CSE 544 - Class Project

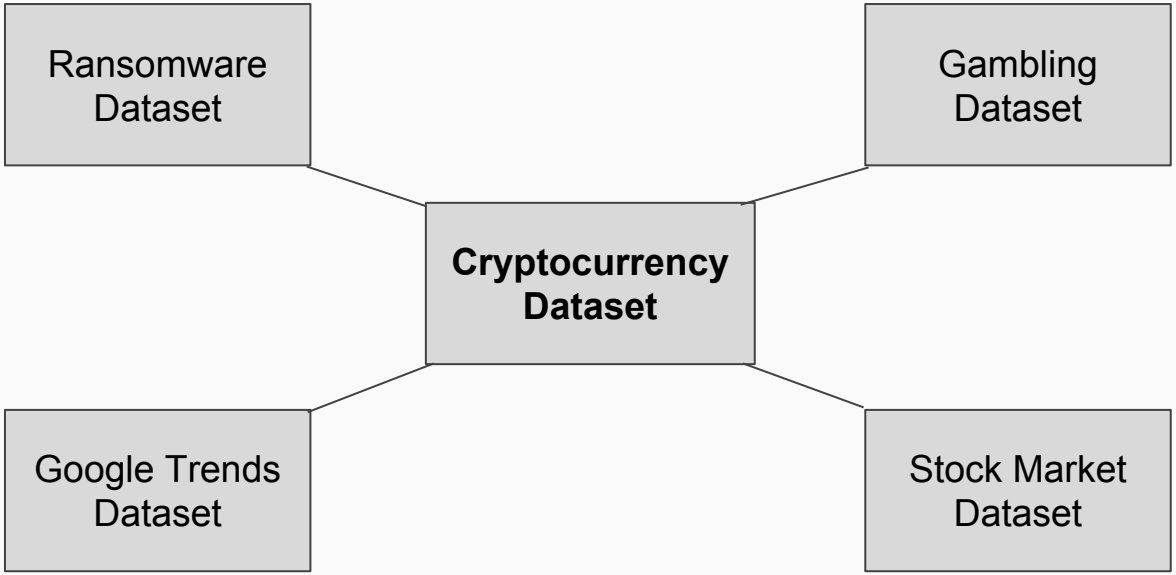
Cryptocurrency Analysis

Nishant Borude
Bhushan Sonawane
Aditya Yele
Dhanashri Patil
Mihir Chakradeo

Motivation

- Cryptocurrencies? All aboard the hype train!
- Affects multiple domains
- Potential to revolutionize currencies
- Volatile in nature,
  hence challenging to analyze

About the Dataset



Main Dataset: bitcoin dataset

- Information about the price, market cap, mining difficulty, etc of bitcoin

Date	btc_market_price (USD)	btc_total_bitcoins (Count)	btc_n_transactions (Count)	btc_market_cap (USD)
2018-02-19	11110.965	16875062.5	187367	187498228810
2018-02-20	11390.3916	16876825	198455	192233646840

- Data size
  - Daily bitcoin data from 2010 to 2018
  - 2920 rows and 24 feature columns
- Data Preprocessing
  - Normalization for analysis

Additional Dataset 1 - Google “Bitcoin” keyword search dataset

- Information about Search frequency (adjusted) , bitcoin daily closing price, and Date

Date	btc_market_price (USD)	Searches
2013-04-28	0.701172	3
2018-02-18	55.127267	26

- Data Size
  - Daily bitcoin data and Google Trends data from 2013 to 2018
  - 261 rows and 3 feature columns
- Data Preprocessing
  - Scaled btc_market_price to 0 -100 (in the range of Searches)

Additional Dataset 2 - Stocks

- Stock details of ASIC companies - NVIDIA, AMD, TSM

Date	Open (USD)	High (USD)	Low (USD)	Close (USD)	Volume (USD)
2017-11-09	205.27	206.33	200.370	205.32	23895006
2017-11-10	213.08	218.67	211.630	216.14	31300857

- Data Size (from 2010 to 2018)
  - NVIDIA - 4734 rows
  - AMD - 8738 rows
  - TSM - 3202 rows
- Data Preprocessing
  - Normalization

Additional Dataset 3 - GDP of Greece

- Contains the year quarter and the increase or decrease in GDP for that quarter in percentage.

