Bitcoin & the Digital Gold Rush

Mining the data for Traditional Finance

Garrick Chu (March 2018)

Executive Summary

- Strong evidence that Bitcoin is a storage-of-value rather than a medium-of-exchange.

- Bitcoin offers portfolio diversification but does not appropriately compensate per unit of risk.

- Strong evidence that Bitcoin is experiencing an economic bubble (and still is despite recent correction).

- It is difficult to estimate future values of Bitcoin using Price/Volume of other coins as a predictor.

Topic Overview

- Brief Introduction to Bitcoin

- The Problem: Bitcoin's Application in Traditional Finance

- Analysis and Results

- Recommendations

- Limitations and Further Research

Introduction to Bitcoin

- Product of a 2009 whitepaper entitled "*Bitcoin: A Peer-to-Peer Electronic Cash System*" penned by nom-de-plume "Satoshi Nakamoto"

- Bitcoin is a peer-to-peer network which transactions can take place, circumventing traditional banking systems (and potential transaction/processing fees)

- Network is underpinned by new technology, Blockchain (decentralized ledger/record-keeping system)
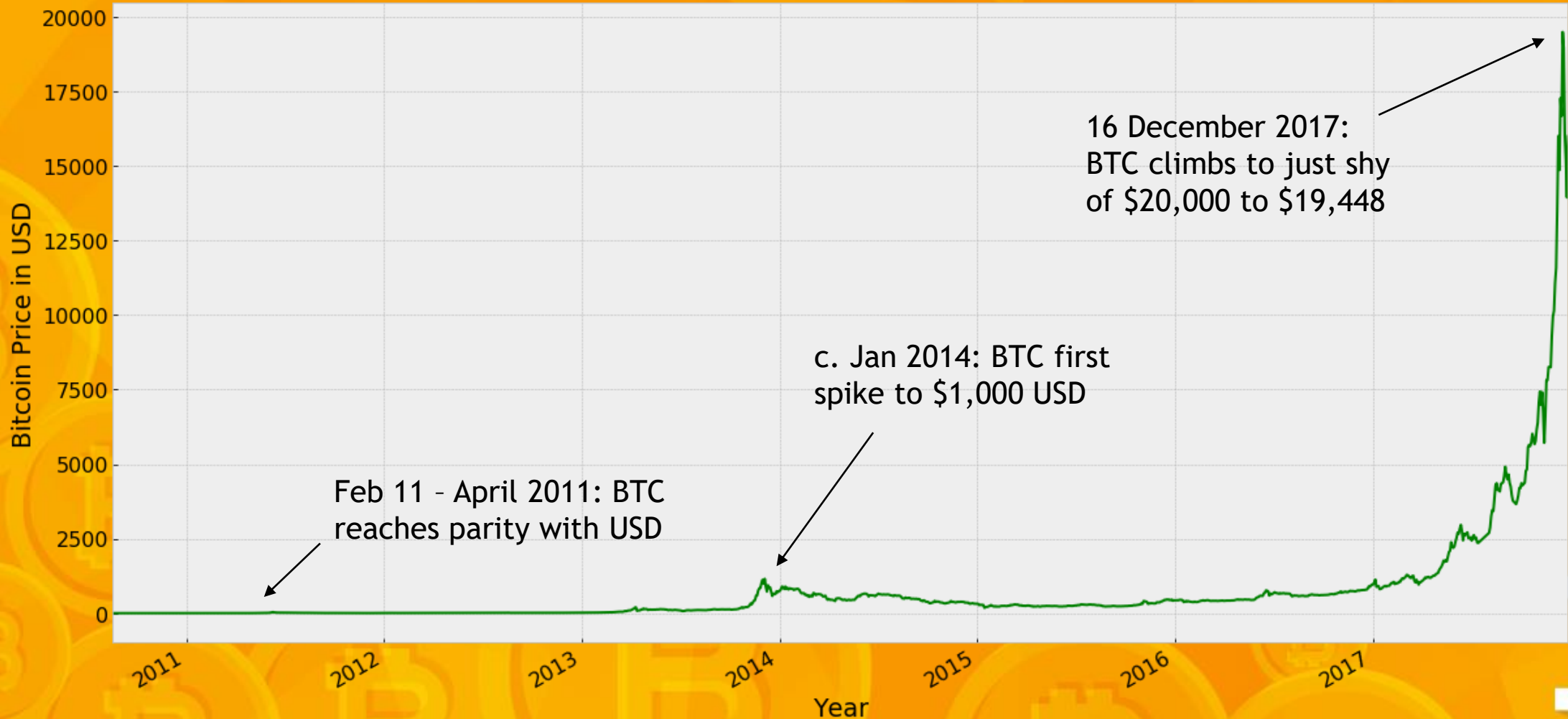
Present Day Bitcoin

- Bitcoin's Market Cap - $200 Billion USD
- CBOE Launches Bitcoin Futures (Dec. 11, 2017)
- Google searches for "Bitcoin" peak in December 2017

Bitcoin by the numbers

Historical Bitcoin Prices

16 December 2017:
BTC climbs to just shy
of $20,000 to $19,448

c. Jan 2014: BTC first
spike to $1,000 USD

Feb 11 – April 2011: BTC
reaches parity with USD

The Problem

- With the emergence of Bitcoin (and other cryptocurrencies), how do we evaluate within and incorporate into traditional financial frameworks?
  - Can Bitcoin serve a purpose in traditional asset/portfolio management?
  - Do we define Bitcoin as a medium-of-exchange (currency) or as a storage-of-value?
  - Are we witnessing yet another asset bubble?
  - Can we model and predict Bitcoin Prices?

Goals of Analysis and Tools

- Assess Bitcoin's adoption as a medium-of-exchange

- Viability of Bitcoin as Portfolio Management tool
  - Statistical methods used to find correlations among assets
  - Domain Knowledge: Modern Portfolio Theory, Sharpe Ratio

- Determine if Bitcoin is experiencing an asset bubble? Prime example of "greater fool theory"?
  - Machine Learning: K-Means (as Clustering Method), Principal Component Analysis
  - Statistical Methods (correlation analysis with other asset bubbles)

- Attempt to predict Bitcoin prices
  - Machine Learning Methods: Linear Regression, Autoregressive Integrated Moving Av

Bitcoin: Medium-of-Exchange (currency) or Storage-of-Value (investment vehicle)

- Methodology
  - Calculate ratio of "trade" transactions (BTC bought/sold for fiat currency) to total transactions (which includes both transactions executed on an exchange such as Coinbase and peer-to-peer payments).

  - If the ratio of trade transactions to total decreases over time, this implies that either trade transactions are decreasing relative to all transactions or that peer-to-peer payments are increasing relatively.

