[Music]

so since I made my last video about

Twitter scraping so many things happened

like Elon bought

Twitter change his name to X and then

did his best to shut down free scraping

from

Twitter and apparently in 2024 that's

one of the hardest thing to do to get

data from Twitter in this video I'm

going to show you how I managed to get

data from the website and along the way

I also show you some tips and tricks to

hopefully prevent you from getting

banned and also how to handle rate limit

exceptions so let's get started in the

terminal I type pip install T kit this

is the package we are going to use to

scrape Twitter data

now I create a new file call it

main.py then I do the Imports I do

from TW kit import client this a class

helps us to interact with Twitter and

get data from it and to handle rate

limit errors I import too many requests

then we also need some other Imports

import

time from date time import date

time import

CSV then from config parser import

config parser and lastly from random

import Rand in

it now I type the constants that I need

which

is minimum

minimum

tweets and let's start with 10 later

we're going to increase it to higher

numbers and also for our

query let's start with something really

simple chat GPT then later again we're

going to make the query more

complex next step is to import our

[Music]

login

credentials now I have highly suggest

you to create a secondary Twitter

account that if it gets banned it's

still okay with you during making the

materials for this video I had several

accounts getting banned so I always made

a secondary account to make sure

everything is fine then we put the

credentials of that Twitter account over

here so I make a new

file call it

config.ini let's inside of it create a

section called X put username over

here put password over

here and also put the email address over

here let's say

XXX sin xx.com after you typed your

credentials over here you can just close

it and come back to the code now that we

have the credentials in config file I

will start reading it over here so first

I create an instant of config

parser then I read my config in

file and then I start creating the

variables that I need so username

is config from the

xsection

username then

email and then

password next we should authenticate

to the website

so authenticate to

x.com now there are two methods that

that we can authenticate the first

method of course is

to use the login

credentials and the second method is to

use cookies what we do is first we log

in to the website once using the

credentials then we save cookies and

then from then on we will just use those

cookies to always have access to the

website and get the data so what I do is

now to create a client

instance and I pass the language to be

English us next I do

client.

login authentication info one is

username authentication info to is email

and then we pass the password

and then I do

client.

saave cookies and I save the cookies

into a Json file now when I run this

piece of code we will have a cookies.

Json file over here add

it so I do in the terminal Python

main.py and there it is now we have a

cookies file over here from now on we

are going to just use these cookies is

to have access so I go back to the

code I comment out these two

lines and I do client. load cookies Next

Step let's yeah get some

data so we

do get

tweets so I do tweets is equal to

client.

[Music]

search tweet

then I pass the query over

here and then the second one is

product which we could have either the

latest Media or top tweets let's start

with

top and then I do for tweet in tweets

tweet is a complicated object so to

print it I do print vs of to it so it

will convert it into to a dictionary and

let's just break the loop over here so

we only see one to it and it's all

details so now back to terminal I run

the

code and there we have it so we got one

tweet over here this is

really dense there's a lot of data over

here this is user that could give you

the ID or for example name of the user

user over here and so many more data so

we can grab all of these data over here

so I just clear this terminal go back to

my code instead of converting the Tweet

to a dictionary what I will do is to get

the important information over here so

before everything I just create a tweet

count constant set it equal to zero then

every time we get a tweet I increase

that number by one

then I get the Tweet data let's create a

list we put tweet

count tweet. user.name

tweet.