Time	Value
2013-Q1	-2.192256
2013-Q2	0.077293

- Data Size: 18 rows
- Quarter 1 of 2013 - Quarter 1 of 2017

Additional Dataset 4 - Ransomware

- Number of different attacks that took place per month

Year	Month	Count
2016	4	14
2016	5	13

- Data Size
  - Data from Jan 2015 to July 2017: 31 rows
- Data Preprocessing
  - Normalization

Additional Dataset 5 - Gambling

- Information about player expenditure in 5 different gambling

Month	Year	Casino (USD)	EGM (USD)	Keno (USD)	Lottery (USD)	Wagering (USD)
7	2004	45662132.51	145766780.77	6836926.58	32451660.78	27298552.41
8	2004	36652961.84	143627917.85	4215418.49	24712127.1	26147748.54

- Data
  - Monthly player expenditure in USD from July 2004 till April 2017
- Data Processing and Cleaning
  - Refactored Game Stream as features (Casino, ..., Wagering)
  - Normalization

List of Topics

- Bitcoin price analysis and effect on other Cryptocurrencies
- Effect of Google Search on Bitcoin price
- Bitcoin and Finance
- Cryptocurrency and Ransomware
- Cryptocurrency and Gambling

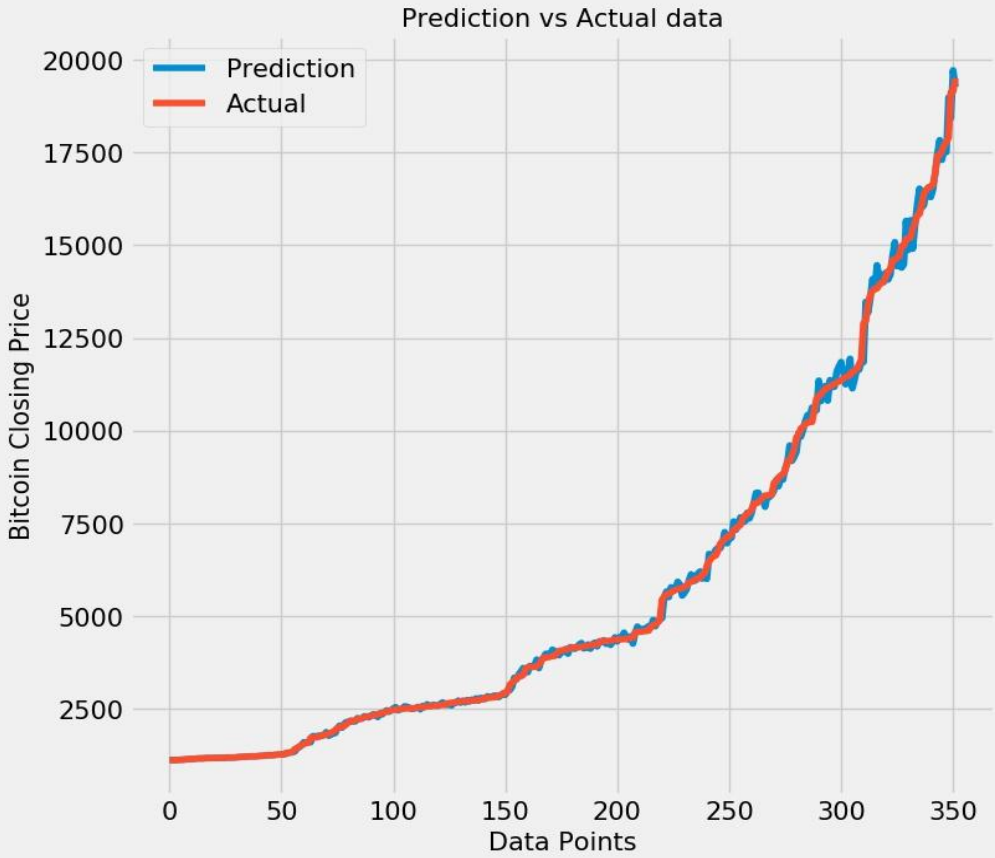
1. Bitcoin price analysis and effect on other Cryptocurrencies

- Time series analysis of Bitcoin closing price
  - EWMA - alpha = 0.5, 0.8
  - AR with p = 352
  - Seasonal with p = 352
- Metric: Mean Absolute Percentage Error