  - If the ratio is increasing, we reject the hypothesis that Bitcoin is increasingly becoming a medium-of-exchange in favor of the theory that Bitcoin is likelier a storage-of-value

Bitcoin: Medium-of-Exchange (currency) or Storage-of-Value (investment vehicle)

Trade Volume to Total Transaction Volume
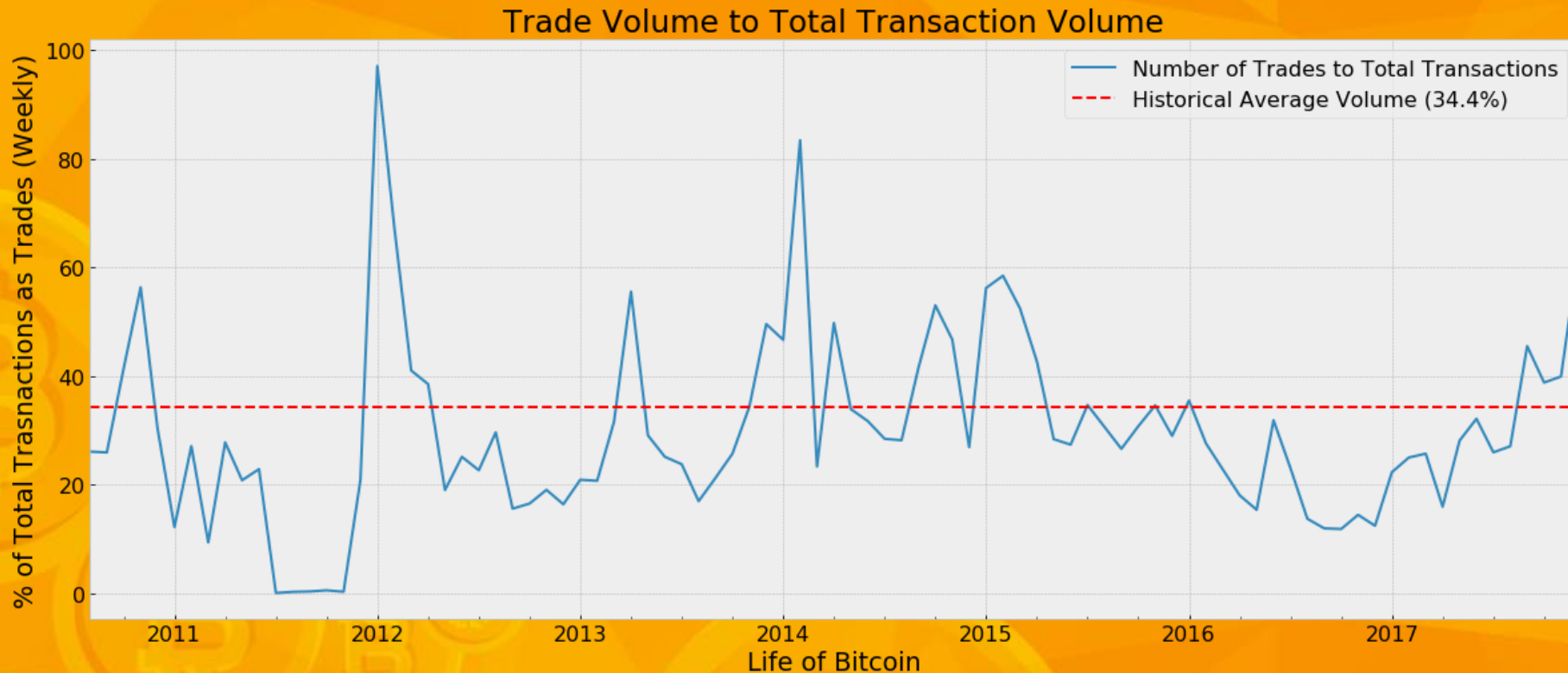
The ratio has no clear pattern over time.

The ratio has climbed nearly three-fold in 2017 and corresponds with the astronomical rise in Bitcoin prices.

Given this measure, we reject the notion that Bitcoin is increasingly being adopted as a medium-of-exchange.

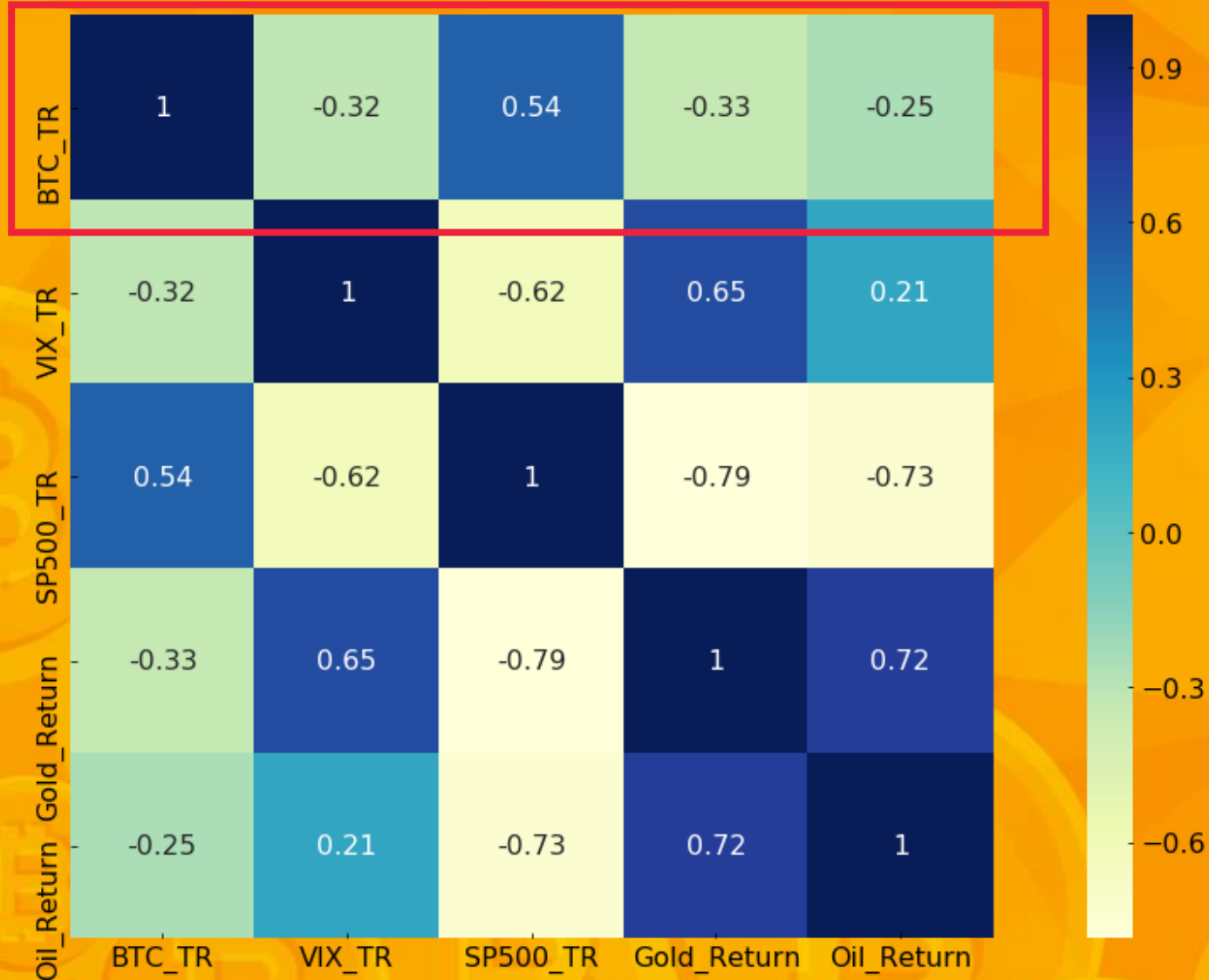Data Sources: Blockchain.info, Bitcoinity.org

So... it's an investment. Should I include in my or a client's portfolio?

