extract the .csv file from the **GitHub Repo(**[**csse\_covid\_19\_data**](https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data)**)**

# [**Is there a way to import csv file from github automatically to my jupyter notebook**](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo)

[Ask Question](https://stackoverflow.com/questions/ask)

Asked 1 year, 9 months ago

Active [1 year, 9 months ago](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo/60626813?lastactivity)

Viewed 3k times

1

2

I am a beginner and I'm developing a code to visualize the spread of corona virus globally, I want to extract the .csv file from the **GitHub Repo(**[**csse\_covid\_19\_data**](https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data)**)** where a new .csv file is uploaded every 2 days. Instead of downloading the file manually is it possible to import the latest csv file to notebook automatically?

I have tried scraping the data but it doesn't help

import requests

url = 'https://github.com/CSSEGISandData/COVID-19/blob/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/03-08-2020.csv'

response = requests.get(url)

print(response.text)

[python](https://stackoverflow.com/questions/tagged/python)[web-scraping](https://stackoverflow.com/questions/tagged/web-scraping)[data-science](https://stackoverflow.com/questions/tagged/data-science)

[Share](https://stackoverflow.com/q/60609760)

[Improve this question](https://stackoverflow.com/posts/60609760/edit)

Follow

asked Mar 9 '20 at 23:45

[[Icon

Description automatically generated](https://stackoverflow.com/users/13035815/yoda)](https://stackoverflow.com/users/13035815/yoda)

[YoDa](https://stackoverflow.com/users/13035815/yoda)

**51**11 silver badge55 bronze badges

* This might be useful :: [python-how-to-download-a-single-specified-file-from-github](https://stackoverflow.com/questions/17843699/bash-or-python-how-to-download-a-single-specified-file-from-github)

– [leopardxpreload](https://stackoverflow.com/users/12489438/leopardxpreload" \o "732 reputation)

[Mar 9 '20 at 23:51](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo/60626813#comment107229533_60609760)

* 1

*I have tried scraping the data but it doesn't help* What does that mean, exactly? Please clarify what the issue is.

– [AMC](https://stackoverflow.com/users/11301900/amc)

[Mar 10 '20 at 0:54](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo/60626813#comment107230347_60609760)

[Add a comment](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo/60626813)

## 2 Answers

[Active](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo?answertab=active#tab-top)[Oldest](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo?answertab=oldest#tab-top)[Votes](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo?answertab=votes#tab-top)

7

## ****Solution 1:****

This solution is specific to your use case:

Install [PyGithub](https://pygithub.readthedocs.io/en/latest/introduction.html) package using the below pip command:

!pip install PyGithub

Generate Github API token from this [page](https://github.com/settings/tokens) by clicking on Generate new token and pass that token as a string in the below code at the place of token to establish a connection with Github:

Github\_personal\_access\_token is: -

ghp\_7wANkSal19vhZabALWetUOk9Ow9Rgu3ENFPK

from github.MainClass import Github

g = Github(token)

Now you are connected with Github using your credentials and you can access all of your repo contents as well as other public repos.

Load the repo in which your CSV files are stored:

repo = g.get\_repo("CSSEGISandData/COVID-19")

Get the list of object of the files stored in the directory where your CSV files are stored:

file\_list = repo.get\_contents("csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports")

Since the directory where these CSV files are stored also contains one .gitignore file and one README.md file and file nomenclature are of the format "mm-dd-yyyy", so README.md is present at the last and the last-second file is your latest updated file. To access that run the below code:

github\_dir\_path = 'https://github.com/CSSEGISandData/COVID-19/raw/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/'

file\_path = github\_dir\_path + str(file\_list[-2]).split('/')[-1].split(".")[0]+ '.csv'

Load the data from the specified path using the read\_csv() method of pandas.

