

How to Query Twitter Data

In this project, you'll be using [Tweepy](#) to query Twitter's API for additional data beyond the data included in the WeRateDogs Twitter archive. This additional data will include retweet count and favorite count.

Some APIs are completely open, like MediaWiki (accessed via the [wptools](#) library) in Lesson 2. Others require authentication. The Twitter API is one that requires users to be authorized to use it. This means that before you can run your API querying code, you need to set up your own Twitter application. Here are the steps to do that on the Twitter site:

- First, if you do not already have one, you need to [sign up for a Twitter account](#).
- Next, to set up a developer account, follow the directions on [Twitter's Developer Portal, in the "How to Apply" section](#).
- You will be guided through the steps, and asked to describe in your own words what you are building. Here is some suggested language you can use: "As a Udacity student, I need to access the Twitter API in order to complete a Data Wrangling student project. In this project, I'll be using Tweepy to query Twitter's API for data included in the WeRateDogs Twitter archive. This data will include retweet count and favorite count. Before I can run my API querying code, I need to set up my own Twitter application. Once I have this set up, I will develop some code to create an API object that I'll use to gather Twitter data. After querying each tweet ID, I will write its JSON data to a tweet_json.txt file with each tweet's JSON data on its own line. I will then read this file, line by line, to create a pandas DataFrame that I will assess and clean. I may post this completed project on my GitHub account, where it will get viewers. Otherwise there will be no other

readers or users of my Twitter data or project analysis beyond the Udacity instructors and reviewers."

- Once you submit your application, you should soon receive an email from Twitter letting you know they have approved your new Twitter developer account. Follow the link in the email from Twitter to a page of directions to get started creating your app.
- If you are asked for an app name, it can be anything appropriate, and if you're asked for a Website URL, it can be anything in a standard URL format. You can do the same with other requested URLs, or perhaps leave them blank.
- If you're asked to explain how your app will be used, you could say something like "I'm creating this for a student Data Wrangling project with Udacity, where we need to query and analyze Twitter data from WeRateDogs."
- You should then be given a Success message, and a new developer page displayed to you where you can manage your app.
- You can then go to the Keys and Tokens tab on this page to find or generate the Consumer API keys, and the Access Token and Access Token Secret that you will need.

Note: If you have any trouble creating this Twitter account or accessing the data, please see the section at the bottom of this page "Accessing Project Data Without a Twitter Account."

Once you have your Twitter account and Twitter app set up, the following code, which is provided in the [Getting started](#) portion of the Tweepy documentation, will create an API object that you can use to gather Twitter data.

```
import tweepy
```

```
consumer_key = 'YOUR CONSUMER KEY'
consumer_secret = 'YOUR CONSUMER SECRET'
access_token = 'YOUR ACCESS TOKEN'
access_secret = 'YOUR ACCESS SECRET'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_secret)

api = tweepy.API(auth)
```

Tweet data is stored in JSON format by Twitter. Getting tweet JSON data via tweet ID using Tweepy is described well in this [StackOverflow answer](#). Note that setting the `tweet_mode` parameter to `'extended'` in the `get_status` call, i.e., `api.get_status(tweet_id, tweet_mode='extended')`, can be useful.

Also, note that the tweets corresponding to a few tweet IDs in the archive may have been deleted. [Try-except blocks](#) may come in handy here.

Do Not Include Your API Keys, Secrets, and Tokens in Your Submission

Do **not** include your API keys, secrets, and tokens in your project submission. This is standard practice for APIs and public code.

Twitter's Rate Limit

Twitter's API has a rate limit. Rate limiting is used to control the rate of traffic sent or received by a server. As per [Twitter's rate limiting info page](#):

Rate limits are divided into 15 minute intervals

To query all of the tweet IDs in the WeRateDogs Twitter archive, 20-30 minutes of running time can be expected. Printing out each tweet ID after it was queried and [using a code timer](#) were both helpful for sanity reasons. Setting the `wait_on_rate_limit` and `wait_on_rate_limit_notify` parameters to `True` in the `tweepy.api` [class](#) is useful as well.

Writing and Reading Twitter JSON

After querying each tweet ID, you will write its JSON data to the required `tweet_json.txt` file with each tweet's JSON data on its own line. You will then read this file, line by line, to create a pandas DataFrame that you will soon assess and clean. This [Reading and Writing JSON to a File in Python](#) article from Stack Abuse, will be useful.

Accessing Project Data Without a Twitter Account

If you can't set up a Twitter developer account using the steps above, or you prefer not to create a Twitter account for some reason, you may instead follow the directions below to access the data necessary for the project. **Note:** We recommend that you follow the steps above to access the data using a Twitter developer account, because with the shortcut detailed below, you will miss practicing the valuable skill of gathering this data on your own. However, Twitter's updated process may not work for everyone, and we realize there are legitimate reasons that some students may prefer this approach, so we provide it to you here. **You choose the approach to access the data that works best for you. This shortcut approach will certainly work for you to pass the project equally well.**

Directions for accessing the Twitter data without actually creating a Twitter account:

At the bottom of this page you can find two files you can download:

- `twitter_api.py`: This is the Twitter API code to gather some of the required data for the project. Read the code and comments, understand how the code works, then copy and paste it into your notebook.
- `tweet_json.txt`: This is the resulting data from `twitter_api.py`. You can proceed with the following part of "Gathering Data for this Project" on the Project Details page: "Then read this `tweet_json.txt` file line by line into a pandas DataFrame with (at minimum) tweet ID, retweet count, and favorite count."