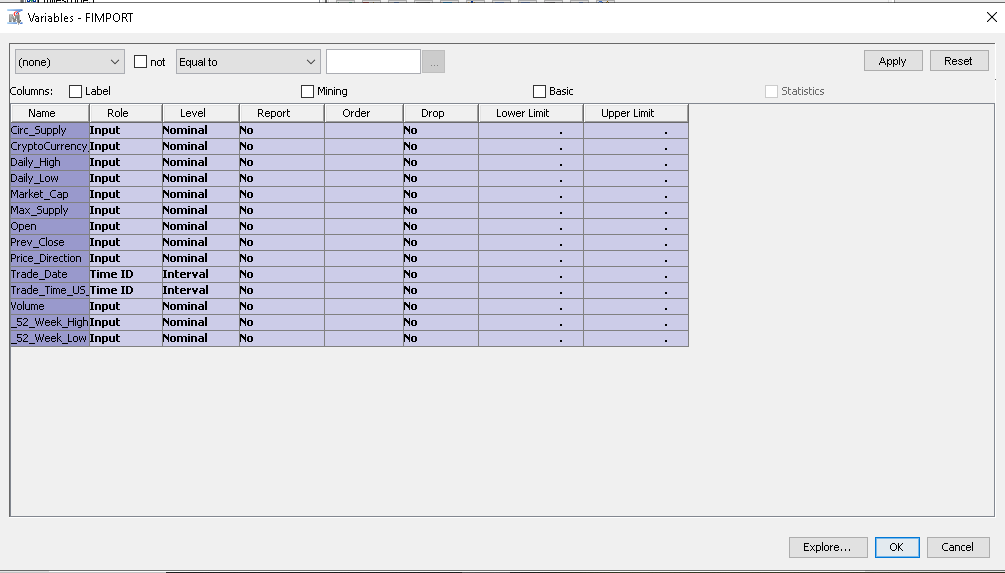
Plot histogram for **Volume**

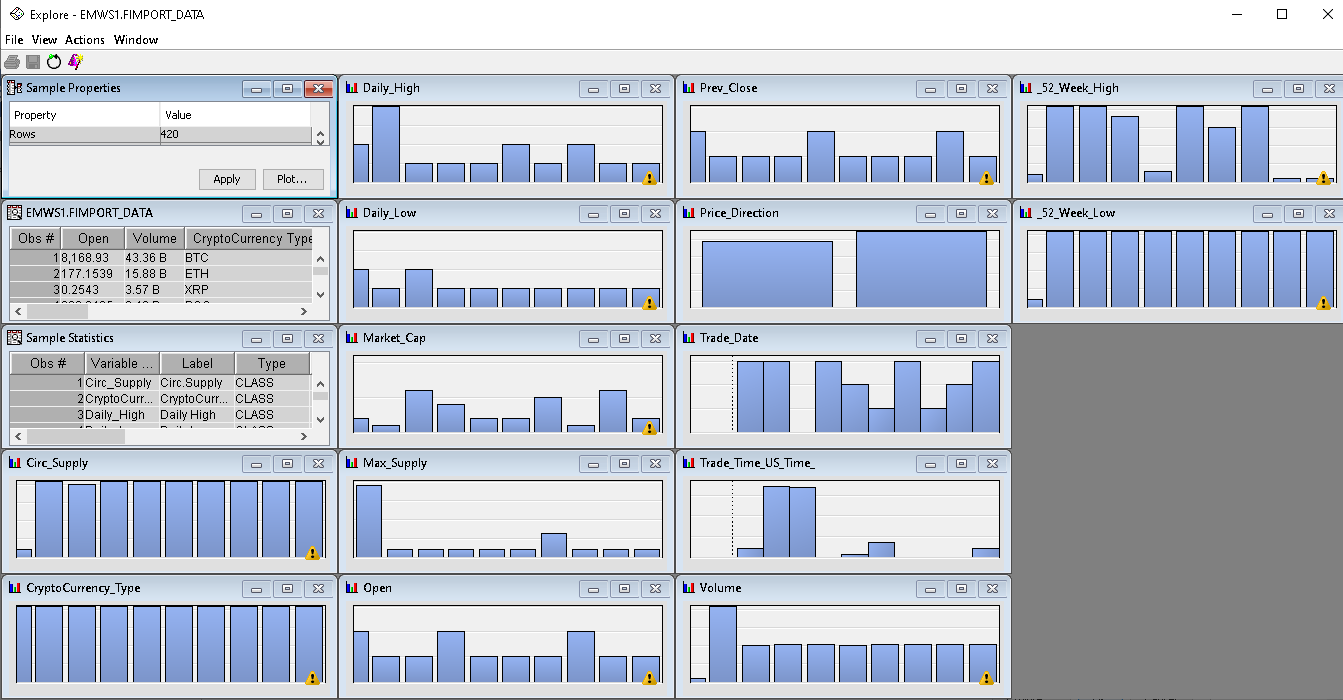
Plot histogram for **Circ.Supply**

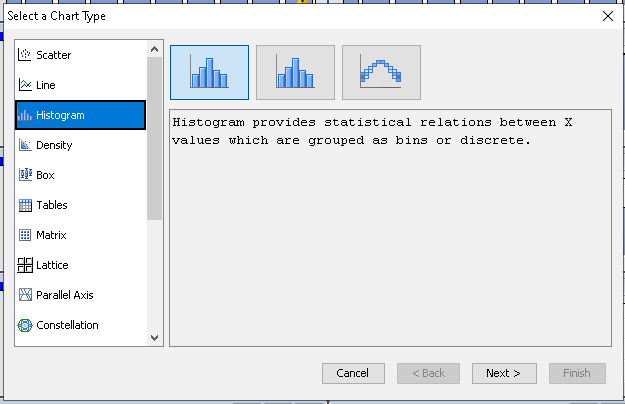
Plot histogram for **Max.Supply**

Show missing bins for histogram plots

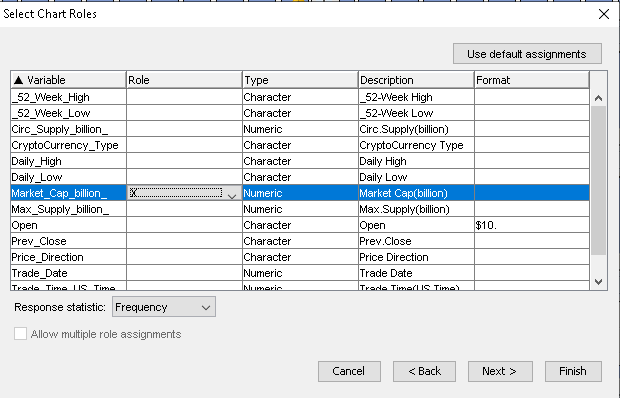
Plot pie chart for **Price\_Direction**

Next, we can do EDA on our datasets. Go to the Diagram, right click on the File Import node , click **Edit Variables**. You should get the screen similar as above. Select all the variables and click **Explore**

