Business Analytics Programming

Lab 2 (Twitter API & TextBlob)

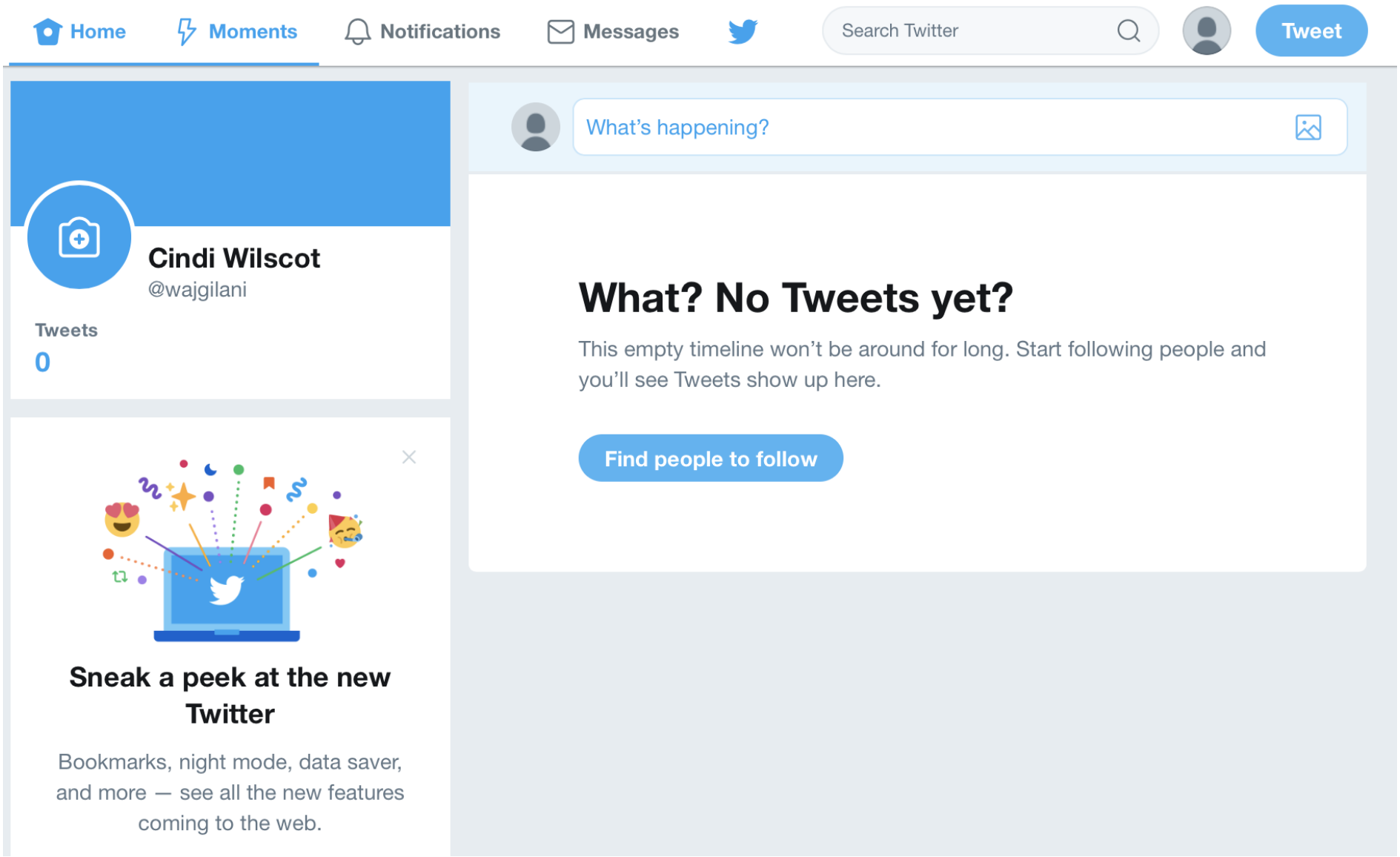
Dr. Wajahat Gilani

Rutgers Business School

September 10, 2019

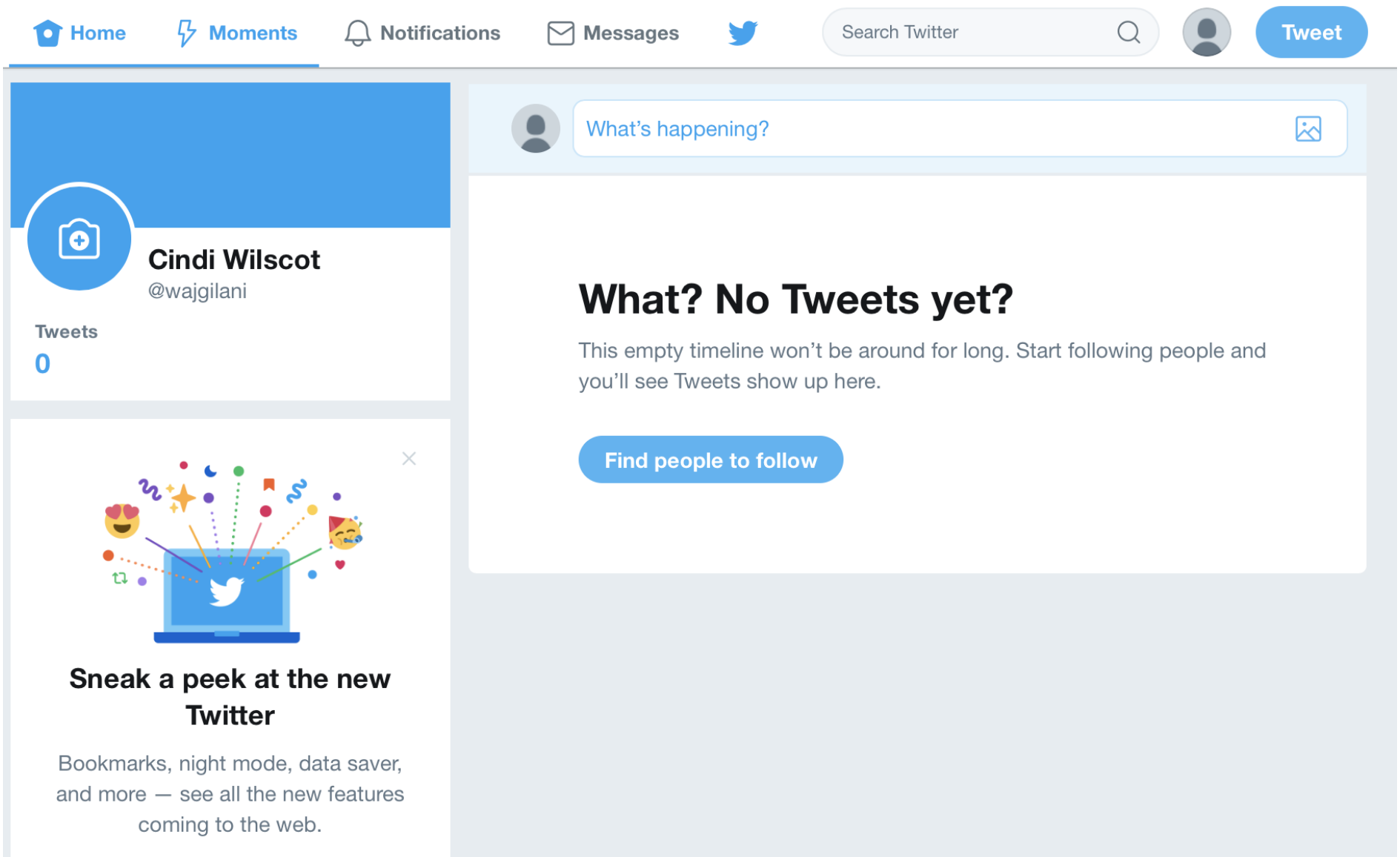
Create Twitter API

Step 1: Log into twitter or create an account



Create A Twitter APP

Step 2: Go to https://developer.twitter.com/en/apps



Create A Developer Account

Step 3: Click on Continue

STATUS: IN PROGRESS

- User profile
- Account details
- Use case details
- Terms of service
- Email verification

Interested in a developer account?

