

# Ethereum: Time Series and Forecasting



## Final Project

Prince Sekyi

# **1. Introduction**

## **a) Background**

Cryptocurrency, a digital currency that can be used for goods and services. Currently the second largest cryptocurrency is Ethereum after bitcoin. Ethereum, is a blockchain that provides a verified and trusted environment for programs to run. Vitalik Buterin invented this cryptocurrency in 2013, it first price was release in July of 2015. was by programmer.

Reference: <https://en.wikipedia.org/wiki/Ethereum>

## **b) Problem Statement**

Generally, any form of investment has a certain risk associated to it: some have high risk and others low risk, cryptocurrency is very risk hence is Ethereum. Knowledge of the price of Ethereum gives directions on investing or the measure of investing in this cryptocurrency.

## **c) Goals**

This project seeks to

- observe the trends of the price of Ethereum over the years through Exploratory Data Analysis.
- predict the price of Ethereum using various Time Series and Machine Learning algorithms.

## 2. Dataset

Historical dataset of the price of Ethereum (ETH) from March 10<sup>th</sup> 2016 to August 9<sup>th</sup> 2021 was used for this research. This daily time series dataset was obtained from Kaggle, it has the following variables

- **Date:** Date of the ETH prices
- **Price:** Prices of ETH (dollars)
- **Open:** Opening price of ETH on the respective date (Dollars)
- **High:** Highest price of ETH on the respective date (Dollars)
- **Low:** Lowest price of ETH on the respective date (Dollars)
- **Vol.:** Volume of ETH on the respective date (Dollars).
- **Change %:** Percentage of Change in ETH prices on the respective date.

## 3. Data Cleaning and Data Wrangling

After the uploading the CSV file of the dataset to python Jupyter notebook, cleaning and preparation of the data must be made.

	Date	Price	Open	High	Low	Vol.	Change %
0	Aug 09, 2021	3,162.93	3,011.88	3,184.84	2,899.24	1.44M	5.01%
1	Aug 08, 2021	3,012.07	3,158.30	3,188.49	2,949.66	1.25M	-4.62%
2	Aug 07, 2021	3,158.00	2,889.58	3,169.74	2,867.58	64.84K	9.29%
3	Aug 06, 2021	2,889.43	2,827.23	2,946.62	2,726.04	1.06M	2.20%
4	Aug 05, 2021	2,827.21	2,725.28	2,842.95	2,533.51	1.65M	3.74%

Table 1

The variable, Date was converted to a datetime format and the commas in the various prices of Ethereum were remove, this makes it easy to work with.

The volume column was change to millions for easy graphing. Missing values were dropped, this reduced the shape of the dataset to (1964, 7).

	Date	Price	Open	High	Low	Vol.	Change %
0	2021-08-09	3162.93	3011.88	3184.84	2899.24	1.44000	5.01
1	2021-08-08	3012.07	3158.30	3188.49	2949.66	1.25000	-4.62
2	2021-08-07	3158.00	2889.58	3169.74	2867.58	0.06484	9.29
3	2021-08-06	2889.43	2827.23	2946.62	2726.04	1.06000	2.20
4	2021-08-05	2827.21	2725.28	2842.95	2533.51	1.65000	3.74

Table 2

New variables (Year, Days and Months) where created for exploratory data analysis (EDA). The final data frame for EDA is below.

	Date	Price	Open	High	Low	Vol.	Change %	Year	Month	Day
0	2021-08-09	3162.93	3011.88	3184.84	2899.24	1.44000	5.01	2021	Aug	Monday
1	2021-08-08	3012.07	3158.30	3188.49	2949.66	1.25000	-4.62	2021	Aug	Sunday
2	2021-08-07	3158.00	2889.58	3169.74	2867.58	0.06484	9.29	2021	Aug	Saturday
3	2021-08-06	2889.43	2827.23	2946.62	2726.04	1.06000	2.20	2021	Aug	Friday
4	2021-08-05	2827.21	2725.28	2842.95	2533.51	1.65000	3.74	2021	Aug	Thursday

Table 3

## 4. Exploratory Data Analysis (EDA).

Data is explored with respect to the various variables.

Plotting various variables, making visualization of the variables in the cleaned dataset.