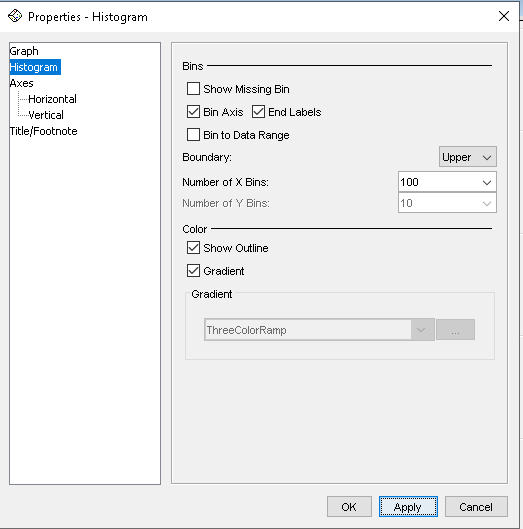
Go to **Sample Properties** Tab, click **Plo**t.

Go to **Histogram** to plot histogram graph.

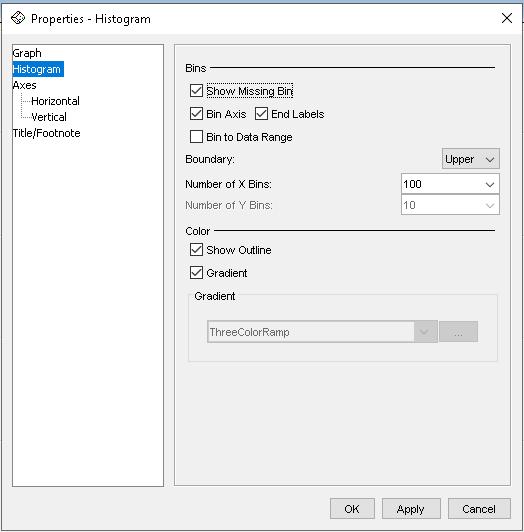
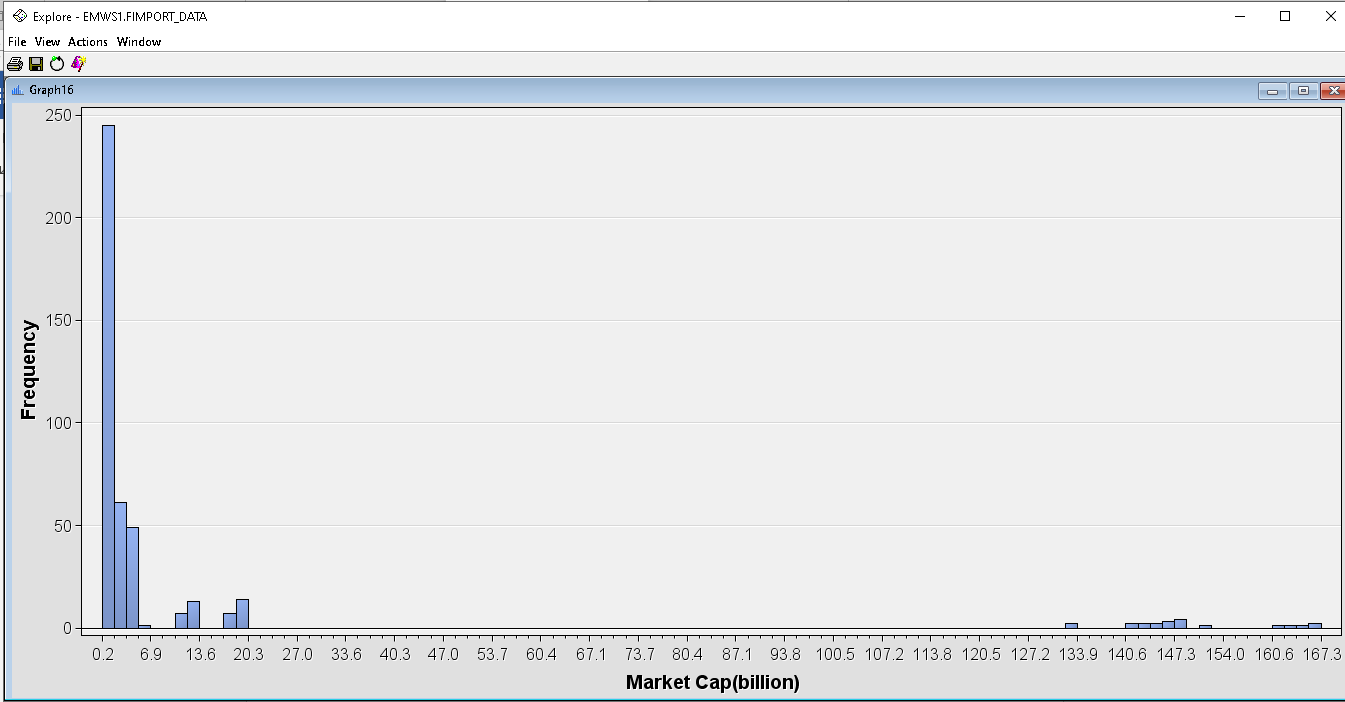
**Market Cap Histogram Plot**

Here, select **Market\_Cap\_billion** as the X input.

It can be seen from the plot that a huge portion of the cryptocurrency market cap are in the range of 0.2256 to 16.9350 billion.



Change the bin by increasing from default 10 to 100.