Method	MAPE
EWMA	0.9990
AR	1.5037
Seasonal	89.6630

Conclusion

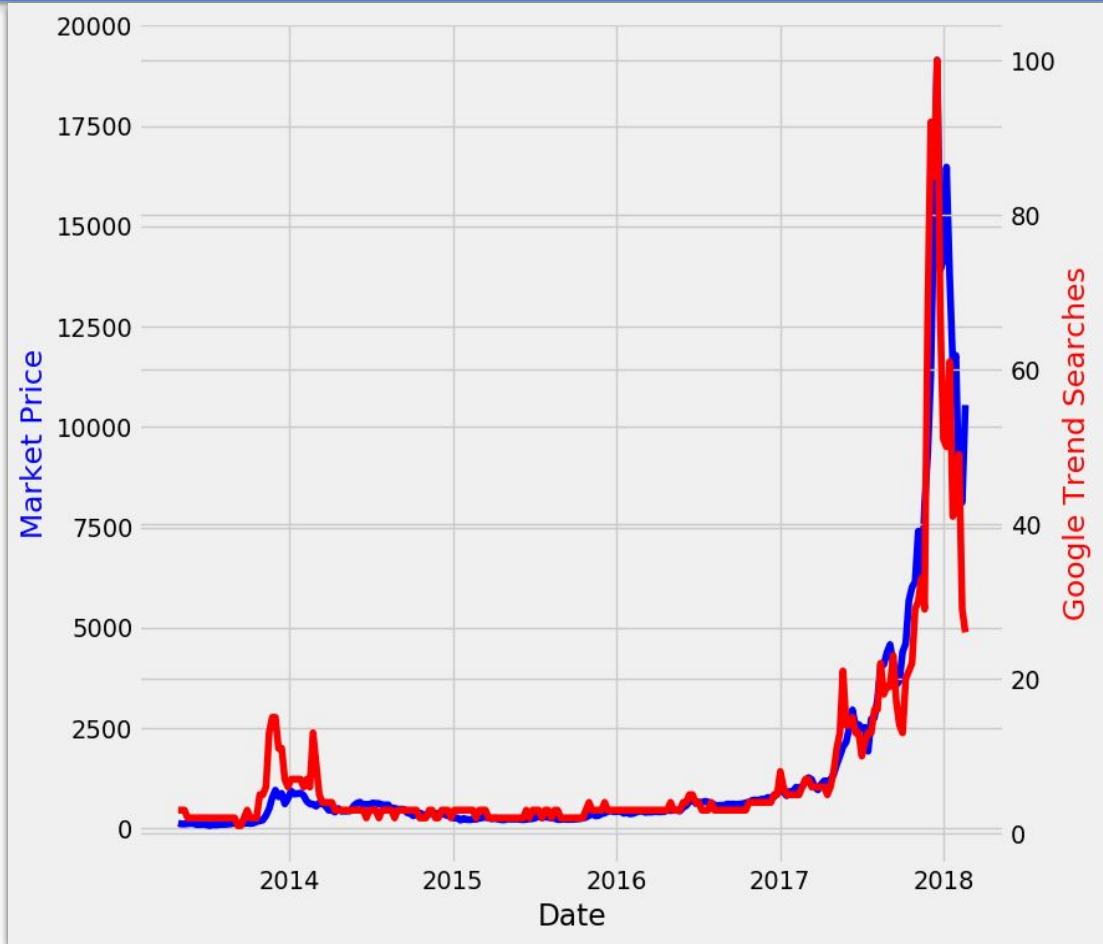
EWMA gives a better estimate for the bitcoin closing price



2. Effect of Google Search on Bitcoin price

Regression Analysis

- Simple Linear Regression to predict bitcoin market price using frequency of “Bitcoin” searches on Google

