Carnegie Mellon University

FINAL PROJECT

ABSTRACT & USER INSTRUCTIONS

COVIDaily

90-819 Intermediate Python

Group Name: Snake Coders

Fabio Beltran

Muhammad Bin Oiad

Lucas Falivene

COVIDaily Abstract

COVIDaily application gathers, processes, displays, and stores real time COVID-19 information. The application obtains the data from the following resources: Johns Hopkins Coronavirus Resource Center's API, Pennsylvania Department of Health and Centers for Disease Control and Prevention (CDC) websites, and Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) CSV datasets. The data is collected using URL request libraries and scraping techniques. Then, the data is processed using pandas dataframes, numpy and regular expressions libraries. Finally, SQLite3 and matplot libraries are used to display, maintain, and save the processed data.

User Instructions

The program does not require any library installation, all the required libraries are managed by the execution of the code. The user will only be required to have all 4 python files that compose the COVIDaily application in the same directory. The python files are:

1. Main COVIDaily function .py file
2. API .py file
3. Web scraping .py file
4. CSV data processing .py file

Note: CSV datasets must be collected beforehand in case there is no internet access

Main Menu & Application Start Up

To start the program, the user should use: python COVIDaily

The program starts by presenting a menu with 3 options. The user has to enter the number corresponding to the desired option (Refer to Figure 1). Below is a brief description of each option:

1. Using API, this option gets updated statistical information for any county and allows to compare data between two counties in any state in the US.
2. Using web scraping, this option retrieves frequently asked questions about COVID-19 from the Pennsylvania Department of Health and CDC. The program scraps both websites so that the user can get updated information, which might be more useful when restrictive measures start relaxing.
3. Using datasets, this option retrieves historic series data from all countries around the world. The user can use the data to generate up-to-date graphs for any country's active COVID-19 cases and store those graphs. In addition, the user can download the data output into new datasets or SQLite database.

Figure 1. Main Menu

1. User Instructions: API Menu

If the user selected option 1 in the main menu, the program would present a second menu (Refer to figure 2) where:

1. This will present a quick statistical summary of COVID-19 in the world.
2. This will present a statistical summary for a desired US county and corresponding state.
3. This will present statistical summaries comparisons in two US counties.

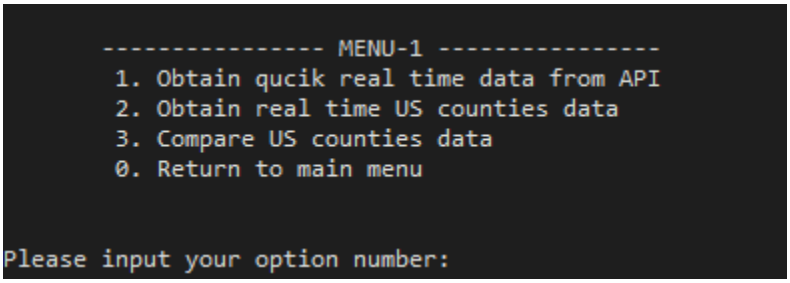

Figure 2: Option 1 menu. Updated information retrieved via API

1.1 API Menu (Option 1)

This option will show the end user a statistical quick summary of the global data trend related to COVID-19 that is obtained from the API.