text let's say tweet. created

ads let's also put how many retweets it

had and also how many likes it had so we

get all of this data over here then I

just

print the Tweet

data let's run again the

code and there it is so we got one tweet

from Team

pole it has 444 retweets and almost

3,000 likes which is great now let's

instead of one tweet let's see more than

one so instead of breaking let's print

all of the tweets and at the end let's

just print how many tweets we got

normally I like to print out the

time let's say

done got this amount of tweets now we

rerun the

code so at the end we received 19 tweets

one issue that we have so far is that we

didn't use this minimum tweets over here

10 is okay what if we wanted

30 then we didn't get it over here so to

solve that after I create the Tweet

count consent I initialize tweets as

none then I write a while loop

while tweet count is less than minimum

tweets let's bring all of this piece

into a while loop then we say

if tweets is none

we search for tweets so let's get this

one bring it inside here let's also do a

print over here so we see where we are

getting

tweets if tweets is not NN that means we

already did one round of search so I

do print daytime getting more tweets or

let's instead call it getting next

tweets because what we will do is to

write tweets equal to

tweets. next so if I run this piece of

code now we should get at least 30

tweets there we have it we got 39 at

some point we got 19 tweets and then we

went for the next patch of tweets until

we got 39 tweets so still there is one

problem over here that we have is that

sometimes this next method returns empty

results because there were no more

results to get so to handle that we say

if we didn't get any

tweets say that no more twst were found

and then break our while loop we still

have one more issue over here is that

every time we are calling this next

method immediately that can be alarming

to Twitter because there are so many

calls one after another which can easily

flag our account and get banned so later

I'm going to show how to make this a bit

better so hopefully we don't get banned

and immediately but before that let's

save our tweets into a CSV file rather

than just printing so for that up here

before our

authentication we start

creating CSV

file so

we open a CSV file called tweet. CSV we

open it into a write mode and we name

name it file and then we create a writer

instant for

that then we write the colum names over

here now I don't like n o I just

put tweet

count the

username the text created

ads number of retweets and number of

likes that's nice then down here instead

of printing tweet data we write it into

a CSV file so we open the tweets. CSV

file into the append

mode and we start writing the rows again

we create a writer instant for the file

and then we write a row which has the

Tweet data over

here and then let's also over here add a

print instance just to see how many

tweets we get in each Loop over here

let's close it here that's great so now

when I run this code we will get all of

the data and save it into a CSV

file so I just clear over

here python

main.py we got 19 Tweets in the first

iteration then 39 tweets then done now

we have this tweets. CSV file over here

if I open it all of the data is here

39 tweets great so I close

this so now we have a functional piece

of code we can get whatever amount of

tweets we want and then write it into a

CSV file the next step is try to modify

the code in a way to simulate a human

behavior so hopefully we just don't get

banned immediately by Twitter for that

I'm going to add some delays to this

piece of code before call the twiter API

so first let's just clean up a bit this

code I take this piece of code I cut it

and create a function so I say tweets is

equal to

get tweets and I pass over here tweets

now I need to Define this get tweets

function so I go at the beginning of my

code after the constants I do def get

tweets and I paste our code over here

then at the end I return

tweets so we still didn't do anything we

just created a new function now every

time before getting new tweets I add

some DeLay So I go to my get tweet

function over here first we do a search

which is fine then I create a random

weight time so I call it weight time

it's a random number

from 5 to 10 so then I say getting new

tweets

after wait time

seconds then we do time.

sleep and we wait for that amount of

time before calling the next method now

let's go instead of 30 call 100 tweets

and run again our code

so we get 19 tweets then we wait for 8

seconds then we call for the next piece

of

data we got 39 tweets then again we wait

for 8 seconds and we just continue this

Loop until we get the 100 tweets that we

want finally after calling the API for

six times we got 19 19 tweets and if we

again check our tweets. CSV

file we should have 119 tweets over

here there it is now if we start

increasing this number of tweets to for

example 1,000 there will be some moment

that we had already called so many times

the Twitter API in a row so what happens

we will get a rate limit exception so we

have to also handle that in our code so

I go down

here where we are getting tweets instead

of getting the tweets directly we use

try and accept method so we first we try

get

tweets but if we got a too many request

exception we write it as e we can type

rate limit reset equal

to date time. from

timestamp e. rate limit read reset and

then print rate limit reached instead of

reset at that time I

type waiting

until then we

do

wait time equal to rate limit reset

minus date time now so that's the amount

of time we have to wait until everything

is fine so we wait until that by calling

the time time do Sleep Method and we

wait for the amount of wait time. total

seconds and then after that's done we

just continue our Loop so this way we

should be able to handle this

exception now we try to get 1,000 twist

related to chat GPT so I open my

terminal let's clean it up and start our

code so we finally got it 116 tweets it

took quite a lot of time I think like 10

minutes and at some point we also hit

the rate limit and we waited until the

rate limit got reset so we could again

get new tweets so this way we can now

get the data from the Twitter and

hopefully not to get banned now the last

piece that I promised to you was

to create a more complex query over here

so let's try to get 10 tweets from Elon

Musk account for creating complex

queries there is a really nice article

by the developer Twitter team and I put

a link to that article down below the

like button now there is a more easy way

of creating complex queries and for

doing that you can go to the X

website here is my account and then

search for something random let's say

coding let's say say coding

101 and then over here hit on advanced

search and then create your search query

let's say we want to search any type of

Tweets in

English from

Elon

Musk and let's make it from a specific

date let's say

January

[Music]

1st

2018 to January

1

2020 so I just hit search and there we

have it here is our query I just copy

this over here bring it inside the code

and

then run the

code done now if I open this tweets.

CSV we see we got all the tweets from

Elon Musk so that's how I get my Twitter

data in 20124 it's not easy we had to

use a lot of tricks here and there to

make it work at least it's for free and

we don't have to pay $100 per month to

get this data I hope it was useful for

you and it can help you in your projects

if you liked it please like the video

and subscribe thanks for watching and

happy scraping