

Bitcoin Analysis

Raw Data

- 4 CSV files:
 - btc_usd: raw data from bitcoin over time by minute from 2012 to 2017
 - btc_hacks: data related to reported hacks
 - S&P500 and NASDAQ: files with data over time for those indicators

Data importing from csv to PostgreSQL

- Data imported from flat file to a DB.
 - Helps accessing and save data transformation intermedium steps
 - This is done for all 4 CSVs each to a table.


```
1 Timestamp,Open,High,Low,Close,Volume_(BTC),Volume_(Currency),Weighted_Price
2 1325317920,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
3 1325317980,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
4 1325318040,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
5 1325318100,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
6 1325318160,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
7 1325318220,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
8 1325318280,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
9 1325318340,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
10 1325318400,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
11 1325318460,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
12 1325318520,4.39,4.39,4.39,4.39,0.45558087,2.0000000193,4.39
```

	123 timestamp_value	123 open	123 high	123 low	123 close	123 volume_btc	123 volume_currency	123 weighted_price
1	1,325,323,200	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
2	1,325,323,260	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
3	1,325,323,320	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
4	1,325,323,380	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
5	1,325,323,440	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
6	1,325,323,500	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
7	1,325,323,560	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
8	1,325,323,620	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
9	1,325,323,680	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
10	1,325,323,740	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900

btc_usd: Preprocessing

- Convert timestamp to datetime
- Resample data from by min to by day:
 - Low = min(low)
 - High = max(high)
 - Volume = sum(volume)
 - Value = mean(value)
 - Open = first(open)
 - Close = last(close)
- Created some statistics about the day:
 - Median, mode, standard deviation, kurtosis and skewness

	123 timestamp_value ∇	123 open ∇	123 high ∇	123 low ∇	123 close ∇	123 volume_btc ∇	123 volume_currency ∇	123 weighted_price ∇
1	1,325,323,200	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
2	1,325,323,260	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
3	1,325,323,320	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
4	1,325,323,380	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
5	1,325,323,440	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
6	1,325,323,500	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
7	1,325,323,560	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
8	1,325,323,620	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
9	1,325,323,680	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900
10	1,325,323,740	4.3900	4.3900	4.3900	4.3900	0.45558087	2.0000000193	4.3900

Column Name	#	Data type
 date_	1	<u>timestamp</u>
123 high	2	<u>float8</u>
123 low	3	<u>float8</u>
123 volume_btc	4	<u>float8</u>
123 volume_currency	5	<u>float8</u>
123 weighted_price	6	<u>float8</u>
123 std_weighted_price	7	<u>float8</u>
123 open	8	<u>float8</u>
123 close	9	<u>float8</u>
123 median_weighted_price	10	<u>float8</u>
123 mode_weighted_price	11	<u>float8</u>
123 kurtosis_weighted_price	12	<u>float8</u>
123 skewness_weighted_price	13	<u>float8</u>

btc_usd: Processing

- Create some ratio variables:
 - A ratio between the high and low in the day
 - A ratio between the open and close value
 - A ratio between today and yesterday mean value

Column Name	#	Data type
date_	12	timestamp
perc_close_open	1	float8
open	2	float8
close	3	float8
perc_high_low	4	float8
high	5	float8
low	6	float8
perc_weighted_price	7	float8
weighted_price	8	float8
std_weighted_price	9	float8
volume_btc	10	float8
volume_currency	11	float8
median_weighted_price	17	float8
mode_weighted_price	18	float8
kurtosis_weighted_price	19	float8
skewness_weighted_price	20	float8

NASDAQ and sp500: Preprocessing

- Clean and convert some variables:

- Vol
- Change_percentage

- Resample daily to be able to fill weekends

- Created some ratios:

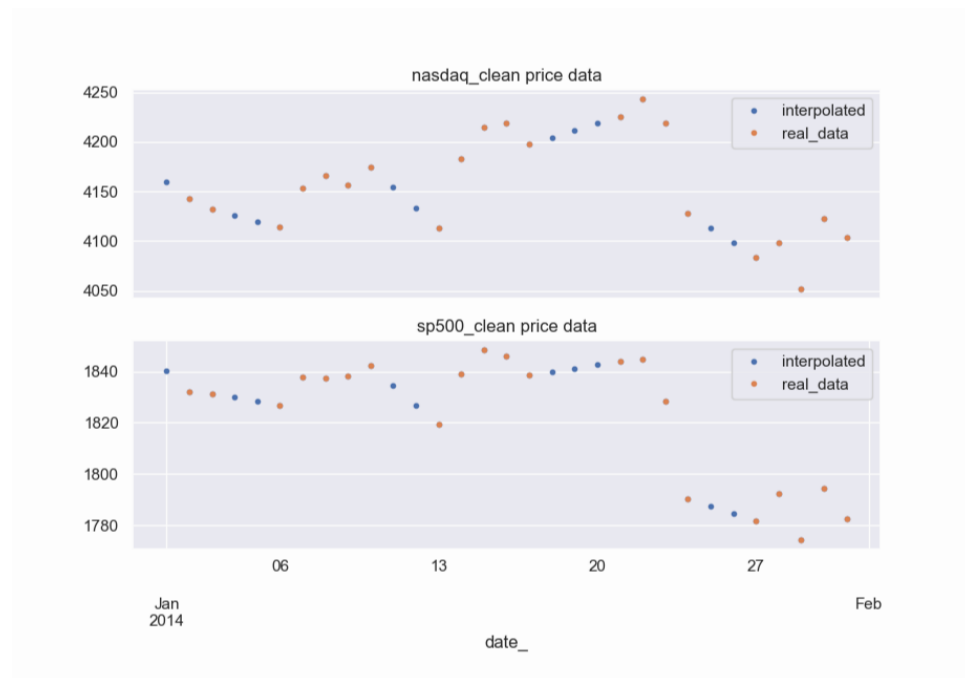
- A ratio between the high and low in the day
- A ratio between today and yesterday mean value

