Overview of analysis

An outline of how I explored the relationship between cryptocurrency prices and Twitter followers of influential figures in crypto.

(GitHub project link)

Background

- Cryptocurrency prices influenced by many factors
- Twitter serves as the online “town hall” of crypto
- Academic research has found evidence that Twitter activity and crypto prices are interrelated
  - [1] Relationship exists between crypto prices and general Twitter sentiment
  - [2] Machine learning and sentiment analysis can predict crypto markets
  - [3] Price direction and magnitude of price change can be predicted by analyzing tweet volumes and sentiment with neural network models
- “Elon effect” has been described in popular media
  - Elon Musk’s tweets cause documented shifts in crypto prices

Issues with prior analyses

- Too specific or too general
  - Many analyses focus on only a single cryptocurrency (e.g. Bitcoin)
  - Focus on all tweets or all tweets related to crypto rather than a more meaningful subset
- Leave questions unanswered
  - Don’t discuss whose tweets have the most impact on crypto markets
  - Analysis using neural networks and machine learning can be a “black box” – we don’t know which specific factors are important
- Lack of actionable conclusions
  - Most conclusions focus on the effectiveness of analysis methods rather than wider impact or use cases

Premise

- By analyzing broader and more meaningful datasets on Twitter and cryptocurrency prices, we can attempt to:
  - Build a broader picture of nuanced relationships
  - Answer questions beyond current understanding
  - Generate actionable insights

Research questions

1. How is the popularity of influential figures in crypto related to cryptocurrency prices?
2. Who has the most influence on/is most influenced by the price of each cryptocurrency?
3. Can we estimate crypto prices based on Twitter followers?

Choosing datasets

- How can we measure Twitter influence?
  - Total followers
  - New followers per day
- How can we best measure cryptocurrency prices?
  - Prices of top cryptocurrencies by volume over the period we are analyzing
  - Not including stablecoins – they are poor indicators of price fluctuation

Extracting raw data – Twitter followers

- Finding list of of influential figures in crypto
  - I used Cointelegraph's 2022 list of the 100 most influential people in crypto
    - I ignored those with no Twitter presence
  - Full list of Twitter usernames in "twitter-list.txt"
- Scraping data from Social Blade using Python
  - As Twitter does not provide follower counts over time, I wrote a script in Python to scrape total follower counts and new follower counts from 2022/08/15 to 2022/09/11 from socialblade.com
    - I did not include users with not enough followers to be tracked by Social Blade

Extracting raw data – crypto prices

- Finding list of top cryptocurrencies by volume
  - I used CoinMarketCap's list and found the top 10 cryptocurrencies by transaction volume (not including stablecoins) during the period analyzed
  - Full list of cryptocurrencies in "crypto-list.txt"
- Scraping data from investing.com using Python
  - My Python script scraped price data from 2022/08/15 to 2022/09/11 for each cryptocurrency from investing.com

Extracting raw data – output

- Python libraries I used for scraping:
  - Selenium for browser automation
  - Beautiful Soup for parsing HTML
- Python script returns scraped data output as:
  - Python pandas dataframe via pickling
  - CSV file
  - Table output to SQL database
- For full details, see “scraper.py” script

Overview of time series data extracted

- All prices and followers recorded for 28 days from 2022/08/15 to 2022/09/11
- Total followers
  - Total followers on each day recorded for 83 Twitter accounts
- New followers
  - Change in followers on each day recorded for 83 Twitter accounts
- Crypto prices
  - Price in USD on each day recorded for 10 cryptocurrencies

Choosing data analysis/visualization methods

- Choosing between analysis/visualization tools I have experience in: Excel, Tableau, and Python's Matplotlib library.
- Matplotlib seems to be the best choice
  - Excel is not very effective for complex visualizations
  - Tableau's strong point is interactivity, which I don't feel is necessary
    - With so many Twitter accounts and cryptocurrencies, I feel it's better to guide the user through the data with more customized plots
  - I extracted my data using Python, and so it makes sense to use Matplotlib for better integration

Data analysis in Python using Matplotlib

- I wrote another script to analyze and visualize data in Python using the Matplotlib library
  - Script manipulates dataframes, generates plots, and saves them as images
- Data was imported from the pickled pandas dataframes I scraped earlier
- For full details, see “matplotlib-analysis.py” script

Findings

- See “Figures” folder for graphs and plots I generated
- See “Overview of findings” slides for a summary of my observations