An econometric analysis of the COVID-19 pandemic

on the Caribbean tourism industry

CSEC 491: Senior Project Report

Andrei Pascu

Advisor: Dr. Kim Blenman

May 4, 2023

Table of Contents

[Section I. Introduction 3](#_Toc133253713)

[Background 3](#_Toc133253714)

[Project Description 3](#_Toc133253715)

[Section II. Development Environment 5](#_Toc133253716)

[Project Resources 5](#_Toc133253717)

[Software Installation 6](#_Toc133253718)

[Downloading the Project Files 7](#_Toc133253719)

[Walkthrough of the Repository 9](#_Toc133253720)

[COVID-19 Data 9](#_Toc133253721)

[Tourism Data 14](#_Toc133253722)

[Analysis Code 16](#_Toc133253723)

# Section I. Introduction

## Background

The COVID-19 pandemic had a profound and long-lasting impact on the economies of the world. One of the industries that was substantially affected by the pandemic is the tourism sector. The number of flights or cruises to tourist destinations plummeted. The introduction of quarantine restrictions aggressively limited incoming visitors and prohibited many recreational activities. A large proportion of jobs in tourism were put on hold and in many cases were terminated. As highlighted by the International Monetary Fund, tourism-dependent regions were most notably and severely impacted by the COVID-19 outbreak and “will likely feel the negative impacts of the crisis for much longer than other economies[[1]](#footnote-1).” One such case study is that of the Caribbean countries, which have a high economic reliance on tourism and thereby have been especially susceptible to the negative financial consequences of the COVID-19 outbreak. Therefore, studying the impact of the COVID-19 pandemic on the Caribbean economy represents an opportunity to analyze the immediate and long-term effects of a health crisis on similar tourism-dependent regions and may help shape predictions about how future pandemic events will affect tourism.

## Project Description

This project seeks to investigate the impact of the COVID-19 pandemic on the tourism industry in the context of the Caribbean region. This senior project is divided into two main parts—an econometric-driven analysis and a software development component—that fulfills the Computer Science and Economics requirements of the CSEC 491 senior thesis course. First, through the investigation of the real-world data, we wish to quantitatively identify the economic effects of the COVID-19 outbreak on the tourist-dependent Caribbean economy. We analyze key economic metrics pertinent to tourism, such as number of visitors, changes in revenue, duration (e.g., overnight; same-day) and mode of travel (e.g., flight; cruise) that we expect to be timely correlated with the onset of the COVID-19 pandemic. Finally, we attempt to estimate the direct impact of COVID-19 cases, deaths and mortality on the Caribbean economy in order to better understand the financial impact of the COVID-19 pandemic on the tourism industry.

Second, using the results from the previous section, we wish to formulate a predictive estimate for future changes in key economic variables associated with COVID-19 rates, leveraging the results found in the first part of this project. The objective is to generalize our findings of the impact of COVID-19 on the tourism-dependent Caribbean economy to other countries that have a similarly heavy reliance on tourism for the growth and generation of tourism volume and revenue. We wish to present this model and other economic insights by building software data visualizations and plots through the use of software programming in order to provide a clear understanding of the impact of a health crisis on the tourism sector.



Figure : Map of the Caribbean countries

[TBD: Add list of Caribbean countries]

[TBD: Add section overview]

[TBD: Add acknowledgement]

[TBD: Review introduction]

# Section II. Development Environment

## Project Resources

The following applications and software tools were used for the development and completion of this project:

* JupyterLab[[2]](#footnote-2): web-based interactive development environment for running Python
* Git: recommended local version control system for keeping track of updating files
* GitHub: web-based version control system used for publishing project
* Visual Studio Code: preferred IDE used for editing and managing the project files

This project was developed on a Unix system and coded primarily in Python 3 (the latest version of Python at the time of this document is recommended) with some programming scripts written in Bash. The following command-line tools were used in the completion of this senior thesis:

* Homebrew[[3]](#footnote-3): package manager for installing command-line and software utilities
* pip: Python package installers

Additionally, the software and programming libraries that were used for the analysis of this project are as follows:

* Matplotlib[[4]](#footnote-4): data visualization library for creating plots
* NumPy[[5]](#footnote-5): mathematical package for data computing
* pandas[[6]](#footnote-6): data analysis and manipulation library
* statsmodels[[7]](#footnote-7): statistical modelling and regression module
* linearmodels[[8]](#footnote-8): complementary library to statsmodels (i.e., fixed-effects regression)
* seaborn[[9]](#footnote-9): data visualization package based on Matplotlib

## Software Installation

*Note: This software installation section is written for a Unix-based system, such as MacOS or Linux, on which this project was developed on.*

First, installing Homebrew is recommended for the installation of all necessary software tools; run the following command to install:

|  |
| --- |
| **$ /bin/bash -c "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"** |

Using Homebrew, install all the necessary software components for this project including Python 3, JupyterLab, Git and the Python package manager:

|  |
| --- |
| **$ brew install python3**  **$ brew install jupyterlab**  **$ brew install git** |

Using Python’s pip package manager, install the latest versions of the remaining Python libraries by running the following set of commands:

|  |
| --- |
| **$ pip install matplotlib**  **$ pip install numpy**  **$ pip install pandas**  **$ pip install statsmodels**  **$ pip install linearmodels**  **$ pip install seaborn** |

More information on these software libraries and packages and their full set of functionalities are available on their respective online websites.

## Downloading the Project Files

All of the project files are hosted on an online GitHub repository available for public viewing and download at <https://github.com/andreiui/csec491>. In order to clone the repository, run the following command:

|  |
| --- |
| **$ git clone https://github.com/andreiui/csec491** |

This command will download all of the project files onto your local machine into a directory named “csec491” by default; alternatively, users can download the compressed ZIP file directly from the GitHub link.

After cloning the online repository or downloading the files onto your machine, open the project folder using the software editor of the user’s choice, although Virtual Studio Code is recommended; Figure 1 shows an example of the project files being loaded into the IDE.

The repository is structured into four main folders with task-specific subdirectories:

