Cleaning

- Python Pandas-- Jupyter Notebook
- Reformat date for Bitcoin API/JSON dataset
  - Innner merge on date
- Columns: dropped due to non-relevance
- Columns: renamed to keep consistentsy
- **Challenge:** Remove rows that contained empty fields in the Market Cap column with (-)

Cleaning

- Once cleaned seperately, cleaned once combined
- New dataframe created seperately for index variable
- Faster/ease of use in database loading phase than searching through string

Bitcoin

- Bitcoin dataset (API/JSON):
  - Bitcoin CoinGecko API/JSON
  - Create loop call 2017-2019 dates
  - Build in try-except
  - Dataframe created from call: date, market cap, volume, current price columns
- **Challenge:** Extraction was difficult because of creation of loop and try-except

Bitcoin Cash

- Bitcoin Cash dataset(CSV):
  - Kaggle
  - Daily dates & stats
  - Date, open, high, low, close, volume, market cap

Transform

Extract

BITCOIN V. BITCASH

LOAD

Purpose

Database

- Pg4Admin SQL used to load data
  - Data relational
    - Only joined on one element-date
  - Two tables created
    - Index- primary key
    - Bitcoin/Bitcoin Cash combined date- foreign key
      - References primary key

Analysis

- To further normalize, used Matplotlib to generate line graph based on series of common dates

Analysis

Analysis

- Bitcoin Cash generally follows the trend of Bitcoin
- Hypothesize: Pulling in any other coin, they would follow a similar trend
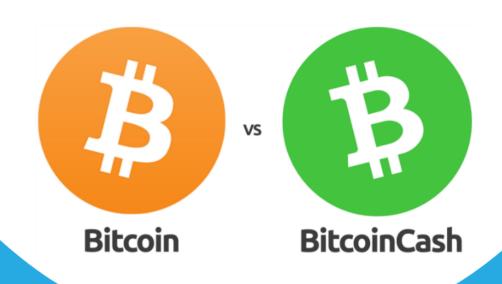 - Bitcoin is the "gold" standard of cryptocurrencies.

BITCOIN V. BITCASH

Christos Ferderigos, Zohna Jones, Omar Kreidieh, Anu Srivastava

Prezi

Purpose

- Bitcoin and Bitcoin Cash: popular and increasingly profitable cryptocurrencies

- Purpose: Perform the ETL process on Bitcoin and Bitcoin Cash data to analyze the relationship.

Bitcoin vs **BitcoinCash**

Extract

Bitcoin Cash

- Bitcoin Cash dataset(CSV):
  - Kaggle
  - Daily dates & stats
  - Date, open, high, low, close, volume, market cap

Bitcoin

- Bitcoin dataset (API/JSON):
    - Bitcoin CoinGecko API/JSON
    - Create loop call 2017-2019 dates
    - Build in try-except
    - Dataframe created from call: date, market cap, volume, current price columns

- **Challenge:** Extraction was difficult because of creation of loop and try-except

Cleaning

- Python Pandas-- Jupyter Notebook
- Reformat date for Bitcoin API/JSON dataset
  - Innner merge on date
- Columns: dropped due to non-relevance
- Columns: renamed to keep consistentsy

- **Challenge:** Remove rows that contained empty fields in the Market Cap column with (-)

Cleaning

- Once cleaned seperately, cleaned once combined
- New dataframe created seperately for index variable
- Faster/ease of use in database loading phase than searching through string

Database

- Pg4Admin SQL used to load data
  - Data relational
  - Only joined on one element-date
- Two tables created
  - Index- primary key
  - Bitcoin/Bitcoin Cash combined date-foreign key
    - References primary key

Analysis

- To further normalize, used Matplotlib to generate line graph based on series of common dates

Analysis