Blog Post #1

Thee Meensuk (tmeensuk), Petros Dawit (pdawit), Dikshyant Rai (drai), Kasemsan Kongsala (kkongsal)

The Data

The data we will be using for our project will be readily accessible to us. We will be pulling data for our bitcoin dataset from,

https://www.quandl.com/collections/markets/bitcoin-data

which lists out the dates, bitcoin daily prices, volume per month, bitcoin price versus many currencies, bitcoin hash rate, miners revenue, total bitcoins and market capitalization since 2009 on a daily basis. Note that in order to download large amount of data, we need to create a free account in order to use the API.

We will also be collecting data from various other assets from quandl.com when looking up other datasets such as gold, stock, other precious metal prices and other currencies, as well as the open datasets from Github, including oil prices, S&P 500 index prices and other stuff.

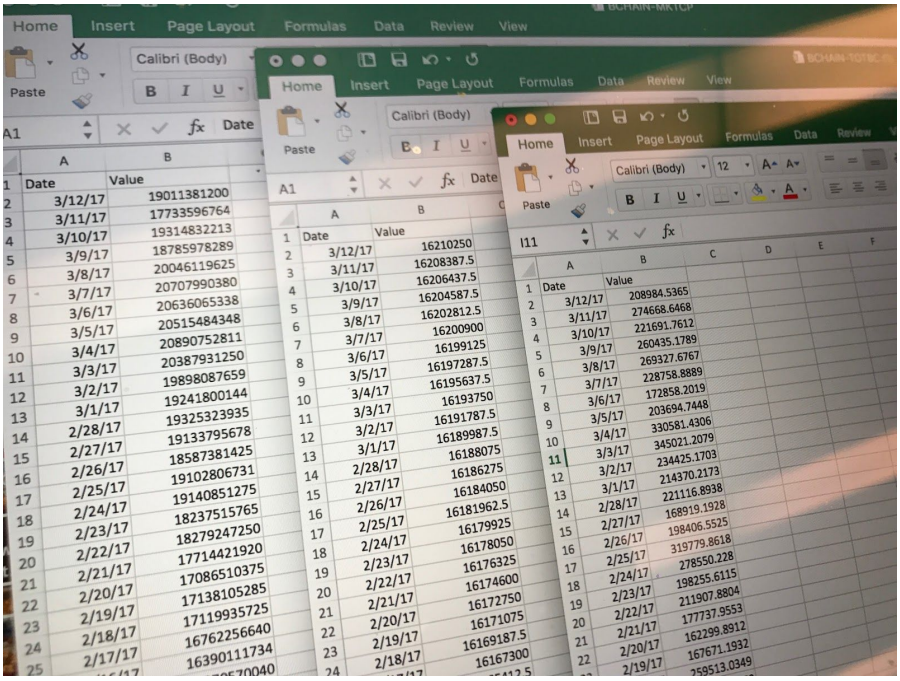

Figure 1: The CSV files we collected from quandl, including

the Market Capitalization, Total Bitcoins, and Estimated Transaction Volume.

Bitcoin Data:

1. Bitcoin Prices in USD (Daily)
2. Trading Volume from xxx exchange (Daily)
3. 1-day BTC price change (Daily)
4. 7-day BTC price change (Daily)
5. 30-day standard deviation on price (Daily)
6. Number of bitcoins mined on that day (Daily)

We also collect other assets data from datasets hosted on Github. This page provides the index for many useful datasets:

https://github.com/datasets

Other Assets Data:

7. Euro/dollar exchange rate (Daily)
8. GB Pound/dollar exchange rate (Daily)
9. Chinese Yuan/dollar exchange rate (Daily)
10. Japanese Yen/dollar exchange rate (Daily)
11. Gold price in USD (Daily)
12. Oil price in USD (Daily)
13. S&P500 stock index (daily)

The Database

For collecting the data, we will be mainly extracting them from Quandl and Github datasets and merge them on attributes. Many of the datasets already are in the csv or json format. For our bitcoin data, we will merge them on dates, creating one table containing all the Bitcoin related data. For our other assets data, if they are not taken from Quandl, we will have to clean them first and format them using the same primary keys and similar formats. Then, we merge them and create tables for them separately i.e: have separate table for bitcoin, gold, oil etc…

From here, we will create our database using the data we cleaned. Then, we will have a complete database where we can necessary information to put in our formulas and create models. We will do some calculations required (the 7-day average prices, for example.) Also, we can illustratrate the data as well as the models using D3 or other visualization tools, which might help us determine which model works best.

The Models

We use many models to predict the price of Bitcoin. We may combine some of the results that are related, or we will report them separately. The reason behind this decision is that different models render different aspects of the environment. If many models agree on the same trend, we can quite confidently concludes that the price is going in that direction. However, if many models disagree, we might need to look at a different period which our models work.

For this blog post, we will just outline the models we intend to use. After we finish cleaning the data and constructing the database, we will apply these approaches using various equations. For example, we might use linear, polynomial, or exponential regressions to fit one of the models outlined below, and see which one makes the most sense.

1. Computational Power Approach

We think that the expected price of Bitcoin is the function of computation power and the market capitalization. In this case, we decide to use the $hashRate$ as the representation of computational power because it is a measurable, direct result of the computational power, and it is directly used to produced Bitcoin. For the $marketCap$, here, for simplicity, we use the market capitalization of Bitcoin against the USD. Thus, we have,

$$E(price) = f(hashRate, marketCap)$$

After we have the cleaned data, we will try to fit the models.

2. Trading Trend Approach

One of the most basic approaches is to predict the future price from historical prices. We will use those historical prices and trading volumes to form a range of indicators, and then use those indicators to do a regression analysis to find the relationship between the future price and those indicators. Specifically, we are looking at something like:

$$E(price) = f(p1, p7, p30, sd7, sd30, av7, av30, v7, v30)$$

where:

$$p1 = 1-day\ BTC\ Price\ Change$$
$$p7 = 7-day\ BTC\ Price\ Change$$
$$sd7 = 7-day\ Price's\ Standard\ Deviation$$
$$v7 = 7-day\ Average\ Trading\ Volume$$

$$av7 = 7 - day\ Moving\ BTC\ PriceAverage$$
$$p30 = 30 - day\ BTC\ Price\ Change$$
$$sd30 = 30 - day\ Price's\ Standard\ Deviation$$
$$v30 = 30 - day\ Average\ Trading\ Volume$$
$$v30 = 30 - day\ Moving\ BTC\ PriceAverage$$

Then we will plot the percentage increase over time and apply regression to predict what the trend of the percentage increase will be. In this way we can use the trend of percentage increase to predict the expected value of bitcoin in the future

Macroeconomics Approach

We will use the macroeconomic indicators that are likely to correlate with the prices of Bitcoin. First we will find the list of countries where the bitcoin trading volume is maximum. Then we use their macroeconomic indicators to see what is its impact on the bitcoin. Apply regression on the changes in those indicators such as the EUR-USD exchange rate, GBP-USD exchange rate, CNY-USD exchange rate, JPY-USD exchange rate, gold price, oil price, and the S&P500 stock market index.

Black Market Approach

We will target the specific countries that use Bitcoin heavily in their dark net markets, such as India, Mexico, and China. We can find some indicators of the dark net activities and find their relationships with the bitcoin price.