Analyzing Bitcoin Data

Sandhya

December 12, 2017

1 Abstract

In this paper, I aimed at analyzing the historical data of Bitcoin. The objective was to get deeper understanding of the underlying mechanism and to identify patterns in the data that may point towards the volatile change in Bitcoin price. My study was divided into three parts. First, to analyze the seasonal trends in the Bitcoin price. The goal was to find patterns in the data that can help predicting the future bitcoin price. Second, to find optimal variables surrounding Bitcoin and see the association of bitcoin price with those features. The goal was to understand how components in underlying network e.g blocksize, hashrate, transaction rate might be affecting the bitcoin price. Lastly, to find the relationship between Bitcoin price and Stock market price. The bitcoin market somehow works similar as stock market. So, I was interested to see how stock markets in different regions and various industry sectors might be affecting the price. The results showed that 1) there might be seasonal patterns in the bitcoin price. 2) some features show high correlation with bitcoin price. 3) there is no apparent relationship between bitcoin price and stock market price.

2 Introduction

Bitcoin is a decentralized digital cryptocurrency based on peer-to-peer transaction with no third party involvement[1]. The network of bitcoin is decentralized network of computers around the world that track of all bitcoin transactions in a public ledger called blockchain[6]. A blockchain is a long list of secured blocks that store the transaction information. Though the transactions are publicly available, nobody can see the source and person behind the transactions. The system was first introduced in 2009 but it has gained a lot of attention recently [1] and has resulted in tremendous surge in price of bitcoin in this year[7].

There are a lot of opinions floating around about price and future of bitcoin. As a data enthusiast, I will try to answer this question through the historical data available and will look for trends in the data. Moreover, I will try to find the association among bitcoin price and various factors that I think might be contributing in the surge in price of bitcoin.

3 Data

For this study, I used three open source datasets from online Kaggle Datasets repository. ***Cryptocurrency Historical Price*** [2] is a dataset consisting of 15 CSV files about various cryptocurrencies. Though all the files are used to find the relationship among prices of various cryptic currencies, there are mainly two CSV files used from this dataset. One of the files with 1655 rows and 7 features contains daily price data of bitcoins e.g close price, market capitalization, high price. Second file with 2920 rows and 24 features contains various features related to the bitcoin transaction data e.g block-size, transaction per block, hash rate. ***S&P 500 stock data***[3] is a U.S stock dataset containing historical price data for S&P 500 companies over the period of last five years. For this study, I selected three S&P companies of different sectors namely 3M Company (Industrial), Apple Inc. (Information Technology) and AFLAC Inc (Financial Life and health Insurance). The subset data for these companies were stored in three separate CSV files. ***Uniqlo(FastRetailing) Stock Price Prediction***[4] is a Japan stock dataset that contains the price data for Uniqlo company from 2012-2017.

3.1 Data Joining and Preprocessing

Since all datasets are time series, I joined the files based on Date. Specifically, I joined Bitcoin price file with Bitcoin transaction file to get the accumulated bitcoin data per day over the period of five years. Later, the Bitcoin data was joined with US and Japan stock data.

Since the data is collected online from automated processes, there is no missing data. The date format is not consistent across all the data files due to having different sources. hence, I made the date format consistent in order to join the data properly.

4 Results

4.1 Analyzing Seasonal Trends

Figure 1 shows the day-of-week analysis for bitcoin closing price over the period of five years and then separately for current year.Since, there has been rapid surge in the price of bitcoin this year, I was interested in comparing these results. Both graphs in figure 1 show that Monday and Tuesday have higher bitcoin price and Wednesday sees the lowest bitcoin price. This indicates that there might be some seasonal trend.

