**Use publicly available data**

**Use data from a GitHub repository**

**Import multiple CSV files from a GitHub repository using a range of dates**

'''

Purpose:

(1) Import CSV files from a repository https://github.com/CSSEGISandData/COVID-19/tree/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports where CSV files are uploaded daily from 01-22-2020 and consistently named as mm-dd-yyyy.csv

Note: column names changed from 03-22-2020 and onward

How to run this file in ipython?

os.chdir(folder\_path\_script)

%run ./import-daily-files\_GitHub-COVID-19-csse-covid-19-daily-reports-CSV.py

'''

import requests

import shutil

import datetime

import os

import pandas as pd

folder\_path\_script='D:/googleDrive/python/scripts/'

# Column names changed from columns\_to\_read\_1 to columns\_to\_read\_1 on 03-22-2020.csv

columns\_to\_read\_1=['Province/State', 'Country/Region', 'Last Update', 'Confirmed','Deaths', 'Recovered']

columns\_to\_read\_2=['Province\_State', 'Country\_Region', 'Last\_Update', 'Confirmed','Deaths', 'Recovered']

# Create date ranges similar to the CSV files

dates=[datetime.date(2020,1,22)+datetime.timedelta(dval) for dval in range(0,366)];

date\_column\_names\_changed=datetime.date(2020,3,22)

# Create an empty dataframe for appending imported CSV files

df\_base=pd.DataFrame(columns=columns\_to\_read\_2)

# Loop thru each CSV file URL

for dateval in dates:

filepath='https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/'+ dateval.strftime('%m-%d-%Y')+'.csv'

r= requests.get(filepath)

if r.status\_code == 200:

if dateval < date\_column\_names\_changed:

df\_temp=pd.read\_csv(filepath,usecols=columns\_to\_read\_1)

df\_temp.columns=columns\_to\_read\_2

else:

df\_temp=pd.read\_csv(filepath,usecols=columns\_to\_read\_2)

df\_base=pd.concat([df\_temp,df\_base],ignore\_index=True)

**Download CSV files (not working, downloaded CSV files are empty)**

'''

Purpose:

(1) Download CSV files from a website https://github.com/CSSEGISandData/COVID-19/tree/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports.

CSV files are uploaded daily and consistently named as mm-dd-yyyy.csv

How to run this file in ipython?

%run D:/googleDrive/python/scripts/download-daily-files\_GitHub-COVID-19-csse-covid-19-daily-reports-CSV.py

'''

# --------------------

import requests

import shutil

import datetime

import os

folder\_path\_script='D:/googleDrive/python/scripts/'

folder\_path\_download='D:/download/GitHub/CSSEGISandData/COVID-19/csse-covid-19-daily-reports-CSV/'

# os.makedirs('dir-to-create',exist\_ok=True) enables the mkdir -p functionality [mkdir -p functionality in Python [duplicate]](https://stackoverflow.com/questions/600268/mkdir-p-functionality-in-python)

os.makedirs(folder\_path\_download, exist\_ok=True)

# -----------------------------------------------------------------------------------

dates=[datetime.datetime(2020,1,22)+datetime.timedelta(dval) for dval in range(0,366)];

# -----------------------------------------------------------------------------------

# Download daily added CSV files from https://github.com/CSSEGISandData/COVID-19/tree/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports

for dateval in dates:

r= requests.get('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/'+ dateval.strftime('%m-%d-%Y')+'.csv')

if r.status\_code == 200:

with open(dateval.strftime('%m-%d-%Y')+".csv", 'wb') as f:

r.raw.decode\_content = True

# shutil.copyfileobj(fsrc, fdst[, length])

## fsrc: A file-like object representing the source file to be copied

## fdst: A file-like object representing the destination file, where fsrc will be copied.

os.chdir(folder\_path\_download)

shutil.copyfileobj(r.raw, f)

# ---------------------------------------------------------------------------------

**Import a CSV file into a DataFrame (working)**

import pandas as pd

url= 'https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/csse\_covid\_19\_daily\_reports/01-22-2020.csv'

df= pd.read\_csv(url)

Published on [**HackMD**](https://hackmd.io/)

 195

[Like](https://hackmd.io/@Chang/use-publicly-available-data) [Bookmark](https://hackmd.io/@Chang/use-publicly-available-data) [Subscribe](https://hackmd.io/@Chang/use-publicly-available-data)