SAN DIEGO STATE UNIVERSITY

Analyzing Twitter Cryptocurrency List URL's To Detect Warning Signs Of Malicious Activity

Terrance Hutchinson
Computer Science Department
San Diego State University

Introduce problem

- There are hundreds of thousands of lists related to cryptocurrencies.
- These lists can be for promotional purposes, malicious scams, or entirely fake.
- How can the URLs in the descriptions of these indicate the intentions of the user that created the cryptocurrency related twitter list?

Background

Twitter is filled with all sorts of information. In recent years, we have learned that Twitter is filled with disinformation and scams. Twitter is a great way to stay up-to-date on what's happening and connect with people all over the world, but it's also important to learn what you can do to protect yourself on Twitter. Since there's a wealth of misinformation and scams out there, one of the best ways to protect your finances . Especially, when it concerns digital assets like cryptocurrency.
By investigating the URL links within the descriptions of cryptocurrency related Twitter lists we can learn indicators that can identify fraudulent lists that are attempting to mislead user to phishing websites or scam crypto projects.

Problem Motivation

In recent years, the news headlines have been filled with stories of cryptocurrency related scams or hacks. The recent hack of the cryptocurrency exchange, Bitfinex, and the fraudulent activities of Sam Bankman-Fried with his company FTX has caused many people to question whether or not cryptocurrencies are legitimate investment vehicles or whether they should be overlooked entirely. It is important to distinguish the difference between a legitimate investment opportunity and a fraudulent one. The everyday user of Twitter could easily fall into a trap by falling victim to a scam because the perpetrators of these schemes are able to easily fool the average person due to their lack of experience in cryptocurrency and blockchain technology. Scammers often play on the average persons propensity to herd mentality and the idea of "fear of missing out" to entice them into making an investment that they know is a scam.
By investigating the the URL links that are contained within these cryptocurrency related Twitter lists a realistic idea of legitimacy can be determined. The data pulled from the URL links can provide the everyday Twitter user with useful insights on the reliability of these links that could potentially be a scam.

Approach

- Connect to the Twitter API
- Pull data from Twitter API related to cryptocurrency list
- Data pulled from the Twitter API include:
  - List ID (int), List Name (string), List Description (string), List Owner ID (int), List Creation Date/Time (date/time), List Follower Count (int)
- Parse data that was pulled from Twitter API. In particular, the description field where the URL links are listed.
- Clean the URL link data by parsing the URL link and formatting them correctly to be requested.
- Parse the HTTP response header data:
  - URL (link), Status Code (string), Content Type (string), Date/Time of HTTP Request (date/time), Strict Transport Security (string), Cache Age (int), Content Security Policy (string), X-Content-Type-Options (string), Request History (string), Error (string), Error Code (string)
- Export all data combined to a CSV file.
- Perform data analysis on data. In particular, the Status Codes, Follower Counts, Strict Transport Security, Content Security Policy, and X Content Type Options.