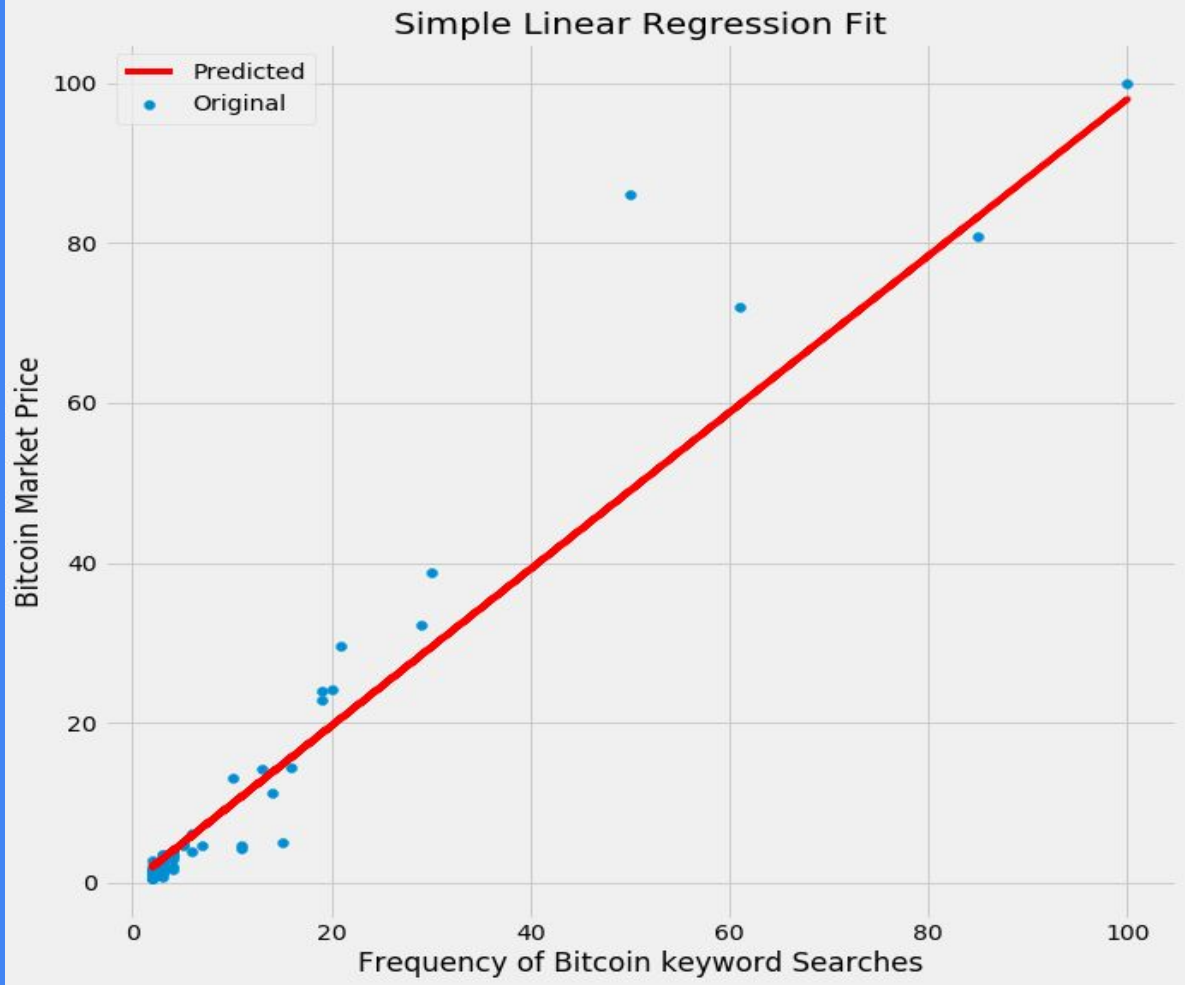


Metrics

- SSE - 2051.36
- MAPE - 52.93

Conclusion

The SSE and MAPE values are very bad, and hence we cannot predict bitcoin price from trends



3. Bitcoin and Finance

Predicting Stocks with Bitcoin

- Multiple linear regression
- Prediction
  - Stock price of an ASIC and GPU making company
- Features used
  - btc_market_price
  - btc_total_bitcoin
  - btc_trade_volume
  - btc_n_transaction