Some of our premium APIs are currently in Beta. By applying, you agree to receive emails from our team requesting feedback on your experience.

Select a user profile to associate

By default, this @username will be the admin of this developer account. If you are creating a developer account on behalf of your organization, you may wish to use your organization's @username as it is most likely to own the Apps you will use to access the API endpoints or warrant special permissions. You'll be able to invite teammates and re-assign roles later within your developer account settings.

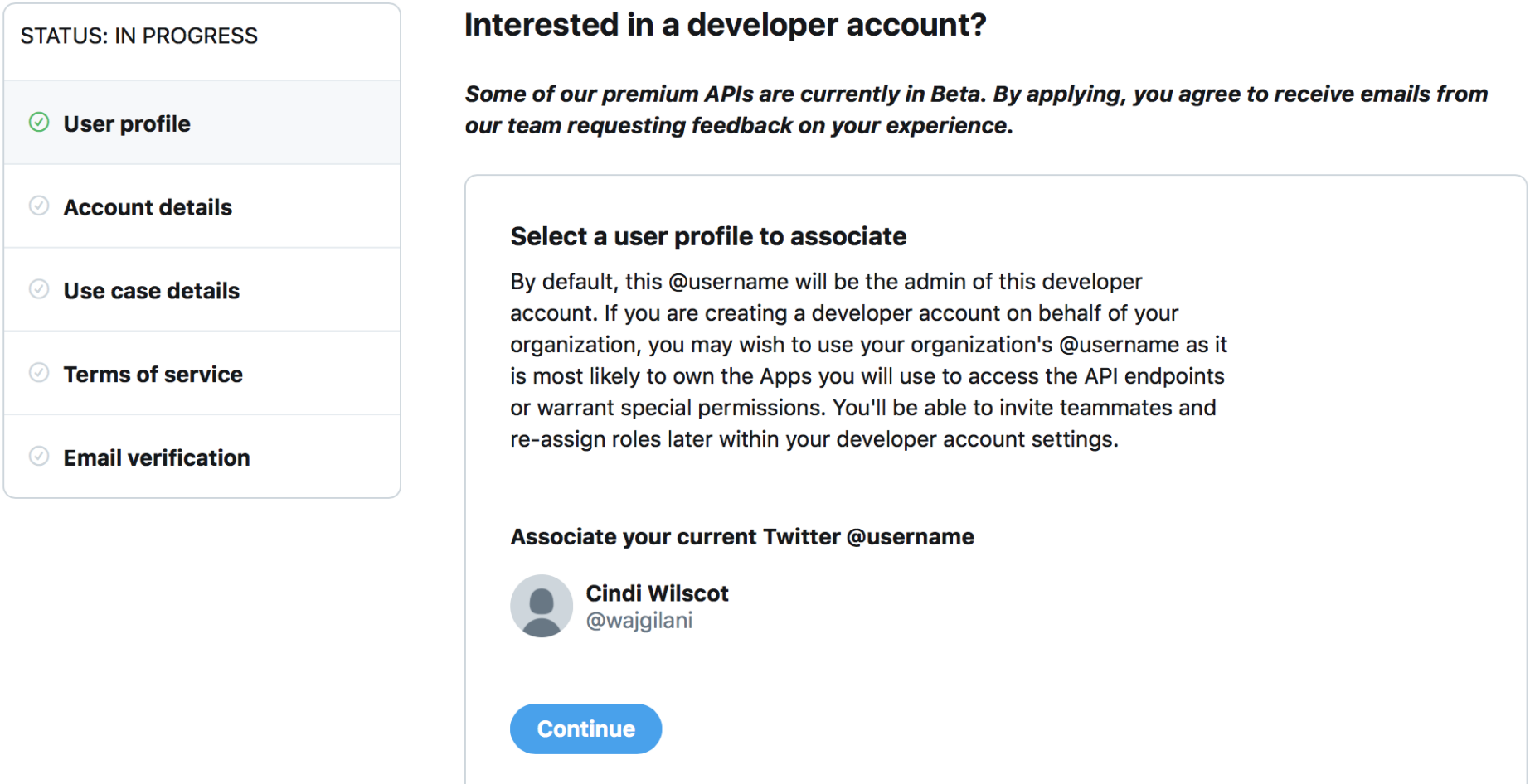
Associate your current Twitter @username

