Ethereum Analysis

By Ilaria Enache, Yasaman Pazhoolideh, Zeyneb M'Hamedi , Taylor Lucero

Ethereum

Ethereum is a decentralized, open-source blockchain with smart contract functionality. Ether (ETH) is the native cryptocurrency of the platform. Among cryptocurrencies, Ether is second only to Bitcoin in market capitalization. Ethereum was proposed in 2013 by programmer Vitalik Buterin.

Ethereum's own purported goal is to become a global platform for decentralized applications, allowing users from all over the world to write and run software that is resistant to censorship, downtime and fraud.

Goals

Business Goal

Data analysis can assist organizations and traders in understanding current market trends and making sound decisions. It provides a secure transaction framework and enables individual investors to determine appropriate market conditions and invest accordingly.

Technical Goal

Using this dataset, we aim to understand and then utilize the trends in the Ethereum Market to predicting future events.. By analyzing the data we will predict the likelihood of future trends in this market by predicting Market Capitalization on Ethereum Coins.

Description of Dataset

This dataset provides the history of daily prices of Ethereum. The data starts from 09-Aug-2015 and ends 07-Jul-2021.

Consisting of 2160 Instances (Rows) and 10 Attributes (Columns)

```
      SNo              Name              Symbol              Date          
 Min.   :   1.0   Length:2160        Length:2160        Length:2160       
 1st Qu.: 540.8   Class :character   Class :character   Class :character  
 Median :1080.5   Mode  :character   Mode  :character   Mode  :character  
 Mean   :1080.5                                                           
 3rd Qu.:1620.2                                                           
 Max.   :2160.0                                                           
      High              Low               Open             Close        
 Min.   :   0.483   Min.   :   0.421   Min.   :   0.432   Min.   :   0.435  
 1st Qu.:  14.265   1st Qu.:  13.191   1st Qu.:  13.758   1st Qu.:  13.819  
 Median : 205.125   Median : 193.303   Median : 198.425   Median : 198.644  
 Mean   : 398.259   Mean   : 365.593   Mean   : 382.880   Mean   : 383.911  
 3rd Qu.: 396.495   3rd Qu.: 375.147   3rd Qu.: 386.265   3rd Qu.: 386.435  
 Max.   :4362.351   Max.   :3785.849   Max.   :4174.636   Max.   :4168.701  
     Volume            Marketcap        
 Min.   :1.021e+05   Min.   :3.221e+07  
 1st Qu.:3.825e+07   1st Qu.:1.136e+09  
 Median :2.149e+09   Median :2.070e+10  
 Mean   :7.057e+09   Mean   :4.172e+10  
 3rd Qu.:9.629e+09   3rd Qu.:4.231e+10  
 Max.   :8.448e+10   Max.   :4.829e+11  
```

Dataset Attributes

SNo - Serial Number: the Unique identifier of each instance of the Ethereum dataset.	Datatype: Integer (Unique identifier for each instance)
Name - the Name of the coin	Datatype : String
Symbol - the Symbol or Acronym associated with the Ethereum Coin	Datatype : String
Date - Time and Date of when the particular instance was created	Datatype : String
High - The greatest value of the coin at that time	Datatype : Float

Dataset Attributes (continued)

Low - The lowest value of the coin at that time	Datatype : Float
Open - Opening price on the given day	Datatype : Float
Close - Closing price on the given day	Datatype : Float
Volume - Volume of transactions on the given day	Datatype : Float
Market Cap - Market capitalization in USD or the total value of all the coins that have been mined in USD	Datatype : Float

Preprocessing

Normalization of numerical data to ensure all attributes scales in the dataset are similar. This helps reduce the unequal importance larger values may tend to show compared to smaller ones.

