

Tweepy save to file

In [2]:

```
1 import pandas as pd
2 from tweepy.streaming import StreamListener
3 from tweepy import OAuthHandler
4 from tweepy import Stream
5 import time
6
7
8 class savetweepyTweets(StreamListener):
9
10     def on_data(self, data):
11
12         try:
13             print (type(data))
14             print ("New tweet")
15             savefile = open('twitter.txt','a')
16             savefile.write(data)
17             savefile.write('\n')
18             savefile.close()
19             #x =x+1
20             return True
21         except BaseException as e:
22             print ("Failed ", str(e))
23             self.disconnect()
24
25     def on_error(self, status):
26         print (status)
27         stream.disconnect()
28
29 tl = savetweepyTweets()
30 au = OAuthHandler(CONSUMER_KEY,CONSUMER_SECRET)
31 au.set_access_token(ACCESS_TOKEN,ACCESS_TOKEN_SECRET)
32 stream = Stream(au,tl)
```

```
In [3]: 1 stream.filter(track=["data scientist","data"], languages=['en'])
```

[illegible]

```
In [4]: 1 stream.disconnect()
```

```
In [5]: 1 import json
2 tweets_data = []
3 json_data = open('twitter.txt','r')
4 #print type(json_data)
5 for line in json_data:
6     #print type(line)
7     try:
8         twe = json.loads(line)
9         #print type(twe)
10        tweets_data.append(twe)
11    except:
12        continue
13
```

```
In [6]: 1 len(tweets_data)
```

Out[6]: 124

```
In [7]: 1 tweet_df = pd.DataFrame(tweets_data)
        2 tweet_df
```

Out[7]:

	contributors	coordinates	created_at	display_text_range	entities	extended_e
0	NaN	NaN	Thu Apr 12 01:07:08 +0000 2018	NaN	{'hashtags': [{'text': 'RHONY', 'indices': [12...	
1	NaN	NaN	Thu Apr 12 01:07:08 +0000 2018	NaN	{'hashtags': [], 'urls': [], 'user_mentions': ...	
2	NaN	NaN	Thu Apr 12 01:07:08 +0000 2018	NaN	{'hashtags': [], 'urls': [], 'user_mentions': ...	
3	NaN	NaN	Thu Apr 12 01:07:08 +0000 2018	NaN	{'hashtags': [], 'urls': [], 'user_mentions': ...	
			Thu Apr 12			

```
In [8]: 1 pd.set_option('display.max_colwidth', 0)
        2 tweet_df[["created_at", "text"]]
```

Out[8]:

	created_at	text
0	Thu Apr 12 01:07:08 +0000 2018	Not surprised that @ramonasinger wasn't bothered by Luann's Halloween costume but I am surprised by Bethenny's reaction. #RHONY
1	Thu Apr 12 01:07:08 +0000 2018	RT @daveyalba: Zuckerberg's talking point of AI shifting to proactively look at content on Facebook is really frustrating. AI solutionism j...
2	Thu Apr 12 01:07:08 +0000 2018	RT @stoned2thabones: me smoking weed is beneficial for everyone, the only reason i haven't clocked a bitch in the face is because i'm too...
3	Thu Apr 12 01:07:08 +0000 2018	RT @QueerMeNow: Angel Rivera Gets Fucked Raw By His Boyfriend Ben Masters @AngelRiveraXXX @BenMastersxxx @HelixStudios \n\n 🍑 https://t.co/KM...
4	Thu Apr 12 01:07:08 +0000 2018	RT @boomfisted: mccree: whats it like... being a cyborg... \ngenji: my life is nothing but pain. i cant look in the mirror because im revolte...
	Thu Apr 12	RT @bevyedevannort: husband and wife drop their daughter off at school \n ICE agents

```
In [ ]: 1
```