JSC370 Final Project - Cryptocurrency Analysis

Jin Shang

Introduction

Cryptocurrency is a digital or virtual currency that uses cryptography for security and operates independently of a central bank. Cryptocurrencies are decentralized, meaning that they are not controlled by a single authority or institution, but instead operate on a distributed ledger system called blockchain.

The first and most well-known cryptocurrency is Bitcoin, which was created in 2009 by an anonymous person or group of people using the pseudonym Satoshi Nakamoto. Since then, many other cryptocurrencies have been developed, including Ethereum, Litecoin, and Ripple.

Cryptocurrencies are created through a process called mining, which involves using computational power to solve complex mathematical problems and verify transactions on the blockchain. Transactions on the blockchain are recorded in blocks, which are then added to the blockchain in a chronological and immutable manner.

Now, cryptocurrency has become an increasingly important investment in recent years, with many investors and traders looking to capitalize on its price volatility and potential for high returns. However cryptocurrencies, like any other investment, carry a number of risks that investors should be aware of before investing. For example, cryptocurrencies are known for their high volatility, with prices often fluctuating significantly over short periods of time and the value of cryptocurrencies depends on their adoption and use by individuals and businesses. If adoption is low or slows down, the value of a cryptocurrency may decrease or become worthless. Thus while they can provide opportunities for high returns, investors should carefully consider timing of buying or selling cryptocurrencies.

In this project, we will explore the historical price data of some cryptocurrencies and try to gain insights into its price trends and factors that may affect its price. We will use various statistical and data visualization techniques to analyze the data and draw conclusions about the cryptocurrency's past and potential future performance.

Methods

Data Access

These cryptocurrency data is collected by Coinranking API. The data contains the current information on 50 different cryptocurrencies, include current price, the change of price in last 24 hours, etc. And for analysis the price trend of cryptocurrency, I have also extracted historyical price data for several cryptocurrencies. The following are explainations of variable in our database extracted from coinranking API.

Table 1: Variable

Name	Definition
ids	Unique id
symbols	Abbreviation or symbols
names	Names of coin
tiers	Seperate coin into 3 tiers based on supply, volume and market cap

Name	Definition
volume24h	The volume over the last 24 hours of coins without the filters
changes	Percentage of change over the the last 24 hours
marketCap	Market capitalization. Price times circulating supply
max	Maximum amount the coin's supply can ever be
current	Current amount the coin's supply
circulatings	Number of coins that are circulating in the public market
market_number	The number of markets that contain the this coin
exchange_number	The number of exchanges that trade this coin
V/MC	volume/marketCap
timestamp	Unix timestamp
price	history price corresponding to timestamp

Data Wrangling

By Coinranking API, we get a long char for all information, so I use regex to extract all these information. I extracted 50 cryptocurrencies with basic information and 6 representative cryptocurrencies with historical prices from 2018-04-27 to 2023-04-27 (5 years) based on current price and last 24 hours change.

Checking for NAs

Below I will show the number of these six cryptocurrencies that have been priced at NA over the last five years:

Cryptocurrency	Proportion
Bitcoin	0.003831418
Tether USD	0.003831418
Shiba Inu	0.6086481
Monero	0.003831418
Cronos	0.1559934
EnergySwap	0.7504105

By observing the historical data of these cryptocurrencies, I found the missing dates for Bitcoin, Tether USD and Monero are same, so this is probably because the api does not have access to price data for those days. And for the remaining three, since they were released late, the pre-release prices are all NA. That's why the missing proportion of them are very high. And for historical price, I converted Unix time stamp to date. In addition, I replace median NA price in my dataset with the mean of the previous and next non-NA values and add log_price for analysis.

Result

Summary Visual

Figure 1

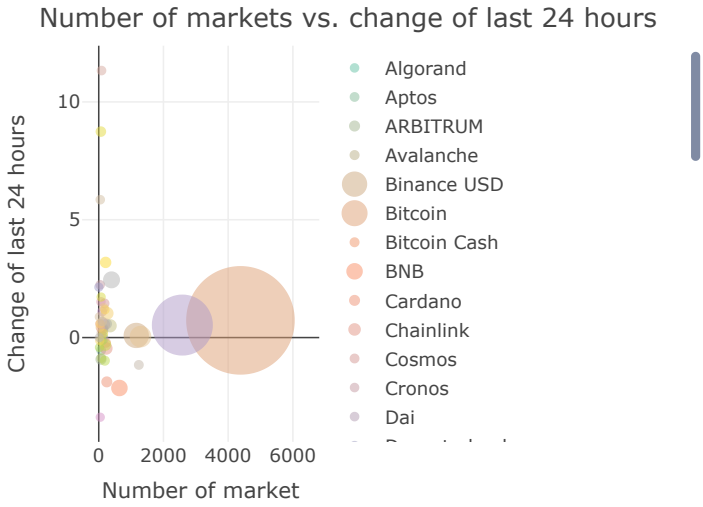


Figure 2

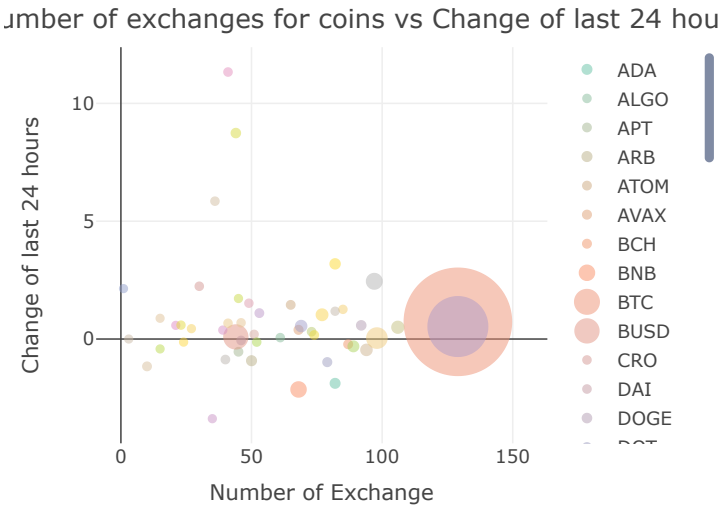

umber of exchanges for coins vs Change of last 24 hou
Change of last 24 hours
10
5
0
0
50
100
150
Number of Exchange
ADA
ALGO
APT
ARB
ATOM
AVAX
BCH
BNB
BTC
BUSD
CRO
DAI
DOGE