	SNo	Name	Symbol	Date	High	Low	Open	Close	Volume	Marketcap
	<int>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	Ethereum	ETH	2015-08-08 23:59:59	2.798810	0.714725	2.793760	0.753325	674188	45486894

	High	Low	Open	Close	Volume	Marketcap
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	-0.6296305	-0.6439648	-0.6337728	-0.6374495	-0.6628663	-0.6031878

Other Steps:

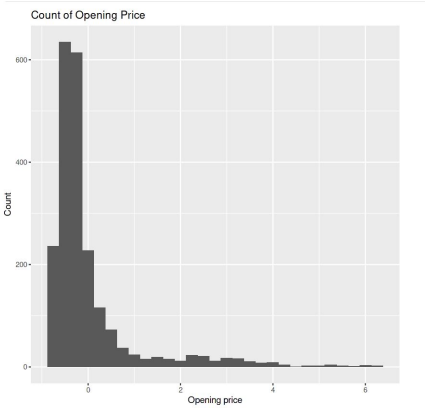
- Removal of redundant attributes like Name.
- Testing/ Removal of NA's
- Converting Date Attribute from String to Date Time Format.

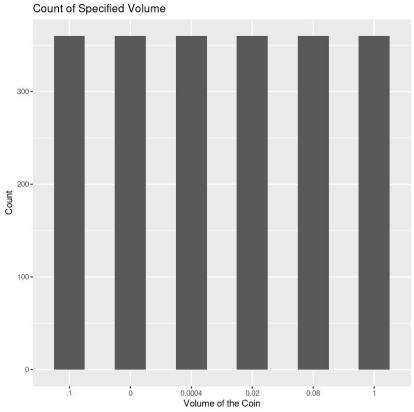
Univariate Distribution & Normality

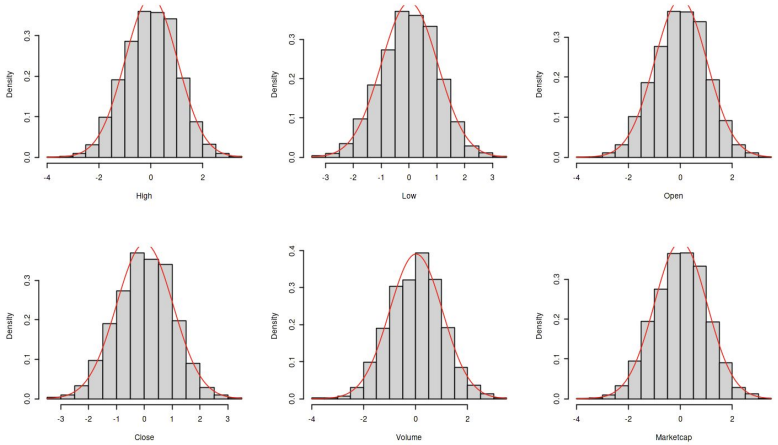
Initial visualization of numerical attributes showed a Non-Normal right skew.

Non-numerical attributes and Volume Count Distribution showed a uniform distribution.

Based on the **Shapiro-Wilks Tes**t Volume shows Normal Distribution, while other are shown to be non-normal.







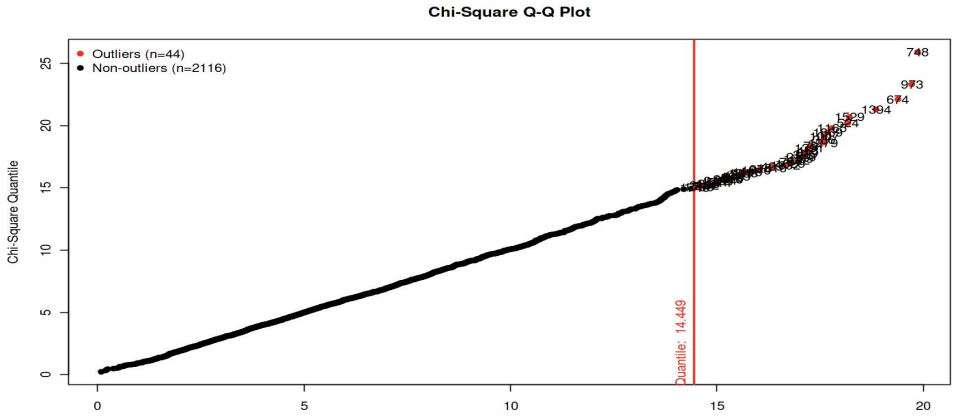
	Test	Variable	Statistic	p value	Normality
	<I<chr>>	<I<chr>>	<I<chr>>	<I<chr>>	<I<chr>>
1	Shapiro-Wilk	High	0.9983	0.0238	NO
2	Shapiro-Wilk	Low	0.9983	0.0263	NO
3	Shapiro-Wilk	Open	0.9983	0.0267	NO
4	Shapiro-Wilk	Close	0.9983	0.0261	NO
5	Shapiro-Wilk	Volume	0.9989	0.2031	YES
6	Shapiro-Wilk	Marketcap	0.9983	0.0292	NO

