

## ResBaz – Scraping Twitter with Python (Instructor: Darla Still)

### Description of Workshop

In this workshop, we will be using Python to scrape Twitter. The script we will run grabs tweets from a single public username – this can be done as many times for as many usernames you would like to analyze; however, we will focus on three handles in this workshop. We will also do some descriptive analysis of these tweets in Python. You should walk away with knowledge of how to scrape Twitter, what elements of a tweet are available in this free service, how to manage data in Python, and how to do some descriptive analysis of text in Python. If you are unable to receive authorization needed to scrape Twitter, no worries! There will be data files provided for this workshop, which will be used for the examples during instruction.

### Getting Started

Before the workshop:

1. To create an app, you must have an approved developer account. Do this first. Provide as much detail as possible in the application process. Presumably, you are scraping Twitter for your research, so it would be best to describe your project in the application.
  - a. Apply here: <https://developer.twitter.com/en/application/use-case>
2. Wait for approval... while you wait...
  - a. Become familiar with Twitter's Terms of Service: <https://developer.twitter.com/en/developer-terms>
  - b. See also:
    - i. <https://www.youtube.com/watch?v=GWpCHh54pXU>
    - ii. <https://www.bitbybitbook.com/en/1st-ed/ethics/>
3. Next, create a Twitter app by clicking here (logging into your Twitter account) and following the steps.
  - a. <https://developer.twitter.com/en/apps>
  - b. More information: <https://developer.twitter.com/en/docs/basics/apps/overview>
4. Locate your individual Consumer API Keys and Access tokens:

The screenshot shows the Twitter Developer Portal interface. At the top, there's a breadcrumb 'Apps > fire-app-hack-tx-2019'. Below this are three tabs: 'App details', 'Keys and tokens' (which is selected), and 'Permissions'. A yellow warning box is present, stating: 'Important notice about your access token and access token secret. To make your API integration more secure, we will no longer show your access token and access token secret beyond the first time that you generate it starting January 21, 2020. You will be able to regenerate it at anytime here, which will invalidate your current access token and secret. Please save this information if you need to access it. This does not affect your consumer API keys, which will still be shown here as they are below. To learn more, visit the Forums.' Below the warning, the 'Keys and tokens' section is displayed. It includes a sub-header 'Consumer API keys' with two fields: one for the API key (redacted) and one for the API secret key (redacted), with a 'Regenerate' button below them. The next section is 'Access token & access token secret', showing an 'Access token' (redacted) and an 'Access token secret' (redacted). At the bottom, it shows the 'Read and write' access level and 'Revoke' and 'Regenerate' buttons.

5. Install “tweepy” in your Python environment
  - a. <https://pypi.org/project/tweepy/>
6. If you already have an app/credentials and have used tweepy to scrape previously, you are set for this workshop!