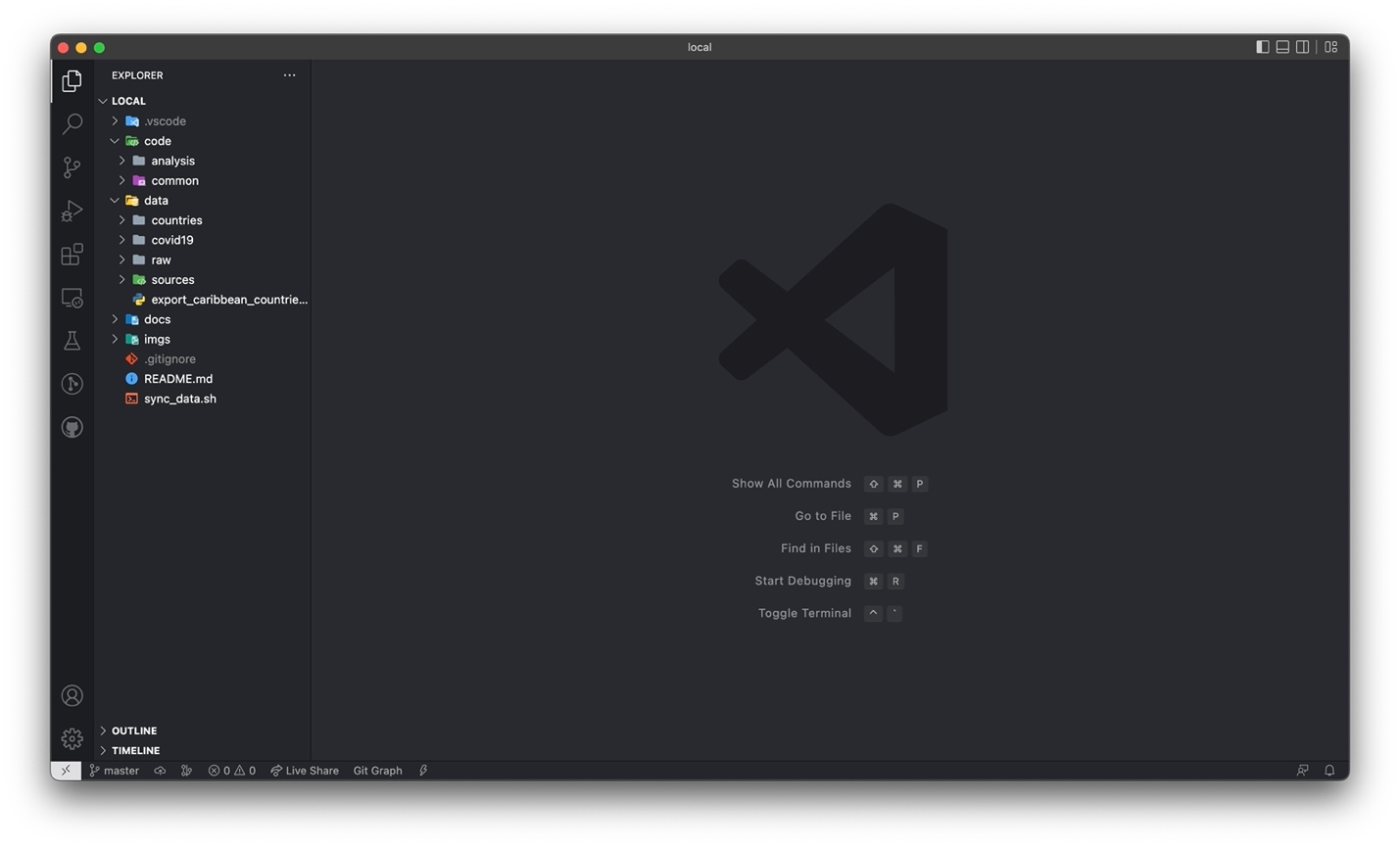


Figure 2: Example of project repository loaded in the Virtual Studio Code IDE

* “code”: contains all project code files for data analysis
  + “code/analysis”: has all of the data analysis code of the Caribbean region using Python and JupyterLab, subdivided into folders named for the corresponding Caribbean country
  + “code/common”: includes a Python file with commonly reused code and functions across analyses
* “data”: contains data-specific files, such as scripts for importing and exporting data and subfolders with raw and processed data files
  + “data/countries”: has all per-country data sorted into Excel files with tourism and COVID-19 numbers
  + “data/covid19”: includes all processed COVID-19 data for all Caribbean countries, divided into files storing daily, monthly and yearly rates
  + “data/raw”: contains all raw COVID-19 data for countries around the world extracted from external online repositories

