Our Summer Experience at IIT

By: Lucas Ciobanu, Stefan Donisa, and Lucas Raicu
Under the Guidance of Prof. Ioan Raicu
July 21st, 2023

ABOUT US

Lucas Ciob: 14 years old and a rising freshman at Lisle High School. Likes computers and gaming on them but doesn't know much about the hardware, how they work, or coding.

Lucas R: 14 years old and a rising sophomore at Glenbrook South High School. Had a CS class my first semester last year where I learned Java. Other than what I knew from my parents and this class, most of what we did this summer was completely new for me.

Stefan: 14 years old and a rising freshman at John Hersey High School. Likes playing computer games but doesn't know much on how computers work. No previous CS experience.

Timeline

1 — 2 — 3 — 4 — 5

Week 1

Learning about the basics of Computer Science along with reading Wikipedia articles and writing reports.

Week 2

Learning basics of all programming languages and learning how to code in Python.

Week 3

Took a deep dive into Python. Built a large calculator that performed functions in our Jupyter Notebook.

Week 4

Learned the basics of HTML and built our personal websites.

Week 5-6

Experimented with Binance's cryptocurrency transaction datasets, created candlesticks up to one second-level granularity, and then created plots for visualization.

A famous quote – "the advent of computation can be compared, in terms of the breadth and depth of its impact on research and scholarship, to the invention of writing and the development of modern mathematics"

— Ian Foster

Calculator Project

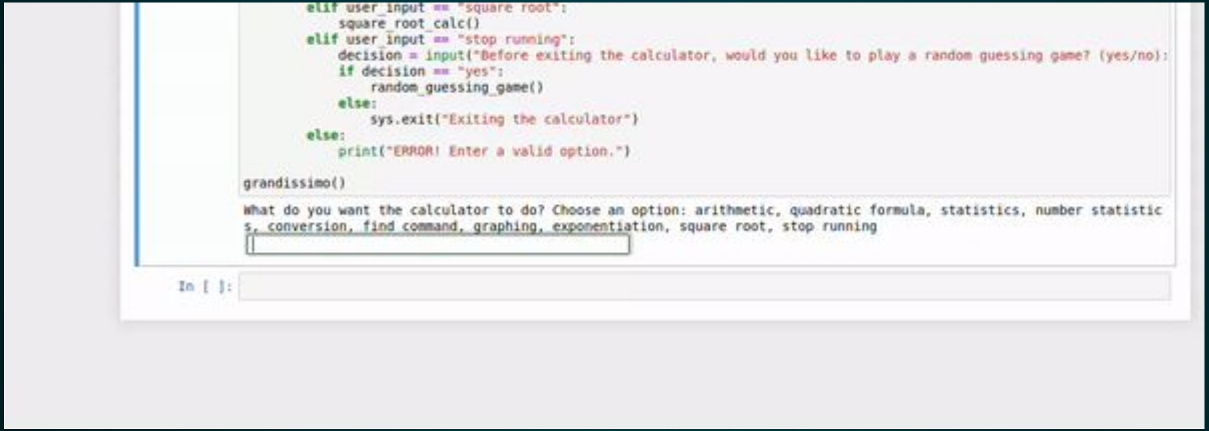
User Input

Computations +
Conversions

Libraries

Random
Guessing Game

The Grand Calculator



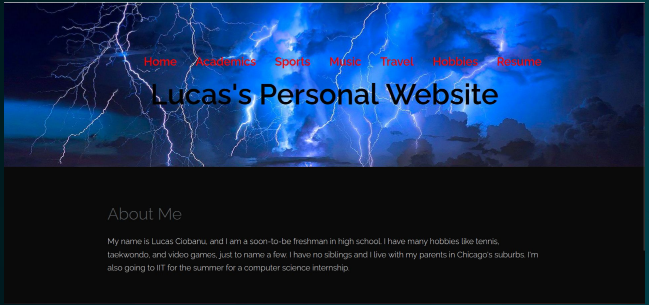
```
        elif user_input == "square root":
            square_root_calc()
        elif user_input == "stop running":
            decision = input("Before exiting the calculator, would you like to play a random guessing game? (yes/no):
            if decision == "yes":
                random_guessing_game()
            else:
                sys.exit("Exiting the calculator")
        else:
            print("ERROR! Enter a valid option.")

grandissimo()
```

```
What do you want the calculator to do? Choose an option: arithmetic, quadratic formula, statistics, number statistic
s, conversion, find command, graphing, exponentiation, square root, stop running
```

```
In [ ]:
```