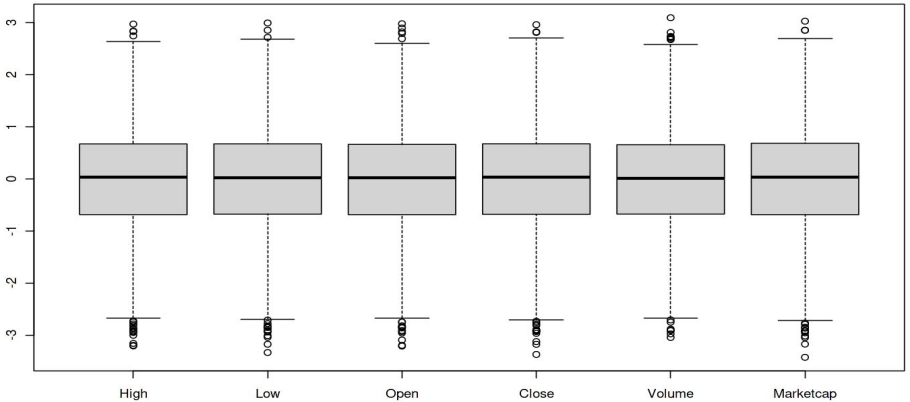
Multivariate Distribution & Normality

Multivariate Normality Distribution tested with Energy test.

The **Energy Test** proved Multivariate Normality.

Test	Statistic	p value	MVN
<chr>	<dbl>	<dbl>	<chr>
E-statistic	1.23816	0.492	YES





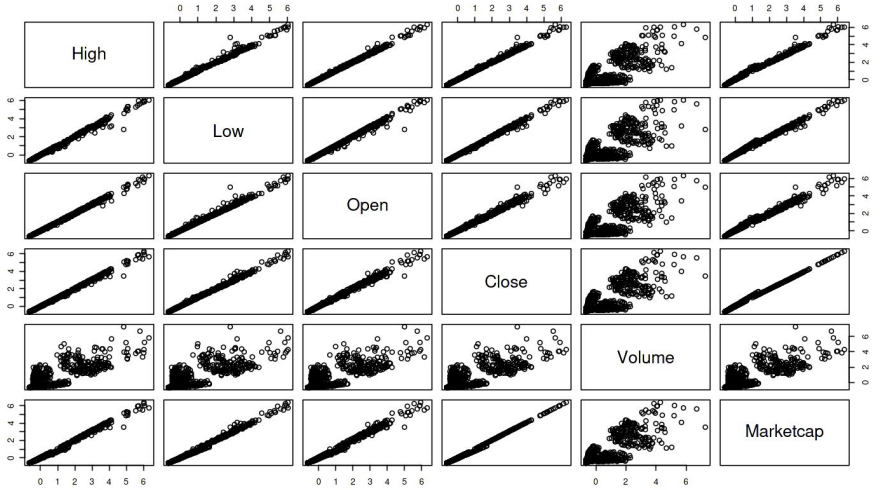
Hoeffding's Test of Independence

A Test based on the population measure of deviation from independence.

The D-static quantifies the dependency between variables.

