Ryerson University

TED ROGERS SCHOOL OF MANAGEMENT

SocialMediaLab.ca
"Making sense of a networked world"

Session 3:

Getting Started with Communalytic

Data Collection from Twitter : How to use Communalytic to collect data from a Twitter Thread and from the Twitter Academic Product Track

June 10, 2021

SocialMediaLab.ca

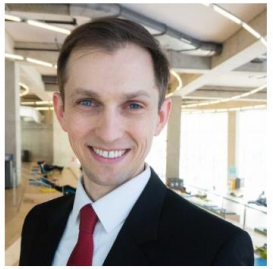
The Lab & The People

"Making Sense of a Networked World"

The Social Media Lab is a Multidisciplinary Research Laboratory at Ryerson University in Toronto, Canada

The lab studies how social media is changing the way people communicate, share information and form communities online, and how these changes impact society.

Instructors



Dr. Anatoliy Gruzd
Canada Research Chair, Associate Professor, Director of Research at the Ryerson University Social Media Lab

Philip Mai M.A., J.D.
Co-Director and Senior Researcher at the Ryerson University Social Media Lab

Socialmedialab.ca

Ryerson University | TED ROGERS SCHOOL OF MANAGEMENT | SocialMediaLab.ca "Making sense of a networked world"

Videos and slides from Session 1 & 2 are now available online at: communalytic.com

communalytic

FAQ TUTORIALS PUBLICATIONS CSS BOOTCAMP

APR 07 2021

Social Media Lab's Computational Social Science (CSS) Bootcamp – Summer 2021

Off

By COMMUNALYTIC

What: CSS Bootcamp on Examining Online Discourse & Networks with Communalytic

When: 2nd and 4th Thursday of the month at 10 am -11:30 am (ET) between May and July, 2021.

Where: Zoom (see details below)
Free Registration via Zoom

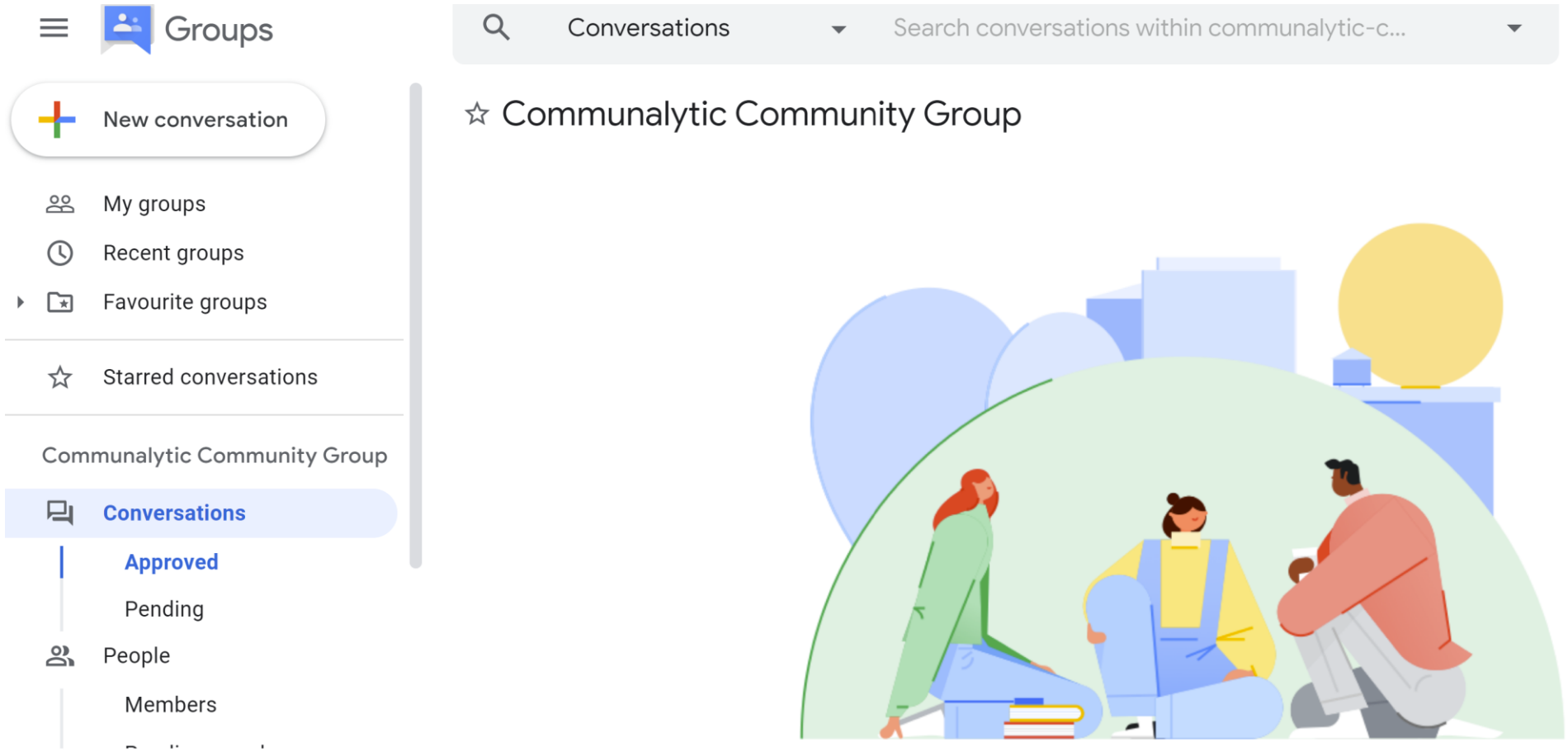
CSS Bootcamp Schedule Summer 2021

Session #1	**Getting Started with Communalytic: Data Collection from Reddit**	**May 13, 2021, 10:00-11:30am (EDT)**
Session #2	**Toxicity Analysis with Reddit Data using Perspective API**	**May 27, 2021**, 10:00-11:30am (EDT)
Session #3	**Getting Started with Communalytic: Data Collection from Twitter (Twitter Thread via API v2.0 and Twitter Academic Track)**	**June 10, 2021**, 10:00-11:30am (EDT)
Session #4	**Toxicity Analysis of Twitter data using Perspective API** Guest talk by Dr. Paola Pascual-Ferra, Assoc. Prof, Dept. of Communication at Loyola Uni. She will be presenting the results from a forthcoming study examining polarized discourse on Twitter about wearing face masks during the COVID-19 pandemic.	**June 24, 2021**, 10:00-11:30am (EDT)
Session #5	**Social Network Analysis of Signed Networks with Reddit and Twitter data**	**July 8, 2021**, 10:00-11:30am (EDT)
Session #6	**Getting Started with Communalytic: Data Collection from Facebook & Instagram (via CrowdTangle API) + Social Network Analysis of Two-mode Semantic Networks with CrowdTangle data**	**July 22, 2021**, 10:00-11:30am (EDT)

@SMLabTO

Join the Communalytic Community Group

https://groups.google.com/u/1/g/communalytic-community-group



Outline

About Communalytic

Twitter Thread Data Collection

(Edu & Pro)

How to request Twitter's Bearer authentication token

Twitter Academic Product Track

(Pro Only)

Two options to collect social media via APIs

Using a script - Requires programming knowledge & access to computing infrastructure

Python

Here is a sample request and response using the Python version of the Google API Client Libraries.

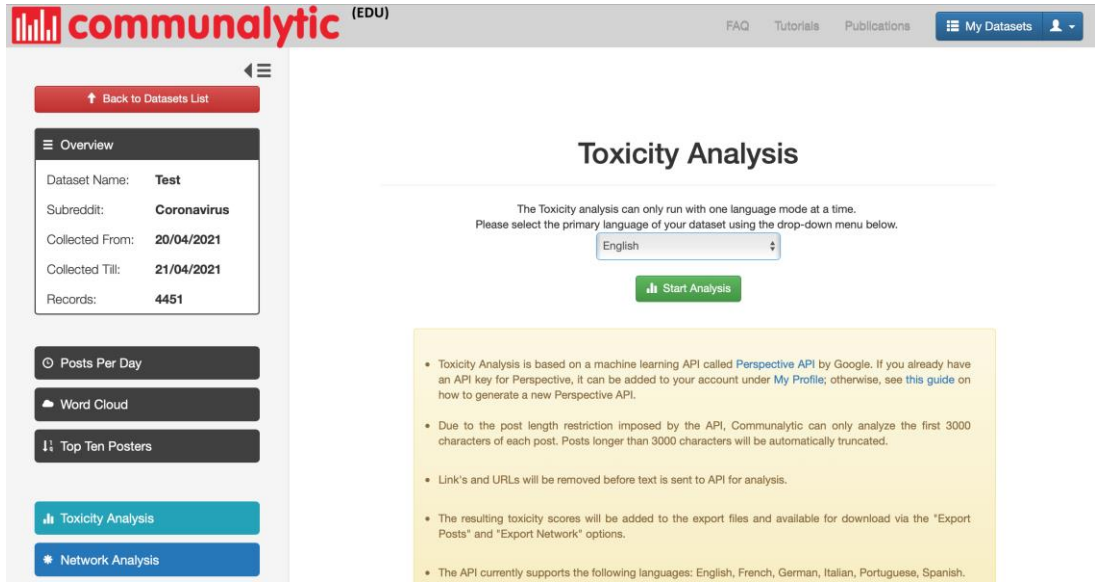
1. Install the Python client library.
2. Run the following commands:

Python | Copy

Docs > Sample Requests > Python

```
from googleapiclient import discovery
import json

API_KEY = 'copy-your-api-key-here'

client = discovery.build(
  "commentanalyzer",
  "v1alpha1",
  developerKey=API_KEY,
   discoveryServiceUrl="https://commentanalyzer.googleapis.com/$discovery/rest?vers
   static_discovery=False,
)

analyze_request = {
  'comment': { 'text': 'friendly greetings from python' },
  'requestedAttributes': {'TOXICITY': {}}
}

response = client.comments().analyze(body=analyze_request).execute()
print(json.dumps(response, indent=2))
```

Using a third-party system such as Communalytic – no programing knowledge required



Communalytic is a research tool for studying online communities and online discourse.

Communalytic can collect and analyze public data from social media platforms. It uses advanced text and social network analysis techniques to automatically pinpoint toxic and anti-social interactions, identify influencers, map shared interests and the spread of misinformation, and detect signs of possible coordination among seemingly disparate actors.

collect

extract

analyze

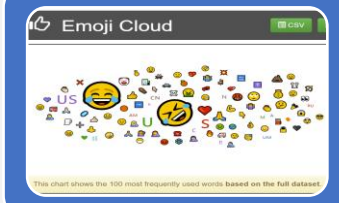
visualize

https://communalytic.com/

How to choose between Communalytic Edu and Pro.	communalytic EDU	communalytic PRO
Account Type	Free	$349/6-mo. to support site infrastructure (server-side data collection, storage, processing, analysis and visualization)
Designed For	**Students and is ideal for teaching and learning about social media analytics**	**Academic researchers and is ideal for large scale academic research projects**
Account Caps	≤ 30K records shared across 3 datasets	≤ 10M records shared across 50 datasets
Reddit	Live-collection* of public posts from any public subreddit for ≤ 7 consecutive days (Limit: ≤ 30K posts)	Live* & historical collection of public posts from any public subreddit for ≤ 31 consecutive days (Limit: Account Caps)
Twitter Threads (API ver.2) req. Twitter developer's account	Public replies to any public tweet posted within the previous 7 days (Limit: ≤ 30K tweets)	Public replies to any public tweet posted within the previous 7 days (Limit: ≤500K tweets/month)
Twitter Academic Track (API ver.2) req. Application to Twitter	Not supported	Full-archive historical-search of tweets back to 2006 (Limit: ≤ 10M tweets/month)
CrowdTangle (FB/IG) URL Search req. a CrowdTangle account	Public Facebook or Instagram posts that shared the same URL (Limit: ≤ 30K posts)	Public Facebook or Instagram posts that shared the same URL (Limit: Account Caps)

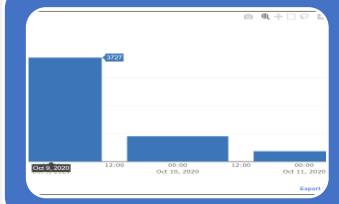
* Live-collection = the collection of posts/tweets posted on or after the date when you initiated the data collection.

communalytic EDU | communalytic PRO

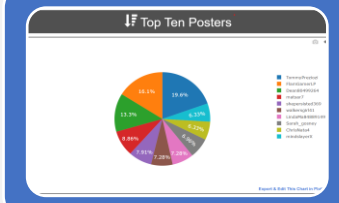


Emoji cloud (freq. used emojis)

Word cloud (freq. used words)

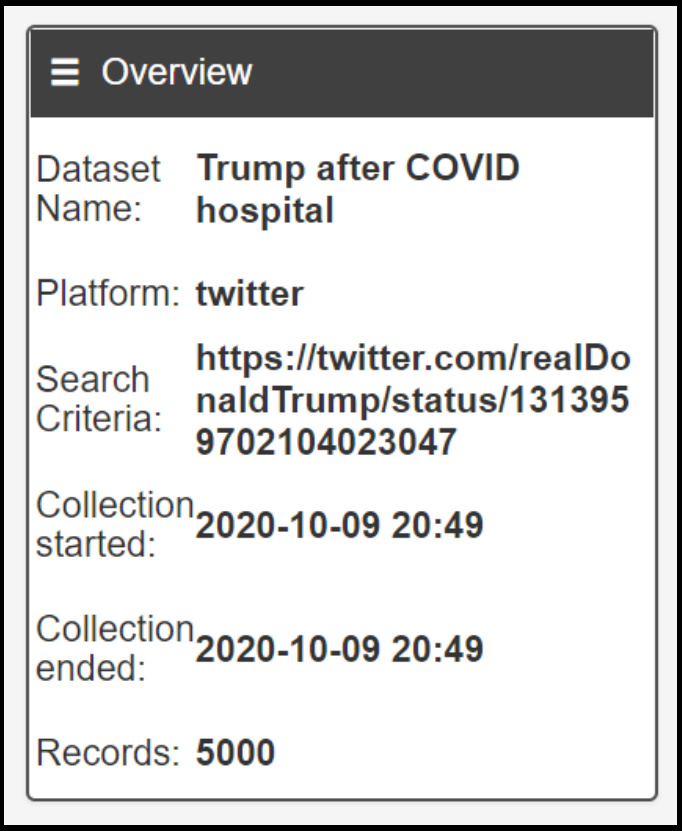


Time series (posts per day)



Top posters (top 10)

Communalytic Features: **Exploratory Data Analysis Module**



communalytic EDU | communalytic PRO

Print | Reset Analysis

There were **12** comments analyzed in **English**

Click on the highest/lowest values to see the top posts of each category

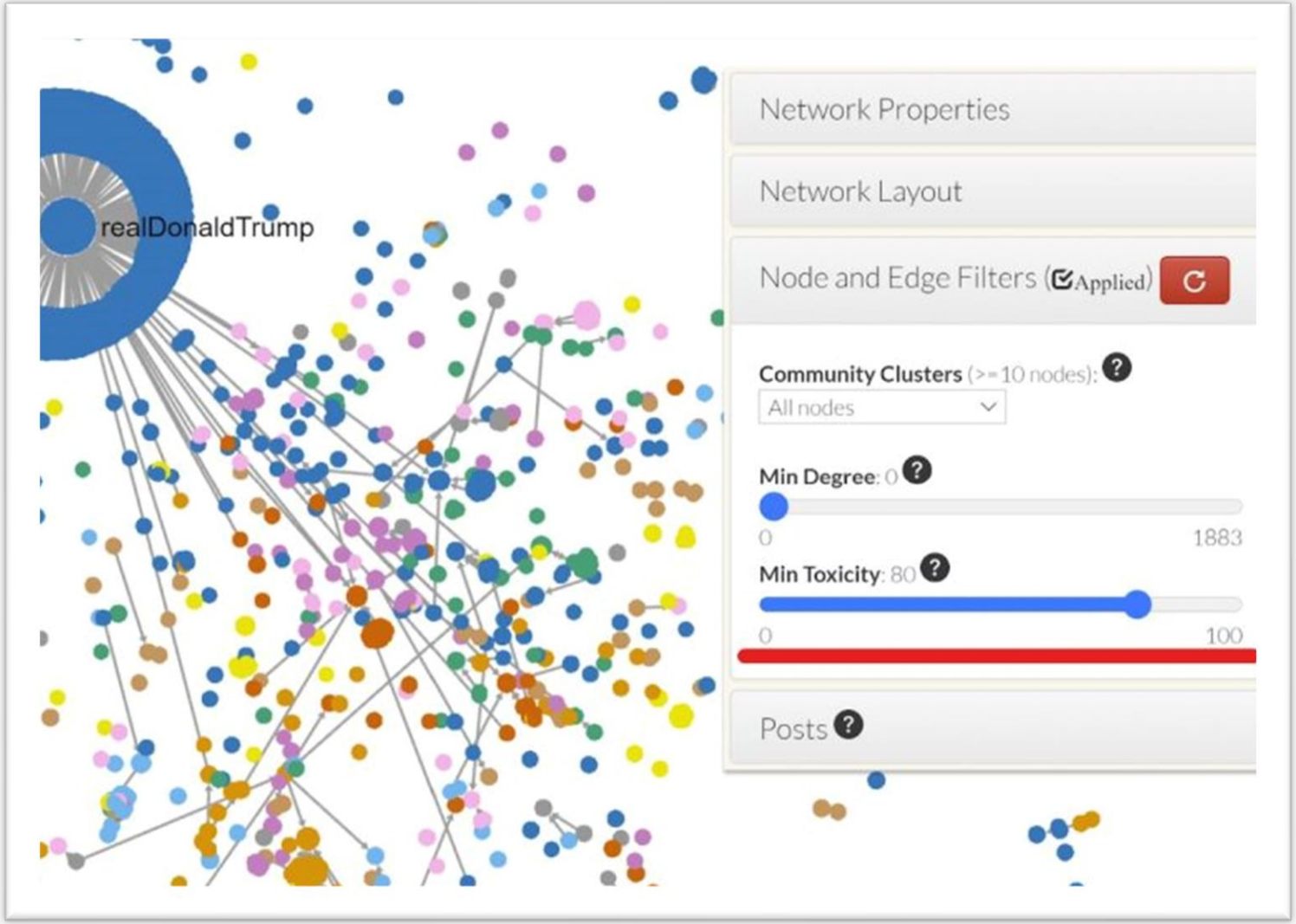
Download the CSV File for Toxicity analysis results of all the posts in the dataset

	Average for dataset	Highest value	Lowest value
Toxicity	0.18	0.53	0.06
Severe toxicity	0.09	0.28	0.01
Identity attack	0.11	0.18	0.03
Insult	0.16	0.76	0.03
Profanity	0.13	0.71	0.02
Threat	0.24	0.69	0.05
Sexually Explicit	0.13	0.23	0.02
Flirtation	0.30	0.44	0.08
Attack On Commenter	0.52	0.87	0.05

Communalytic Features:
Text Analysis Module

Communalytic can automatically pinpoint toxic and anti-social interactions using a machine learning API called ‘Perspective ‘ developed by Google.

communalytic EDU | communalytic Pro



* - a graph that describes connections between two types of nodes, where one of the node types represents **Social Actors** and the other node types represents **Semantic Concepts**.

Communalytic Features: Network Analysis & Visualization Module

Communalytic built-in social network analysis (SNA) module can render network w/up to 100K edges.

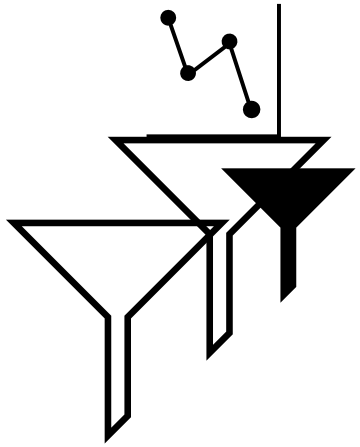
- Reply Network
- Reply Network w/Toxicity Scores
- 2-Mode Semantic Network (for CrowdTangle FB / IG data) *

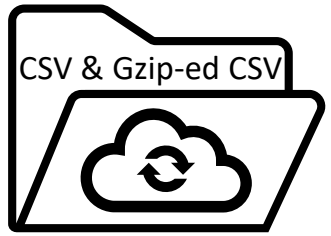
It can also

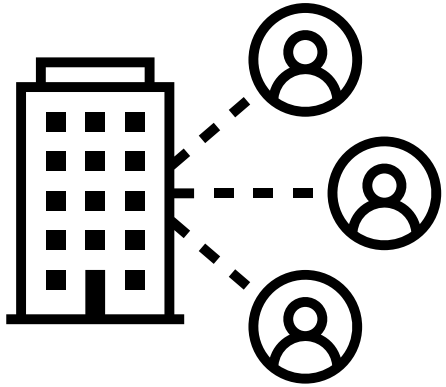
- Identify influencers
- Map shared interests and the spread of misinformation, and
- Detect signs of possible coordination among seemingly disparate actors.

Communalytic Features:
Additional Features

communalytic EDU	communalytic PRO
• **Simultaneously Run Multiple Data Collectors:** 1 Reddit, 1 Twitter & 1 CrowdTangle • **Import existing datasets** (CSV and Gzip-ed CSV) • **Collaboration Friendly:** Access to various team collaboration features	• **Simultaneously Run Multiple Data Collectors:** 2 Reddit, 1 Twitter & 1 CrowdTangle • **Import existing datasets** (CSV and Gzip-ed CSV) • **Collaboration Friendly:** Access to various team collaboration features







Outline

About Communalytic

Twitter Thread Data Collection

(Edu & Pro)

How to request Twitter's Bearer authentication token

Twitter Academic Product Track

(Pro Only)

communalytic (EDU)

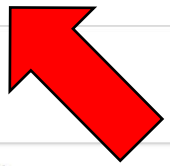
FAQ Tutorials Publications My Datasets

My Datasets

Server Time: June 10, 2021 12:19 UTC

Collect data from

Reddit (LIVE) Twitter Thread CrowdTangle CSV File



Search dataset list...