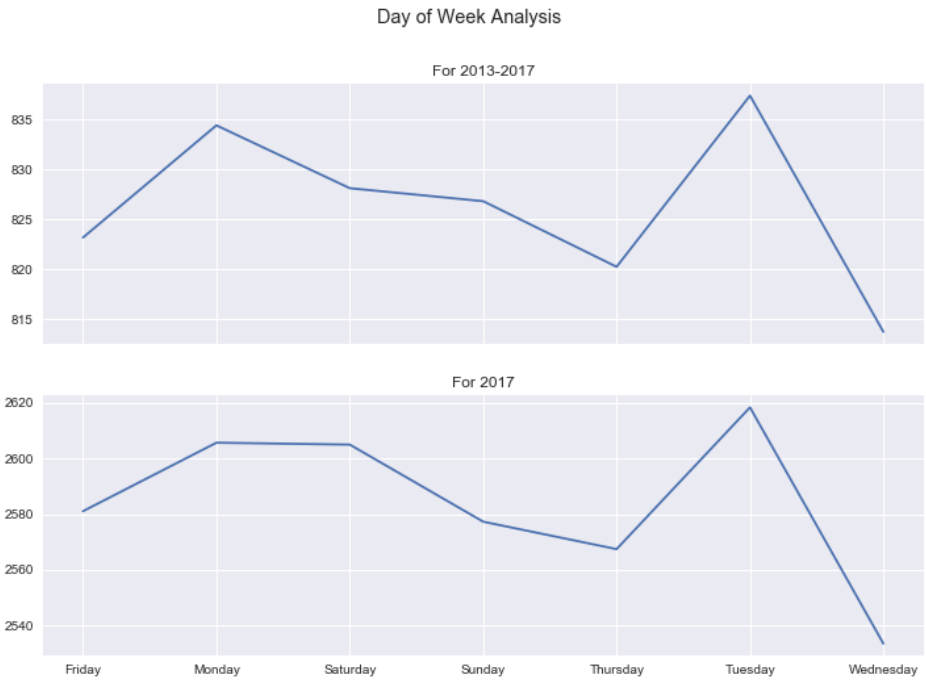

Figure 1: day-of-week analysis

But looking only at the average price may not be a good idea. Figure 2 shows the day-of-week analysis for the past five years along with the spread of price over those years.If we see left to right, First and third column show bitcoin closing price per week-day per year; second and fourth show bitcoin closing price along with its standard deviation. The difference in four graphs is the scale. First two columns are plotted focusing on that year's price range while last two columns are plotted using the scale of recent price of bitcoin. We can consider first two columns as zoomed in version focusing on the specific year while last two column focusing on overall picture. The figure shows some interesting results that I would want to highlight.

- column 1 in the figure 2 shows that from three of five years, Wednesdays see price drops and Mondays see price surge. It also shows that the trend of alternating years is comparable as well. For example year 2013, 2015 and 2017 somehow follow the similar trend.
- Column 4 in the figure 2 shows the bitcoin price is the most volatile in this year. The past years have also shown price variation[column 2 in figure 4] but price in 2017 is highly spread

out. So, there might be something different happening in this year which is affecting the price.