	🕒 date_ 📉	123 price 📉	123 open 📉	123 high 📉	123 low 📉	abc vol 📉	abc change_percentage 📉
1	2018-07-24 00:00:00	7,840.7700	7,914.1300	7,928.6500	7,814.3300	469.84M	-0.01%
2	2018-07-23 00:00:00	7,841.8700	7,806.9300	7,846.7600	7,776.5500	395.31M	0.28%
3	2018-07-20 00:00:00	7,820.2000	7,843.0800	7,860.2500	7,815.8600	464.63M	-0.07%
4	2018-07-19 00:00:00	7,825.3000	7,829.8000	7,849.8500	7,811.1500	434.39M	-0.37%
5	2018-07-18 00:00:00	7,854.4400	7,859.4300	7,863.7700	7,822.8300	437.89M	-0.01%
6	2018-07-17 00:00:00	7,855.1200	7,751.9800	7,867.1500	7,749.6000	406.40M	0.63%
7	2018-07-16 00:00:00	7,805.7200	7,831.7400	7,838.8200	7,791.9800	429.29M	-0.26%
8	2018-07-13 00:00:00	7,825.9800	7,827.6200	7,843.5200	7,803.3400	413.13M	0.03%
9	2018-07-12 00:00:00	7,823.9200	7,752.7100	7,825.6700	7,746.7600	453.33M	1.39%
10	2018-07-11 00:00:00	7,716.6100	7,698.5100	7,748.1600	7,696.5800	420.72M	-0.55%
11	2018-07-10 00:00:00	7,759.2000	7,770.6300	7,777.4800	7,731.9800	419.19M	0.04%
12	2018-07-09 00:00:00	7,756.2000	7,731.7400	7,757.2800	7,702.0600	453.76M	0.88%
13	2018-07-06 00:00:00	7,688.3900	7,595.9300	7,695.8100	7,588.6500	431.50M	1.34%

	🕒 date_ 📉	123 price 📉	123 open 📉	123 high 📉	123 low 📉	123 volume_currency 📉	123 perc_price 📉
21	2009-07-14 00:00:00	1,799.7300	1,790.7000	1,800.7300	1,782.8500	583,380,000.0000	0.0036
22	2009-07-15 00:00:00	1,862.9000	1,827.6100	1,863.2900	1,824.3100	788,230,000.0000	0.0351
23	2009-07-16 00:00:00	1,885.0300	1,856.0500	1,887.8900	1,854.7000	657,840,000.0000	0.0119
24	2009-07-17 00:00:00	1,886.6100	1,880.3100	1,887.3400	1,873.8100	630,410,000.0000	0.0008
25	2009-07-18 00:00:00	1,929.2600	1,912.9850	1,930.9200	1,909.7350	667,640,000.0000	0.0221069218
26	2009-07-19 00:00:00	1,957.5370	1,942.7120	1,959.3630	1,940.0620	668,710,000.0000	0.0144451931
27	2009-07-20 00:00:00	1,909.2900	1,896.9900	1,909.8900	1,890.0000	618,050,000.0000	0.0120
28	2009-07-21 00:00:00	1,916.2000	1,917.3900	1,917.4600	1,892.1700	650,100,000.0000	0.0036
29	2009-07-22 00:00:00	1,926.3800	1,908.8300	1,934.6100	1,907.2500	676,120,000.0000	0.0053
30	2009-07-23 00:00:00	1,973.6000	1,924.9600	1,979.3400	1,924.7700	869,630,000.0000	0.0245
31	2009-07-24 00:00:00	1,965.9600	1,943.2300	1,966.2600	1,937.6400	644,280,000.0000	-0.0039

NASDAQ and sp500: weekends

- 1st approach: interpolation
 - Suffers from data leakage because we need next Monday
- 2nd approach: linear regression using last 5 days



btc_hacks: Preprocessing


- Cleaning loss_usd
- Resample:
 - Create a variable to count days since the last hack
 - Create a variable which is the value of the last hack divided by the days without hacks

	name	loss_crypto	loss_usd	date_	source
1	allinva wallet theft	25,000 BTC	\$500,000	2011-06-13 00:00:00	https://bitcointalk.org/index.p
2	MyBitCoin Theft	154,406 BTC	\$2,000,000	2011-08-01 00:00:00	https://observer.com/2011/08/
3	Bitcoin7	11,000 BTC	\$50,000	2011-10-05 00:00:00	https://bitcointalk.org/index.p
4	Linode Hack	46,700 BTC	\$228,000	2012-03-03 00:00:00	https://arstechnica.com/inform
5	Bitcoinica 2	18,000 BTC	\$91,000	2012-05-21 00:00:00	https://news.bitcoin.com/bitcc
6	Bitcoinica 3	40,000 BTC	\$305,200	2012-07-13 00:00:00	https://bitcointalk.org/index.p
7	Bitcoin Savings & Trust	265,000 BTC	\$2,800,000	2012-08-27 00:00:00	https://www.theverge.com/2012/08/27/1214285714
8	BitFloor Hack	24,000 BTC	\$250,000	2012-09-04 00:00:00	https://www.coindesk.com/bit
9	Bitmarket.eu (related to bitcoinica hack in 2012)	20,000 BTC	\$260,000	2012-12-21 00:00:00	https://www.newsbtc.com/2012/12/21/1214285714
10	Virucurex	1,454 BTC	\$160,000	2013-05-10 00:00:00	https://www.coindesk.com/for
11	Inputs.io Hack	4,100 BTC	\$1,200,000	2013-10-26 00:00:00	https://www.theguardian.com/

