DSCI 510 HW5
(Final Project)

Tejas Sujit Bharambe (2160 8491 14)

Scope of Project

Find three (3) data sets on the web such that

- One must require “scraping” (i.e. not available via external API)
- One must be available via external public API
- The third can be either (with API, or scraped)

Describe what analysis or presentation you’d like to do with the combined data

Motivation

Lately, cryptocurrencies have gained a lot of popularity in the finance sector owing to the massive returns, they have been offering since last few years. There are more than thousands of cryptocurrencies that have been generated.

This project focuses on the top 25 cryptocurrencies as they constitute more than 90% of the cryptocurrency market capitalization. The idea is to focus on the cryptocurrencies which are currently trending based on the historical returns as well as based on the sentiments across various social media platforms.

This is an interesting problem to make public aware of the highly trending cryptocurrencies and guide them while doing the investments. The project identifies top 10 cryptocurrencies based on market cap, social media attention on Reddit, Twitter and Cryptocompare. Additionally, based on the current price of the cryptocurrency and the technical indicator, the project also suggests whether a particular cryptocurrency shows a bullish signal or a bearish signal or is neutral.

Data Sets and their brief description

The following three data sets will be used in the project:

Data Set	Description	API/Website	API Auth	API HTTPS
Nomics	Historical and real time cryptocurrency prices and market data across exchanges	API	API key needed	Yes
CryptoCompare	Comparison of cryptocurrencies. It contains cryptocurrency trade data, order book data, blockchain data, social data and historical data	API	API key needed	Yes
Coinmarketcap	Price-Tracking website for crypto assets. It also contains the ranking, market cap, gainers, losers, trending and most visited crypto assets	Website	-	-

Project Objective (Analysis Performed)

Perform the following analytics and visualizations based on the data using python concepts like Data structures, Class, Functions, Web scraping websites, File concepts:

1. Identify top 25 cryptocurrencies based on market capitalization and their market capture. Insights and visualizations added using scatter plot and bar plot

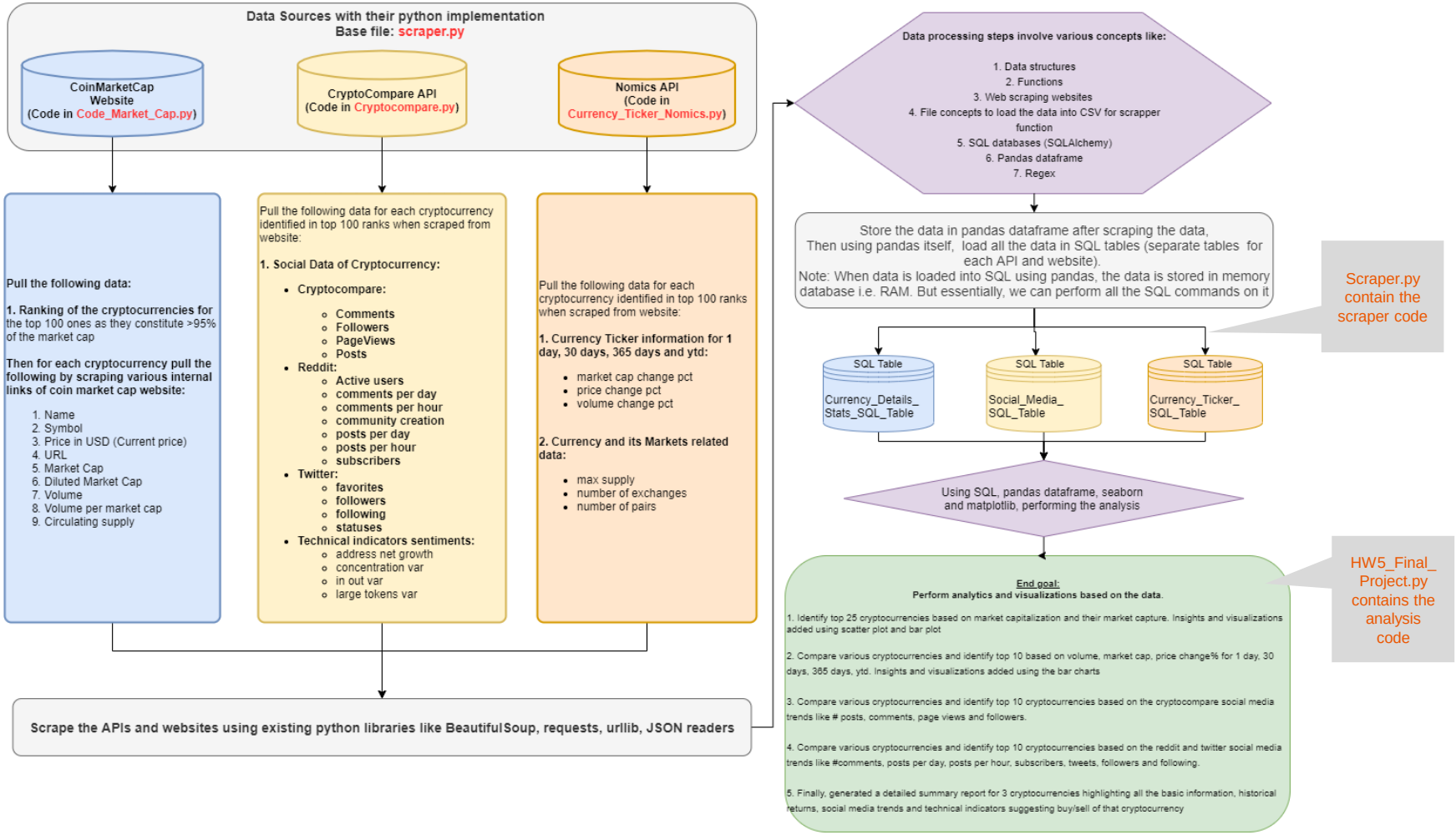
2. Compare various cryptocurrencies and identify top 10 based on volume, market cap, price change% for 1 day, 30 days, 365 days, YTD. Insights and visualizations added using the bar charts