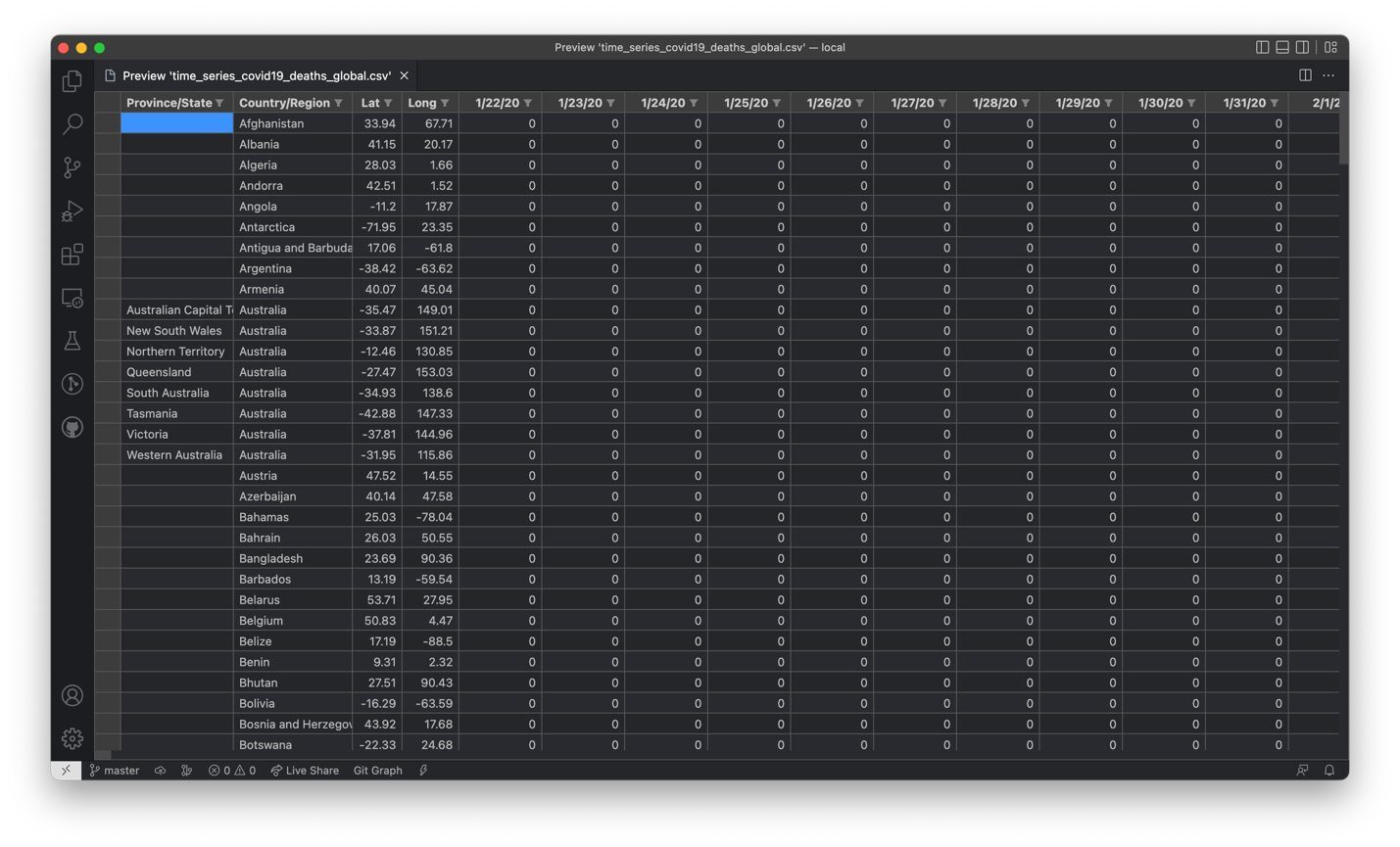


Figure 3: Raw time series of daily COVID-19 deaths data for global countries

* “docs”: has all documents relating to the development of this senior project, such as the initial senior thesis proposal and final project report
* “imgs”: includes all manually curated and generated images for this project

## Walkthrough of the Repository

In this part, we provide a comprehensive walkthrough of the repository and data-generation steps for ease of use by any user who wishes to expand on and use this project as a template. The steps outlined below serve as an informational and step-by-step guide for generating the Caribbean COVID-19 and tourism data for the scope of this project; however, these files have already been created and included in the project repository for ease and readiness of use.

Each code file has been well-commented to assist in readability and ease of understanding of the structure and functionality of the program. Additionally, subdirectories often include sources, either in Markdown, Excel or HTML format, outlining in great detail where each external piece of data was collected from that contributed to the fulfillment of this project.

### COVID-19 Data

In the “data/raw” directory, there are four CSV files containing raw unformatted information on the number of COVID-19 cases and deaths recorded in the U.S. and countries around the world and are descriptively titled as follows:

* “time\_series\_covid19\_confirmed\_global.csv”
* “time\_series\_covid19\_confirmed\_US.csv”
* “time\_series\_covid19\_deaths\_global.csv”
* “time\_series\_covid19\_deaths\_US.csv”