	date_	loss_usd	lost_by_days_without_hacks	days_without_hacks
1	2011-06-13 00:00:00	500,000.0000	500,000.0000	0.0000
2	2011-06-14 00:00:00	[NULL]	250,000.0000	1.0000
3	2011-06-15 00:00:00	[NULL]	166,666.6666666667	2.0000
4	2011-06-16 00:00:00	[NULL]	125,000.0000	3.0000
5	2011-06-17 00:00:00	[NULL]	100,000.0000	4.0000
6	2011-06-18 00:00:00	[NULL]	83,333.3333333333	5.0000
7	2011-06-19 00:00:00	[NULL]	71,428.5714285714	6.0000
8	2011-06-20 00:00:00	[NULL]	62,500.0000	7.0000
9	2011-06-21 00:00:00	[NULL]	55,555.5555555556	8.0000
10	2011-06-22 00:00:00	[NULL]	50,000.0000	9.0000
11	2011-06-23 00:00:00	[NULL]	45,454.5454545455	10.0000
12	2011-06-24 00:00:00	[NULL]	41,666.6666666667	11.0000
13	2011-06-25 00:00:00	[NULL]	38,461.5384615385	12.0000

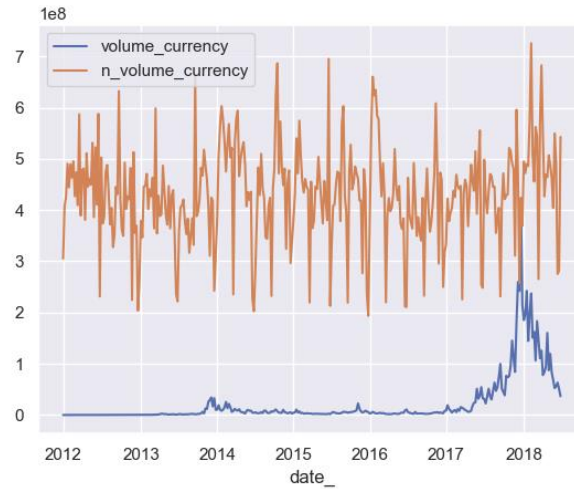
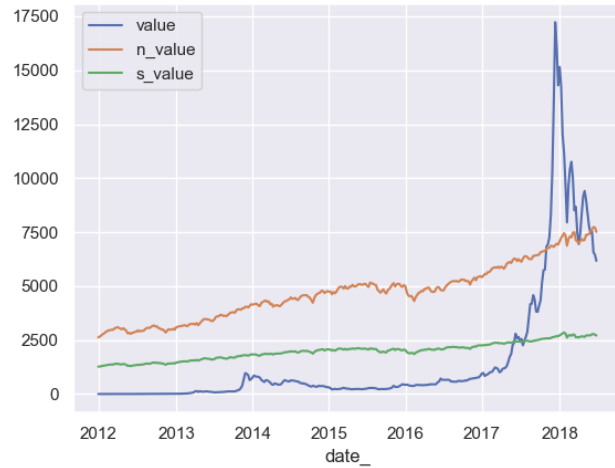
Data fusion in one table

- Data is then joined in one table using the date as a global key.
- Adding prefixes to avoid columns with the same name:
 - b_: btc_usd and btc_hacks
 - s_: sp500
 - n_: nasdaq

Column Name	#	Data type
¹²³ b_perc_close_open	1	float8
¹²³ b_open	2	float8
¹²³ b_close	3	float8
¹²³ b_perc_high_low	4	float8
¹²³ b_high	5	float8
¹²³ b_low	6	float8
¹²³ b_perc_value	7	float8
¹²³ b_value	8	float8
¹²³ b_std_value	9	float8
¹²³ b_volume_btc	10	float8
¹²³ b_volume_currency	11	float8
 date_	12	timestamp
¹²³ b_median_value	13	float8
¹²³ b_mode_value	14	float8
¹²³ b_kurtosis_value	15	float8
¹²³ b_skewness_value	16	float8
¹²³ n_open	17	float8
¹²³ n_perc_high_low	18	float8
¹²³ n_high	19	float8
¹²³ n_low	20	float8
¹²³ n_perc_value	21	float8
¹²³ n_value	22	float8
¹²³ n_volume_currency	23	float8
¹²³ s_open	24	float8
¹²³ s_perc_high_low	25	float8
¹²³ s_high	26	float8
¹²³ s_low	27	float8
¹²³ s_perc_value	28	float8
¹²³ s_value	29	float8
¹²³ b_loss_usd	30	float8
¹²³ b_lost_by_days_without_hacks	31	float8
¹²³ b_days_without_hacks	32	float8