The plot and table show that Volume is the only Independent variable

	High	Low	Open	Close	Volume	Marketcap
High	1.0000000	0.9281891	0.9466572	0.9522706	0.2497214	0.8947064
Low	0.9281891	1.0000000	0.9429575	0.9478031	0.2383958	0.9099302
Open	0.9466572	0.9429575	1.0000000	0.9194813	0.2431668	0.8823541
Close	0.9522706	0.9478031	0.9194813	1.0000000	0.2453015	0.9226737
Volume	0.2497214	0.2383958	0.2431668	0.2453015	1.0000000	0.2594075
Marketcap	0.8947064	0.9099302	0.8823541	0.9226737	0.2594075	1.0000000



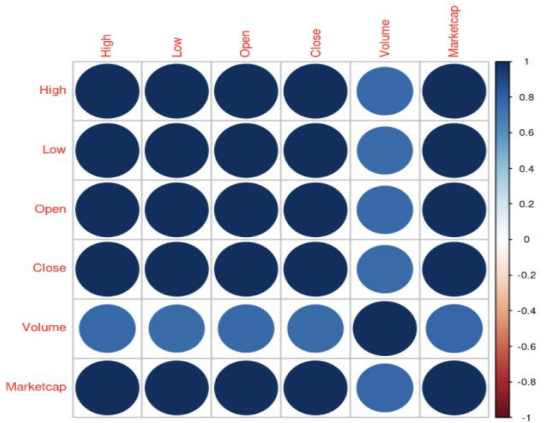
The Covariance/Correlation Matrix

The Normalized Covariance matrix shows the relationship between the variables.

The correlation between Volume and other attributes is still fairly high, however, it has less importance compared to other attributes with correlation values above 0.9

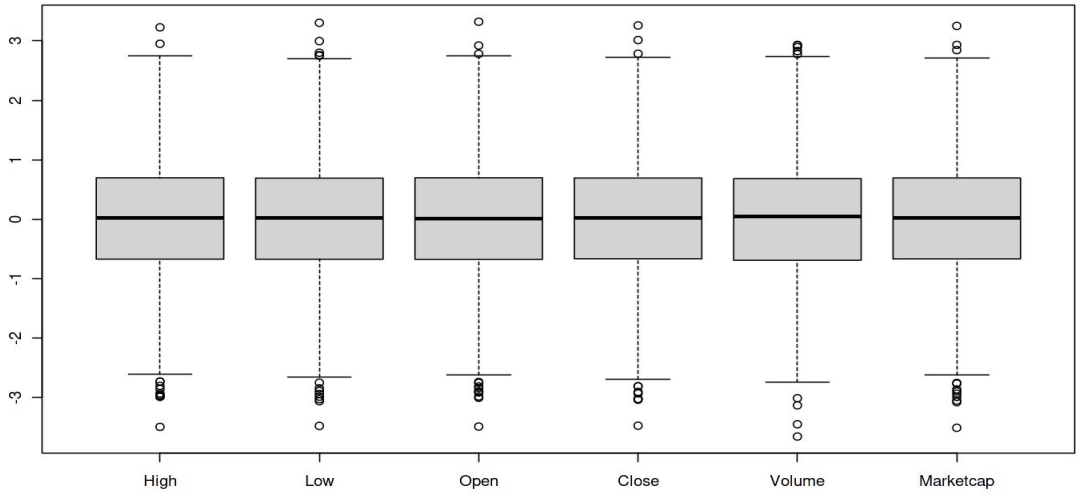
Normalized Cov/Cor matrix

	High	Low	Open	Close	Volume	Marketcap
High	1.0000000	0.9970698	0.9988632	0.9984573	0.7752340	0.9967876
Low	0.9970698	1.0000000	0.9970488	0.9985516	0.7611793	0.9968869
Open	0.9988632	0.9970488	1.0000000	0.9967218	0.7715921	0.9950781
Close	0.9984573	0.9985516	0.9967218	1.0000000	0.7691999	0.9983515
Volume	0.7752340	0.7611793	0.7715921	0.7691999	1.0000000	0.7854115
Marketcap	0.9967876	0.9968869	0.9950781	0.9983515	0.7854115	1.0000000