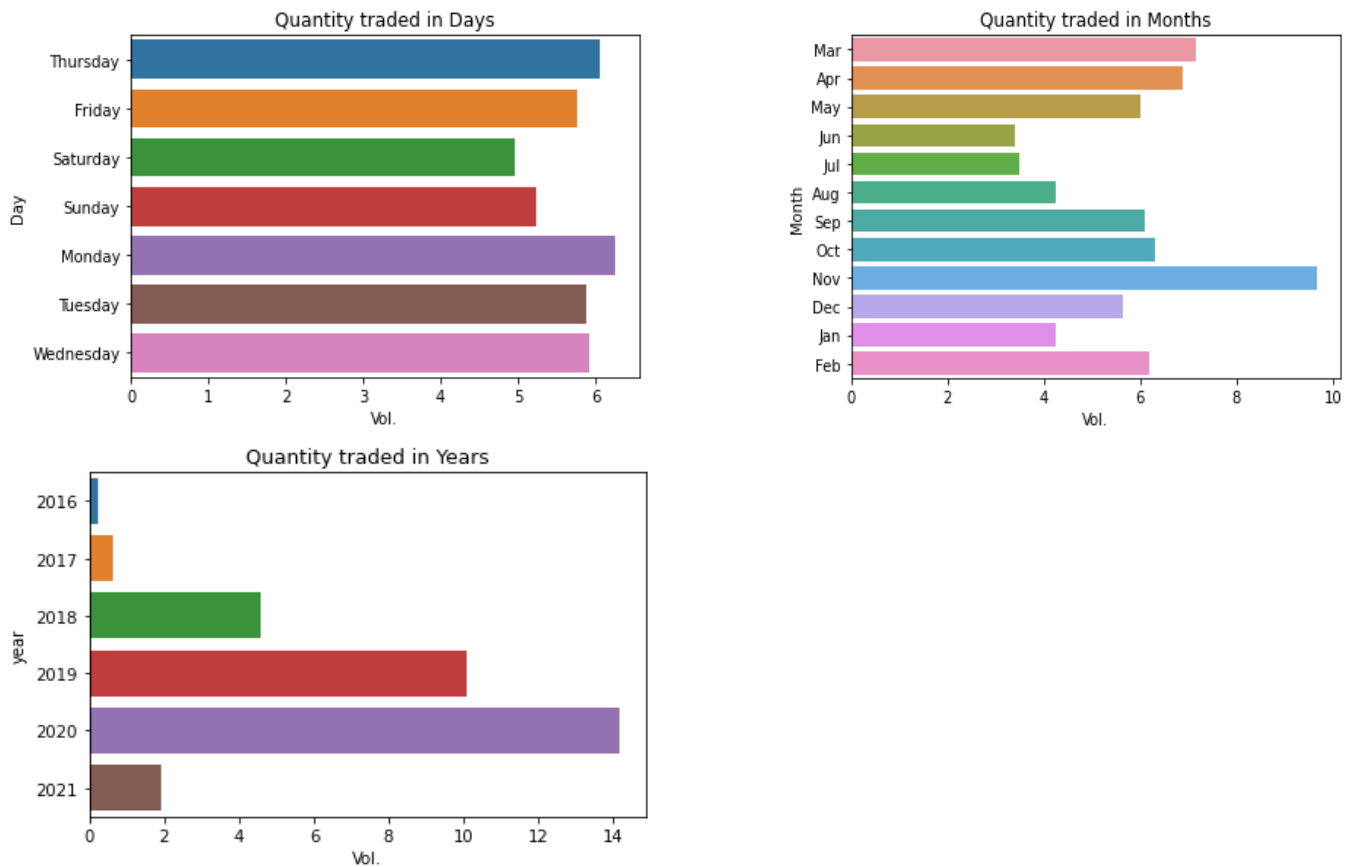


Figure 1

The daily time series dataset categorized into months, days and years. The volume of Ethereum traded were checked with respect to days, months and years.

It can be observed that more Ethereum are traded on Mondays, in the month November and in the year 2020. Less were traded in the year 2016, this must be due to it starting and less popularity.