3. Compare various cryptocurrencies and identify top 10 cryptocurrencies based on the cryptocompare social media trends like # posts, comments, page views and followers.

4. Compare various cryptocurrencies and identify top 10 cryptocurrencies based on the reddit and twitter social media trends like #comments, posts per day, posts per hour, subscribers, tweets, followers and following.

5. Finally, generated a detailed summary report for 3 cryptocurrencies highlighting all the basic information, historical returns, social media trends and technical indicators suggesting buy/sell of that cryptocurrency

Project Flow Chart – Showcasing what the script does



Conclusions (Part 1)

There are more than thousands of cryptocurrencies in the market. But the top 25 cryptocurrencies control more than 90% of this market. The top 3 cryptocurrencies (Bitcoin, Ethereum and Binance Coin) based on market cap control ~68% of the market. The top 5 cryptocurrencies based on market cap control ~75% of the market.

The trending cryptocurrencies based on the market cap, volume and prices across last 30 days, 365 days and 2021 YTD vary a lot. For market cap, the most trending cryptocurrencies were Gala, Solana and Avalanche whereas in terms of price returns, they were Shiba, Gala and TerraUSD.

Shiba has given the highest returns of ~3150% in the last 1 year followed by Gala with 1631%

The cryptos which are most widely talked about on the cryptocompare social media platform are: Bitcoin, XRP and Ethereum

The most famous cryptos based on the reddit trends (posts per hour, comments, subscribers) are Bitcoin, Dogecoin, Loopring and Helium

Bitcoin has highest comments per day n Reddit with a total of 7735.0 comments per day, followed by Loopring with a total of 6206.9 comments per day And then Dogecoin with a total of 5073.4 comments per day

**** The numbers in this slide are based on the market price of the crypto currencies as of 12/3**

Conclusions (Part 2)

The most famous cryptos based on the twitter trends (favorites, followers and tweets) are Binance, Dogecoin, XRP, KuCoin and Huobi. Huobi Token has the highest favorites on Twitter with a total of 38831.0 favourites. Binance Coin has the highest followers with a total of 6454110.0 followers, followed by Dogecoin with a total of 2603426.0 followers

Bitcoin may be one of the most discussed cryptocurrency but there are more than 50 other cryptocurrencies which have given greater returns in the past 1 year. Currently, the technical analysis of Bitcoin suggests a bearish signal

Ethereum may be second most discussed cryptocurrency on social media but there are more than 30 cryptocurrencies which have given greater returns in the past 1 year. For Ethereum also, the current technical analysis suggest that it will fall in the near future.

