

UNIVERSITY OF WARSAW

WEBSCRAPPING AND SOCIAL MEDIA SCRAPPING

**Ensuring the quality of the data set and
combining it with the situation of the trading
industry for cryptocurrencies.**

Author

Mateusz Baryła 427778

Supervisor

Anna Lewczuk
Przemysław Kurek

May 8, 2021



1 Organisation

This project will be covered by myself only.

2 Project Goals

Taking into account the business perspective there is a need to investigate the current perspectives of cryptocurrencies industry. The main page used for scrapping is <https://conotoxia.com/cryptocurrencies/cryptocurrency-rates>. This the desired scrapping outcome.

Currency	Average rate	Market value	Exchange (24h)
Bitcoin	\$ 57 260.00 USD	1 070 923 584 403.00 USD	1.11 %
Bitcoin	\$ 57 260.00 USD	1 070 923 584 403.00 USD	1.11 %

3 Scraper mechanics and technical description

Each of three scrapers use very similar mechanism that allows for gathering data in the robust way. We can specify the number of pages that limit scrapers. At the end file is saved in csv format for soup and selenium. Taking into account scrapy we specify file name in command line.

4 Data analysis

The important part for each data science project is of course exploratory data analysis. Keeping in mind the data analysis part that is one page limit we will address the issue of limited resource provided.

First rows

	currency	average_rate	market_value	volume	exchange
0	1 Bitcoin (BTC)	57821.000	1.082933e+12	6.755888e+10	3.99
1	2 Ethereum (ETH)	3528.470	4.089562e+11	4.642398e+10	2.23
2	3 Binance Coin (BNB)	633.040	9.802898e+10	4.644952e+09	0.97
3	4 Dogecoin (DOGE)	0.642	8.294119e+10	2.357125e+10	12.11
4	5 XRP (XRP)	1.600	7.415394e+10	1.054338e+10	1.22
5	6 Tether (USDT)	1.000	5.494157e+10	2.011920e+11	0.21
6	7 Cardano (ADA)	1.670	5.368205e+10	8.288167e+09	7.60
7	8 Polkadot (DOT)	41.040	4.073998e+10	2.749088e+09	1.91
8	9 Bitcoin Cash (BCH)	1360.240	2.558066e+10	1.203662e+10	-5.34
9	10 Litecoin (LTC)	345.960	2.318772e+10	1.410422e+10	4.98

