

Data Analysis using

...

By Eddie Aguilar, Mohammed Gour, Pranik Chainani, Shawn Manalel
and Zachariah Dohogne

Introduction

What is twitter?

How will be using Twitter data?

What are we trying to achieve?

What data are we using from twitter...

What are the requirements

What is needed to Extract Data?

1. Make a Twitter Account
2. Create a Twitter Application
3. Install Packages/Libraries
4. Data Extraction File
5. Plotting the Data

Twitter Account & Application

Why is an account needed?

- To have access to the twitter API

What is the Twitter Application?

- To communicate with Twitter API
- <https://apps.twitter.com/>
- It gives consumer keys and access tokens

Setting up Python (Windows users)

- This tutorial is using Python 3.6
 - Different Python versions have different syntax
 - From Python 2 to Python 3
- Install the latest version of python

Download the latest version for Windows

Download Python 3.6.1

Download Python 2.7.13

Wondering which version to use? [Here's more about the difference between Python 2 and 3.](#)

Looking for Python with a different OS? Python for [Windows](#),
[Linux/UNIX](#), [Mac OS X](#), [Other](#)

Want to help test development versions of Python? [Download](#)



Setting up Python (Mac Users)

- The tools are needed to get started

1) Get XCode : <https://developer.apple.com/xcode/downloads/>

2) Open the terminal and run the following

a) Make a new directory , in case lets call it “code”

```
$ mkdir Code
```

3) Change into your directory and install homebrew

```
$ ruby -e "$(curl -fsSL https://raw.githubusercontent.com/mxcl/homebrew/go)"
```

4) Install python3 with homebrew

```
$ brew install python3
```

Setting up Python (Mac users, cont.)

5) After successfully installing python install pip

```
$ curl -O http://python-distribute.org/distribute_setup.py
$ python distribute_setup.py
$ curl -O https://raw.githubusercontent.com/pypa/pip/master/contrib/get-pip.py
$ python get-pip.py
```

Setting up Python (Linux users)

- Python 2.7 comes installed by default
- Python 3.6 Is needed
- If Ubuntu/Debian :
 - `sudo apt-get install python3`

Make a virtual environment (Mac and Linux only)

- Install virtualenv through pip
 - `pip install virtualenv`
- Once installed make a virtual environment on the terminal
 - `virtualenv -p /usr/bin/python3.6 my_project`
- Activate the virtual environment
 - `source my_project/bin/activate`

Installing Packages

Tweepy: `pip install tweepy`

Pandas: `pip install pandas`

Matplotlib: `pip install matplotlib`

Extract File


```
1  from tweepy import Stream
2  from tweepy import OAuthHandler
3  from tweepy.streaming import StreamListener
4  import time
5
6  ckey = 'Enter Your API Key'
7  csecret = 'Enter Your API Secret'
8  atoken = 'Enter Your Access Token'
9  asecret = 'Enter Your Access Toekn Secret'
10
```

1




Extract File (2nd part)

```
1  #This is a basic listener that just prints received tweets to stdout.
2  class StdOutListener(StreamListener):
3
4      def on_data(self, data):
5
6          try:
7              print (data)
8              saveFile = open('ENTER-NAME-OF-YOUR-TEXT-FILE.txt', 'a')
9              saveFile.write(data)
10             saveFile.write('\n')
11             saveFile.close()
12             return True
13         except BaseException:
14             time.sleep(5)
15
16     def on_error(self, status):
17         print (status)
```



Extract File (3rd part)

```
def on_error(self, status):  
    print (status)  
  
if __name__ == '__main__':  
  
    #This handles Twitter authentication and the connection to Twitter Streaming API  
    auth = OAuthHandler(ckey, csecret)  
    auth.set_access_token(accessToken, accessTokenSecret)  
    twitterStream = Stream(auth, StdOutListener())  
  
    #This line filter Twitter Streams to capture data by the keywords: 'pokemonGO', 'gameofThrones'  
    twitterStream.filter(track=["pokemonGo", "gameofThrones"])
```