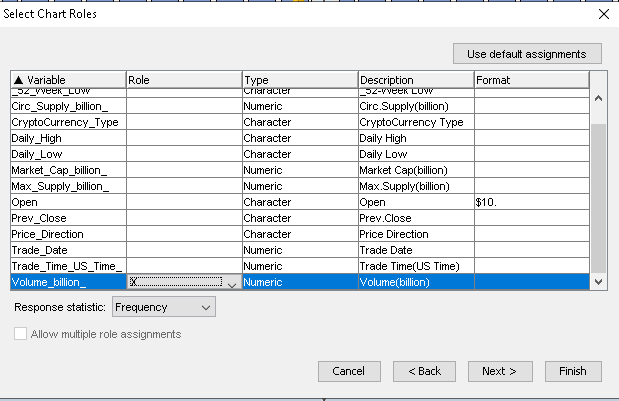


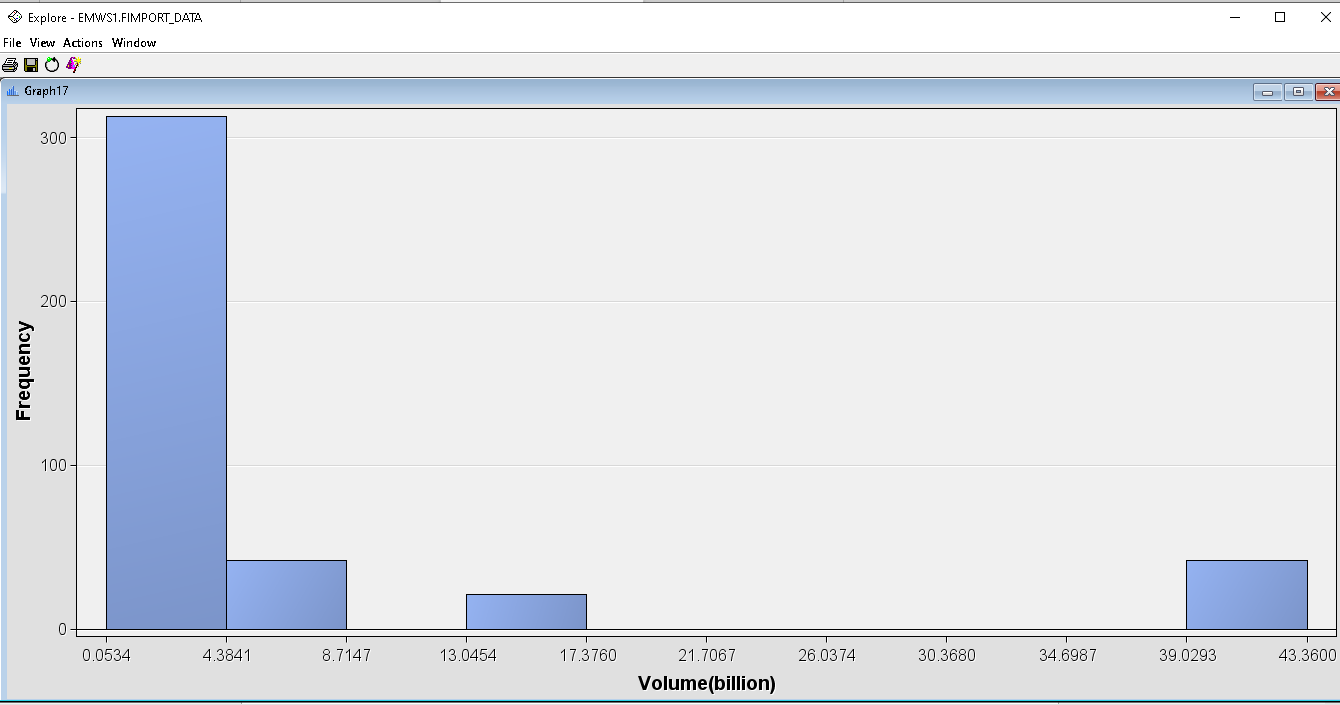
Check for missing values in variables by clicking **Show Missing Bin**.

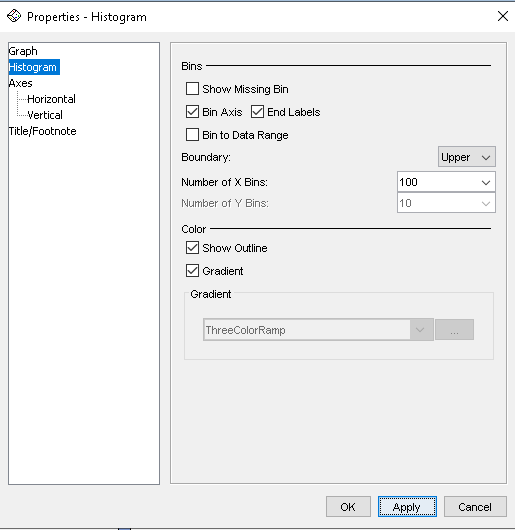
The missing bin can be seen at the far-left bottom of the histogram plot.

