

Scraping Twitter with Python

By: Darla Still

(slides borrowed from CSS 2019 Mini Conference)

Have you interacted with an API before?

Yes

No

I don't
know



A web API is designed to allow users to interact with the back-end data streams of a platform or service.

All web scraping requires a user to interact with an API.

True

False



The Twitter API allows for large-scale data collection but is rate-limited.

- It is different depending on which endpoint, or type of inquiry, you make.
- Twitter only gives two weeks back in time, limits the number of requests you can make in a given time period, and limits how much data you can get from those requests.
- It is also possible to scrape and bypass rate-limiting but it usually violates Terms of Service.

<https://developer.twitter.com/en/docs/basics/rate-limits.html>

Tweepy is the most popular Python Twitter API Tool.

- Tweepy is a pretty easy to learn package in Python that allows you to make a range of requests from the Twitter API.
- We will use Tweepy today!

<http://www.tweepy.org/>

Do you have the approved Twitter credentials for scraping as of today's workshop?

Yes

No

I don't
know



Each tweet is considered an object with scrapable attributes.

- Not all tweet attributes are available with the free developer account.

(line 94 of code)

```
outtweets = [[tweet.id_str, tweet.user, tweet.created_at,  
tweet.full_text, tweet.in_reply_to_screen_name,  
tweet.is_quote_status, tweet.source, tweet.lang, tweet.favorite_count,  
tweet.retweet_count, tweet.favorited, tweet.retweeted] for tweet in  
all_the_tweets]
```

<https://developer.twitter.com/en/pricing>

<https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object>

🗨️ When poll is active, respond at **PollEv.com/dstill283**

📱 Text **DSTILL283** to **37607** once to join

Any questions so far?



Let's scrape Twitter!

What is your general research interest and what handles are you planning to scrape?

Breakout rooms for a couple of minutes? Answer above question as icebreaker/community building.

Use this as a short break/pause while folks pull up necessary windows to begin scraping...