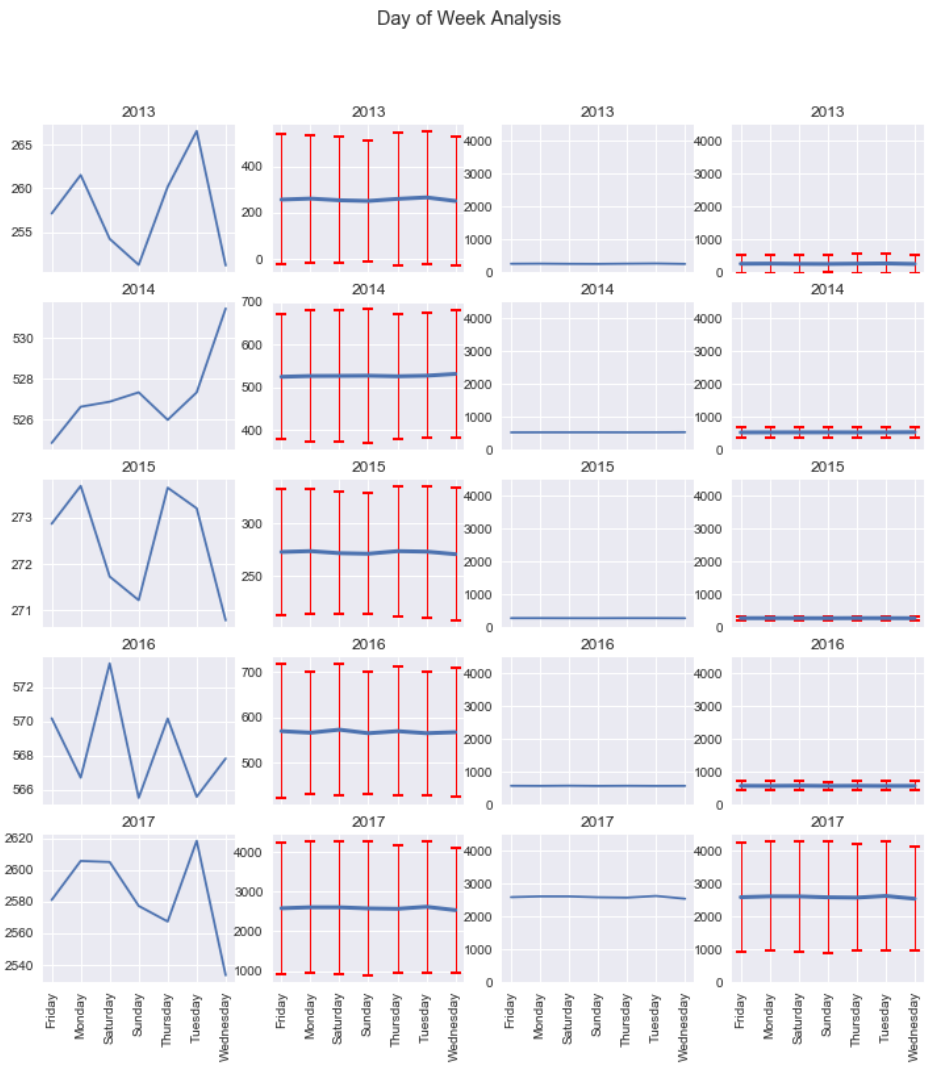

Figure 2: day-of-week analysis per year

Figure 3 shows the month-of-year analysis of closing price of bitcoin over the period of five years and then separately for current year. The top plot in the figure shows that bitcoin price goes up earlier in the year and starts to drop in October. But if we look at the bottom graph that is the most recent trend which shows that bitcoin price is going up throughout the year of 2017.