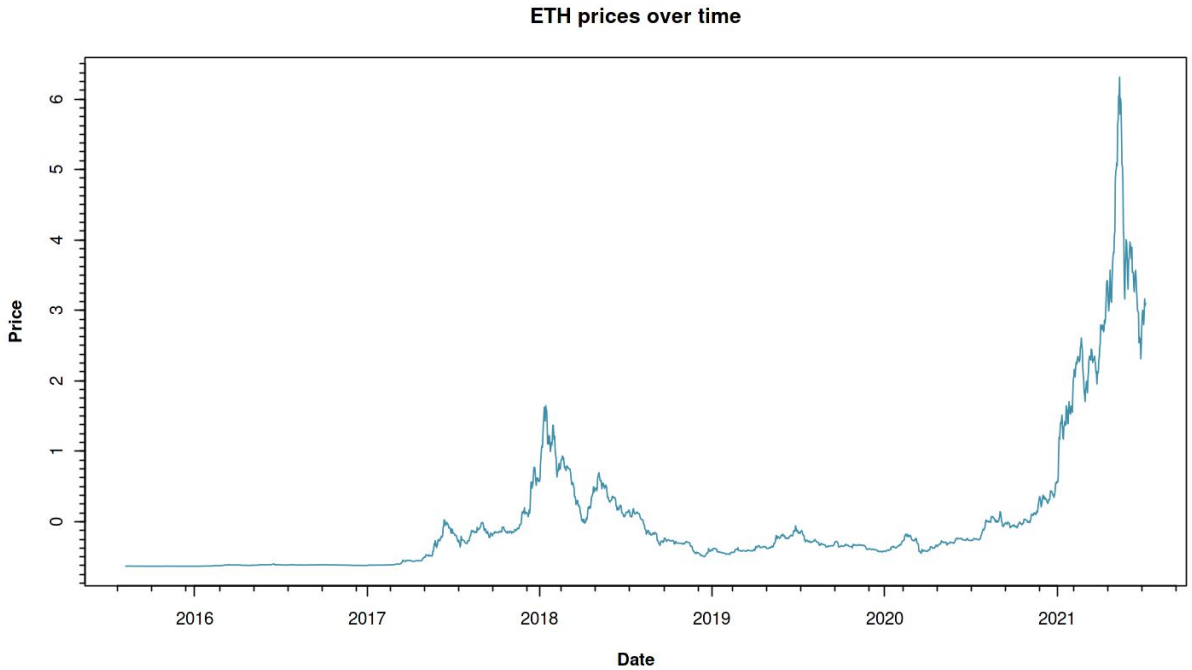


Outliers

Due to the nature of the Cryptocurrency and the extreme volatility in this domain, outlier will be kept. The previous boxplot shows the 44 outliers that lay outside the IQR.