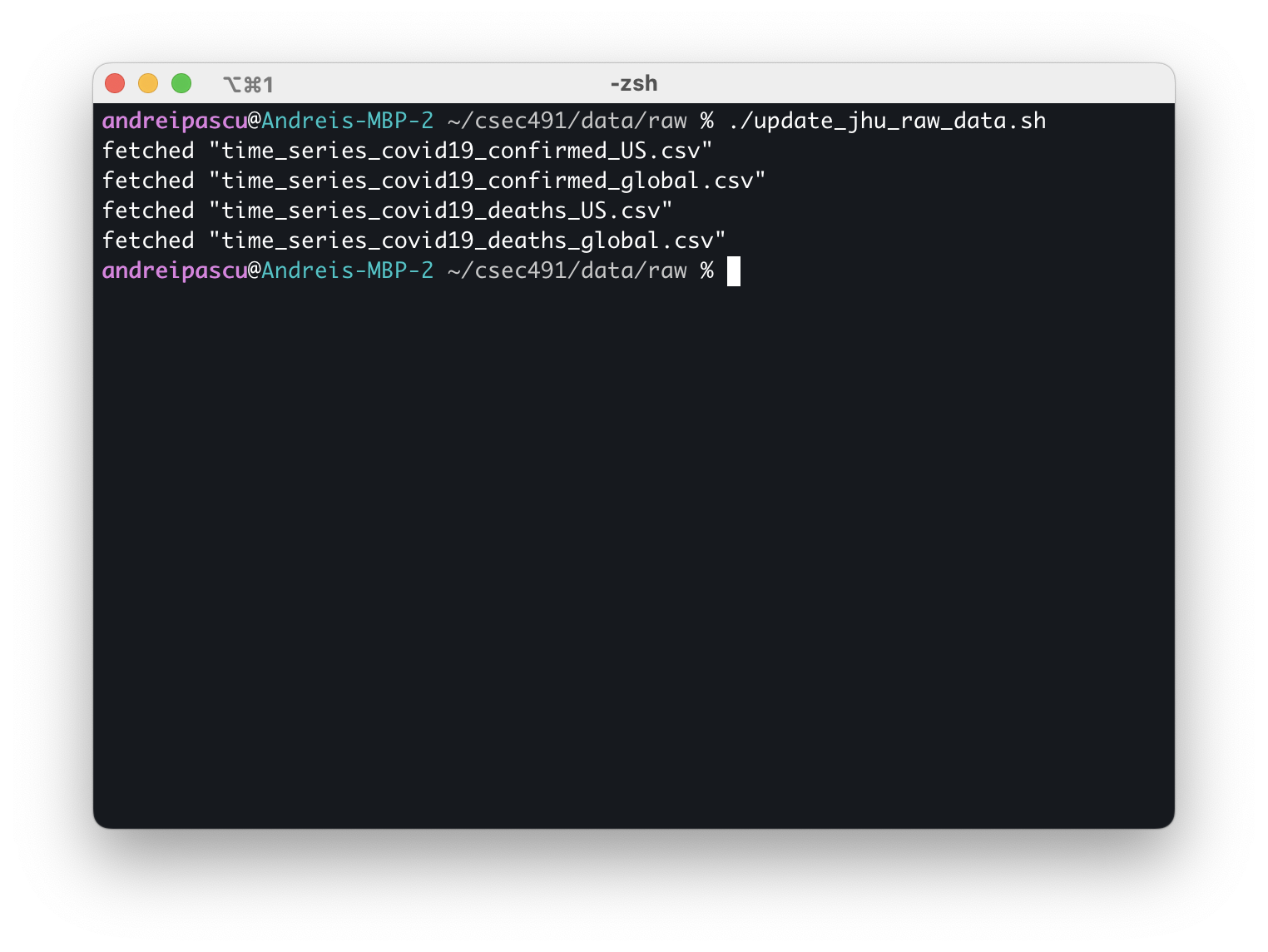


Figure : Sample run of the “update\_jhu\_raw\_data.sh” Bash script

These files have been downloaded from the Johns Hopkins University online GitHub repository on COVID-19 data made available by the Center for Systems Science and Engineering (CSSE), which include cumulative daily COVID-19 rates starting on January 20, 2022. The repository is made public and available at <https://github.com/CSSEGISandData/COVID-19>. Figure 2 illustrates the structure of the raw file for the time series of COVID-19 global cases as an example.

The “update\_jhu\_raw\_data.sh” file is a Bash script that automatically downloads and updates these time series files contained in the “csse\_covid\_19\_data/csse\_covid\_19\_time\_series” subdirectory on JHU’s GitHub (last download snapshot occurred on March 1, 2023 at 12:30 PM EST). Figure 3 presents a sample run and output of this executable file.

In the parent “data” folder, the “export\_caribbean\_countries.py” Python file parses through the raw COVID-19 data and extracts the relevant information relevant for the Caribbean region; the list of countries selected is identical to that provided in Section I. The program reads from the “data/raw” subdirectory and parses all of the countries, porting each of the Caribbean countries to two new CSV files—for the incidence of cases and deaths. During its execution, the script outputs the start and end of processing each raw file and signals when it has finished generating the new files in the “data/covid19” folder; additionally, it will output any data discrepancies such as misaligned dates across the raw files and output the full count and set of countries that were processed. Figure 4 shows an example of running the Python script.