Figure 3

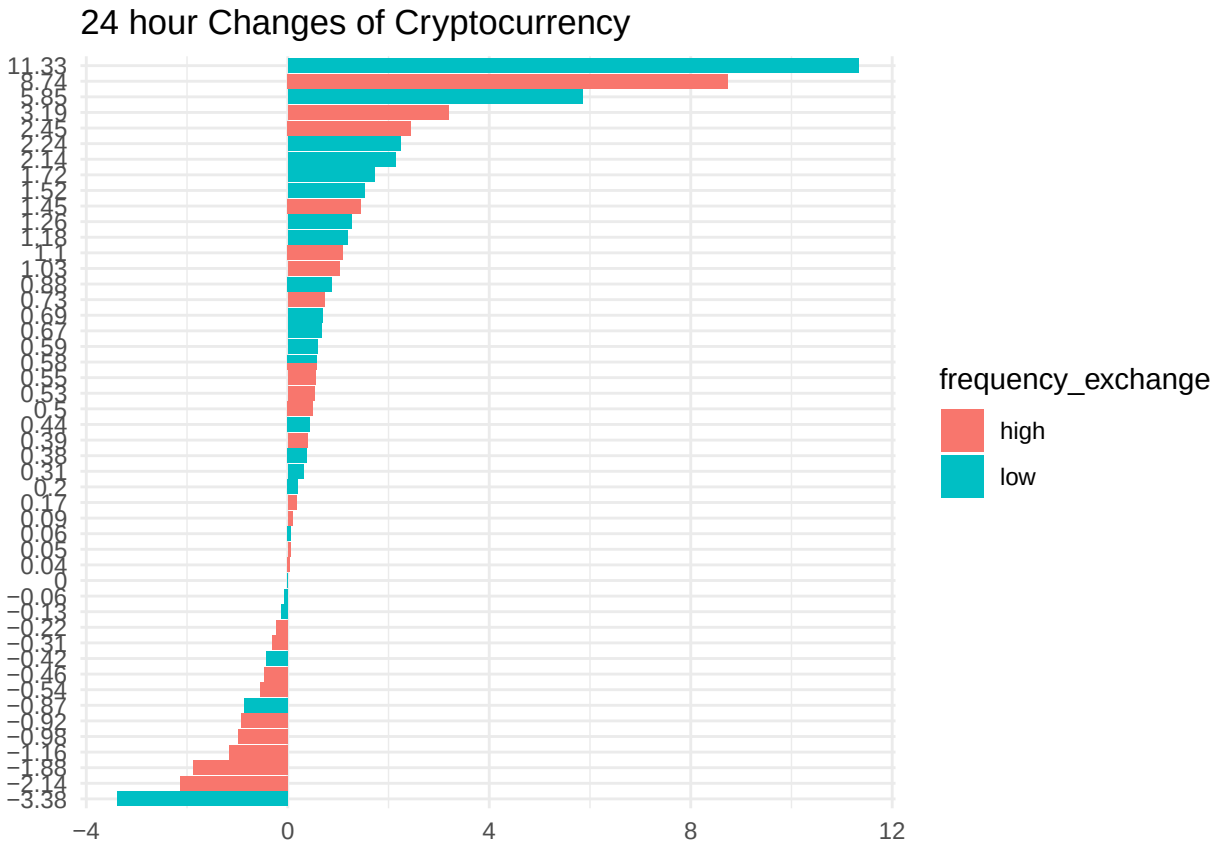


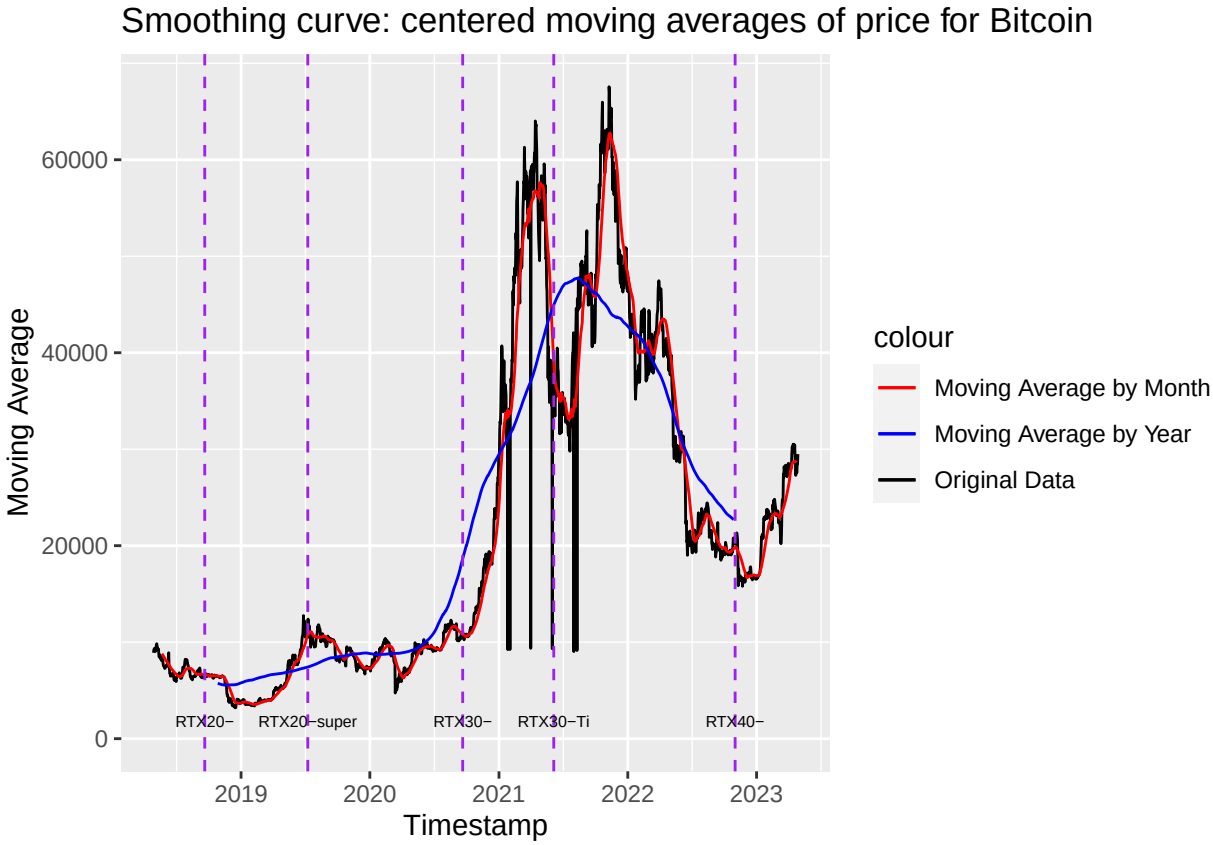
Table 3: Summary of Cryptocurrency group by volume

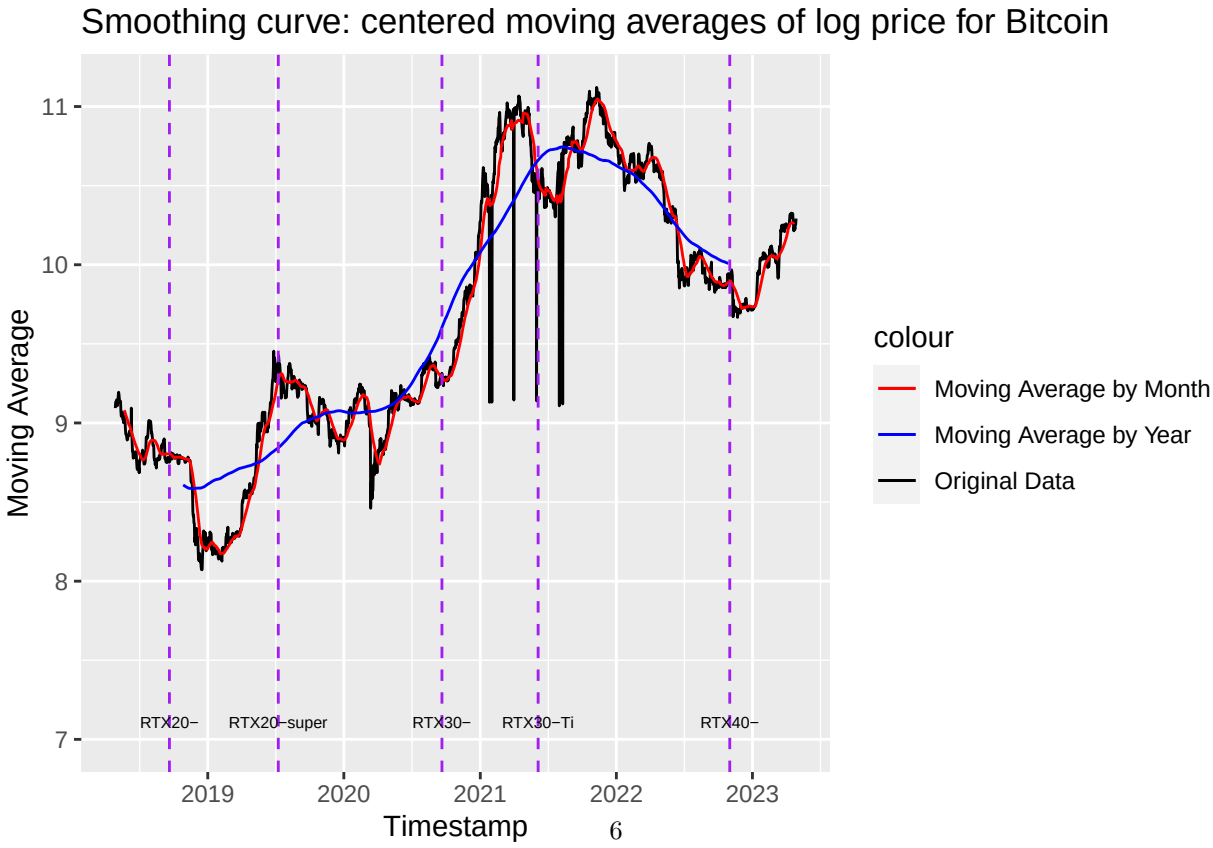
frequency_exchange	median_price	max_price	min_price	median_exchange_number	mean_exchange_number	mean_ratio
high	5.854824	29250.99	0.0000103	79	78.24	0.1193940
low	1.898785	29249.63	0.0223896	40	39.24	0.0284924

From the charts and table above, we can see that most of variables including number of markets, number of exchanges, frequency, etc., will not significantly affect the price change during last 24 hours. This is to be expected, and this is why these types of questions like predicting stock or cryptocurrency prices are so difficult to answer. Even finding the factors associated with the price is quite difficult. But we can see that the volume in last 24 hours (size of marker) is positively correlated with number of exchange and number of market from interactive plots.