- Let's evaluate within the context of Modern Portfolio theory...

- Modern Portfolio Theory (Markowitz) seeks to maximize diversification and expected return based on a given level of market risk.

- For Bitcoin to be an effective portfolio management tool, its returns will need to be uncorrelated to other assets and offer an attractive risk-adjusted return.

In search of diversification…

	BTC_TR	VIX_TR	SP500_TR	Gold_Return	Oil_Return
BTC_TR	1	-0.32	0.54	-0.33	-0.25
VIX_TR	-0.32	1	-0.62	0.65	0.21
SP500_TR	0.54	-0.62	1	-0.79	-0.73
Gold_Return	-0.33	0.65	-0.79	1	0.72
Oil_Return	-0.25	0.21	-0.73	0.72	1

- Using statistical methods in Python, determined the correlation of returns among Bitcoin and traditional assets the S&P 500, Oil, Gold and the CBOE Volatility Index[1].

- Out of the bunch, we find that Bitcoin's returns are most correlated with the S&P 500 although <u>not</u> to a substantial degree.

- As a result, there is diversification benefit from integrating Bitcoin in a traditional portfolio

Data Sources: Yahoo! Finance, Blockchain.info, Chicago Board of Exchange

[1] CBOE Volatility Index: the market's expectation of 30-day volatility as measured by implied volatility in S&P 500 Index options

… and risk-adjusted returns

$$S = \left(\frac{R_p - R_f}{\sigma_p} \right)$$

- Using the formula for the Sharpe Ratio as:
  - R(p) will be the Mean Return of Bitcoin
  - R(f), known as the Risk-Free Rate, is a theoretical rate of return of an investment with zero risk. For demonstrative purposes, we will consider the return of the S&P 500 as the "risk-free rate" in lieu of the more commonplace usage of US Gov't Debt Yields.

- With the mean return of Bitcoin of 10,769.98% and the mean return of the S&P 500 (over the same time period as Bitcoin's existence) of 1.64%, we arrive at an excess return value of = 10,768.34%. Sounds promising, right?

- Not quite…The standard deviation of Bitcoin returns is a staggering 29107.12% (compared to the S&P 500's 0.38% over the same period).

- **Our resulting Sharpe Ratio for Bitcoin = 0.37.** Compare this with the S&P 500's ratio of 4.32.

- Sharpe Ratio > 1.0 is considered good and represents a typical decision boundary (accept the investment if > 1, reject if < 1).

Is Bitcoin another asset bubble?

- Two analytical methods:
  - 1) Determining correlation between Bitcoin and prior asset bubbles (Dot-com bubble, US Housing Bubble, Dutch Tulip mania)
  - 2) Employing Machine Learning methods to classify "eras" of Bitcoin and attempt to identify asset bubble "territory"

- Introduction to the data:
  - For this analysis, I am using a 12 month period of Bitcoin Prices from Jan 2017 to Jan 2018 which captures the rise of Bitcoin prices that has garnered the most media attention.
  - The Dot-Com Bubble is represented by the Nasdaq Composite Index spanning a 12-month period leading up to when the index peaked at 5,048.62 on March 10, 2000.
  - The Dutch Tulip index data contains only 14 data points and is adopted from the work of: Garber, 1990; Krelage, 1942; Thompson, 2007.
  - The US Housing Bubble is represented by the Case-Shiller US Home Price Index. The index is measured on a monthly basis and is the product of slower market dynamics and price discovery mechanisms. Therefore, I am using a 12 year period to mimic the 1 year time scale as the other datasets.

Bitcoin Bubble: Method #1
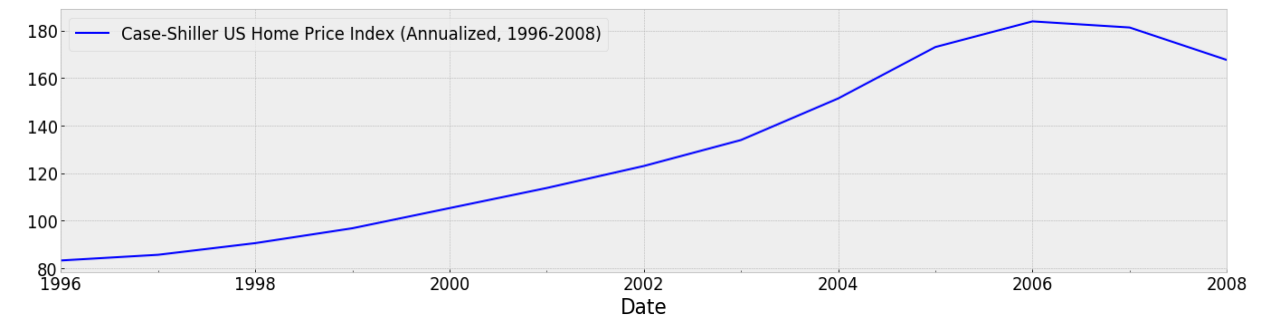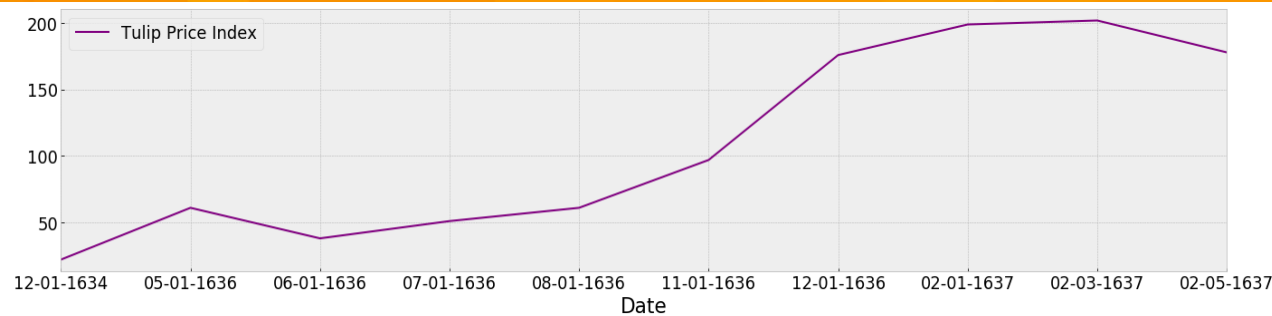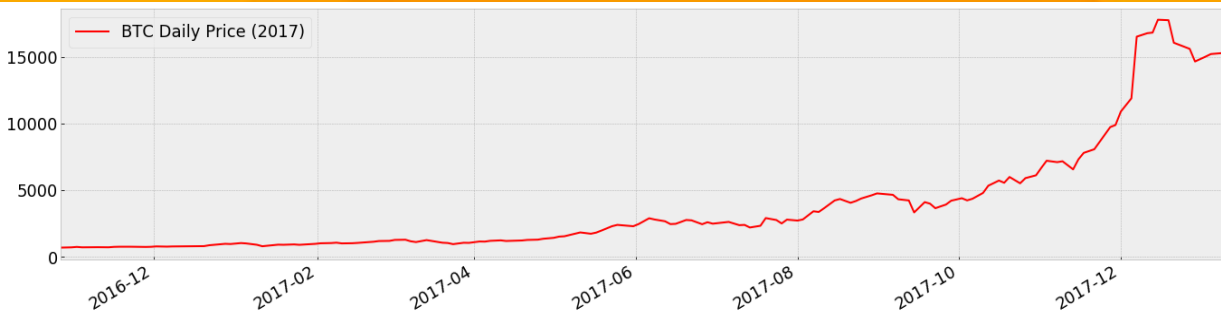
Top to bottom, left to right: Bitcoin Prices from December 2016 to present, Nasdaq Composite from 1999-2000, Tulip Price Index from 1634 to 2642, Case-Shiller US Home Price Index from 1996-2008.