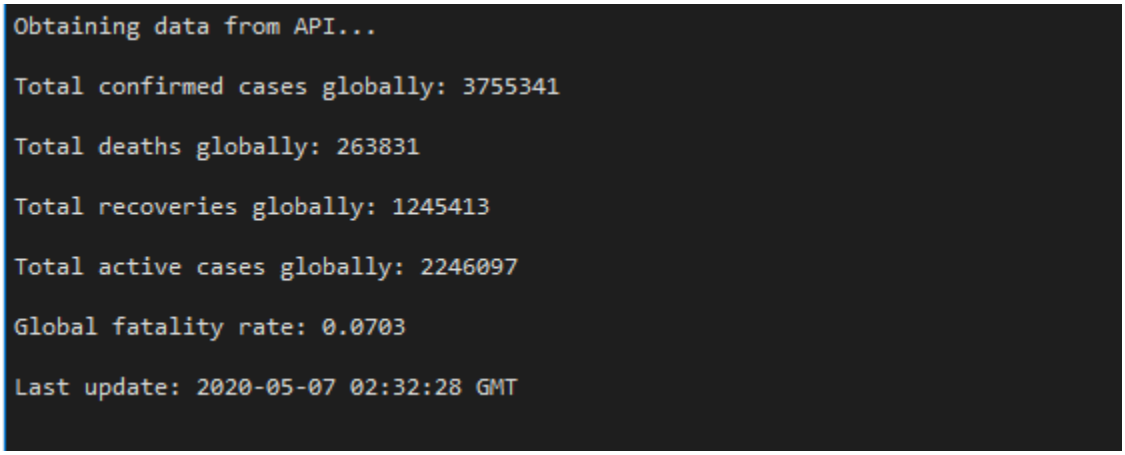

Figure 3: Option 1 menu 1. Quick global information retrieved via API

1.2 API Menu (Option 2)

This option will show the confirmed cases and deaths related to COVID-19 for any US county and state pair. The user can choose to get the latest data (yesterday's data upload) or get historical data for a certain custom date in time.

```
Enter the state where the county is located: Pennsylvania

Enter the US county you want to obatin data from: allegheny

Do you want to obtain yesterday's data?

Input yes or no: yes

Obtaining Allegheny data from API...

COVIDaily for Allegheny in the state of Pennsylvania
Confirmed cases in Allegheny: 1394
Deaths in Allegheny: 111
Last updated: 2020-05-07 02:32:28 GMT

State information
Confirmed cases in Pennsylvania: 54800
Deaths in Pennsylvania: 3345
Last updated: 2020-05-07 02:32:28 GMT
```

Figure 4: Option 2 menu 1. Us county and state information retrieved via API

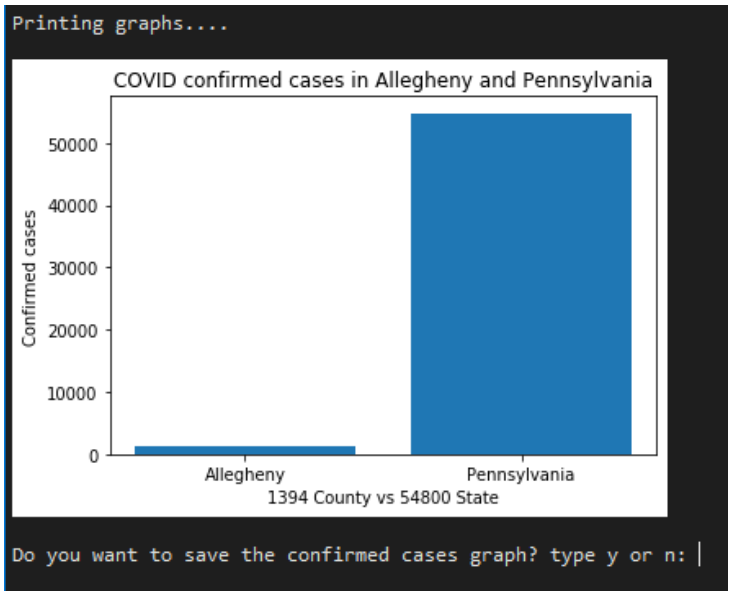

Figure 5: Option 2 menu 1. Us county and state information retrieved via API

The user can select to only visualize the graphs or also store them in the current running directory. Moreover, the findings of the API query can be also stored in a database.