**** The numbers in this slide are based on the market price of the crypto currencies as of 12/3**

Requirements to run the project

1. Do not change the folder structure. Also, keep/extract all the folders present in the ZIP file

- **'Code'** folder contains all the .py files. There are 5 files there. The HW5_Final_Project.py is the main file and it imports from
  - → **scraper.py** - This is to scrape the data from API/websites and (or) csv files. It uses:
    - → **Coin_Market_Cap.py** - This is to scrape the data from website
    - → **Cryptocompare.py** - This is to scrape the data from API #1
    - → **Currency_Ticker_Nomics.py** - This is to scrape the data from API #2
  - → **HW5_Final_Project.py** – This contains the code for performing the analytics and insights
- **'Data'** folder contains the scrapped data from APIs and website in csv format
- **'Data_Generated_For_Processing_Not_Output'** folder contains the intermediate data which is needed for processing of the data

Requirements to run the project

2. Following packages needed for implementing the code:

- Requests
- Bs4 (Beautiful soup)
- Pandas
- Urllib
- JSON
- Time
- Sys
- CSV
- SQL Alchemy
- Seaborn
- Matplotlib.pyplot
- Numpy
- Squarify
- Warnings

Note: The pip install commands are already included in the code using the os commands. Hence, this shouldn't be a concern.

Requirements to run the project

3.1 Following versions of packages needed for implementing the code:

- beautifulsoup4==4.10.0
- bs4==0.0.1
- certifi==2021.10.8
- charset-normalizer==2.0.7
- cycler==0.11.0
- edx-dl==0.1.13
- fonttools==4.28.2
- greenlet==1.1.2
- html5lib==1.1
- idna==3.3
- kiwisolver==1.3.2
- matplotlib==3.5.0
- numpy==1.21.4
- packaging==21.3
- pandas==1.3.4
- Pillow==8.4.0
- pyparsing==3.0.6

Note: The pip install commands are already included in the code using the os commands. Hence, this shouldn't be a concern.

Requirements to run the project

3.2 Following versions of packages needed for implementing the code:

- pyparsing==3.0.6
- python-dateutil==2.8.2
- pytz==2021.3
- requests==2.26.0
- ruamel.yaml==0.17.17
- ruamel.yaml.clib==0.2.6
- scipy==1.7.3
- seaborn==0.11.2
- setuptools-scm==6.3.2
- six==1.16.0
- soupsieve==2.3.1
- SQLAlchemy==1.4.27
- squarify==0.4.3
- tabulate==0.8.9
- tomli==1.2.2
- urllib3==1.26.7
- webencodings==0.5.1
- youtube-dl==2021.6.6

Note: The pip install commands are already included in the code using the os commands. Hence, this shouldn't be a concern.

Steps to run the project

3.1 Go to the code folder

3.2 Open command prompt in this folder location i.e. in 'code' folder

3.3 Run either of the following commands (Snapshots in the next slides):

--> python3 HW5_Final_Project.py

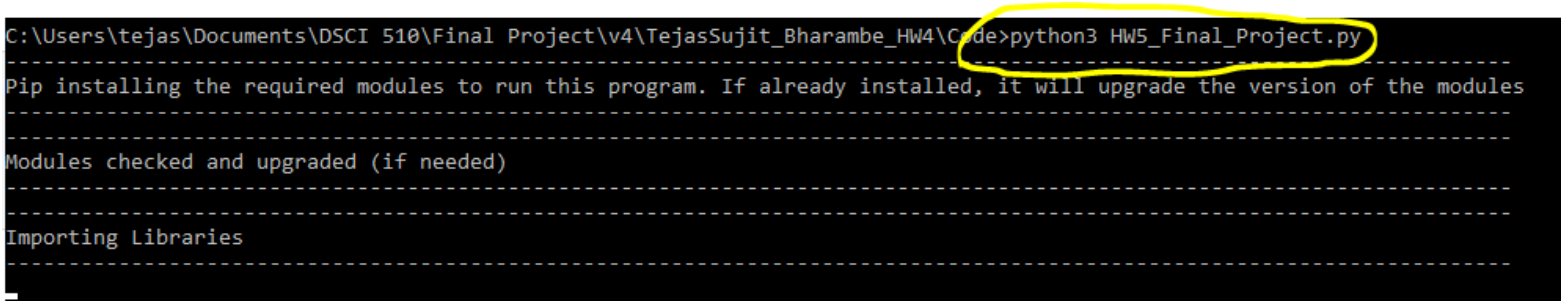
--> python3 HW5_Final_Project.py --static