My Datasets Shared With Me

Donald J. Trump @realDonaldTrump · Sep 10

Sending out 80 MILLION BALLOTS to people who aren't even asking for a Ballot is unfair and a total fraud in the making. Look at what's going on right now!

20.4K 30.5K 98.6K

David Former Trump Supporter Weissman @davidm... · Sep 10

Replying to @realDonaldTrump

Tweet reported for misinformation on the elections.

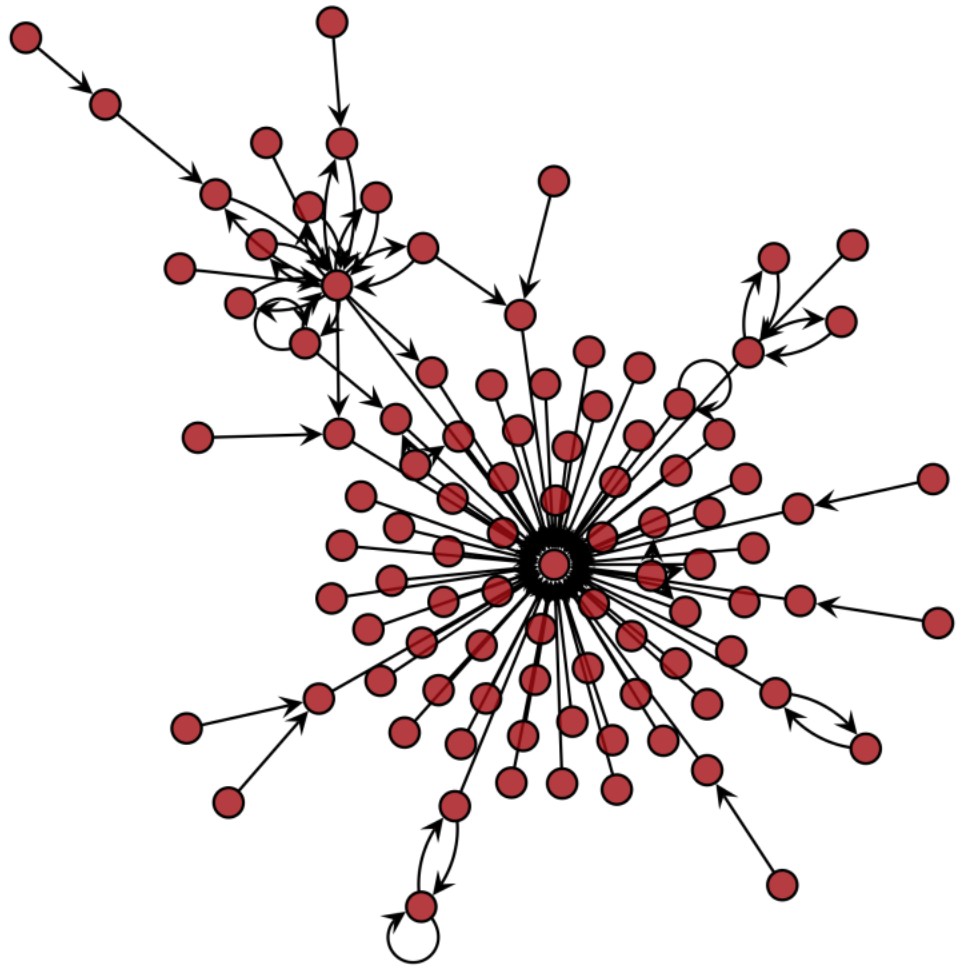
104 146 2.3K

Osha Davidson @OshaDavidson · Sep 10

Me too.

5 2 181

Show replies



A network representation of replies to a given tweet

The Associated Press on Twitter:

twitter.com/AP/status/1312063294665117696

Home

Explore

Notifications

Messages

Bookmarks

Lists

Profile

More

Tweet

Tweet

The Associated Press
@AP

First lady Melania Trump says she is experiencing “mild symptoms” after she and President Trump tested positive for COVID-19. “Thank you for the love you are sending our way. I have mild symptoms but overall feeling good,” she tweeted.

The Latest: Putin offers 'sincere support' to Trump
WASHINGTON (AP) — The Latest on President Donald Trump and first lady Melania Trump testing positive for the coronavirus (all times EDT): 5:15 a.m. ...
apnews.com

12:14 PM · Oct 2, 2020 · SocialFlow

18 Retweets **3** Quote Tweets **83** Likes

communalytic (EDU)

FAQ Tutorials Publications My Datasets

Server Time: May 11, 2021 14:04 UTC

Collect data from

Reddit (LIVE) Twitter Thread CrowdTangle CSV File

Search dataset list...

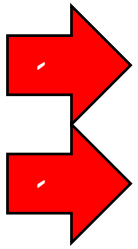
communalytic (EDU)

FAQ Tutorials Publications My Datasets

Back to My Datasets

Twitter Thread Data Collection

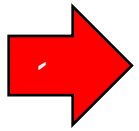
Please enter the following details to begin collection



Dataset name

Enter the URL of a popular tweet (with many replies) ?

Email me once job completes



Start Data Collection

Reddit (LIVE) Twitter Thread CrowdTangle CSV File

Search dataset list...

My Datasets | Shared With Me

Total records in the account: 186 / 30000

1%

Dataset Name	Number of Records	Toxicity Analysis	Network Analysis	Download Dataset	Collaborators	Delete
Test Reddit reddit : CanadaPolitics Click to check progress	Extracting until 2021-05-12 100					
Test Twitter twitter : https://twitter.com/NBA/status/1392123045968498692	86					

Back to My Datasets

Overview

Dataset Name: **Test Twitter**

Platform: **twitter**

Search Criteria: **https://twitter.com/NBA/status/1392123045968496892**

Collection started: **05/11/2021**

Collection ended: **05/11/2021**

Records: **86**

Posts Per Day

Word Cloud

Emoji Cloud

Top Ten Posters

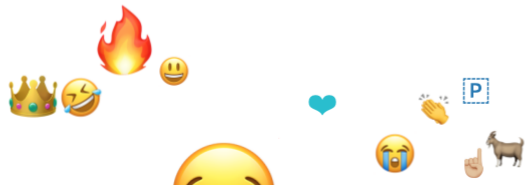
Word Cloud

CSV EXCEL PNG

This chart shows up to 100 most frequently used words in your dataset. (Excluding numbers, URLs, and stop words in 15 different languages.)

Emoji Cloud

CSV EXCEL PNG



Outline

About Communalytic

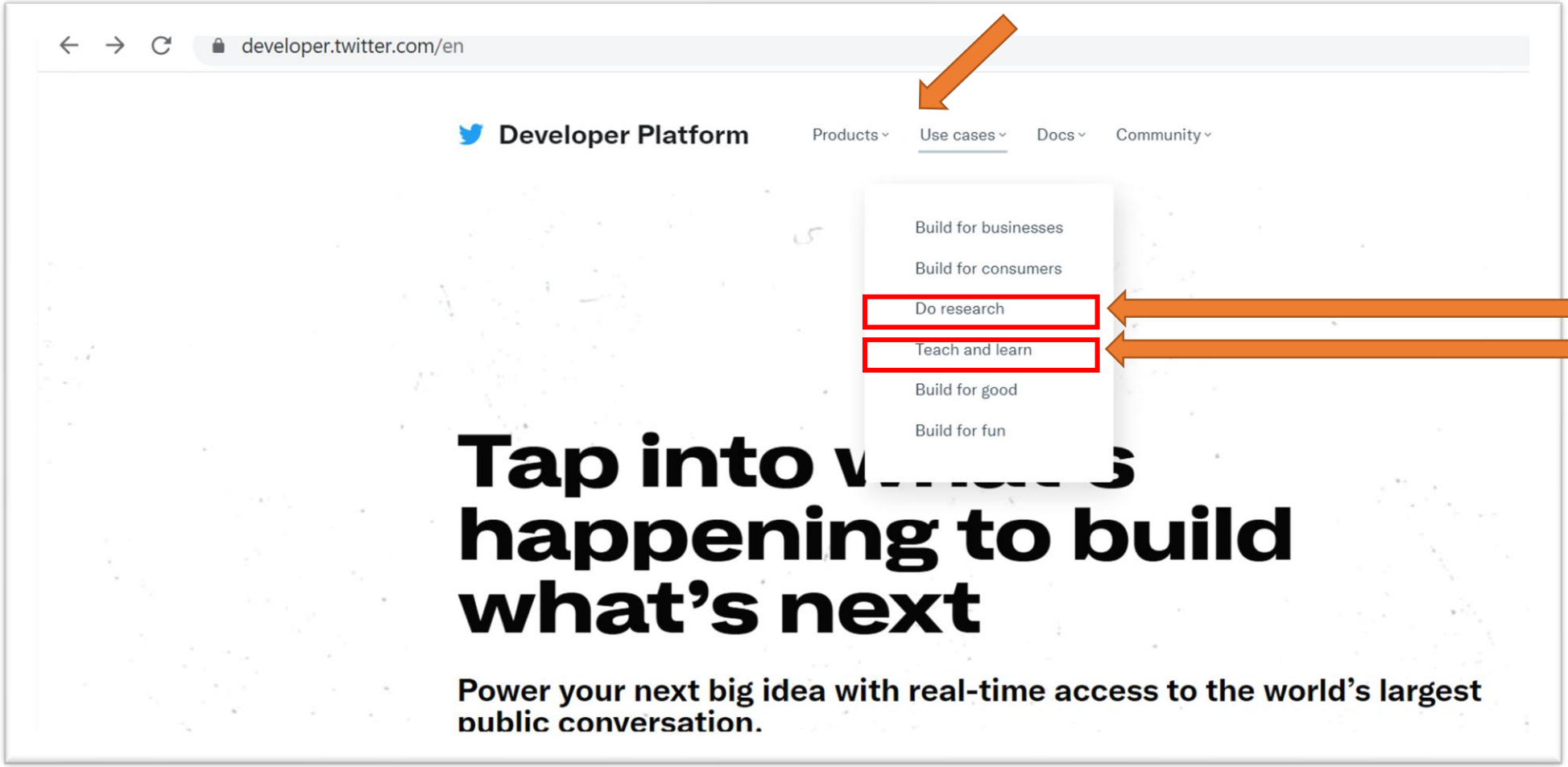
Twitter Thread Data Collection

(Edu & Pro)

How to request Twitter's Bearer authentication token

Twitter Academic Product Track

(Pro Only)



https://developer.twitter.com

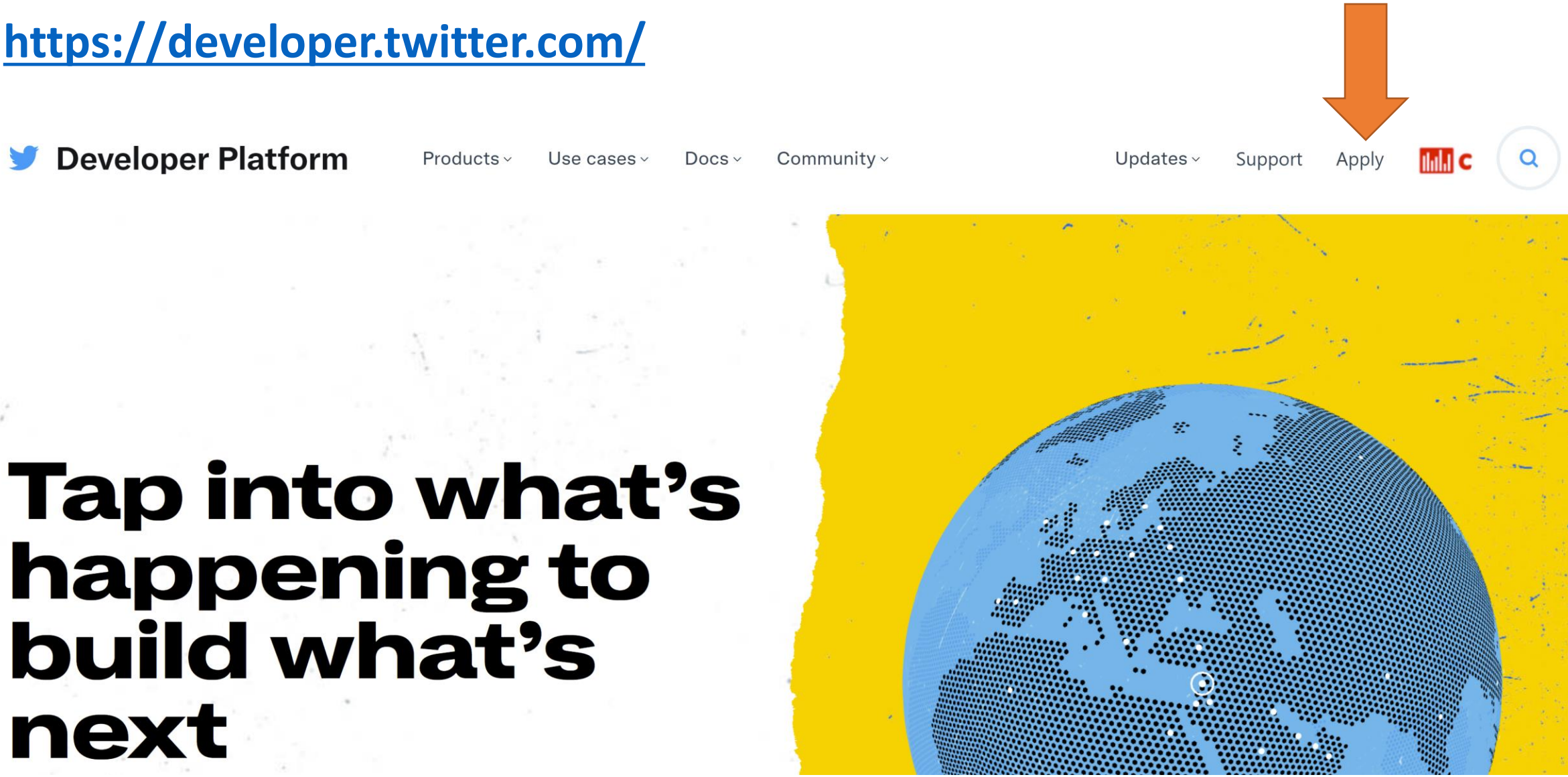
Academic product track

Standard product track

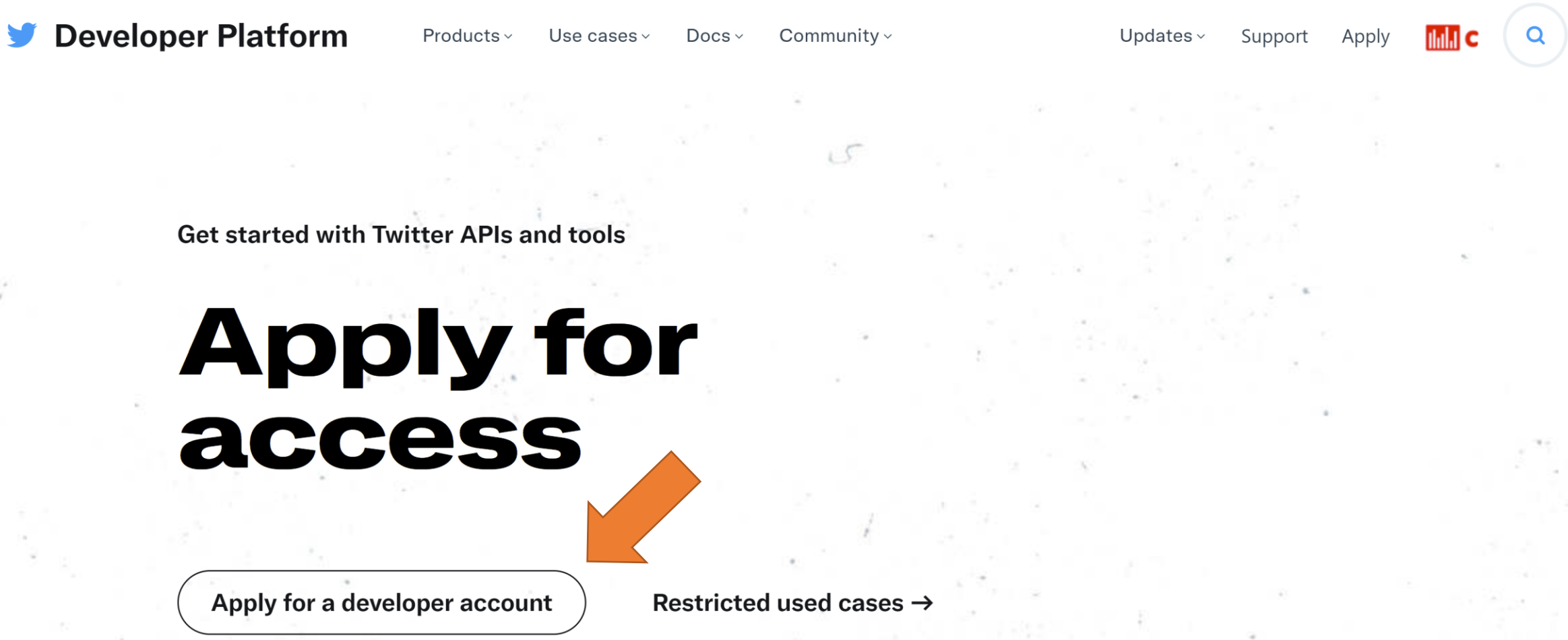
Academic Product Track is new. It provides access to both historical and real-time public tweets for research projects. 10M/Monthly Cap – Requires a more detailed application, harder to qualify for.

Standard Product Track is the "default product track for most developers, including those getting started, building for fun, for a good cause, for consumers, or to teach and learn." 500k/Monthly Cap - Easier to get approved.

https://developer.twitter.com/



https://developer.twitter.com/



https://developer.twitter.com/

Developer Portal

#UseCases

First things first, let's get you the right application.

Pick the use case that most closely relates to the type of work you intend to do while using the Twitter developer platform.

There's some overlap between these so focus on the specifics of your main objective.

Docs Community Updates Support

Which best describes you?

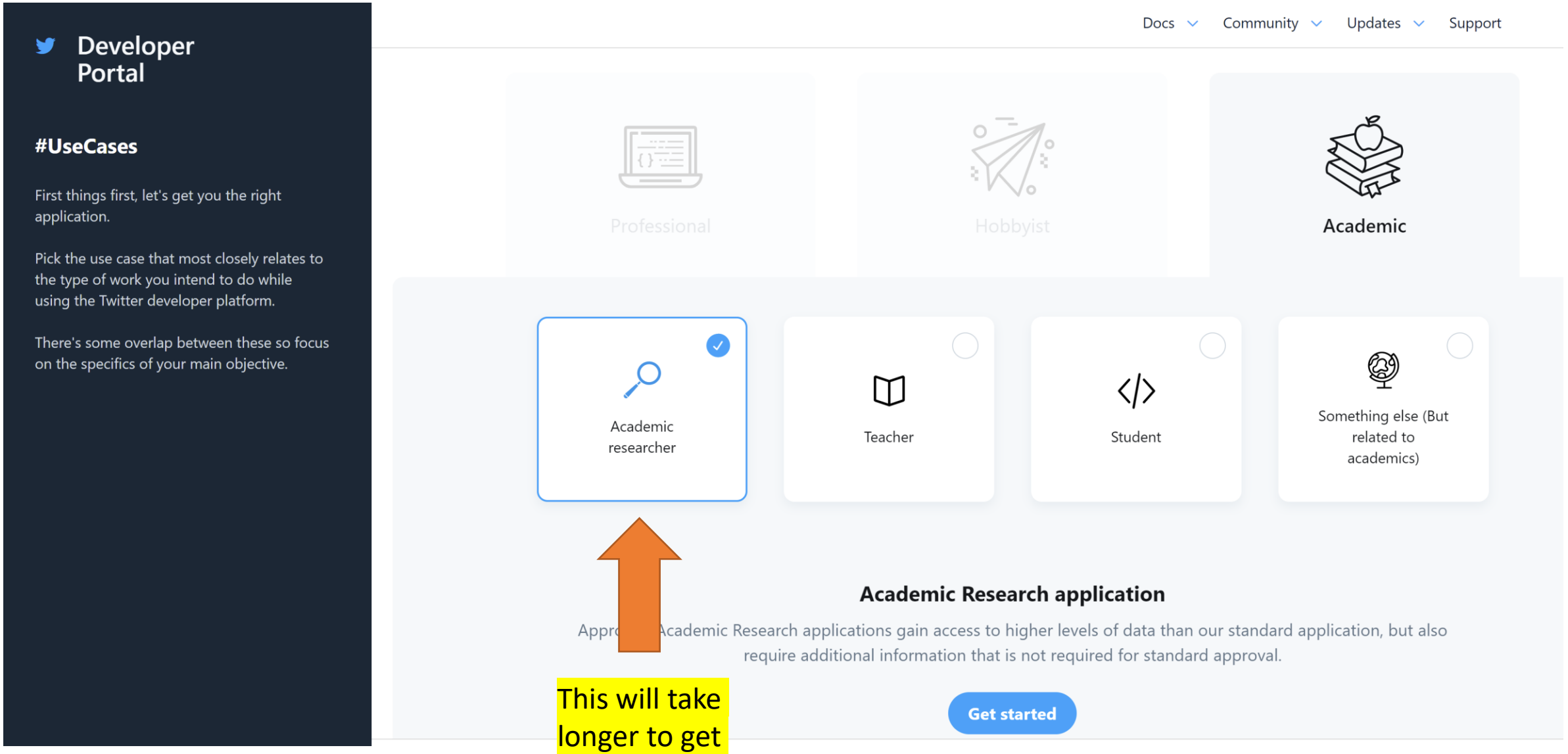
This is how you intend to use the Twitter developer platform

