**Finding Walmart’s Image using Twitter Data**

**(User Manual and Dataset)**

**Team Name : #JD**

**Github :** [**https://github.com/supark/data\_mining\_project**](https://github.com/supark/data_mining_project)

**Dewan Chowdhury, dchowdhury3@student.gsu.edu**

**Jeongsu Park, jpark93@student.gsu.edu**

**Tools**

-Python 3.6

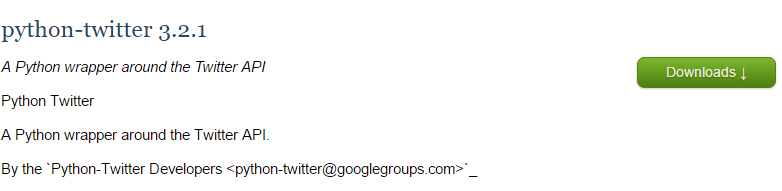
-Python twitter library

-Terminal

**Procedures**

1. Install Python 3.6 - Python 3.6 is the latest version. All our Python codes of this project are followed by 3.6 rule. Therefore, Python 3.6 is necessary in order to compile the following codes. Download python 3.6 from the following link <https://www.python.org/downloads/>
2. Python Twitter Tools - to connect to Twitter API which is then used to download data from Twitter.

2.1 - Python Twitter Tools is downloaded from <https://pypi.python.org/pypi/python-twitter/3.2.1>

2.2. - 

On this page, click download button.

2.3 - After downloading, please type following command lines on the terminal.