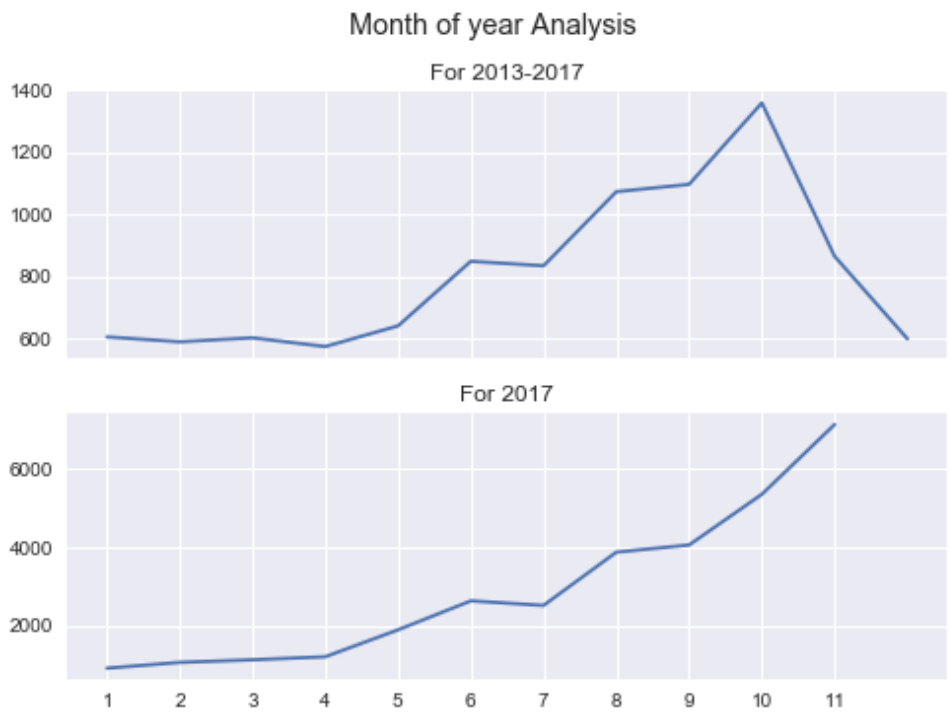

Figure 3: month-of-year analysis

But again, looking at just the mean can be misleading because of the high variance in the data. So, it is important to consider the spread of the data. Figure 4 shows the same analysis individually for five years along with the distribution of price in month. We can notice that there is no dip in price as it was in figure 3 when we considered only the average price of bitcoin.

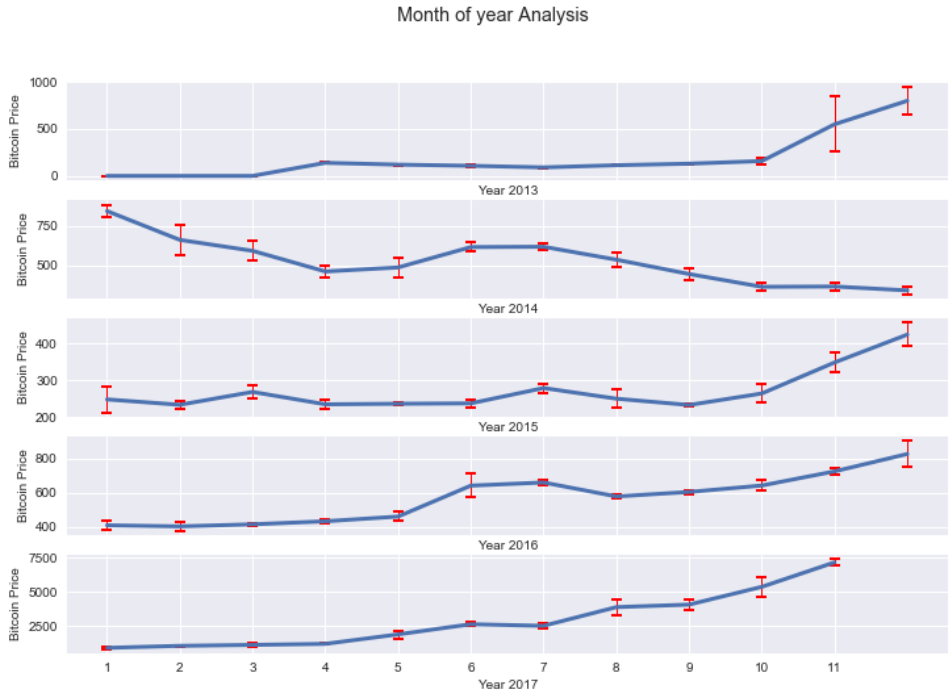

Figure 4: month-of-year analysis per year

4.2 Association among bitcoin price and related variables

Figure 5 shows how price changes with the number of bitcoins mined per day. This column was not actually present in the actual data rather I computed it by taking difference with previous day value of total bitcoins mined. Total bitcoins mined represents the total number of bitcoins mined till date but I wanted to have the number of bitcoins mined per day to see the relationship. It can be seen in figure 5 that there is an inverse relationship between bitcoin price and bitcoin mined per day. As number of bitcoins increases, the price of bitcoin decreases. This also seems to follow the supply-demand law of economics. Higher supply results in lower demand and lower demand results in reduced price.