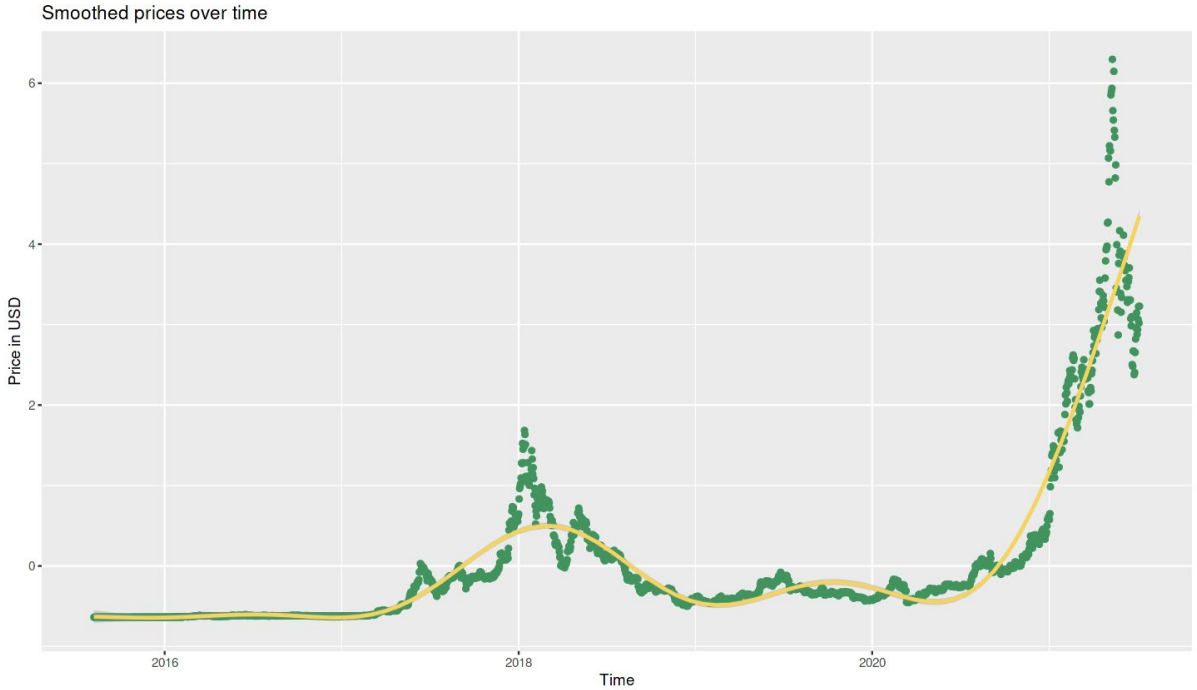


Ethereum High Prices Over Time



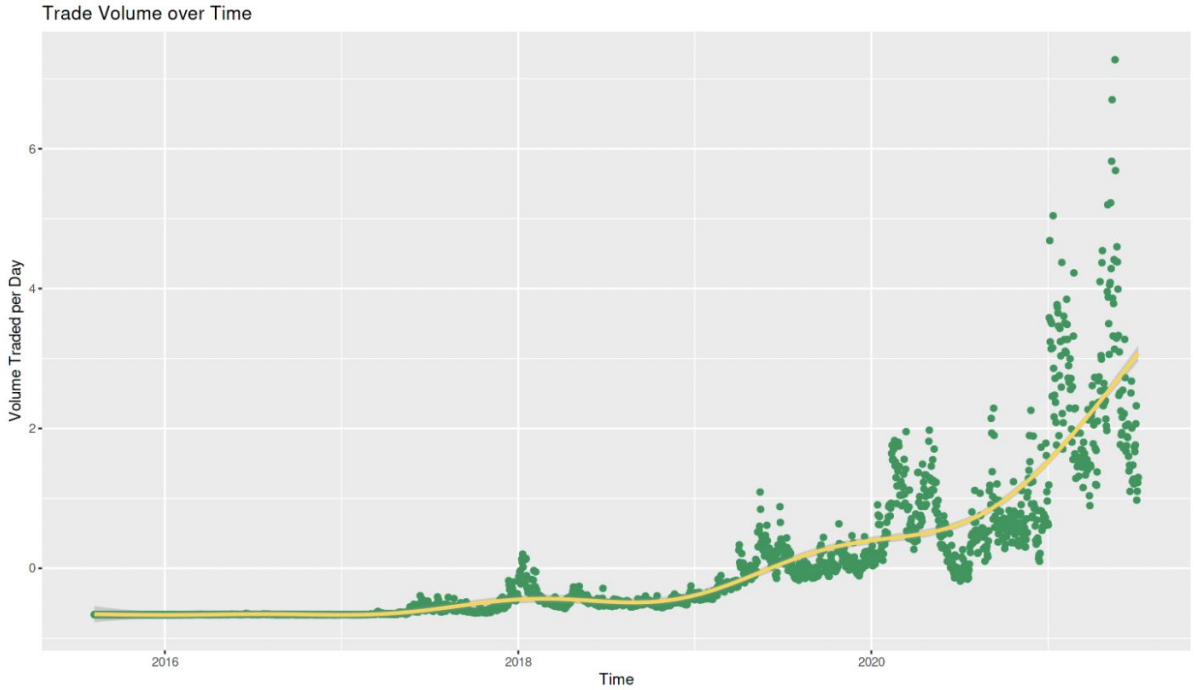
Closing ETH Prices Over Time

From visual inspection we can observe that ETH tends to have an explosive growth and a relatively high volatility.