Ways to run the code

Using the following command:

python3 HW5_Final_Project.py

This will run the analytics based on the csv data stored in the Data folder. The csv data in the Data folder is the already scrapped data.

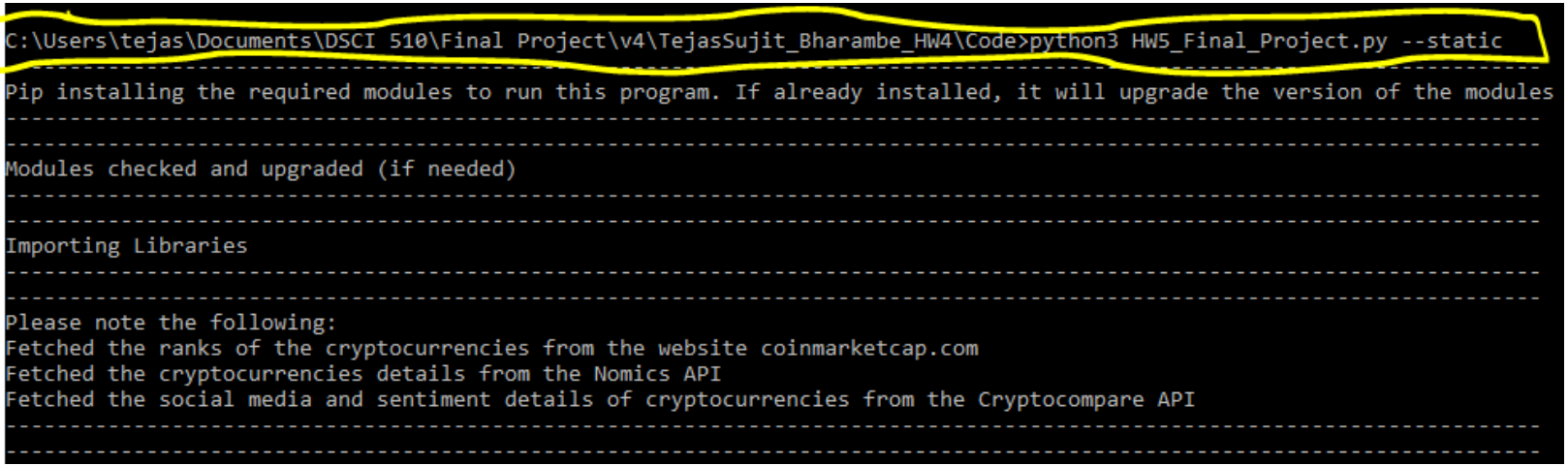


Ways to run the code

Using the following command:

python3 HW5_Final_Project.py --static

This will run the analytics based on the csv data stored in the Data folder. The csv data in the Data folder is the already scrapped data.



Extensibility of code

The output from this code can be used directly to draw some additional insights, analytics and conclusions from the data

Some modules of the code can be further used for cleaning, grouping and even extracting the new data from website/API if needed

Maintainability of code

The crytocurrency data that my code uses is updated at each second. Hence, there might be some ambiguity which might arise while running the code. However, I have done as much exception handling as possible.

Because the data is based on the current price, the data that you might see may vary in the next run.

The code is purely dependent on the API and website's information. If website crashes or the website changes its format of storing the data/HTML changes.
This will further impact the code. As long as the website and APIs are not updated drastically, the code should run correctly

Code is also dependent on the API Key which is assigned to me. But given this is a public API, it should work

Thanks!