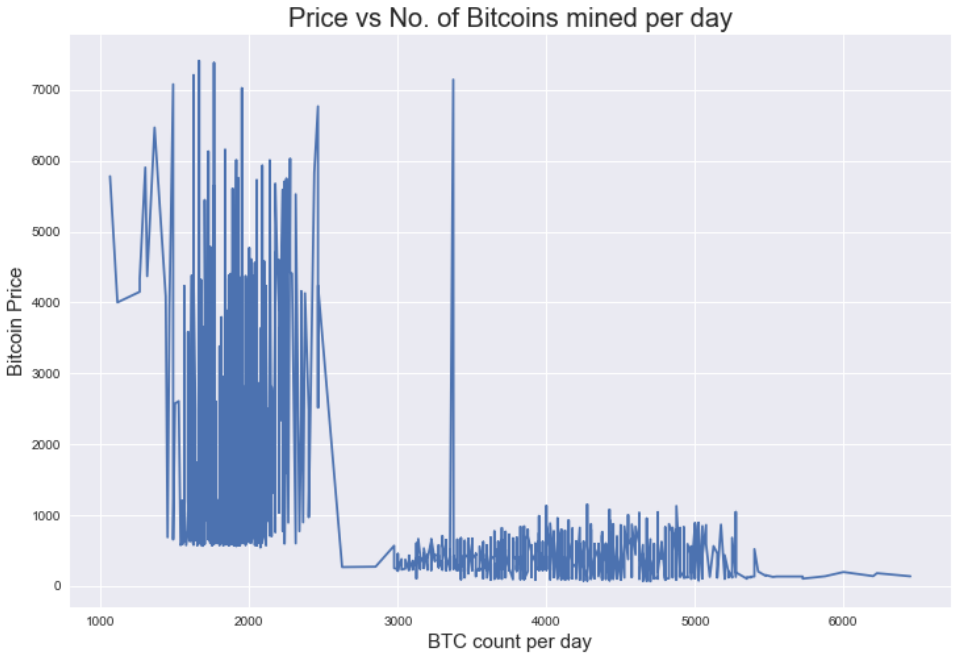

Figure 5: BTC price vs No. of BTC

After some thorough research, I considered 7 features that may be correlated with the price of bitcoin. The features are summarized in the table below with description taken from [2].

FEATURE	DESCRIPTION
btc_market_cap	market capitilization of btc
btc_n_transactions_per_block	transactions per block per day
abs_btc_count	bitcoin mined per day
btc_hash_rate	hash power per day
btc_difficulty	difficulty in finding a new block.
btc_cost_per_transaction	miners revenue per transaction
btc_n_transactions	number of Bitcoin transactions.