Data Sources: Blockchain.info, Yahoo! Finance, The Tulip Price Index (Garber, 1990)(Krelage, 1942)(Thompson, 2007), Federal Reserve Bank of St. Louis. Additional reference from Goldman Sachs January 2018 Investment Outlook (link).

Method #1 (Cont'd)

Pearson R correlation for Bitcoin (2017) and the Dot-Com Bubble: 0.85.

R-squared of BTC and Dot-Com Bubble: 0.72.

Pearson R correlation for Bitcoin (2017) and US Housing Bubble: 0.75.

R-squared of BTC and US Housing Bubble: 0.57

- Conclusion:

- With respect to the US Housing Bubble, there is a firm correlation but unfortunately the assumed fitted regression line only explains ~57% of the variance in the data.

- The Pearson R correlation suggests that 2017 Bitcoin is strongly correlated to the Dot-Com Bubble with the assumed fitted regression line explains ~72% of the variance in the data. As a result, one should consider the recent price action of Bitcoin carefully and would be fair to draw similarities and inferences to the Dot-Com bubble and resulting crash.

  - In Goldman Sach's Investment outlook: "We should also add that we do not believe a collapse in bitcoin will have major contagion effects on the global economy or financial markets.

Method #2: Finding Stages of Bitcoin

- Objective: Determine if there are distinctive "eras" in the Bitcoin's lifespan and if one of those "eras" can be distinctively isolated as a "bubble" era/phase.

- Method: Framing as a classification problem and attempting to classify "eras" in Bitcoin market dynamics (irrespective of time). I suspect price and transaction volume to be significant factors.
  - Within the subset of classification models, I chose to use an unsupervised method of classification called K-Means
  - K-Means aims to label data into discrete clusters based on given characteristics
  - Also leveraging a Scaling function and Principal Component Analysis to exclude the effect of scale in the dataset

Method #2: Data Preparation

- Input Data: Daily Bitcoin Metrics data including: Average Block Size, Cost Per Transaction, Number of Transactions, Volume and Average Bitcoin Qty per Transaction (total of 10 features/predictors)

- Data Process: Obtain and format, scale, split data into test/training set, train and test Clustering model (K-Means)

- Data Preparation: Due to differing scale and unit measurements, used Scaling function to transform data set. Separated dataset into training and testing sub-sets.

Data Sources: Quandl (Blockchain.info)

Method #2: Reducing Dimensions

- Motivation: Given 10 possible predictor variables, find the N number of most "insightful" predictors within our dataset.

- To accomplish, used Principal Component Analysis testing on all possible number of components. Choose using "elbow-method" and where majority of variance can be explained (see below figure).