HTML and CSS

Lucas C Website

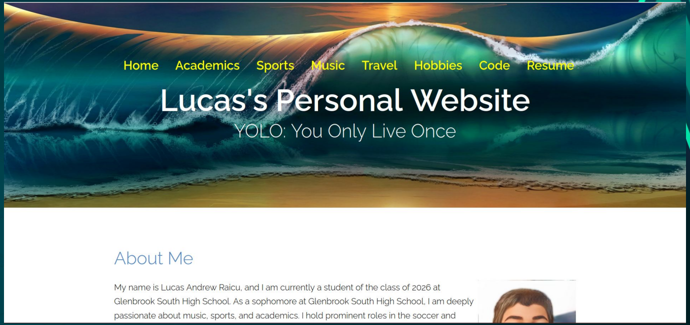


Lucas's Personal Website

About Me

My name is Lucas Ciobanu, and I am a soon-to-be freshman in high school. I have many hobbies like tennis, taekwondo, and video games, just to name a few. I have no siblings and I live with my parents in Chicago's suburbs. I'm also going to IIT for the summer for a computer science internship.

Lucas R Website



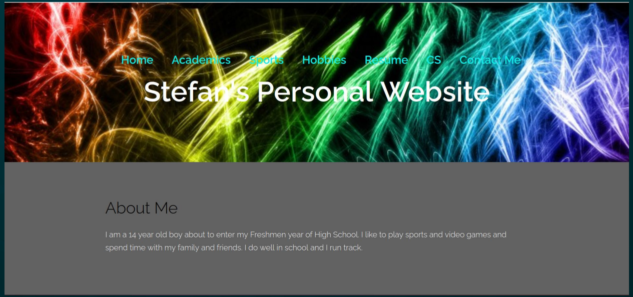
Lucas's Personal Website

YOLO: You Only Live Once

About Me

My name is Lucas Andrew Raicu, and I am currently a student of the class of 2026 at Glenbrook South High School. As a sophomore at Glenbrook South High School, I am deeply passionate about music, sports, and academics. I hold prominent roles in the soccer and

Stefan's Website



Stefan's Personal Website

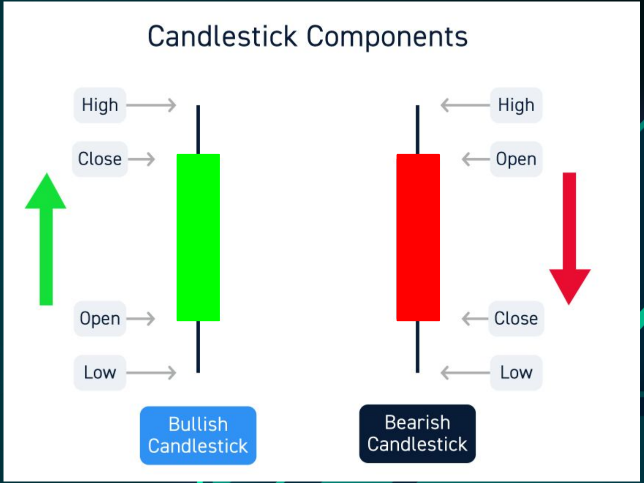
About Me

I am a 14 year old boy about to enter my Freshmen year of High School. I like to play sports and video games and spend time with my family and friends. I do well in school and I run track.