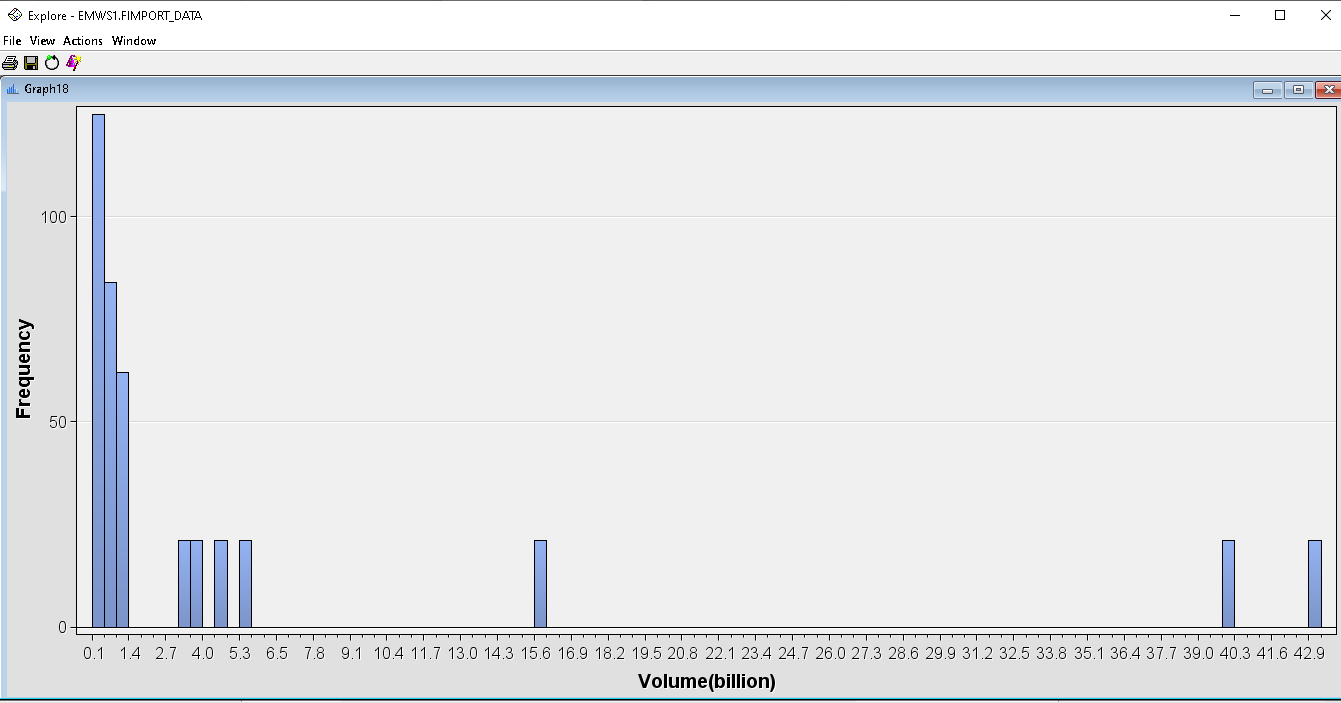
**Volume Histogram Plot**

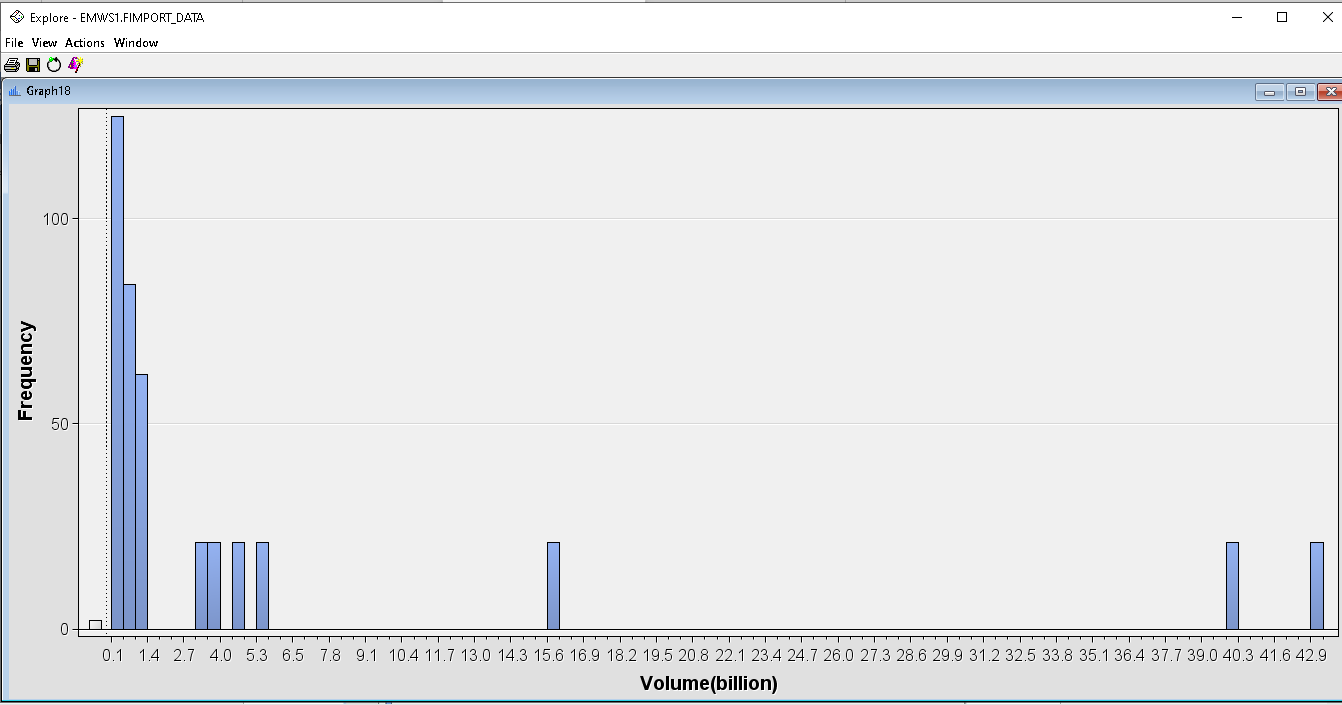
We can also plot histogram for Volume



From the plot, the volume of cryptocurrency in the market is skewed to the right due to a higher upper boundary. Showing that certain cryptocurrency have high trading volume in the markets.

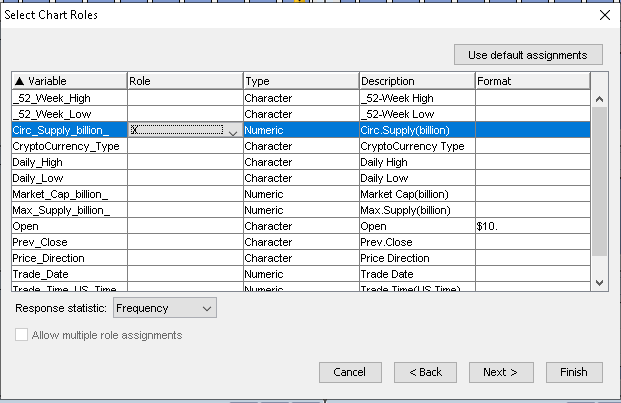


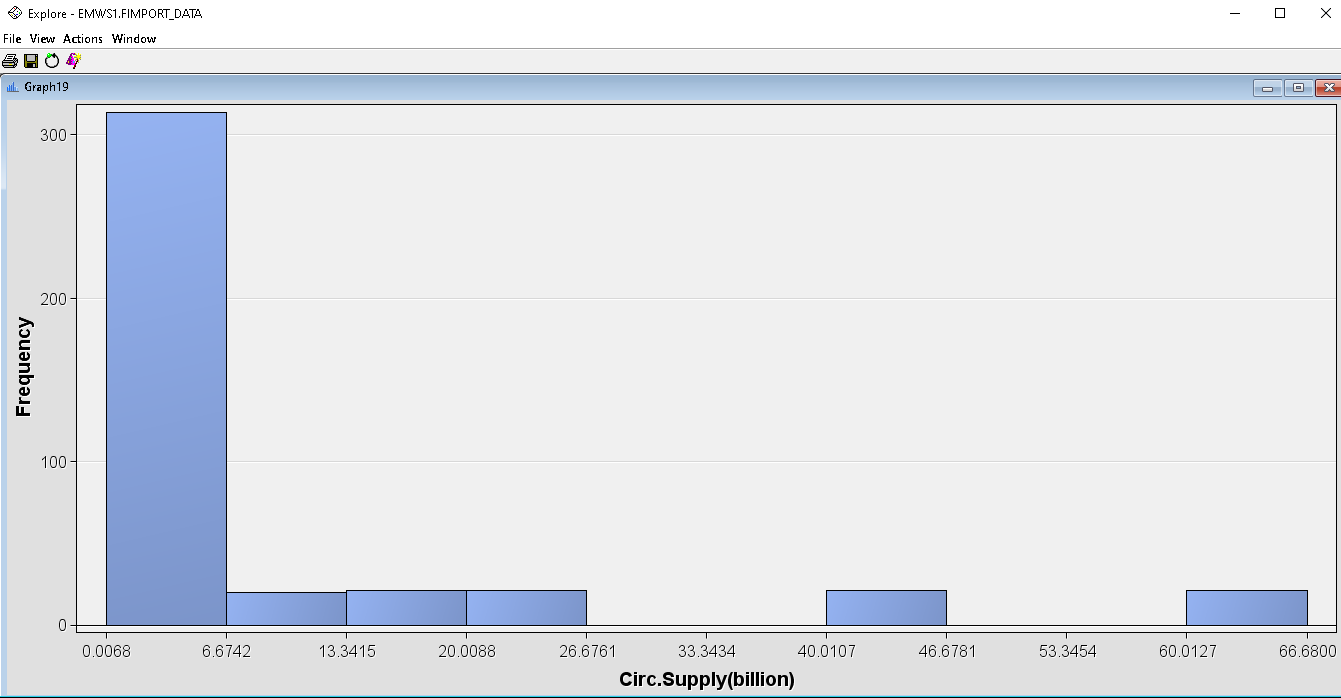
After changing the bin number to 100, a much more detailed histogram plot can be observed

