



# Session 3: Getting Started with Commanalytic

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

June 10, 2021

SocialMediaLab.ca



# The Lab & The People

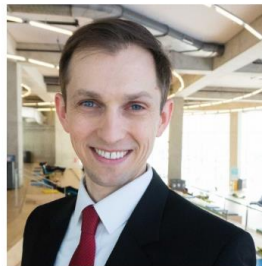
“Making Sense of a Networked World”

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 Gruz**  
Canada Research  
Chair, Associate  
Professor, Director of  
Research at the  
Ryerson University  
Social Media Lab



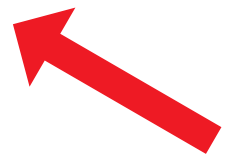
**Philip Mai M.A., J.D.**  
Co-Director and  
Senior Researcher at  
the Ryerson  
University Social  
Media Lab

Videos and slides from Session 1 & 2 are now available online at: [communalytic.com](https://communalytic.com)



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: 2<sup>nd</sup> and 4<sup>th</sup> 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


<b>Session #1</b>	<b>Getting Started with Communalytic: Data Collection from Reddit</b>	<b>May 13, 2021, 10:00- 11:30am (EDT)</b>
<b>Session #2</b>	<b>Toxicity Analysis with Reddit Data using Perspective API</b>	<b>May 27, 2021, 10:00- 11:30am (EDT)</b>
<b>Session #3</b>	<b>Getting Started with Communalytic: Data Collection from Twitter (Twitter Thread via API v2.0 and Twitter Academic Track)</b>	<b>June 10, 2021, 10:00- 11:30am (EDT)</b>
<b>Session #4</b>	<b>Toxicity Analysis of Twitter data using Perspective API</b> Guest talk by <a href="#">Dr. Paola Pascual-Ferra</a> , 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.	<b>June 24, 2021, 10:00- 11:30am (EDT)</b>
<b>Session #5</b>	<b>Social Network Analysis of Signed Networks with Reddit and Twitter data</b>	<b>July 8, 2021, 10:00- 11:30am (EDT)</b>
<b>Session #6</b>	<b>Getting Started with Communalytic: Data Collection from Facebook &amp; Instagram (via CrowdTangle API) + Social Network Analysis of Two-mode Semantic Networks with CrowdTangle data</b>	<b>July 22, 2021, 10:00- 11:30am (EDT)</b>


@SMLabTO





# Join the Commanalytic Community Group


<https://groups.google.com/u/1/g/commanalytic-community-group>

☰  Groups

 New conversation

 My groups

 Recent groups


▶  Favourite groups

---

☆ Starred conversations


---

Commalytic Community Group

 **Conversations**

Approved

Pending

 People

Members

🔍 Conversations Search conversations within commanalytic-c...