Improving Crypto Currency Analysis through Fine-Grained Historical Datasets

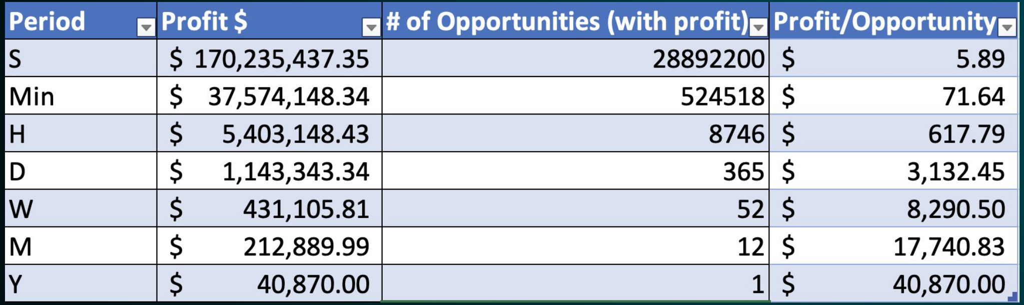
Background

- Time series Data:
  - A timeline of events or measurements
  - Example: A record of cryptocurrency price over time
- Candlestick Data and Plots
  - A way to visualize time series data
  - Represents the price movement of a cryptocurrency during a specific time period (e.g., a day or an hour)
  - More efficient and concise than time series data, since not all data points are shown
- Crypto Currencies
  - Digital assets for secure transactions
- Binance.us
  - A cryptocurrency exchange
  - A subsidiary of Binance, only catering to US users

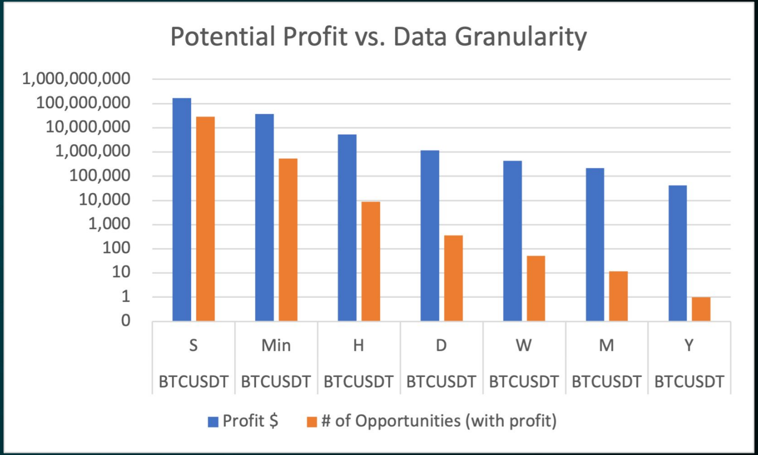


Motivations

Precise Data and Profit



Period	Profit $	# of Opportunities (with profit)	Profit/Opportunity
S	$ 170,235,437.35	28892200	$ 5.89
Min	$ 37,574,148.34	524518	$ 71.64
H	$ 5,403,148.43	8746	$ 617.79
D	$ 1,143,343.34	365	$ 3,132.45
W	$ 431,105.81	52	$ 8,290.50
M	$ 212,889.99	12	$ 17,740.83
Y	$ 40,870.00	1	$ 40,870.00



Problem Statement

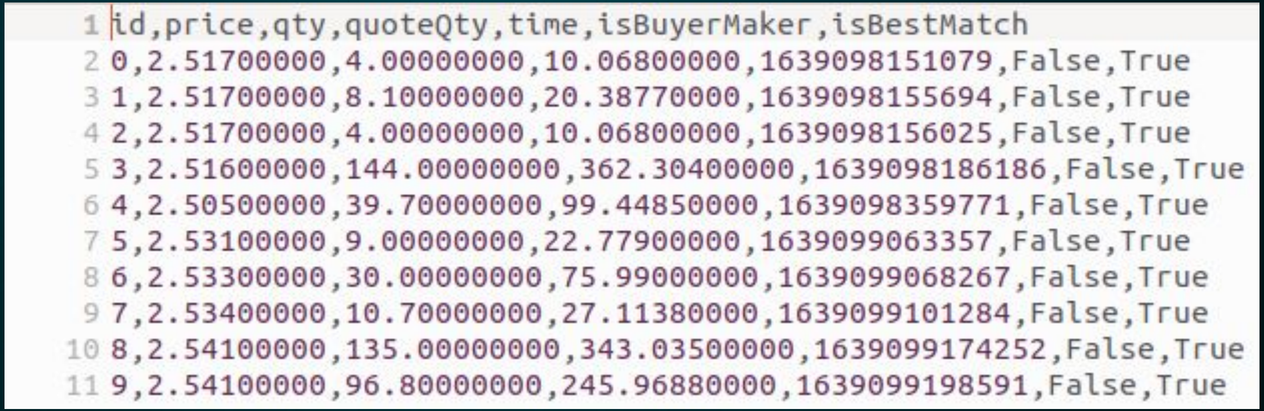
Fine grained crypto currency pricing data is not readily available in a form that can be used to conduct data analysis