Cindi Wilscot
@wajgilani

Continue

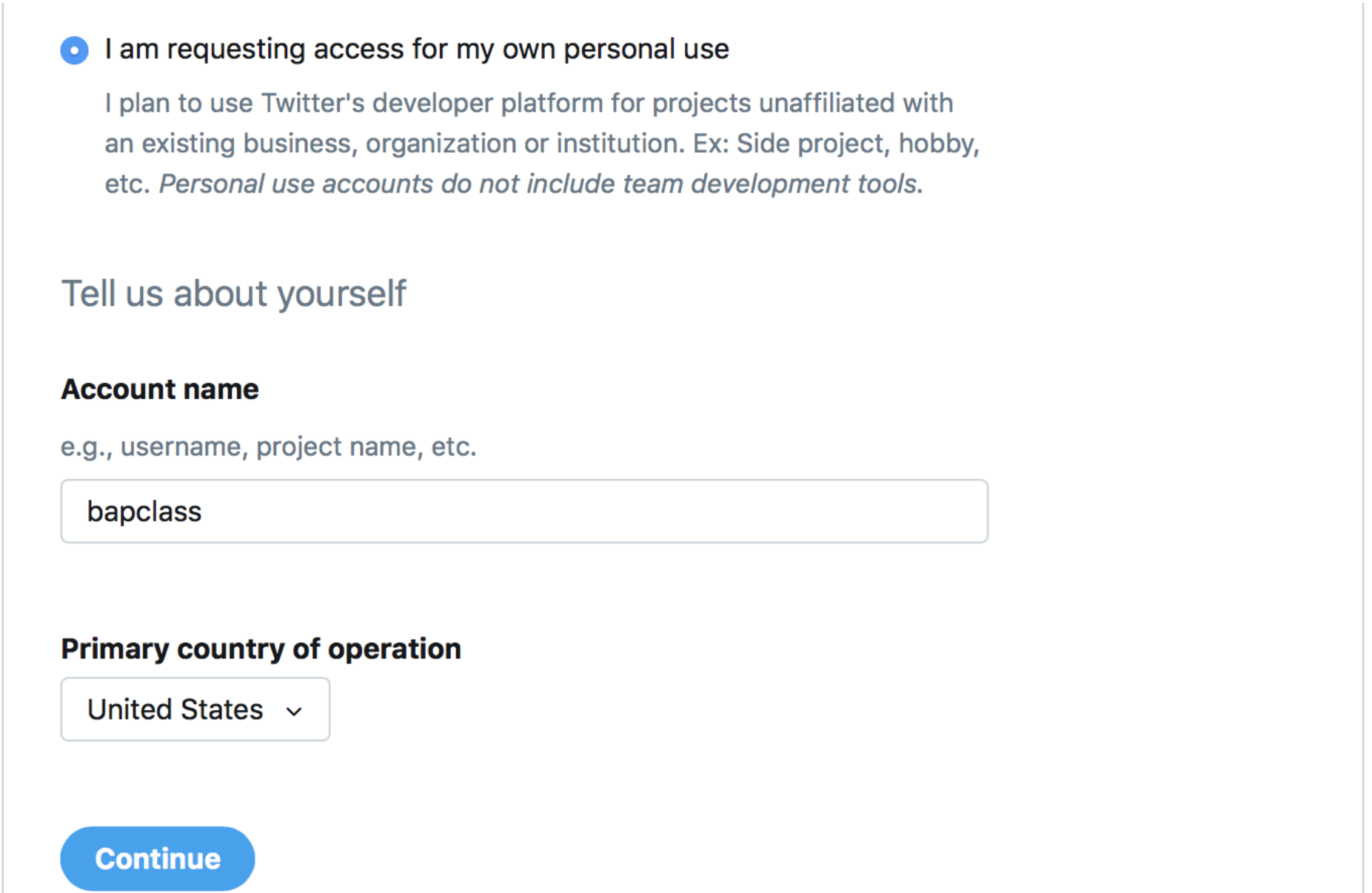
Choose Account Details

Step 4: Click on "I am requesting access for my own personal use"



Choose Account Name

Step 5: Choose an Account Name and Country



Choose Project Description

Step 6: Choose an Account Name and Country

STATUS: IN PROGRESS

- User profile
- Account details
- Use case details
- Terms of service
- Email verification

Why the questions?

We empower freedom of expression by providing a platform that protects the voices of our users — both on Twitter, and via our developer products. To help verify that all uses of Twitter data comply with our policies, we require additional information from developers signing up to use this service. Providing thorough answers will help us understand your use cases and will help expedite the evaluation of your application. Learn more about our restricted use cases.

Tell us about your project

What use case(s) are you interested in?

Select all that apply

- [x] Academic research
- [ ] Advertising
- [ ] Audience analysis
- [ ] Chatbots and automation
- [ ] Consumer / end-user experience
- [ ] Engagement and customer service
- [ ] Publish and curate Tweets
- [ ] Student project / Learning to code
- [ ] Topic analysis
- [ ] Trend and event detection
- [ ] Other

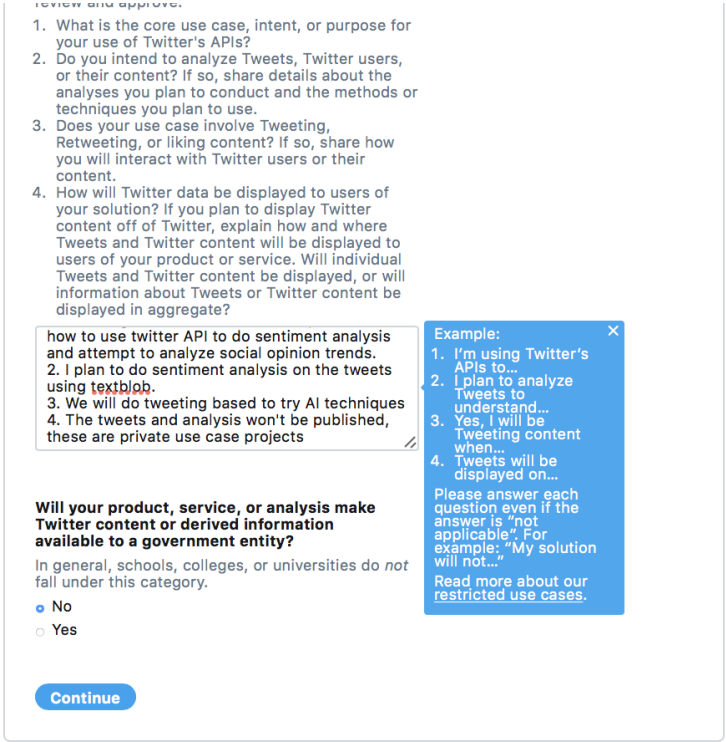
Describe in your own words what you are building

Please describe what you would like to build with Twitter's APIs. Be sure to give detailed answers to the following questions. If the question does not apply to your solution, please explicitly state that. The more detailed the response, the easier it is to review and approve.

1. What is the core use case, intent, or purpose for your use of Twitter's APIs?
2. Do you intend to analyze Tweets, Twitter users, or their content? If so, share details about the analyses you plan to conduct and the methods or techniques you plan to use.

Write a Project Description

Step 7: Fill in Text Box (Copy and Paste, or write your own Description)



1. I'm using Twitter's API to teach my MBA class how to use twitter API to do sentiment analysis and attempt to analyze social opinion trends.
2. I plan to do sentiment analysis on the tweets using textblob.
3. We will do tweeting based to try AI techniques
4. The tweets and analysis won't be published, these are private use case projects

Confirm Email

Step 8: Confirm Email

One more step...

To complete your application, please check your inbox to confirm your email address.

Didn't receive your verification email? Resend it now.

Get Started

Step 9: Create App

Get started

Create an app

To use an API, we require you create an app as part of our OAuth authorization scheme. Visit the Apps page of this developer portal to create one. Then, return to this page to complete the next step.

Set up a dev environment

To begin using the new Premium APIs, you need to set up one or more dev environments for the endpoint — and connect it to an app. Dev environments can be used to isolate usage, rules, rate limits, and more. If you are planning to use our standard APIs, you can skip this step.

