

Twitter

Table of Contents

[1 About the Platform](#)

[2 Available Data](#)

[3 Restrictions and Limitations](#)

[4 Available Metadata](#)

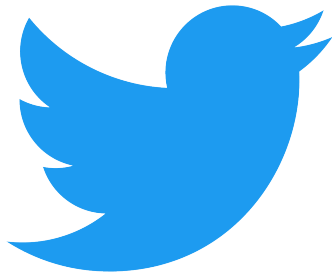
[5 Relevant Application Programming Interface \(API\)](#)

[6 Searching by Geolocation](#)

[7 Filtered Content](#)

[8 References](#)

About the Platform:



Twitter is a fast-growing social media platform that allows for users to interact with people from across the globe. It is a multiple-media platform that allows for the sharing of content whether that be posts, news, music, etc. Economic and Social Research (2021) did an overview of Twitter and its impactful functions. Twitter allows users to easily promote ideas or opinions within 240 characters where users can also provide links to blogs, news, journal articles, while reaching a large number of people quickly through tweets and retweets. Users can also follow the thoughts and actions of others, build relationships with other followers, keep up to date with the latest news and developments, and share it with others instantly, reach new audiences, seek feedback

and give feedback to others, follow and contribute to discussions on events, and engage in self-expression (2021). Users can discuss hot button topics and debate on real world issues. This allows for different forms of activism, communication, and even education to occur. People can interact with celebrities and world leaders even over a simple joke. Twitter prides itself on creating a “safe space” (Twitter, 2021) for people to express themselves and create content at their leisure.

1) What data is available from this platform?

Twitter data allows for access to tweets and associated meta-data (i.e., data of the data) as well as information about its users. Researchers have noted in previous work that there are 144 unique pieces of metadata associated with each tweet. Each tweet includes meta-data for the post as well as meta-data on the account from which it was posted. Meta-data also includes entities or associated data such as hashtags, media links, or user mentions (See: Perez, Musolesi, & Stringhini, 2018). A list of variables associated with Tweets (i.e., Tweet Data Dictionary), User (User Data Dictionary), and Entities (Entities Data Dictionary) can be found under question three. The Application Programming Interface (API) that this project is using is open to the public so that developers have access and are able to use the program free of charge. This API allows researchers to examine user, tweet, and entity-related variables.

There can be any number of variables pulled from a tweet such as geolocation, hashtag, and actual content from the tweet, such as human codification of or sentiment analysis on user’s feelings. This allows researchers to answer a variety of questions related to the COVID-19 pandemic using a multi-method approaches (e.g., content/spatial/network analyses, data mining, machine learning). Knowing what data is freely available in advance can allow programmers and researchers to plan for and make use of several variables for analyses.

2) What are the restrictions/limitations of the data?

For the public API, there is a 450 request per rate limit window. There is a tweet cap limit for 500,000 tweets per month.

3) What variables are available? What meta-data (i.e., the data of the data) is included? For example, what information is available about user profiles (e.g., username, demographic data, number of followers/friends)? What information is linked to the content/post/blog/tweet (e.g., timestamps, re-tweets, location, device/platform that content was shared from, number of likes)?

Information linked to public posts include information about the user that posted it (e.g., profile pictures, bios), the comments, the retweets including the retweets with comments, the likes, the hashtags, followers/friends, associated hyperlinks, and other

data associated with that tweet including the location and timestamp. We provide the specific variable types and definitions by Tweet, User, and Entities dictionaries below.

Tweet Data Dictionary

(See: <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/tweet>)

- `created_at`: The time that the tweet was created (Date)
- `id`: The unique identifier of the tweet (ID #)
- `text`: The text of the tweet in 280 characters or less (Content)
- `source`: Device used to post the tweet (e.g., Android vs. iPhone)
- `in_reply_to_user_id`: ID# of respondent's account
- `lang`: Twitter includes information on language source for 47 types of languages (for examples see this section of their site via the link above).
- `attachments`: Different types of files linked with the tweet (video/news/media)
- `author_id`: Author account of the tweet (User1234)
- `context_annotations`: Contents of the tweet (text)
- `conversation_id`: ID of original tweet to follow conversations (retweets with reply)
- `entities`: Different trait identifiers for the tweet such as hashtags, letter count, word count, etc.
- `geo`: Determines if a tweet is close in location (geolocation tag, Memphis, TN)
- `non_public_metrics`: Looks at nonphysical impacts from the tweet (Who viewed the tweet)
- `organic_metrics`: Original tweet impact (Likes of original)
- `possibly_sensitive`: Only surfaces when the tweet involves a link, allows for Twitter to circulate with similar topics (Media Link)
- `promoted_metrics`: Looks at nonphysical impacts from tweet when promoted (News)
- `referenced_tweets`: Used to follow conversations of retweets (retweets and reply)
- `reply_settings`: Settings to allow replies to a tweet
- `withheld`: Holds back items that might cause copyright infringement

User Data Dictionary

(See: <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/user>)

