

Weekly Update 1

Hien An Ngo

February 9, 2026

Weekly Progress (5 points)

1. Realized there is no inexpensive option to scrape X accounts using a third-party solution. Decided I would try to build my own scraper.
2. Tried to mimic a scraper I found online. Here's the tutorial: [YouTube](#).
3. Ran into several issues with the original script. The library the script has newer versions and I needed to modify a few things. I also ran into issues logging into X remotely, so I troubleshooted by using cookies to run the script instead of logging in with account information.
4. Was able to successfully scrape around 14,000 rows from former Rep. Marjorie Taylor Greene's personal profile. This took about four hours.
5. Currently working on scaling this to more profiles.

Challenges (3 points)

1. I had a hard time logging into X remotely using my original config file as shown in the tutorial.
2. Because of wait times and rate limits the script has a very long runtime depending on the contents of the profile.
3. I was only able to scrape about 14,000 posts, but her X account says she has about 18,900 in total. I'm not sure why I'm only getting about three quarters of posts.

Code Commits (2 points)

<https://github.com/GW-datasci/26-spring-ngo-a/commits/main/>

1. Committed X scraper
2. Committed script to clean/deduplicate csv and the cleaned csv

References

- [1] Mehran Shakarami. Twikit main.py. https://github.com/mehranshakarami/AI_Spectrum/blob/main/2024/Twikit/main.py, 2024. GitHub repository.
- [2] Twikit Documentation. Twikit documentation. <https://twikit.readthedocs.io/en/latest/twikit.html>, 2024.