Professional

Hobbyist

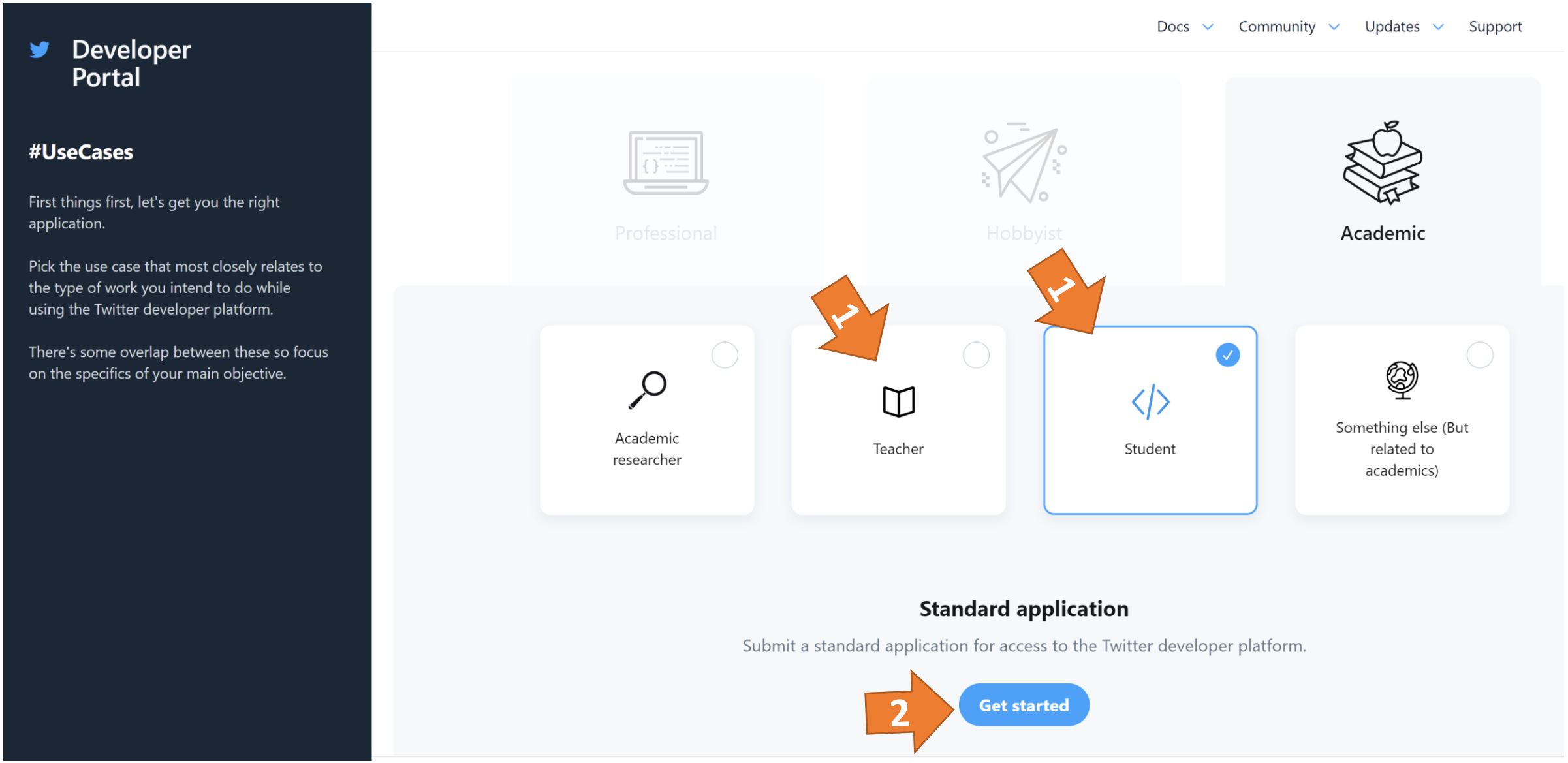
Academic

https://developer.twitter.com/

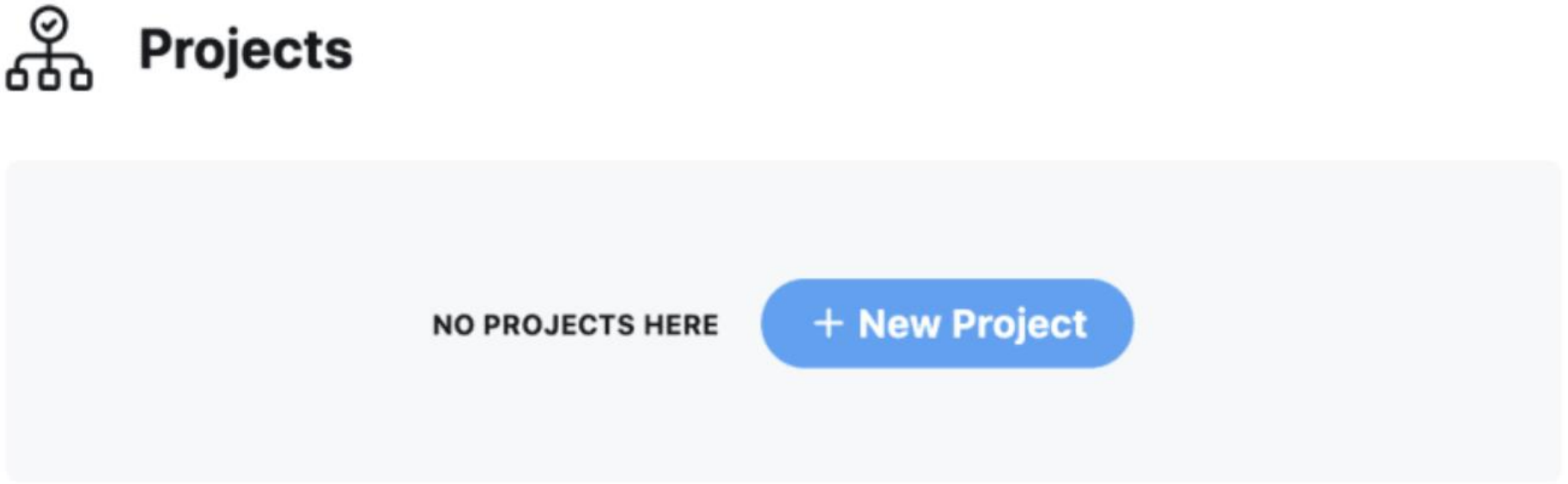


This will take longer to get approved

https://developer.twitter.com/

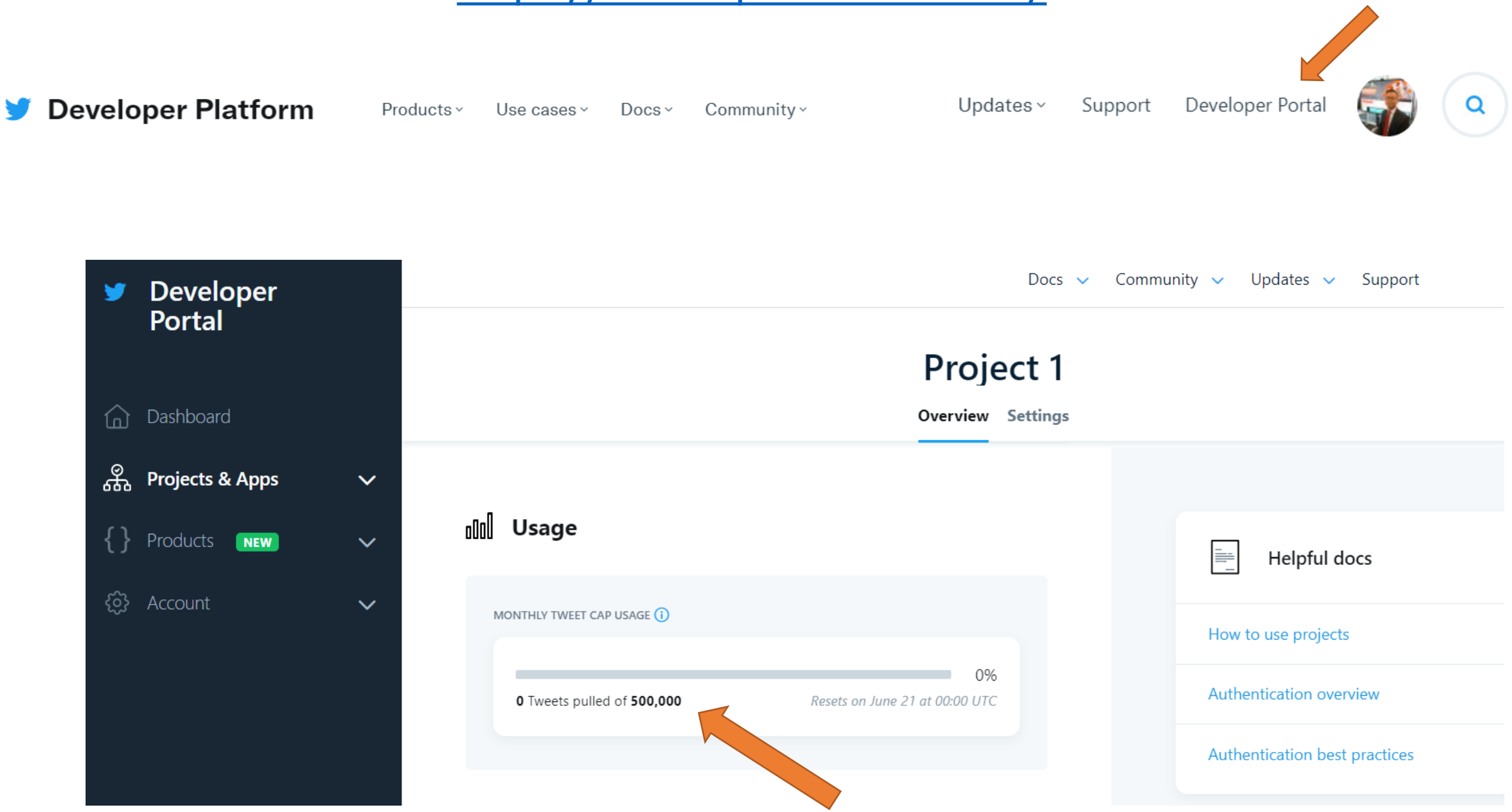


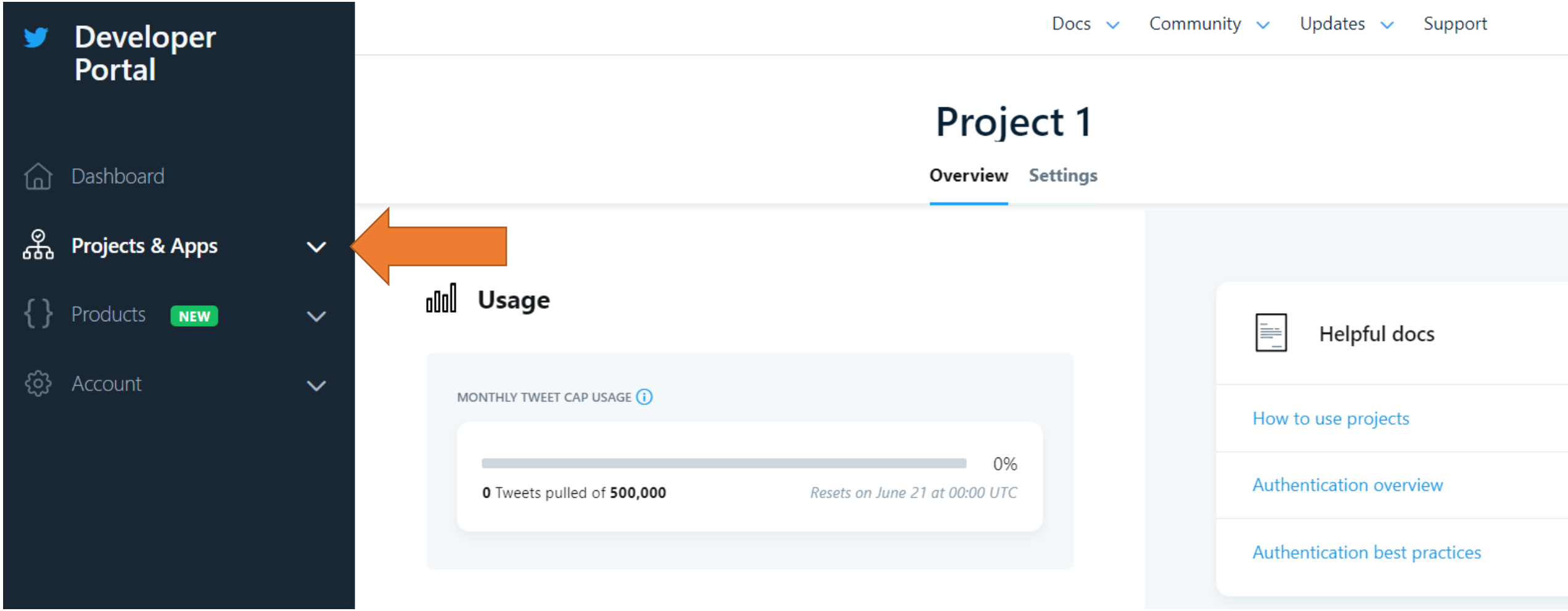
https://developer.twitter.com/

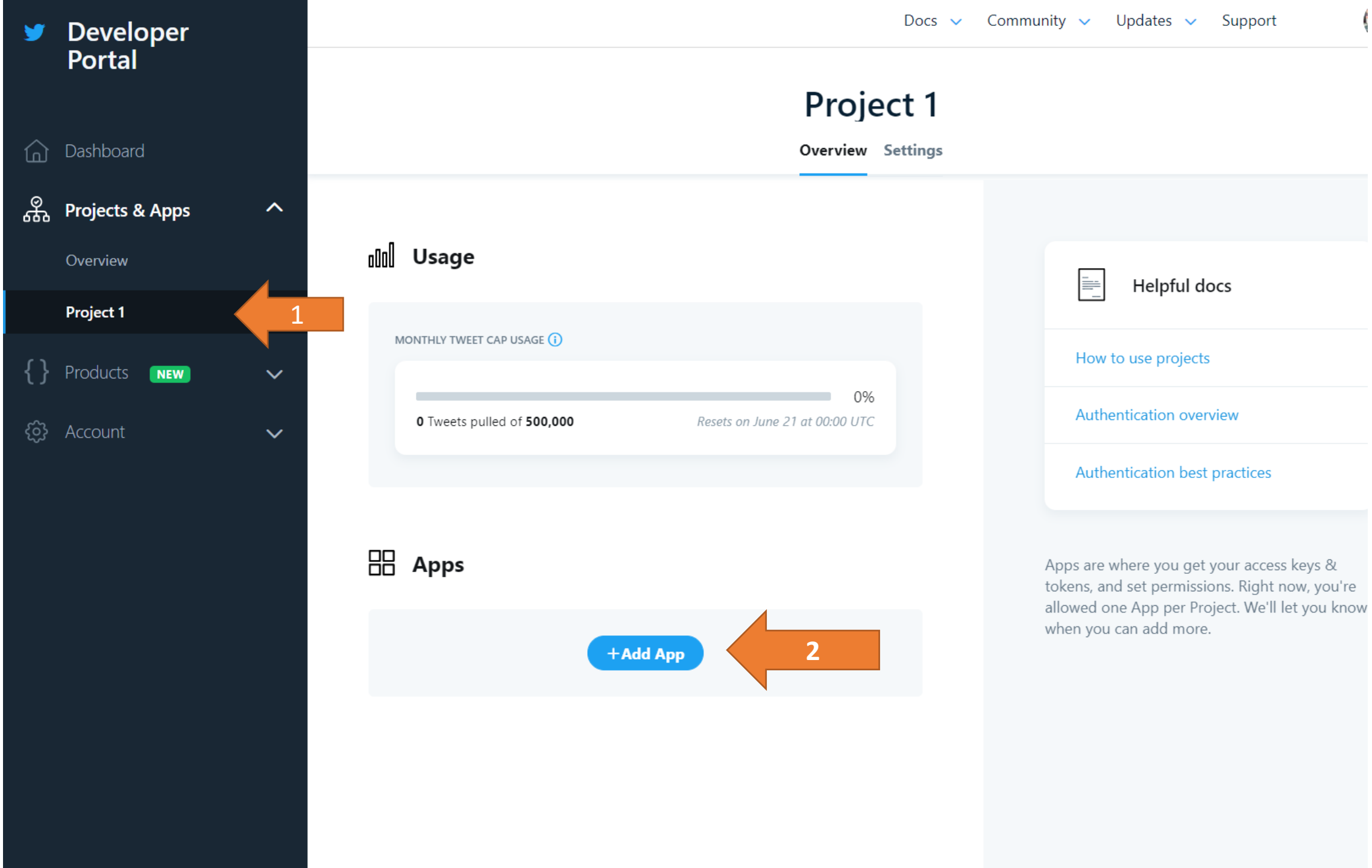


***See more info here:
https://developer.twitter.com/en/docs/tutorials/step-by-step-guide-to-making-your-first-request-to-the-twitter-api-v2

https://developer.twitter.com/




Developer Portal
Dashboard
Projects & Apps
Products
NEW
Account
Docs
Community
Updates
Support
Project 1
Overview
Settings
Usage
MONTHLY TWEET CAP USAGE
0%
0 Tweets pulled of 500,000
Resets on June 21 at 00:00 UTC
Helpful docs
How to use projects
Authentication overview
Authentication best practices


Developer Portal
Dashboard
Projects & Apps
Overview
Project 1
1
Products
NEW
Account
Docs
Community
Updates
Support
Project 1
Overview
Settings
Usage
MONTHLY TWEET CAP USAGE
0%
0 Tweets pulled of 500,000
Resets on June 21 at 00:00 UTC
Apps
+Add App
2
Helpful docs
How to use projects
Authentication overview
Authentication best practices
Apps are where you get your access keys & tokens, and set permissions. Right now, you're allowed one App per Project. We'll let you know when you can add more.