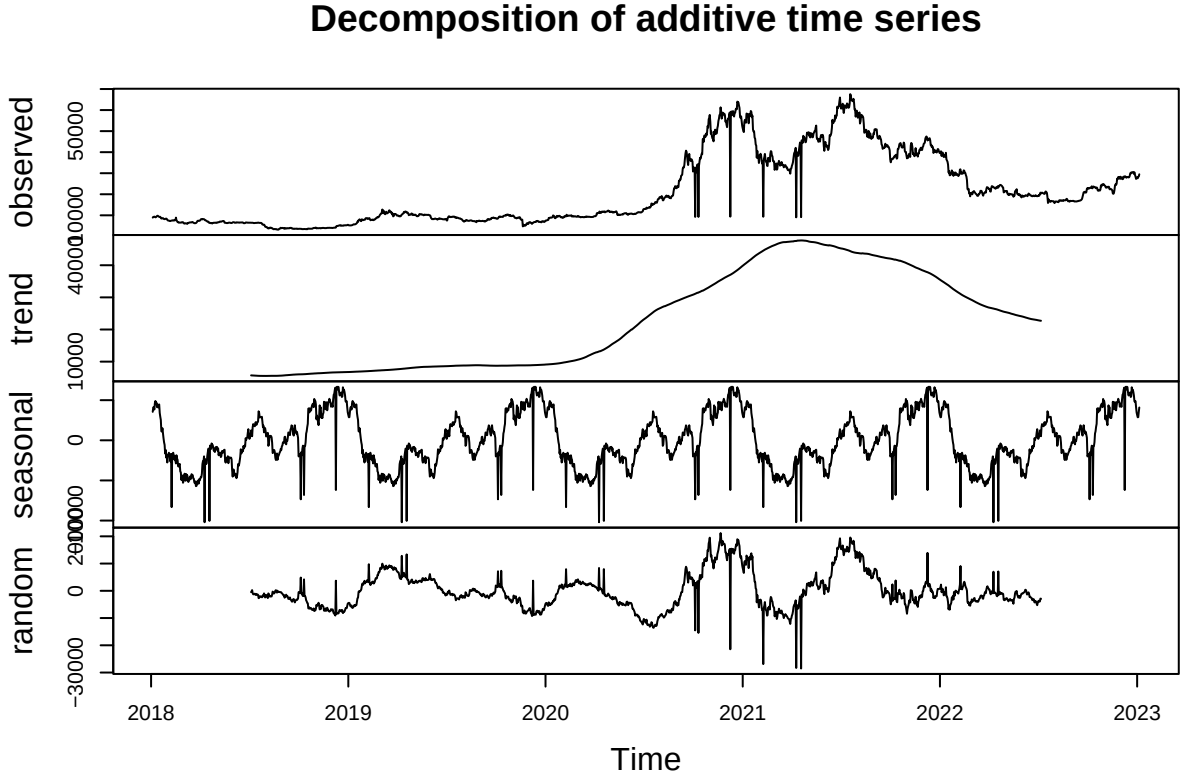
Time seriers analysis

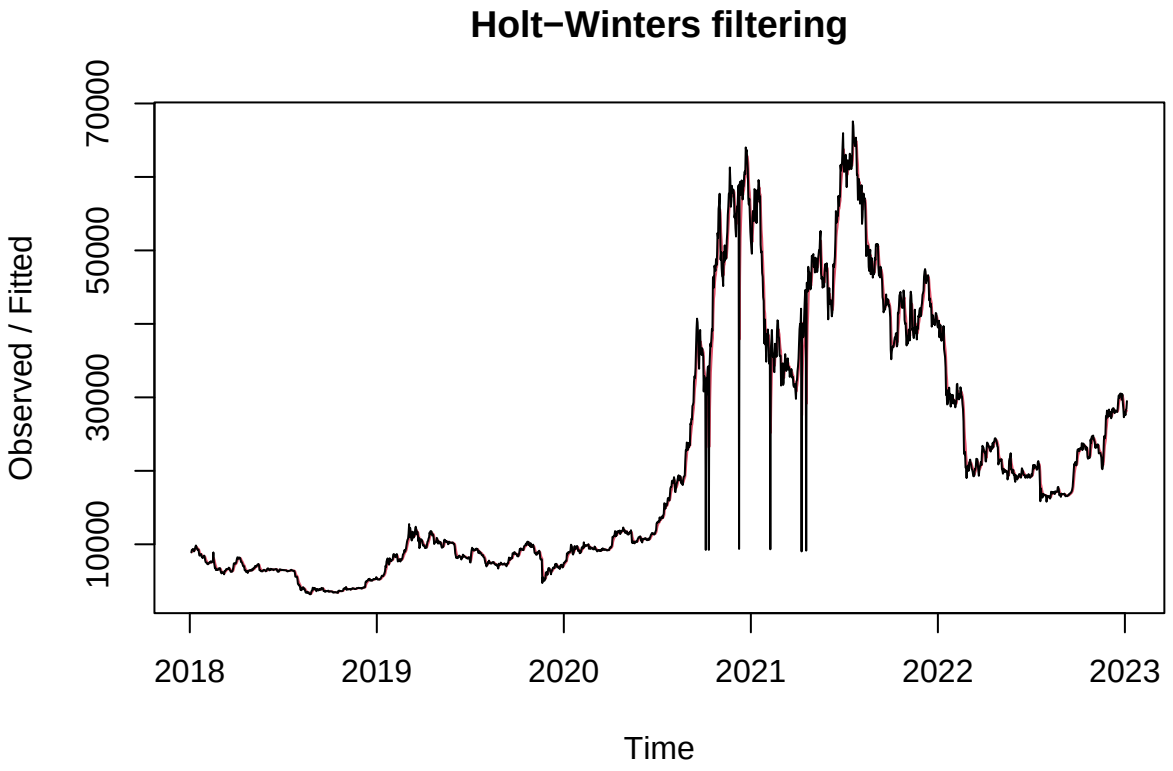
Time series for Bitcoin





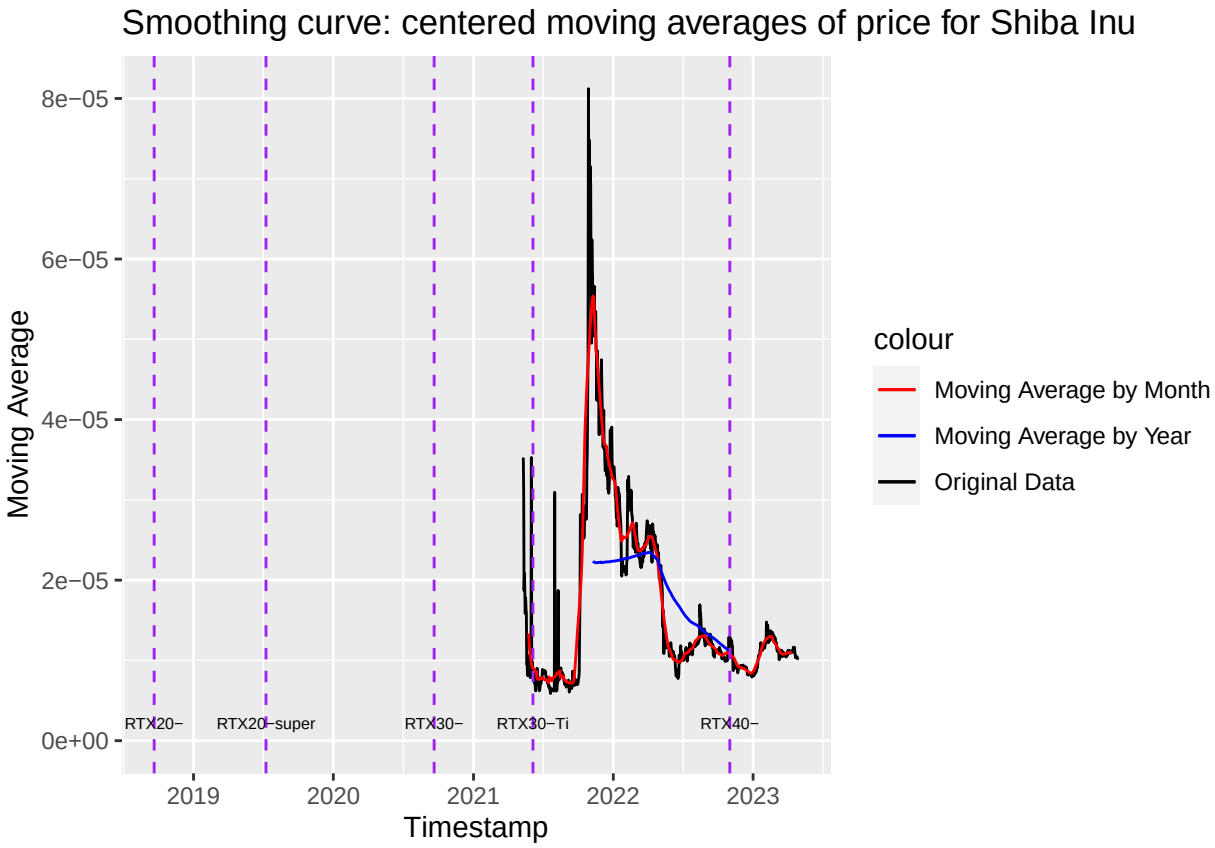
From the centered moving average plots, we can easily look the general trend of Bitcoin's price. Before the mid of 2021, the general price trend was increasing. But after the mid of 2021, the price is heading for a downward trend. In addition, as I mentioned in midterm project, price are rising rapidly in these days. In addition, since producing new cryptocurrencies requires a lot of computing resources, I had speculated in the midterm project that the price of cryptocurrencies might be related to the release of new GPUs. Now from the graph, I can not see a direct relationship between the release date of Nvidia GPU (main GPU for mining) and price of Bitcoin. My guess is that since the max amount of bitcoin remains the same, its price is not significantly affected by the current stock and therefore the GPU does not have a significant impact on its price. In addition, I would have liked to observe the historical supply versus historical price and the number of tweets for #Bitcoin hashtag versus price, but extracting historical supply from all cryptocurrency APIs and extracting number of tweets out of 7 days need premium. So I have to give up these ideas. Instead, I try to use additive decomposition model to explain the price versus date.