$ easy\_install pip

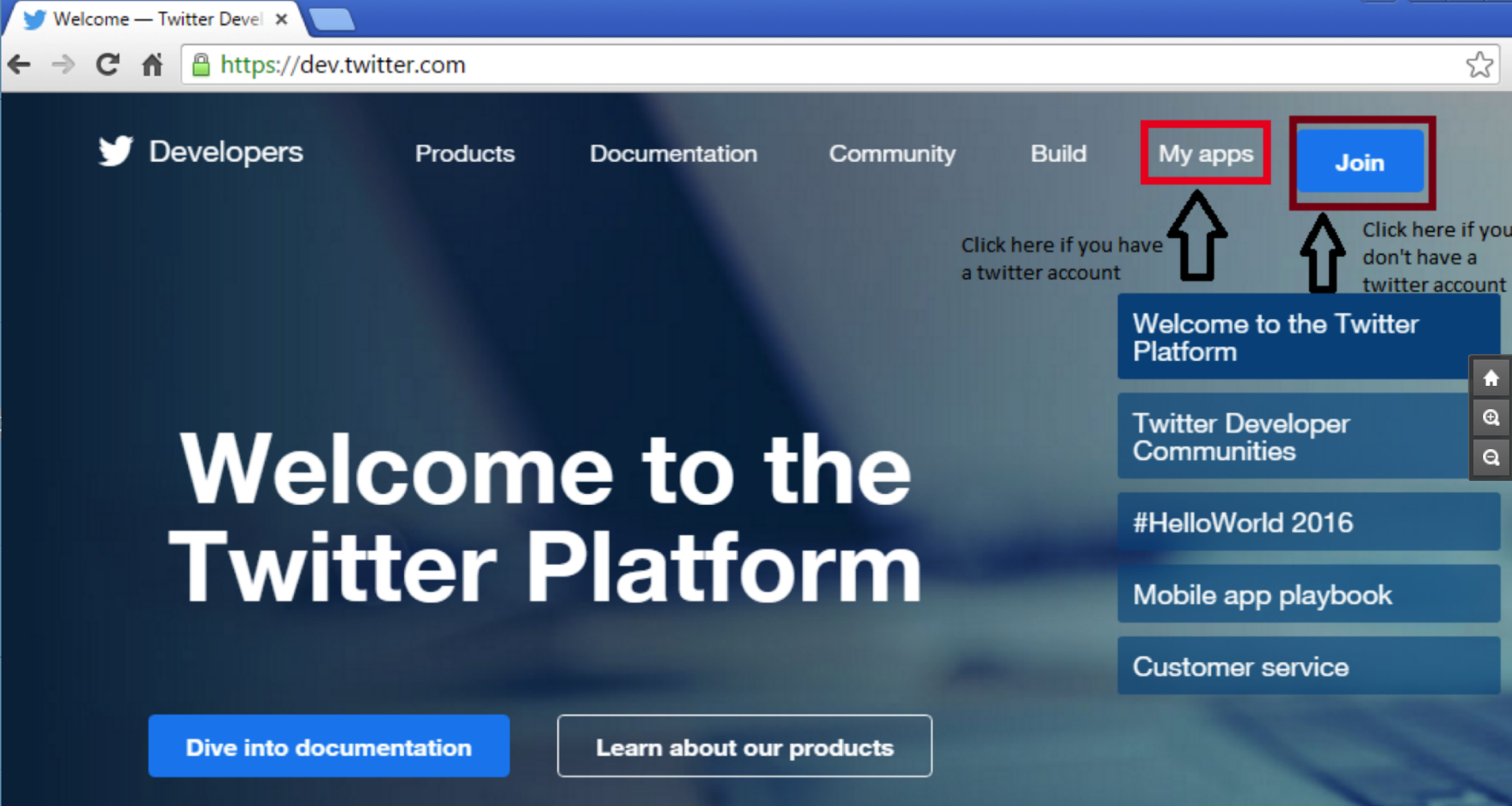
$ python get-pip.py

$ pip install simplejson

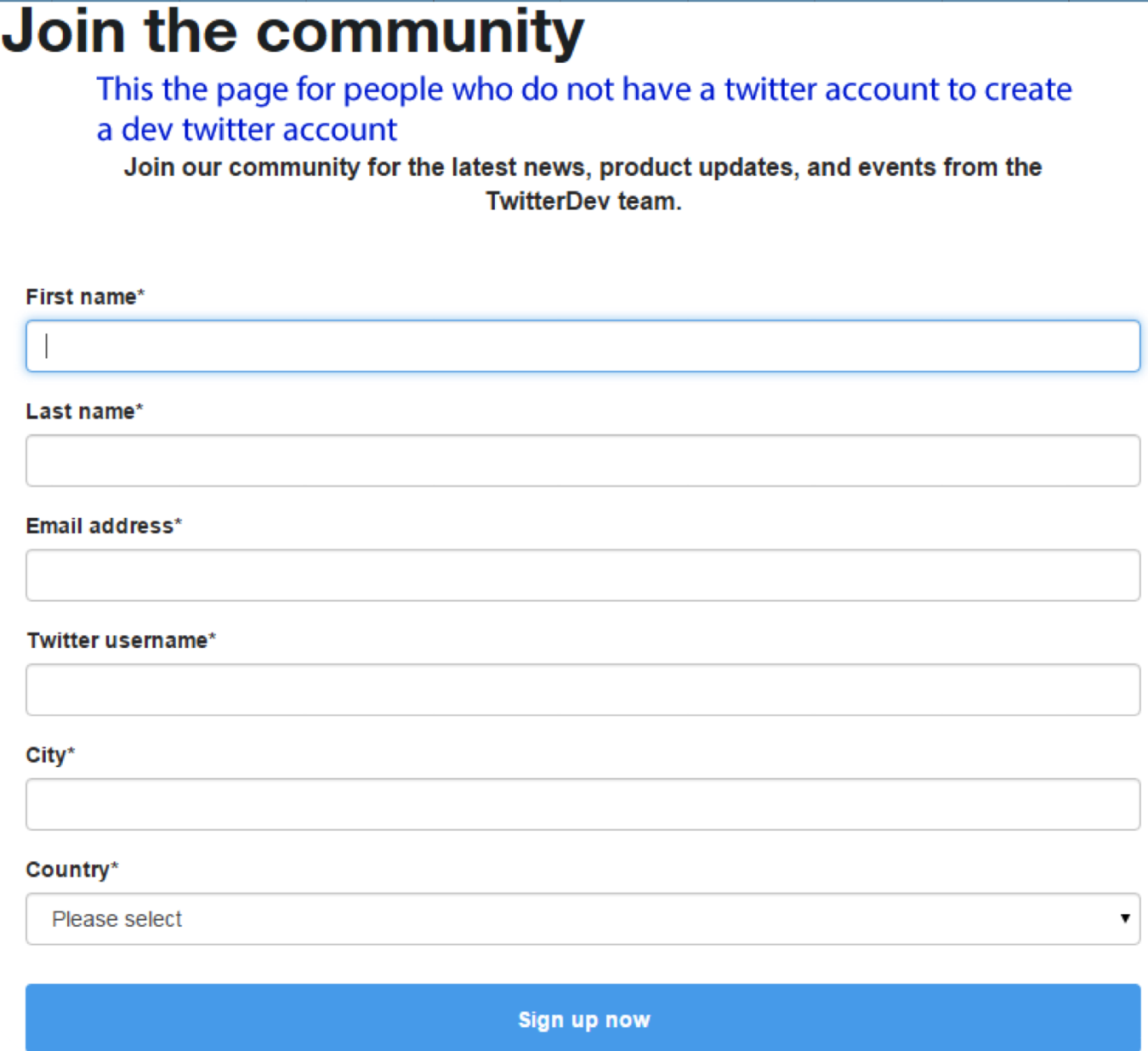
$ pip install twitter

