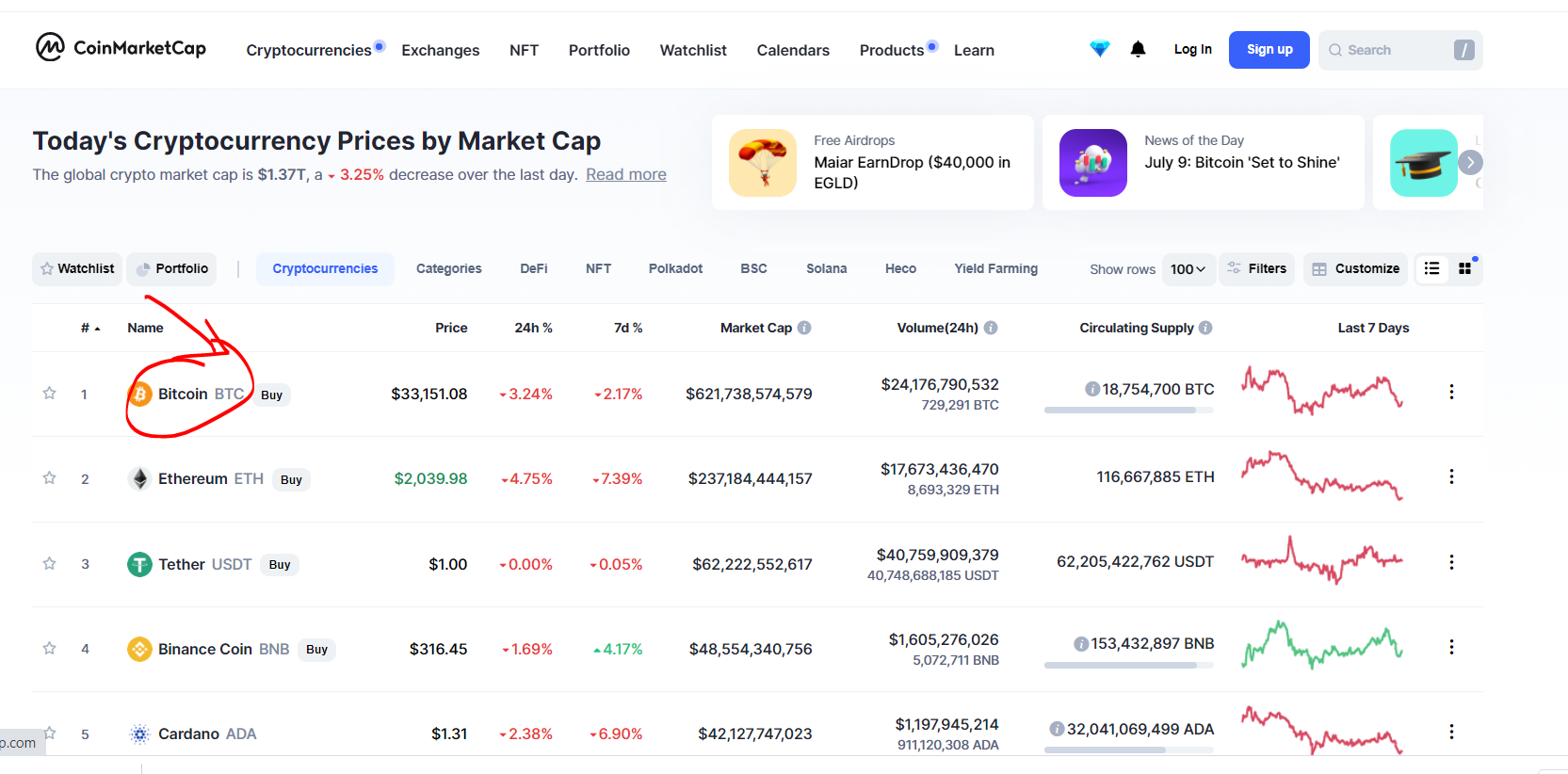
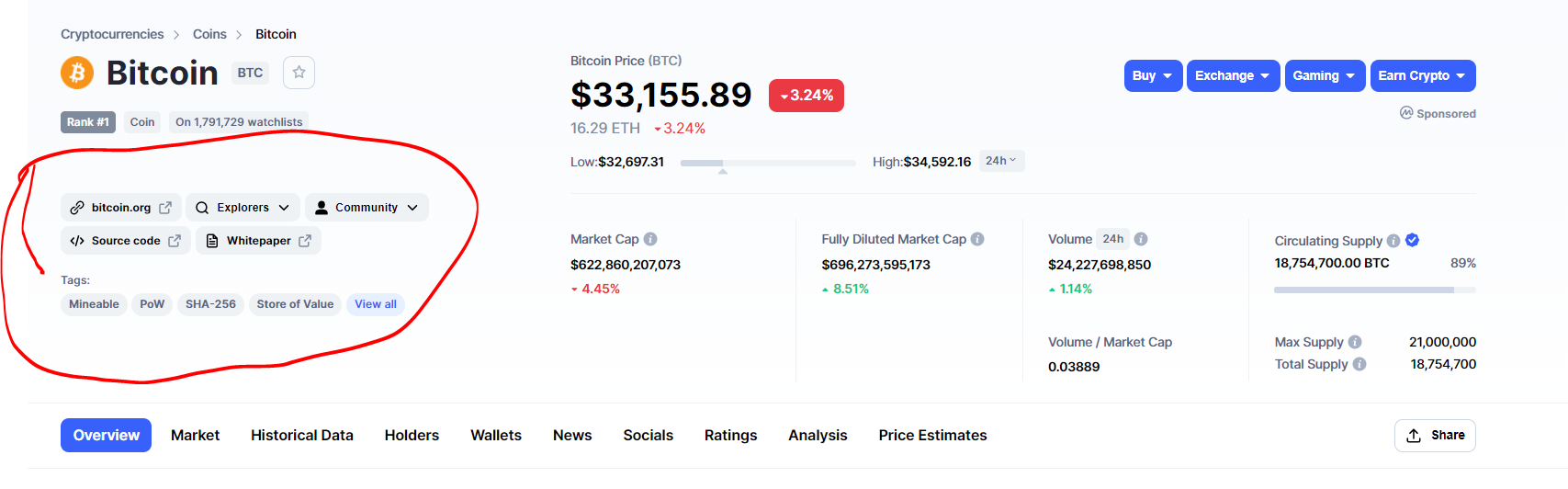
Hi Team,