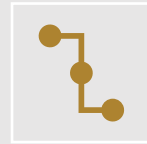
## ☆ Commalytic Community Group



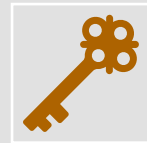
# Outline



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
1 from googleapiclient import discovery
2 import json
3
4 API_KEY = 'copy-your-api-key-here'
5
6 client = discovery.build(
7     "commentanalyzer",
8     "v1alpha1",
9     developerKey=API_KEY,
10    discoveryServiceUrl="https://commentanalyzer.googleapis.com/$discovery/rest?vers
11    static_discovery=False,
12 )
13
14 analyze_request = {
15     'comment': { 'text': 'friendly greetings from python' },
16     'requestedAttributes': {'TOXICITY': {}}
17 }
18
19 response = client.comments().analyze(body=analyze_request).execute()
20 print(json.dumps(response, indent=2))
```

**Using a third-party system such as Commanalytic – no programming knowledge required**

The screenshot shows the Commanalytic (EDU) web interface. The top navigation bar includes links for FAQ, Tutorials, Publications, and My Datasets. The main content area is titled "Toxicity Analysis" and displays a summary for a dataset named "Test" on the subreddit "Coronavirus". The data shows it was collected from 20/04/2021 to 21/04/2021, resulting in 4451 records. Below the summary, there are buttons for "Posts Per Day", "Word Cloud", "Top Ten Posters", "Toxicity Analysis", and "Network Analysis". A "Start Analysis" button is prominently displayed. A yellow information box at the bottom provides details about the toxicity analysis process, including language selection (currently set to English) and a list of supported languages: English, French, German, Italian, Portuguese, and Spanish.

# 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**

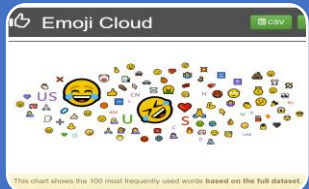
<https://communalytic.com/>



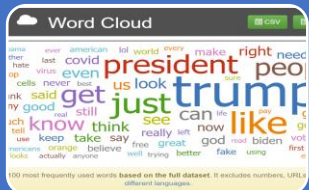
How to choose between Communalytic Edu and Pro.

	 <b>communalytic</b> <b>EDU</b>	 <b>communalytic</b> <b>PRO</b>
<b>Account Type</b>	Free	\$349/6-mo. to support site infrastructure (server-side data collection, storage, processing, analysis and visualization)
<b>Designed For</b>	<b>Students and is ideal for teaching and learning about social media analytics</b>	<b>Academic researchers and is ideal for large scale academic research projects</b>
<b>Account Caps</b>	≤ 30K records shared across 3 datasets	≤ 10M records shared across 50 datasets
<b>Reddit</b>	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)
<b>Twitter Threads (API ver.2)</b> req. <a href="#">Twitter developer's account</a>	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)
<b>Twitter Academic Track (API ver.2)</b> req. <a href="#">Application to Twitter</a>	Not supported	Full-archive historical-search of tweets back to 2006 (Limit: ≤ 10M tweets/month)
<b>CrowdTangle (FB/IG) URL Search</b> req. a <a href="#">CrowdTangle account</a>	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)

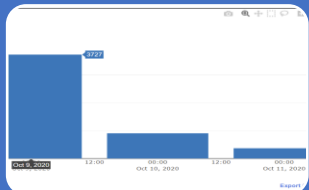
\* Live-collection = the collection of posts/tweets posted on or after the date when you initiated the data collection.



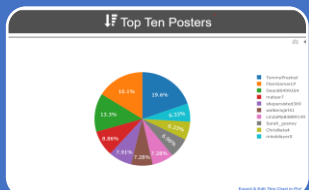
Emoji cloud (freq. used emojis)



Word cloud (freq. used words)



Time series (posts per day)



Top posters (top 10)

## Communalytic Features: Exploratory Data Analysis Module

**Overview**

**Dataset Name:** Trump after COVID hospital

**Platform:** twitter

**Search Criteria:** [https://twitter.com/realDonaldTrump/status/1313959702104023047](https://twitter.comrealDonaldTrump/status/1313959702104023047)

**Collection started:** 2020-10-09 20:49

**Collection ended:** 2020-10-09 20:49

**Records:** 5000

[Print](#)
[Reset Analysis](#)

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

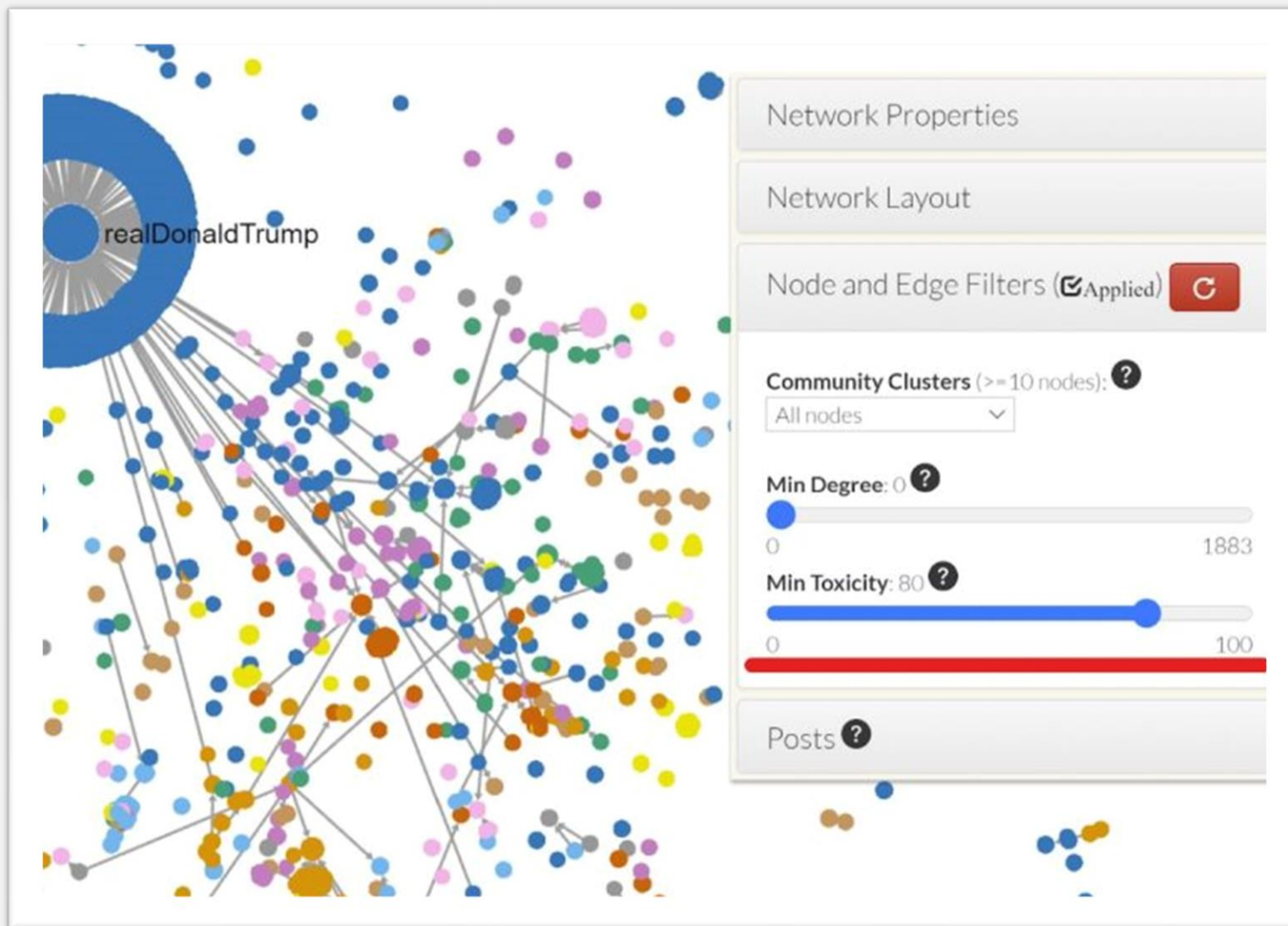
Click on the highest/lowest values to see the top posts of each category

[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	<a href="#">0.53</a>	<a href="#">0.06</a>
Severe toxicity ⓘ	0.09	<a href="#">0.28</a>	<a href="#">0.01</a>
Identity attack ⓘ	0.11	<a href="#">0.18</a>	<a href="#">0.03</a>
Insult ⓘ	0.16	<a href="#">0.76</a>	<a href="#">0.03</a>
Profanity ⓘ	0.13	<a href="#">0.71</a>	<a href="#">0.02</a>
Threat ⓘ	0.24	<a href="#">0.69</a>	<a href="#">0.05</a>
Sexually Explicit ⓘ	0.13	<a href="#">0.23</a>	<a href="#">0.02</a>
Flirtation ⓘ	0.30	<a href="#">0.44</a>	<a href="#">0.08</a>
Attack On Commenter ⓘ	0.52	<a href="#">0.87</a>	<a href="#">0.05</a>

## Communalytic Features: Text Analysis Module

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



\* - 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) \*

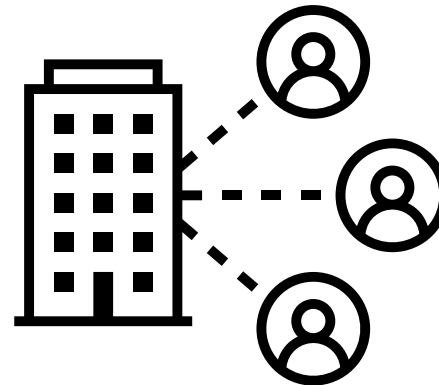
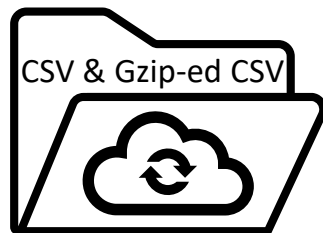
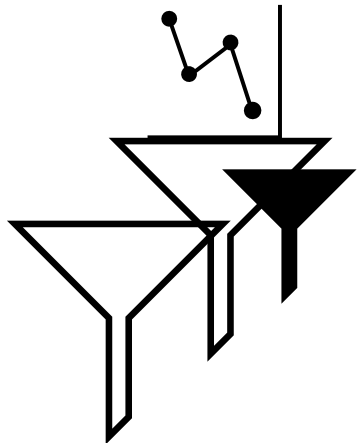
It can also

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

## Communalytic Features: Additional Features

- **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)

# My Datasets

Server Time: June 10, 2021 12:19 UTC

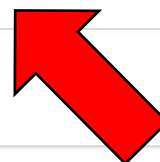
Collect data from

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

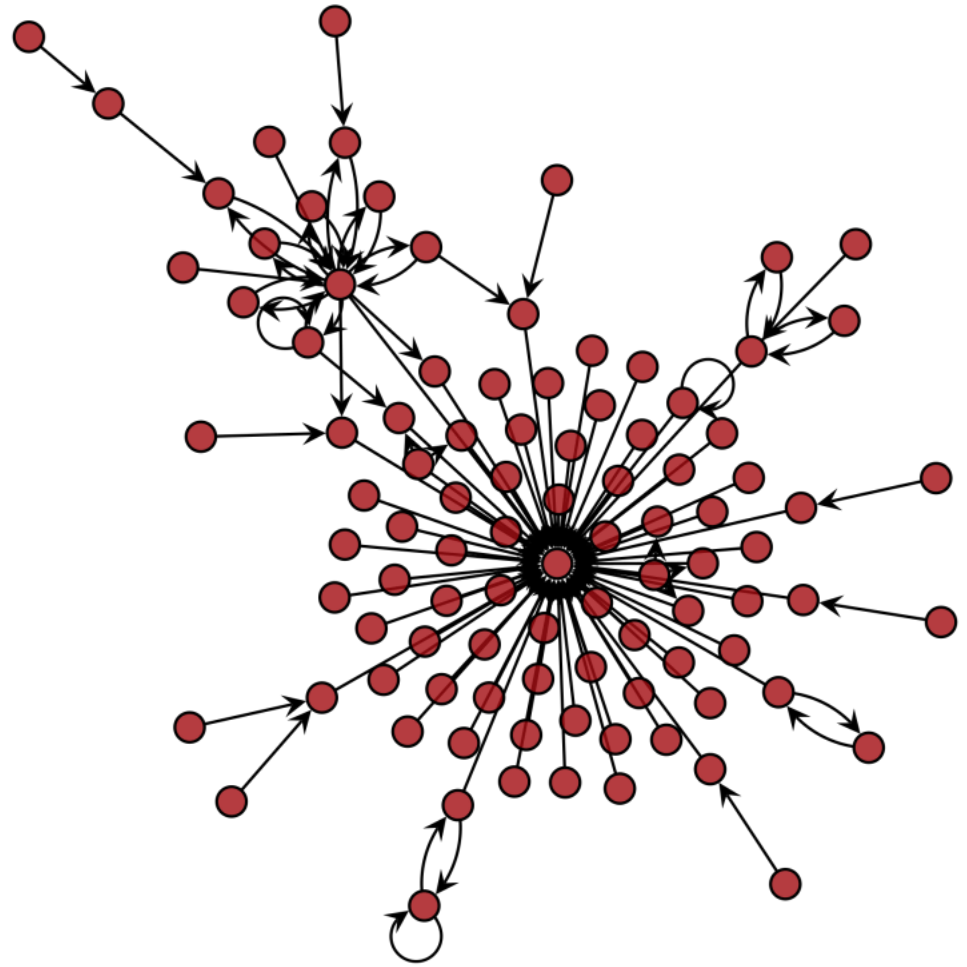
 2

 181



Show replies





A network representation of replies to a given tweet



- Home
- Explore
- Notifications
- Messages
- Bookmarks
- Lists
- Profile
- More

Tweet

← Tweet



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](https://apnews.com)

12:14 PM · Oct 2, 2020 · SocialFlow