Appendix

Nomics API Data Snippets

Website - https://nomics.com/docs/

Currencies Ticker


```
"currency": "BTC",
"id": "BTC",
"status": "active",
"price": "8451.36516421",
"price_date": "2019-06-14T00:00:00Z",
"price_timestamp": "2019-06-14T12:35:00Z",
"symbol": "BTC",
"circulating_supply": "17758462",
"max_supply": "21000000",
"name": "Bitcoin",
"logo_url": "https://s3.us-east-2.amazonaws.com/nomics-api/stat
"market_cap": "150083247116.70",
"market_cap_dominance": "0.4080",
"transparent_market_cap": "150003247116.70",
"num_exchanges": "357",
"num_pairs": "42118",
"num_pairs_unmapped": "4591",
"first_candle": "2011-08-18T00:00:00Z",
"first_trade": "2011-08-18T00:00:00Z",
"first_order_book": "2017-01-06T00:00:00Z",
"first_priced_at": "2017-08-18T18:22:19Z",
"rank": "1",
"rank_delta": "0",
"high": "19404.81116899",
"high_timestamp": "2017-12-16",
"1d": { … }
```


Currencies Metadata


```
"id": "BTC",
"original_symbol": "BTC",
"name": "Bitcoin",
"description": "Bitcoin cryptocurrency",
"website_url": "https://bitcoin.org/en",
"logo_url": "https://s3.us-east-2.amazonaws.com/nomics-api/stat
"blog_url": "https://bitcoin.org/en/blog",
"discord_url": "https://discord.gg/bitcoin",
"facebook_url": "https://www.facebook.com/bitcoin",
"github_url": "https://github.com/bitcoin",
"medium_url": "https://medium.com/bitcoin",
"reddit_url": "https://www.reddit.com/r/bitcoin",
"telegram_url": "https://t.me/bitcoin",
"twitter_url": "https://twitter.com/bitcoin",
"whitepaper_url": "https://bitcoin.org/en/bitcoin-paper",
"youtube_url": "https://www.youtube.com/channel/bitcoin",
"linkedin_url": "",
"block_explorer_url": "",
"bitcointalk_url": "",
"replaced_by": null,
"cryptocontrol_coin_id": "bitcoin",
"used_for_pricing": true
```


Currencies Markets & Market Cap History


```
[
- {
    "exchange": "binance",
    "market": "XMRBTC",
    "base": "XMR",
    "quote": "BTC"
  }
]
```



```
"timestamp": "2018-07-05T15:00:00Z",
"market_cap": "269222501959",
"transparent_market_cap": "270122591259"
```

CryptoCompare API Data (Keys) Snippets

Website - https://min-api.cryptocompare.com/

Toplists

- Toplist by 24H Volume Full Data
- Toplist by 24H Top Tier Volume Full Data
- Toplist by Market Cap Full Data
- Top Exchanges Volume Data by Pair
- Top Exchanges Full Data By Pair
- Toplist by Pair Volume
- Toplist of Trading Pairs
- Exchange Top Symbols By Volume

Social Data

- Latest Coin Social Stats Data
- Historical Day Social Stats Data
- Historical Hour Social Stats Data

News

- Latest News Articles
- List News Feeds
- News Article Categories
- List News Feeds and Categories

Order Book

- Exchanges With Order Book Data
- Order Book L1 Top
- Order Book L2 Snapshot

$ Price (inc. CCCAGG)

- Single Symbol Price
- Multiple Symbols Price
- Multiple Symbols Full Data
- Generate Custom Average

Historical (inc. CCCAGG)

- Daily Pair OHLCV
- Hourly Pair OHLCV
- Minute Pair OHLCV
- Day All Pairs OHLCV (CSV)
- Minute Pair OHLCV (CSV)
- Day Pair OHLCV by TS
- Day Pair OHLCV (Hourly VWAP) by TS
- Daily Exchange Vol.
- Hourly Exchange Vol.
- Daily Symbol Vol.
- Hourly Symbol Vol.
- Daily Symbol Vol. Single Exchange
- Hourly Symbol Vol. Single Exchange

Coin Market Cap Website Snippets

Website - https://min-api.cryptocompare.com/