3.Twitter API - Access token, access secret, consumer key, and consumer secret are needed to access to Twitter API.

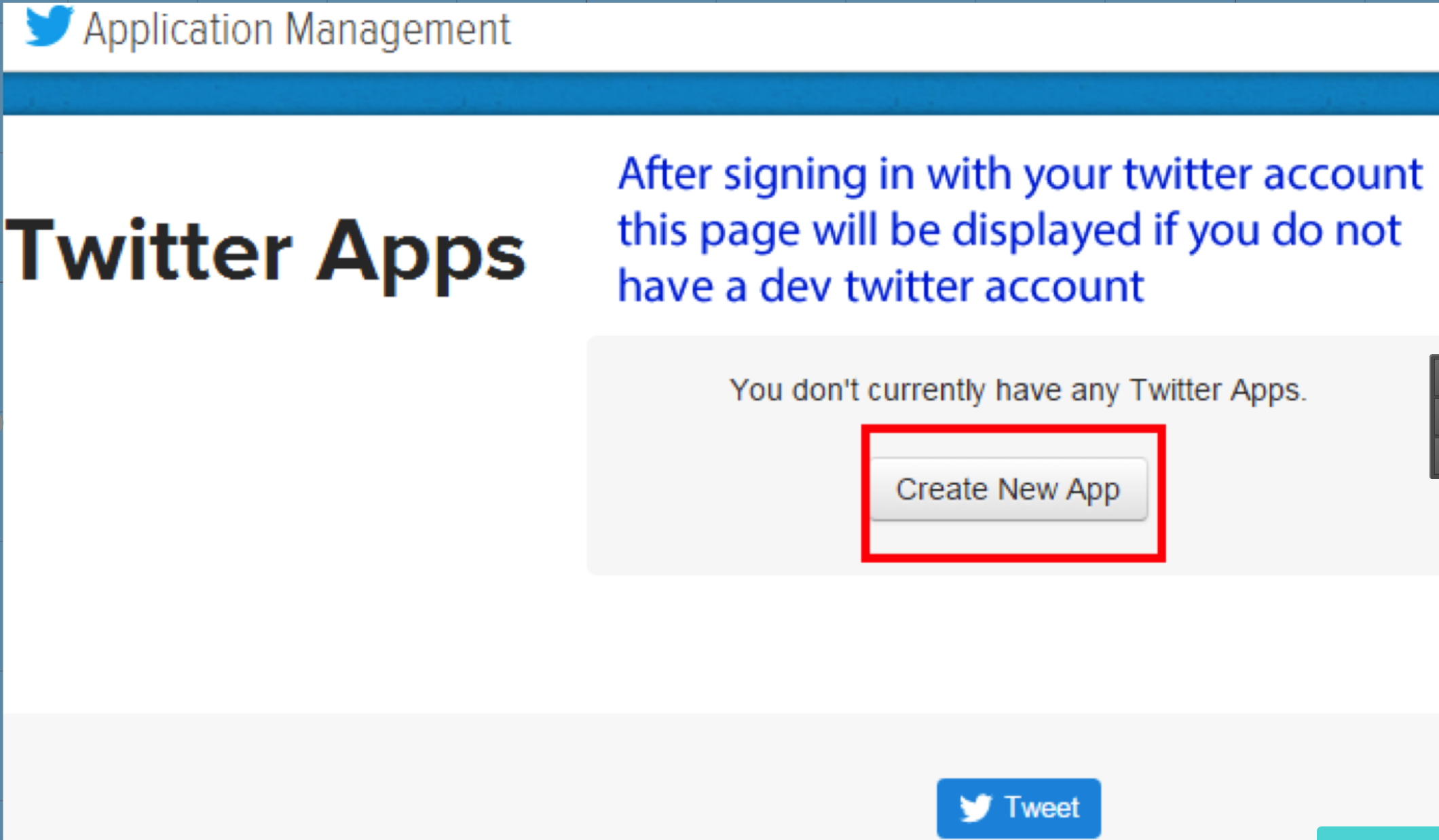
3.1 - Create twitter account unless you have one