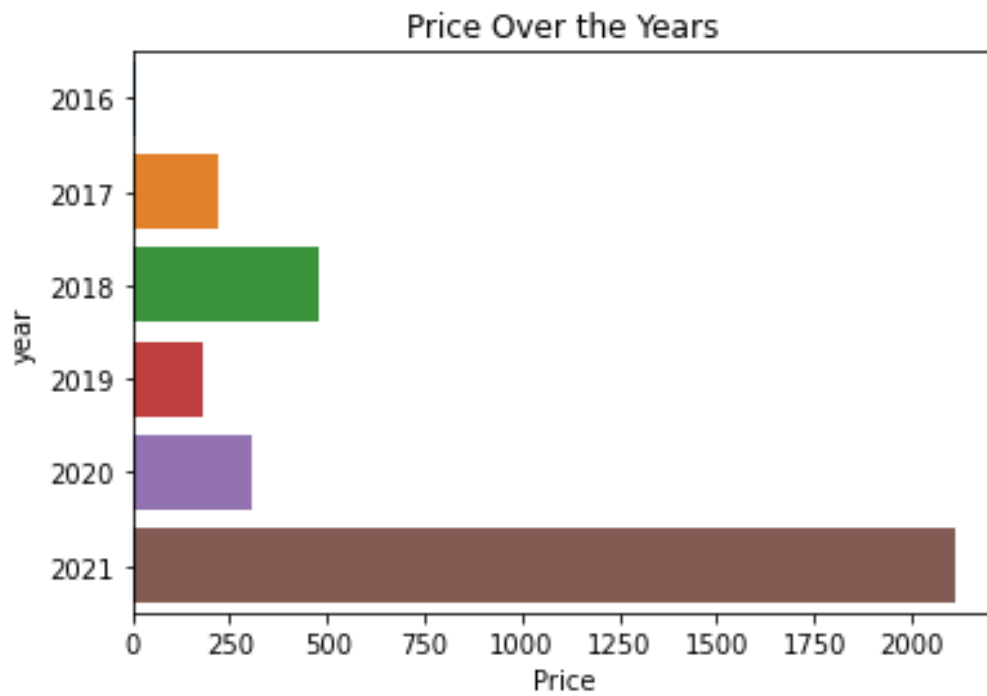


Figure 2

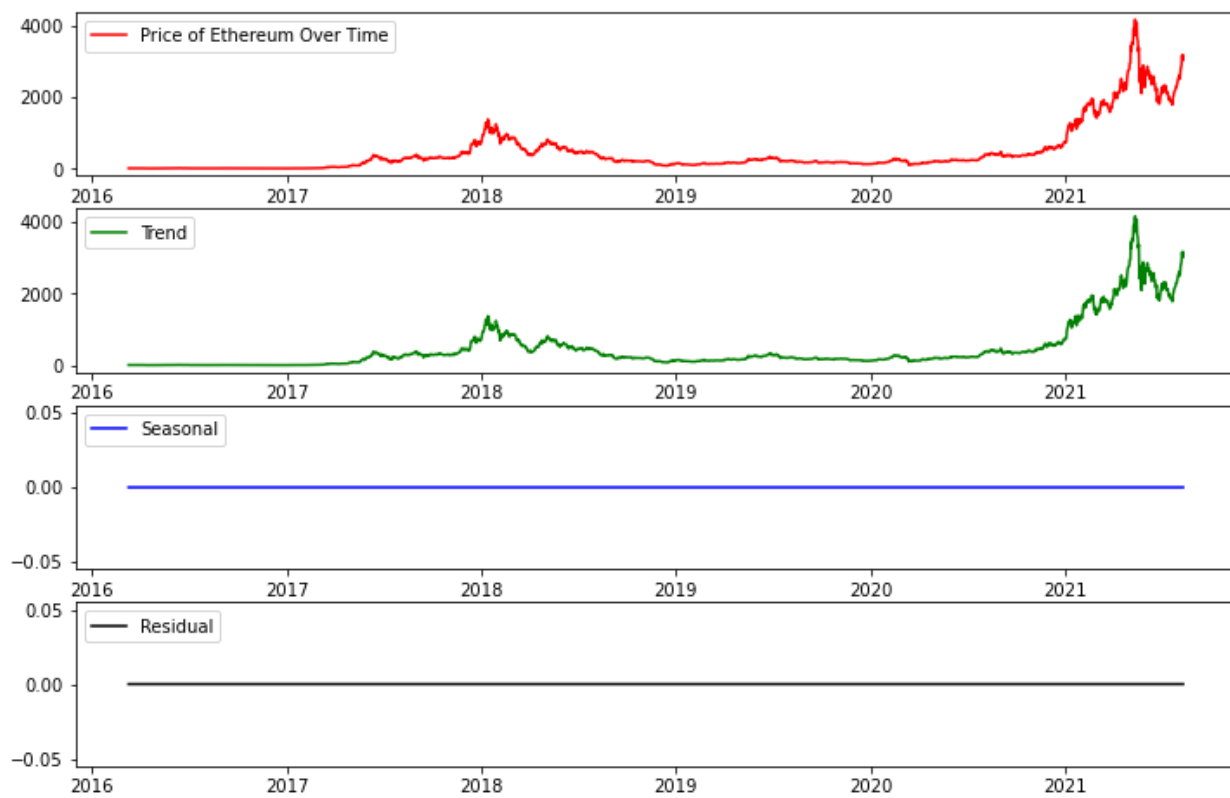


Figure 3

Figure 3 shows the decomposed time series dataset, this displays a plot of the price of Ethereum over time, the trend, seasonality and residual. From the plot of the prices over time, it can be observed that the price shoots up in the year 2021, then it declined but not to the lowest price since 2016.

There is an increasing trend in this dataset, even though it is showing a little decline. The data happens to be non-seasonal hence the graph of seasonality and residual.