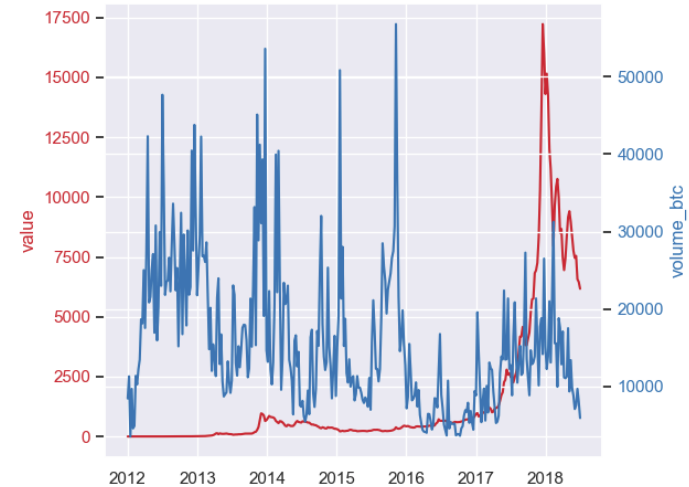
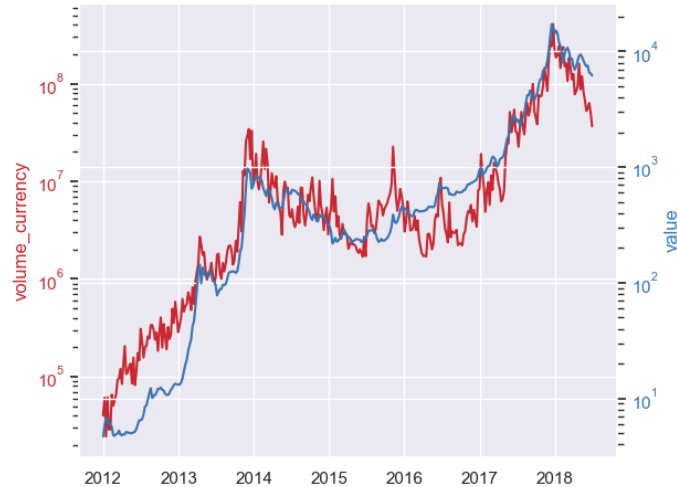
Dataset analysis

- Bitcoin value vs stock value
- Bitcoin volume vs Nasdaq Volume



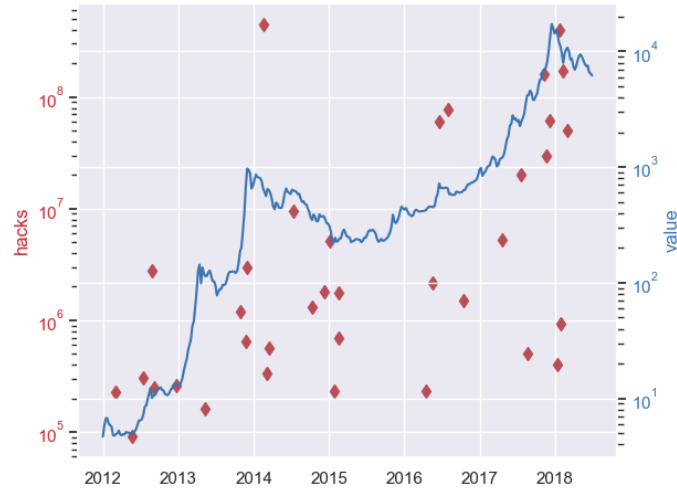
Bitcoin value vs volume

- Vs volume in currency
- Vs volume in btc units

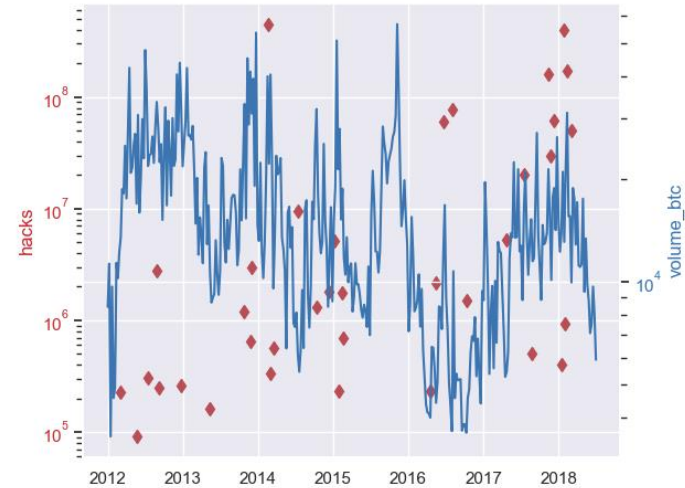


Bitcoin vs Hacks

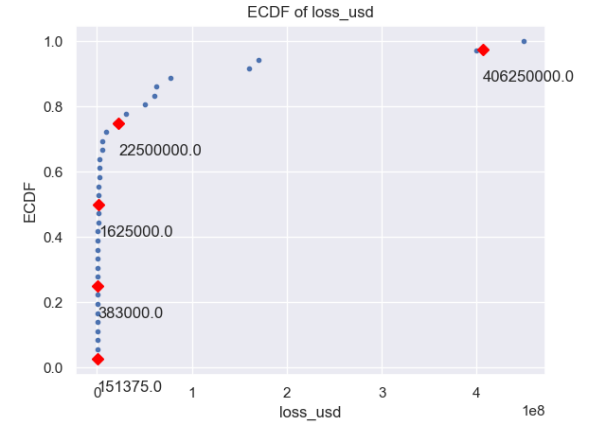
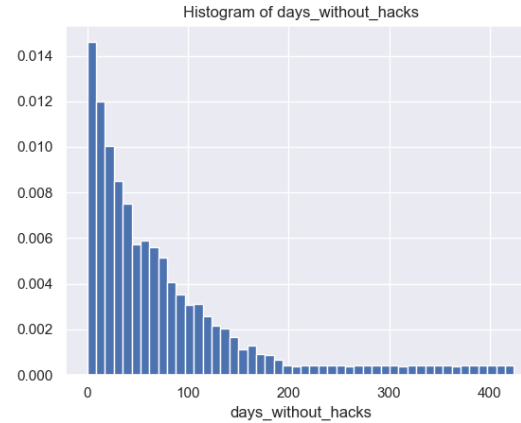
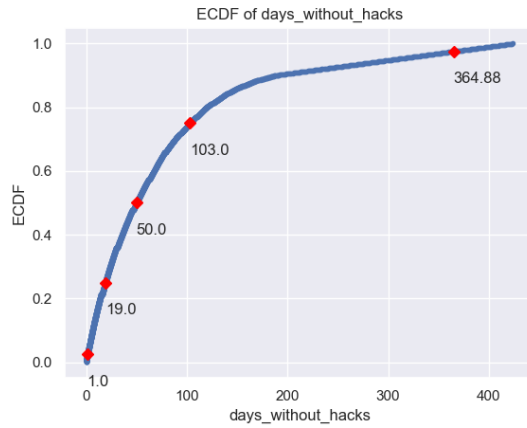
- Value vs Hacks