Developer Portal

Dashboard

Projects & Apps

Overview

Project 1

Products NEW

Account

Docs Community Updates Support

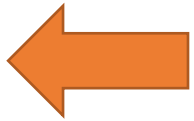
Add an existing App or create a new App

Add an existing App

Add an existing App

Create a new App

Create new



Developer Portal

Dashboard

Projects & Apps

Overview

Project 1

Products NEW

Account

Docs Community Updates Support

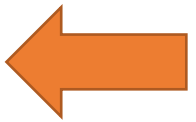
Name your App

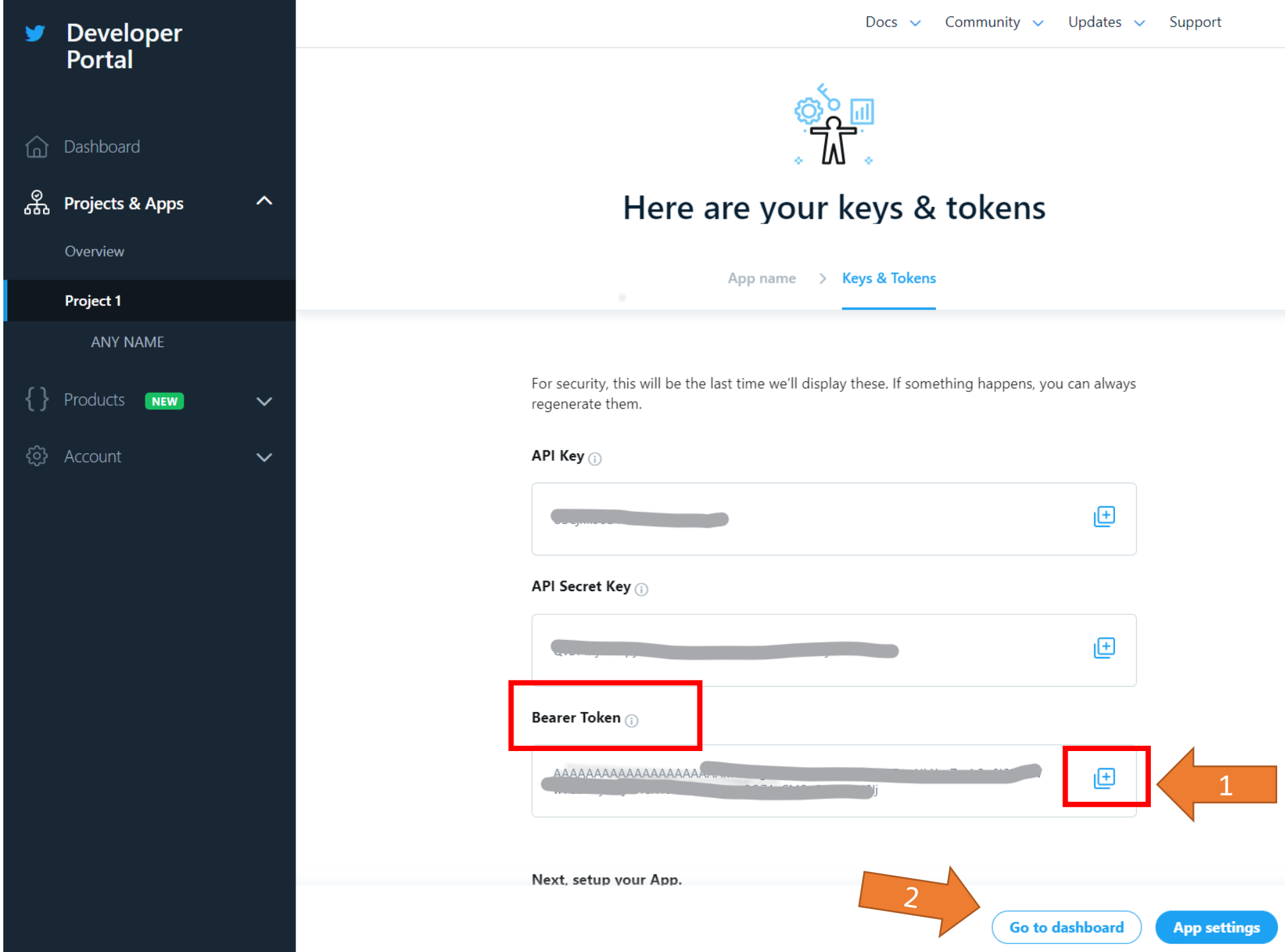
App name > Keys & Tokens

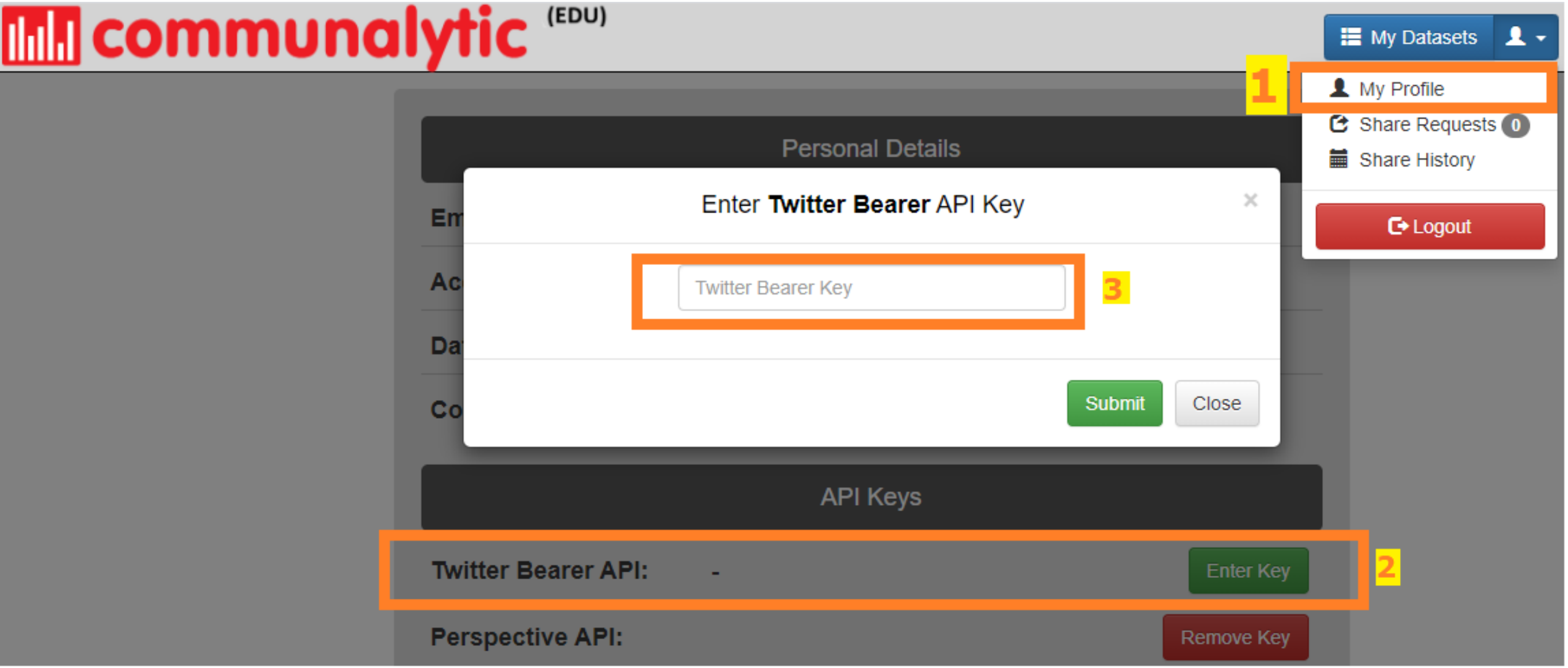
Apps are where you get your access **keys & tokens**, plus set permissions. You can find them within your Projects.

ANY NAME

23




Developer Portal
Dashboard
Projects & Apps
Overview
Project 1
ANY NAME
Products
NEW
Account
Docs
Community
Updates
Support
Here are your keys & tokens
App name
Keys & Tokens
For security, this will be the last time we'll display these. If something happens, you can always regenerate them.
API Key
API Secret Key
Bearer Token
1
Next, setup your App.
2
Go to dashboard
App settings


communalytic
(EDU)
My Datasets
1
My Profile
Share Requests 0
Share History
Logout
Personal Details
Enter Twitter Bearer API Key
Twitter Bearer Key
3
Submit
Close
API Keys
Twitter Bearer API: -
Enter Key
2
Perspective API:
Remove Key

Outline

About Communalytic

Twitter Thread Data Collection

(Edu & Pro)

How to request Twitter's Bearer authentication token

Twitter Academic Track

(Pro only)

Working with Twitter's Academic Product Track in Communalytic Pro

https://pro.communalytic.com

Twitter API: Academic Research Product Track

Eligibility requirements:

- Master's student, doctoral candidate, postdoc, faculty, or research-focused employee at an academic institution
- A clearly defined research project (i.e., you have a specific plan for Twitter data use, analysis, and sharing)
- Non-commercial use

Source: https://developer.twitter.com/en/products/twitter-api/academic-research

Twitter API: Academic Research Product Track

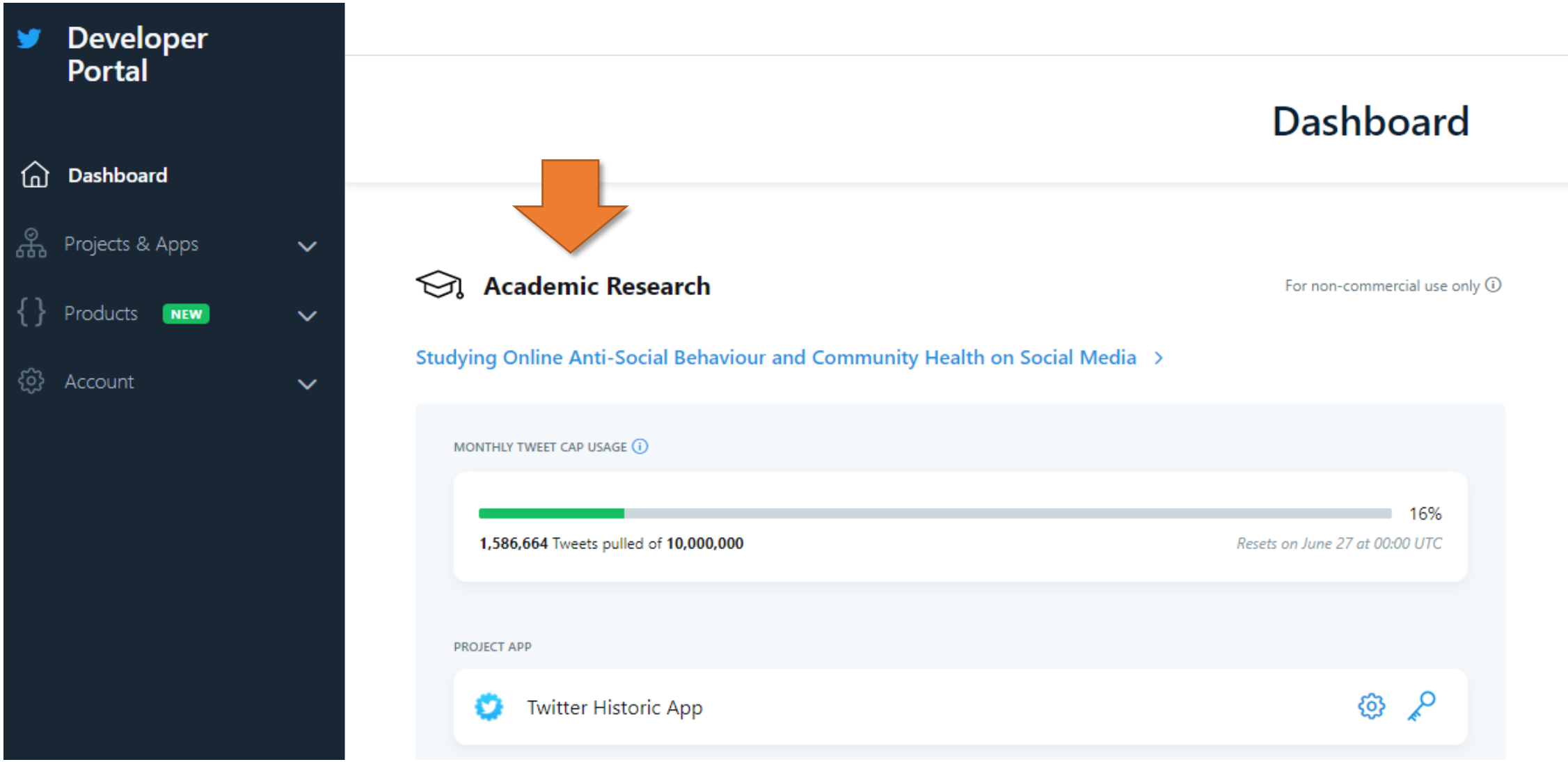
Key parameters:

Key benefits	Access Twitter's real-time and historical public data with additional features and functionality that support collecting more precise, complete, and unbiased datasets. More details on included endpoints
Data volume	10 million Tweets / month
Query rules	1024 characters, 1000 streaming rules

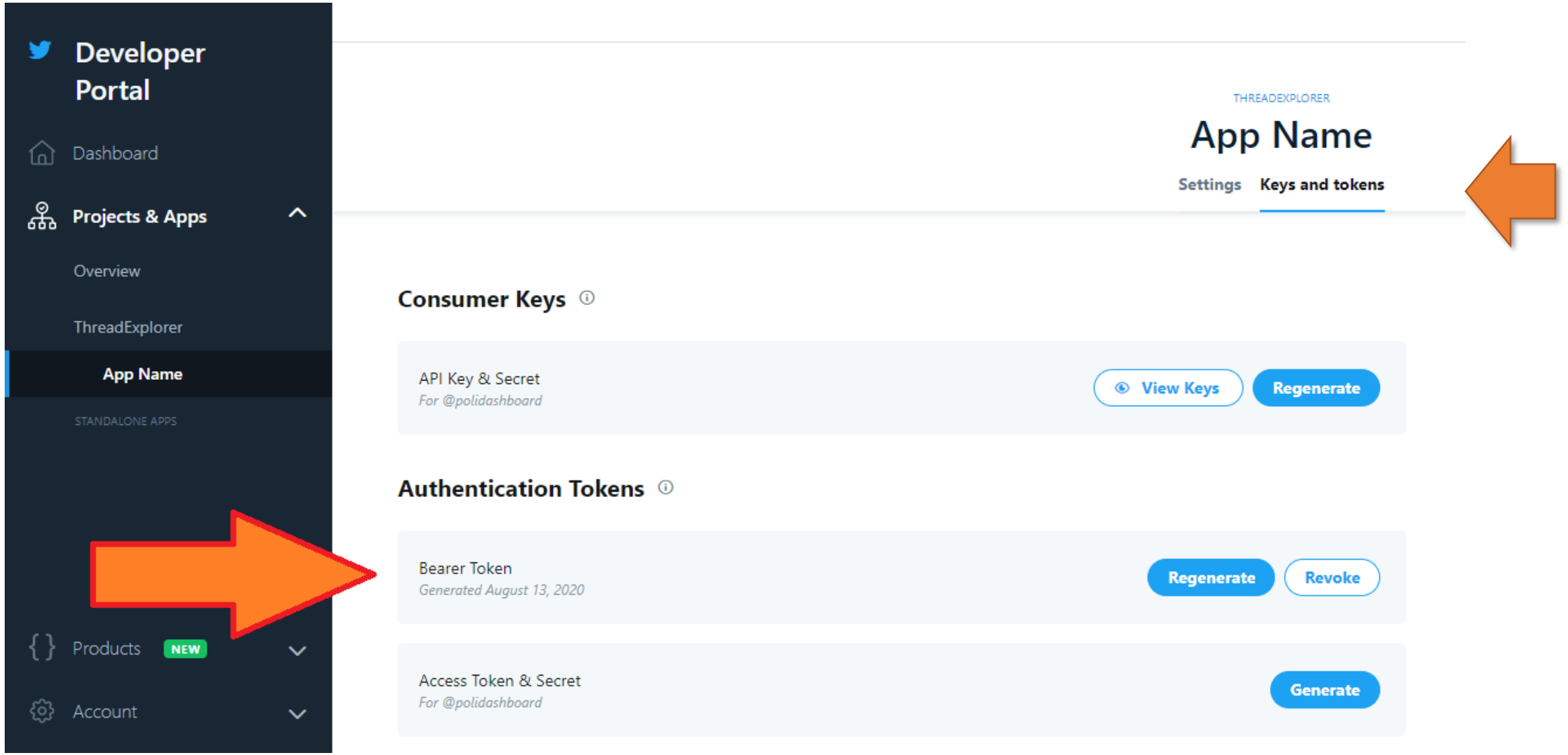
Note: Communalytic Pro only supports the historical data collection with this API

Apply at https://developer.twitter.com/en/portal/petition/academic/is-it-right-for-you

Twitter's Developer Dashboard: Confirm Access to the Academic API

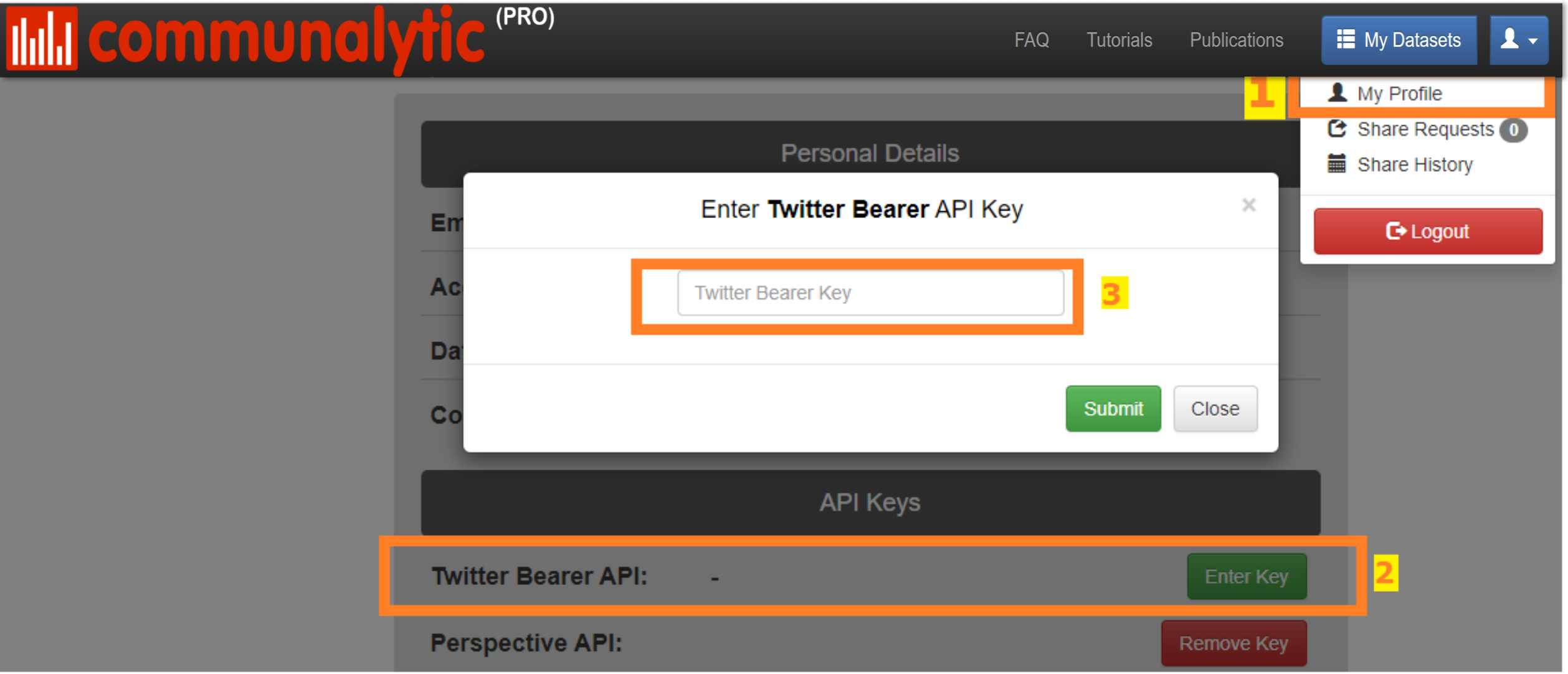


Twitter's Developer Dashboard: Retrieve your "Bearer" access token



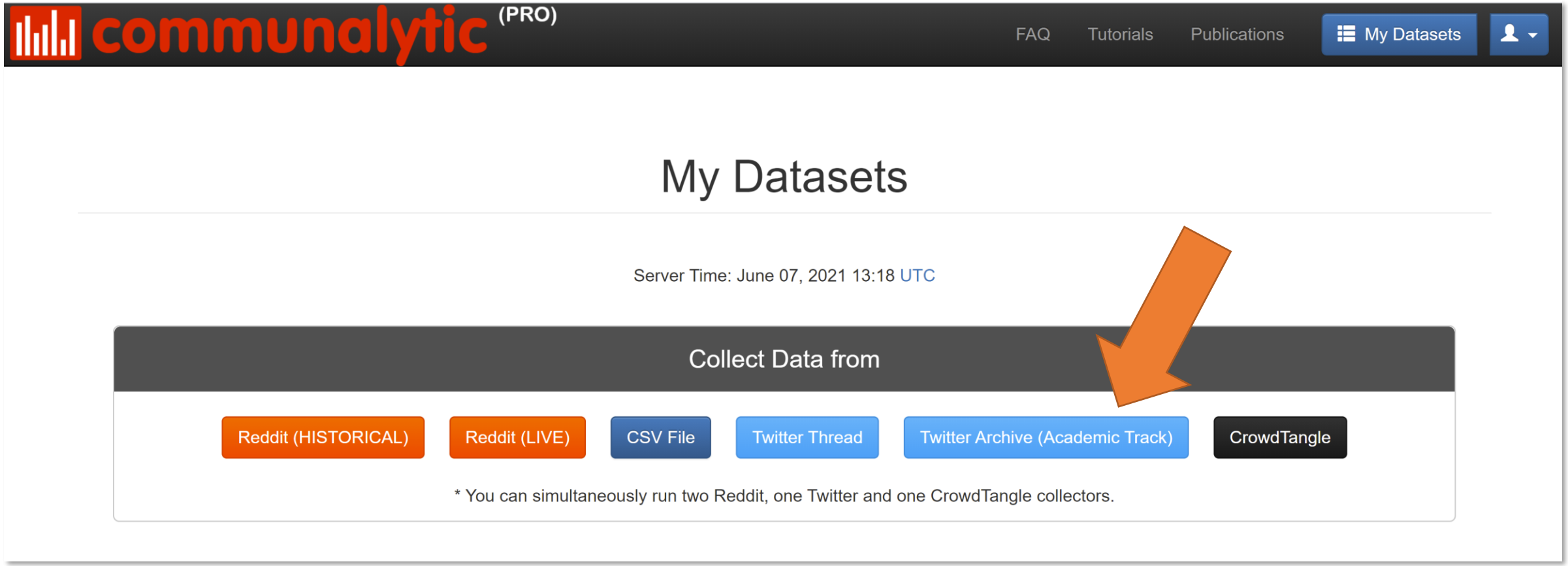
1. Twitter API → 2. Collection → 3. Metadata → 4. Export → 5. Sharing