Thanks for your help regarding scraping the Coinmarketcap price archival data. Now I need your help to get more detailed information about each coin.

For each coin, you can click the link to check the detailed info about each coin.

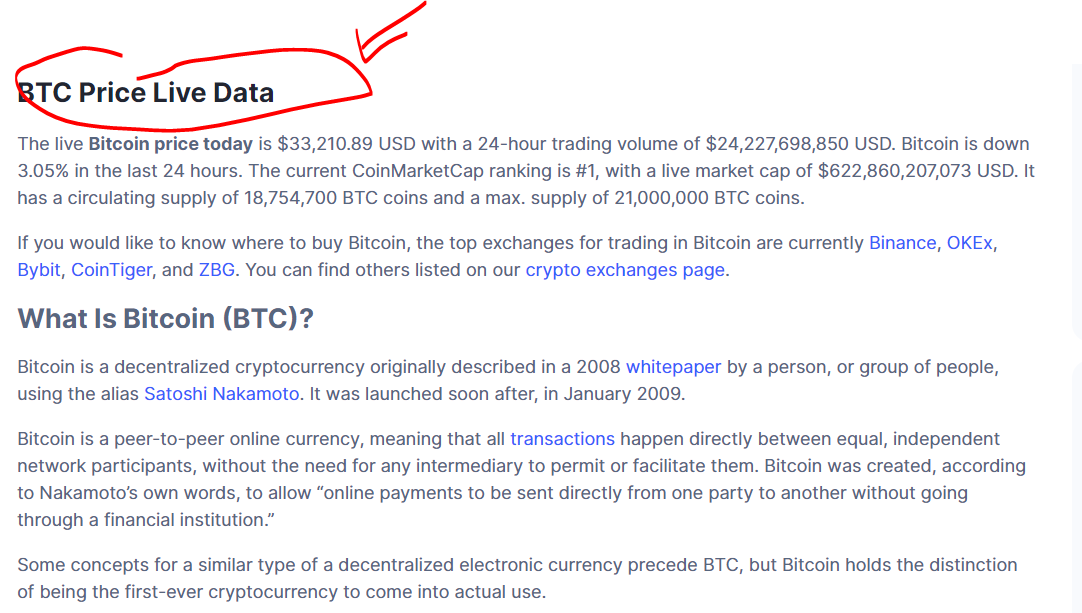




E.g., regarding bitcoin, we want to save the following data:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Crypto | link | Source Code | Whitepaper | explorers | Community | Tags | Description | audit | Contracts: |
| BTC | https://bitcoin.org/ | https://github.com/bitcoin/ | https://bitcoin.org/bitcoin.pdf | <https://blockchain.coinmarketcap.com/chain/bitcoin>; <https://blockchain.info/>; <https://live.blockcypher.com/btc/>; <https://blockchair.com/bitcoin>; <https://explorer.viabtc.com/btc> | <https://bitcointalk.org/>; <https://reddit.com/r/bitcoin> | Mineable, PoW, SHA-256, … | XXX XXX | CertiK, StowMist |  |

When we save the description, we may want to remove all the special characters. For example, we can directly save the tokenized descriptions. (If we don’t do this, there could be some format issue).



Note: We already have the basic codes for you guys to start with.

The notes by previous student volunteers.:

***I was not able to completely overcome the slight format change from page to page****. Whatever results I have been able to scrape are in D:\Vasudev-Jingbo-Peiyu\data\Coinmarket details\details.csv. I have used '|' (pipe) as delimiter instead of ',' (comma) since there were a lot of commas in the description and the json representation of contracts.*

*contract\_string - [{"Name": "Ethereum", "Link": "https://...", {"Name": "Binance Smart Chain", "Link": "..."}]. This can be imported like json.loads(contract\_string)*

*The code for everything is in D:\Vasudev-Jingbo-Peiyu\src\Scraper. Let me know if you have any questions.*

***So the changes that were hard to fix were related to Explorer, Community. Sometimes they are plain objects with no dropdown, other times they have multiple items within them. There is also an issue with finding the 'X' button in the View All popup window for tags.*** *I am not able to uniquely identify it. Sometimes that seem to be links which are similar that open up new tabs. I think this also comes from the Source.*

***The latest updates:***

***All prior students cannot use code to get the detailed information of the top 100 cryptos.***

***Shuai, could you give it a try?***

***Sreeram manually got 68 out of top 100 (cryptoDetails\_sample01\_10.xlsx) . If we can not use the code do it, manually get all the information is not a bad call.***

***However, if we can use the code to get it, we can get the all cryptos’ info or top 2000 crypto’s info.***

***Crowd wisdom… …***

***NLP: sentiment polarity, sentiment subjectivity,***

***Time series: RNN, LSTM, ARMA, AR, MA, PANLE VAR,***

***Panel Regression***