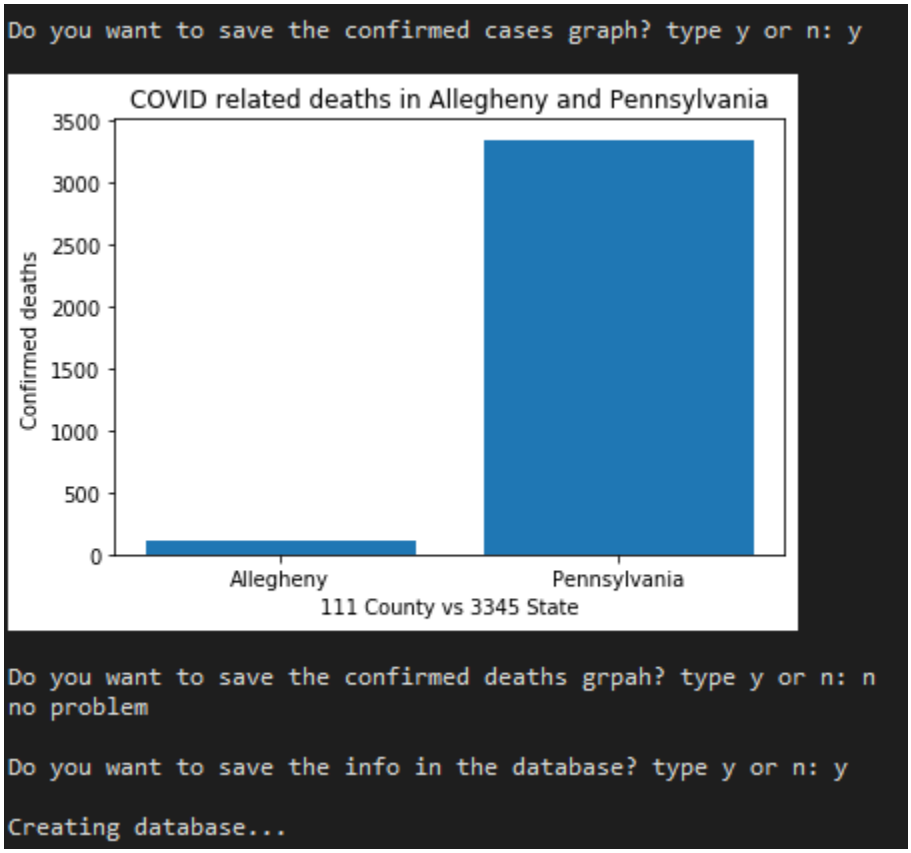

Figure 6: Option 2 menu 1. Us county and state information retrieved via API, database storage

There is also an option for the user to store all the data query to the API in a sqlite database.

1.3 API Menu (Option 3)

In this option, the end user will be able to select two different US counties in order to do a data comparison between the confirmed cases and death toll numbers related to COVID-19. The user must input two pairs of US county and state in order for the API to retrieve the information (as seen in Figure 7).

```
Enter the state where the county is located: Pennsylvania

Enter the first US county you want to compare: Allegheny

Enter the state where the county is located: Pennsylvania

Enter the second US county you want to compare: Philadelphia

Obtaining Allegheny data from API...

COVIDaily for Allegheny in the state of Pennsylvania
Confirmed cases in Allegheny: 1394
Deaths in Allegheny: 111
Last updated: 2020-05-07 02:32:28 GMT

State information
Confirmed cases in Pennsylvania: 54800
Deaths in Pennsylvania: 3345
Last updated: 2020-05-07 02:32:28 GMT
```

```
COVIDaily for Philadelphia in the state of Pennsylvania
Confirmed cases in Philadelphia: 16697
Deaths in Philadelphia: 803
Last updated: 2020-05-07 02:32:28 GMT

State information
Confirmed cases in Pennsylvania: 54800
Deaths in Pennsylvania: 3345
Last updated: 2020-05-07 02:32:28 GMT

Printing graphs....
```

Figure 7: Option 3 menu 1. US county information comparison retrieved via API.

The user will be prompted with two comparison graphs - one for confirmed cases and the other for the death toll - which he can choose or not to store as a file. An example of the death toll graph comparison is displayed in Figure 8.