Adding your “Bearer” access token to your Communalytic **PRO** account



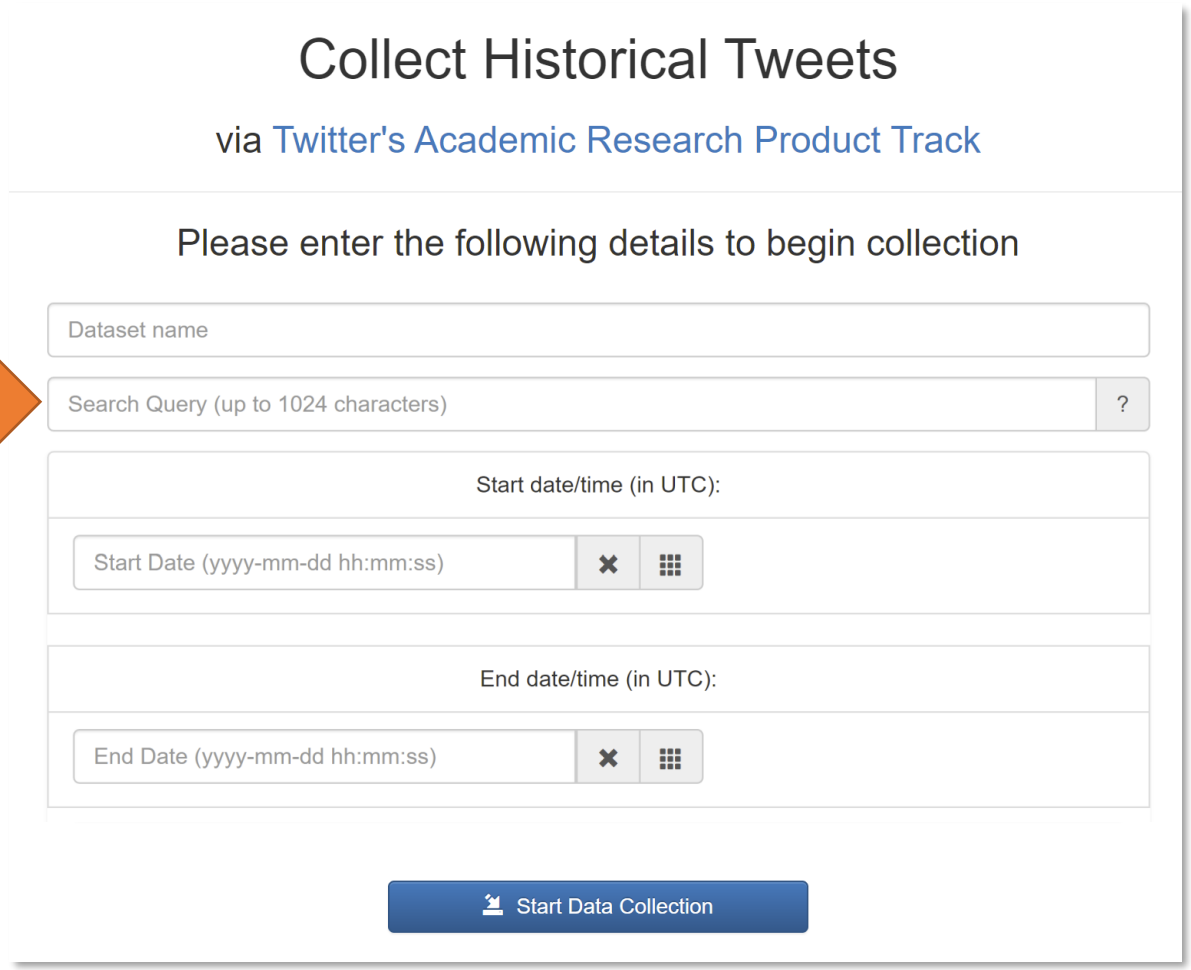
1. Twitter API → 2. Collection → 3. Metadata → 4. Export → 5. Sharing

Full-archive Search API: Create a new search (1)

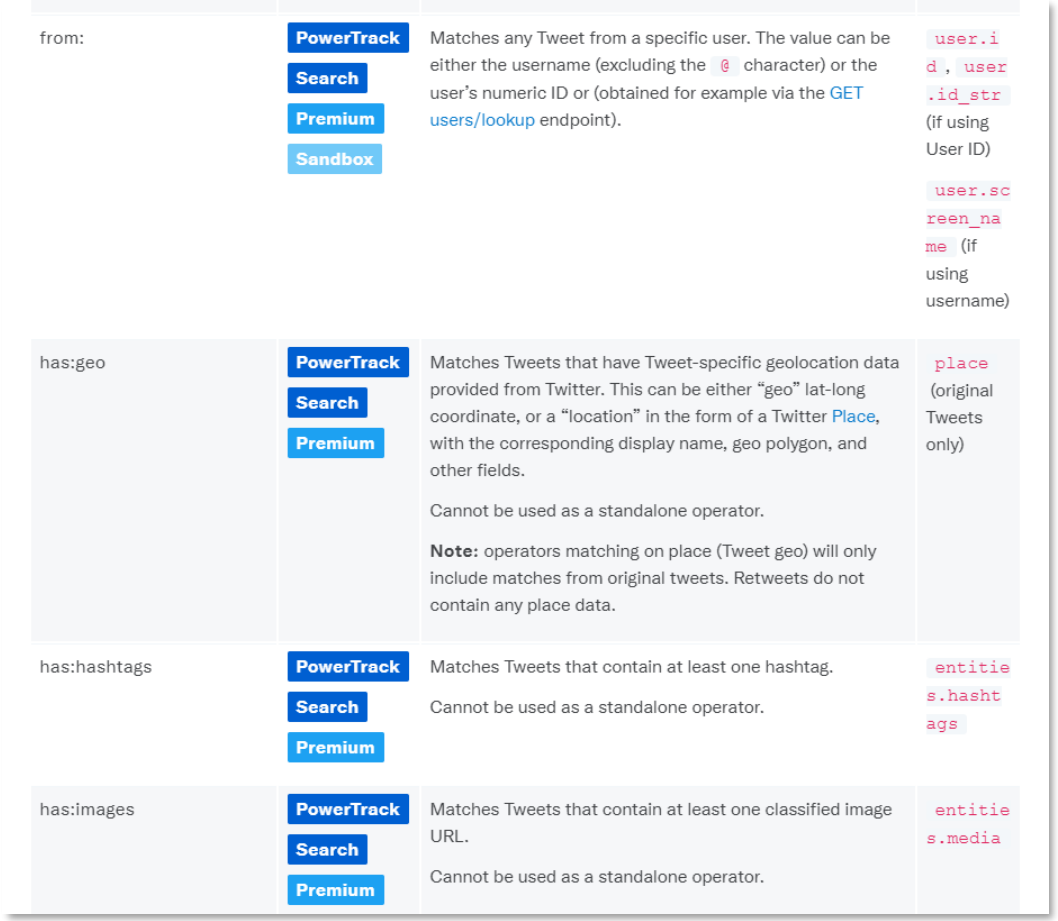


Full-archive Search API: Create a new search (2)

Search Query
(up to 1024 chars)



Advanced Search Rules and Operators



from:	PowerTrack Search Premium Sandbox	Matches any Tweet from a specific user. The value can be either the username (excluding the `@` character) or the user's numeric ID or (obtained for example via the GET users/lookup endpoint).	`user.id`, `user.id_str` (if using User ID) `user.screen_name` (if using username)
has:geo	PowerTrack Search Premium	Matches Tweets that have Tweet-specific geolocation data provided from Twitter. This can be either "geo" lat-long coordinate, or a "location" in the form of a Twitter Place, with the corresponding display name, geo polygon, and other fields. Cannot be used as a standalone operator. **Note:** operators matching on place (Tweet geo) will only include matches from original tweets. Retweets do not contain any place data.	`place` (original Tweets only)
has:hashtags	PowerTrack Search Premium	Matches Tweets that contain at least one hashtag. Cannot be used as a standalone operator.	`entities.hashtags`
has:images	PowerTrack Search Premium	Matches Tweets that contain at least one classified image URL. Cannot be used as a standalone operator.	`entities.media`

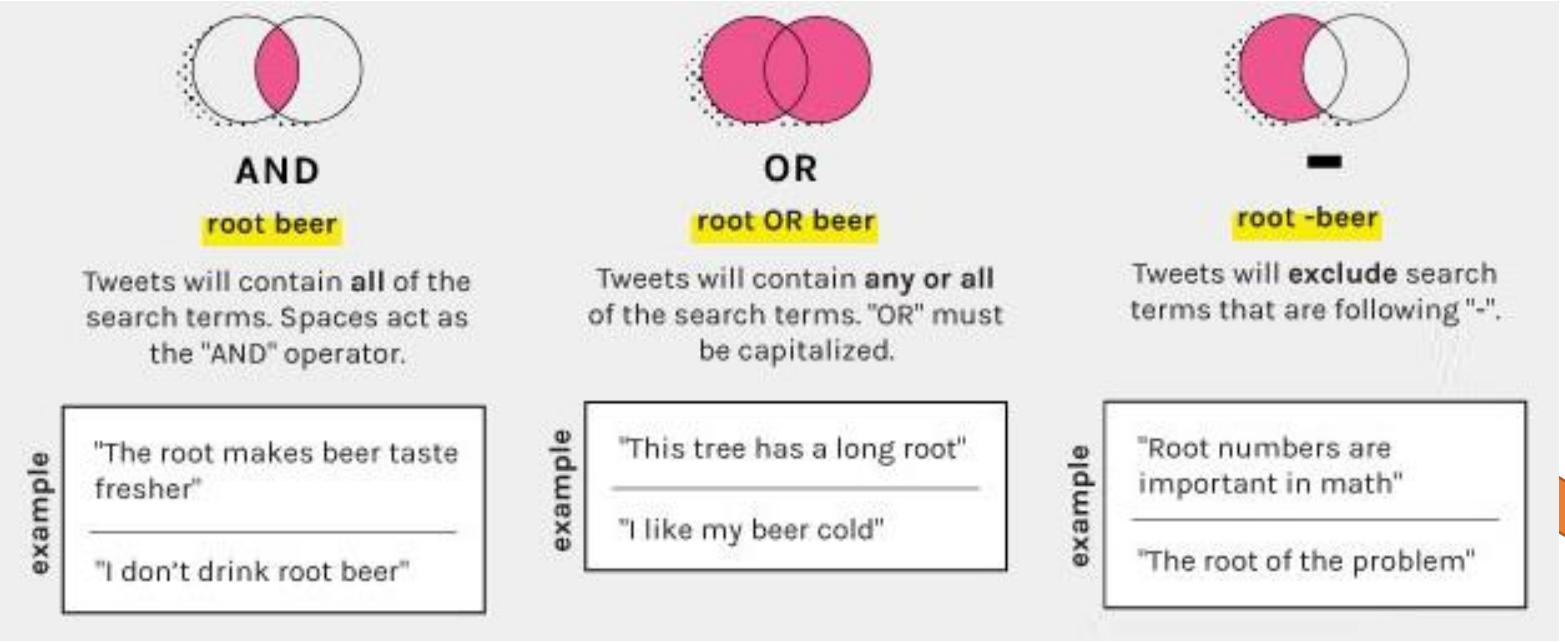
General data collection tips:

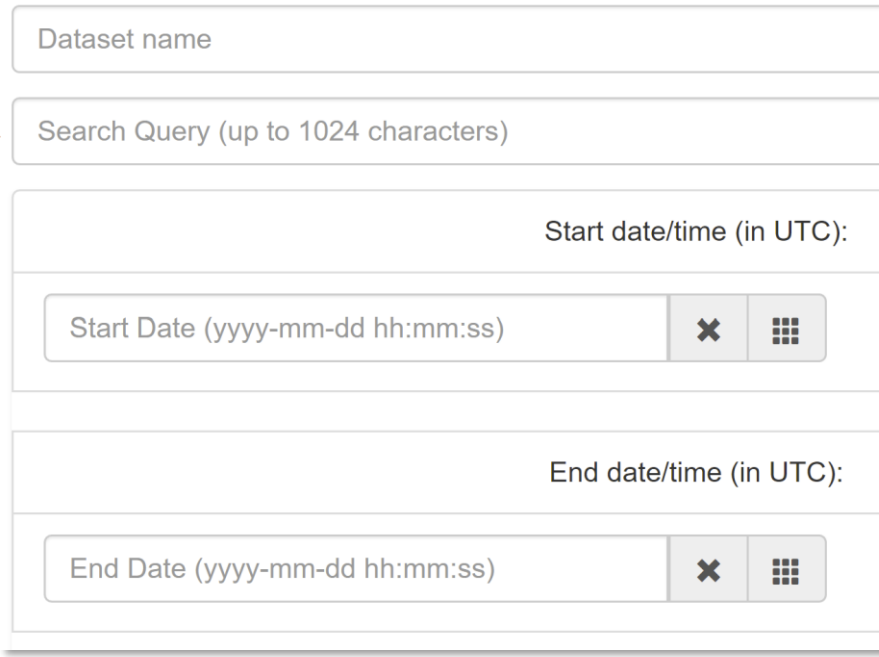
- https://developer.twitter.com/en/docs/tutorials/building-high-quality-filters

Search operators

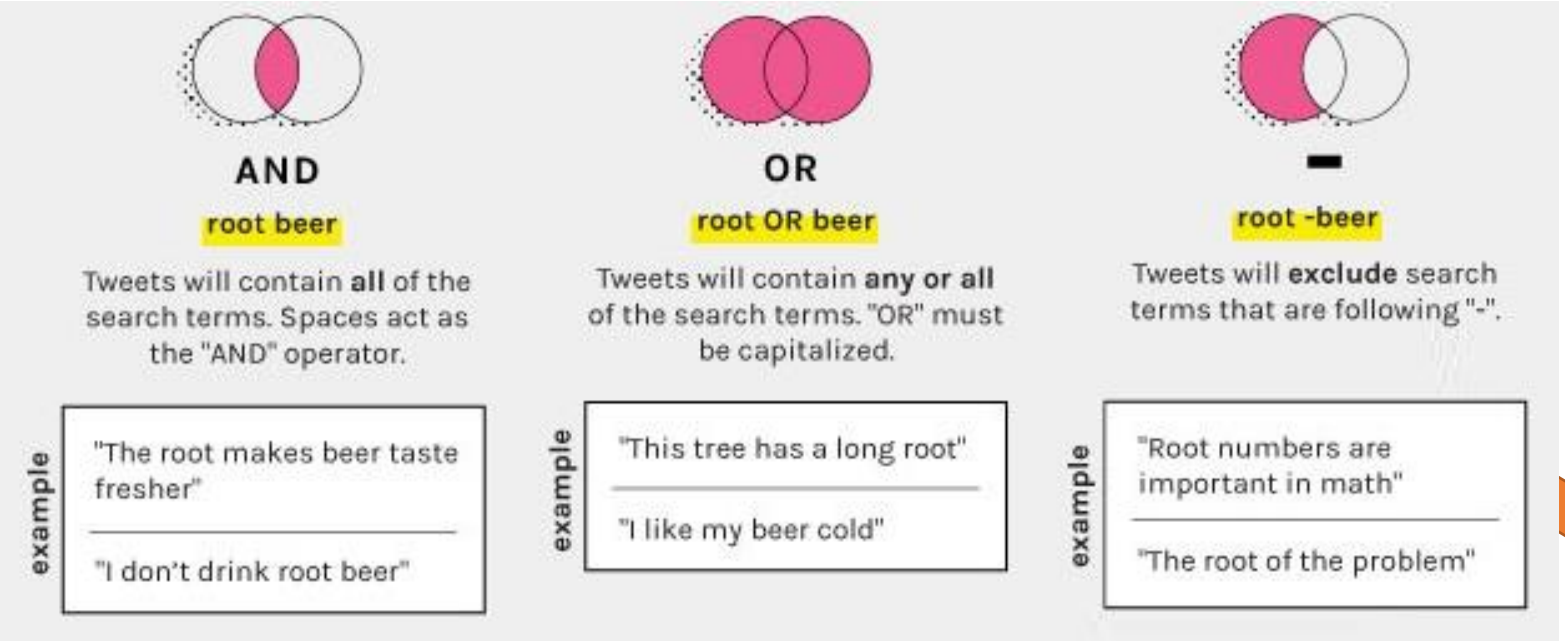
- https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query

Full-archive Search API: Common Search Operators (1)





Full-archive Search API: Common Search Operators (2)

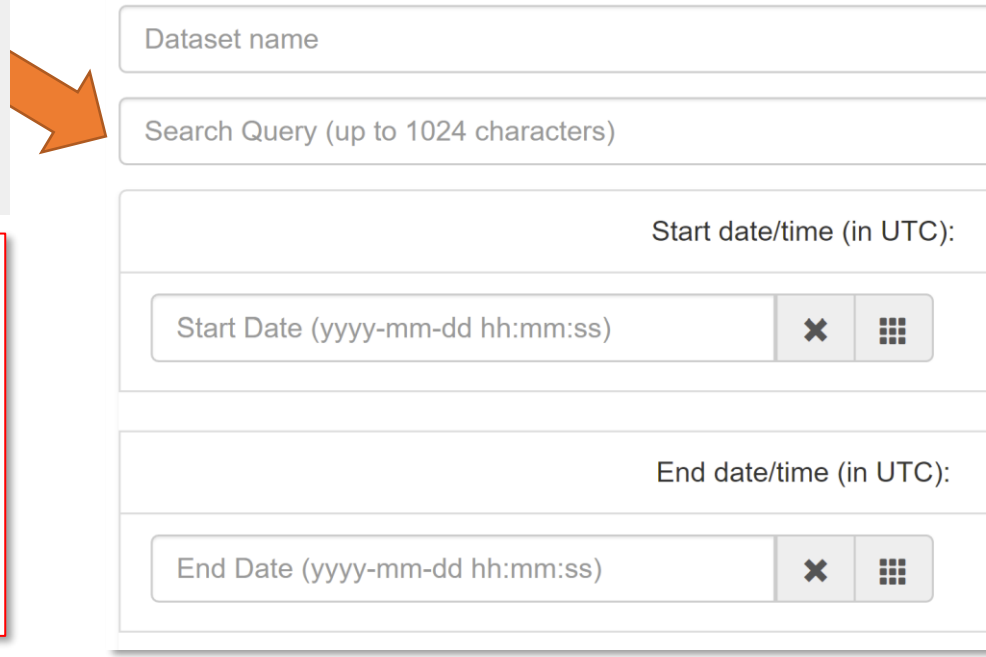


Precedence of AND before OR

One key difference to keep in mind when using operators with the new recent search and filtered stream endpoints is that

- In the old v1.1 API (search/tweets), OR is applied before logical AND(which is denoted by a space between terms or operators)
- In the new Twitter API (recent search and filtered stream), AND is applied before OR

Source: https://developer.twitter.com/en/docs/tutorials/building-high-quality-filters

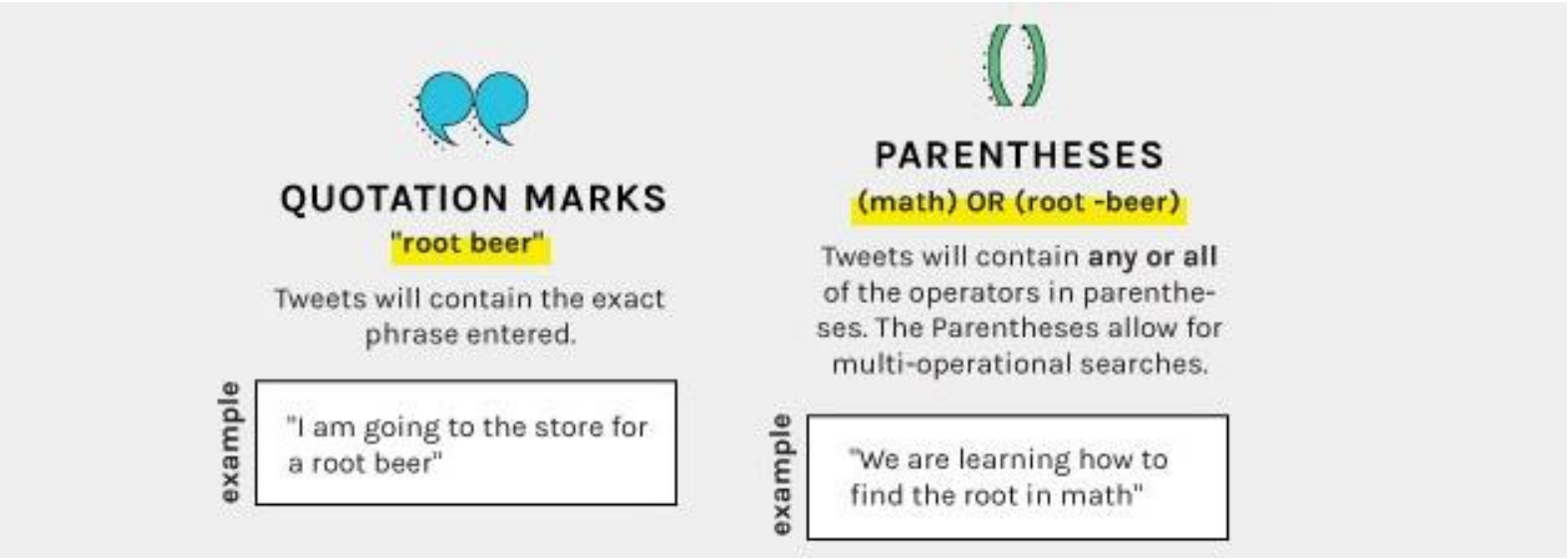


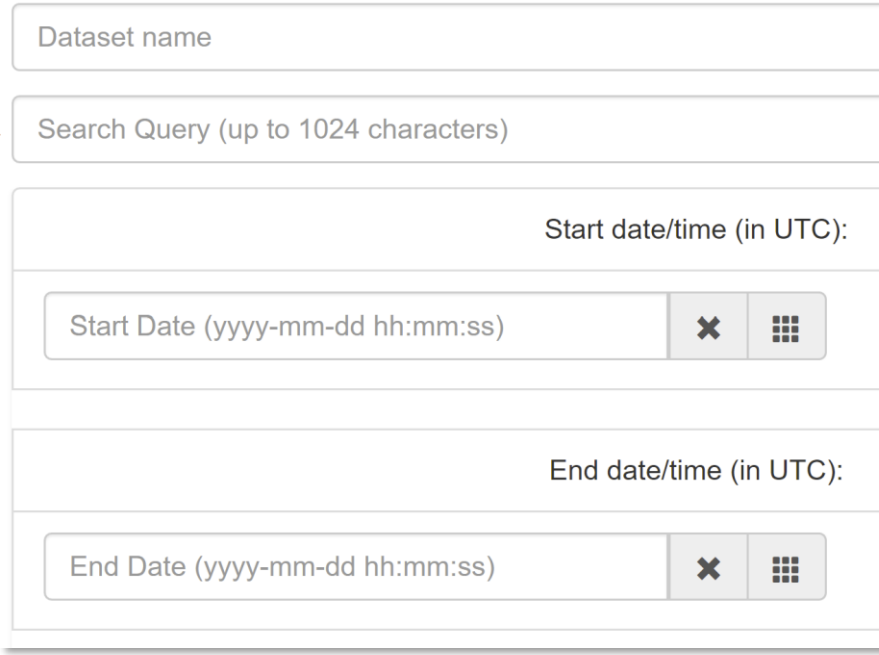
Full-archive Search API: Common Search Operators (3)

vaccine Covid OR Coronavirus

This query will be interpreted by Twitter as …

Twitter API V1.1: **vaccine AND (Covid OR Coronavirus)**
Twitter API V2.0: **(vaccine AND Covid) OR Coronavirus**





Full-archive Search API: Common Search Operators (3)

vaccine (Covid OR Coronavirus)

Examples of matching tweets:

Daily Express
@Daily_Express

#Spain backlash: Expat fury as foreigners NOT vaccinated despite it 'being their turn' #coronavirus #vaccine

Spain backlash: Expat fury as foreigners NOT vaccinated despite it 'being their t...
EXPATS living in Spain have claimed they are being denied their coronavirus vaccinations - even when it is their turn.
express.co.uk

5:12 PM · Jun 9, 2021 · TweetDeck

3 Retweets

Laurie Garrett
@Laurie_Garrett

Meet Dr. Sherri Tenpenny. In testimony to Ohio legislators she said #COVID19 #vaccines contain magnets that connect your body to 5G cell towers. Yeah, she's pals with the #MyPillow guy.