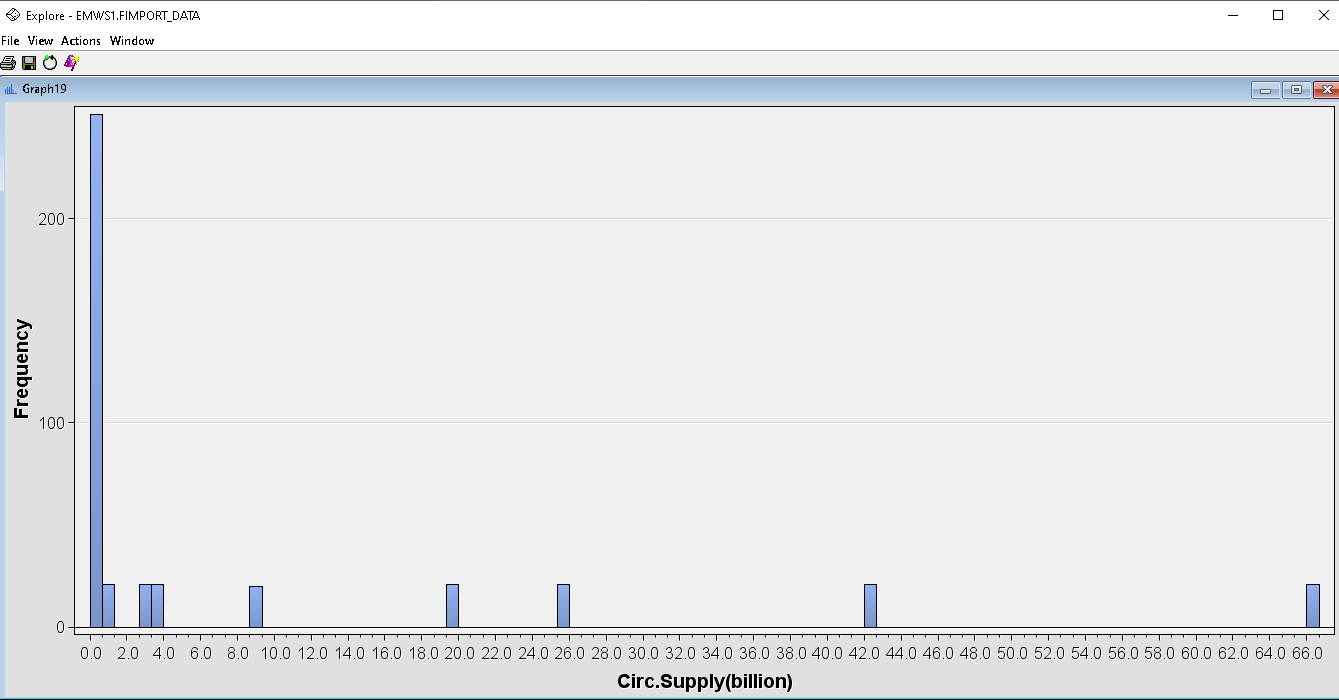
The missing bins can also be observed to be fairly insignificant to the overall plot.

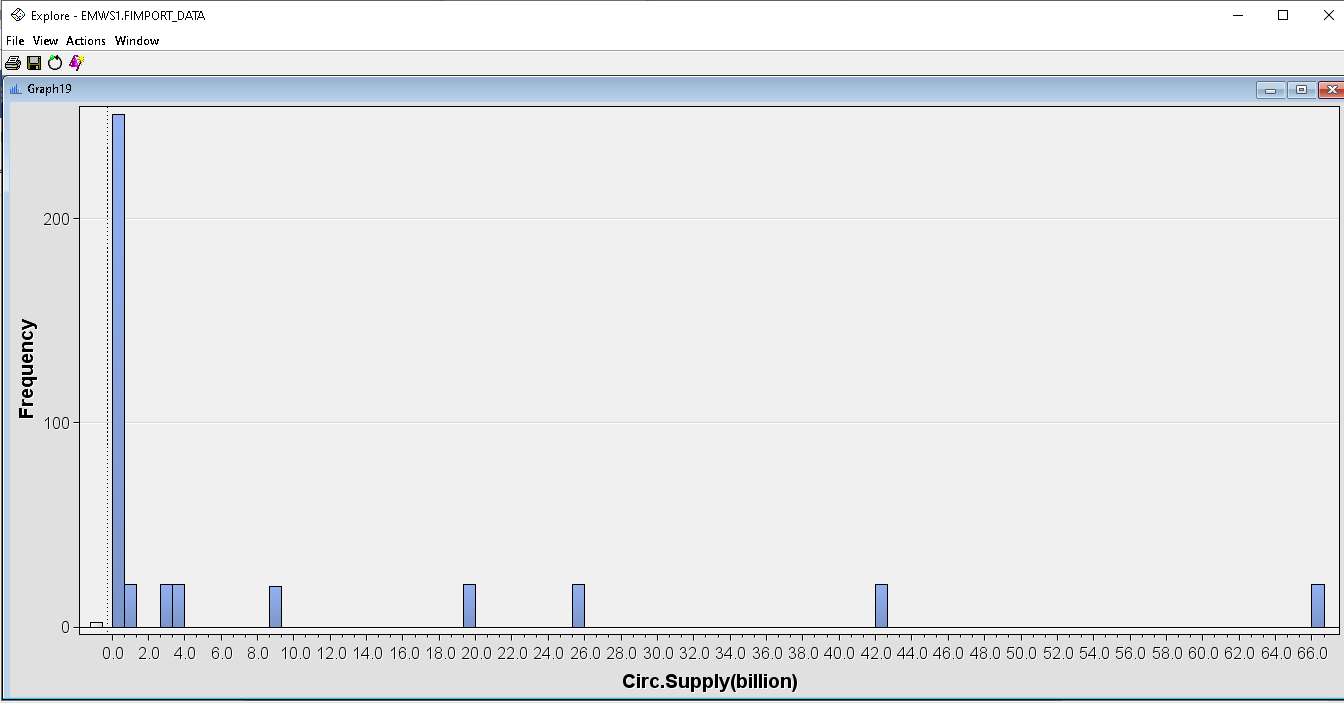
**Circ.Supply Histogram Plot**

Plotting histogram for Circ.Supply



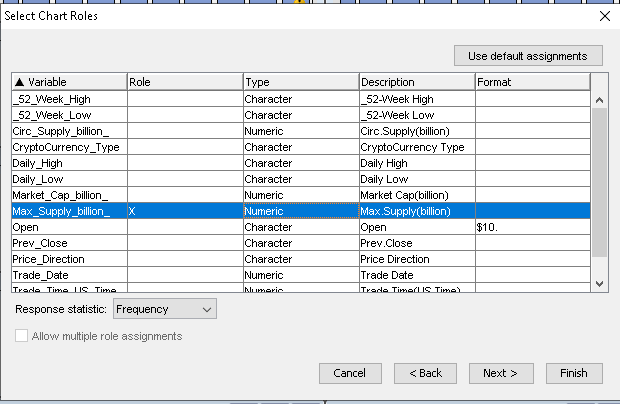
The histogram plot above show almost the same patterns as previous plots, most of the circulating cryptocurrency supply are in the range of 0.0068 to 6.6742 billion in the market. There are also huge amounts of market circulation for other cryptocurrencies ranging from 60.0127 to 66.68 billion.