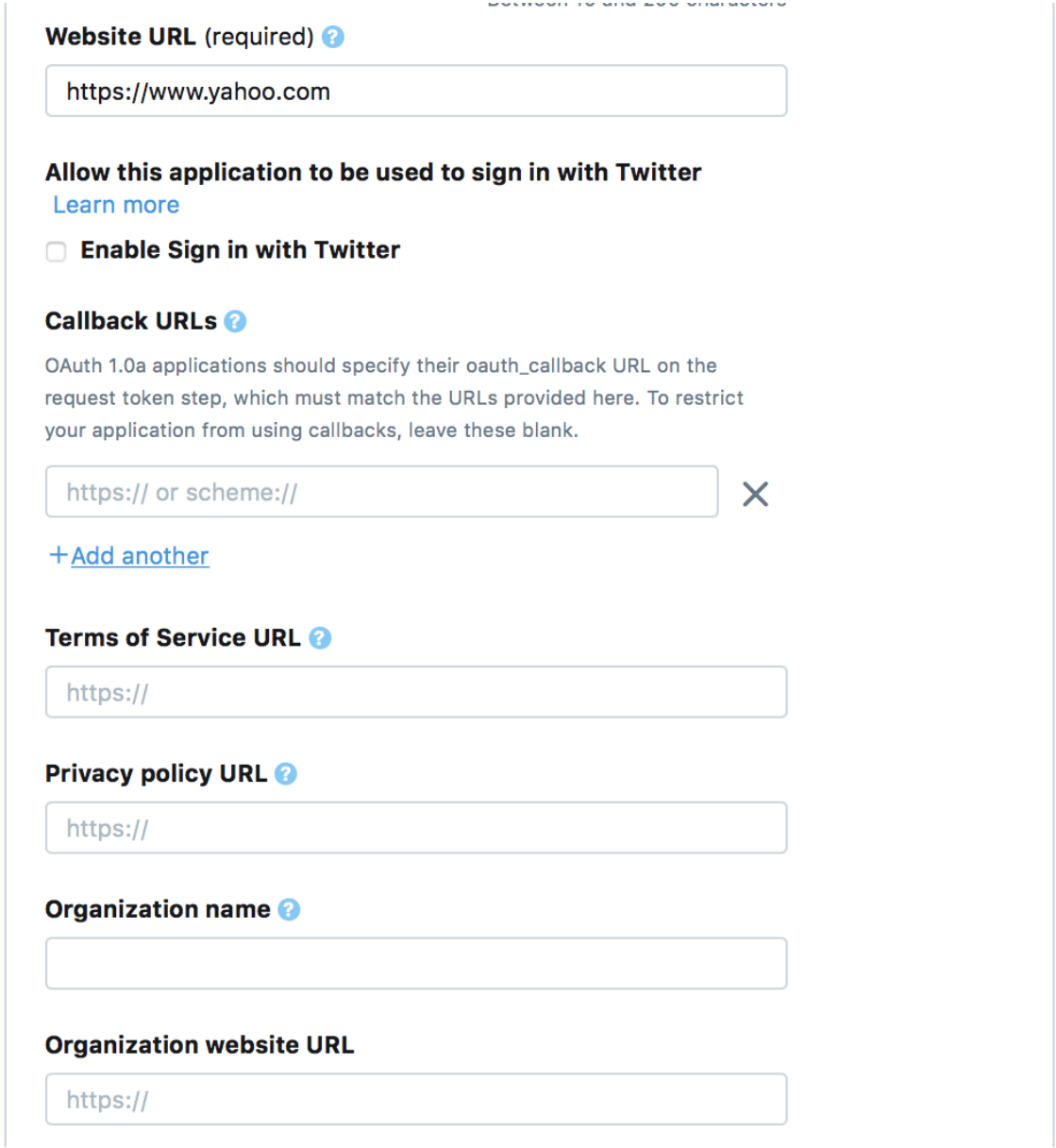
Start using the endpoints!

Once you've set up your account, accessing the endpoint is super simple. Check out our documentation and API reference for additional details about each endpoint.

Manage / upgrade your access level

All new developer accounts are provisioned with a default, sandbox level of access to our premium APIs. Explore increased access or enhanced functionality available via these endpoints.

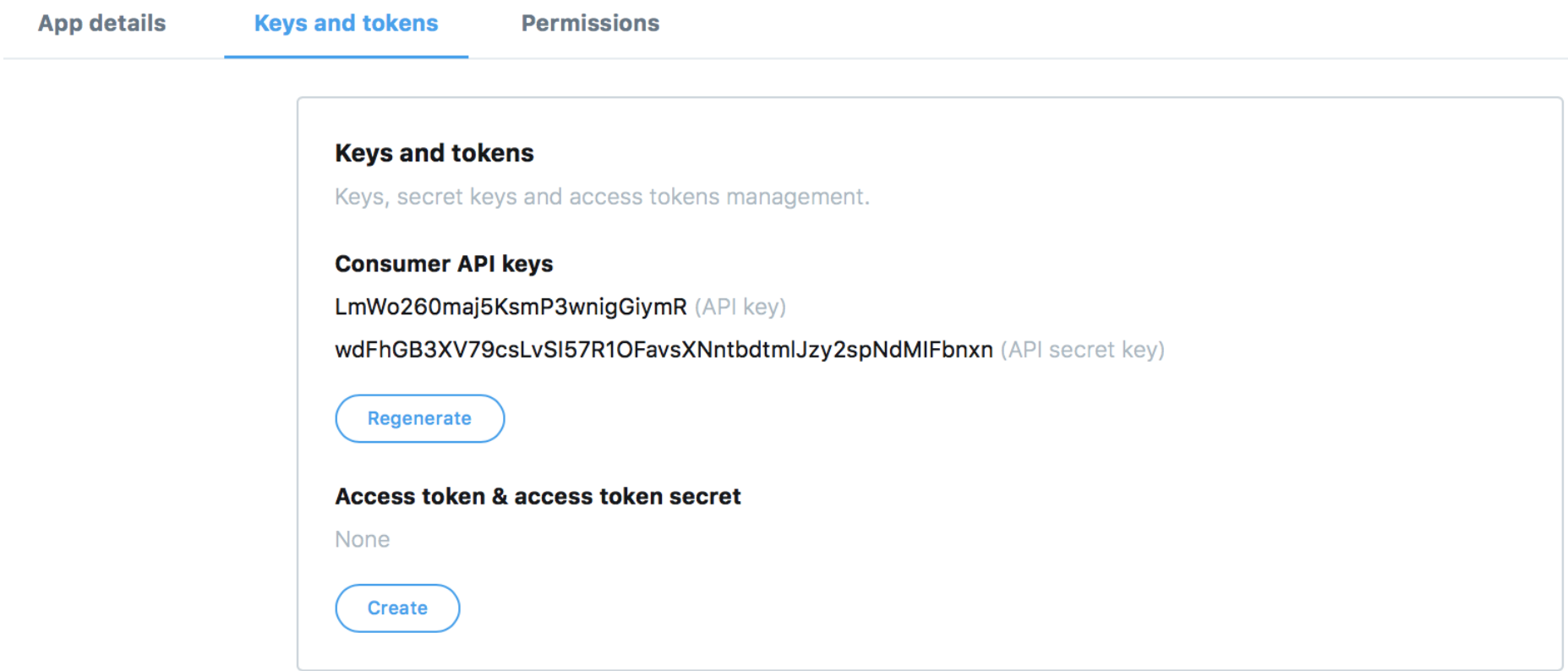
Get Started



Step 10: Create App

Get Started

Step 11: Copy the API key and API secret key



Get Started

Step 12: Copy the API key and API secret key

App details | Keys and tokens | Permissions

Keys and tokens

Keys, secret keys and access tokens management.

Consumer API keys

LmWo260maj5KsmP3wnigGiymR (API key)

wdFhGB3XV79csLvSI57R1OFavsXNntbdtmlJzy2spNdMIFbnxn (API secret key)

Regenerate

Access token & access token secret

4012083173-pdPffs50tApeBURWR9QQt22rlhEp0sdEaCFwBvR (Access token)

PVwr62zsdUF0QQpcRMkPDBLxJ4HhAG4Cjccy49GwPv8pK (Access token secret)

Read and write (Access level)

Revoke Regenerate

```
1 apikey='LmWo260maj5KsmP3wnigGiymR'
2 apisecretkey='
      wdFhGB3XV79csLvSI57R1OFavsXNntbdtmlJzy2spNdMIFbnxn'
3 accesstoken='4012083173-
      pdPffs50tApeBURWR9QQt22rlhEp0sdEaCFwBvR'
4 accesstokensecret='
      PVwr62zsdUF0QQpcRMkPDBLxJ4HhAG4Cjccy49GwPv8pK'
```

Setting Up Your Application Programming Interface (API)

The information on websites are powered by databases. Companies allow us access to their databases through the use of API's.

```
import numpy as np
import pandas as pd

import pip
```

Pip allows you to download python packages (libraries of code written by another company) on your computer. In RStudio we would simply use the packages GUI, but we don't have an equivalent in python. You now have to download the *twitter* library.

```
!pip install twitter
```

You can execute the pip command in your console or script code, but just execute it once.

