My Interest in Bitcoin
Sandhya

Cryptocurrency especially Bitcoin is a new buzzword in media. Everyone except those who might be in the woods away from social network, is talking about it these days. This might be because the surge in price or might be to maintain the social status. The reason is not apparent yet. Some time ago, my friend was a bit concerned and asked me, 'Hey, I have bought 10 bitcoins but I am not sure how that works and what is going to happen with its price? Should I sell since price is going low or keep it considering it might go up.' I was little surprised at first because he bought the coins while having limited knowledge about it but I was unsure as well as I did not have any answer to it. From that time on, I have been reading about bitcoin and a process through which it seems to be working. It seems interesting and worth exploring.

I plan to work on bitcoin data as my final project for Data Science class. The idea of bitcoin is new to me. I do not claim to know everything about underlying process but I plan to learn about it more as part of the project. Luckily enough, I got multiple datasets on Kaggle.com about historic data on bitcoin and cryptocurrency in general. So, I am planning of analyzing that data as my final project for this class and will try to answer scientific questions that I have.

There are two datasets that I want to combine and work with.

- A dataset that contains the data regarding price of bitcoin per day from 2013 to 2017
  It contains various columns like high price, low price, closing price per each row.

Date	Open	High	Low	Close	Volume	Market Cap
Sep 30, 2013	137.15	138.35	130.27	133	-	1614350000
Sep 29, 2013	134.9	140.61	134.71	137.34	-	1587290000
Sep 28, 2013	133.77	135.63	132.56	134.78	-	1573400000
Sep 27, 2013	128.94	134.74	128.49	133.78	-	1516020000
Sep 26, 2013	128.21	134.93	127.32	128.38	-	1506920000
Sep 25, 2013	127.38	129.69	127.33	128.22	-	1496510000
Sep 24, 2013	126.05	127.46	125.39	127.25	-	1480210000
Sep 23, 2013	128.98	132.72	125.66	125.95	-	1513940000
Sep 22, 2013	127.87	133.94	126.59	129.12	-	1500260000
Sep 21, 2013	126.95	128.61	126.24	127.43	-	1488870000
Sep 20, 2013	129.7	135.62	126.68	127.04	-	1520470000
Sep 19, 2013	131.37	131.77	128.45	129.65	-	1539430000

- A detailed dataset about bitcoin containing bitcoin block size, average block size, market capital, hash rate, estimated transaction volume, cost per transaction etc. This dataset is from 2009 to 2017, relatively larger dataset than the first one.

Date	btc_market_price	btc_total_bitcoins	btc_market_cap	btc_trade_volume	btc_blocks_size	btc_avg_block_size	btc_n_orphane
2009-11-10 00:00:00	0	1339450	0	0	0	0.000215422535211	
2009-11-11 00:00:00	0	1342900	0	0	0	0.000323072463768	
2009-11-12 00:00:00	0	1346400	0	0	0	0.000215357142857	
2009-11-13 00:00:00	0	1349900	0	0	0	0.000241728571429	
2009-11-14 00:00:00	0	1354050	0	0	0	0.000215638554217	
2009-11-15 00:00:00	0	1358850	0	0	0	0.000215583333333	
2009-11-16 00:00:00	0	1363850	0	0	0	0.00021569	
2009-11-17 00:00:00	0	1368400	0	0	0	0.000215681318681	
2009-11-18 00:00:00	0	1372350	0	0	0	0.000215544303797	
2009-11-19 00:00:00	0	1376000	0	0	0	0.000215342465753	
2009-11-20 00:00:00	0	1380100	0	0	0	0.000351951219512	
2009-11-21 00:00:00	0	1383950	0	0	0	0.000215714285714	
2009-11-22 00:00:00	0	1388950	0	0	0	0.00021569	
2009-11-23 00:00:00	0	1393750	0	0	0	0.000215875	
2009-11-24 00:00:00	0	1398850	0	0	0	0.000215607843137	

Some of the questions that I want answer for from the data are:

- What makes people interested in buying bitcoin?
- what are the factors that lead to increase/decrease in price?
- can we predict the future price looking at the historic data?
- Is there a seasonal trend in price fluctuation?
- Are there social, economic factors affecting the price? (this might require another dataset)
- Change in price overtime?
- How many cryptocurrencies? (this will require additional data outside the two files)
- Is there relation between hourly price fluctuations?
- Can we get useful trends by looking at data by day-of-week?

Challenges I might face:

- Since the format of files is different, some preprocessing will be required to combine these two files.
- There are about 34 features combined in both datasets. It will require some time to understand each feature to know its importance in overall picture.
- There are missing values in the data. Techniques to impute data with meaningful value might be time consuming.
- Since the topic is fairly new to me, it will be required of me to do some research, read literature review and articles about crypto currency and how their price is being affected.
- Effective data visualization techniques.