- Volume in btc vs hacks



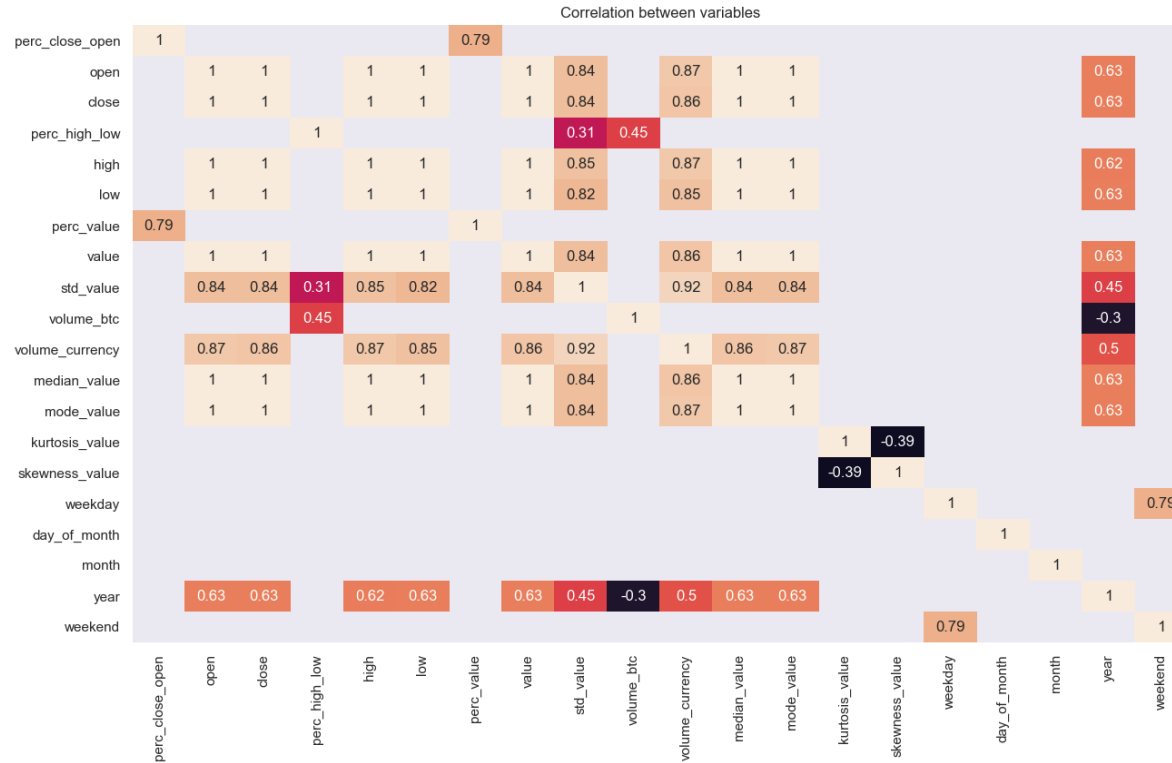
Bitcoin vs Hacks



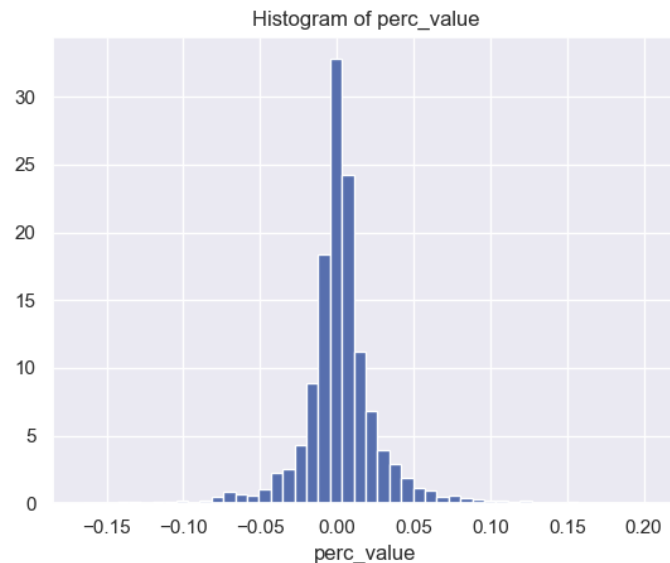
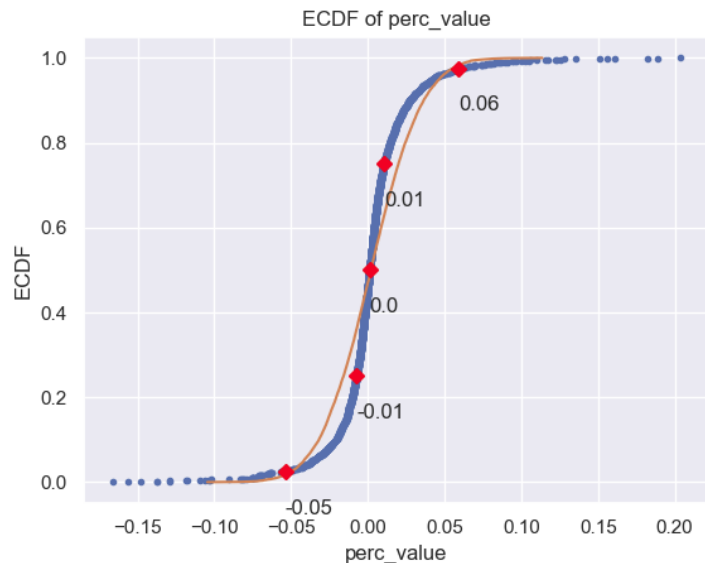
Correlation