Twitter Libraries

We need to reference the libraries from the *twitter* package that we plan to use:

```
1 from twitter import Twitter
2 from twitter import OAuth
```

- Notice the from keyword before the import keyword. The reason is that we want to be able to use the *Twitter* and *OAuth* function calls without having to use a variable pointer (i.e. np or pd) or having to call the package before the functions, i.e. *twitter.Twitter*.
- *OAuth* takes our 4 keys and access codes and authenticates them.
- *Twitter* uses the *OAuth* object to establish a connection to the twitter database.

```
1 oauth = OAuth(accesstoken,accesstokensecret,apikey,
      apisecretkey)
2 api = Twitter(auth=oauth)
```

The variable *api* now is going to allow us to call functions that get data from twitter.

Many Languages With Different API

If you go to the URL:
https://developer.twitter.com/en/docs/developer-utilities/twitter-libraries.html

Python

- python-twitter *maintained by @bear* — this library provides a pure Python interface for the Twitter API (documentation)
- tweepy *maintained by @applepie & more* — a Python wrapper for the Twitter API (documentation) (examples)
- TweetPony *by @Mezgrman* — A Python library aimed at simplicity and flexibility.
- Python Twitter Tools *by @sixohsix* — An extensive Python library for interfacing to the Twitter REST and streaming APIs (v1.0 and v1.1). Also features a command line Twitter client. Supports Python 2.6, 2.7, and 3.3+. (documentation)
- twitter-gobject *by @tchx84* — Allows you to access Twitter's 1.1 REST API via a set of GObject based objects for easy integration with your GLib2 based code. (examples)
- TwitterSearch *by @crw_koepp* — Python-based interface to the 1.1 Search API.
- twython *by @ryanmcgrath* — Actively maintained, pure Python wrapper for the Twitter API. Supports both normal and streaming Twitter APIs. Supports all v1.1 endpoints, including dynamic functions so users can make use of endpoints not yet in the library. (docs)
- TwitterAPI *by @boxnumber03* — A REST and Streaming API wrapper that supports python 2.x and python 3.x, TwitterAPI also includes iterators for both API's that are useful for processing streaming results as well as paged results.
- Birdy *by @sect2k* — "a super awesome Twitter API client for Python"

We Are Using the Native Twitter API

Type and run: (picture of console)

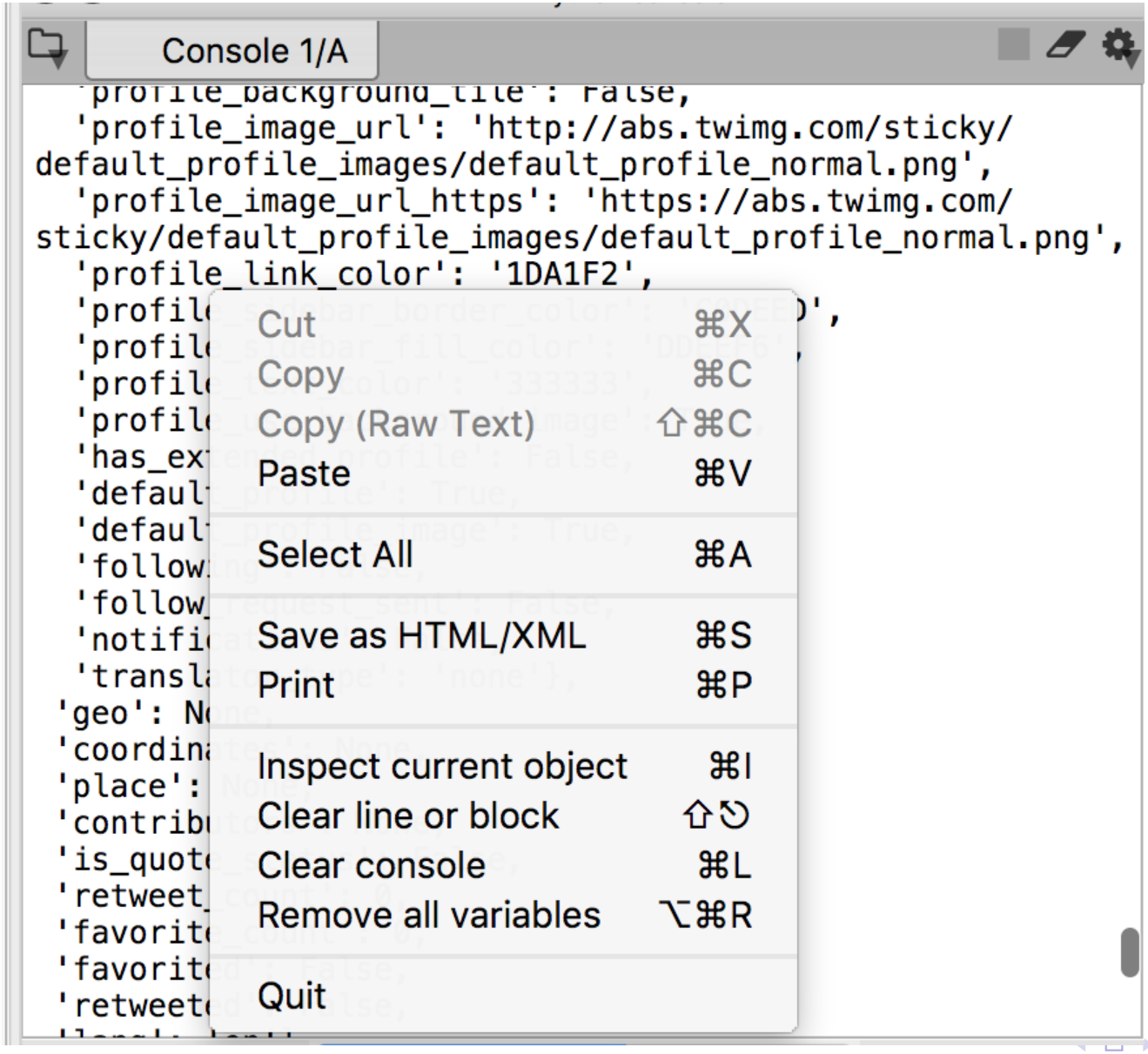
```
1 help(api)
```

```
In [21]: help(api)
Help on Twitter in module twitter.api object:

class Twitter(TwitterCall)
 |  Twitter(format='json', domain='api.twitter.com', secure=True,
auth=None, api_version=<class 'twitter.api._DEFAULT'>, retry=False)
 |
 |  The minimalist yet fully featured Twitter API class.
 |
 |  Get RESTful data by accessing members of this class. The result
 |  is decoded python objects (lists and dicts).
 |
 |  The Twitter API is documented at:
 |
 |    https://dev.twitter.com/overview/documentation
 |
 |  The list of most accessible functions is listed at:
 |
 |    https://dev.twitter.com/rest/public
 |
 |
 |  Examples::
 |
 |      from twitter import *
```

Save Console Output as HTML File

Double Click Mouse to Save Output as an HTML File: (picture of console)