Anti-vax doctor, an adviser to Mike Lindell, claims COVID vaccine will "magnetiz...
Dr. Sheri Tenpenny, ally and adviser to MyPillow Guy, also says COVID vaccines will connect you to 5G towers
salon.com

7:11 PM · Jun 9, 2021 · Twitter Web App

123 Retweets **42** Quote Tweets **461** Likes

Hindustan Times
@htTweets

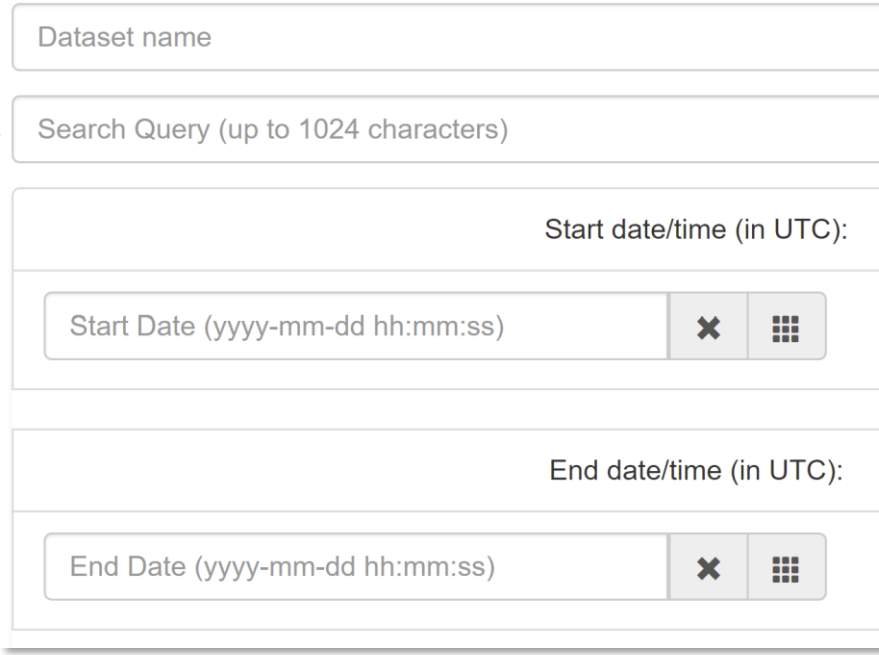
Congress asked Karnataka government to pay heed to its vaccine advisor Gagandeep Kang and alleged that the BJP government is only doing headline management.

Congress asks Karnataka's BJP govt to listen to its own Covid-19 vaccine advisor
"We thought the BJP government was finally taking scientific advice. However, it is clear it was only headline management because the government's actions do ...
hindustantimes.com

10:39 AM · Jun 2, 2021 · Twitter Web App

5 Retweets **1** Quote Tweet **39** Likes

Full-archive Search API: Common Search Operators (4)



Twitter's Advanced Search Operators: Filter by **Language**

Twitter currently supports 70 language and dialect identifiers:

lang:language_code

Examples:

#Coronavirus lang:**en**

#Coronavirus lang:**fr**

#Coronavirus lang:**uk**

Amharic: am	German: de	Malayalam: ml	Slovak: sk
Arabic: ar	Greek: el	Maldivian: dv	Slovenian: sl
Armenian: hy	Gujarati: gu	Marathi: mr	Sorani Kurdish: ckb
Basque: eu	Haitian Creole: ht	Nepali: ne	Spanish: es
Bengali: bn	Hebrew: iw	Norwegian: no	Swedish: sv
Bosnian: bs	Hindi: hi	Oriya: or	Tagalog: tl

Source: https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query

Twitter's Advanced Search Operators: Filter by **Tweet Type**

- **is:retweet**
- **is:reply**
- **is:quote**
- **is:verified**
  (includes tweets from verified accounts only)
- **-is:nullcast**
  (excludes promotional tweets)

Examples:

#Coronavirus lang:en **is:retweet**

#Coronavirus lang:fr **-is:reply**

#Coronavirus lang:uk **is:verified**

`is:retweet`	Conjunction required	Core	Matches on Retweets that match the rest of the specified rule. This operator looks only for true Retweets (for example, those generated using the Retweet button). Quote Tweets will not be matched by this operator. Example: `data @twitterdev -is:retweet`
`is:reply`	Conjunction required	Core	Deliver only explicit replies that match a rule. Can also be negated to exclude replies that match a query from delivery. Note: This operator is also available with the filtered stream endpoint. When used with filtered stream, this operator matches on replies to an original Tweet, replies in quoted Tweets, and replies in Retweets. Example: `from:twitterdev is:reply`
`is:quote`	Conjunction required	Core	Returns all Quote Tweets, also known as Tweets with comments. Example: `"sentiment analysis" is:quote`
`is:verified`	Conjunction required	Core	Deliver only Tweets whose authors are verified by Twitter. Example: `#nowplaying is:verified`
`-is:nullcast`	Conjunction required	Advanced	Removes Tweets created for promotion only on ads.twitter.com that have a `"source":"Twitter for Advertisers (legacy)"` or `"source":"Twitter for Advertisers".` This operator must be negated.

Source: https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query

Twitter's Advanced Search Operators: Filter by **Tweet Content**

- has:mentions
- has:media
- has:images
- has:videos
- has:geo
  (includes geo-coded tweets and excludes retweets)

Examples:

(Coronavirus **OR** COVID**) has:mentions**

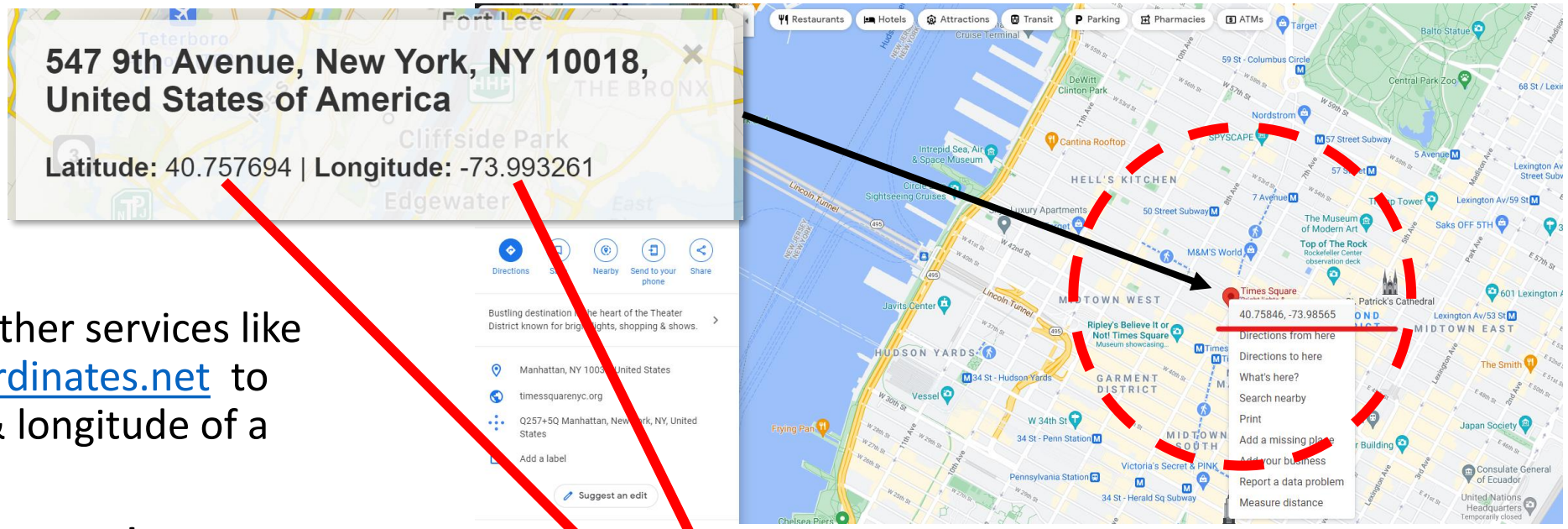
#Coronavirus is:reply **-has:media**

#Coronavirus lang:uk **has:geo**

`has:mentions`	Conjunction required	Core	Matches Tweets that mention another Twitter user. Example: `#nowplaying has:mentions`
`has:media`	Conjunction required	Core	Matches Tweets that contain a media object, such as a photo, GIF, or video, as determined by Twitter. This will not match on media created with Periscope, or Tweets with links to other media hosting sites. Example: `(kittens OR puppies) has:media`
`has:images`	Conjunction required	Core	Matches Tweets that contain a recognized URL to an image. Example: `#meme has:images`
`has:videos`	Conjunction required	Core	Matches Tweets that contain native Twitter videos, uploaded directly to Twitter. This will not match on videos created with Periscope, or Tweets with links to other video hosting sites. Example: `#icebucketchallenge has:videos`
`has:geo`	Conjunction required	Advanced	Matches Tweets that have Tweet-specific geolocation data provided by the Twitter user. This can be either a location in the form of a Twitter place, with the corresponding display name, geo polygon, and other fields, or in rare cases, a geo lat-long coordinate.

Source: https://developer.twitter.com/en/docs/twitter-api/tweets/search/integrate/build-a-query

Twitter's Geo-based Search Operators: **Point Radius**



Use Google Map or other services like https://www.gps-coordinates.net to identify the latitude & longitude of a desired location.

Example for **Times Square, Manhattan, NY**

Latitude: 40.757694

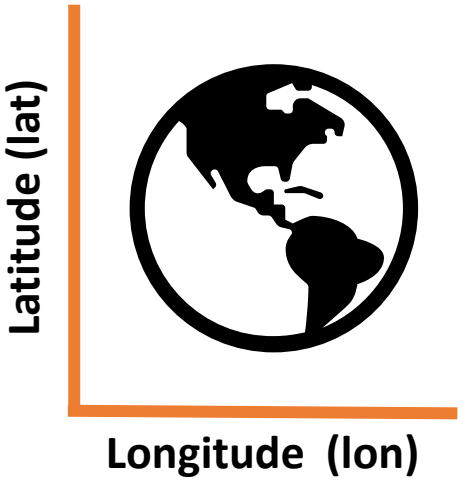
Longitude: -73.9932615

Example search:

Trump point_radius:[-73.9932615 40.757694 40km]

point_radius

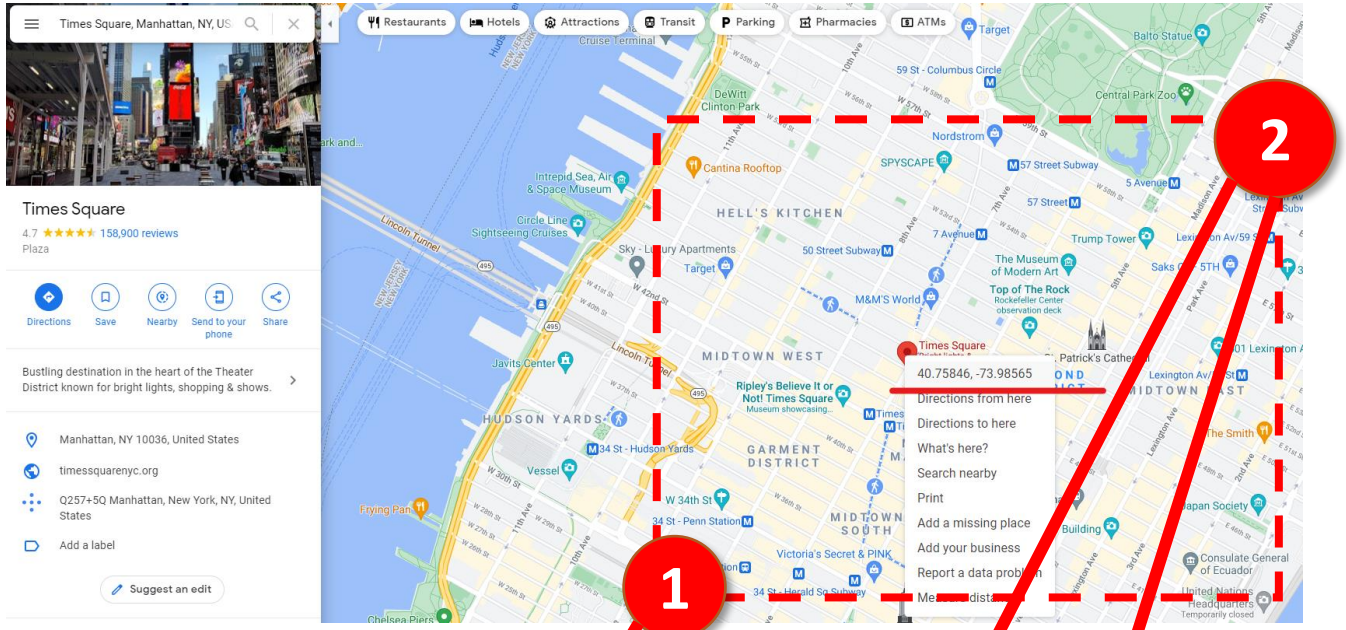
- This operator specifies a circular geographic area to match geo-coded tweets that fall within that area
- Define a central lon-lat coordinate, and then set the radius (up to 25 miles).
- Use: point_radius:[lon lat radius]



Twitter's Geo-based Search Operators: **Bounding Box**

Sample Query:

Trump bounding_box:[west_long south_lat east_long north_lat]



bounding_box

- Specify a 4-sided geographic area and match geo-coded tweets that fall within that area.

bounding_box:[west_long **south_lat** east_long **north_lat**]

Twitter's Geo-based Search Operators: **Country Code**

Based on a country code

Examples:

place_country:CA

place_country:GB

https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2

Officially assigned code elements [edit]

The following is a complete list of the 249 current officially assigned ISO 3166-1 alpha-2 codes, with the following columns:[1]

- **Code**: ISO 3166-1 alpha-2 code
- **Country name**: English short name officially used by the ISO 3166 Maintenance Agency (ISO 3166/MA)
- **Year**: Year when alpha-2 code was first officially assigned (1974, first edition of ISO 3166)
- **ccTLD**: Corresponding country code top-level domain (note that some are inactive); exceptions where another ccTLD is assigned for the country are shown in pa
- **ISO 3166-2**: Corresponding ISO 3166-2 codes
- **Notes**: Any unofficial notes

Code	Country name (using title case)	Year	ccTLD	ISO 3166-2	Notes
AD	Andorra	1974	.ad	ISO 3166-2:AD	
AE	United Arab Emirates	1974	.ae	ISO 3166-2:AE	
AF	Afghanistan	1974	.af	ISO 3166-2:AF	
AG	Antigua and Barbuda	1974	.ag	ISO 3166-2:AG	
AI	Anguilla	1985	.ai	ISO 3166-2:AI	AI previously represented French Afars and Issas
AL	Albania	1974	.al	ISO 3166-2:AL	
AM	Armenia	1992	.am	ISO 3166-2:AM	
AO	Angola	1974	.ao	ISO 3166-2:AO	
AQ	Antarctica	1974	.aq	ISO 3166-2:AQ	Covers the territories south of 60° south latitude Code taken from name in French: *Antarctique*
AR	Argentina	1974	.ar	ISO 3166-2:AR	
AS	American Samoa	1974	.as	ISO 3166-2:AS	
AT	Austria	1974	.at	ISO 3166-2:AT	
AU	Australia	1974	.au	ISO 3166-2:AU	Includes the Ashmore and Cartier Islands and the Coral Sea Islands
AW	Aruba	1986	.aw	ISO 3166-2:AW	
AX	Åland Islands	2004	.ax	ISO 3166-2:AX	An autonomous province of Finland

Twitter Data & Metadata

+

1. Twitter API → 2. Collection → **3. Metadata** → 4. Export → 5. Sharing

Twitter Data & Metadata: **Tweet Object**

Tweet as shown on Twitter.com

Metadata associated with the tweet


```
{
 "created_at":"Thu Apr 06 15:24:15 +0000 2017",
 "id": 850006245121695744,
 "id_str": "850006245121695744",
 "text": "1/ Today we're sharing our vision for the future of the Twitter API platt
 "user": {},
 "entities": {}
}
```


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

Tweet object: Location related fields

Tweet as shown on Twitter.com

Metadata associated with the tweet

```
"place":
{
    "attributes":{},
     "bounding_box":
    {
        "coordinates":
        [[
                [-77.119759,38.791645],
                [-76.909393,38.791645],
                [-76.909393,38.995548],
                [-77.119759,38.995548]
        ]],
        "type":"Polygon"
    },
     "country":"United States",
     "country_code":"US",
     "full_name":"Washington, DC",
     "id":"01fbe706f872cb32",
     "name":"Washington",
     "place_type":"city",
     "url": "http://api.twitter.com/1/geo/id/01fbe706f872cb32.json"
}
```

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

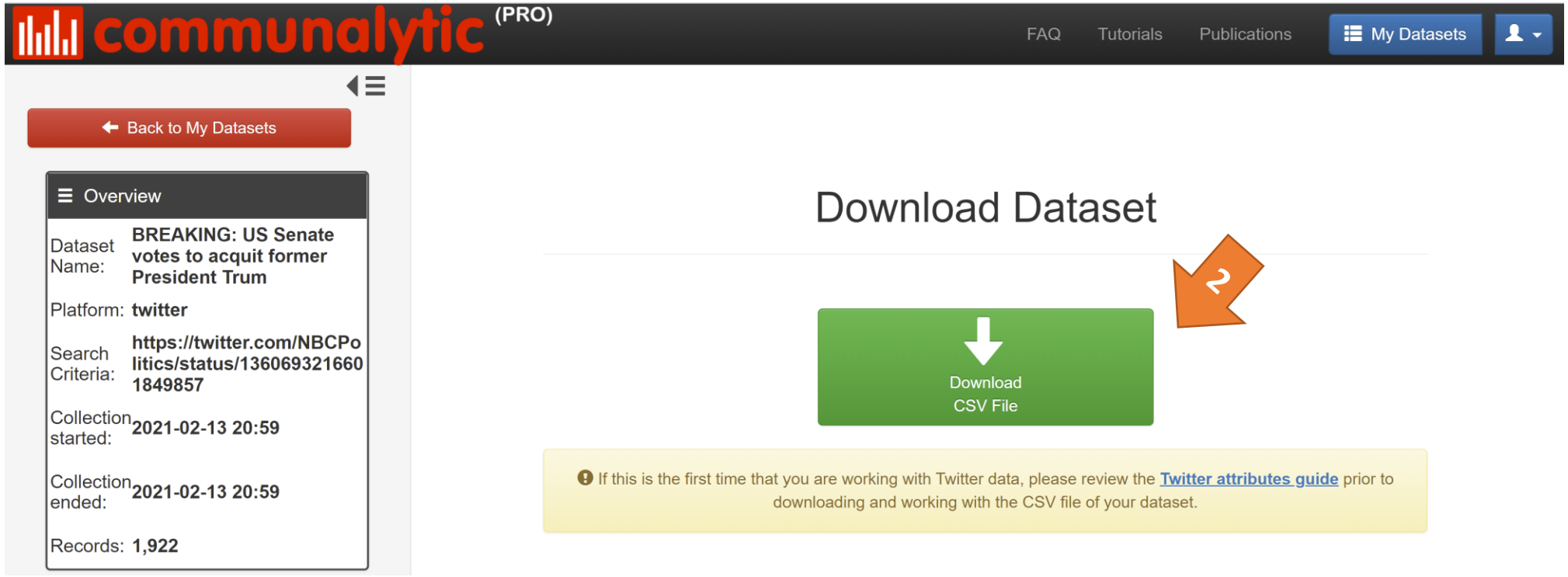
Tweet object: User related fields

Tweet as shown on Twitter.com

```
{
  "user": {
    "id": 2244994945,
    "id_str": "2244994945",
    "name": "TwitterDev",
    "screen_name": "TwitterDev",
    "location": "Internet",
    "url": "https://dev.twitter.com/",
    "description": "Your official source for Twitter Platform news, up
    "verified": true,
    "followers_count": 477684,
    "friends_count": 1524,
    "listed_count": 1184,
    "favourites_count": 2151,
    "statuses_count": 3121,
    "created_at": "Sat Dec 14 04:35:55 +0000 2013",
    "utc_offset": -25200,
    "time_zone": "Pacific Time (US & Canada)",
    "geo_enabled": true,
    "lang": "en",
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/5
  }
}
```