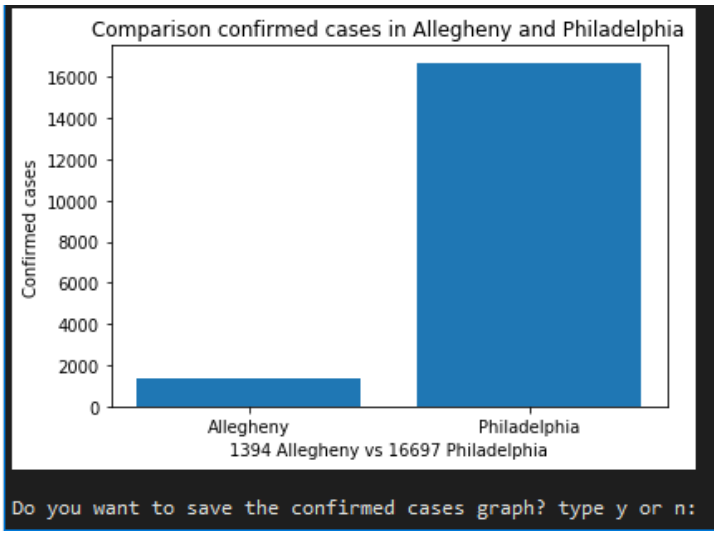

Figure 8: Option 3 menu 1. Us county information comparison retrieved via API.

User Instructions: Web Scraping Menu

This option will present a menu so that the end user selects to see Frequently Asked Questions from PA Department of Health or CDC site about COVID-19. This information will be scraped when option 1 is selected from the main menu.

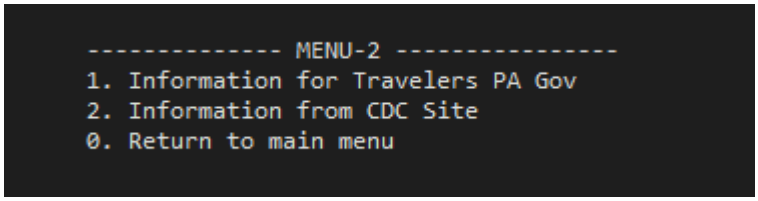

Figure 9: Menu 2. Information for travelers retrieved via Scraping from PA Department of Health.

Web Scraping Sub-Menu (Option 1)

For Frequently Asked Questions for Travelers, the program will show the available options from the PA Department of Health. The user has to select a valid option.

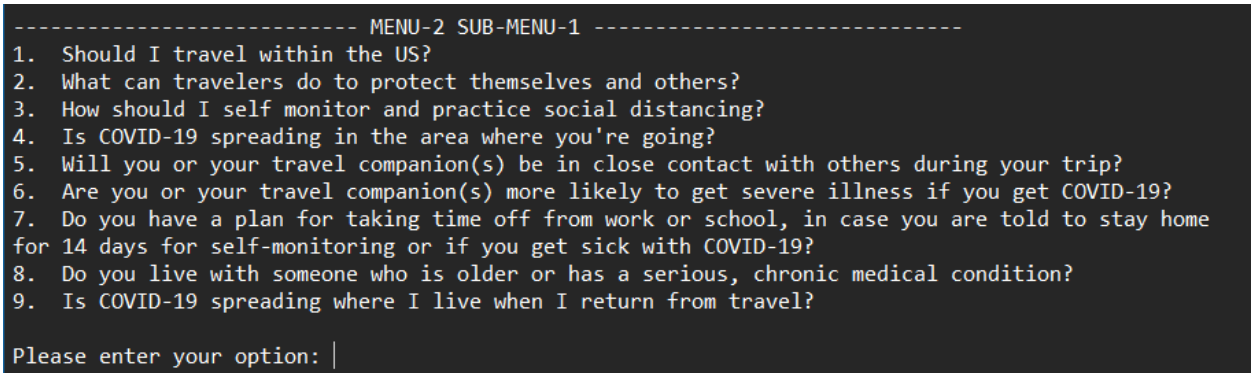

Figure 10: FAQ from PA Department of Health site

Once the user selects a valid option, the program will show the answer to the selected question (Refer to Figure 11).