What Topic is Trending Around World - Variable Explorer

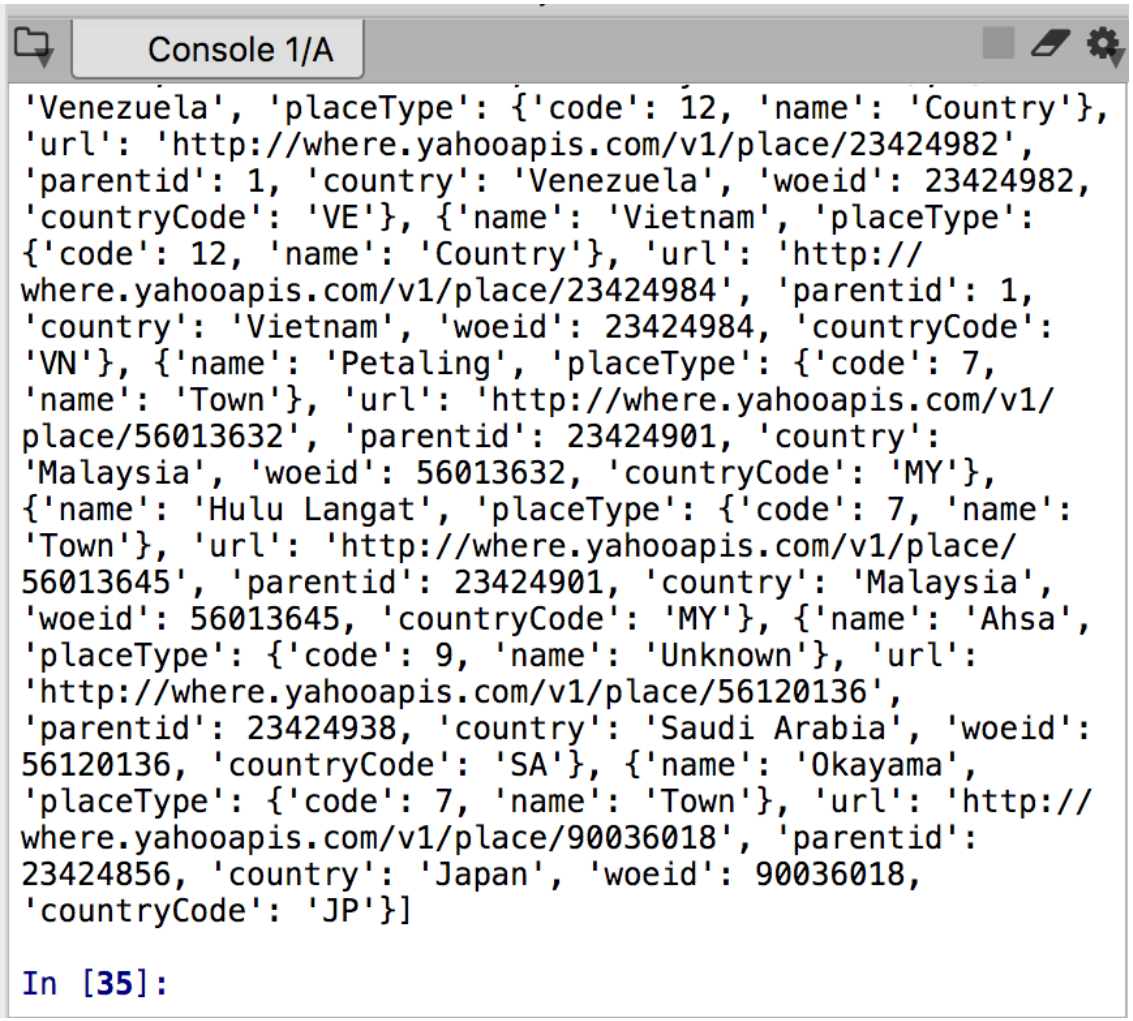
```
1 t_loc = api.trends.available()
```

ıde ▲	Type	Size	Value
0	dict	7	{'name':'Worldwide', 'placeType':{'code':19, 'name':'Supername'}, 'url ...
1	dict	7	{'name':'Winnipeg', 'placeType':{'code':7, 'name':'Town'}, 'url':'http ...
2	dict	7	{'name':'Ottawa', 'placeType':{'code':7, 'name':'Town'}, 'url':'http:/ ...
3	dict	7	{'name':'Quebec', 'placeType':{'code':7, 'name':'Town'}, 'url':'http:/ ...
4	dict	7	{'name':'Montreal', 'placeType':{'code':7, 'name':'Town'}, 'url':'http ...
5	dict	7	{'name':'Toronto', 'placeType':{'code':7, 'name':'Town'}, 'url':'http: ...
6	dict	7	{'name':'Edmonton', 'placeType':{'code':7, 'name':'Town'}, 'url':'http ...
7	dict	7	{'name':'Calgary', 'placeType':{'code':7, 'name':'Town'}, 'url':'http: ...
8	dict	7	{'name':'Vancouver', 'placeType':{'code':7, 'name':'Town'}, 'url':'htt ...
9	dict	7	{'name':'Birmingham', 'placeType':{'code':7, 'name':'Town'}, 'url':'ht ...
10	dict	7	{'name':'Blackpool', 'placeType':{'code':7, 'name':'Town'}, 'url':'htt ...
11	dict	7	{'name':'Bournemouth', 'placeType':{'code':7, 'name':'Town'}, 'url':'h ...
12	dict	7	{'name':'Brighton', 'placeType':{'code':7, 'name':'Town'}, 'url':'http ...
13	dict	7	{'name':'Bristol', 'placeType':{'code':7, 'name':'Town'}, 'url':'http: ...
14	dict	7	{'name':'Cardiff', 'placeType':{'code':7, 'name':'Town'}, 'url':'http: ...
15	dict	7	{'name':'Coventry', 'placeType':{'code':7, 'name':'Town'}, 'url':'http ...

Close

What Topic is Trending Around World - Console

```
1 t_loc = api.trends.available()
2 print(t_loc)
```



The results are passed to us as a JSON Object.

JavaScript Object Notation (JSON)

JSON (JavaScript Object Notation) is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate.

JSON is built on two structures:

- A collection of name/value pairs. In various languages, this is realized as an object, record, struct, dictionary, hash table, keyed list, or associative array.
- An ordered list of values. In most languages, this is realized as an array, vector, list, or sequence.

Normalize JSON Object

Pandas has a function that normalizes JSON objects, into tables. First, import the *json_normalize* function, remember run it once by itself:

```
from pandas.io.json import json_normalize
```

Now simply call the function, with any JSON object in the parameter:

```
df_loc=json_normalize(t_loc)
```

You can now view the table *df_loc*

Normalize JSON Object - Global Trends

Index	country	countryCode	name	parentid	placeType.code	placeType.name	url	woeid
362	United States	US	Long Beach	23424977	7	Town	http://where.yahoo…	2441472
363	United States	US	Los Angeles	23424977	7	Town	http://where.yahoo…	2442047
364	United States	US	Louisville	23424977	7	Town	http://where.yahoo…	2442327
365	United States	US	Memphis	23424977	7	Town	http://where.yahoo…	2449323
366	United States	US	Mesa	23424977	7	Town	http://where.yahoo…	2449808
367	United States	US	Miami	23424977	7	Town	http://where.yahoo…	2450022
368	United States	US	Milwaukee	23424977	7	Town	http://where.yahoo…	2451822
369	United States	US	Minneapolis	23424977	7	Town	http://where.yahoo…	2452078
370	United States	US	Nashville	23424977	7	Town	http://where.yahoo…	2457170
371	United States	US	New Haven	23424977	7	Town	http://where.yahoo…	2458410
372	United States	US	New Orleans	23424977	7	Town	http://where.yahoo…	2458833
373	United States	US	New York	23424977	7	Town	http://where.yahoo…	2459115
374	United States	US	Norfolk	23424977	7	Town	http://where.yahoo…	2460389
375	United States	US	Oklahoma City	23424977	7	Town	http://where.yahoo…	2464592
376	United States	US	Omaha	23424977	7	Town	http://where.yahoo…	2465512
377	United States	US	Orlando	23424977	7	Town	http://where.yahoo…	2466256
378	United States	US	Philadelphia	23424977	7	Town	http://where.yahoo…	2471217