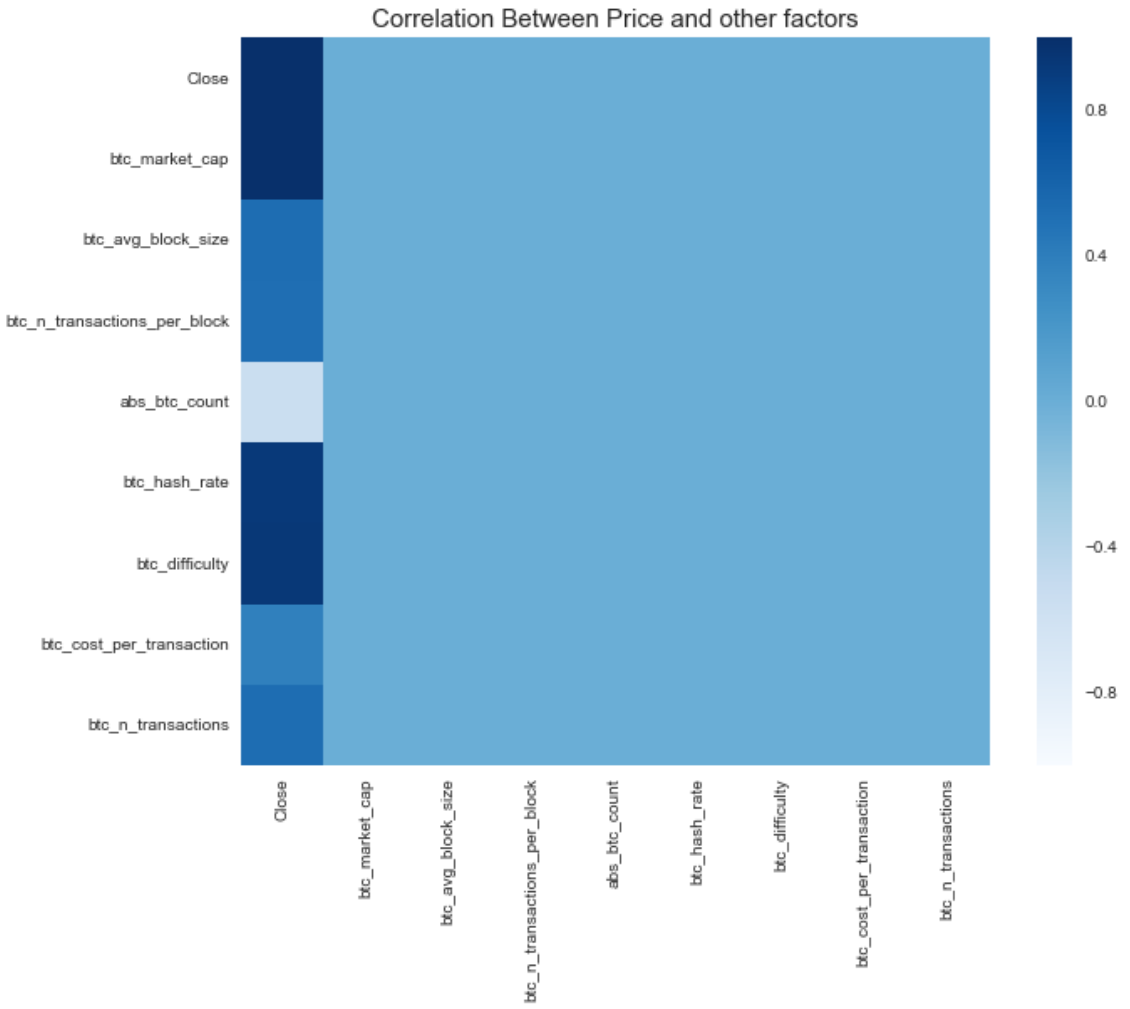

Figure 6: BTC price vs various features

Figure 6 shows the Pearson Correlation Coefficient of bitcoin price with above mentioned features. It can be seen that bitcoin is positively correlated with market capitilization value, hash rate and bitcoin difficulty measure. The results make sense because the more difficult it is to mine the bitcoin, the greater hash rate it would be needed. We can also imply that greater difficulty should result in less number of bitcoins mined per day hence price should go up as there is less supply and probably more demand. Again, this is what I hypothesize about the results. It may be possible that more research and indicators should be considered to reach any conclusion.

4.3 Association between bitcoin and Stock market price

This section shows whether there is any association between bitcoin price and stock market price. For this study, I compared bitcoin price with three different stock sectors in US market:

- Apple Inc (Information Technology)
- 3M company (Industrial)
- AFLAC Inc (Financial Life and health Insurance)