Our Solution

1. Collect all transactions from Binance.us
2. Generate fine-grained datasets with 1-sec and up candle sticks
3. Generate plots to visualize data
4. Automate process to ensure datasets are up-to-date
5. Host datasets on website and Kaggle

Our Solution: Part 1

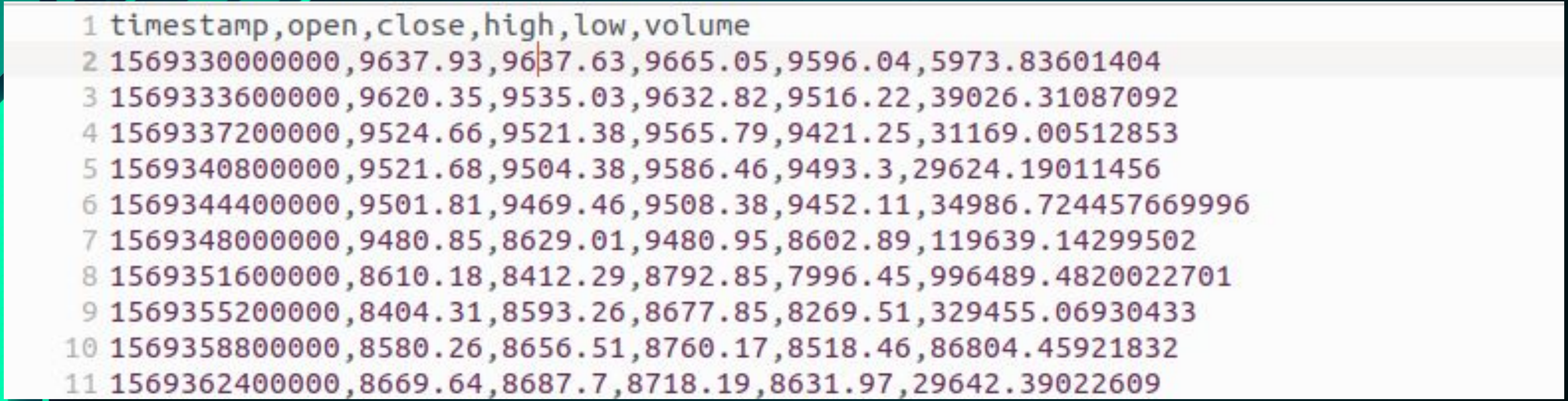
- Getting the transactions from Binance using our api key and importing the requests library



```
id,price,qty,quoteQty,time,isBuyerMaker,isBestMatch
0,2.51700000,4.00000000,10.06800000,1639098151079,False,True
1,2.51700000,8.10000000,20.38770000,1639098155694,False,True
2,2.51700000,4.00000000,10.06800000,1639098156025,False,True
3,2.51600000,144.00000000,362.30400000,1639098186186,False,True
4,2.50500000,39.70000000,99.44850000,1639098359771,False,True
5,2.53100000,9.00000000,22.77900000,1639099063357,False,True
6,2.53300000,30.00000000,75.99000000,1639099068267,False,True
7,2.53400000,10.70000000,27.11380000,1639099101284,False,True
8,2.54100000,135.00000000,343.03500000,1639099174252,False,True
9,2.54100000,96.80000000,245.96880000,1639099198591,False,True
```