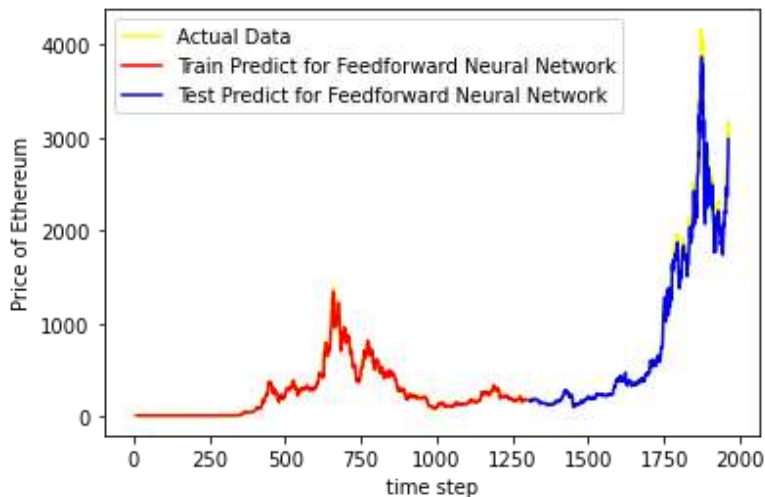
## 5. Data Preprocessing, Modelling and Forecasting

The dataset is now changed to a univariate time series data frame as shown below. This is used for modelling and forecasting

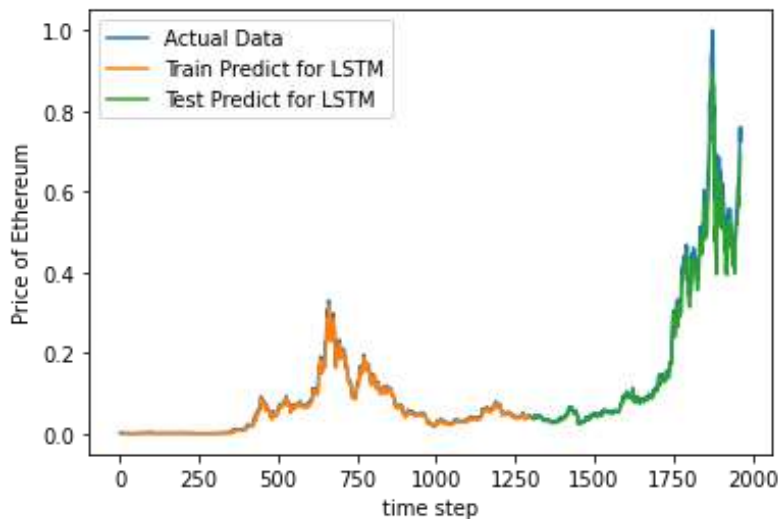
Price	
Date	
2016-03-10	11.75
2016-03-11	11.95
2016-03-12	12.92
2016-03-13	15.07
2016-03-14	12.50

## Model

Deep Learning algorithm, Feedforward Neural Network and Long short-term memory were used to fit the model the graphs below were obtained. Root Mean Square Error (RMSE) was used to evaluated the two models, below we have each model and their RMSE.