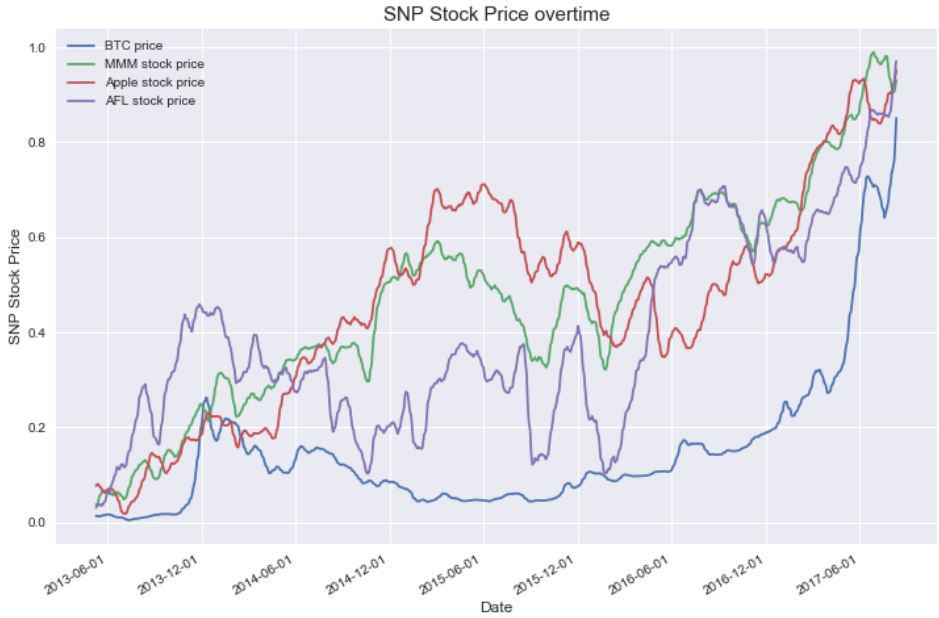

Figure 7: BTC vs US companies stock price

Figure 7 shows the trend of bitcoin price and three S&P companies' stock price. There seems to be no apparent relationship between bitcoin price and the stock price. Please note that the plots show the moving average of price over the 10 day window to smoothen the data.

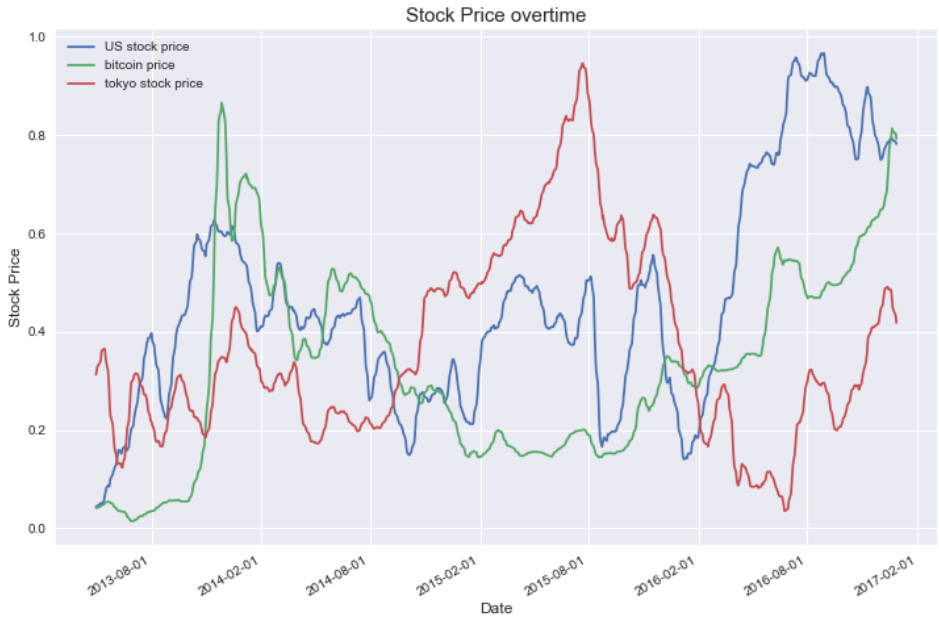

Figure 8: BTC vs Tokyo and US stock price

Figure 8 shows the bitcoin price, tokyo stock price and US stock price distribution over the same time period. There does not seem to be a strong association between bitcoin and stock price.