Normalize JSON Object - Global Trends (Continued)

Which country has the most cities being tracked?

```
print(df_loc.country.value_counts())
```

What if we wanted to search for a particular city, using a partial word?

```
dfNew=df_loc[df_loc['name'].str.contains('New')]
dfNew[['name','woeid']]
```

	name	woeid
25	Newcastle	30079
371	New Haven	2458410
372	New Orleans	2458833
373	New York	2459115
440	New Zealand	23424916

We want to use the woeid to see what's currently trending in any one city. Get the value for New York and put it in a variable ny.

```
ny=dfNew.loc[dfNew.name=='New York','woeid']
```

New York Trends

```
ny_trend = api.trends.place(_id=ny)
```

This will give an error:

"ValueError: The truth value of a Series is ambiguous. Use a.empty, a.bool(), a.item(), a.any() or a.all()."

The reason is because the parameter *_id* requires a number, but *ny* is a Series.

```
type(ny)
```

The result is "pandas.core.series.Series". How do we get the value from a Series?

```
ny.values
```

array([2459115])

Will an array work? NO! The function requires a single number. So how do we get it?

New York Trends (Continued)

If *ny.values* is an array of a single number, then we simply need to choose the first element of *ny.values*.

```
ny.values[0]
```

Output:
2459115

```
ny_trend = api.trends.place(_id=ny.values[0])
```

ny_trend is a JSON object

```
dfny=json_normalize(ny_trend)
```

Index	as_of	created_at	locations	trends
0	2019-02-14T…	2019-02-14T…	[{'name': 'New York',…	[{'name': '#TheMasked…

The values in the columns "locations" and "trends" are also JSON objects. The "trends" column has the terms trending in New York.

New York Trends Table (Continued)

Create a table *dftrends* from the JSON object. How do we select the "trends" column?

```
dfny.trends
type(dfny.trends)
```

What type of data container is *dfny.trends*?
What is the size of *dfny.trends.values*? (Its an array)

```
dfny.trends.shape
```

Output:
(1,)

Its an array of size 1, therefore to process the JSON object in the trends column of dfny:

```
dftrends=json_normalize(dfny.trends.values[0])
```

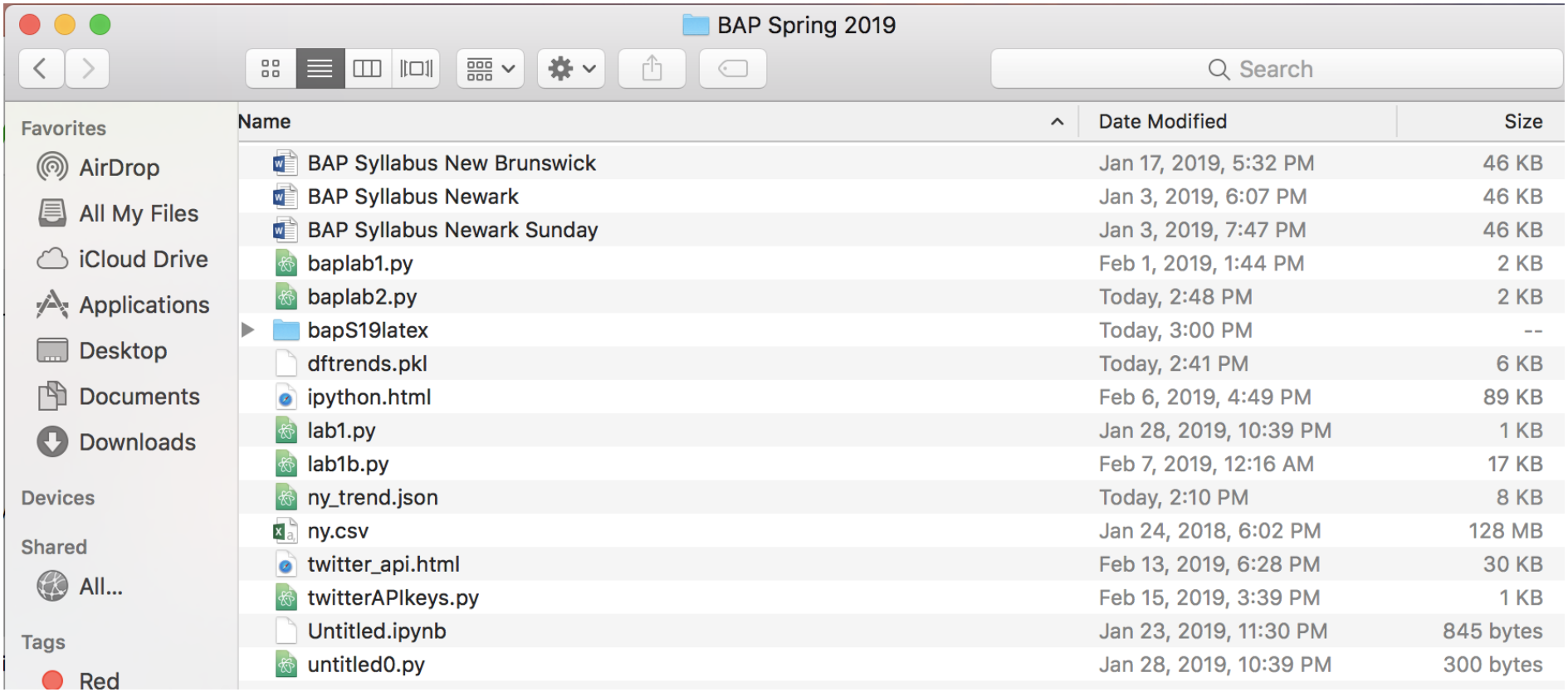
New York Trends Table - dftrends

Index	name	promoted_content	query	tweet_volume	url
0	#TheMaskedSinger	None	%23TheMaske…	25999	http://twitter.com…
1	#RHONJ	None	%23RHONJ	nan	http://twitter.com…
2	#CBBUS2	None	%23CBBUS2	48132	http://twitter.com…
3	Ben Simmons	None	%22Ben+Simmons%22	30001	http://twitter.com…
4	Valentine's Day	None	%22Valentine%27s+Day%22	1.10758e+06	http://twitter.com…
5	Duke	None	Duke	58283	http://twitter.com…
6	Joe Flacco	None	%22Joe+Flacco%22	65709	http://twitter.com…
7	So Far Gone	None	%22So+Far+Gone%22	88877	http://twitter.com…
8	Tamar	None	Tamar	33591	http://twitter.com…
9	#CelebrityBigBrother	None	%23Celebrit…	nan	http://twitter.com…
10	YNW Melly	None	%22YNW+Melly%22	51589	http://twitter.com…
11	Ricky	None	Ricky	41184	http://twitter.com…
12	#TheChallenge33	None	%23TheChall…	nan	http://twitter.com…
13	Ryan Adams	None	%22Ryan+Adams%22	28206	http://twitter.com…
14	Big Brother	None	%22Big+Brother%22	18655	http://twitter.com…
15	Tetris 99	None	%22Tetris+99%22	16366	http://twitter.com…
16	Link's Awakening	None	%22Link%27s+Awakening%	94625	http://twitter.com…

Pandas - Saving/Reading a Pandas Table

Saving a pandas dataframe to your directory:

```
1 dftrends.to_pickle('dftrends.pkl')
```




```
1 dftrends = pd.read_pickle('dftrends.pkl')
```

Twitter - Post and Read a Tweet

```
1 api.statuses.update(status="Their is an invasion at the
      border, someone get Jon Snow!!!")
2
3 mytweets=api.statuses.home_timeline()
4
5 dfmyt=json_normalize(mytweets)
6
7 mytweets1=api.statuses.home_timeline(count=1)
8 dfmyt1=json_normalize(mytweets1)
```

Line 1: Posts tweets onto your twitter page
Line 3: Downloads your tweets into a data container *mytweets* (it will be a JSON object)
Line 5: Convert JSON object into pandas table *dfmyt*
Lines 7&8: Download a certain number of the tweets from your twitter, according to the latest tweets.

