

Project Goal

Our project aims to create a SQL database of monthly aggregate prices from 2018 - 2021 of 6 crypto-currencies and 14 major stock indices from around the world. This database is poised for potential EDA on crypto prices.

Project Organization:

Group members tackled the following:

| Ben | Chris | Emily | Hana |
|---------------------|------------------|---------------|---------------------|
| Dogecoin Binance | Tether Monero | Major indices | Bitcoin Ethereum |

Project Report

Extract:

We obtained our raw datasets from two sources: Kaggle.com and Yahoo! Finance. These sources were either direct download or filtered from the web interface (.csv), totaling 7 datasets with over 125k raw data points.

[Kaggle Stock Exchange Data](#)

[Yahoo!Finance Cryptocurrency Data Site](#)

Transform:

The raw dataset consisted of daily prices in the following categories:

Prices at x and y times, opening price, closing price, high price, low price, marketcap, and volume.

- We binned all the daily prices by month and then ran these aggregation functions:
 - max of high price per month
 - max of marketcap per month (calculated by multiplying:)
 - max of volume per month
 - max of high price per month
 - min of low price per month
 - average of opening and closing price
 - rounded the floats to 3 decimal places
- Dropped duplicate values

- Dropped Index when exporting cleaned .csv files

Load:

We chose to load our data into a .sql file. Our goal was to create a database that could answer which crypto prices were correlated with various stock indexes. We chose to create a relational database using the schemas below.

The following schemata was used to create an identifier table, allowing us to easily view all of the different crypto and indices available in our database:

- ID INT NOT NULL
- Symbol VARCHAR(10) NOT NULL
- Name VARCHAR NOT NULL
- PRIMARY KEY (ID)

The following schemata:

- Date INTEGER NOT NULL
- Open FLOAT NOT NULL
- Close FLOAT NOT NULL
- High FLOAT NOT NULL
- Low FLOAT NOT NULL
- Marketcap DECIMAL NOT NULL
- Symbol VARCHAR(10) NOT NULL
- Name VARCHAR NOT NULL
- PRIMARY KEY (Date, Symbol)

The following tables were created:

- | | |
|---------------|---------|
| • Identifier | • J203 |
| • Monero | • KS11 |
| • Tether | • N100 |
| • Bitcoin | • N225 |
| • Dogecoin | • NSEI |
| • BinanceCoin | • NYA |
| • Ethereum | • oneSS |
| • GDAXI | • SSMI |
| • GSPTSE | • SZ |
| • HSI | • TWII |
| • IXIC | |

Limitations:

We made assumptions that the Kaggle dataset we are using for the stock prices accurately reflects what Yahoo!Finance reported for the years we needed. We did a quick inspection of the data, and it appeared to be accurate. There is a possibility we missed days that were not as reported on the original Yahoo site.

To create an intersection with the Stock Market Index(SMI), we formatted it the same way as the Cryptocurrency dataset. The SMI .csv file contained all indexes within one file. Manual

formatting and extraction of each index was needed to keep it consistent. This could lead to potential human errors as some naming conventions or data could have been transformed or formatted differently than expected.

While the dataset can be leveraged for a comparison of cryptocurrency and international index market prices and capitalization, more data and different data would likely be needed for a full comparison between cryptocurrency and stock market indexes (or their constituents). Information covering a longer time period beyond two years would likely also be needed to better understand the correlation (if any) between the two kinds of investments.