18 Retweets 3 Quote Tweets 83 Likes



Server Time: May 11, 2021 14:04 UTC

### Collect data from

Reddit (LIVE)

Twitter Thread

CrowdTangle

CSV File



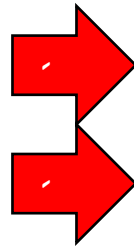
Search dataset list...



[← Back to My Datasets](#)

## Twitter Thread Data Collection

Please enter the following details to begin collection

 ?

Email me once job completes

 [Start Data Collection](#)

Reddit (LIVE)

Twitter Thread

CrowdTangle

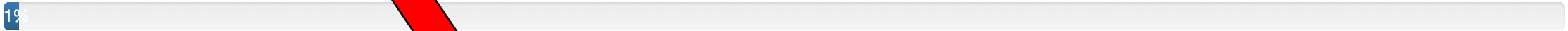
CSV File

Search dataset list...

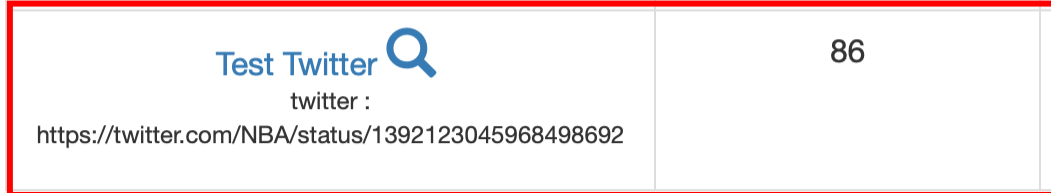
My Datasets

Shared With Me

Total records in the account: 186 / 30000



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




[← Back to My Datasets](#)

### Overview

Dataset Name: **Test Twitter**

Platform: **twitter**

Search Criteria: **https://twitter.com/NBA/status/139212304596849869**  
**2**

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

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

Records: **86**

🕒 Posts Per Day

☁️ Word Cloud

👍 Emoji Cloud

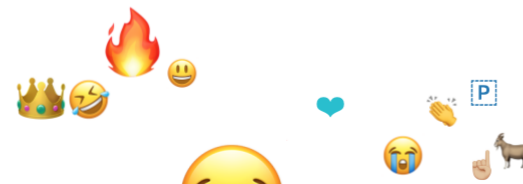
⏴ Top Ten Posters

## 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](#).)

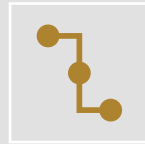
## Emoji Cloud

[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)

Build for businesses

Build for consumers

Do research

Teach and learn

Build for good

Build for fun

# Tap into what's happening to build what's next

Power your next big idea with real-time access to the world's largest public conversation.

<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/>

 **Developer Platform**

Products ▾

Use cases ▾

Docs ▾

Community ▾

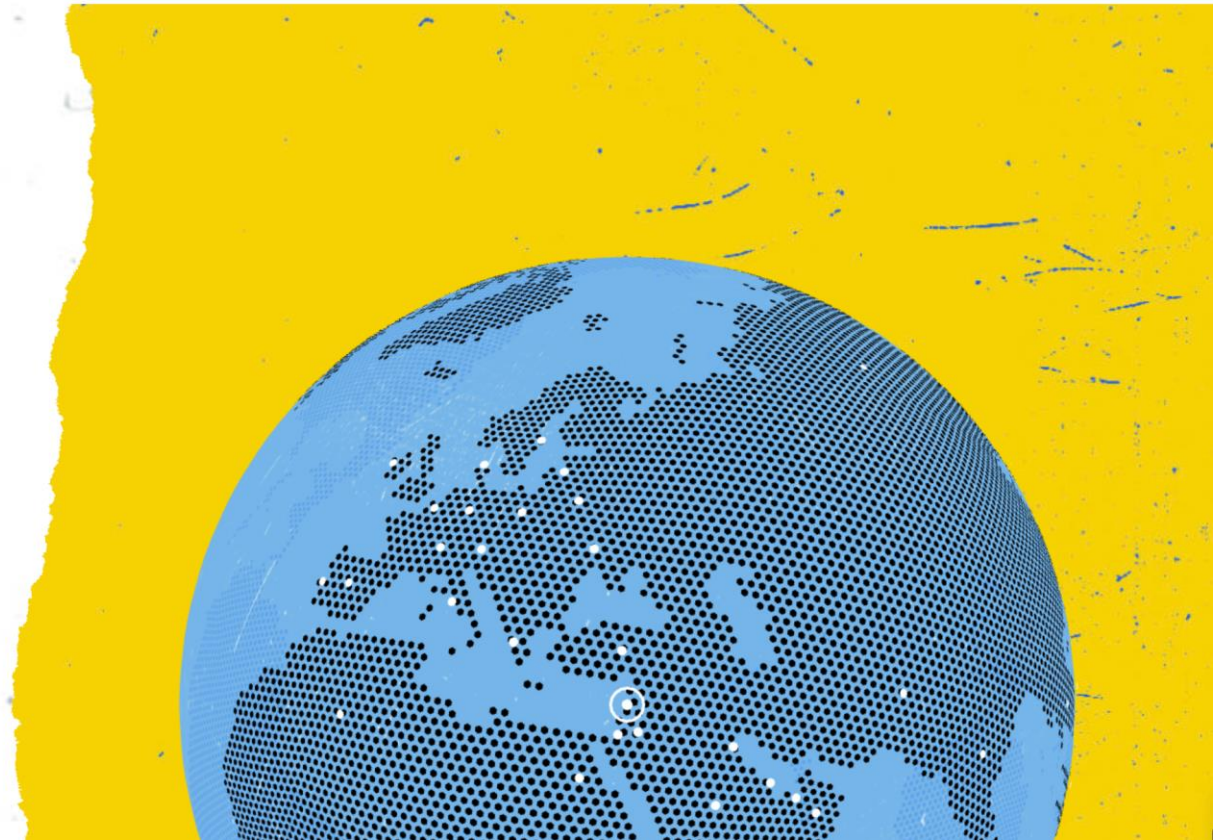
Updates ▾

Support

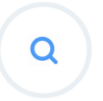
Apply



**Tap into what's  
happening to  
build what's  
next**



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



Get started with Twitter APIs and tools

# Apply for access

Apply for a developer account



Restricted used cases →

## #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.

## Which best describes you?

This is how you intend to use the Twitter developer platform



Professional



Hobbyist



Academic



## 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.



Professional



Hobbyist



Academic



Academic researcher



Teacher



Student



Something else (But related to academics)



### Academic Research application

Approved Academic Research applications gain access to higher levels of data than our standard application, but also require additional information that is not required for standard approval.

Get started

This will take longer to get approved

## #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.



Professional



Hobbyist



Academic



Academic researcher



Teacher



Student



Something else (But related to academics)

## Standard application

Submit a standard application for access to the Twitter developer platform.

2

[Get started](#)

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

 **Projects**

NO PROJECTS HERE


[+ New Project](#)





\*\*\*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 ▾

## Project 1

Overview **Settings**

### Usage

MONTHLY TWEET CAP USAGE ⓘ



Helpful docs

[How to use projects](#)

[Authentication overview](#)

[Authentication best practices](#)



# 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](#)



# Project 1

Overview Settings

## Usage

MONTHLY TWEET CAP USAGE ⓘ



0 Tweets pulled of 500,000

Resets on June 21 at 00:00 UTC

## Apps

+ Add App

### 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.



# Add an existing App or create a new App



Add an existing App

Add an existing App



Create a new App

Create new



# 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.

23






# 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 ⓘ

[Redacted API Key] 

API Secret Key ⓘ

[Redacted API Secret Key] 

Bearer Token ⓘ

AAAAAAAAAAAAAAAAAAAA [Redacted Bearer Token] 



Next, setup your App.



Go to dashboard

App settings

Personal Details

Enter **Twitter Bearer** API Key ×

API Keys

Twitter Bearer API: -

Perspective API:

1

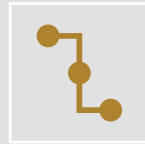
3

2

# 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>



## 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>

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing



# 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](#)

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

### Data volume

10 million Tweets / month

### Query rules

1024 characters, 1000 streaming rules

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

1. Twitter API



2. Collection



3. Metadata

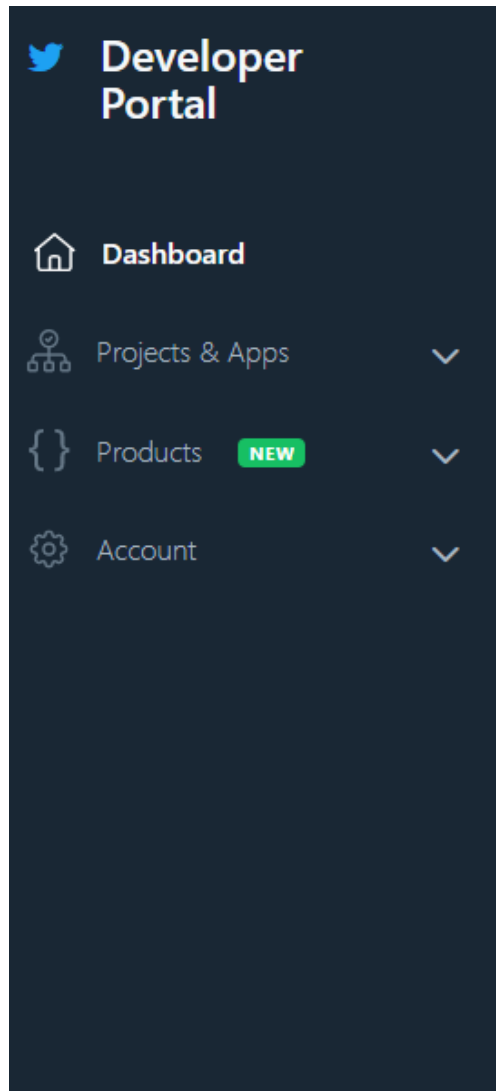


4. Export

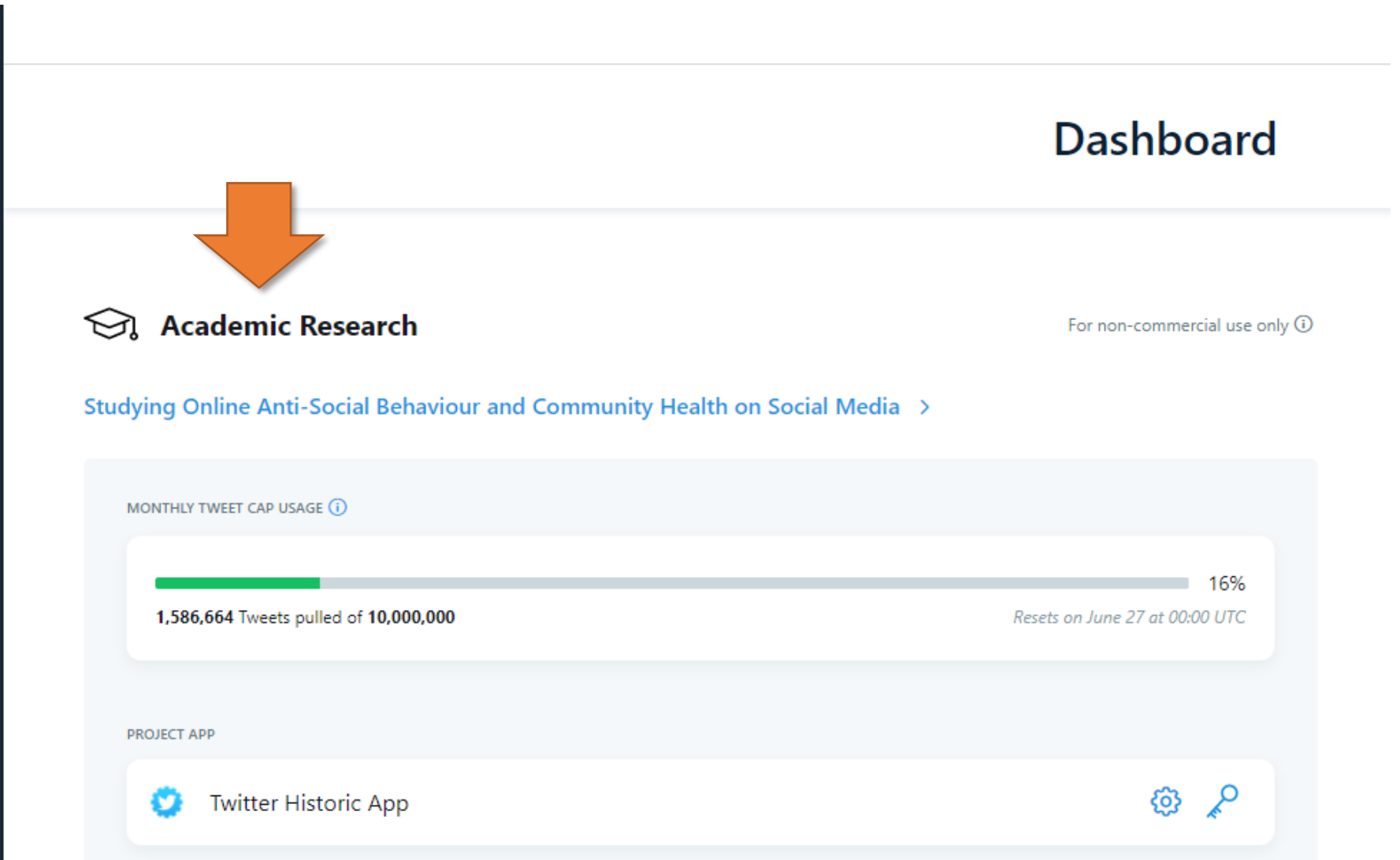


5. Sharing

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



- Developer Portal
- Dashboard
- Projects & Apps
- Products **NEW**
- Account



## Dashboard

**Academic Research** For non-commercial use only ⓘ

[Studying Online Anti-Social Behaviour and Community Health on Social Media >](#)

MONTHLY TWEET CAP USAGE ⓘ

1,586,664 Tweets pulled of 10,000,000 16% Resets on June 27 at 00:00 UTC

PROJECT APP

Twitter Historic App ⚙️ 🔑

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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