Our Solution: Part 2

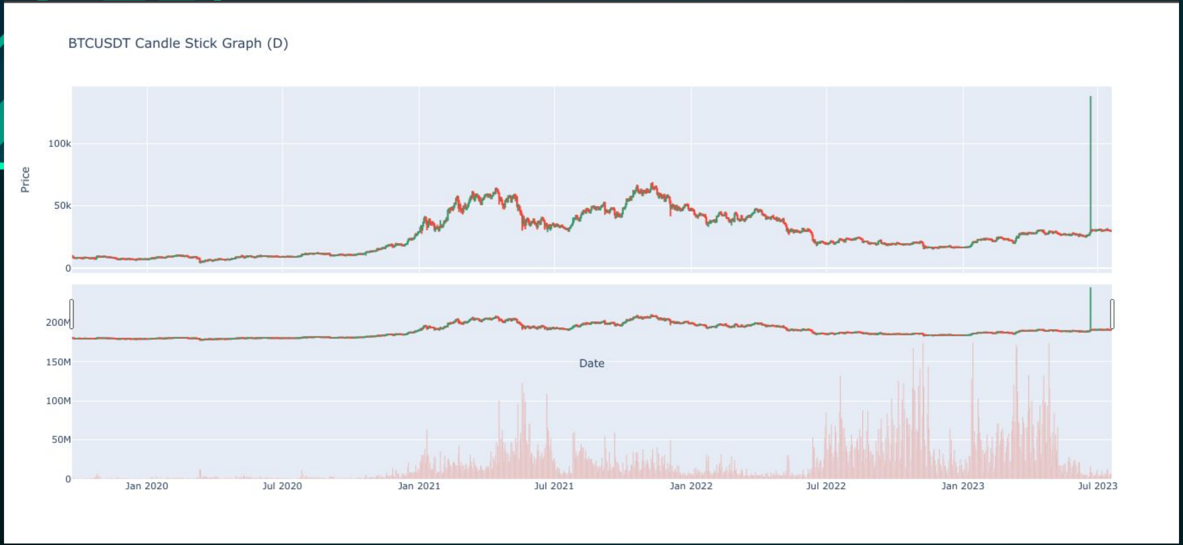
- Extract the prices and timestamps from the transactions and make candlesticks.
- Did candlesticks for seconds, minutes, hours, days, weeks, months, and years.



```
timestamp,open,close,high,low,volume
1569330000000,9637.93,9637.63,9665.05,9596.04,5973.83601404
1569333600000,9620.35,9535.03,9632.82,9516.22,39026.31087092
1569337200000,9524.66,9521.38,9565.79,9421.25,31169.00512853
1569340800000,9521.68,9504.38,9586.46,9493.3,29624.19011456
1569344400000,9501.81,9469.46,9508.38,9452.11,34986.724457669996
1569348000000,9480.85,8629.01,9480.95,8602.89,119639.14299502
1569351600000,8610.18,8412.29,8792.85,7996.45,996489.4820022701
1569355200000,8404.31,8593.26,8677.85,8269.51,329455.06930433
1569358800000,8580.26,8656.51,8760.17,8518.46,86804.45921832
1569362400000,8669.64,8687.7,8718.19,8631.97,29642.39022609
```

Our Solution: Part 3

- Generated plots for the data
- The smaller the granularity, the longer it took to render the graph



Our Solution: Part 4

- Automated process every midnight
  - Transactions, candlesticks, and plots all update
  - New data is appended to files

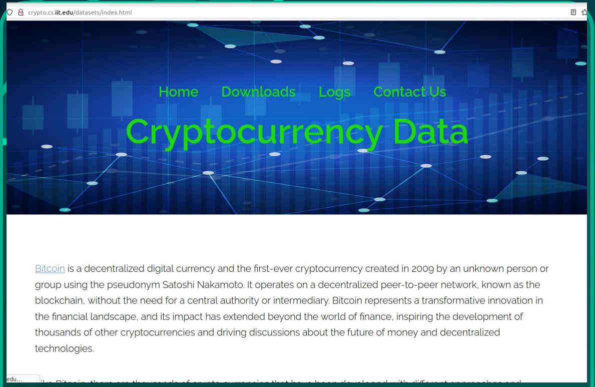
Our Python Code:

Our Solution: Part 5

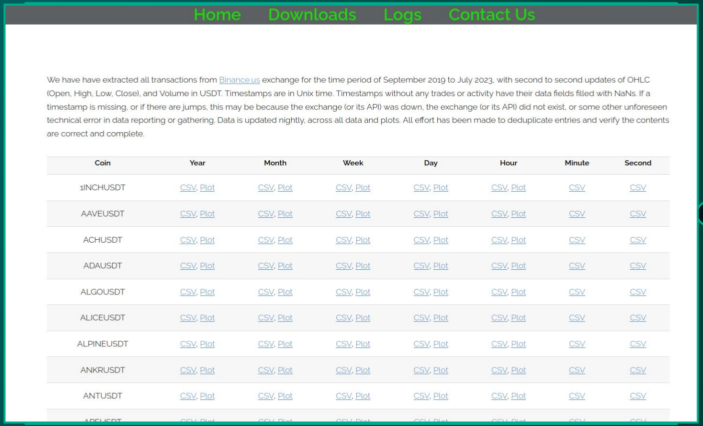
- We used one of the Mystic computers
- Installed a RAID5 array spanning 6 disks ⇒ 1.2TB storage