4.4 Bitcoin Price Spread

Figure 9 plots the price standard deviation individually for five years. The goal is to check the stability in bitcoin price. It seems that bitcoin price has been highly unstable in 2017. This deduction is important in concluding that the surge in price of bitcoin may just be a bubble which will burst when the price becomes robust enough to not react to external factors.

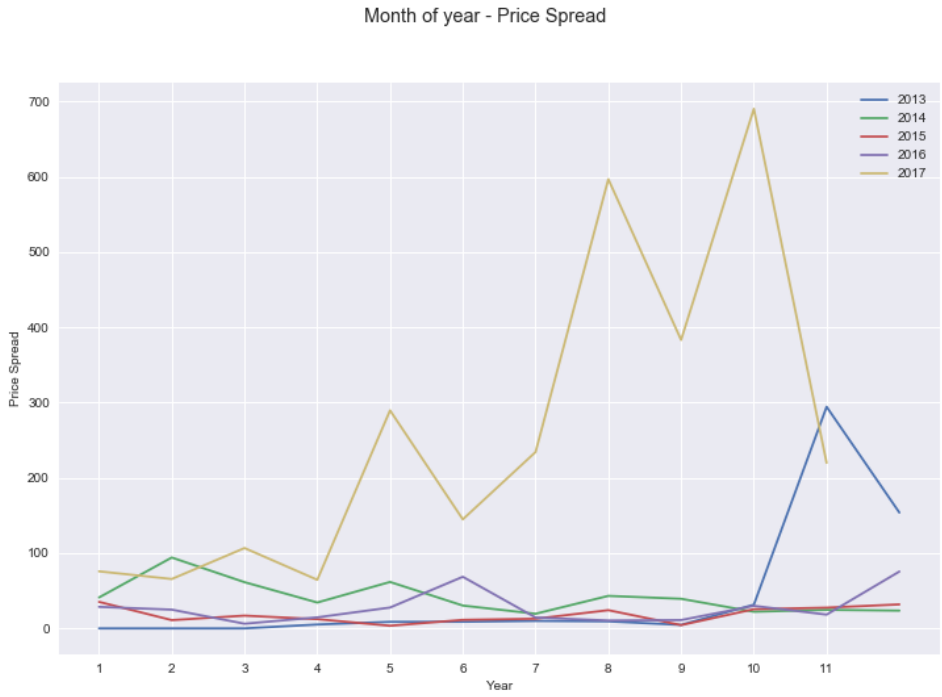

Figure 9: Bitcoin price stability over the period of 5 years

5 Discussion

The study shows that the price of bitcoin follows a seasonal pattern. however, it seems that more recent data would be helpful in reaching any solid conclusion. The study also shows that the underlying features of the bitcoin network might be good predictors for the bitcoin price. A small subset of features was chosen based on author's understanding, but it might be a good idea to look at all the features and look for unseen relationship that might be there. Bitcoin is a fixed supply of 21 million, so it will be interesting to see what will happen to bitcoin once all the bitcoins are mined. Lastly, it will be interesting to see how news affects the price of bitcoin.

References

[1] Madan, Isaac, Shaurya Saluja, and Aojia Zhao. *Automated Bitcoin Trading via Machine Learning Algorithms*. (2015): 1-6.

[2] Kaggle: Cryptocurrency Historical Prices
`https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory`

[3] Kaggle: S&P 500 stock data
`https://www.kaggle.com/camnugent/sandp500`

[4] Kaggle: Uniqlo(FastRetailing) Stock Price Prediction
`https://inclass.kaggle.com/daiearth22/uniqlo-fastretailing-stock-price-prediction/data`

[5] Russ Wilcox: Blockchain—Why Care?
`https://medium.com/pillar-companies/blockchain-why-care -c0890c71454c`

[6] Nathaniel Popper: What Is Bitcoin?
https://www.nytimes.com/2017/05/15/business/all-about-bitcoin-the-mysterious-digital-currency.html

[7] Coindesk: Bitcoin Price
https://www.coindesk.com/price/