Results

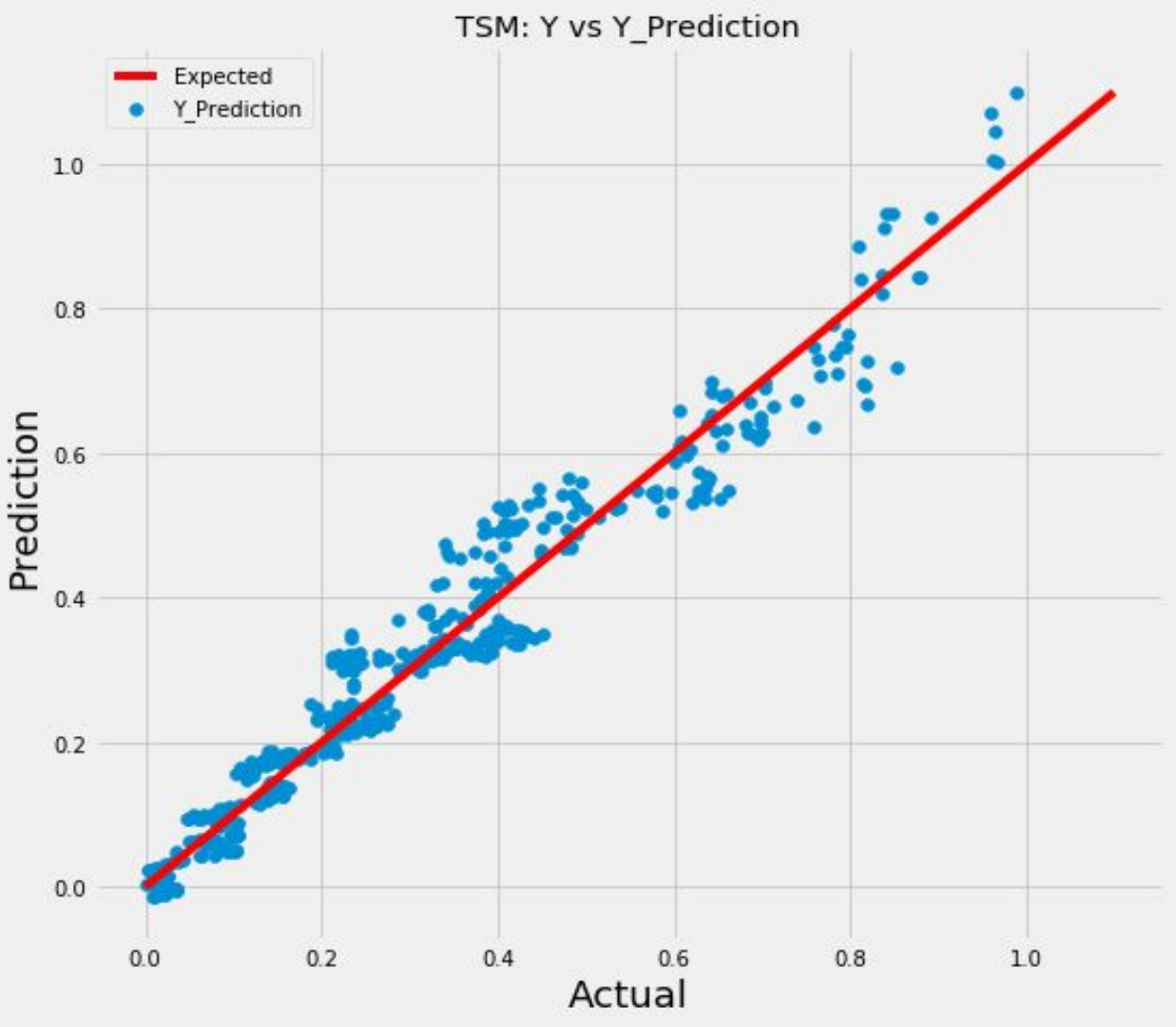
Company	SSE	MAPE
NVIDIA	1.4475	126.0607
AMD	9.2806	103.6849
TSM	1.0901	23.0741

Conclusion

Given bitcoin features linearly fit the stock market price data with the highest weight for btc_market_price

Weights For TSM:

btc_market_price: 0.6518
btc_total_bitcoins: 0.2934
btc_trade_volume: -0.1899
btc_n_transactions: 0.3536



4. Cryptocurrency and Ransomware attacks

Comparing the distribution of number of attacks and bitcoin per transaction cost

- Average data per month:
  - Number of attacks vs bitcoin cost per transaction
- Wald's 2 population test for alpha = 5%