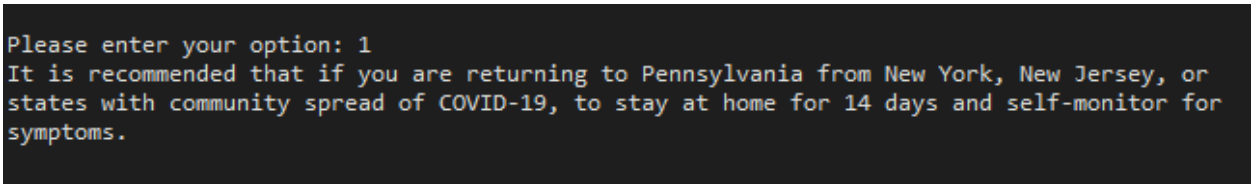

Figure 11: Answer for the selected option from PA Department of Health site

Web Scraping Menu (Option 2)

If the user selected Information from the CDC website (figure 9), the program will show a sub-menu (Figure 12) with the available FAQ's sections from the CDC site. Each section has its own list of questions. Thus, when the user selects each of the following options, the program will present the list of questions available for the selected option.

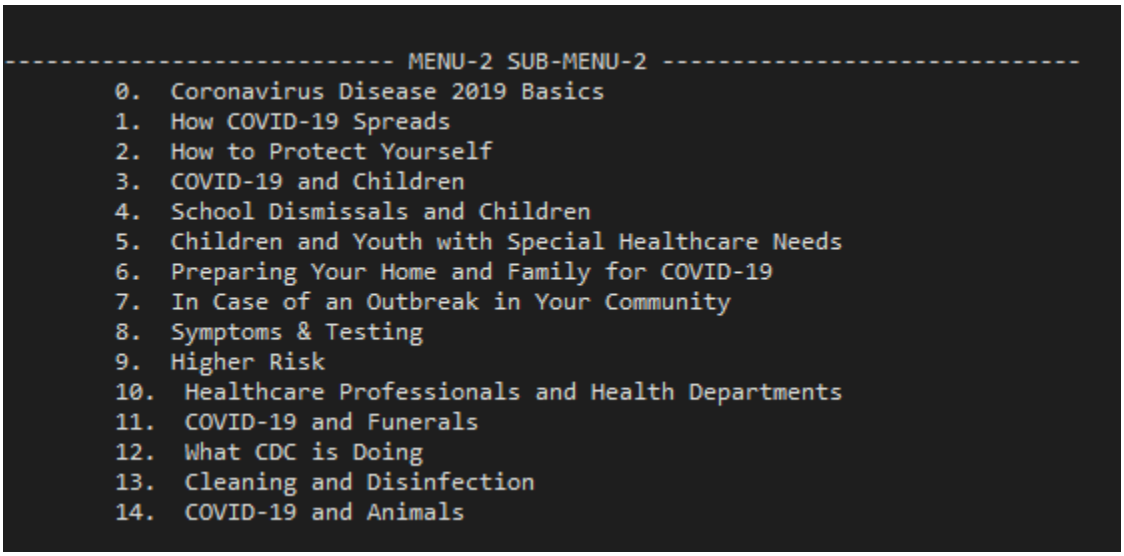

Figure 12: Answer for the selected option from CDC site

3. User Instructions: CSV Datasets Menu:

If the user selected option 3 in the main menu, the program would present a second menu (Refer to figure 13 below) where:

1. This will present a timeline plot for the active COVID-19 cases in a country
2. Save a SQLite database for the statistics of active COVID-19 cases in all countries worldwide
3. Save a CSV dataset for the statistics of active COVID-19 cases in all countries worldwide

```
---------------------------------------- MENU-3 ----------------------------------------
1. Get active COVID-19 cases in a specific country with timeline plot
2. Create a Database of the confirmed/deaths/recoveries/active cases in all countires
3. Create a dataset of the confirmed/deaths/recoveries/active cases in all countires
0. Return to main menu
```

Figure 13: Dataset Menu

3.1 CSV Menu (Option 1):

In this option, the user will be prompted to enter a specific country name, today's total number and the graph of the active COVID-19 cases since January 22nd (dataset start time) will be displayed as seen in the figure 14 below:

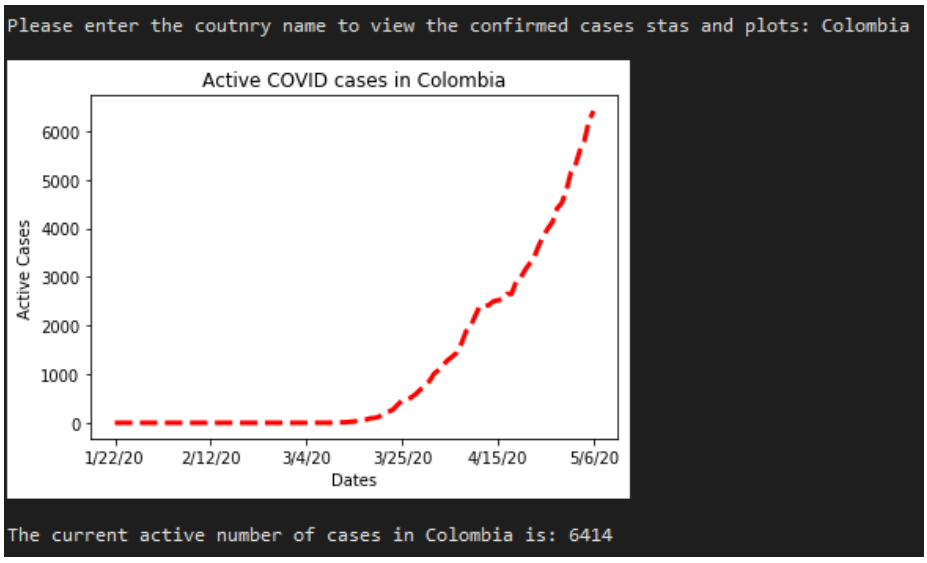

Figure 14: Active COVID-19 in Colombia

3.2 CSV Menu (Option 2):

In this option, the user will be able to create/update a database that has all the up-to-date information about the active COVID-19 cases in the world. Figure 15 below shows a snapshot of the new dataset:

	country	confirmed	deaths	recovered	active
0	Afghanista	2335	68	310	1957
1	Albania	782	31	488	263
2	Algeria	4154	453	1821	1880
3	Andorra	745	43	468	234
4	Angola	30	2	11	17
5	Antigua an	25	3	15	7
6	Argentina	4532	225	1292	3015
7	Armenia	2148	33	977	1138
8	Australia	6778	93	5775	910

Figure 15: Active COVID-19 in the world (dataset)

3.3 CSV Menu (Option 3):

In this option, the user will be able to create/update a database that has all the up-to-date information about the active COVID-19 cases in the world. Figure 16 below shows a snapshot of the new database:

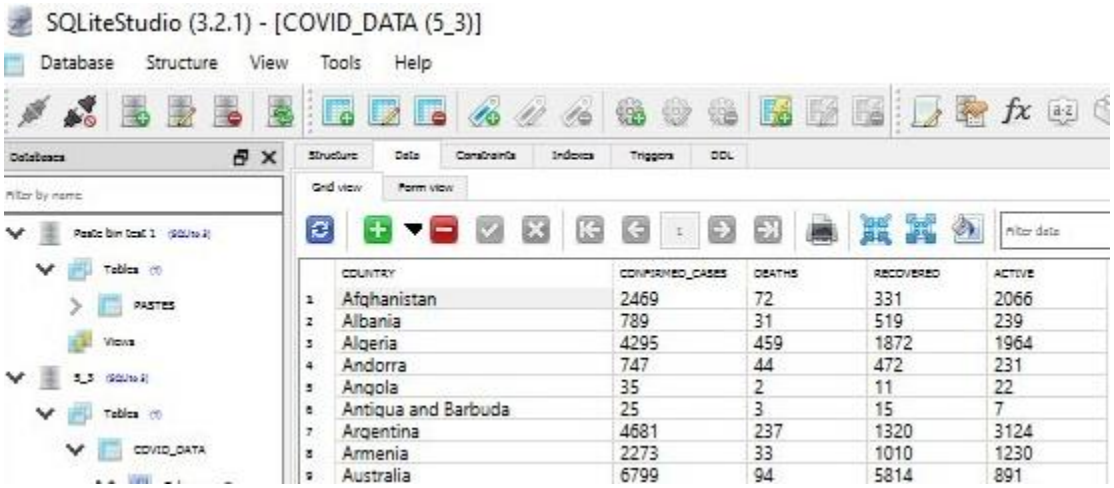

Figure 16: Active COVID-19 in the world (SQLite)