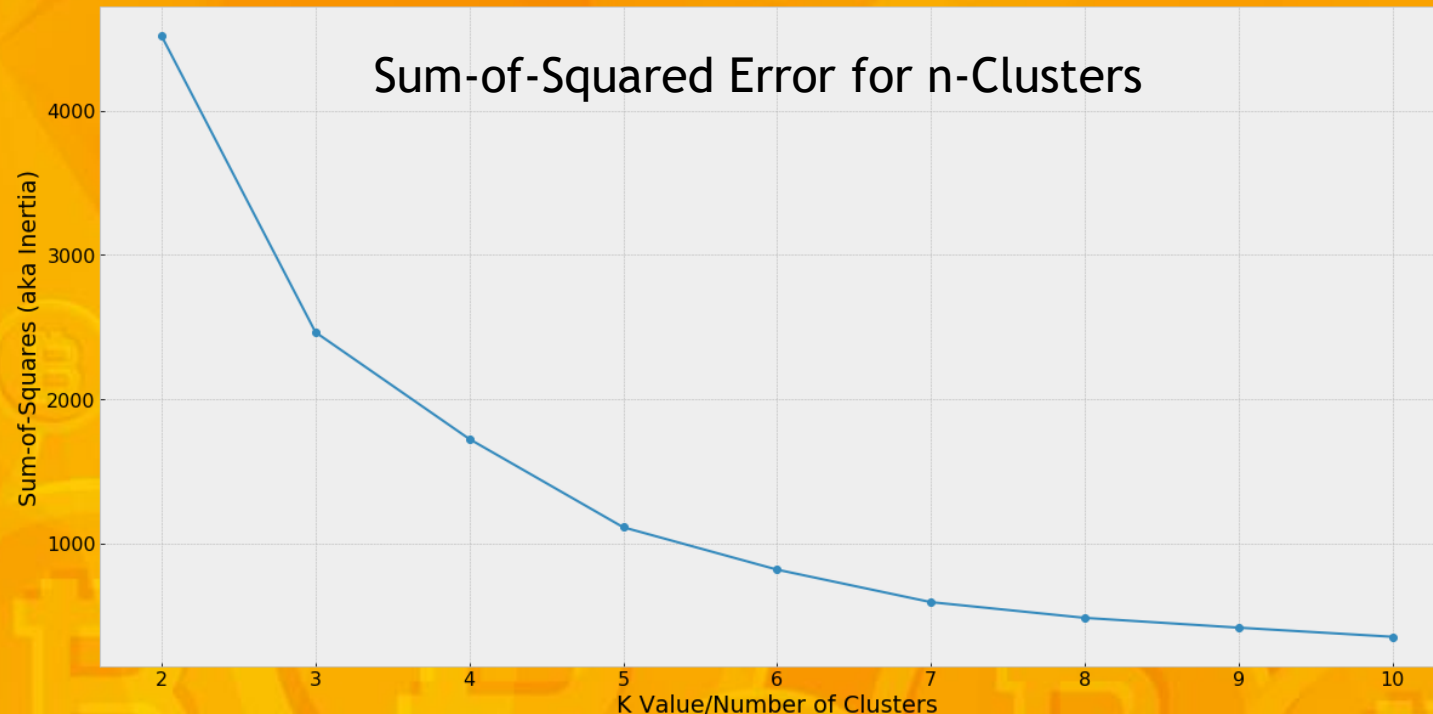Unexplained Variance for n-PCA Features
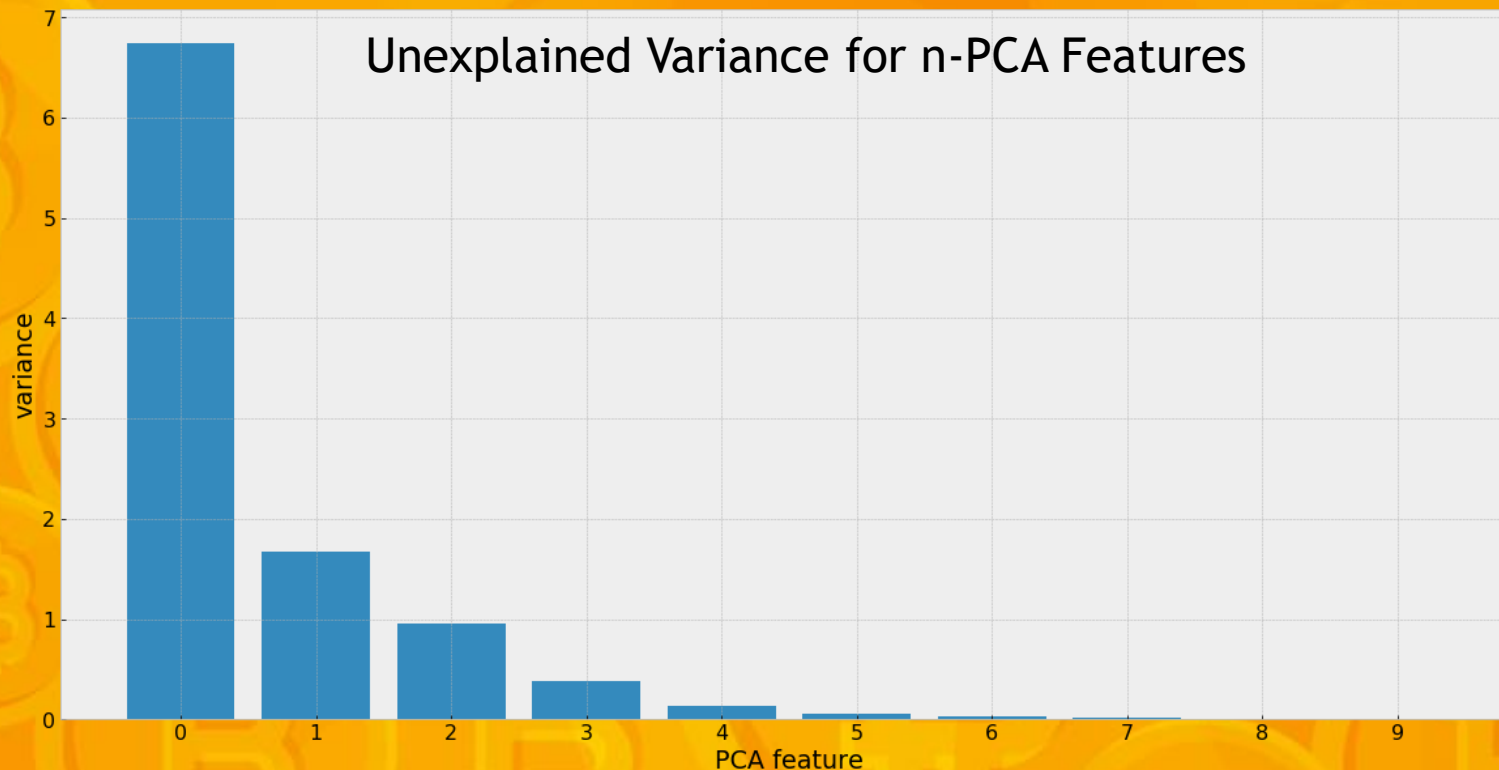
2 Principal Components explains the majority of the variance and thus, these two features will be used for our Clustering model.
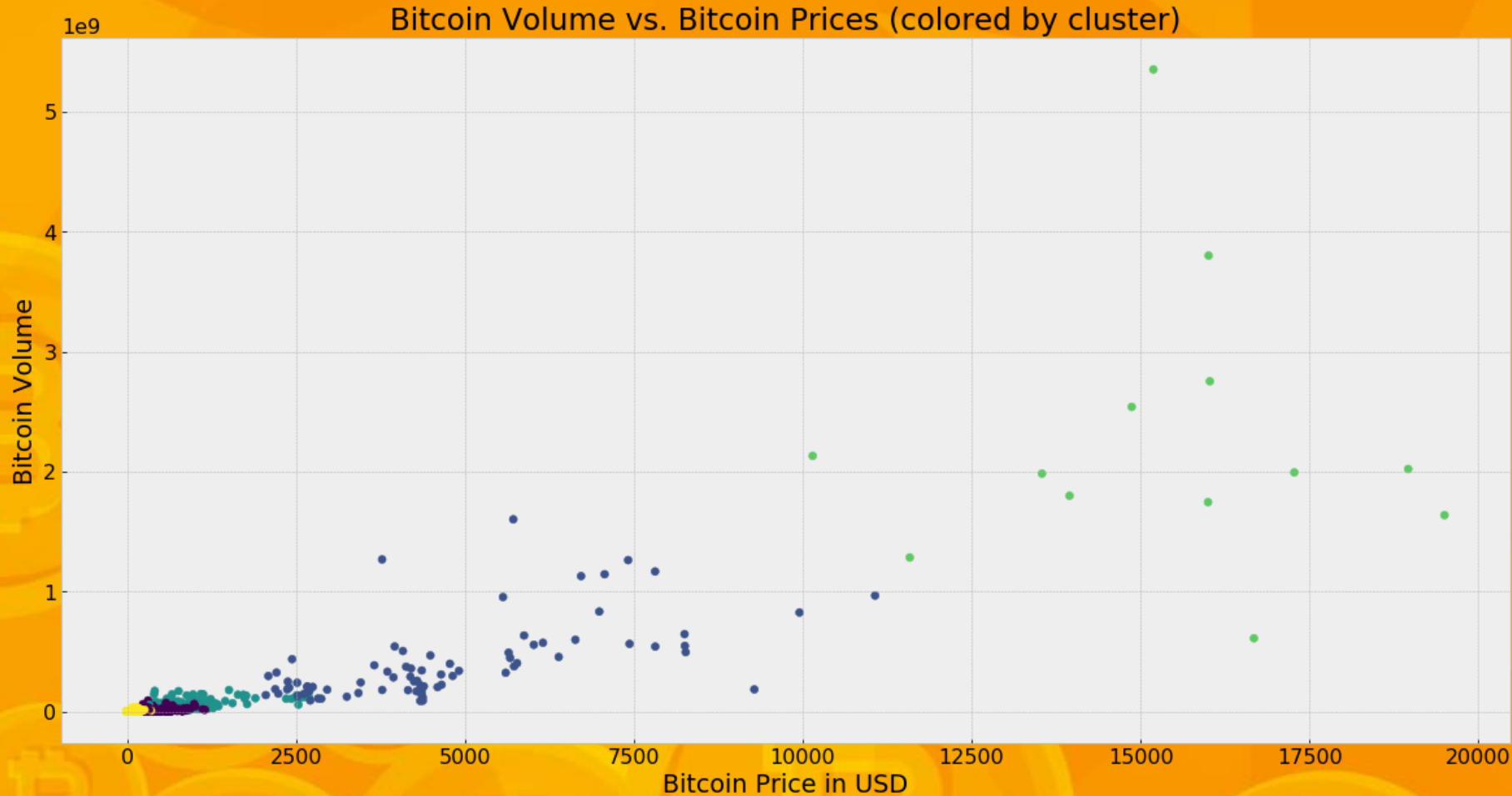
Method #2: Determining # of Clusters/Classes

- Finding optimal number of Clusters/classes. Again, choose using "elbow-method" and where majority of variance can be explained and the decreasing rate that the Sum-of-Squares begins to plateau.