The screenshot shows the Twitter Developer Dashboard interface. On the left is a dark sidebar with the Twitter logo and 'Developer Portal' header. Below it are navigation items: 'Dashboard', 'Projects & Apps' (with a dropdown arrow), 'Overview', 'ThreadExplorer', 'App Name' (highlighted with a red arrow), and 'STANDALONE APPS'. At the bottom of the sidebar are 'Products' (with a 'NEW' badge and a dropdown arrow) and 'Account' (with a dropdown arrow). The main content area is white and shows the 'App Name' page for 'THREDEXPLORER'. The 'Keys and tokens' tab is selected, indicated by an orange arrow. Below the tabs, there are three sections: 'Consumer Keys' with a 'View Keys' button and a 'Regenerate' button; 'Authentication Tokens' with a 'Regenerate' button and a 'Revoke' button; and 'Access Token & Secret' with a 'Generate' button. The text 'API Key & Secret For @polidashboard' is visible in the first section, and 'Bearer Token Generated August 13, 2020' is visible in the second section.

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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

The screenshot shows the Communalytic PRO user interface. At the top left is the logo. The top right contains navigation links for FAQ, Tutorials, and Publications, along with a user menu. The user menu is open, showing options like My Profile, Share Requests, Share History, and Logout. A modal dialog titled "Enter Twitter Bearer API Key" is centered on the screen, with a text input field and "Submit" and "Close" buttons. In the background, the "API Keys" section is visible, with a "Twitter Bearer API" entry highlighted by an orange box and a yellow "2". The input field within the modal is also highlighted by an orange box and a yellow "3".

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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

commuanalytic (PRO)

FAQ Tutorials Publications My Datasets

## My Datasets

Server Time: June 07, 2021 13:18 UTC

Collect Data from

Reddit (HISTORICAL) Reddit (LIVE) CSV File Twitter Thread Twitter Archive (Academic Track) CrowdTangle

\* You can simultaneously run two Reddit, one Twitter and one CrowdTangle collectors.

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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

**Search Query**  
(up to 1024  
chars)



## Collect Historical Tweets

via [Twitter's Academic Research Product Track](#)

Please enter the following details to begin collection

Dataset name

Search Query (up to 1024 characters)

Start date/time (in UTC):

Start Date (yyyy-mm-dd hh:mm:ss)

End date/time (in UTC):

End Date (yyyy-mm-dd hh:mm:ss)

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# Advanced Search Rules and Operators

from:	<b>PowerTrack</b> <b>Search</b> <b>Premium</b> <b>Sandbox</b>	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 <a href="#">GET users/lookup</a> endpoint).	<code>user.id</code> , <code>user.id_str</code> (if using User ID)  <code>user.screen_name</code> (if using username)
has:geo	<b>PowerTrack</b> <b>Search</b> <b>Premium</b>	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 <a href="#">Twitter Place</a> , with the corresponding display name, geo polygon, and other fields.  Cannot be used as a standalone operator.  <b>Note:</b> operators matching on place (Tweet geo) will only include matches from original tweets. Retweets do not contain any place data.	<code>place</code> (original Tweets only)
has:hashtags	<b>PowerTrack</b> <b>Search</b> <b>Premium</b>	Matches Tweets that contain at least one hashtag.  Cannot be used as a standalone operator.	<code>entities.hashtags</code>
has:images	<b>PowerTrack</b> <b>Search</b> <b>Premium</b>	Matches Tweets that contain at least one classified image URL.  Cannot be used as a standalone operator.	<code>entities.media</code>

## 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>

1. Twitter API



2. Collection



3. Metadata

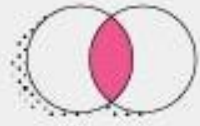


4. Export



5. Sharing

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



**AND**

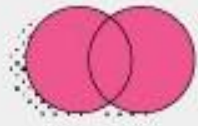
**root beer**

Tweets will contain **all** of the search terms. Spaces act as the "AND" operator.

example

"The root makes beer taste fresher"

"I don't drink root beer"



**OR**

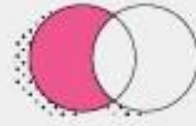
**root OR beer**

Tweets will contain **any or all** of the search terms. "OR" must be capitalized.

example

"This tree has a long root"

"I like my beer cold"



**-**

**root -beer**

Tweets will **exclude** search terms that are following "-".

example

"Root numbers are important in math"

"The root of the problem"

Dataset name

Search Query (up to 1024 characters)

Start date/time (in UTC):

Start Date (yyyy-mm-dd hh:mm:ss)

End date/time (in UTC):

End Date (yyyy-mm-dd hh:mm:ss)

1. Twitter API



2. Collection



3. Metadata



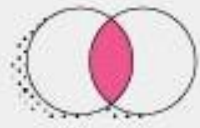
4. Export



5. Sharing



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



**AND**

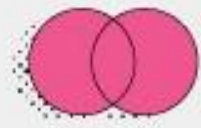
**root beer**

Tweets will contain **all** of the search terms. Spaces act as the "AND" operator.

example

"The root makes beer taste fresher"

"I don't drink root beer"



**OR**

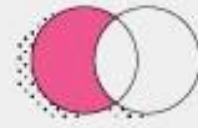
**root OR beer**

Tweets will contain **any or all** of the search terms. "OR" must be capitalized.