Correlation between variables																				
perc_close_open	1	-0.03	0	-0.22	-0.01	-0.01	0.79	-0.01	-0.04	-0.09	-0.05	-0.01	-0.01	-0.04	0.13	-0.04	0.01	0.03	-0.01	-0.03
open	-0.03	1	1	0.16	1	1	-0.06	1	0.84	-0.05	0.87	1	1	-0.07	0.05	-0	-0	-0.02	0.63	0
close	0	1	1	0.15	1	1	-0.03	1	0.84	-0.06	0.86	1	1	-0.07	0.05	-0	-0	-0.01	0.63	0
perc_high_low	-0.22	0.16	0.15	1	0.16	0.14	-0.02	0.15	0.31	0.45	0.29	0.15	0.16	0.02	-0.04	-0.04	-0.03	-0.07	-0.03	-0.08
high	-0.01	1	1	0.16	1	1	-0.04	1	0.85	-0.05	0.87	1	1	-0.07	0.05	-0	-0	-0.02	0.62	-0
low	-0.01	1	1	0.14	1	1	-0.03	1	0.82	-0.07	0.85	1	1	-0.07	0.05	-0	-0	-0.02	0.63	0
perc_value	0.79	-0.06	-0.03	-0.02	-0.04	-0.03	1	-0.04	-0.07	-0.05	-0.09	-0.04	-0.04	-0.02	-0.03	-0.07	0	0.01	-0.04	-0.06
value	-0.01	1	1	0.15	1	1	-0.04	1	0.84	-0.06	0.86	1	1	-0.07	0.05	-0	-0	-0.02	0.63	0
std_value	-0.04	0.84	0.84	0.31	0.85	0.82	-0.07	0.84	1	0.09	0.92	0.84	0.84	-0.07	0.03	-0	-0.01	-0	0.45	-0.02
volume_btc	-0.09	-0.05	-0.06	0.45	-0.05	-0.07	-0.05	-0.06	0.09	1	0.14	-0.06	-0.06	-0.01	-0.02	-0.11	-0.04	0.11	-0.3	-0.16
volume_currency	-0.05	0.87	0.86	0.29	0.87	0.85	-0.09	0.86	0.92	0.14	1	0.86	0.87	-0.06	0.03	-0.03	-0.01	0.01	0.5	-0.06
median_value	-0.01	1	1	0.15	1	1	-0.04	1	0.84	-0.06	0.86	1	1	-0.07	0.05	-0	-0	-0.02	0.63	-0
mode_value	-0.01	1	1	0.16	1	1	-0.04	1	0.84	-0.06	0.87	1	1	-0.07	0.05	-0	-0	-0.02	0.63	-0
kurtosis_value	-0.04	-0.07	-0.07	0.02	-0.07	-0.07	-0.02	-0.07	-0.07	-0.01	-0.06	-0.07	-0.07	1	-0.39	-0.02	0.01	-0.04	-0.1	-0.03
skewness_value	0.13	0.05	0.05	-0.04	0.05	0.05	-0.03	0.05	0.03	-0.02	0.03	0.05	0.05	-0.39	1	0.03	-0.01	0.04	0.11	0.04
weekday	-0.04	-0	-0	-0.04	-0	-0	-0.07	-0	-0	-0.11	-0.03	-0	-0	-0.02	0.03	1	-0	0	-0	0.79
day_of_month	0.01	-0	-0	-0.03	-0	-0	0	-0	-0.01	-0.04	-0.01	-0	-0	0.01	-0.01	-0	1	0.01	-0.01	-0
month	0.03	-0.02	-0.01	-0.07	-0.02	-0.02	0.01	-0.02	-0	0.11	0.01	-0.02	-0.02	-0.04	0.04	0	0.01	1	-0.11	0
year	-0.01	0.63	0.63	-0.03	0.62	0.63	-0.04	0.63	0.45	-0.3	0.5	0.63	0.63	-0.1	0.11	-0	-0.01	-0.11	1	-0
weekend	-0.03	0	0	-0.08	-0	0	-0.06	0	-0.02	-0.16	-0.06	-0	-0	-0.03	0.04	0.79	-0	0	-0	1
perc_close_open	open	close	perc_high_low	high	low	perc_value	value	std_value	volume_btc	volume_currency	median_value	mode_value	kurtosis_value	skewness_value	weekday	day_of_month	month	year	weekend	