- In this case, I chose 5 clusters for the model to classify the data into.

Sum-of-Squared Error for n-Clusters

Method #2: Visualizing the clusters

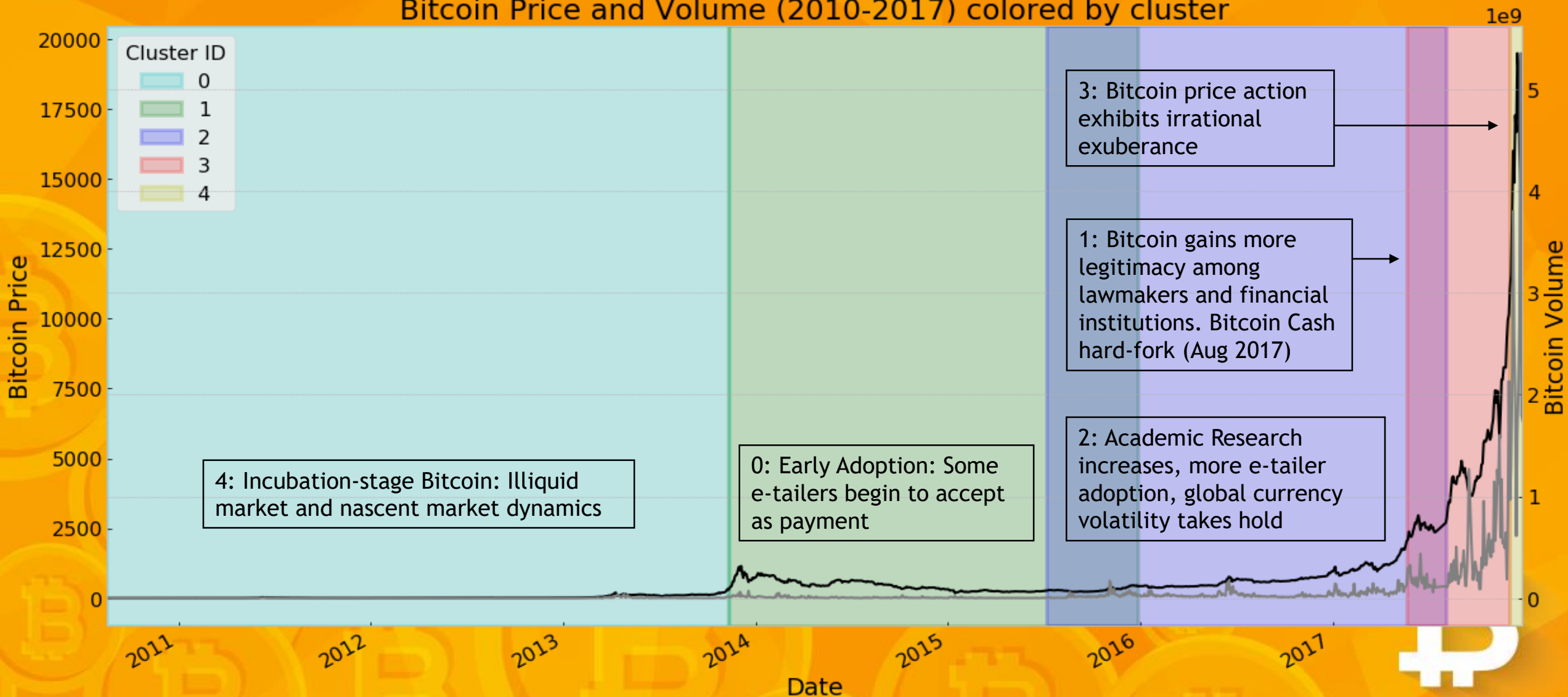Bitcoin Volume vs. Bitcoin Prices (colored by cluster)

It appears the Principal Components to classifying "eras" of Bitcoin are Price and Volume as the clusters are logically grouped.

For example, the yellow cluster classifies this stage of Bitcoin when price and volume were very low and when Price and Volume are very high, the model labels Bitcoin within the light green cluster.

Method #2: Visualizing the clusters over time

Bitcoin Price and Volume (2010-2017) colored by cluster

Cluster ID
- 0
- 1
- 2
- 3
- 4

3: Bitcoin price action exhibits irrational exuberance

1: Bitcoin gains more legitimacy among lawmakers and financial institutions. Bitcoin Cash hard-fork (Aug 2017)

4: Incubation-stage Bitcoin: Illiquid market and nascent market dynamics

0: Early Adoption: Some e-tailers begin to accept as payment

2: Academic Research increases, more e-tailer adoption, global currency volatility takes hold

Cluster Breakdown from Method #2

- Cluster 0 (Cyan)– Early Growth Stage. Market dynamics mostly uniform and nascent.
- Cluster 1 (Green) – Cryptocurrency exchanges raise funding, some hallmark retailers such as Overstock and Newegg begin accepting Bitcoin.
- Cluster 2 (Purple) – Academic research increases, Steam begins accepting Bitcoin. Chinese Renminbi depreciates against the USD, correlating with a surge in price from $600 to $780 USD
- Cluster 3 (Red) - Pre-cursor to the December 2017 rally gaining more attention among lawmakers and legacy financial institutions.
  - Similar to "rumor phase" in equity markets
- Cluster 4 (Yellow)– "Greater Fool Theory"/Irrational Exuberance Phase

Method #2: Summary

- The objective of the model is to group the data into distinct clusters differentiated by any number of characteristics. That is, data points within a given cluster are more alike to other data in the same cluster than data in another cluster.

- I did not use time as a criteria for classification and yet we see that clusters corresponds to discrete "eras" of Bitcoin's history

- As a result, our model primarily used Price and Volume as criteria for clustering. With some overlap, the model is also grouping the data in distinctive "eras" as I assumed at the onset.

  - I suspect the overlap of a few clusters is due to "profit-taking" where elevated price/volume action is from long-term holders liquidating their stakes.