example

"This tree has a long root"

"I like my beer cold"



**-**

**root -beer**

Tweets will **exclude** search terms that are following "-".

example

"Root numbers are important in math"

"The root of the problem"

## 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>

Dataset name

Search Query (up to 1024 characters)

Start date/time (in UTC):

Start Date (yyyy-mm-dd hh:mm:ss) [X] [Grid]

End date/time (in UTC):

End Date (yyyy-mm-dd hh:mm:ss) [X] [Grid]

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

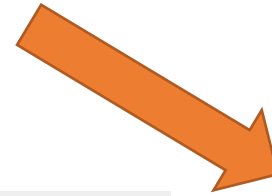
# Full-archive Search API: Common Search Operators (3)

**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**



Dataset name

Search Query (up to 1024 characters)

Start date/time (in UTC):

Start Date (yyyy-mm-dd hh:mm:ss)

End date/time (in UTC):

End Date (yyyy-mm-dd hh:mm:ss)



## QUOTATION MARKS

**"root beer"**

Tweets will contain the exact phrase entered.

example

"I am going to the store for a root beer"



## PARENTHESES

**(math) OR (root -beer)**

Tweets will contain **any or all** of the operators in parentheses. The Parentheses allow for multi-operational searches.

example

"We are learning how to find the root in math"

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# 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](https://www.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.



SUB. H. B. NO. 248  
ENACT VACCINE CHOICE AND ANTI-DISCRIMINATION ACT  
JUN 8, 2021  
HOUSE HEALTH  
THE OHIO CHANNEL

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](https://www.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](https://www.hindustantimes.com)

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

5 Retweets 1 Quote Tweet 39 Likes

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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



## MENTION

**@Canada**

Tweets will mention the entered user.

example

"So excited to visit @Canada tomorrow!"



## TO:

**to:Canada**

Tweets will be replies to the entered user.

example

@Canada Let's go Raptors!



## FROM:

**from:Canada**

Tweets will be from the entered user.

example

"We The North!"

Dataset name

Search Query (up to 1024 characters)

Start date/time (in UTC):

Start Date (yyyy-mm-dd hh:mm:ss)

End date/time (in UTC):

End Date (yyyy-mm-dd hh:mm:ss)

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# 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**

<code>is:retweet</code>	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: <code>data @twitterdev -is:retweet</code>
<code>is:reply</code>	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: <code>from:twitterdev is:reply</code>
<code>is:quote</code>	Conjunction required	Core	Returns all Quote Tweets, also known as Tweets with comments.  Example: <code>"sentiment analysis" is:quote</code>
<code>is:verified</code>	Conjunction required	Core	Deliver only Tweets whose authors are verified by Twitter.  Example: <code>#nowplaying is:verified</code>
<code>-is:nullcast</code>	Conjunction required	Advanced	Removes Tweets created for promotion only on ads.twitter.com that have a <code>"source":"Twitter for Advertisers (legacy)"</code> or <code>"source":"Twitter for Advertisers"</code> . This operator must be negated.

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

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# 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**

<code>has:mentions</code>	Conjunction required	Core	Matches Tweets that mention another Twitter user.  Example: <code>#nowplaying has:mentions</code>
<code>has:media</code>	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: <code>(kittens OR puppies) has:media</code>
<code>has:images</code>	Conjunction required	Core	Matches Tweets that contain a recognized URL to an image.  Example: <code>#meme has:images</code>
<code>has:videos</code>	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: <code>#icebucketchallenge has:videos</code>
<code>has:geo</code>	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>

1. Twitter API



2. Collection



3. Metadata

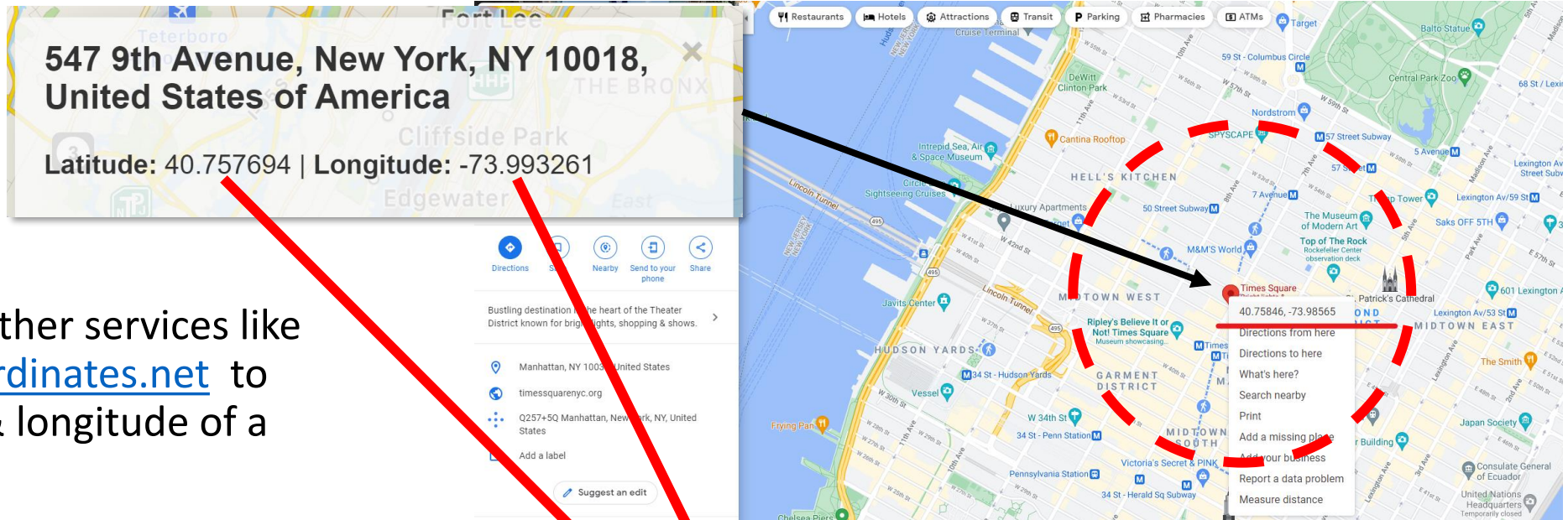


4. Export



5. Sharing

# 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]**

Latitude (lat)



Longitude (lon)

1. Twitter API



2. Collection



3. Metadata



4. Export



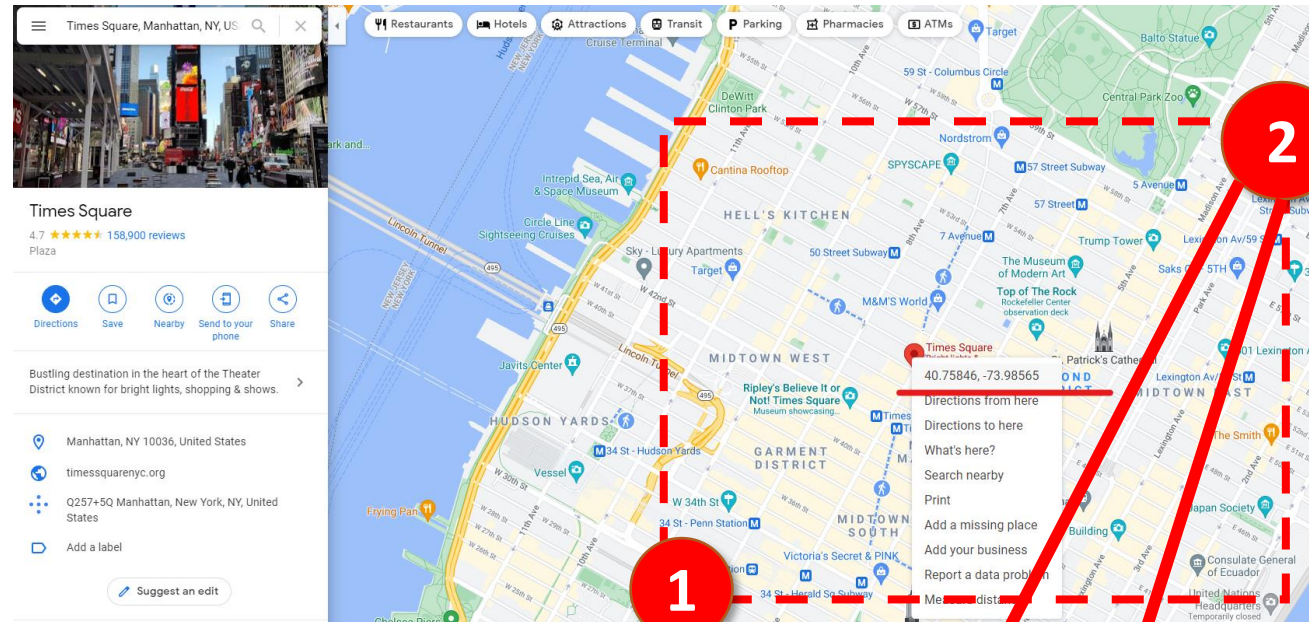
5. Sharing



# 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]**

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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

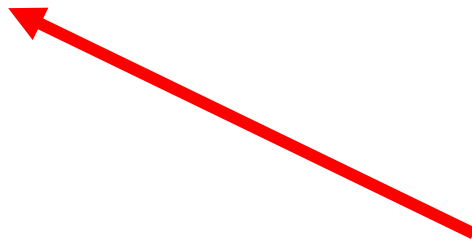
[https://en.wikipedia.org/wiki/ISO\\_3166-1\\_alpha-2](https://en.wikipedia.org/wiki/ISO_3166-1_alpha-2)

## Based on a country code

Examples:

`place_country:CA`

`place_country:GB`



### 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:<sup>[1]</sup>

- **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 p
- **ISO 3166-2:** Corresponding [ISO 3166-2](#) codes
- **Notes:** Any unofficial notes

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

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# Twitter Data & Metadata



1. Twitter API



2. Collection



3. Metadata



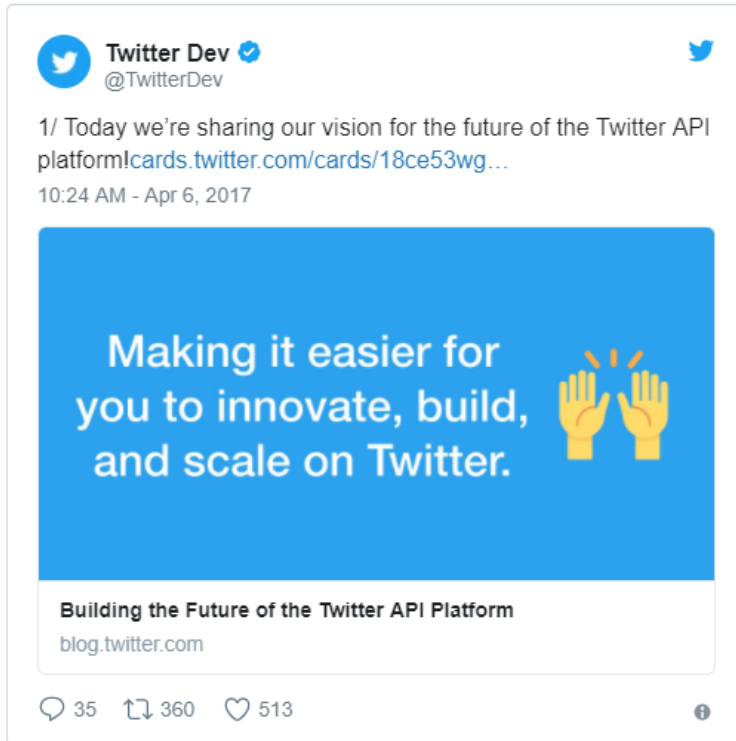
4. Export



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 plat",  
  "user": {},  
  "entities": {}  
}
```