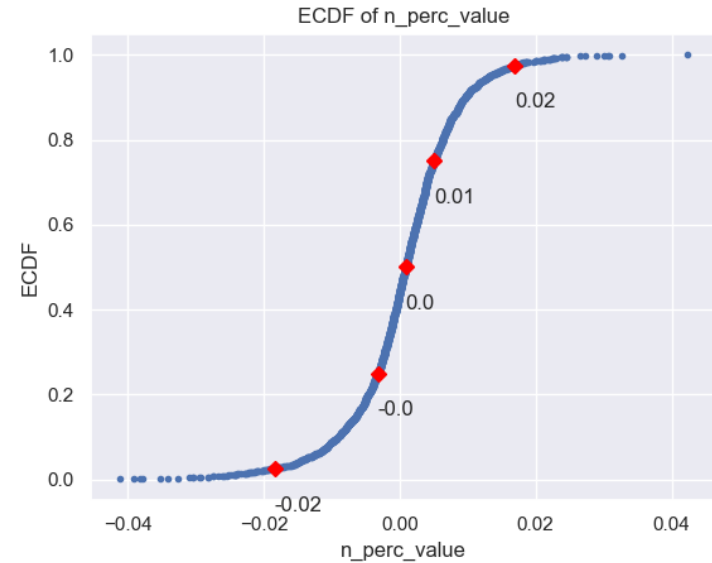
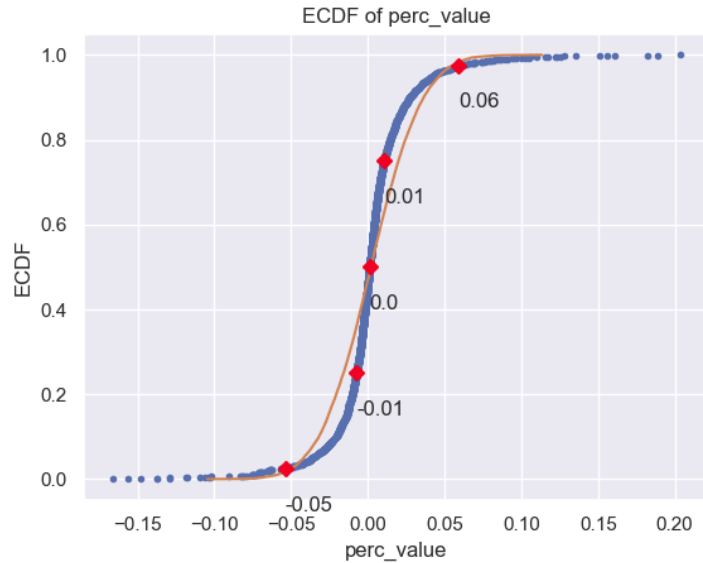
Correlation