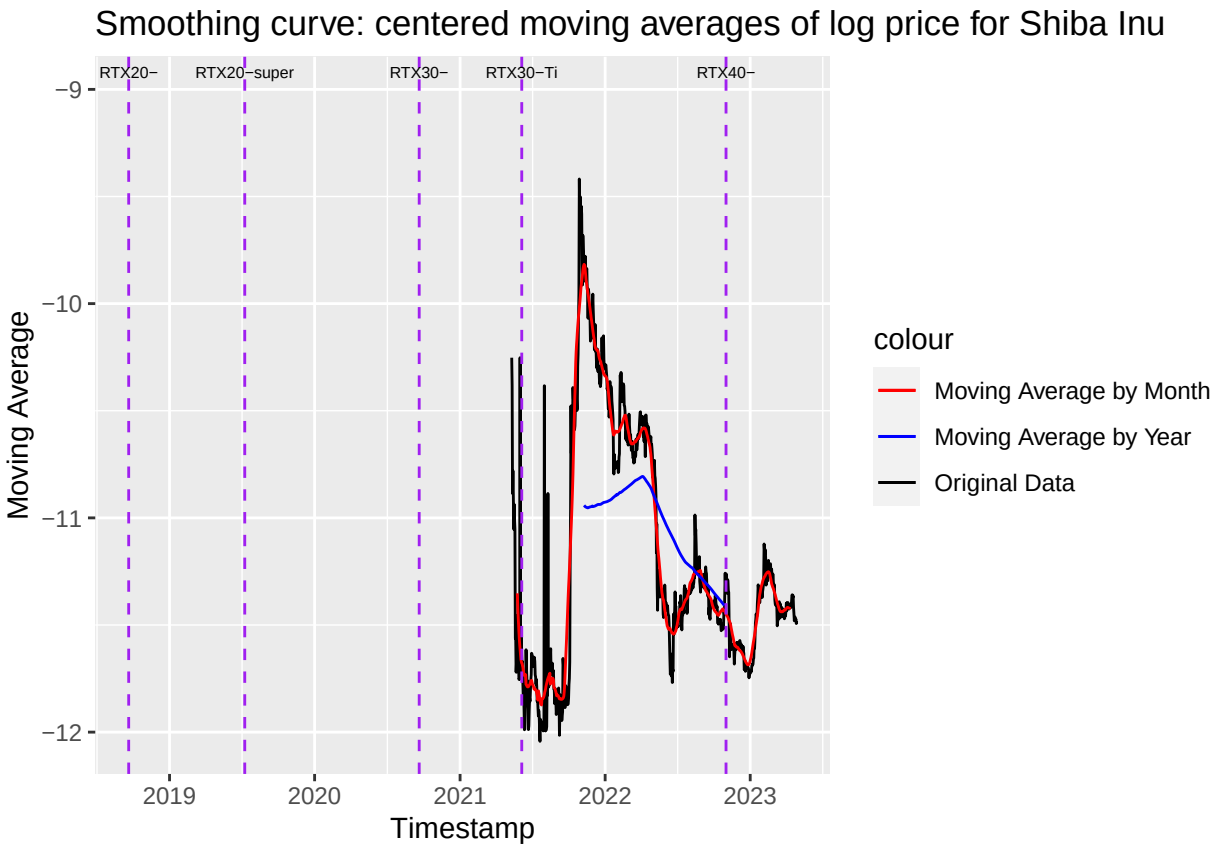




The irregular component can not be ignore in this case, so the accuracy of this model is relatively low. But we can still see that this additive time series model with simple exponential smoothing ($\alpha \approx 0.406$, a little higher than typical value, 0.2, which means less smoothing than usual case) is able to explain to some extent the effect of price by date. And this model is apparently overfitted, so we can not predict future price by this model. Additionaly, the sum of squared error is very high, 10967652000, because there are some outliers around 2021.