Statistic	P-Value	Confidence Interval	Output	Conclusion
Wald's Test	0.5767	[-0.0796, 0.1430]	0.5581	**Accepted**

5. Cryptocurrency and Gambling

Use of bitcoin has steadily risen with increase in gambling expenditure

- Monthly averaged of amount (in USD) spent in 5 different gambling types
- Wald's two population test for alpha 5%
- Permutation test for alpha 5%

Gambling Type	Wald's Test (P-Value)	Confidence Interval	Permutation Test (P-Value)	Conclusion
Casino	0.3965	[-0.0473, 0.1194]	0.4026	**Accepted**
EGM	0.1755	[-0.0276, 0.1515]	0.1802	**Accepted**
Lottery	0.0165	[0.0194, 0.1945]	0.0185	**Rejected**
Keno	0.0063	[0.0375, 0.2292]	0.0082	**Rejected**
Wagering	0.4054	[0.0492, 0.1218]	0.4111	**Accepted**

Conclusion

1. **Bitcoin price analysis and effect on other Cryptocurrencies**
   a. Techniques Used:
      i. (Parametric Inference) Estimated the mean of bitcoin price using MME.
      ii. Time Series Analysis
      iii. Multiple Linear Regression to predict ethereum closing price based on bitcoin dataset (miners_revenue, hashing_difficulty, bitcoin_market_price)
   b. Conclusion:
      i. lambda_MME = 0.3580
      ii. EWMA performs better than Seasonal and AR model for TSA
      iii. Given features are closely correlated with ethereum closing value hence give good prediction for ethereum value
2. **Effect of Google Search on Bitcoin price**
   a. Techniques Used:
      i. Simple Linear Regression to predict bitcoin closing price based on frequency of keywords searched
      ii. (Non parametric) KS Test - Statistic, P value, Wald's 2 population Test - Statistic, P value
      iii. Multiple Linear Regression - find out which feature of bitcoin is influenced most by google searches
   b. Conclusion:
      i. Linear relationship exists between frequency of keywords searched and bitcoin_closing_price
      ii. Frequency of keywords searched and bitcoin closing price come from same the distribution
      iii. Number of transactions was influenced the most.

Conclusion

3. **Bitcoin and Finance**
    a. Techniques Used:
        i. Multiple Linear Regression
        ii. Permutation Test
        iii. Simple Linear Regression
    b. Conclusion
        i. Bitcoin features linearly fit stock prices with highest dependency on bitcoin market price
        ii. The distribution of TSM and bitcoin market price are not same
        iii. There is no relation between GDP of Greece and prices of bitcoin

4. **Cryptocurrency and Ransomware**
    a. Techniques Used:
        i. (Non parametric) Wald's 2 population test - Statistic, P value, 95% Confidence Interval
        ii. (Parametric Inference) Find the MLE parameter for Normal Distribution
        iii. (Bayesian Inference) Posterior Distribution analysis
    b. Conclusion:
        i. Two samples come from the same distribution
        ii. Found mu and sigma parameters.
        iii. The posterior distribution follows Normal distribution given Normal prior

Conclusion

5. **Cryptocurrency and Gambling**
    a. Techniques Used:
        i. Simple and Multiple Linear Regression
        ii. (Non parametric) Wald's 2 population test - Statistic, P value, 95% Confidence Interval
        iii. (Non parametric)Permutation Test - P value
    b. Conclusion:
        i. There is linear relationship between expenditure in gambling and total bitcoins per year and price.
        ii. Wagering contributes most to increase in number of bitcoins per year
        iii. EGM contributes most to increase in price of bitcoins per year

References

1. Cryptocurrency Dataset:
   https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory
2. Google Trends Dataset:
   https://trends.google.com/trends/explore?q=Bitcoin
3. Stock Market Dataset:
   https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs
4. Ransomware Dataset:
   http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/ISTR2016_Ransomware_and_Businesses.pdf
5. All gambling dataset Queensland:
   https://data.qld.gov.au/dataset/all-gambling-data-queensland
6. Wagering and bitcoin news article:
   https://news.bitcoin.com/bitcoin-gamblers-wagered-4-5-billion-btc-2014/
7. Quarterly GDP of Greece
   https://data.oecd.org/gdp/quarterly-gdp.htm