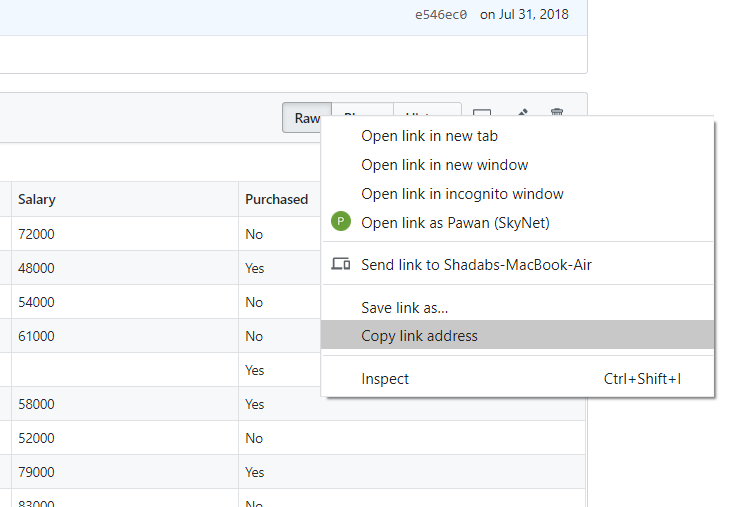
import pandas as pd

df = pd.read\_csv(file\_path, error\_bad\_lines=False)

## ****Solution 2:****

Try this code if you want to specify the path manually:

Get the path of your CSV file from Github by right-clicking on raw as shown below and assign its value to the file\_path:

[](https://i.stack.imgur.com/UYE8V.png)

file\_path = 'https://github.com/CSSEGISandData/COVID-19/raw/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/03-08-2020.csv'

Load the data from the specified path using the read\_csv() method of pandas:

import pandas as pd

df = pd.read\_csv(file\_path, error\_bad\_lines=False)

## ****Solution 3:****

Try this code if you want to specify the path automatically:

Set a time when you want to refresh your code and integrate the below-given solution with that.

Since you know the directory where the latest files are getting stored and how frequently new files are getting added to that directory, you can just change the date dynamically for the current date in the mm-dd-yyyy format:

from datetime import date

file\_date = str(date.today().strftime('%m-%d-%Y'))

file\_date

Output: 03-11-2020

Similarly, just change the value of file\_date if you want to run your code for yesterday's date:

from datetime import date, timedelta

file\_date = str((date.today() - timedelta(days = 1)).strftime('%m-%d-%Y'))

file\_date

Output: 03-10-2020

Since currently in that directory, the last file uploaded is on 9th March 2020, so we are going to use that date:

from datetime import date, timedelta

file\_date = str((date.today() - timedelta(days = 2)).strftime('%m-%d-%Y'))

file\_date

Output: 03-09-2020

Generate file\_path dynamically:

github\_dir\_path = 'https://github.com/CSSEGISandData/COVID-19/raw/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/'

file\_path = github\_dir\_path + file\_date + '.csv'

Load the data from the specified path using the read\_csv() method of pandas.

import pandas as pd

df = pd.read\_csv(file\_path, error\_bad\_lines=False)

[Share](https://stackoverflow.com/a/60626813)

[Improve this answer](https://stackoverflow.com/posts/60626813/edit)

Follow

[edited Mar 10 '20 at 23:58](https://stackoverflow.com/posts/60626813/revisions)

answered Mar 10 '20 at 22:30

[[A picture containing text, indoor, room, floor

Description automatically generated](https://stackoverflow.com/users/5718264/shadab-hussain)](https://stackoverflow.com/users/5718264/shadab-hussain)

[Shadab Hussain](https://stackoverflow.com/users/5718264/shadab-hussain)

**676**44 silver badges2121 bronze badges

[Add a comment](https://stackoverflow.com/questions/60609760/is-there-a-way-to-import-csv-file-from-github-automatically-to-my-jupyter-notebo/60626813)

2

Use:

https://raw.githubusercontent.com/CSSEGISandData/COVID19/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/03-08-2020.csv **[The 'raw' text]**

Example:

import requests

url = 'https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/03-08-2020.csv'

resp = requests.get(url)

print(resp.text)

[Share](https://stackoverflow.com/a/60610930)

[Improve this answer](https://stackoverflow.com/posts/60610930/edit)

Follow

answered Mar 10 '20 at 2:51