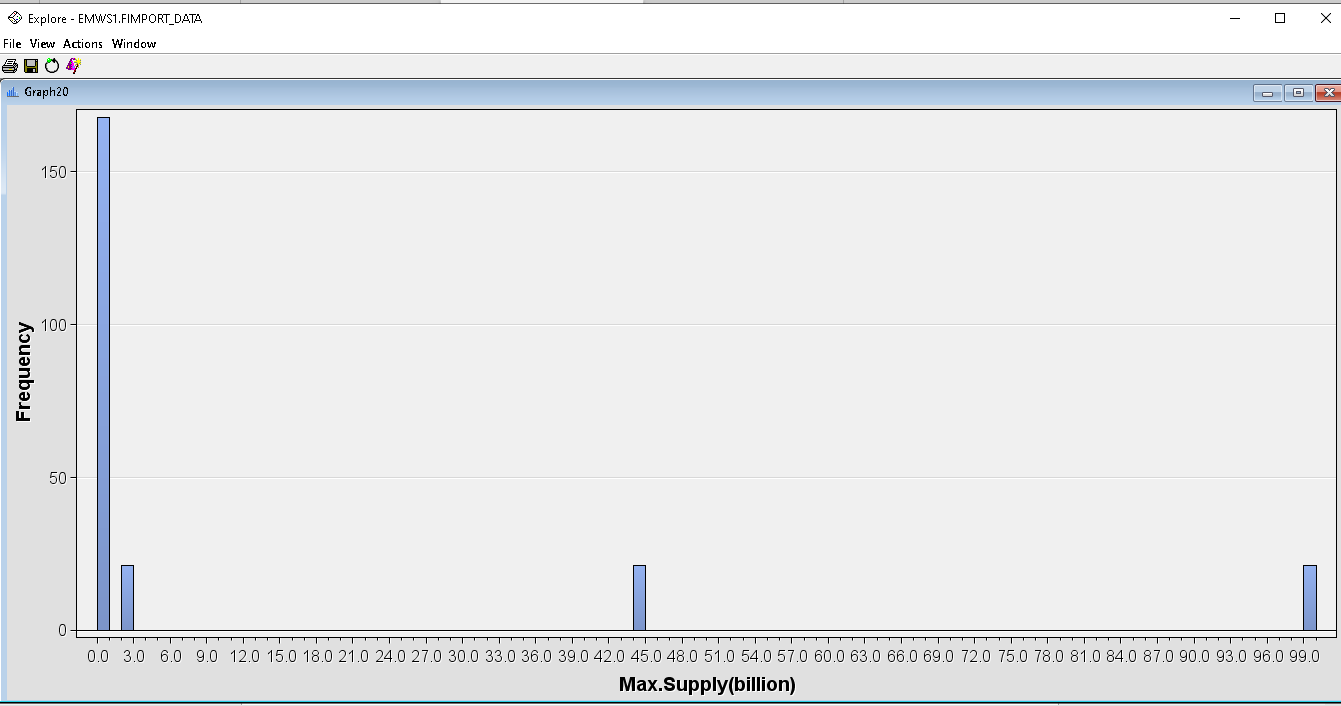
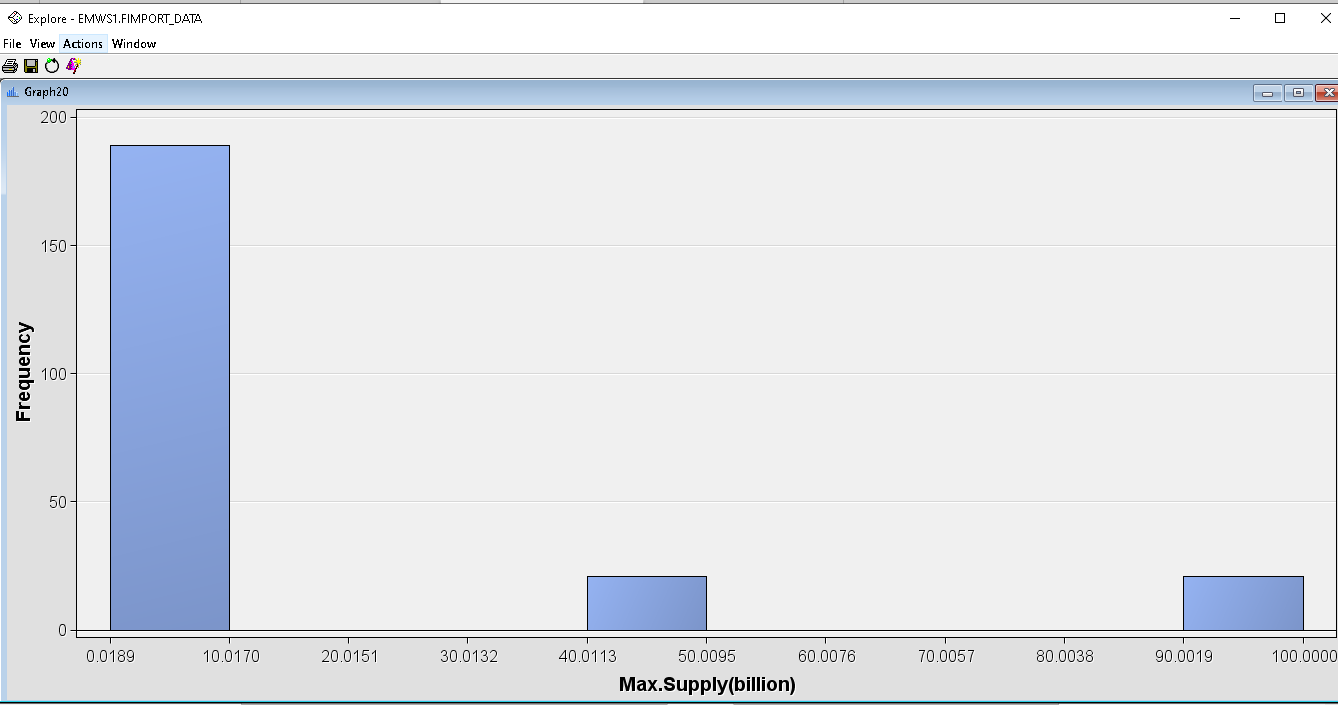
Show plot after 100 bins. However, after increasing the bin number, a more granular plot can be seen.