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

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# 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": "01f8e706f872cb32",
    "name": "Washington",
    "place_type": "city",
    "url": "http://api.twitter.com/1/geo/id/01f8e706f872cb32.json"
  }
}
```

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

1. Twitter API



2. Collection



3. Metadata



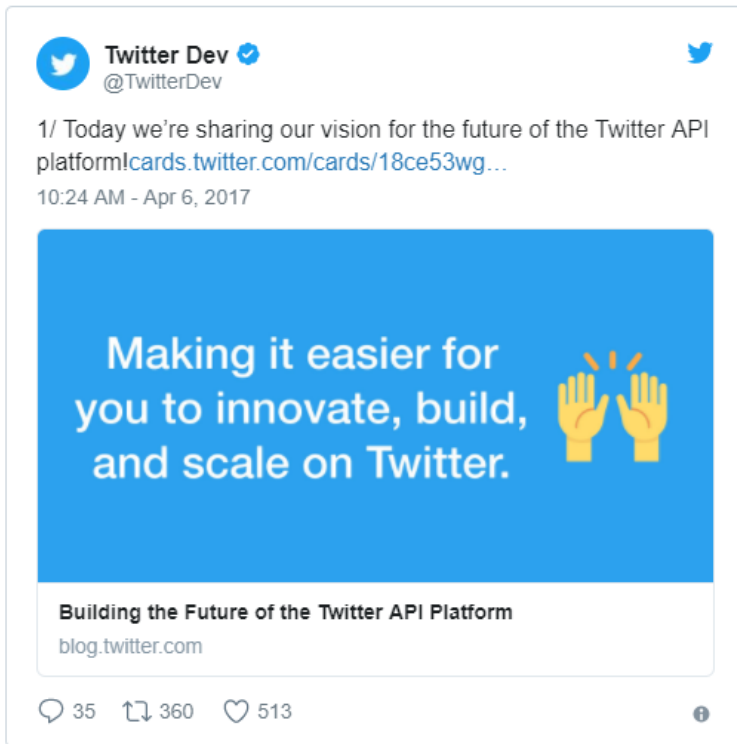
4. Export



5. Sharing

# 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, updates, and more.",
    "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/504414100000000000/504414100000000000.jpg"
  }
}
```

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# Export Twitter Dataset from Commanalytic

BREAKING: US Senate votes to acquit former

President Trum 

twitter :

<https://twitter.com/NBCPolitics/status/1360693216601849857>

1,922



FAQ

Tutorials

Publications

My Datasets



← Back to My Datasets

Overview

Dataset Name: **BREAKING: US Senate votes to acquit former President Trum**

Platform: **twitter**

Search Criteria: **<https://twitter.com/NBCPolitics/status/1360693216601849857>**

Collection started: **2021-02-13 20:59**

Collection ended: **2021-02-13 20:59**

Records: **1,922**

## Download Dataset



Download CSV File



 If this is the first time that you are working with Twitter data, please review the [Twitter attributes guide](#) prior to downloading and working with the CSV file of your dataset.

1. Twitter API



2. Collection



3. Metadata

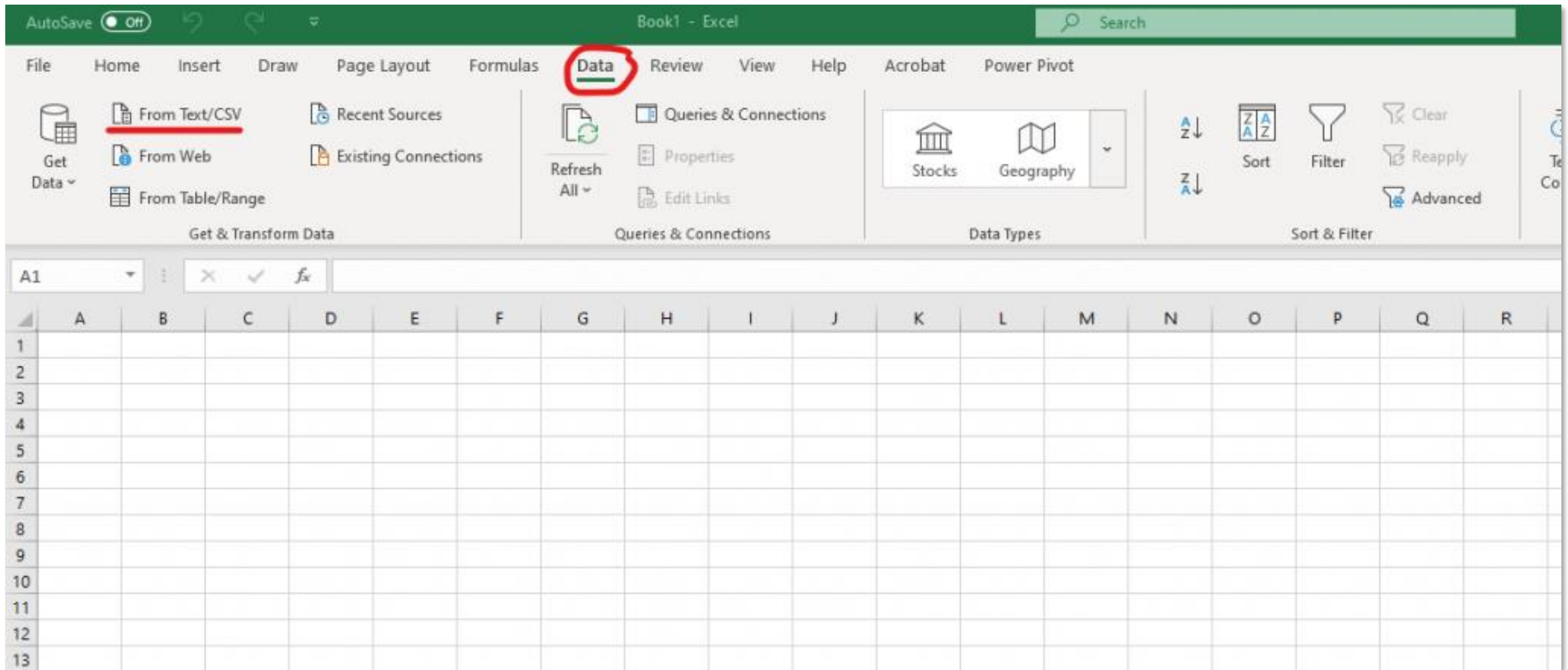


4. Export



5. Sharing

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



1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing



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

1

UTF-8

2

3

File Origin	Delimiter	Data Type Detection
1252: Western European (Windows)	Comma	Based on first 200 rows