- `id`: The unique identifier of the account holder (e.g., ID #)
- `name`: Name of the user (e.g., John Smith)
- `username`: Username handle (e.g., JaneDoe1234)
- `created_at`: Time stamp the tweet was posted at (e.g., 2013-1-14-01:13)
- `location`: The user defined location of the account (e.g., Memphis, TN)
- `description`: Bio of the account holder
- `entities`: Provides additional information like hashtags, links media, etc.
- `pinned_tweet_id`: The unique identifier of the 'pinned' tweet on the profile

- protected: User makes account private to the public. Only people following that account can see activity
- profile_image_url_https: Profile picture used by account
- public_metrics: Can be used to determine Twitter user's reach/influence on the application
- verified: User account is verified by Twitter (e.g., variable is most applicable to celebrities/famous persons, etc.)
- url: A URL provided on the Twitter user's bio
- withheld: Content that hasn't been consented to release

Media Object Data Dictionary

(See: <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/media>)

- media_key: Identifier of media content involved in the post (e.g., news link/music link)
- type: Category of Media attached (e.g., GIF, Photo, Link, etc.)
- duration_ms: Only available when type is video, length of video (e.g., Music Video - 2:33)
- non_public_metrics: Identifies public engagement (e.g., How many users watched the video)
- organic_metrics: Determines how much engagement is happening with the user posting
- preview_image_url: Preview of news/media attachments (e.g., A link to an article)
- promoted_metrics: Determines media engagement when the post is advertised (e.g., How many users view the tweet once it is cycled/trending)
- public_metrics: Determines the total number of users who viewed the video (e.g., 1,000 users watched the video)
- width: Contents in pixel (e.g., 750 pixels)

Poll Object Data Dictionary

(See: <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/poll>)

- id: Identifier of the poll (e.g., 12345678)
- options: Different options for the poll (e.g., Chocolate or Vanilla)
- duration_minutes: How long the poll lasts (e.g., 160 minutes)
- end_datetime: Date the poll ends (e.g., 1/2/2021)
- voting_status: Indicates if the voting is still happening or if the poll has ended

Place Objects Data Dictionary

(See: <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/place>)

- full_name: Name of place the tweet was posted at (e.g., Memphis, TN)
- id: Unique identifier of the place in the tweet (e.g., 12345678)
- contained_with: Other places associated with the location of the tweet (e.g., Nashville, TN)
- country: The country the tweet was posted in (e.g., United States)
- country_code: Number code applied to the country (e.g., 800001)
- geo: Geolocation/coordinates the tweet was posted at (e.g., -74.0000000)
- name: Short nickname for the location the tweet was posted at (e.g., Bluff City)
- place_type: Type of place the tweet was posted at (e.g., city, mountains, country, national park, etc.)

4) Is there an Application Programming Interface (API) available for this platform? If yes, please provide the link for the platform.

Yes. The official API link can be found here: <https://developer.twitter.com/en/docs/api-reference-index>

If you are interested in access to the full firehose, you can read about it here: <https://brightplanet.com/2013/06/25/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care/>

5) Does your assigned API provide a way to limit the search based on geolocation?

Yes, tweets can be collected if the user has enabled geotagging.

6) Does the platform use any kind of filters to remove the content on their sites?

Yes. Twitter's filtering methods include:

- Filtering content by different rules and restrictions such as duplication of posts, reporting of posts/accounts, and whether the tweet has sensitive language
- The tweet must follow Twitter guidelines for safety, or it could be issued a content filter and removed for not following guidelines.

Twitter recently created a COVID-19 streamline that filters a large number of tweets. Twitter became very active because of the pandemic so the developers made a way to group different tweets and allow the release of information through a filtering system.

References

- Perez, B., Musolesi, M., & Stringhini, G. (2018). You are your metadata: Identification and obfuscation of social media users using metadata information. *arXiv preprint arXiv:1803.10133*.
- Twitter (2020a). About Search Rules and Restrictions. (2020). Retrieved from <https://help.twitter.com/en/rules-and-policies/twitter-search-policies>
- Twitter (2021b). Data Dictionary. Retrieved from <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/tweet>
- Twitter (2021c). Data Dictionary. Retrieved from <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/user>
- Twitter (2021d). Data Dictionary. Retrieved from <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/media>
- Twitter (2021e). Data Dictionary. Retrieved from <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/poll>
- Twitter (2021f). Data Dictionary. Retrieved from <https://developer.twitter.com/en/docs/twitter-api/data-dictionary/object-model/place>
- Twitter (2021g). GET /labs/1/tweets/stream/covid19. (2020). Retrieved from <https://developer.twitter.com/en/docs/labs/covid19-stream/api-reference/get-tweets-stream-covid19>
- Twitter (2020h). Tweet Location FAQs. Retrieved from <https://help.twitter.com/en/safety-and-security/tweet-location-settings>
- Twitter (2021i). About Twitter. Retrieved from <https://about.twitter.com/>
- “What is Twitter and Why should you use it?”. (2021). *Economic and Social Research*. Retrieved from: <https://esrc.ukri.org/research/impact-toolkit/social-media/twitter/what-is-twitter/>