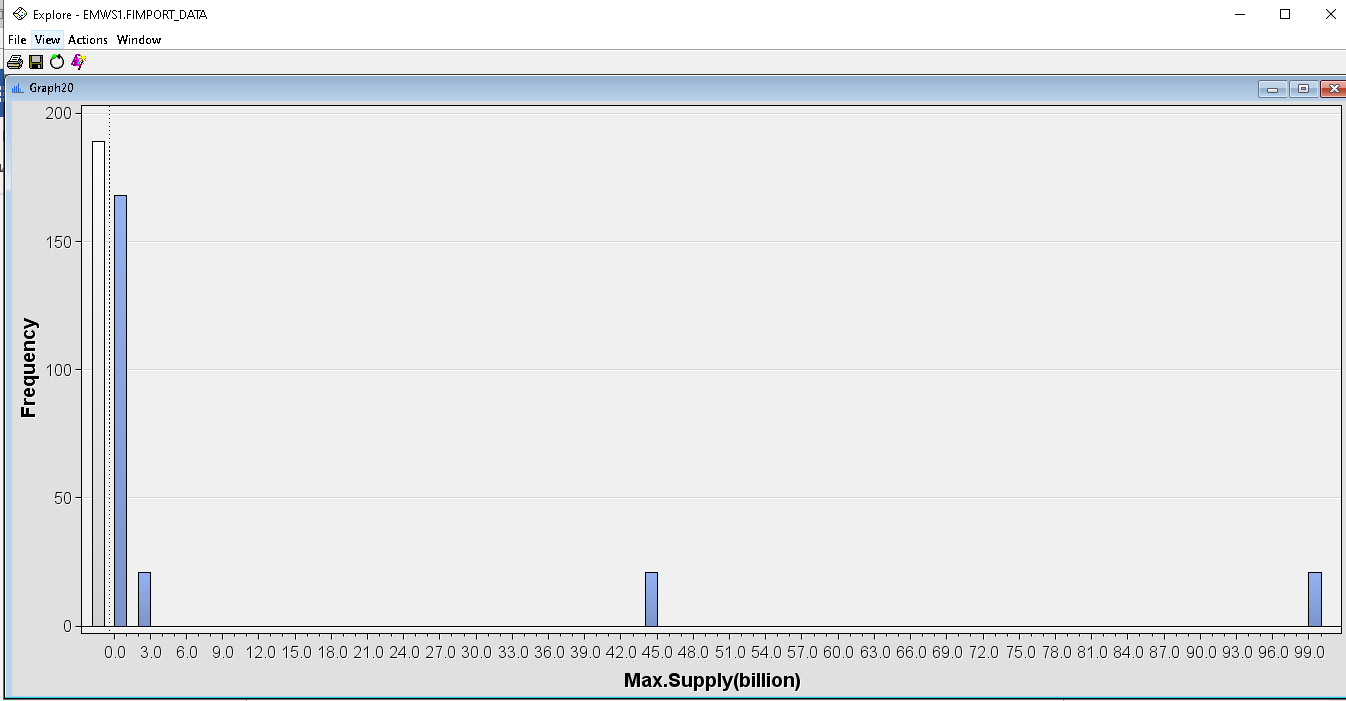
Show missing bins

**Max Supply Histogram Plot**

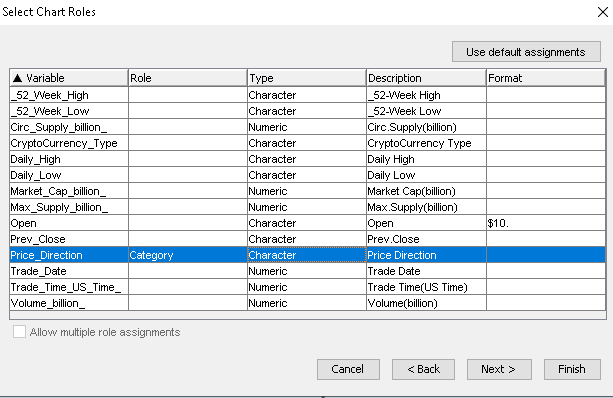
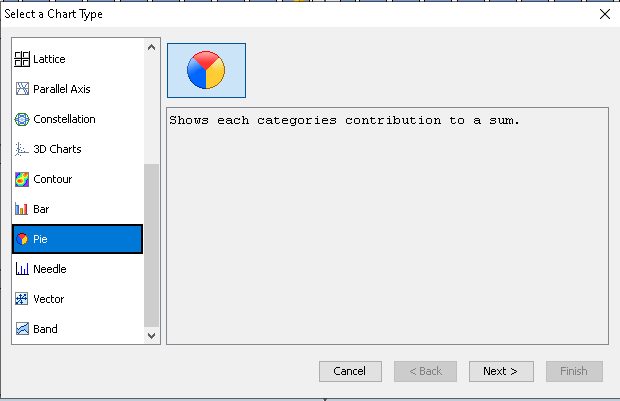
And finally plotting histogram for Max.Supply



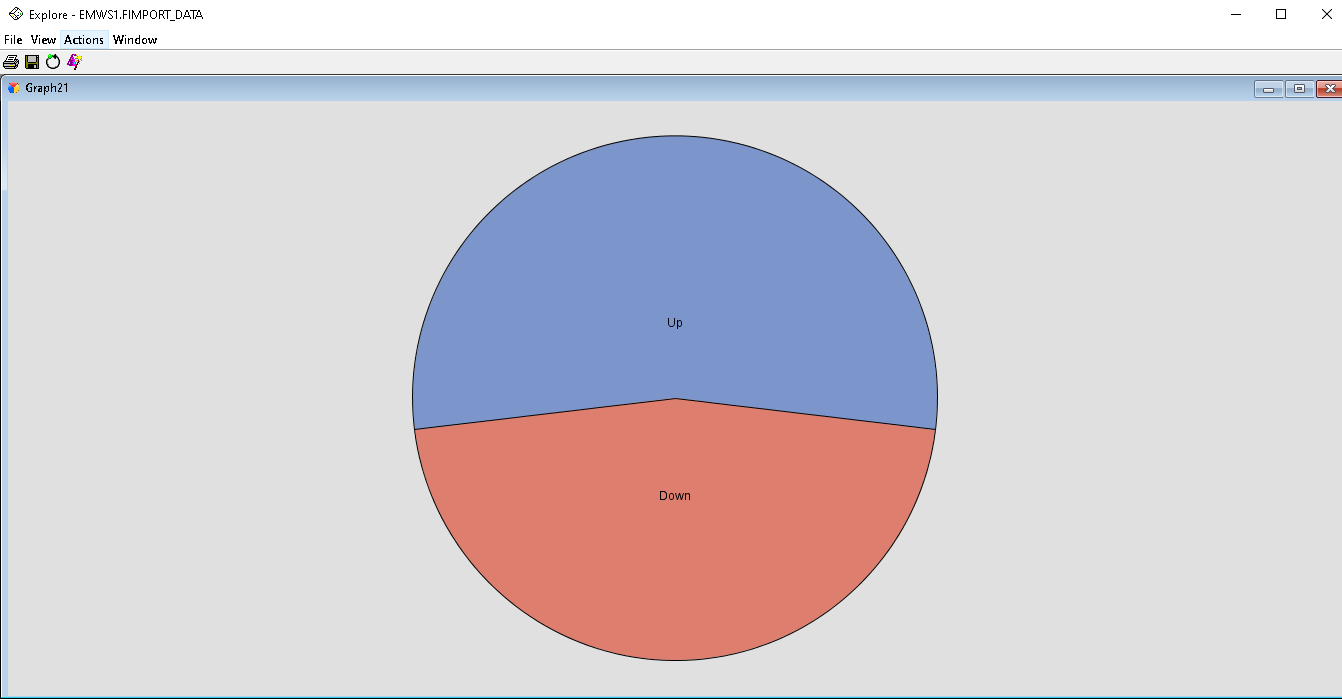
Increasing the bin number to 100 gives you a more detailed plot.