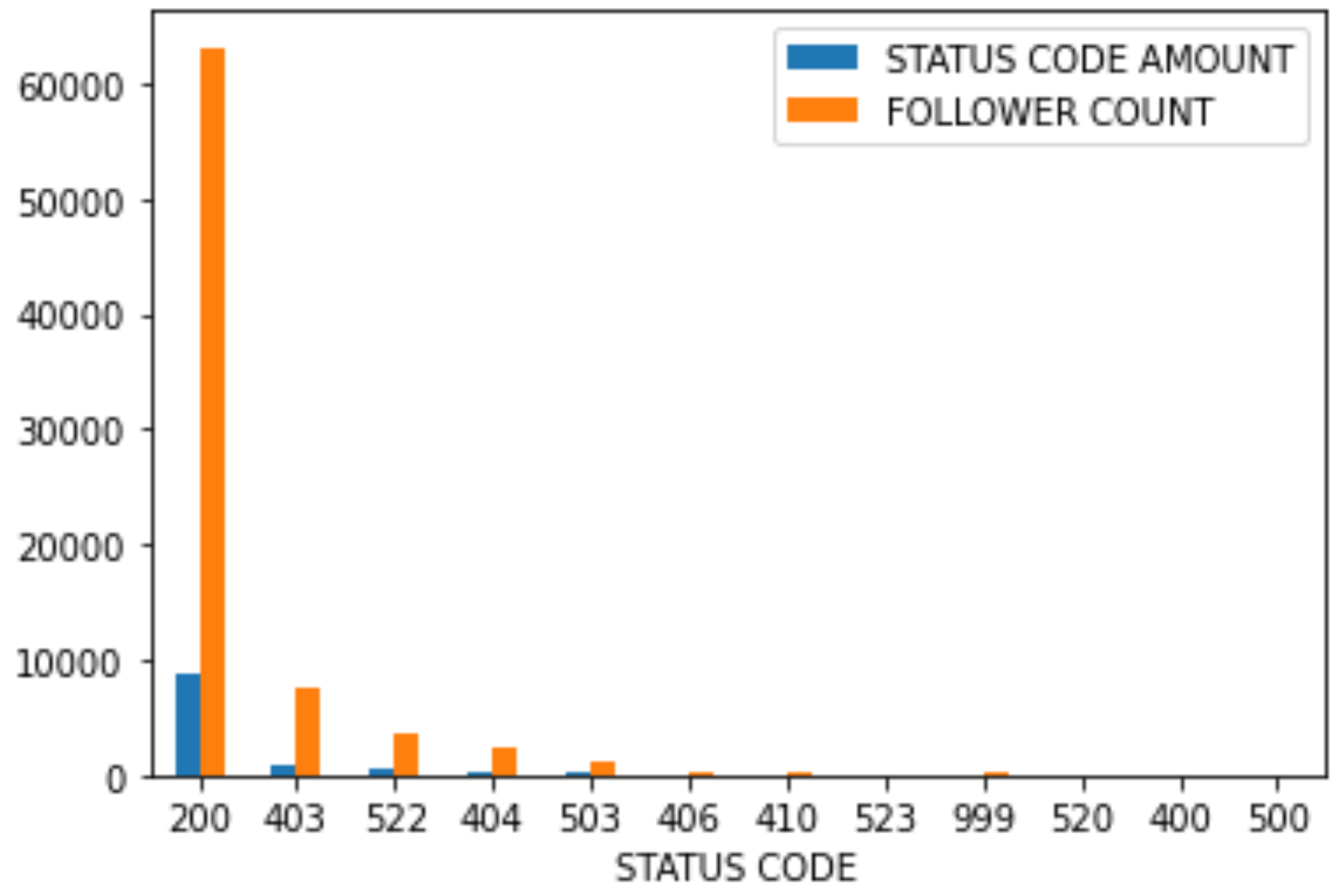
Approach details

- Create a python program that continuously pulls the cryptocurrency Twitter list information utilizing the Twitter API.
- Create a python program that parses the Twitter data for URL links, makes requests on the URL links, parses the HTTP Header data, and exports all data to CSV file.
- Utilizing these python libraries:
  - URLExtract, Requests, URLParse, CSV, Datetime
- Investigate which HTTP Response Headers with pandas to how many URL links indicate malicious behavior or are fake.
- The status codes indicate what HTTP connection the URL links are.
- Strict Transport Security enforces the use of encrypted HTTPS connections instead of plain-text HTTP communication. The preload directive indicates the site is present on a global list of HTTPS only sites. The purpose of preloading is to speed up page loads and eliminate the risk of man in the middle attacks when a site is visited for the first time.
- Content Security Policy lets you precisely control permitted content sources and many other content parameters and is recommended way to protect your websites and applications against XSS attacks.
- X-Content-Type-Optoins s specifically intended to protect websites from cross-site scripting attacks that abuse MIME sniffing to supply malicious code masquerading as a non-executable MIME type. Indicated by the directive "nosniff".
- The understanding of how safe these URL links are can be represented by analyzing the frequency of these security headers from the URL links provided in the Twitter list descriptions.
- These security headers are intended to keep both the user and site safe from malicious activity, so if the lack of them would indicate malicious behavior.