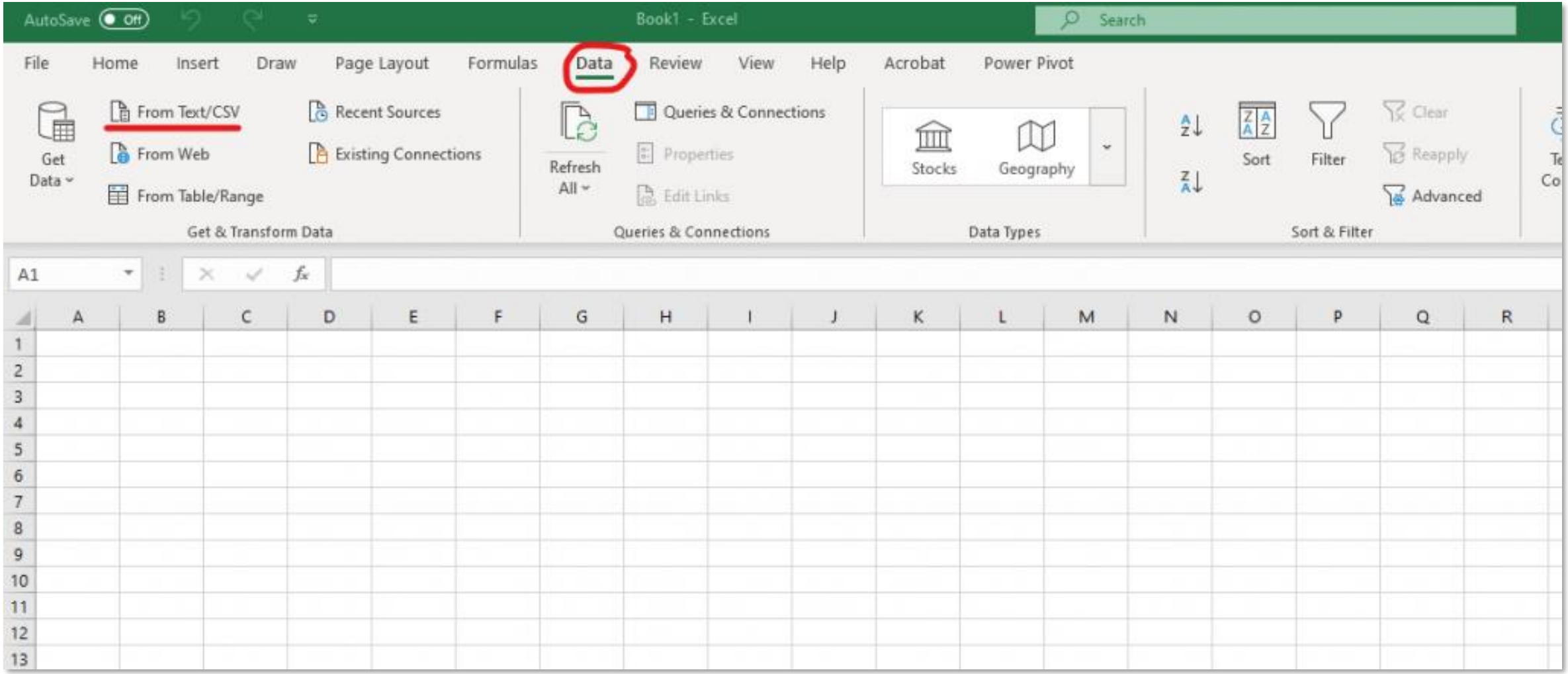
Export Twitter Dataset from Communalytic

BREAKING: US Senate votes to acquit former President Trum twitter : https://twitter.com/NBCPolitics/status/1360693216601849857	1,922					



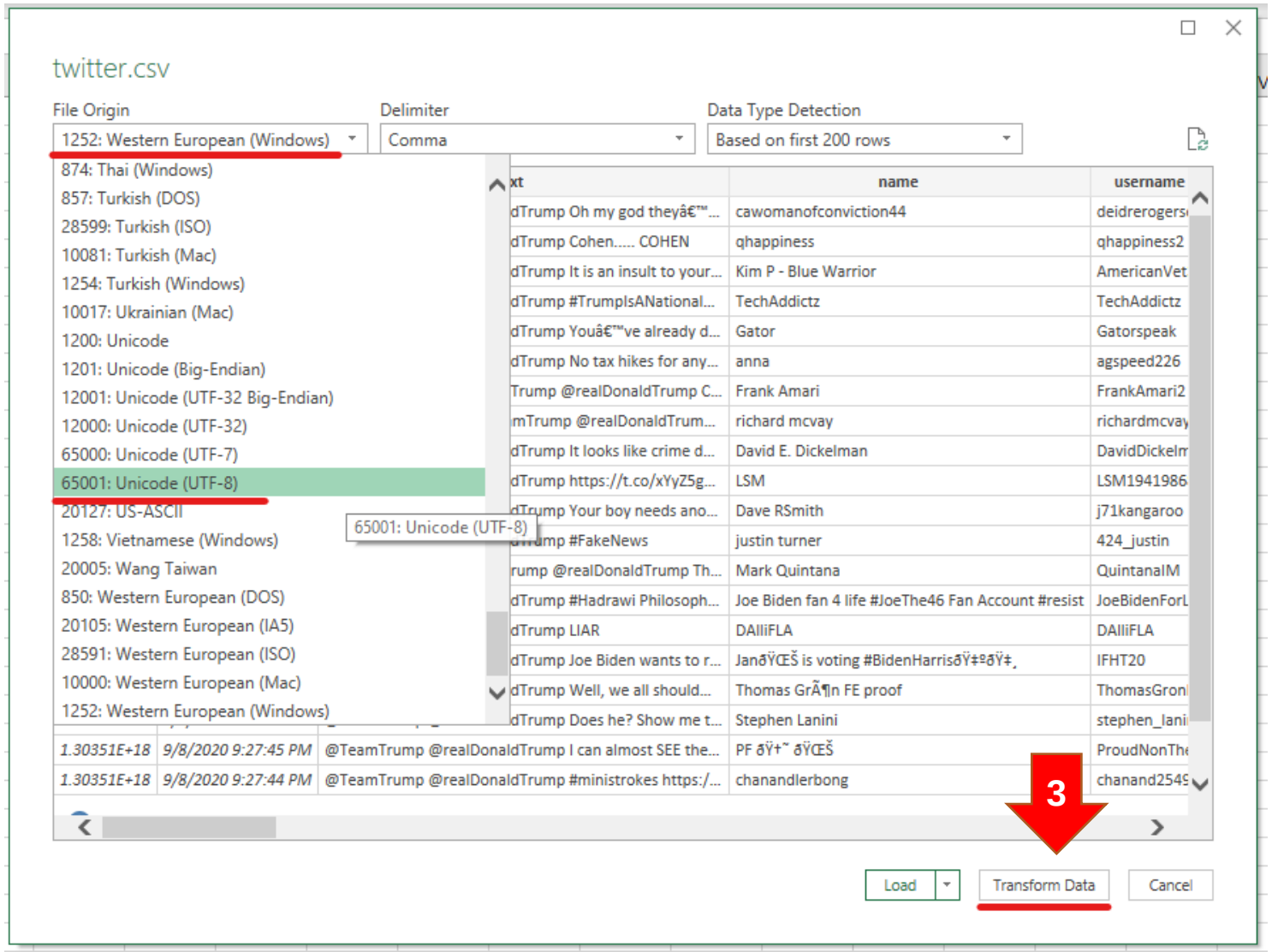
1. Twitter API → 2. Collection → 3. Metadata → **4. Export** → 5. Sharing

How to open the exported CSV file in Excel (1)

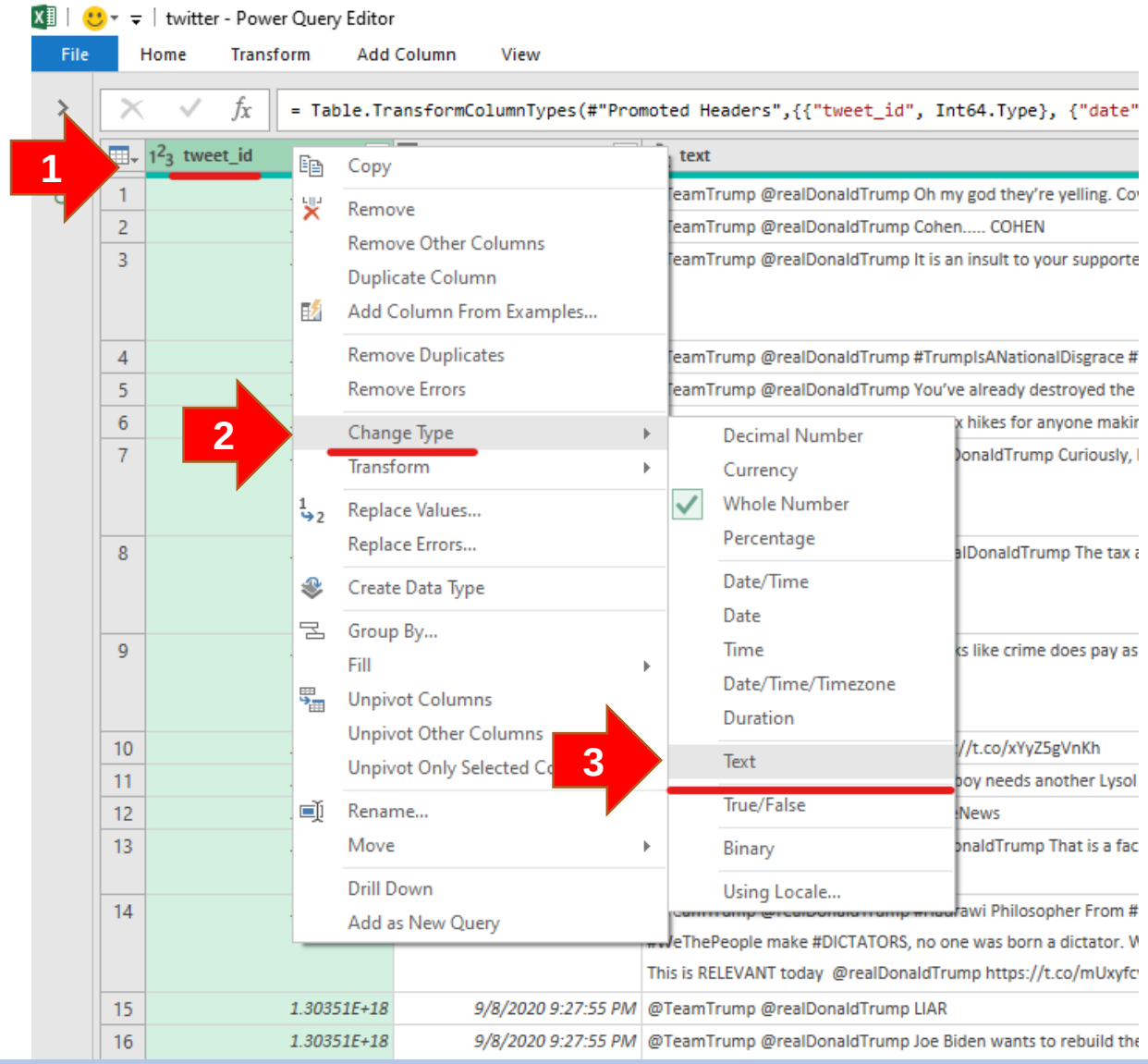


How to open the exported CSV file in Excel (2)

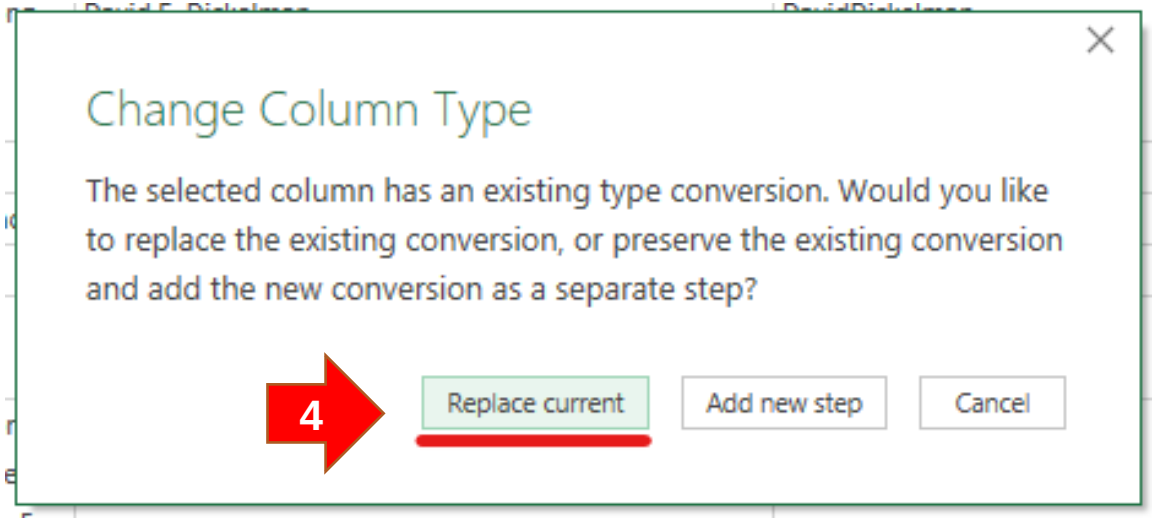
UTF-8



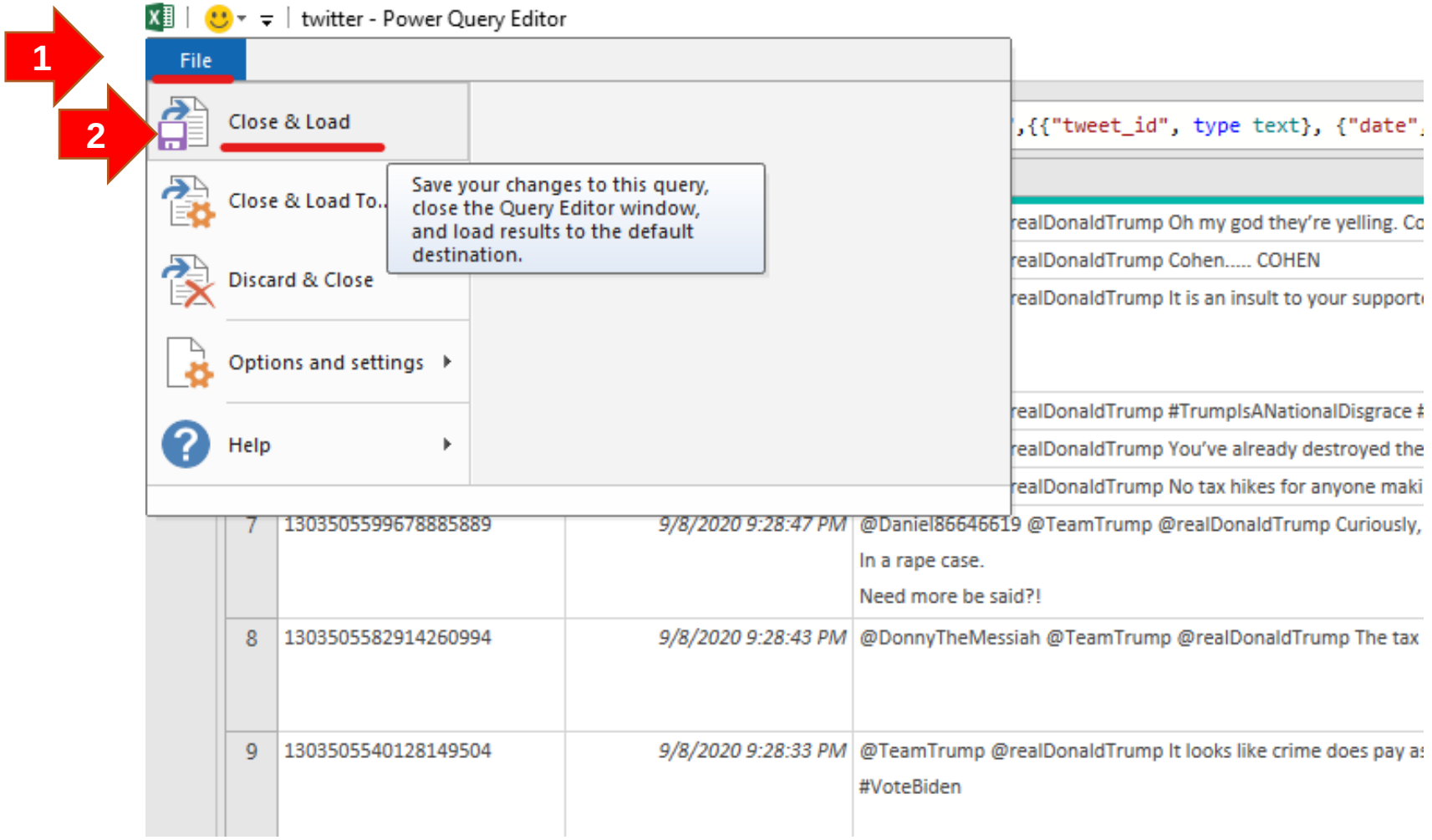
How to open the exported CSV file in Excel (3)



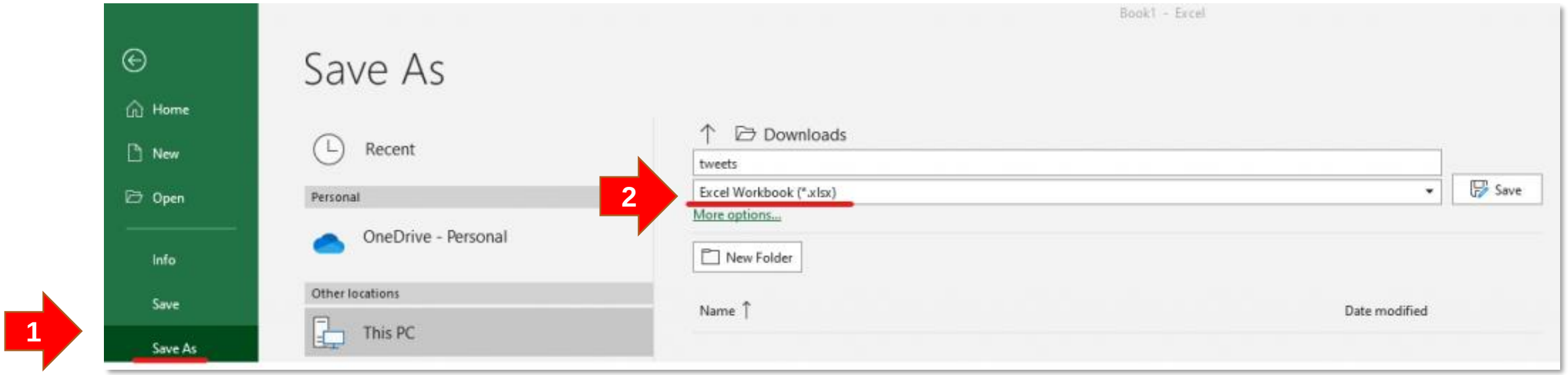
When prompted to click **[Replace current]**



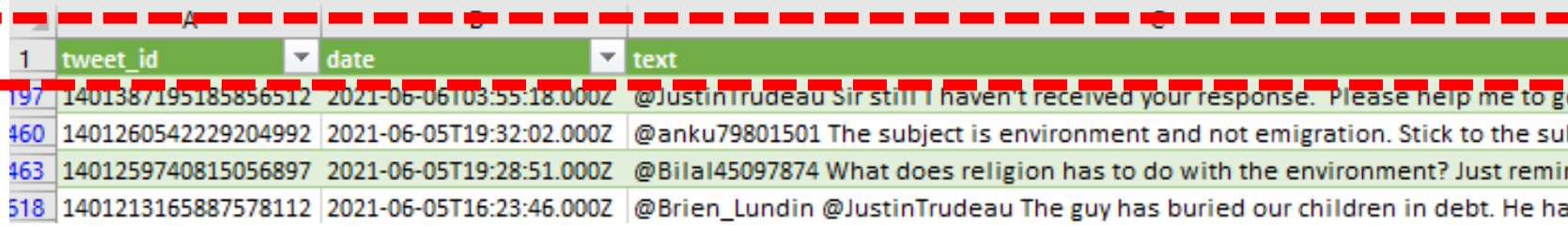
How to open the exported CSV file in Excel (4)



How to open the exported CSV file in Excel (5)



Understanding Twitter Metadata in the exported file (part 1/3)



1	tweet_id	date	text
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received your response. Please help me to g
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment and not emigration. Stick to the su
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to do with the environment? Just remi
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy has buried our children in debt. He ha

tweet_id	1401387195185856512
date	2021-06-06T03:55:18.000Z
text	@JustinTrudeau Sir still I haven't received your response. […]
name	First Last Name
username	username123
userid	1217946015
in_reply_to_tweet_id	1401162543142481920
in_reply_to_username	JustinTrudeau
in_reply_to_userid	14260960
language	en
quote_count	0
retweet_count	0
reply_count	0
like_count	0
possibly_sensitive	False
user_verified	False
user_location	Hyderabad, India
user_created_at	2013-12-28T11:39:37.000Z
user_protected	False
user_description	
user_followers_count	0
user_following_count	29

Understanding Twitter Metadata in the exported file (part 2/3)

1	tweet_id	date	text
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received your response. Please help me to g
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment and not emigration. Stick to the su
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to do with the environment? Just remi
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy has buried our children in debt. He ha

in_reply_to_user_verified	True
in_reply_to_user_location	Papineau
in_reply_to_user_created_at	2008-03-30T21:04:14.000Z
in_reply_to_user_protected	False
in_reply_to_user_description	Father, husband, 23rd Prime Minister of Canada [...]
in_reply_to_user_followers_count	5640556
in_reply_to_user_following_count	941
in_reply_to_user_listed_count	14269
in_reply_to_user_tweet_count	34853
source	Twitter for Android
referenced_tweet_id	1401162543142481920
referenced_tweet_type	replied_to
reply_settings	everyone
referenced_tweet	False
notes	
geo_coded	True
geo_country_code	CA
geo_country	Canada
geo_place_type	city
geo_full_name	Vancouver, British Columbia
geo_name	Vancouver

Referenced tweet

Retweet

1	tweet_id	date	text
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy ha

in_reply_to_user_verified	
in_reply_to_user_location	
in_reply_to_user_created_at	
in_reply_to_user_protected	
in_reply_to_user_description	
in_reply_to_user_followers_count	
in_reply_to_user_following_count	
in_reply_to_user_listed_count	
in_reply_to_user_tweet_count	
source	
referenced_tweet_id	
referenced_tweet_type	retweeted | replied_to | qu
reply_settings	
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	

1	tweet_id	date	text
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy ha

Referenced tweet

CoinMarketCap @CoinMarketCap · Mar 23

Do any of you use @steemit? If you do, why not give us a follow?