The Website

- Website has data for many different cryptos
- Many granularities to choose from
- Graphs and organization
- http://crypto.cs.iit.edu/datasets/index.html



Bitcoin is a decentralized digital currency and the first-ever cryptocurrency created in 2009 by an unknown person or group using the pseudonym Satoshi Nakamoto. It operates on a decentralized peer-to-peer network, known as the blockchain, without the need for a central authority or intermediary. Bitcoin represents a transformative innovation in the financial landscape, and its impact has extended beyond the world of finance, inspiring the development of thousands of other cryptocurrencies and driving discussions about the future of money and decentralized technologies.



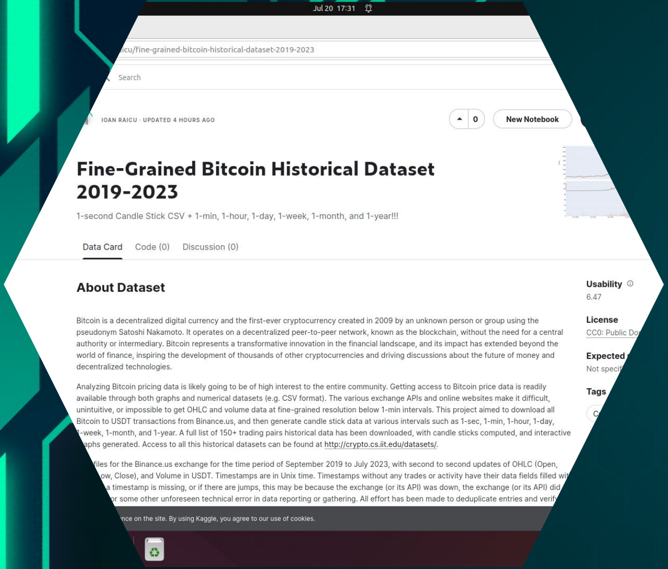
We have have extracted all transactions from Binance.us exchange for the time period of September 2019 to July 2023, with second to second updates of OHLC (Open, High, Low, Close), and Volume in USDT. Timestamps are in Unix time. Timestamps without any trades or activity have their data fields filled with NaNs. If a timestamp is missing, or if there are jumps, this may be because the exchange (or its API) was down, the exchange (or its API) did not exist, or some other unforeseen technical error in data reporting or gathering. Data is updated nightly, across all data and plots. All effort has been made to deduplicate entries and verify the contents are correct and complete.

Coin	Year	Month	Week	Day	Hour	Minute	Second
1INCHUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
AAVEUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
ACHUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
ADAUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
ALGOUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
ALICEUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
ALPINEUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
ANKRUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV
ANTUSDT	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV, Plot	CSV	CSV

Our Web Site HTML:

Kaggle

- An online community platform for data scientists and people who are interested machine learning
- Available DataSets related to our project
  - Bitcoin Historical Data of 1-min candles since 2012 till 2021
- New Page Published
  - https://www.kaggle.com/datasets/iraicu/fine-grained-bitcoin-historical-dataset-2019-2023


Jul 20 17:31
icu/fine-grained-bitcoin-historical-dataset-2019-2023
Search
IOAN RAICU · UPDATED 4 HOURS AGO
0
New Notebook
Data Card
Code (0)
Discussion (0)

Fine-Grained Bitcoin Historical Dataset 2019-2023

1-second Candle Stick CSV + 1-min, 1-hour, 1-day, 1-week, 1-month, and 1-year!!!

About Dataset

Bitcoin is a decentralized digital currency and the first-ever cryptocurrency created in 2009 by an unknown person or group using the pseudonym Satoshi Nakamoto. It operates on a decentralized peer-to-peer network, known as the blockchain, without the need for a central authority or intermediary. Bitcoin represents a transformative innovation in the financial landscape, and its impact has extended beyond the world of finance, inspiring the development of thousands of other cryptocurrencies and driving discussions about the future of money and decentralized technologies.