Structuring The File

```
# Reading Tweets
print('Reading Tweets\n')
# File Name be sure to change!!!!
tweets_data_path = 'twitDB.txt'

tweets_data = []
tweets_file = open(tweets_data_path, "r")
for line in tweets_file:
    try:
        tweet = json.loads(line)
        tweets_data.append(tweet)
    except:
        continue
```

Attributes

```
# Structuring Tweets
print('Structuring Tweets\n')

tweets = pd.DataFrame()
tweets['id'] = list(map(lambda tweet: tweet.get('id', None), tweets_data))
tweets['country'] = list(map(lambda tweet: tweet['place']['country'] if tweet['place'] != None else None, tweets_data))
tweets['full_name'] = list(map(lambda tweet: tweet['place']['full_name'] if tweet['place'] != None else None, tweets_data))
tweets['lang'] = list(map(lambda tweet: tweet['lang'], tweets_data))
tweets['favorite_count'] = list(map(lambda tweet: tweet.get('favorite_count', None), tweets_data))
tweets['favourites_count'] = list(
    map(lambda tweet: tweet['user']['favourites_count'] if tweet['user'] != None else None, tweets_data))
tweets['followers_count'] = list(
    map(lambda tweet: tweet['user']['followers_count'] if tweet['user'] != None else None, tweets_data))
tweets['friends_count'] = list(
    map(lambda tweet: tweet['user']['friends_count'] if tweet['user'] != None else None, tweets_data))
tweets['statuses_count'] = list(
    map(lambda tweet: tweet['user']['statuses_count'] if tweet['user'] != None else None, tweets_data))
tweets['hashtags'] = list(
    map(lambda tweet: tweet['entities']['hashtags'] if tweet['entities'] != None else None, tweets_data))
tweets['entities'] = list(map(lambda tweet: tweet.get('entities', None), tweets_data))
tweets['retweet_count'] = list(map(lambda tweet: tweet.get('retweet_count', None), tweets_data))
tweets['text'] = list(map(lambda tweet: tweet.get('text', None), tweets_data))
```

Excel

```
pip install xlwt
```

```
pip install xlswriter
```

```
# Bonus!!! Write to Excel
```

```
# Create a Pandas Excel writer using XlsxWriter as the engine.
```

```
writer = pd.ExcelWriter('twitData.xlsx', engine='xlsxwriter')
```

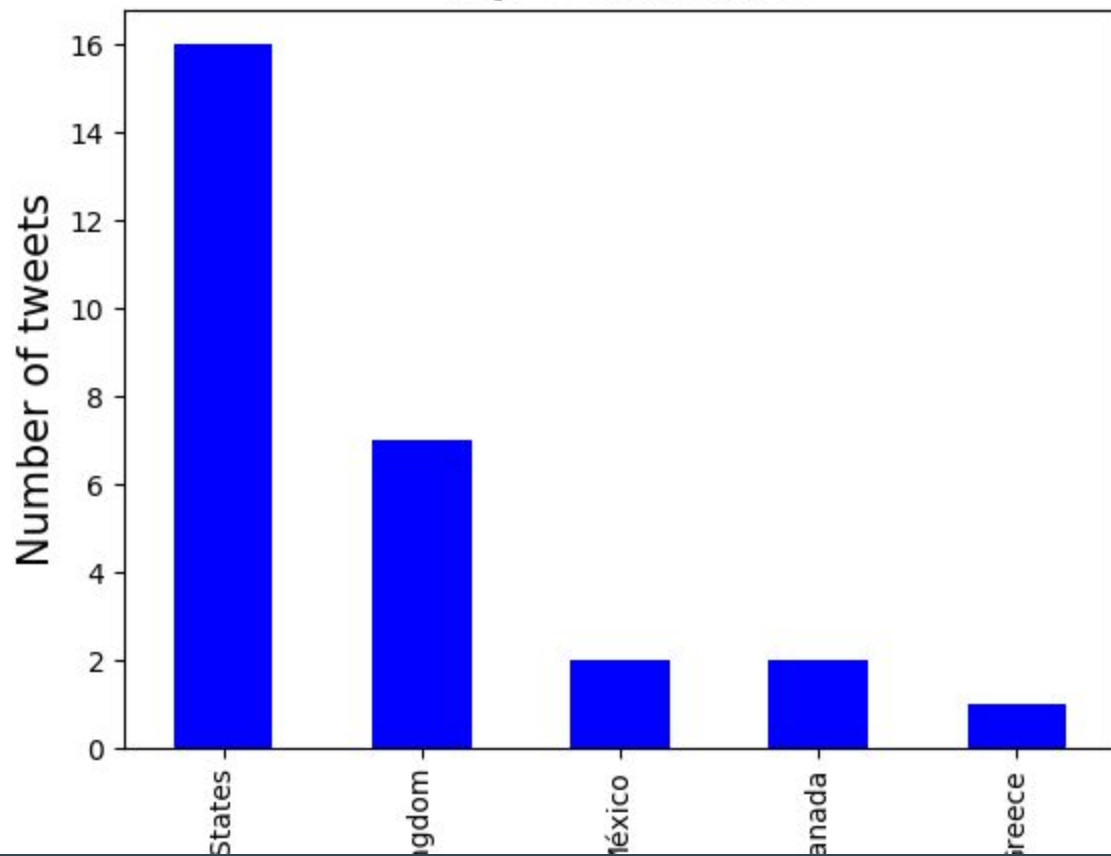
```
# Convert the dataframe to an XlsxWriter Excel object.
```

```
tweets.to_excel(writer, sheet_name='Sheet1')
```

