

Week 9 Lab Assignment Goals

- Learn to use Tweepy (and thereby, the Twitter API)
- Learn to model n-to-n relationships in databases

Step 0 : Create a GitHub repository

- Go to <https://classroom.github.com/assignment-invitations/7b53c71e4b7722cf310fc87dc2563104>
- Accept the assignment invite and clone the assignment repository onto your machine
- Open a Terminal window or command prompt and 'cd' to the cloned directory

Step 1 : Get access to the Twitter API

- Sign in to your Twitter account
- Go to <https://apps.twitter.com> and select 'Create New App'
- Enter an application name, description, and website; and click 'Create your Twitter Application'
- On the next page, click 'manage keys and access tokens', then **copy and save your Consumer Key and Consumer Secret**.
- Click 'Create my access token', then **copy and save your Access Token and Access Token Secret**.

Step 2 : Starting out

- **Add your consumer key, consumer secret, access token, and access token secret to lab9_config.py**
- Install the tweepy module by running 'pip3 install tweepy' in the command prompt/Terminal

Step 3 : Save Taylor Swift's tweets to a text file

- Edit fetch_tweets.py to fetch the 300 most recent tweets by Taylor Swift (taylorswift13) and save them to a text file named 'tweets.json'. This file should contain one JSON-format tweet per line.
 - Hint: You will need one of Tweepy's [timeline methods](#)
- Run fetch_tweets.py and verify that tweets.json contains the right set of tweets. Here is what [this tweet](#) looks like in JSON format:

```
{  
  "id": 822929499851526146, "place": null, "coordinates": null, "in_reply_to_screen_name": null, "source": "<a  
href='http://twitter.com/download/iphone' rel='nofollow'>Twitter for iPhone</a>", "lang": "en", "text": "So much  
love, pride, and respect for those who marched. I'm proud to be a woman today, and every day.  
\\n#WomensMarch", "is_quote_status": false, "in_reply_to_status_id_str": null, "in_reply_to_status_id": null,  
  "in_reply_to_user_id_str": null, "id_str": "822929499851526146", "truncated": false, "geo": null, "user":  
  {"utc_offset": -28800, "profile_image_url_https":  
    "https://pbs.twimg.com/profile_images/720767103712645122/6XEBAXLj_normal.jpg", "contributors_enabled": false,  
    "profile_background_color": "C0DEED", "id": 17919972, "id_str": "17919972", "followers_count": 83659804,  
    "default_profile": false, "description": "Born in 1989.", "statuses_count": 4161, "favourites_count": 2103, "lang":  
    "en", "entities": {"description": {"urls": []}, "url": {"urls": [{"expanded_url": "http://smarturl.it/1989TourLIVE",  
      "indices": [0, 23], "url": "https://t.co/blhi4NNEJr", "display_url": "smarturl.it/1989TourLIVE"}]}}, "protected":  
    false, "is_translation_enabled": false, "following": false, "profile_background_image_url":  
    "http://pbs.twimg.com/profile_background_images/687293757/6d2ec27f32fa8cc2fcb7e6a9eada9945.jpeg", "profile_image_url":  
    "http://pbs.twimg.com/profile_images/720767103712645122/6XEBAXLj_normal.jpg", "geo_enabled": false, "screen_name":  
    "taylorswift13", "name": "Taylor Swift", "notifications": false, "translator_type": "regular", "verified": true,  
    "profile_background_tile": false, "default_profile_image": false, "profile_background_image_url_https":  
    "https://pbs.twimg.com/profile_background_images/687293757/6d2ec27f32fa8cc2fcb7e6a9eada9945.jpeg",  
    "profile_link_color": "0084B4", "friends_count": 244, "profile_sidebar_fill_color": "DDEEF6", "listed_count": 122740,  
    "is_translator": false, "location": "", "profile_text_color": "333333", "time_zone": "Pacific Time (US & Canada)",  
    "profile_use_background_image": false, "url": "https://t.co/blhi4NNEJr", "profile_banner_url":  
    "https://pbs.twimg.com/profile_banners/17919972/1409286315", "profile_sidebar_border_color": "FFFFFF", "created_at":  
    "Sat Dec 06 10:10:54 +0000 2008", "has_extended_profile": false, "follow_request_sent": false}, "contributors": null,  
  "retweet_count": 51623, "in_reply_to_user_id": null, "entities": {"hashtags": [{"indices": [103, 115], "text":
```

```
"WomensMarch"}], "urls": [], "symbols": [], "user_mentions": []}, "favorited": false, "favorite_count": 171417, "retweeted": false, "created_at": "Sat Jan 21 22:10:56 +0000 2017")
```