name	username
dTrump Oh my god theyâ€™...	deidrerogers
dTrump Cohen..... COHEN	qhappiness2
dTrump It is an insult to your...	AmericanVet
dTrump #TrumpIsANational...	TechAddictz
dTrump Youâ€™ve already d...	Gatorspeak
dTrump No tax hikes for any...	agspeed226
Trump @realDonaldTrump C...	FrankAmari2
mTrump @realDonaldTrump...	richardmcvay
dTrump It looks like crime d...	DavidDickelr
dTrump https://t.co/xYyZ5g...	LSM1941986
dTrump Your boy needs ano...	j71kangaroo
dTrump #FakeNews	424_justin
rump @realDonaldTrump Th...	QuintanaIM
dTrump #Hadrawi Philosoph...	JoeBidenForL
dTrump LIAR	DAIIFLA
dTrump Joe Biden wants to r...	IFHT20
dTrump Well, we all should...	ThomasGroni
dTrump Does he? Show me t...	stephen_lani

Load Transform Data Cancel

1. Twitter API

2. Collection

3. Metadata

4. Export

5. Sharing

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

The screenshot shows the Power Query Editor interface. The 'tweet\_id' column is selected, and the 'Change Type' menu is open. The 'Text' option is highlighted. Red arrows indicate the steps: 1. Click on the 'tweet\_id' column header, 2. Click on 'Change Type', and 3. Click on 'Text'.

tweet_id	text
1	TeamTrump @realDonaldTrump Oh my god they're yelling. Co
2	TeamTrump @realDonaldTrump Cohen..... COHEN
3	TeamTrump @realDonaldTrump It is an insult to your supporte
4	TeamTrump @realDonaldTrump #TrumpIsANationalDisgrace #
5	TeamTrump @realDonaldTrump You've already destroyed the
6	TeamTrump @realDonaldTrump x hikes for anyone makir
7	DonaldTrump Curiously, I
8	realDonaldTrump The tax
9	is like crime does pay as
10	//t.co/xYyZ5gVnKh
11	boy needs another Lysol
12	News
13	realDonaldTrump That is a fac
14	realDonaldTrump mawawi Philosopher From #
15	iveThePeople make #DICTATORS, no one was born a dictator. V
16	This is RELEVANT today @realDonaldTrump https://t.co/mUxyfc

When prompted to click [Replace current]

The dialog box titled 'Change Column Type' asks: 'The selected column has an existing type conversion. Would you like to replace the existing conversion, or preserve the existing conversion and add the new conversion as a separate step?'. The 'Replace current' button is highlighted with a red arrow and underlined.

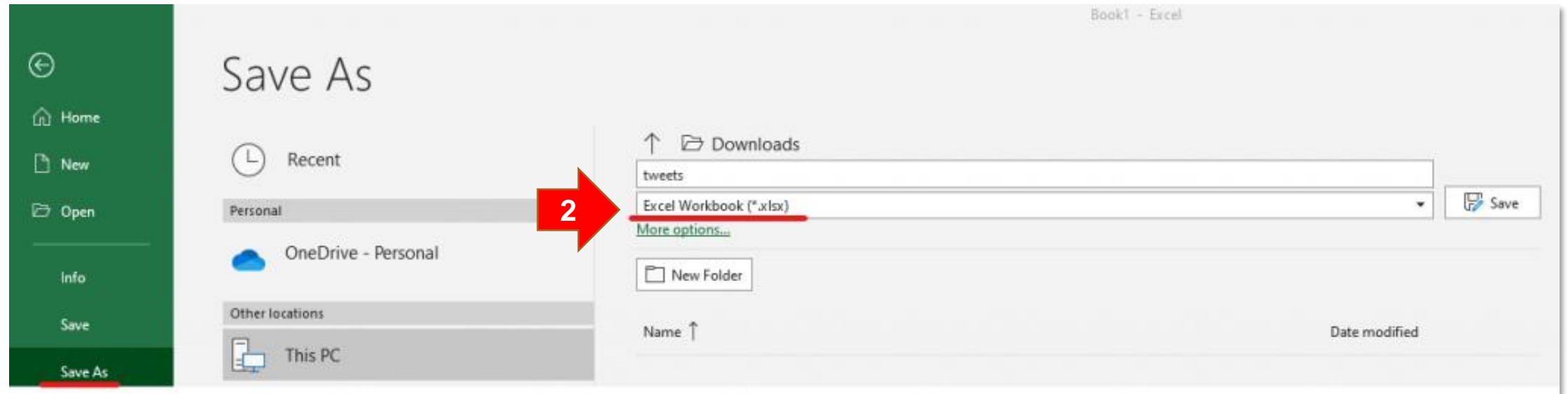


# How to open the exported CSV file in Excel (4)

The screenshot shows the 'File' menu in the Power Query Editor. Two red arrows point to the 'Close & Load' option (labeled '1') and the 'Close & Load To...' option (labeled '2'). A tooltip for 'Close & Load To...' reads: 'Save your changes to this query, close the Query Editor window, and load results to the default destination.' The background displays a table of tweet data with columns for ID, timestamp, and text.

7	1303505599678885889	9/8/2020 9:28:47 PM	@Daniel86646619 @TeamTrump @realDonaldTrump Curiously, In a rape case. Need more be said?!
8	1303505582914260994	9/8/2020 9:28:43 PM	@DonnyTheMessiah @TeamTrump @realDonaldTrump The tax
9	1303505540128149504	9/8/2020 9:28:33 PM	@TeamTrump @realDonaldTrump It looks like crime does pay a: #VoteBiden

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



1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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

	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)

	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



axinte costel aurelian Retweeted

**Retweet**



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 place for crypto ...  
[steemit.com](https://steemit.com)

5:47 PM · Mar 23, 2021 · Twitter Web App


15 Retweets 14 Quote Tweets 106 Likes

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

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	
<b>referenced_tweet_type</b>	<b>retweeted</b>   replied_to   qu
reply_settings	
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	




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](https://steemit.com)

72 29 106

**They Call Me Stevenson**   @...

Replying to @CoinMarketCap and @steemit

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

Reply

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

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	
<b>referenced_tweet_type</b>	retweeted   <b>replied_to</b>   qu
reply_settings	
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	





Referenced

quote

tweet



We'd like to work with you @justinsuntron. Freezing any accounts is just not right. May be there's miscommunicated & misinterpreted intention, we are where we are today, but the only way to move forward is through conversations.



Justin Sun @justinsuntron · Mar 7, 2020

Our intention is never to take over any blockchain but to protect the sanctity of private property! On Feb. 22, some ppl froze 65 million STEEM legally owned by Steemit. On March 3, the same ppl are trying to freeze exchanges assets. We need to stop it! medium.com/@TronFoundatio...

Show this thread

2:41 PM · Mar 7, 2020 · Twitter Web App

65 Retweets 37 Quote Tweets 254 Likes

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

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	
<b>referenced_tweet_type</b>	<b>quoted</b>
reply_settings	
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing



What's happening?

Everyone can reply

### Who can reply?

Choose who can reply to this Tweet.  
Anyone mentioned can always reply.