3.2. <https://apps.twitter.com/> go to this link and sign in with your twitter account.

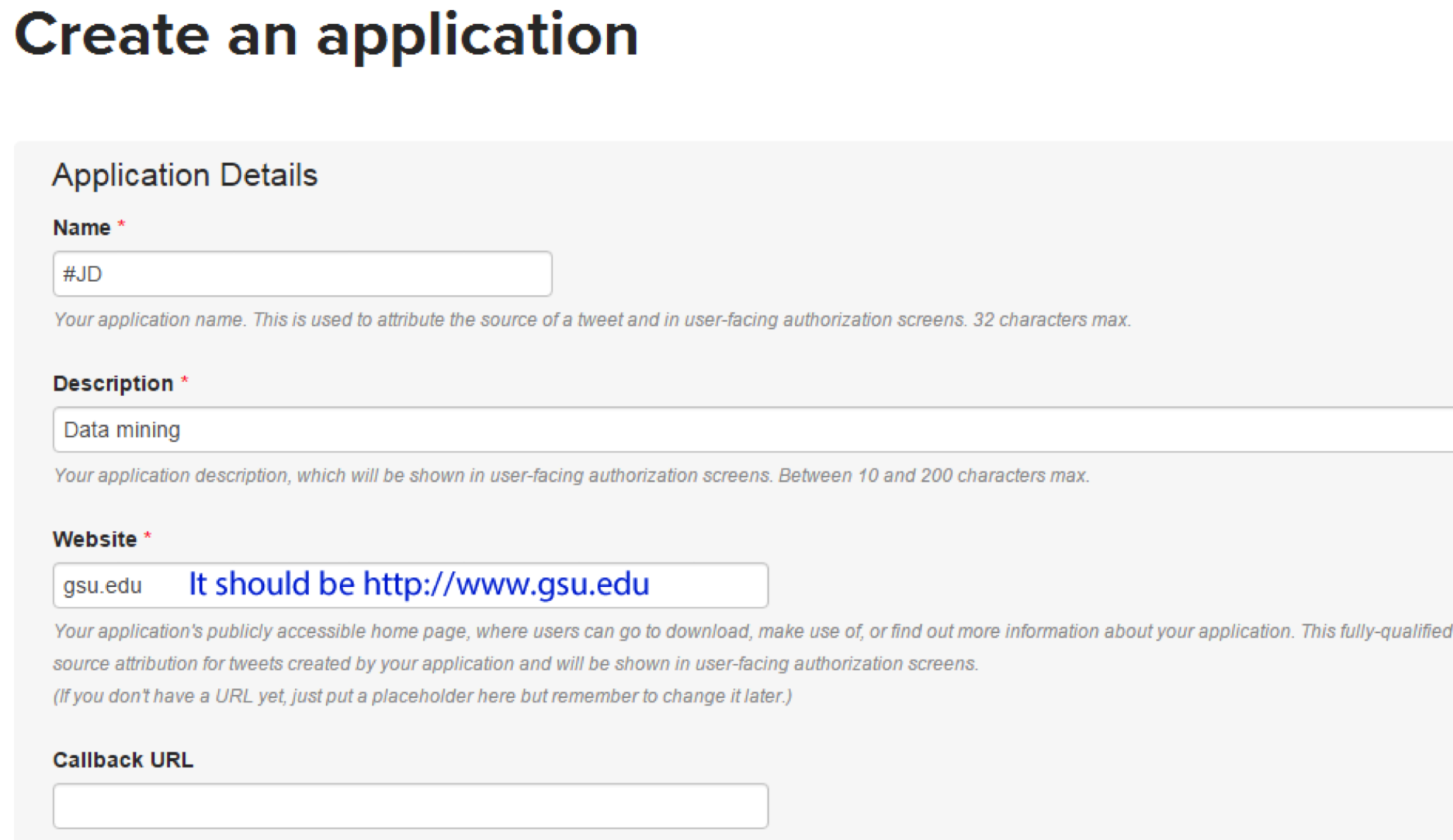


* 1. This allows you to create a Twitter developer account with the same user name.