- Notice that there's a lot of information in each tweet. You can use an online JSON viewer to see what the JSON tree structure looks like, as we did in lab 6. For this lab, we're interested in the tweet ID, text, favorite_count, and hashtags.

Step 4 : Create a SQLite database

We want to store the tweets and hashtags in a database, such that we can retrieve all the tweets that contain a particular hashtag; and separately, all the hashtags that a tweet contains.

- init_db.py creates a SQLite database called 'tweets.db'
- Edit init_db.py to:
 - Create a table named Tweets with the following fields:
 - tweet_id : the ID of a tweet
 - tweet_text : the text of the tweet
 - likes : the number of likes the tweet received
 - Create a table named Hashtags with the following fields:
 - hashtag_id : a unique integer ID for each hashtag
 - hashtag_text : the hashtag text
 - num_occurrences : the number of times this hashtag appears in our dataset
 - **Tweets and hashtags have an n-to-n relationship** (a tweet can contain multiple hashtags; and a hashtag can be used in multiple tweets). **How will you model this in the DB?**
 - Hint: n-to-n relationships can be modeled using a 'map' table containing the id fields of the tables you are mapping between. Thus, a row in the map table would indicate that the tweet with ID tweet_id contains the hashtag with ID hashtag_id.

Step 5 : Store Taylor Swift's tweets in your database

Write a program 'add_to_db.py' that iterates over tweets.json, and for each tweet:

- Saves it to the Tweets table
- Saves its hashtags, if any, to the Hashtags table, making sure to maintain the n-to-n relationship and update num_occurrences if the hashtag is used multiple times in our dataset.
 - Your Hashtags table should contain 51 entries, e.g. fiftyshadesdarker

Step 6 : Query the database!

Write a program 'query_db.py' containing two methods:

- *most_common_hashtags()* should print the 20 most commonly-used hashtags in our dataset
 - Our output looks like this:


```
Most common hashtags:
idontwannaliveforever
fiftyshadesdarker
GRAMMYS
1989TourNashville
```

```

OOTWMusicVideo
1989TourStLouis
ryanadams1989
1989WorldTourLIVE
iHeartAwards
BadBlood
HappyBirthday1989
RyanAdams1989
1989TourTampa
1989TourGreensboro
Billboard200
FiftyShadesDarker
WomensMarch
LiveByNight
TaylorSwiftNOW
drakevsbenchpress

```

- *fifty_shades_darker()* should print the ID and text of all tweets containing #fiftyshadesdarker
 - Our output looks like this

```

Tweets containing #fiftyshadesdarker:
825736192654188544      '#idontwannaliveforever #fiftyshadesdarker\nhttps://t.co/PwER46Gt3F'
825107367624900610      '#idontwannaliveforever #fiftyshadesdarker https://t.co/Tl9TAHIUTZ'
824366759281131525      'Video tomorrow night midnight est #idontwannaliveforever
#fiftyshadesdarker https://t.co/fjWhrnNnSf'
824259659712307200      'Coming soon... #idontwannaliveforever\n#fiftyshadesdarker
https://t.co/BVoMeAQiv'
807079149810700288      '#fiftyshadesdarker\n#idontwannaliveforever https://t.co/Q51nqZ4ppX'
807076804687183873      'Z | T | 50\n\n@fiftyshades #fiftyshadesdarker \n@itunes
https://t.co/m4meFS4F79'

```

Step 7 : Commit code to GitHub

- Commit and push everything EXCEPT lab9_config.py to GitHub
- Once again: **do not push lab9_config.py** [because it contains your secret Twitter API keys]

Step 8 : Bonus round!

- Compare Taylor Swift's most commonly-used hashtags with those for Kim Kardashian's 300 most recent tweets. Would this require any modification to your table structures?