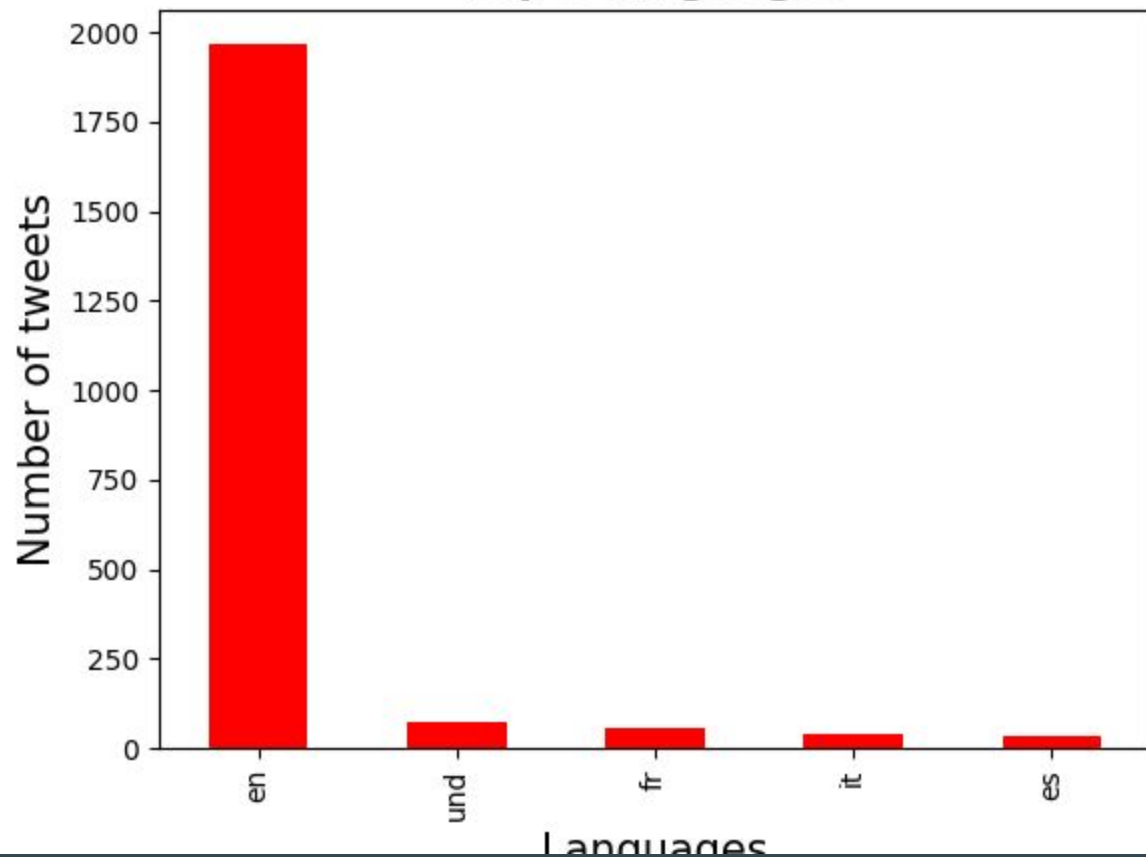
```
# Close the Pandas Excel writer and output the Excel file.
```

```
writer.save() # tweets_by_languages = tweets['lang'].value_counts()
```