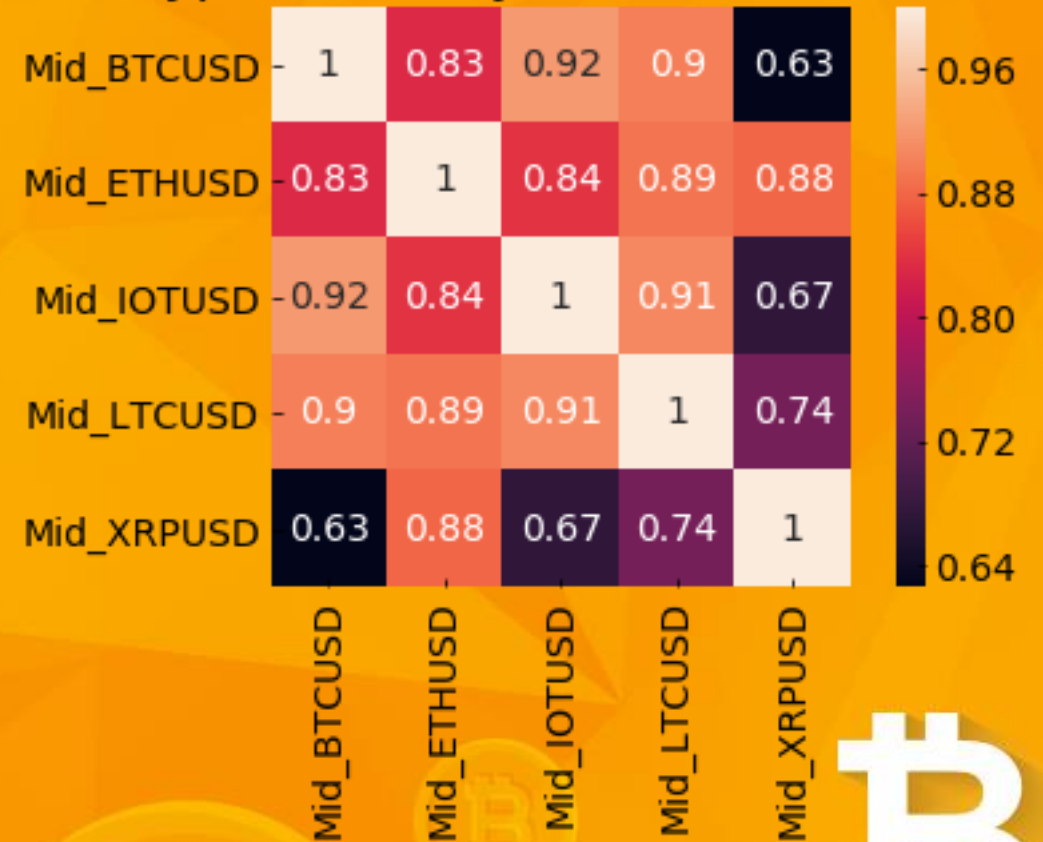# Looking ahead… Can we predict Bitcoin Prices?

- Motivations:
  - Can Bitcoin Prices be modeled in a theoretical framework like equities (akin to Capital Asset Pricing Model or Dividend Discount Model)?
  - Assuming Bitcoin is in the midst of a bubble, can we estimate or predict future Bitcoin prices?

Predicting Bitcoin: Overview

- <u>Input Data:</u> Daily Average Prices and Daily Volume for Bitcoin and next largest coins, Ethereum, Litecoin, Ripple and Iota.

- <u>Methodology:</u> Use predictors that have a strong correlation to Bitcoin Prices (see figure) and employ Linear Regression Model

- <u>Data Preparation:</u> Due to differing scale and unit measurements, used Scaling function to transform data set.

Data Sources: Quandl (Brave New Coin)
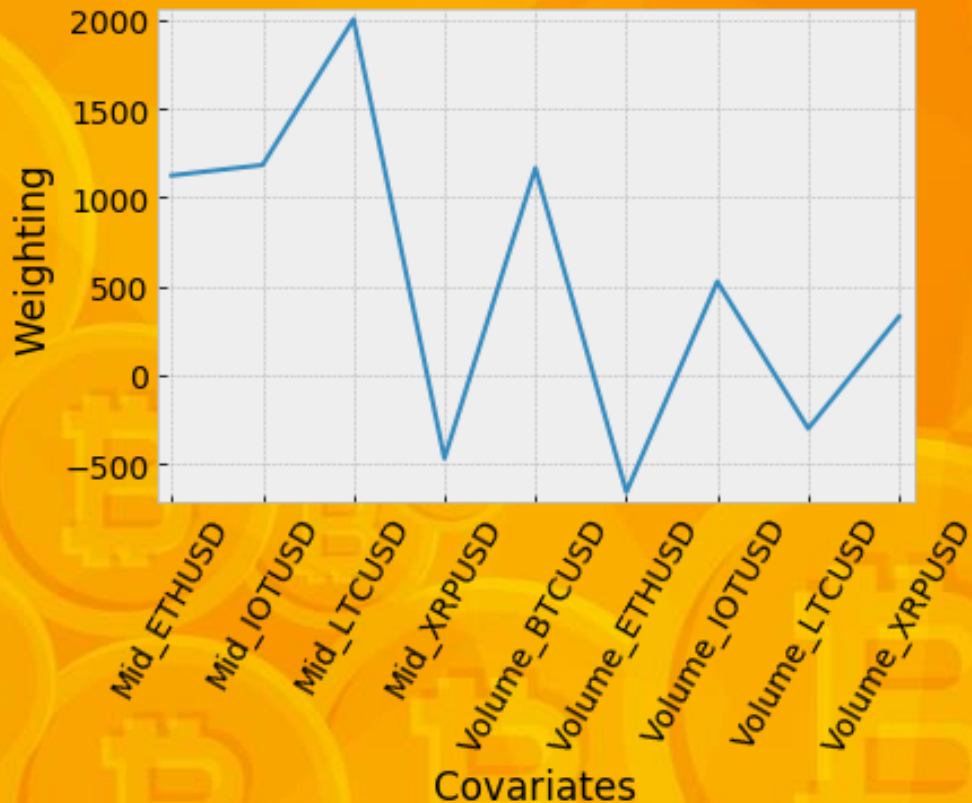

Cryptocurrency Price Correlations

	Mid_BTCUSD	Mid_ETHUSD	Mid_IOTUSD	Mid_LTCUSD	Mid_XRPUSD
Mid_BTCUSD	1	0.83	0.92	0.9	0.63
Mid_ETHUSD	0.83	1	0.84	0.89	0.88
Mid_IOTUSD	0.92	0.84	1	0.91	0.67
Mid_LTCUSD	0.9	0.89	0.91	1	0.74
Mid_XRPUSD	0.63	0.88	0.67	0.74	1

Understanding Covariates in Model

- After scaling the data and fitting the model, the optimized coefficients of the model are as follows:

We see that our substantially overweighted covariates are Prices of Ether and Iota with Volumes of Bitcoin and Ripple.

Conversely, Ripple Prices and Ether volumes are substantially underweighted.

Linear Regression Model Results

- Same-Day Accuracy score: 89.37%
  - On new/unseen data, the model predicted prices within +/-$3,500

Using lagged variables:

- Predicting BTC Prices One-Day ahead Accuracy: 70.2%

- Predicting BTC Prices One-Month ahead Accuracy: 40.2%

Note: Using ARIMA model showed no autoregressive correlation with any other periods of time (see figure on right) and thus one-day and one-month were chosen for demonstrative purposes.
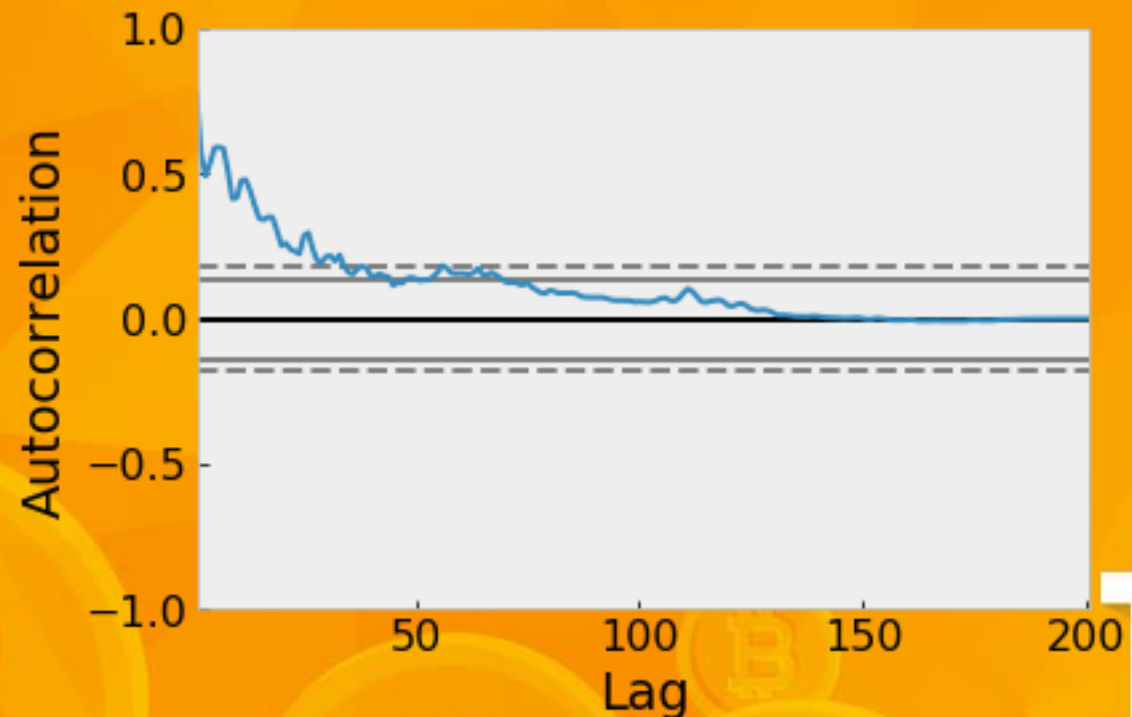
Attempt at Prediction - Takeaways

- Using highly correlated metrics and auto-regression, Bitcoin prices and volatility proves difficult to accurately estimate and predict.

Historical Bitcoin Prices and Volatility ranges, forecasted through 2019
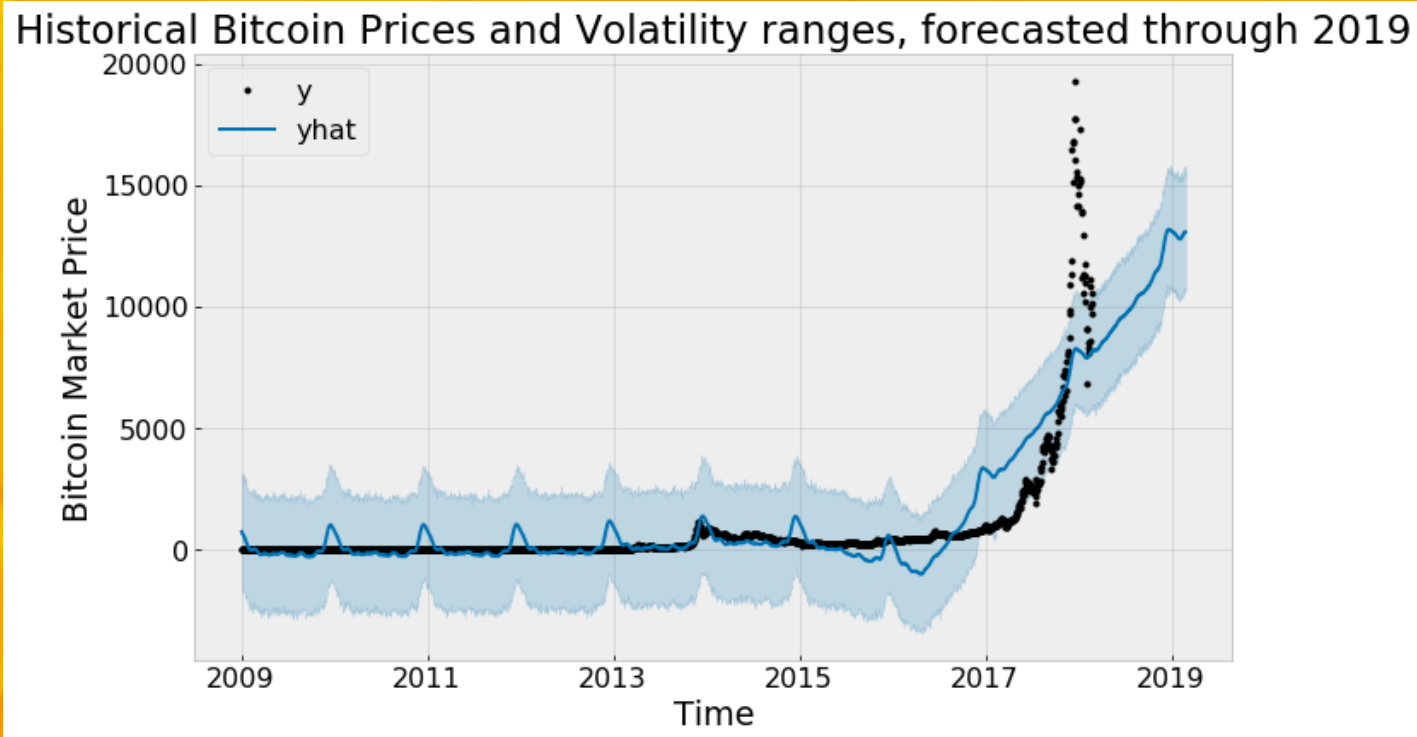
This can be best illustrated with the forecasting model from Prophet, a recently made public forecasting model from Facebook.

In this plot, we see this model has does not accurately capture the 2017 data points.

Conclusions and Outlook

- Bitcoin can be a source of diversification for a risk-tolerant portfolio containing traditional assets but may not be adequately compensating for the added risk.

- Bitcoin is likely a storage-of-value than a medium of exchange.
  - Treat Bitcoin as an investment and not as a currency or hedging tool.

- Strong evidence that Bitcoin is presently in treacherous territory, likely seeing witnessing "Greater Fool" market dynamics.

- Similar to traditional investments, Bitcoin prices and returns cannot be consistently/accurately modeled and predicted.

Limitations and Further Research

- Limitations:
  - Datetime index limitations with Pandas: unable to manipulate and further analyze the Dutch Tulip data as the years of interest (1636-1642) falls out of range.

- Further Research:
  - More research on what data can be used to improve Bitcoin Price Prediction Model (perhaps exogenous covariates like momentum of stock market, twitter mentions, etc.)
  - More data on traditional asset classes for more accurate portfolio modelling

Questions?

- Email: chu.garrick@gmail.com

- Code and Python Notebooks:

https://github.com/thegarrickchu/Springboard/tree/Capstone-Project-1

- Detailed Report: https://github.com/thegarrickchu/Springboard-Bitcoin_Digital_Gold_Rush-Capstone-1-Project-/blob/master/Springboard%20Capstone%20Project%201%20-%20Final%20Report%20(Garrick%20Chu).pdf

"our descendants doubtless will laugh at the human insanity of our Age, that in our times, the tulip flowers have been so revered"

- Theodorus Schrevelius, Dutch historian 11 years after the collapse of tulip prices.

Semper Augustus variety of Tulip. It is said that prior to the collapse, the price for a single bulb was equivalent to a "grand house" along the Amsterdam canal.