Also what sort of content would you like us to share on the platform?

#Crypto #Cryptocurrency #Altcoin #Steemit $STEEM

@officialcmc

The latest posts from Unofficial Channel. Follow me at @officialcmc. CoinMarketCap is your go-to plac...

steemit.com

72 29 106

Reply

They Call Me Stevenson

Replying to @CoinMarketCap and @steemit

I'd give you a follow if you sign up on #hive.

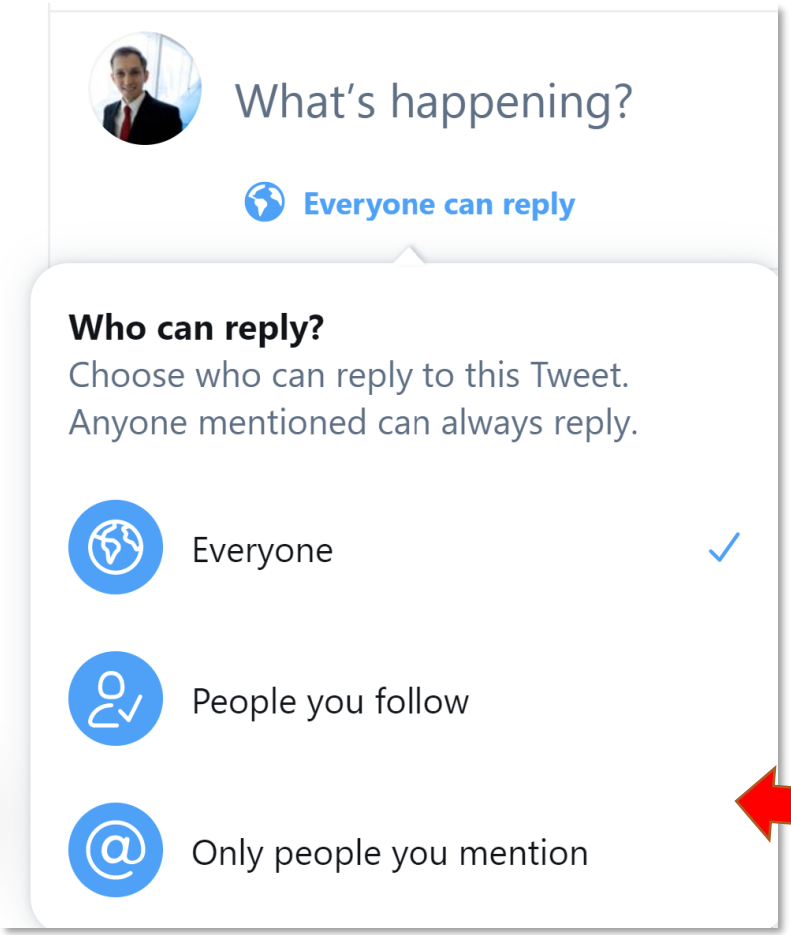
in_reply_to_user_verified	
in_reply_to_user_location	
in_reply_to_user_created_at	
in_reply_to_user_protected	
in_reply_to_user_description	
in_reply_to_user_followers_count	
in_reply_to_user_following_count	
in_reply_to_user_listed_count	
in_reply_to_user_tweet_count	
source	
referenced_tweet_id	
referenced_tweet_type	retweeted | **replied_to** | qu
reply_settings	
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	

1	tweet_id	date	text
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received y
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy ha

quote

Referenced tweet

in_reply_to_user_verified	
in_reply_to_user_location	
in_reply_to_user_created_at	
in_reply_to_user_protected	
in_reply_to_user_description	
in_reply_to_user_followers_count	
in_reply_to_user_following_count	
in_reply_to_user_listed_count	
in_reply_to_user_tweet_count	
source	
referenced_tweet_id	
referenced_tweet_type	quoted
reply_settings	
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	



https://help.twitter.com/en/using-twitter/twitter-conversations

1	tweet_id	date	text
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received your response. Please help me to g
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment and not emigration. Stick to the su
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to do with the environment? Just remi
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy has buried our children in debt. He ha

in_reply_to_user_verified	
in_reply_to_user_location	
in_reply_to_user_created_at	
in_reply_to_user_protected	
in_reply_to_user_description	
in_reply_to_user_followers_count	
in_reply_to_user_following_count	
in_reply_to_user_listed_count	
in_reply_to_user_tweet_count	
source	
referenced_tweet_id	
referenced_tweet_type	
reply_settings	**everyone | following | mentionedUsers**
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	

1	tweet_id	date	text
197	1401387195185856512	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received your response. Please help me to g
460	1401260542229204992	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment and not emigration. Stick to the su
463	1401259740815056897	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to do with the environment? Just remi
518	1401213165887578112	2021-06-05T16:23:46.000Z	@Brien_Lundin @JustinTrudeau The guy has buried our children in debt. He ha

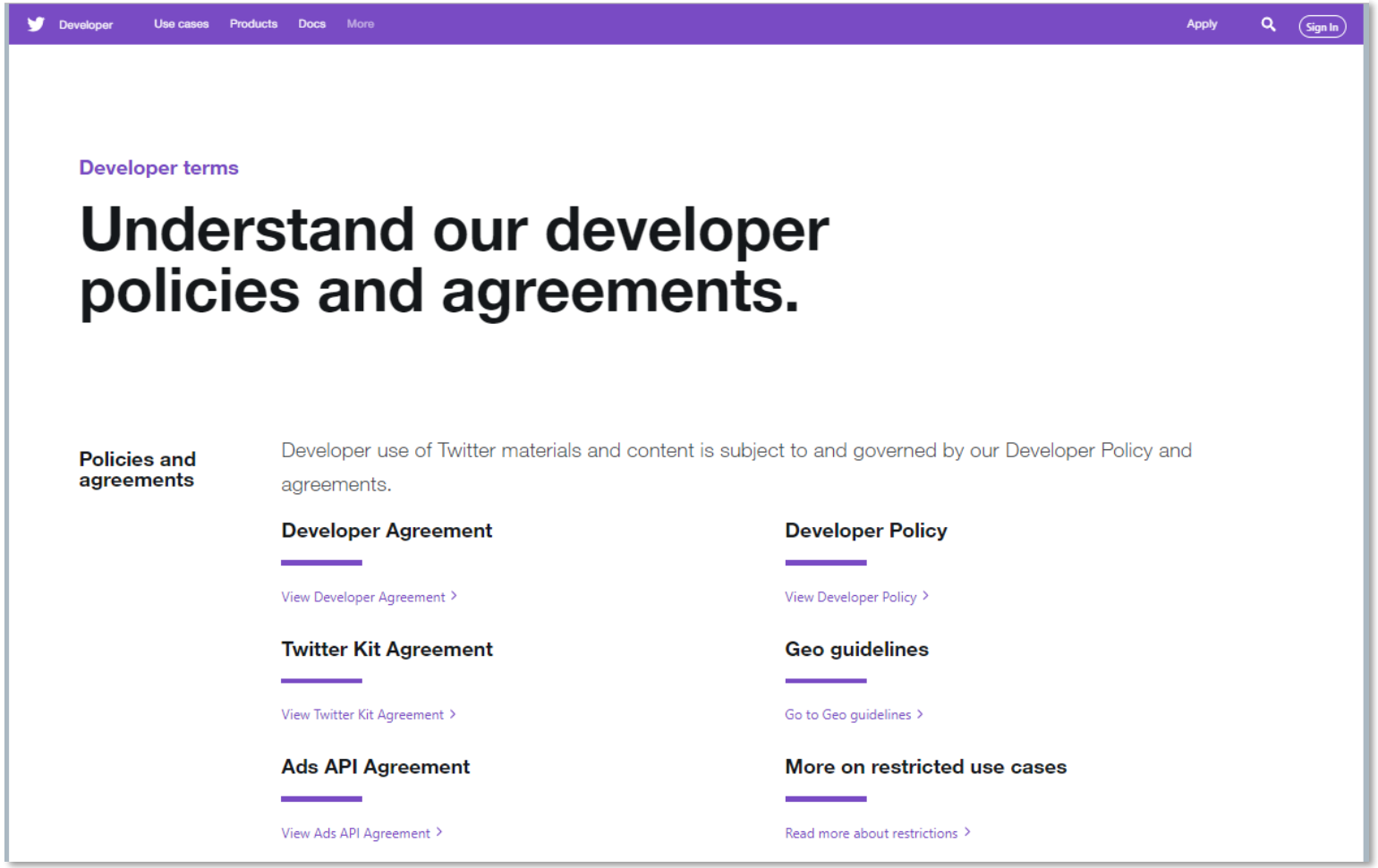
in_reply_to_user_verified	
in_reply_to_user_location	
in_reply_to_user_created_at	
in_reply_to_user_protected	
in_reply_to_user_description	
in_reply_to_user_followers_count	
in_reply_to_user_following_count	
in_reply_to_user_listed_count	
in_reply_to_user_tweet_count	
source	
referenced_tweet_id	
referenced_tweet_type	
reply_settings	
referenced_tweet	**True | False**
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	

Understanding Twitter Metadata in the exported file (part 3/3)

https://twitter.com/JustinTrudeau/status/1401162541686984706

	C	D	G	H	I	P	Q	AI	AJ	AL	AM
1	text	name	in_reply_t	in_reply_t	in_reply_t	user_verif	user_locat	referenced_tweet_id	referenced_tweet_typ	referenced_tweet	notes
	It's #WorldEnvironmentDay. It's a day to celebrate the environment - and it's a day for us to recommit ourselves to protecting it. Here's some of what we've done and what we're doing to build a cleaner future for our kids and grandkids:	Justin Trudeau	null			True	Papineau			True	This tweet is included because it was referenced by another tweet in this dataset. It may not correspond directly to your search criteria.

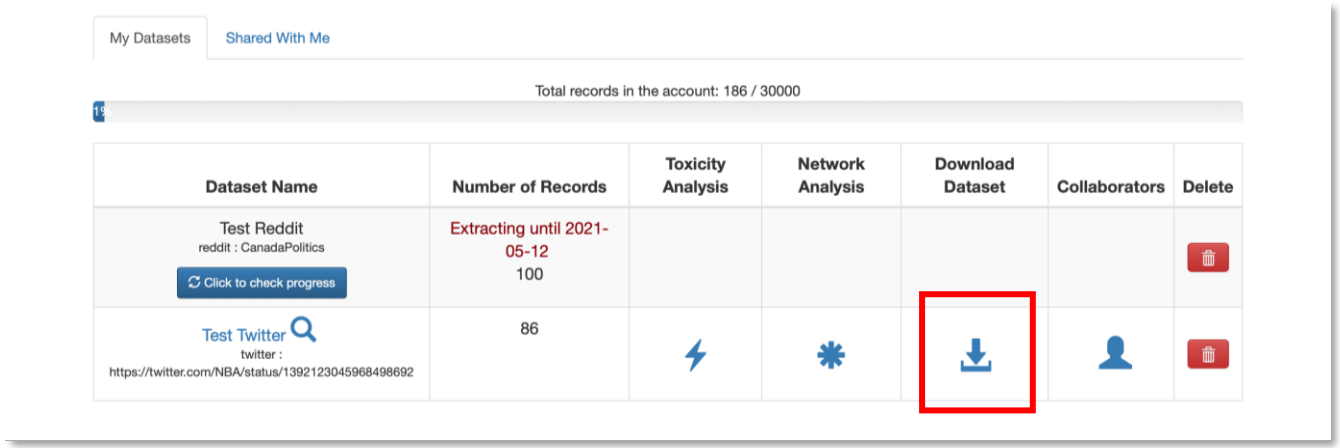
Data Sharing: Twitter's Terms of Service

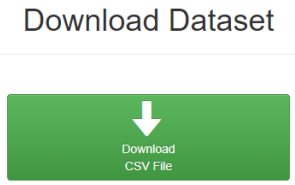


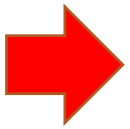
1. Twitter API → 2. Collection → 3. Metadata → 4. Export → 5. Sharing

Data Sharing via Tweet Ids

Collect / Export tweets as CSV







Only share the 'tweet_id' column

	A	B	
1	tweet_id	date	text
2	1401598862201180162	2021-06-06T17:56:23.000Z	@JustinTrudeau
3	1401528927466639360	2021-06-06T13:18:30.000Z	This is real #genocide.https://t.co/
4	1401599908059267073	2021-06-06T18:00:33.000Z	@JustinTrudeau My hobby is a barb
5	1401598610551181312	2021-06-06T17:55:23.000Z	@JustinTrudeau @JustinTrudeau y
6	1401597964896722945	2021-06-06T17:52:49.000Z	@capano_rosa @JustinTrudeau @J
7	1401597739339718659	2021-06-06T17:51:56.000Z	@JustinTrudeau @marcomendicin
8	1401597661833142275	2021-06-06T17:51:37.000Z	@capano_rosa @JustinTrudeau @J
9	1401596728000892933	2021-06-06T17:47:55.000Z	@JustinTrudeau Open the fucking
10	1401591680587812865	2021-06-06T17:27:51.000Z	@JustinTrudeau Everyone is pretty
11	1401589438581415936	2021-06-06T17:18:57.000Z	@JustinTrudeau Yeah maybe... but
12	1401589398773374983	2021-06-06T17:18:47.000Z	@JustinTrudeau Science is a proces
13	1401586720320090117	2021-06-06T17:08:09.000Z	@JustinTrudeau how many of the
14	1401586531563827200	2021-06-06T17:07:24.000Z	@JustinTrudeau Shame on you and
15	1401579144132108290	2021-06-06T16:38:02.000Z	@JustinTrudeau NDP & Conse
16	1401578885427482624	2021-06-06T16:37:01.000Z	@JustinTrudeau Taxing Cdns to dea
17	1401577177846538245	2021-06-06T16:30:13.000Z	@JustinTrudeau Thank you PM Tru
18	1401574285643157504	2021-06-06T16:18:44.000Z	@JustinTrudeau https://t.co/na4uA
19	1401573815537176576	2021-06-06T16:16:52.000Z	@JustinTrudeau https://t.co/6ez0K
20	1401573306952728577	2021-06-06T16:14:51.000Z	@JustinTrudeau This is a great effo
21	1401572148536221705	2021-06-06T16:10:14.000Z	@JustinTrudeau It's a humble requ
22	1401571865877913606	2021-06-06T16:09:07.000Z	@JustinTrudeau It's a humble requ
23	1401571804200673281	2021-06-06T16:08:52.000Z	@JustinTrudeau @marcomendicino
24	1401571047640555528	2021-06-06T16:05:52.000Z	@JustinTrudeau Those trees were
25	1401568121698373641	2021-06-06T15:54:14.000Z	@JustinTrudeau So Happyyyy dear

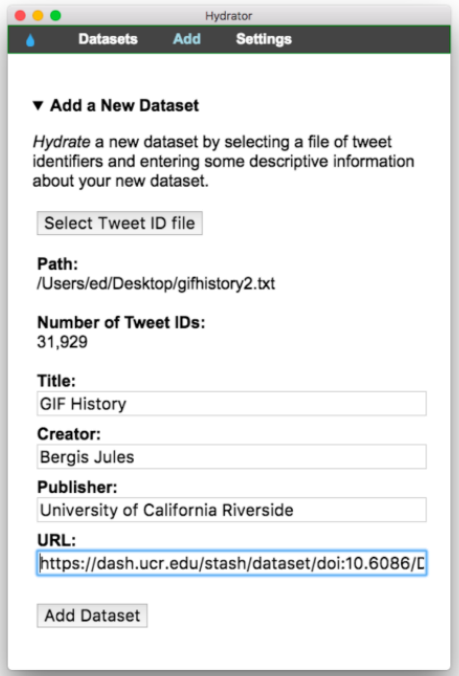
1. Twitter API → 2. Collection → 3. Metadata → 4. Export → **5. Sharing**

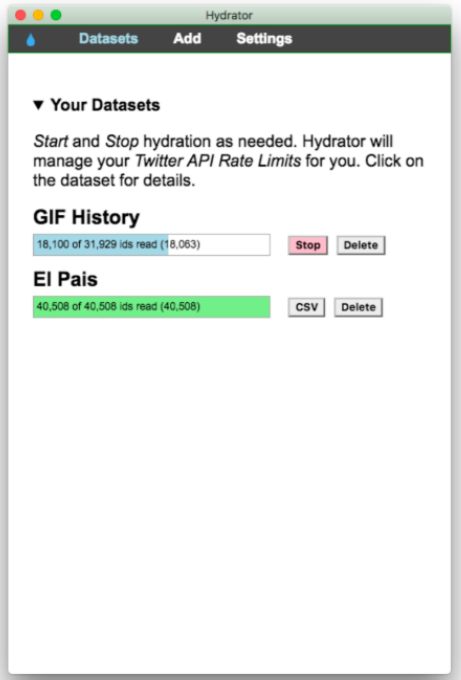
Others can recollect tweets using programs like Hydrator

tweet_id
1401598862201180162
1401528927466639360
1401599908059267073
1401598610551181312
1401597964896722945
1401597739339718659
1401597661833142275
1401596728000892933
1401591680587812865
1401589438581415936
1401589398773374983
1401586720320090117
1401586531563827200

Hydrator

https://github.com/DocNow/hydrator





1. Twitter API → 2. Collection → 3. Metadata → 4. Export → 5. Sharing

Sample Tweet ID Datasets at https://catalog.docnow.io

Catalog

The DocNow Catalog is a collectively curated listing of Twitter datasets. Public datasets are shared as Tweet IDs, which can be hydrated back into full datasets using our Hydrator desktop application.

133 Records comprising 5,884,834,305 tweets ADD RECORD

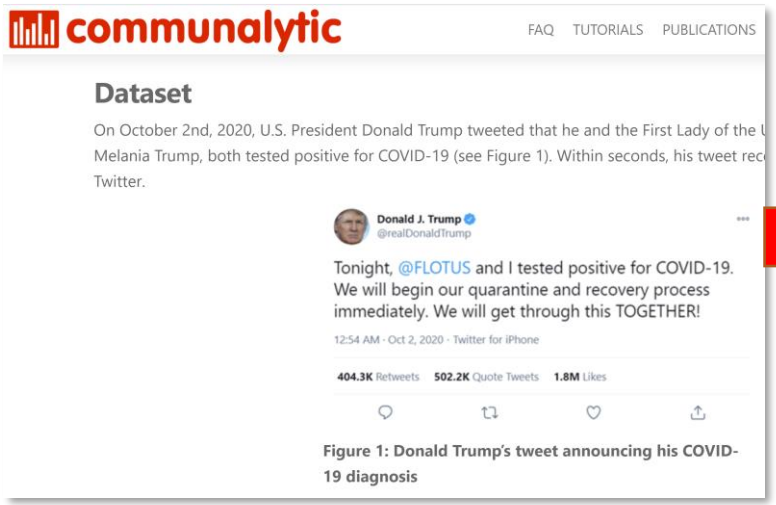
Subjects: All | Tweets Start: 03/21/2006 | Tweets End: 06/07/2021 | Search

Note: all metadata is shared under a CC0 license. Please read our Code of Conduct for more information about contributing datasets.

ADDED	DATE RANGE	TITLE	TWEET COUNT	CREATORS	SUBJECTS	REPOSITORY
05/06/2021	03/21/2006 to 01/31/2021	JOSA - Jopara (Guarani-dominant mixed with Spanish) Sentiment Analysis corpus.	3,941	Agüero-Torales, Marvin M., Vilares, David, López-Herrera, Antonio G.	jopara, guarani, sentiment analysis, code-switching, natural language processing	Harvard Dataverse
01/29/2021	05/25/2020 to 07/25/2020	Fortune 100 response to 2020 BLM Protests	6,184	Kevin McElwee	Black Lives Matter, Fortune 100	GitHub
01/03/2021	03/01/2020 to 12/31/2020	2020 US Presidential Election	965,620,919	Elizabeth Chen, Emilio Ferrara	United States, Politics, Elections	GitHub
01/01/2021	11/18/2020 to 11/30/2020	Diego Maradona	1,644,242	Alejandra Josiowicz, Ed Summers	Football, Argentina	Internet Archive
12/30/2020	12/10/2020 to 12/19/2020	SolarWinds	364,170	Ed Summers	Cybersecurity, Russia, United States	Internet Archive

1. Twitter API → 2. Collection → 3. Metadata → 4. Export → 5. Sharing

Sample Dataset of Tweet IDs



https://communalytic.com/2021/01/13/toxicity-analysis-of-a-twitter-thread/

https://doi.org/10.5683/SP2/BMCGI3

1. Twitter API → 2. Collection → 3. Metadata → 4. Export → 5. Sharing