



**Data**

I am finished with my project, unless there is something you see that you think I should change.

In picture one, the top half shows all three of my tables for the database “historical\_data.db.”

On the bottom half, the middle and right windows show all two of my cache files.

The reddit files are ONLY cached. Whether the file updates or not is dependent on if you requested the data on the same day (harder to do than I initially expected). I did not db them, as I did not see it necessary.

The Coindesk files are cached and databased. These articles are less likely to be updated, it’s mainly an archive newsite. Since there isn’t a lot of new content for each coin on this site, I did not make it date-dependent like for reddit.

Essentially every coin is loaded on the “Coin” table (1500+).

There is also 1-month historical price data that gets inserted the first time a user requests it.

**Interface**

The bottom-right photo is the main page of my program. It consists of a search bar where the user can enter a coin (capitalizations matter). There are also about 200 coins below it that a user can click on if they are just pursuing.

The bottom photo is what loads when a coin is searched. It consists of four data-types:

1. Current information- Right below the name presents the current price, 24hr change, and rank. This information is not cached or databased since they rely on the most up-to-date information.
2. 1-month chart. This uses Plotly to chart the last 30 days, depending on when the information is initially requested.
3. Latest Reddit News- This pulls the top articles within the hour under the coin’s subreddit
4. Coindesk Articles- This pulls articles from either the database or coin desk directly depending on if it has been requested before.

**Bug Concerns/Error Handling**

Since I pull different information from multiple sources, sometimes there won’t be any articles for it. This happens often for less popular coins. In this case the articles are left empty, but the graph and current info still exists. This happens to approximately 50% of coins.

Some coin names are so uniquely spelled that each website handles the name differently in their calls, and it’s impossible to track all 1500+ of these names, so in this case- If I cannot pull numerical data, It goes to a custom 404 page. This only happens to approximately 5% of coins.

**Changes to proposal**

I will not be collecting data from twitter anymore. I will also only present one graph. Coindesk will be the only critic-based site that I scrape from. There will be no csv caching. Additionally, Reddit does not require OAuth. They only require that you provide a reddit username when scraping data.

Comments are appreciated.

Thanks.