#IIIT DHARWAD

#COVID TRACKER 1

import pandas as pd

import matplotlib.pyplot as plt

import sys

import requests

from bs4 import BeautifulSoup

#creating variables for urls of the datasets and parsing the necessary information

confirmed\_url = 'https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series/time\_series\_covid19\_confirmed\_global.csv'

death\_url = 'https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse\_covid\_19\_data/csse\_covid\_19\_time\_series/time\_series\_covid19\_deaths\_global.csv'

r = requests.get("https://en.wikipedia.org/wiki/Developed\_country#Rankings")

intern = list()

soup = BeautifulSoup(r.text, "html.parser")

table = soup.find('table',{'class':'multicol'})

table\_1 = table.find\_all('a')

for i in table\_1:

    intern.append(i.get('title'))

intern\_countries = [i for i in intern if i]

del intern

#reading the csv file and shaping it into a 2D dataframe

df\_c = pd.read\_csv(confirmed\_url, error\_bad\_lines=False)

df\_d = pd.read\_csv(death\_url, error\_bad\_lines=False)

#removing the province column

df\_c = df\_c.loc[:,'Country/Region':]

df\_d = df\_d.loc[:,'Country/Region':]

#summing the values, so that theres only one row for each country

df\_c = df\_c.groupby('Country/Region').sum()

df\_d = df\_d.groupby('Country/Region').sum()

#making countries as index and all the dates as columns

df\_c = df\_c.loc[:,'1/22/20':]

df\_d = df\_d.loc[:,'1/22/20':]

#writing all the country names into a list

countries = list(df\_c.index)

final\_countries = list(set(intern\_countries) & set(countries))

del intern\_countries

while (1):

    print("\nEnter 1 to display graphs")

    print("Enter 2 to display the top twenty high-risk countries")

    print("Enter 3 to exit\n")

    num = int(input())

    if (num==1):

        print("\nEnter the country for which the graph is to be displayed\n")

        c = str(input())

        dates = list(df\_c.columns.values.tolist())

        death\_growth\_per\_day = list()

        if c in countries:

            for i,j in zip(df\_c,df\_d):

                temp = float((df\_d.loc[c][j]/df\_c.loc[c][i])\*100)

                death\_growth\_per\_day.append(temp)

            plt.plot(dates,death\_growth\_per\_day)

            plt.title(c)

            plt.show()

        else:

            print("\nThe country name isn't valid")

    if (num==2):

        high\_risk\_countries = list()

        for i in final\_countries:

            for j,x in zip(df\_c,df\_d):

                if (df\_d.loc[i][x]==0) | (df\_c.loc[i][j]==0):

                    death\_rate = 0

                elif df\_c.loc[i][j]!=0:

                    death\_rate = (df\_d.loc[i][x]/df\_c.loc[i][j])\*100

                else:

                    death\_rate = 0

                two\_percent = (2/100)\*df\_c.loc[i][j]

                if (death\_rate!=0) & (death\_rate>=two\_percent):

                    if i in high\_risk\_countries:

                        break

                    else:

                        high\_risk\_countries.append(i)

        print("\nTop high risk countries to travel to for internships are:\n")

        for i in high\_risk\_countries:

            print(i)

    if (num==3):

        sys.exit(1)