  2. Even if someone does not own a twitter account, one can still create Twitter developer account.



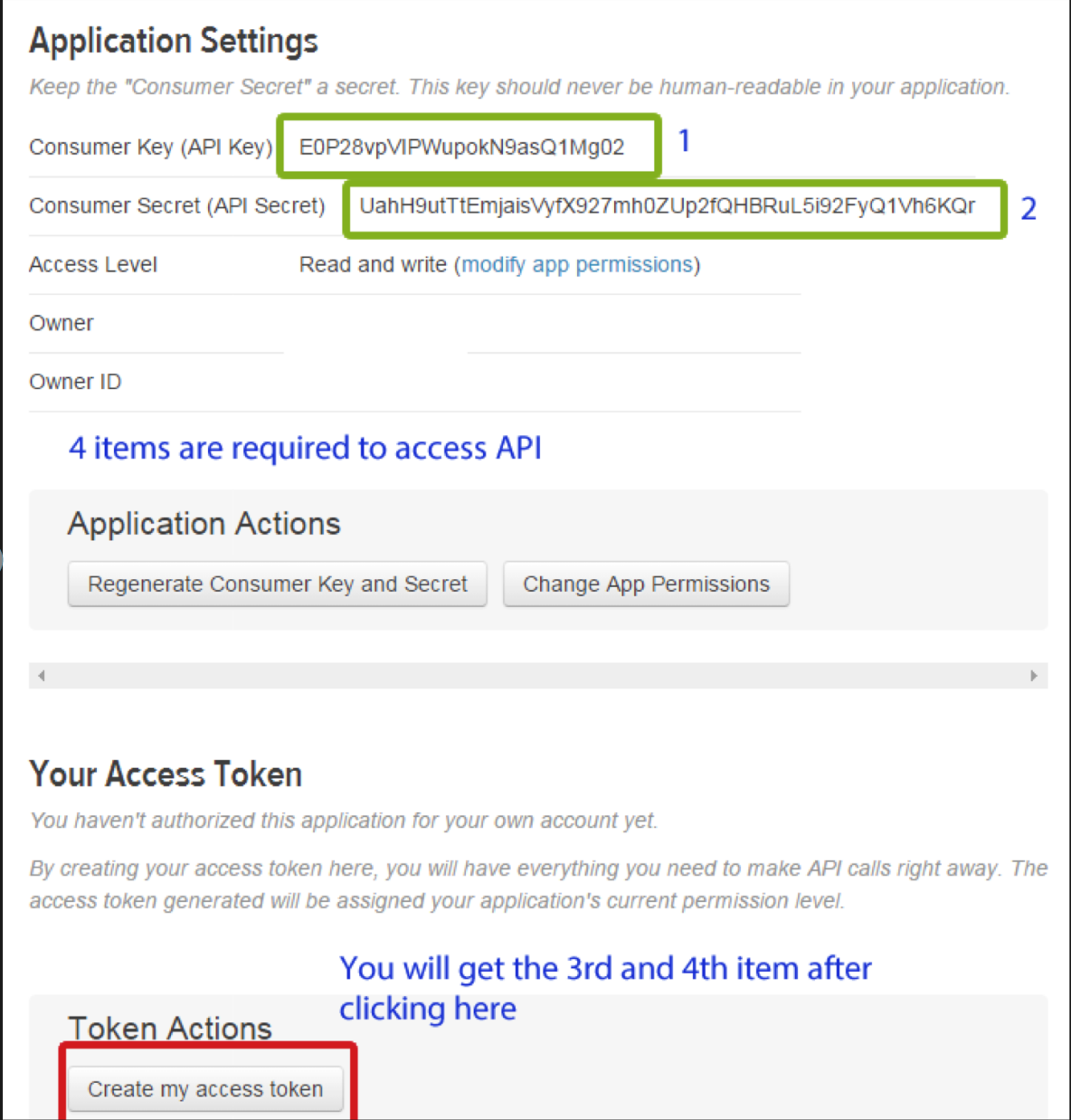
3.3. Click “Create New App”.