Everyone



People you follow



Only people you mention

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

tweet_id	date	text
197	2021-06-06T03:55:18.000Z	@JustinTrudeau Sir still I haven't received your response. Please help me to g
460	2021-06-05T19:32:02.000Z	@anku79801501 The subject is environment and not emigration. Stick to the su
463	2021-06-05T19:28:51.000Z	@Bilal45097874 What does religion has to do with the environment? Just remi
518	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	
<b>reply_settings</b>	<b>everyone   following   mentionedUsers</b>
referenced_tweet	
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

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	
<b>referenced_tweet</b>	True   False
notes	
geo_coded	
geo_country_code	
geo_country	
geo_place_type	
geo_full_name	
geo_name	

1. Twitter API



2. Collection



3. Metadata

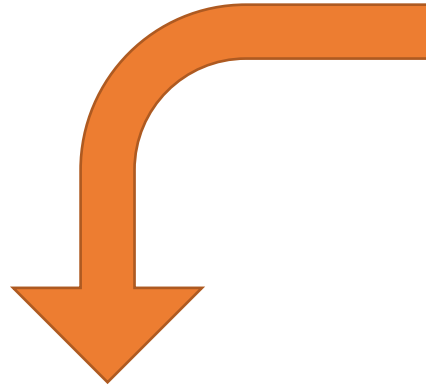


4. Export



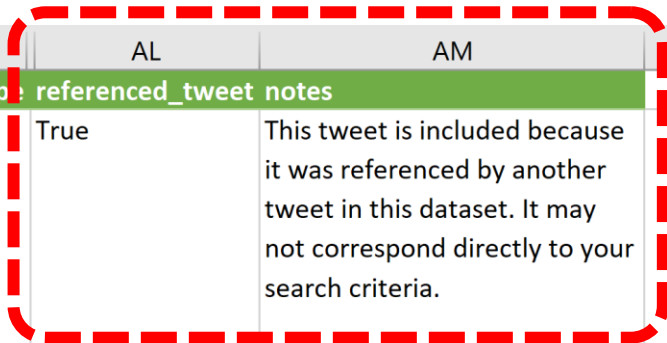
5. Sharing

# 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_local	referenced_tweet_id	referenced_tweet_type	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

The screenshot shows the Twitter Developer Terms of Service page. At the top, there is a purple navigation bar with the Twitter logo, 'Developer', 'Use cases', 'Products', 'Docs', 'More', 'Apply', a search icon, and a 'Sign In' button. Below the navigation bar, the page title is 'Developer terms' followed by the main heading 'Understand our developer policies and agreements.' The main content area is titled 'Policies and agreements' and contains a paragraph: 'Developer use of Twitter materials and content is subject to and governed by our Developer Policy and agreements.' Below this paragraph, there are six links, each with a purple underline and a right-pointing chevron: 'Developer Agreement', 'Developer Policy', 'Twitter Kit Agreement', 'Geo guidelines', 'Ads API Agreement', and 'More on restricted use cases'.

1. Twitter API



2. Collection



3. Metadata



4. Export



5. Sharing

# Data Sharing via Tweet Ids

## Collect / Export tweets as CSV

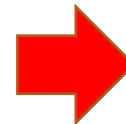
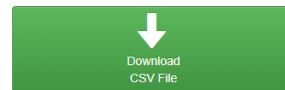
My Datasets Shared With Me

Total records in the account: 186 / 30000

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



Download Dataset



## 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 2
14	1401586531563827200	2021-06-06T17:07:24.000Z	@JustinTrudeau Shame on you and
15	1401579144132108290	2021-06-06T16:38:02.000Z	@JustinTrudeau NDP & Conser
16	1401578885427482624	2021-06-06T16:37:01.000Z	@JustinTrudeau Taxing Cdns to de
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/na4u
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 @marcomendicin
24	1401571047640555528	2021-06-06T16:05:52.000Z	@JustinTrudeau Those trees were
25	1401568121698373641	2021-06-06T15:54:14.000Z	@JustinTrudeau So Happyvvv dea

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

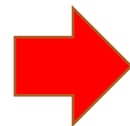
1401591680587812865

1401589438581415936

1401589398773374983

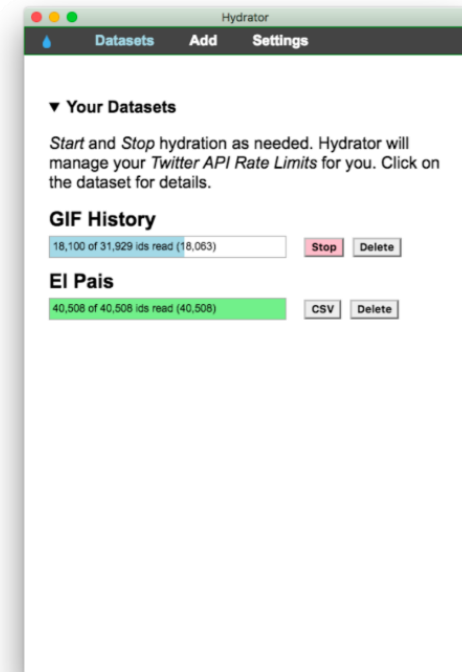
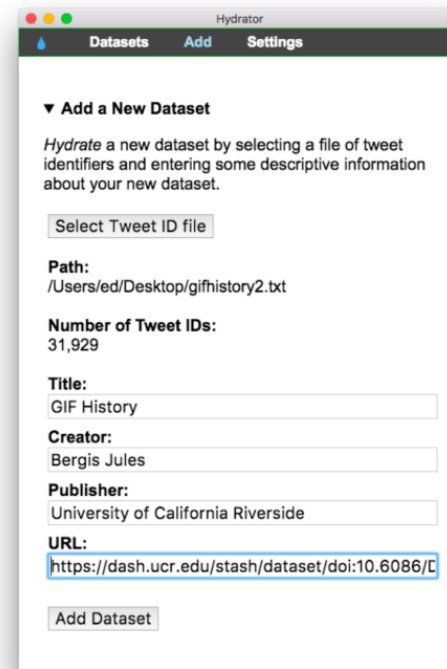
1401586720320090117

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	<a href="#">JOSA - Jopara (Guarani-dominant mixed with Spanish) Sentiment Analysis corpus.</a>	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	<a href="#">Fortune 100 response to 2020 BLM Protests</a>	6,184	Kevin McElwee	Black Lives Matter, Fortune 100	GitHub
01/03/2021	03/01/2020 to 12/31/2020	<a href="#">2020 US Presidential Election</a>	965,620,919	Elizabeth Chen, Emilio Ferrara	United States, Politics, Elections	GitHub
01/01/2021	11/18/2020 to 11/30/2020	<a href="#">Diego Maradona</a>	1,644,242	Alejandra Josiowicz, Ed Summers	Football, Argentina	Internet Archive
12/30/2020	12/10/2020 to 12/19/2020	<a href="#">SolarWinds</a>	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



**communalytic**    FAQ   TUTORIALS   PUBLICATIONS

### Dataset

On October 2nd, 2020, U.S. President Donald Trump tweeted that he and the First Lady of the United States (Melania Trump), both tested positive for COVID-19 (see Figure 1). Within seconds, his tweet received thousands of replies.

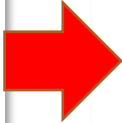
**Donald J. Trump** @realDonaldTrump

Tonight, @FLOTUS and I tested positive for COVID-19. We will begin our quarantine and recovery process immediately. We will get through this TOGETHER!

12:54 AM · Oct 2, 2020 · Twitter for iPhone

404.3K Retweets   502.2K Quote Tweets   1.8M Likes

Figure 1: Donald Trump's tweet announcing his COVID-19 diagnosis



**Scholars Portal Dataverse**    Search   User Guide   Support   English   Log In

Social Media Lab (Ryerson University)

Scholars Portal Dataverse > Ryerson University Dataverse > Social Media Lab >

## Twitter Dataset: Replies to President Trump's Tweet about contracting the Novel Coronavirus

Version 1.1

Download Options

- Original Format ZIP (34.3 MB)
- Archival Format (.tab) ZIP (35.1 MB)

Gruzd, Anatoliy, 2021, "Twitter Dataset: Replies to President Trump's Tweet about contracting the Novel Coronavirus", <https://doi.org/10.5683/SP2/BMCGI3>, Scholars Portal Dataverse, V1, UNF:6:ZoGBrMO4f8/aPycUi+bk+A== [fileUNF]

Cite Dataset    Learn about Data Citation Standards.

**Description**

On October 2nd, 2020, U.S. President Donald Trump tweeted that he and the First Lady of the United States (FLOTUS), Melania Trump, both tested positive for COVID-19 (<https://twitter.com/realDonaldTrump/status/1311892190680014849>). Within seconds, his tweet received thousands of replies. The dataset consists of 298,172 replies to Donald Trump's tweet announcing his COVID-19 diagnosis, posted on October 2nd between 6am and 12:30pm (ET). It contains tweet ids, the toxicity scores (as calculated by Google's Perspective API via <https://Communalytic.com>) and tweet availability status values (via twarc library). Following Twitter's API policy, we stripped metadata associated with each tweet. So, if you'd like to examine potential relationships between other metadata elements, you would need to recollect original tweets using tools like DocNow's Hydrator first. The only downside of this approach is that tweets that have been blocked or deleted will not be recollect. To help you get started, we also shared our Exploratory Data Analysis (EDA) Python script at <https://github.com/RUSocialMediaLab/toxicityanalysis>

**Subject**    Computer and Information Science

**Keyword**    COVID, Donald Trump, Toxicity Analysis, Twitter

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

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