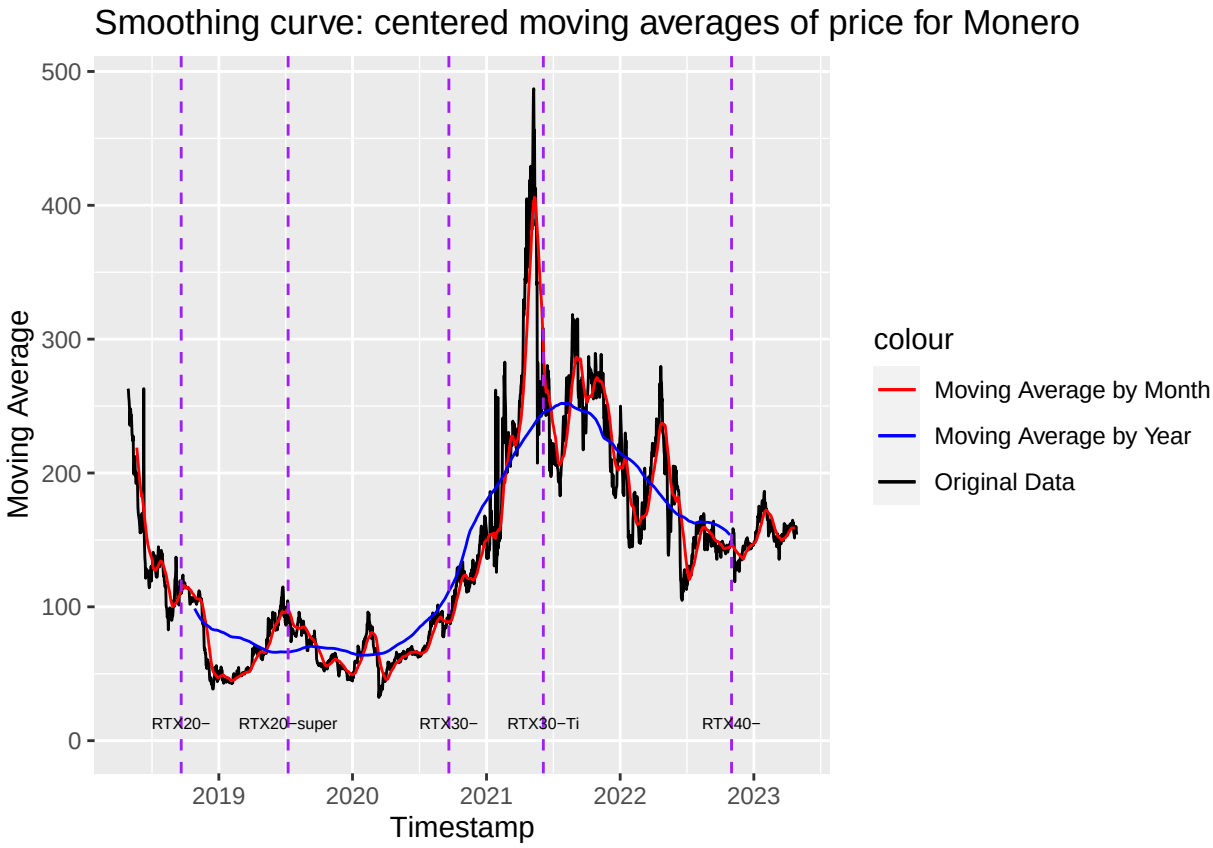
Time series for Shiba Inu

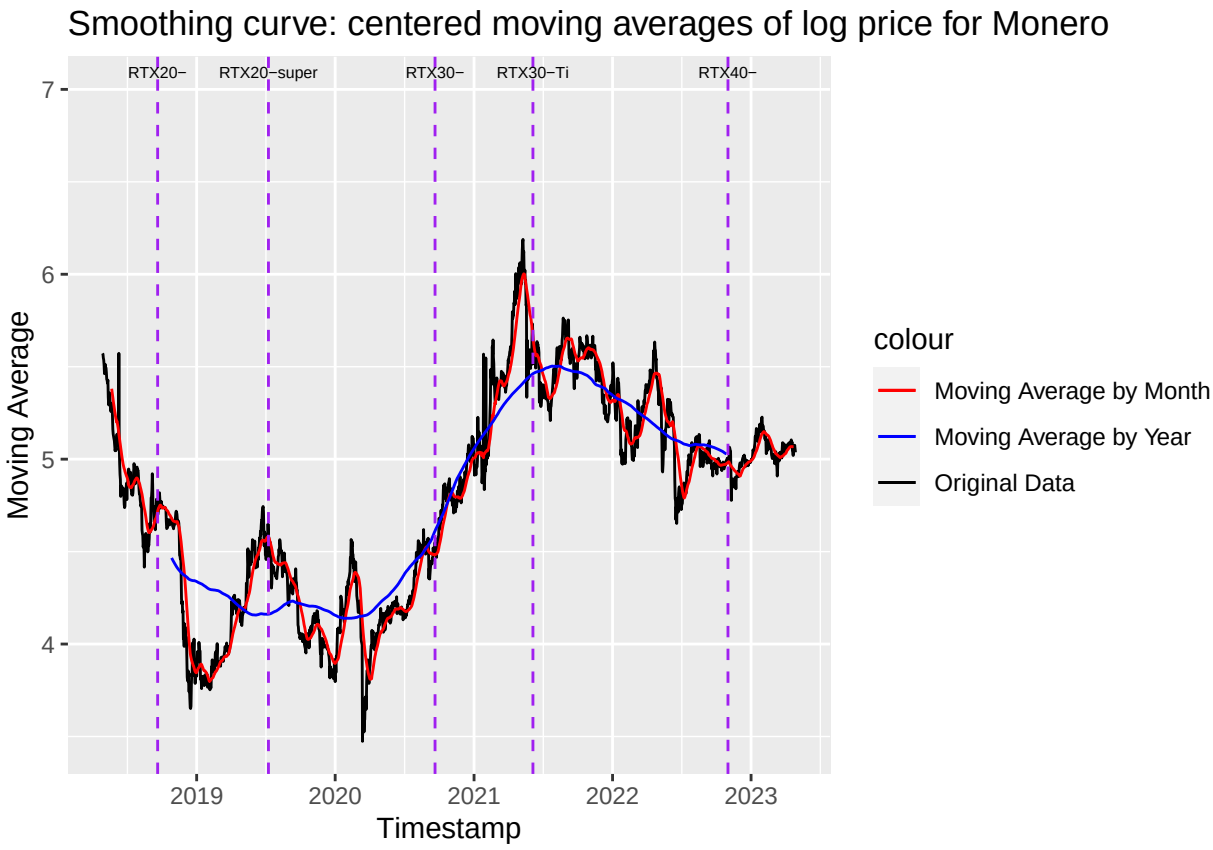


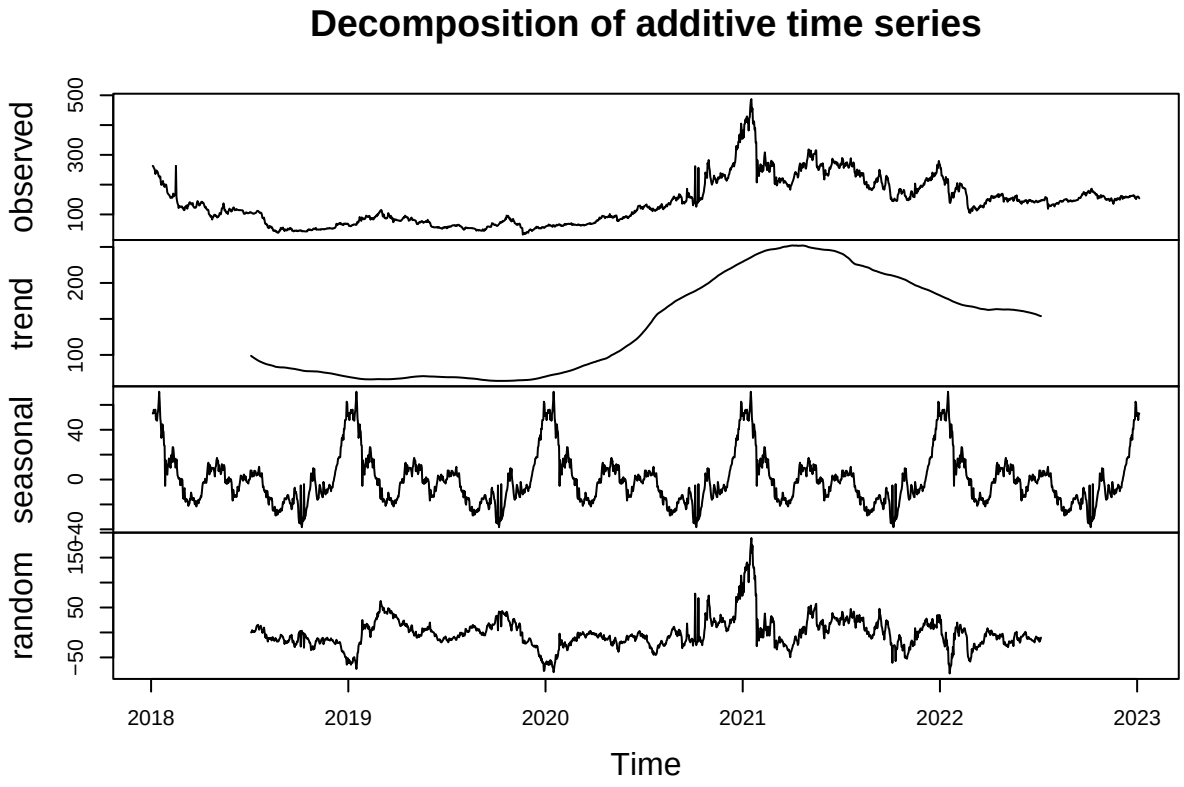


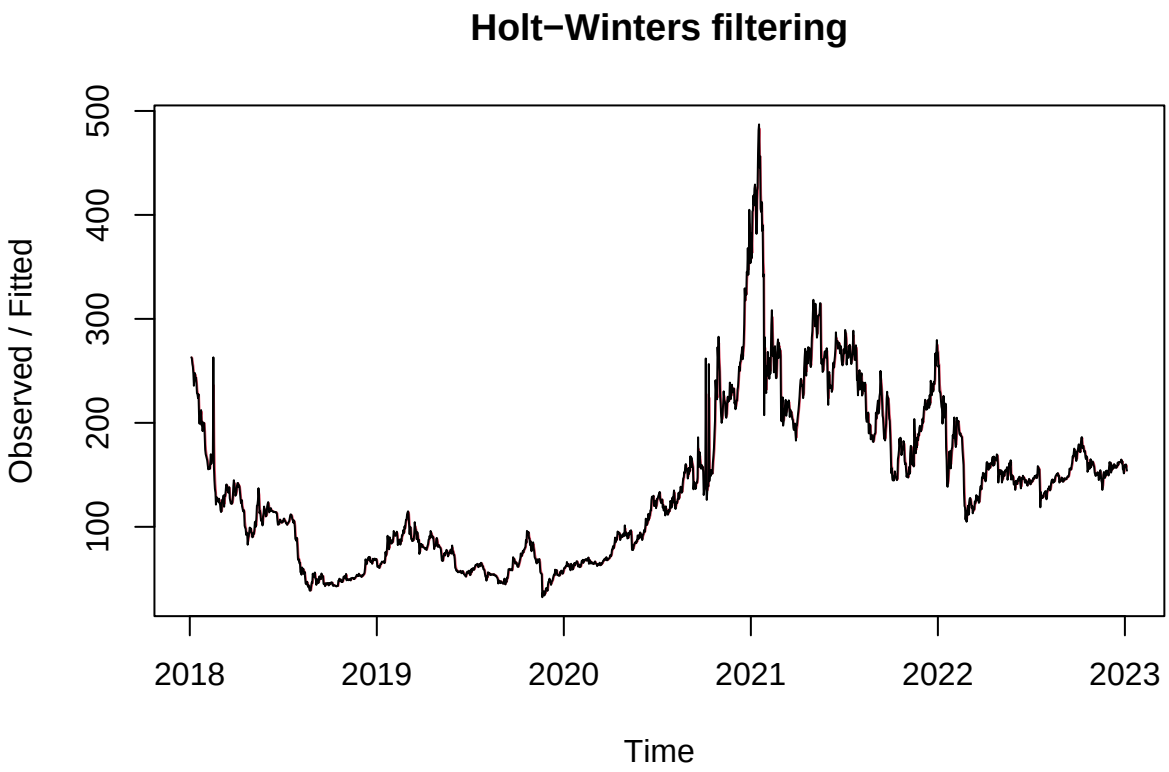
Shiba Inu's price trend is very much like a bit late Bitcoin's price trend. Increasing before 2021 November, then decrease after that. And rising after 2023. So I may say the price of Shiba Inu will rise in next several month. However, since there is just less than two years data, this is not suitable and actually can not fit an additive decomposition model.

Time series for Monero






Decomposition of additive time series
observed
trend
seasonal
random
100
300
500
100
200
40
0
−40
150
50
−50
2018
2019
2020
2021
2022
2023
Time



We can see that the trend component and seasonal component of Monero and Bitcoin are quite similar that is to be expected!

Conclusion

Due to the complexity of this problem, we did not find variables or factor that significantly affect the price or changes of cryptocurrency. But we still get some insights, such as the strong correlation among the prices of most cryptocurrencies, except for some stablecoins like Tether USD. Additionally, we find no strong correlation between gpu released time and cryptocurrency prices, which is not consistent with our guesses. Fortunately, additive decomposition model can explain to some extent the effect of price by date. However, this model tend to overfit the training data, especialy in this case. Thus we should not predict the future price by this model.

Limitations

There are two main limitations in our project. Firstly, we have explored only a few cryptocurrencies chosen based on price, change etc. So our conclusions may not be representative of all cryptocurrencies, but we can still get some insights from this project. Secondly, the variables we extracted is not enough. Like I said in methods part, the historical supply and number of tweets may give us more informative insights. However, due to lack of premium, I have no access to use these APIs.

Further Research

According to the limitations, we can analyze more historical prices of cryptocurrencies. And extracting those data I state above to do further research.