3.4. Fill out the form, agree to terms and conditions.

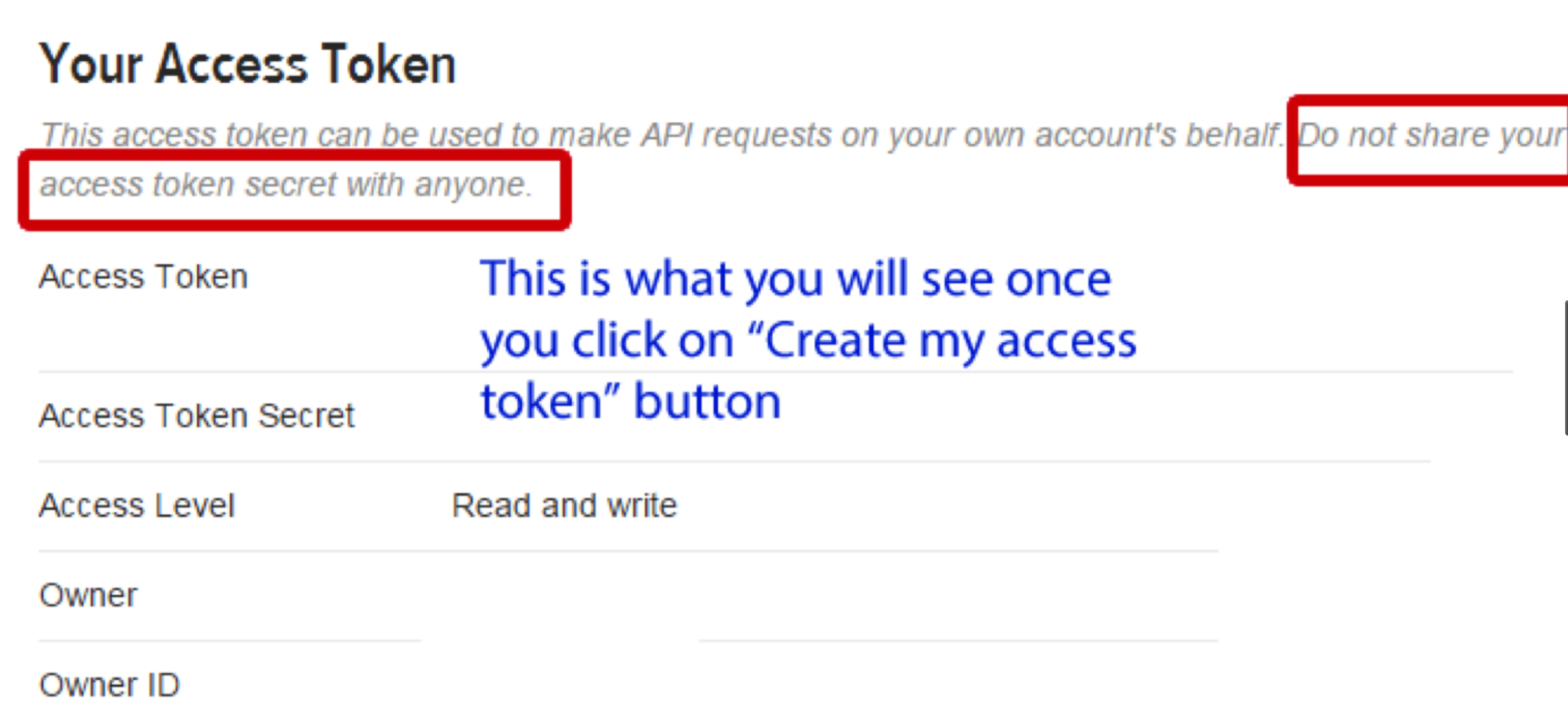


3.5. Click “Create your Twitter application”.

3.6. Click on “Keys and Access tokens” tab and copy the “API key” and “API secret”.



3.7. As you can the picture, consumer key and consumer secret are automatically generated but access token and access secret. Scroll down and click “Create my access token” and copy the “Access token” and “Access token secret”.