Index	retweet_count	retweeted	source	text	truncated	.contributors_ena	user.created_at	ıser.default_profil	.default_
0	0	False	<a href="https…	Their is an invasion at the …	False	False	Wed Oct 21 20:52:50 +0…	True	True

Twitter - Highest Tweet Volume

Find the 5 topics with the highest tweet volumes (New York).

```
dftrends.columns
dftrends.nlargest(5,'tweet_volume')[['name','tweet_volume']]
```

	name	tweet_volume
2	Jordyn	979162.0
7	Bernie	696298.0
1	Kylie	563129.0
3	Khloe	308276.0
5	Tristan Thompson	182668.0

```
search_result = api.search.tweets(q='Trump',count = 100,
    tweet_mode='extended')
dfsr=json_normalize(search_result)
```

Index	_metadata.comple	arch_metadata.cou	rch_metadata.max	h_metadata.max_	ı_metadata.next_r	arch_metadata.que	h_metadata.refres	ch_metadata.since	ı_metadata.sin
0	0.081	100	10983669698…	10983669698…	?max_id=1098…	Bernie	?since_id=10…	0	0

```
dfst=json_normalize(dfsr.statuses.values[0])
```

Twitter - Content Length

The table *dfst* is extremely large. It shows all the information a single tweet holds. The actual text message is in the column *full_text*, but an easier way to understand the type of columns in the table is to just view a single row.

```
1 df0=pd.DataFrame({'Value':dfst.loc[0]})
```

Index	Value
entities.media	nan
entities.symbols	[]
entities.urls	[]
entities.user_mentions	[{'screen_name': 'welovetherays', 'name': 'TICK TOCK MOFO', 'id': 284530669, 'id '284530669', 'indices': [0, 14]}, {'screen_name': 'prichaaarrrddd', 'name': 'Pri
extended_entities.media	nan
favorite_count	0
favorited	False
full_text	@welovetherays @prichaaarrrddd @RepSwalwell @benjaminwittes @JoshuaMZeitz Way to go completely off topic. Trump had every Right to not release his tax returns bu
geo	None
id	1098436205346279425
id_str	1098436205346279425
in_reply_to_screen_name	welovetherays
in_reply_to_status_id	1.09844e+18
in_reply_to_status_id_str	1098435799736041473
in_reply_to_user_id	2.84531e+08
in_reply_to_user_id_str	284530669
is_quote_status	False

Twitter - Following a Twitter Account

```
1 tjson=api.statuses.user_timeline(screen_name="realDonaldTrump
      ",tweet_mode='extended')
2 dftrump=json_normalize(tjson)
```

Index	nded_entities.me	favorite_count	favorited	full_text	geo
0	an	38758	False	THE WALL IS UNDER CONSTRUCTION RIGHT NOW! https://t.co/exUJCiITsz	None
1	an	38470	False	....something that is so obviously the future. I want the United States to win through competition, not by blocking out currently mor…	None
2	an	44246	False	I want 5G, and even 6G, technology in the United States as soon as possible. It is far more powerful, faster, and smarter than the curr…	None
3	an	167300	False	I have instructed Secretary of State Mike Pompeo, and he fully agrees, not to allow Hoda Muthana back into the Country!	None
4	{'id': 0982951388…	162735	False	We have just built this powerful Wall in New Mexico. Completed on January 30, 2019 – 47 days ahead of schedule! Many miles more now un…	None
5	an	116797	False	California now wants to scale back their already failed “fast train” project by substantially shortening the distance so that it no longe…	None
6	an	102576	False	The New York Times reporting is false. They are a true ENEMY OF THE PEOPLE!	None
7	an	86301	False	“If thinking that James Comey is not a good FBI Director is tantamount to being an agent of Russia, than just list all the peopl…	None
8	an	112286	False	“The Washington Post ignored basic journalistic standards because it wanted to advance its well-known and easily documented biased agenda…	None
9	an	130864	False	The Press has never been more dishonest than it is today. Stories are written that have absolutely no basis in fact. The writers don’t eve…	None
10	an	154525	False	Crazy Bernie has just entered the race. I wish him well!	None
11	an	66082	False	“Andrew McCabe gave absolutely no evidence of any threat to substantiate his ABSURD claim.” @LouDobbs	None
12	an	0	False	RT @GeraldoRivera: This is crazy scary. A cabal of unelected bureaucrats-angered & upset that @realDonaldTrump fired their bo…	None
13	{'id': 0979009040…	84823	False	https://t.co/xRbxknI4Nf	None
14	an	110992	False	I never said anything bad about Andrew McCabe’s wife other than she (they) should not have taken large amounts of campaign money from a …	None
15	an	116877	False	The Washington Post is a Fact Checker only for the Democrats. For the Republicans, and for your all time favorite President, it is a Fake …	None
16	an	108482	False	The failed Fast Train project in California, where the cost overruns are becoming world record setting, is hundreds of times more expensi…	None

Twitter - Followers of a Twitter Account

```
1 tfollow=api.followers.ids(screen_name="realDonaldTrump")
2 dffol=json_normalize(tfollow)
```

Index	ids	next_cursor	next_cursor_str	previous_cursor	revious_cursor_st	total_count
0	[1006168414652682240, 984612908234555393, 2327005869, 168493502, 109841920530765414…	16260837426…	16260837426…	0	0	None

Another way to do it is to directly normalize the nested JSON object that has the information we want.

```
1 tfollow=api.followers.ids(screen_name="realDonaldTrump")
2 dffol2=json_normalize(tfollow,'ids')
```

Index	0
0	10061684146…
1	98461290823…
2	2327005869
3	168493502
4	10984192053…
5	10986073768…
6	10897467098…
7	10895617253…
8	10986076440…
9	10108852183…
10	10986083157…
11	10986089611…

Twitter - Followers of a Twitter Account (Continued)

```
dfst2=json_normalize(search_result,'statuses')

u0=api.users.lookup(user_id=dffol2.loc[0,0])
dfu0=json_normalize(u0)
```

Index	mage_url_htt	profile_link_color	e_sidebar_border_	file_sidebar_fill_cc	profile_text_color	_use_background_	protected	screen_name	tatus.contributors	st
0	s://twimg.c…	1DA1F2	C0DEED	DDEEF6	333333	True	False	inmyfatassy	None	N

Using the id's you can get the messages from a twitter account.

```
user1=api.statuses.user_timeline(id=dffol2.loc[1,0],
    tweet_mode='extended')
dfuser1=json_normalize(user1)
```