Evaluation

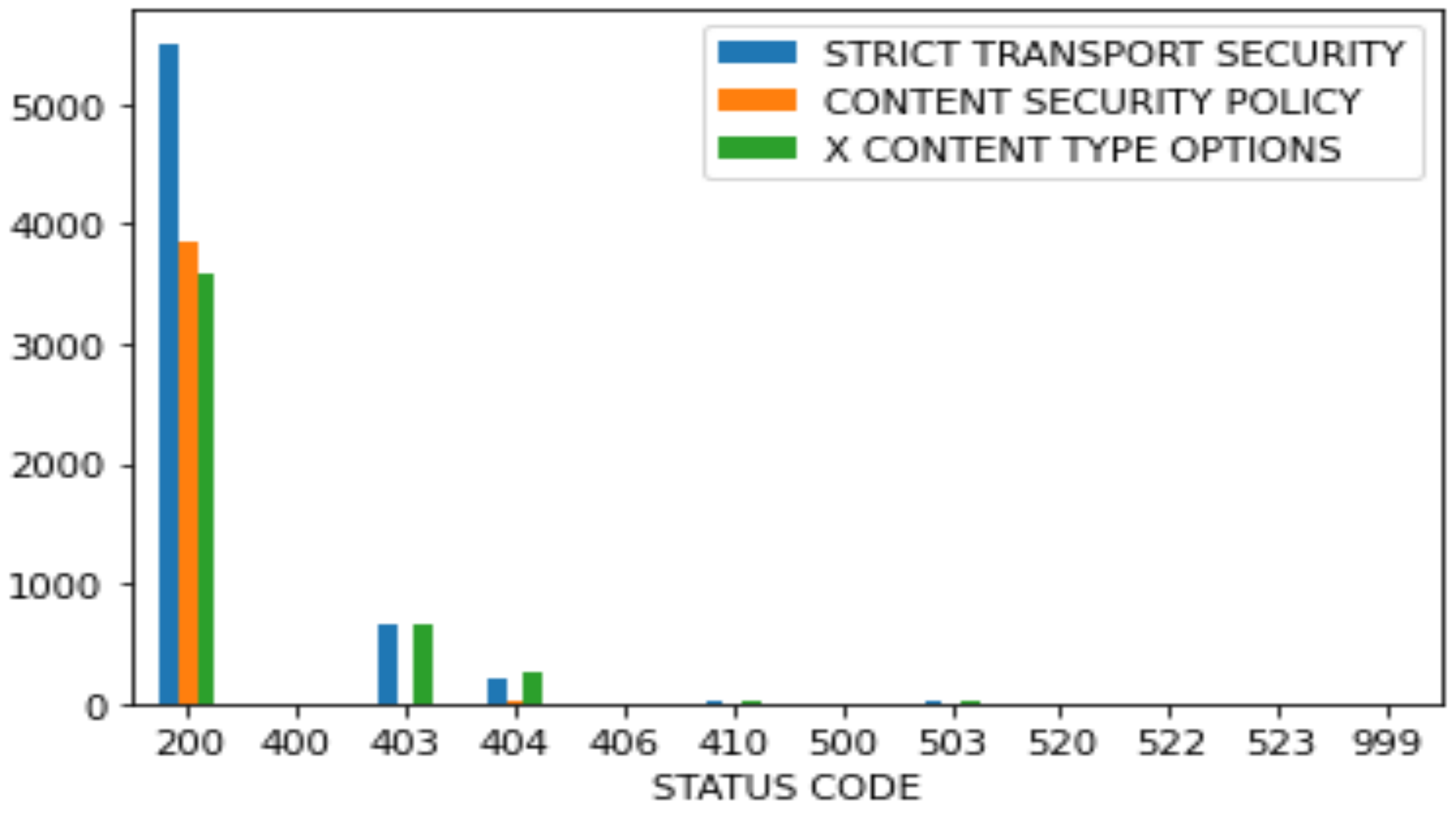
First, the status code of every URL analyzed. There were a total of 10,701 links with 78,447 total followers. The HTTP status code returned were:

- **200 OK**: The request succeeded. 8,701 of the links were 200 status code with 63,145 followers.
- **403 Forbidden** : 907 of the links were returned with the 403 status code with 7680 followers.
- **522 Connection Timed Out:**. 566 of the links were 522 status code with 3664 followers.
- **404 Not Found:** 314 of the links were 404 status code with 2311 followers.
- **503 Service Unavailable:** 128 of the links were 503 status code with 1168 followers.
- **406 Not Acceptable:** 42 of the links were 406 status code with 152 followers
- **410 Gone:** 18 of the links were 410 status code with 151 followers.
- **523 Origin is Unreachable:** 11 of the links were 523 status code.
- **999 Request Denied:** 7 of the links were 999 status code with 123 followers.
- **520 Web Server Returned an Unknown Error:** 5 of the links were 520 status code with 15 followers
- **400 Bad Request:** 1 of the links were 400 status code with 0 followers.
- **500 Internal Server Error:** 1 of the links were 500 status code with 2 followers.



This graph is of importance because it shows that 19% of the links are unreachable. Indicating that the 19% of URL links relating to cryptocurrency on Twitter lists are unusable. Leaving 19.5% of followers following lists with unusable links.

Results



This graph indicates that which status code responses most frequently follow crucial HTTP security protocols that ensure both safety for the URL link and the the users of the URL link. For the analysis of this data ,only the 200 status code data will be considered because those are the only sites However, its important to include the other status codes because it indicates that those unreachable servers practice safe HTTP security protocols, possibly due past promotions of credible URL links. Twitter users can reach.

- The 200 status code has 5,510 URL links that follow the strict transport security protocol which accounts for 63.33% of 200 status code URL links. Strict transport security informs any visiting web browser that the site and all its subdomains use only SSL/TLS communication. This helps eliminate the risk of man-in-the-middle attacks on users.
- The 200 status code has 3,845 URL links that follow content security policy which accounts for 44.19% of 200 status code URL links. Content security policy protects websites and applications from XSS attacks.
- The 200 status code has 3,585 URL links that follow x-content-type-options which accounts for 41.20% of 200 status code URL links. X-content-type-options are to protect users and websites from cross-site scripting attacks that abuse MIME sniffing to supply malicious code masquerading as a non-executable MIME type.

Conclusions

From the data of the URL links that are provided within Twitter cryptocurrency lists it can be concluded that 36.67% did not practice strict transport security protocols, 55.81% did not practice content security policy protocols, and 58.80% did not practice x-content-type-options security protocols. As a result, this indicates that a large majority of the links are unreliable or unsafe for everyday users to use. This is something that the general everyday user of Twitter would not know and could be easily compromised because of it.

References

[1] Zbigniew Banach, 2022, HTTP security headers: An easy way to harden your web applications
, Venue, https://www.invicti.com/blog/web-security/http-security-headers/
[2] Xiao-Lei Liu, 2021, HTTP-Based APT Malware Infection Detection Using URL Correlation Analysis
, Hindawi, https://www.hindawi.com/journals/scn/2021/6653386/
[3] MDN Web Docs, 2020, HTTP response status codes, MDN Web Docs, https://developer.mozilla.org/en-US/docs/Web/HTTP/Status - see_also

Source Code:
(Please request for access)
https://colab.research.google.com/drive/1UBL_Gx91_xljFrLmpKiGakQb11M5fJNZ?usp=sharing
https://drive.google.com/drive/folders/1M8_aGCEMtXQ1Y5gpMkwOgMgnOgf7zS8T?usp=sharing