3.8. Now you have 4(access token, access secret, consumer key, and consumer secret). You can change yours when you execute the python files. However, you can still use ones in the code.

4. Execution

All of the files are available in the Github.

Please download all of the Python files first.

Now you are ready for executing the all the python files.



4.1. Open the terminal and type the below command line.

$ python collect.py > collect.json

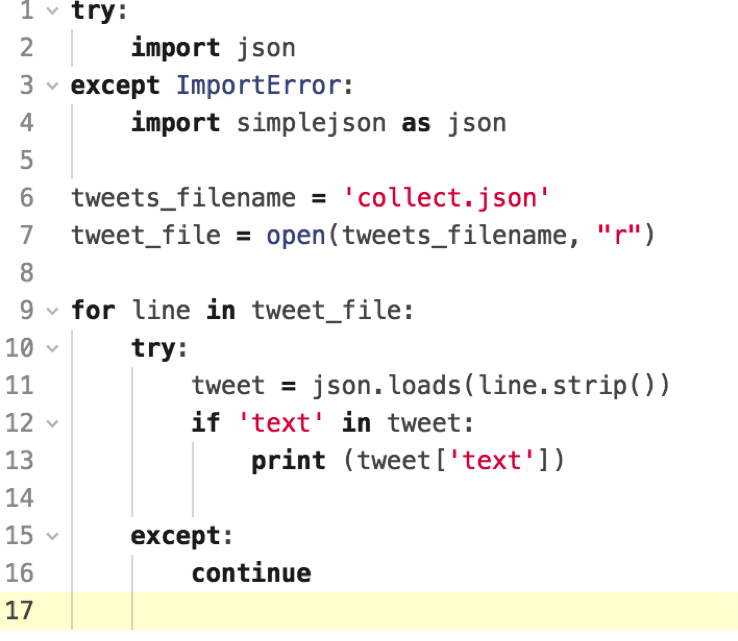
It took approximately 4 hours because of Streaming API twitter. If you do not want to spend such that time, you can tweak the code. On the line 20, change the number instead of 10000 because the execution will be finished until finding 10000 of the keyword, walmart.

Open collect.json to check out the result of the dataset like below.



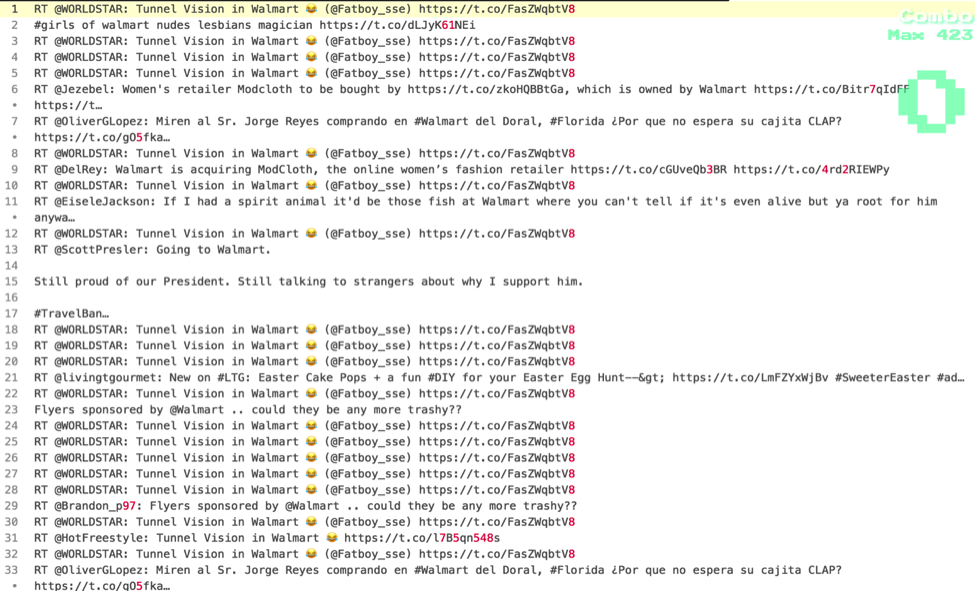
4.2. Open the terminal and type the below command line:

$python extractText.py > extractText.json



The above code extracts only text area.