**Train Score: 22.45 RMSE, Test Score: 118.38 RMSE**



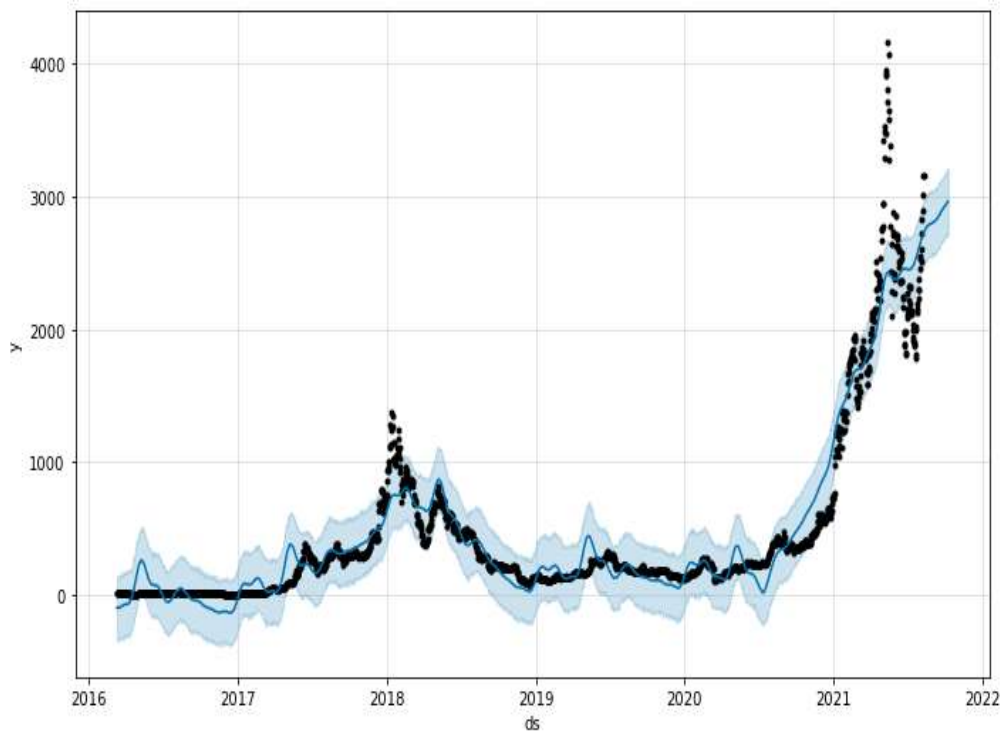
**Train Score: 0.01 RMSE, Test Score: 0.04 RMSE**

Mostly smaller RMSE is preferred since it measures the distance between the residual. For this project LSTM is preferred to Feedforward Neural Network.



## Forecasting

Facebook Prophet was used to do the forecast the price of Ethereum for the rest of the year 2021. The dates and the price column in on data were renamed to y and ds respectively, this is a requirement for Facebook prophet to work in python.



The table in the next page shows the statistical description of the forecasted values.

Statistics	Values
count	114.00
mean	3141.10
std	208.49
min	2826.66
25%	2959.90
50%	3112.47
75%	3305.63
max	3600.77

There were 114 forecasted values. From the forecasting the mean price of Ethereum till the end of the year is \$3,141.10 with a standard deviation of 208.49. The minimum price is estimated to be \$2,826.66 with a maximum of \$3600.77.

## **6. Future Recommendation**

Below are recommendations for further study

- I will consider using more and most current data of the historical price of Ethereum for future study, since the more the data, since this can give a lower RMSE hence a good predictions,
- I will consider other Machine Learning time series algorithm, Darts, Theta and more
- I will also use other ways to evaluate Deep learning models like ROC, Confusion Matrix and more