Bitcoin change in value by day

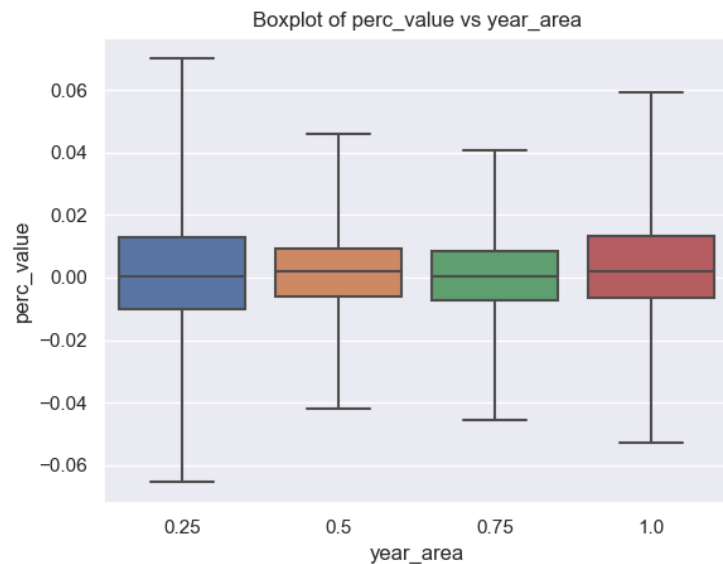


Bitcoin vs Nasdaq change value

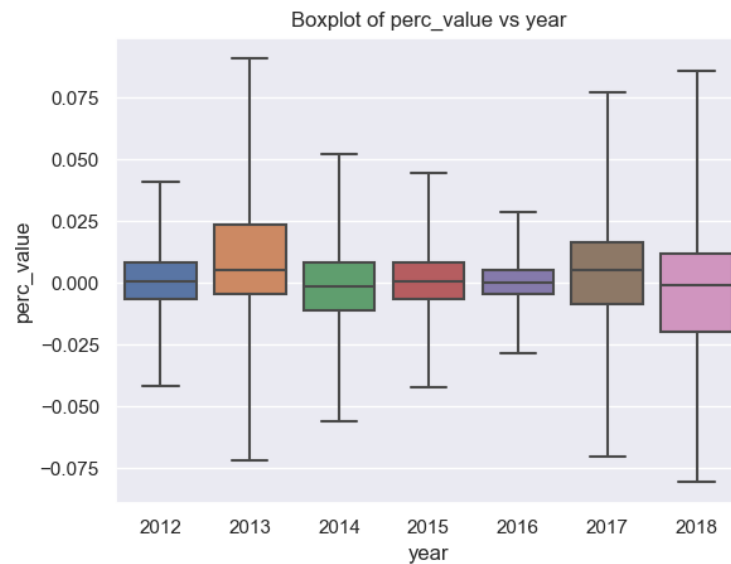


Bitcoin change value by aggregation

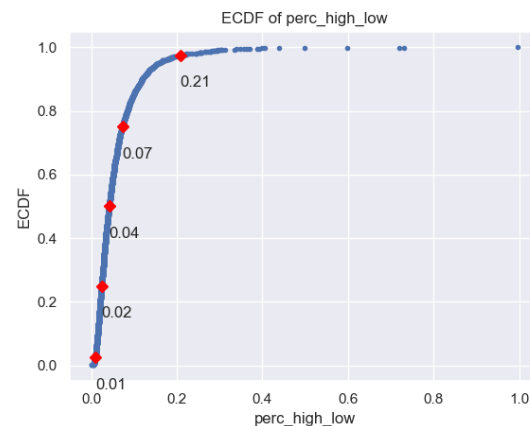
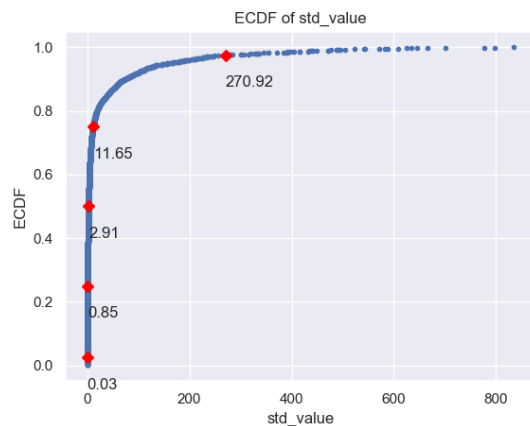
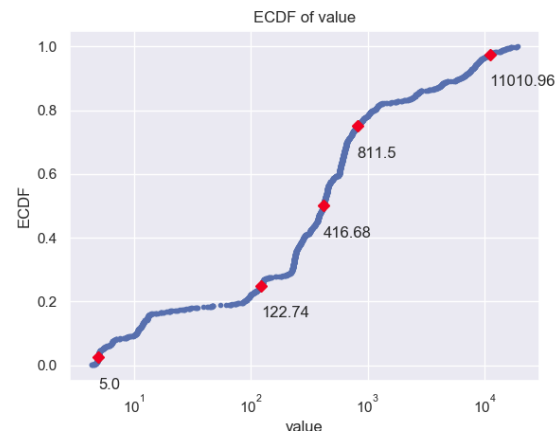
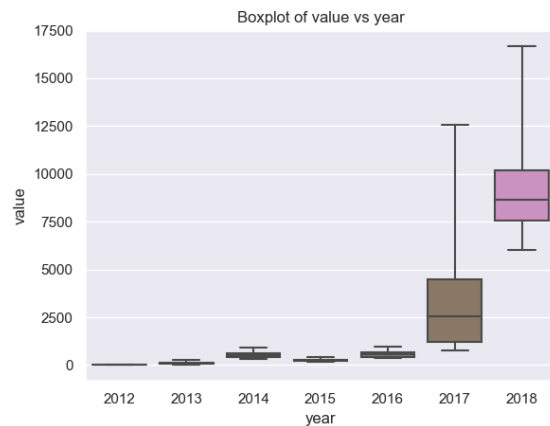
- By year area



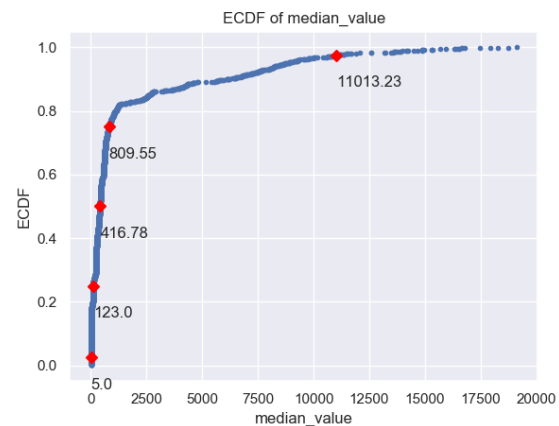
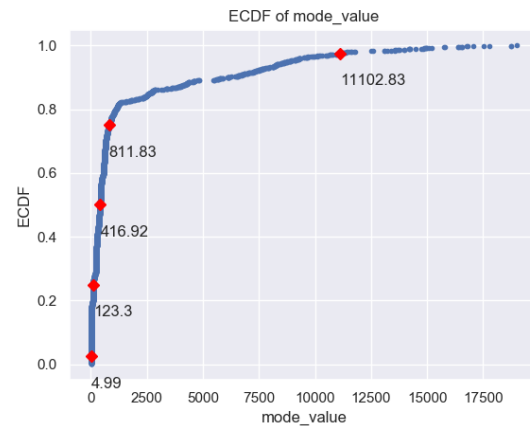
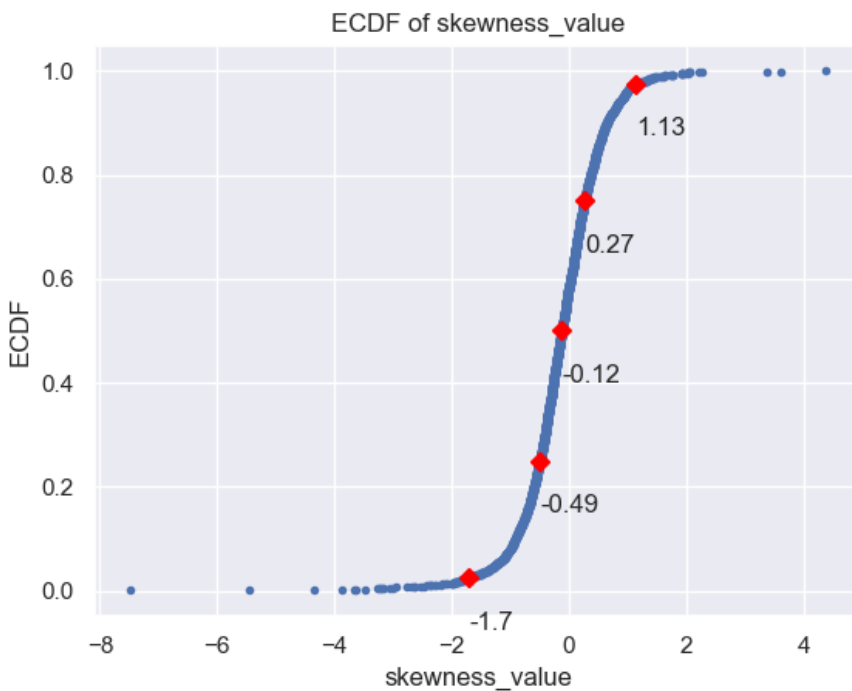
- By year value



Bitcoin value



Bitcoin: Skewness



Bitcoin: volume