Open extractText.json to see like below.

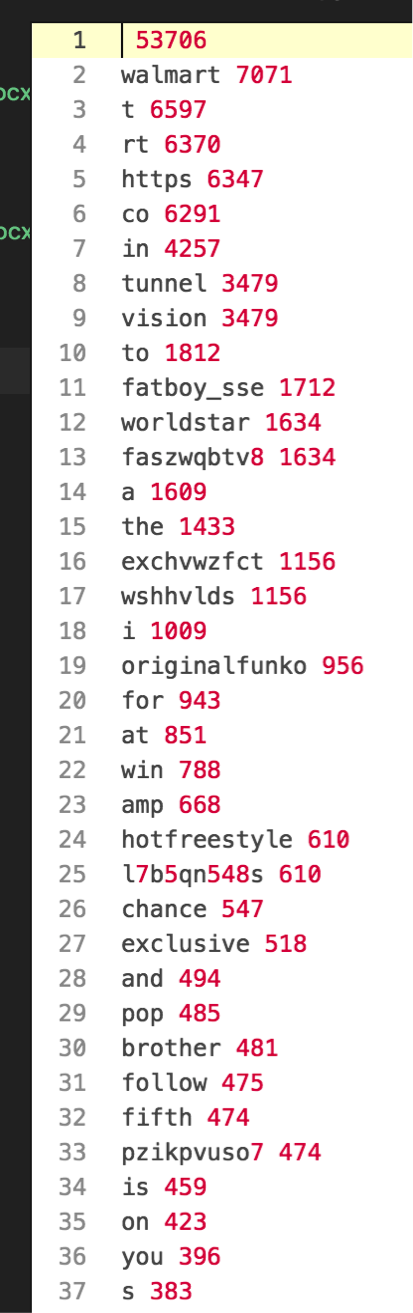


4.3. Open the terminal and type the below command line:

$ python ranking.py > ranking.json

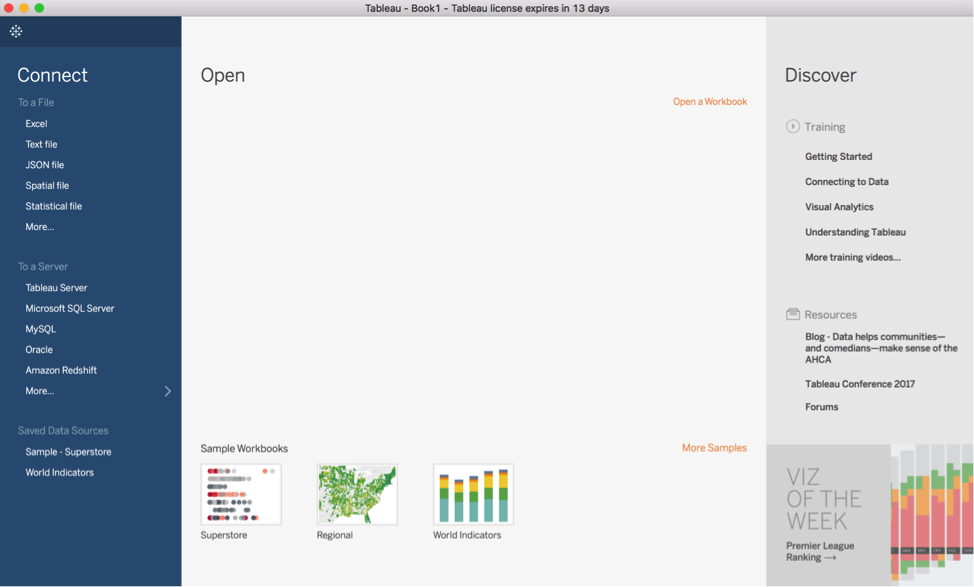


Open ranking.json to see like below.



4.4. All unnecessary words has been eliminated from ranking.json. Unnecessary words like articles, conjunctions and etc were extracted manually and saved in the excel file, [result.xlsx](https://github.com/supark/data_mining_project/blob/master/result.xlsx).

Install tableau from<https://www.tableau.com/> in order to create wordcloud image file.



Upload the excel file, which is in the Github, then convert the results into a word cloud format using Tableau.