Analyzing Bitcoin pricing data is likely going to be of high interest to the entire community. Getting access to Bitcoin price data is readily available through both graphs and numerical datasets (e.g. CSV format). The various exchange APIs and online websites make it difficult, unintuitive, or impossible to get OHLC and volume data at fine-grained resolution below 1-min intervals. This project aimed to download all Bitcoin to USDT transactions from Binance.us, and then generate candle stick data at various intervals such as 1-sec, 1-min, 1-hour, 1-day, 1-week, 1-month, and 1-year. A full list of 150+ trading pairs historical data has been downloaded, with candle sticks computed, and interactive aphs generated. Access to all this historical datasets can be found at http://crypto.cs.iit.edu/datasets/.

files for the Binance.us exchange for the time period of September 2019 to July 2023, with second to second updates of OHLC (Open, ow, Close), and Volume in USDT. Timestamps are in Unix time. Timestamps without any trades or activity have their data fields filled wit a timestamp is missing, or if there are jumps, this may be because the exchange (or its API) was down, the exchange (or its API) did or some other unforeseen technical error in data reporting or gathering. All effort has been made to deduplicate entries and verify

Usability
6.47

License
CC0: Public Do

Expected
Not specif

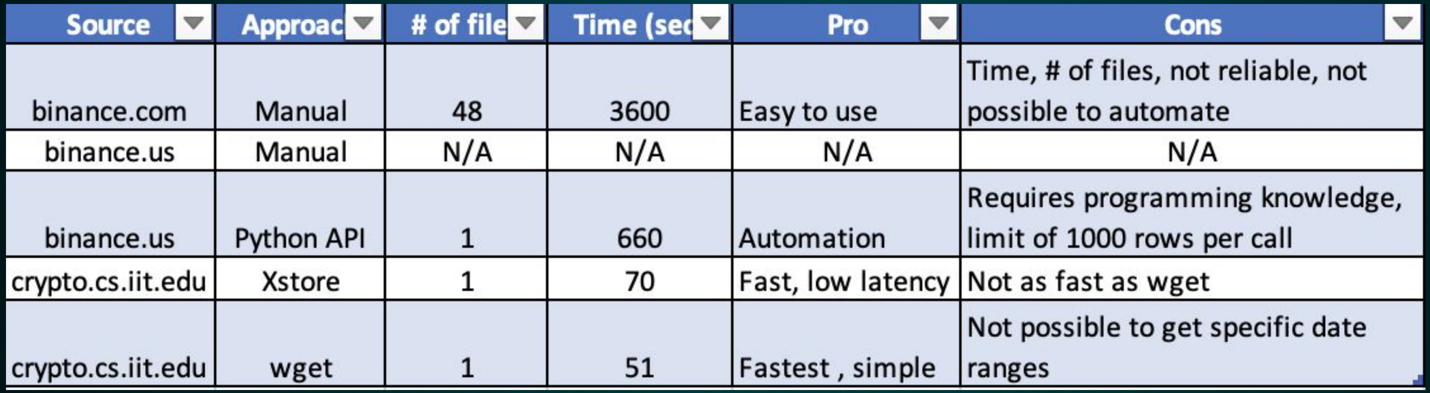
Tags

nce on the site. By using Kaggle, you agree to our use of cookies.

Obstacles along the Way

- Jupyter notebook environment
- Large text files
- Binance.us vs Binance.com
- Finding the right time zone
- Long update wait time when writing
- Github pages storage

Different Approaches



Source	Approac	# of file	Time (sec	Pro	Cons
binance.com	Manual	48	3600	Easy to use	Time, # of files, not reliable, not possible to automate
binance.us	Manual	N/A	N/A	N/A	N/A
binance.us	Python API	1	660	Automation	Requires programming knowledge, limit of 1000 rows per call
crypto.cs.iit.edu	Xstore	1	70	Fast, low latency	Not as fast as wget
crypto.cs.iit.edu	wget	1	51	Fastest , simple	Not possible to get specific date ranges

So what?

- Predictive Analytics
- Backtesting Strategies

Any Questions?