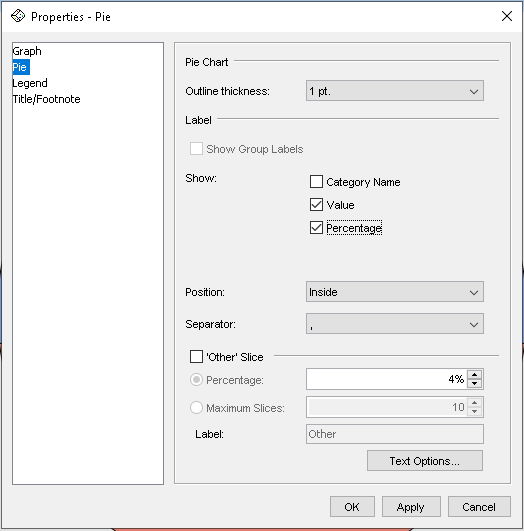
It can be observed that there is a significant amount of values lost from **Max.Supply** variable as the missing bin frequency is almost 190.

**Plotting Pie Chart for Price Direction**

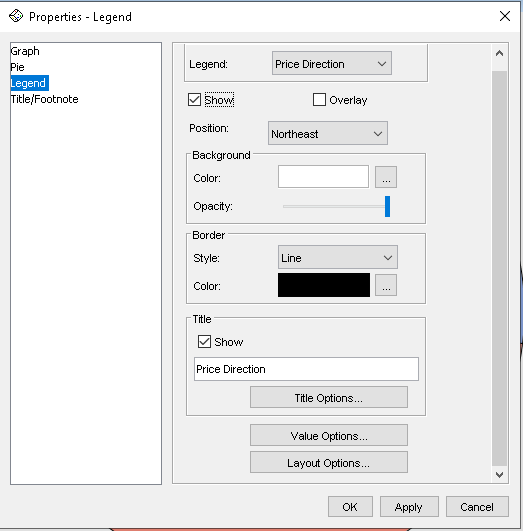


Select **Price Direction** as our **Category**

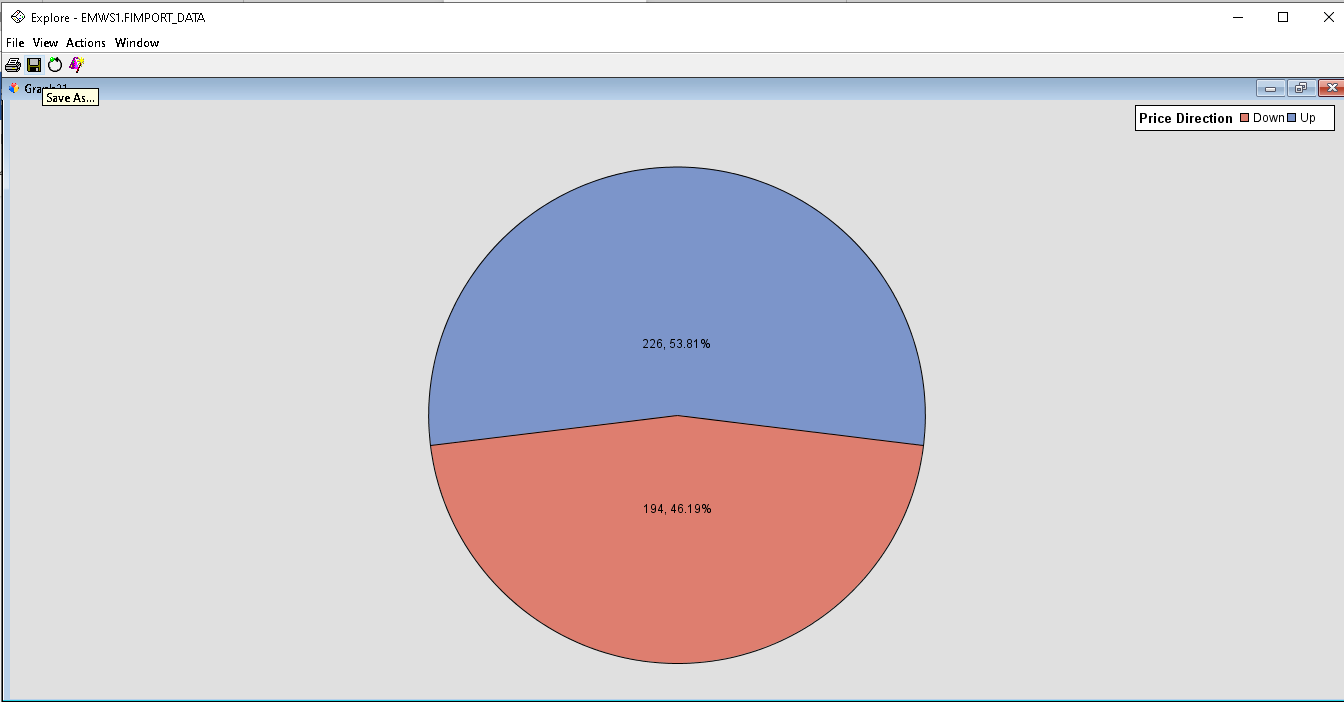
It can be seen that cryptocurrency data as almost half of them showing **Price Direction Up** the other half as Down. However this plot does not have labels, lets go head and add them.



Right click on the graph and select **Graph Properties.** Go to **Pie** and tick **Value** and **Percentage** in the box



Go to **Legend** and tick **Show**

As can seen on the labelled pie chart, more than half or 53.81% of cryptocurrencies show prices Up while 46.19% shows Down