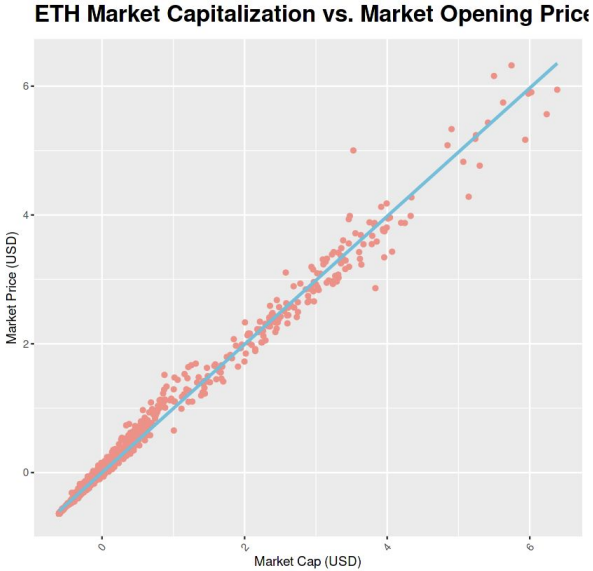


Traded Volume Over Time

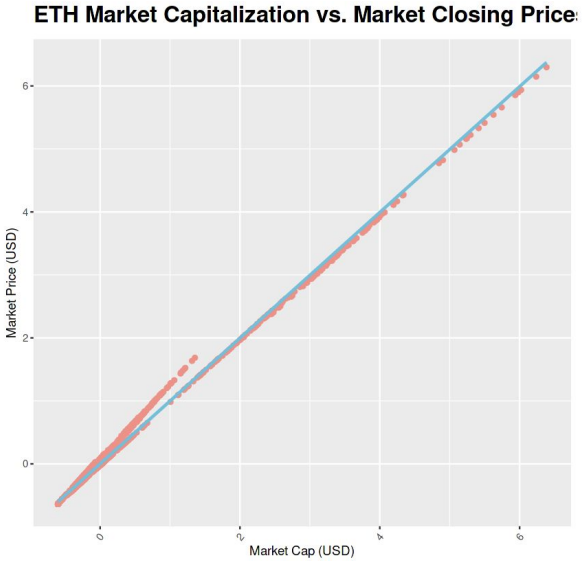
We see that ETH is being traded fairly frequently, and is trending upward in general.



Market Capitalization



We have observed that opening prices are more volatile than closing prices, they have larger variation. However, both of these charts show a very similar line and upward trend. In fact, despite the variation in the Opening price, the line of best fit, very closely resembles the line from the more stable closing cost.



Price vs Volume

Ethereum has a relatively nice volume to price correlation. We can also observe that our regression begins to fail at the very high end: it works better where our data points are clustered tightly.

Future Goal and Analysis

Analysis

Developing a Multivariate Multiple Regression Model/ Non-Linear Regression model

a. Maximum Likelihood Estimation (MLE)
b. Residual Analysis
c. Heteroscedasticity
d. Testing Accuracy

Goal

Produce a model that can accurately predict the Market Cap. on the Ethereum Coin.

Thank you for your time !

Ilaria Enache, Taylor Lucero, Yasaman Pazhoolideh and Zeyneb M’hamedi

Vocab and statistical information

Correlation Coefficient
Coefficient of determination
E-statistic
D-statistic
P-value
Null-hypothesis
Alternative hypothesis
Statistically significant
Central Tendancies
Normality
Trends
Z-score
Sample and Population

For your set of slides always include the reasoning behind why something was done. If you don’t know as or search the internet.