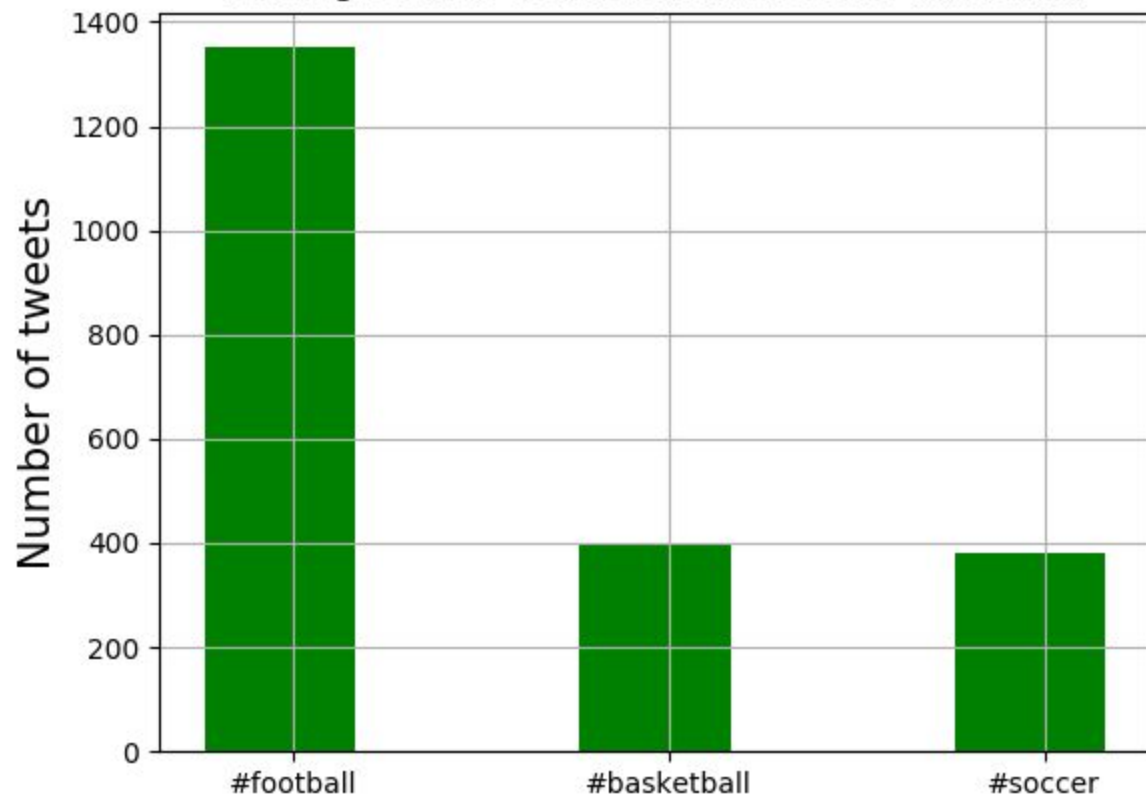

Top 5 Countries



Top 5 languages



Ranking: football vs. basketball vs. soccer (Raw data)



Questions & Comments