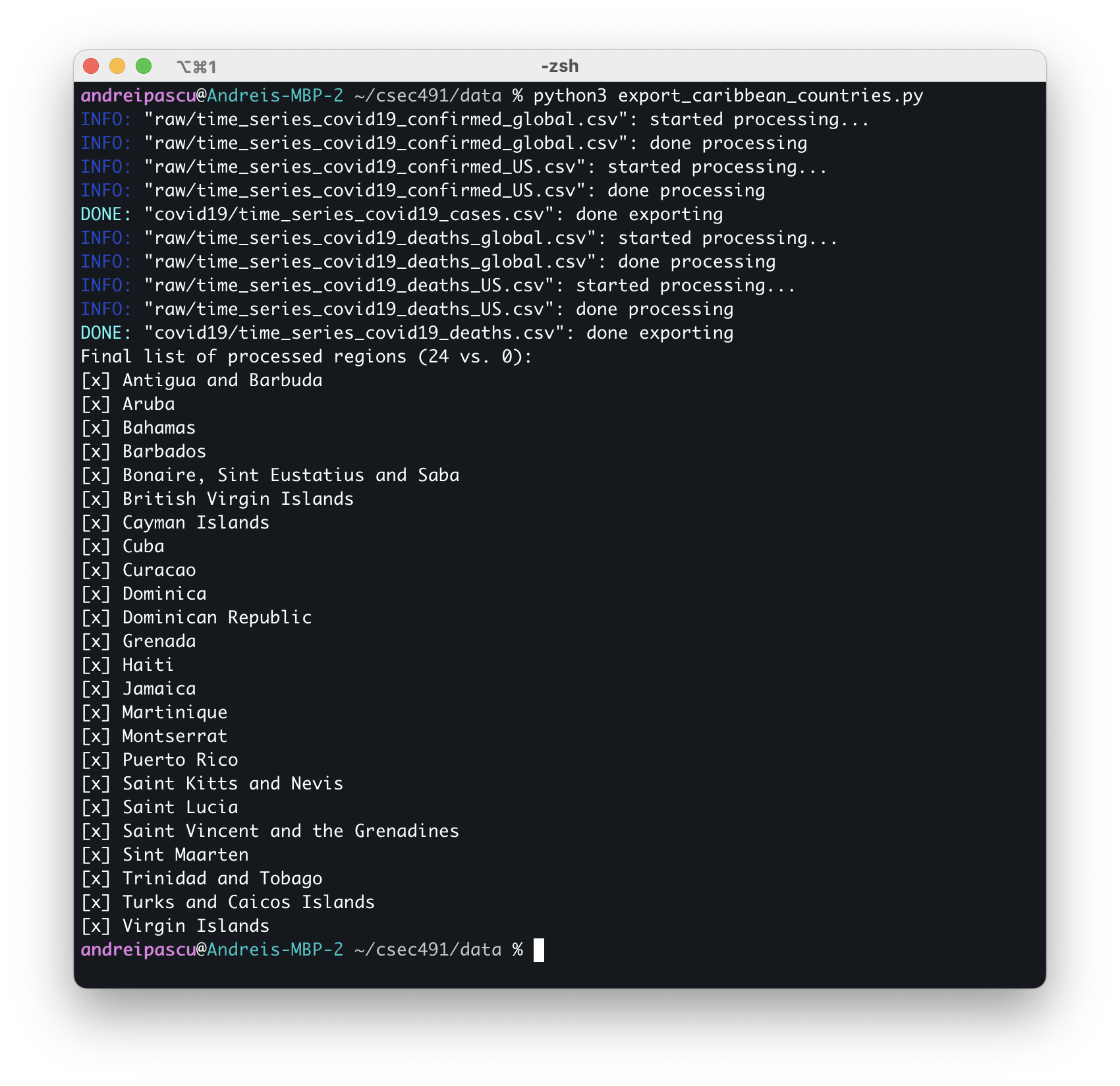


Figure : Sample run of the “export\_caribbean\_countries.py” Python file

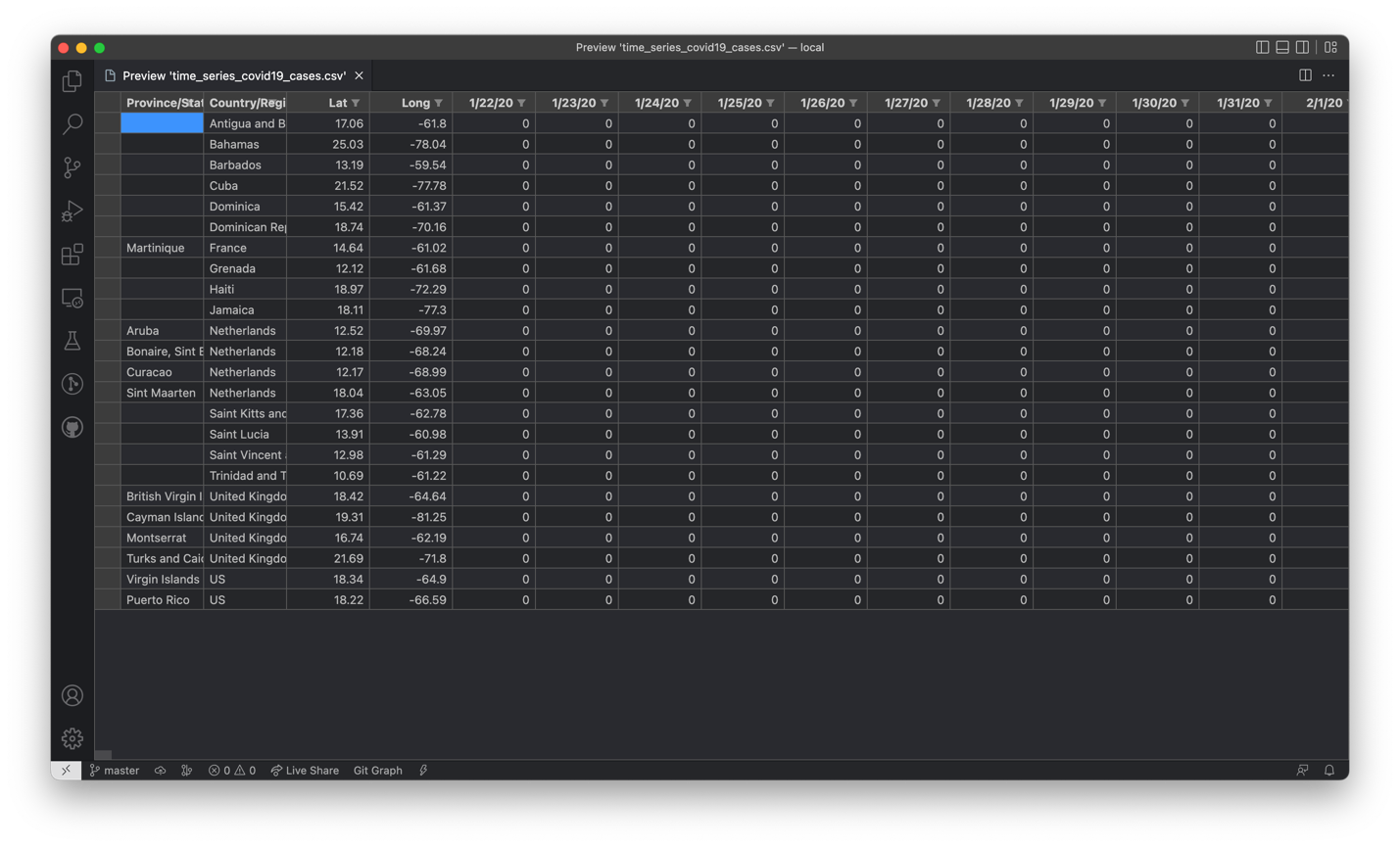


Figure 6: Generated COVID-19 cases for the Caribbean countries from the raw JHU data

The “data/covid19” subdirectory contains the generated files from the raw COVID-19 data after running the “export\_caribbean\_countries.py” Python script: “time\_series\_covid19\_cases.csv” and “time\_series\_covid19\_deaths.csv;” Figure 5 illustrates a preview of the file generated for COVID-19 cases pertinent to the Caribbeans. The numbers in these files correspond to daily cumulative rates; however, for the scope of this project, monthly and yearly rates are preferred for statistical analysis. The “generate\_from\_daily.py” Python file performs this conversion, transforming Caribbean cumulative daily data into monthly and yearly equivalents in four new CSV files:

* “cases\_yearly.csv”
* “cases\_monthly.csv”
* “deaths\_yearly.csv”
* “deaths\_monthly.csv”

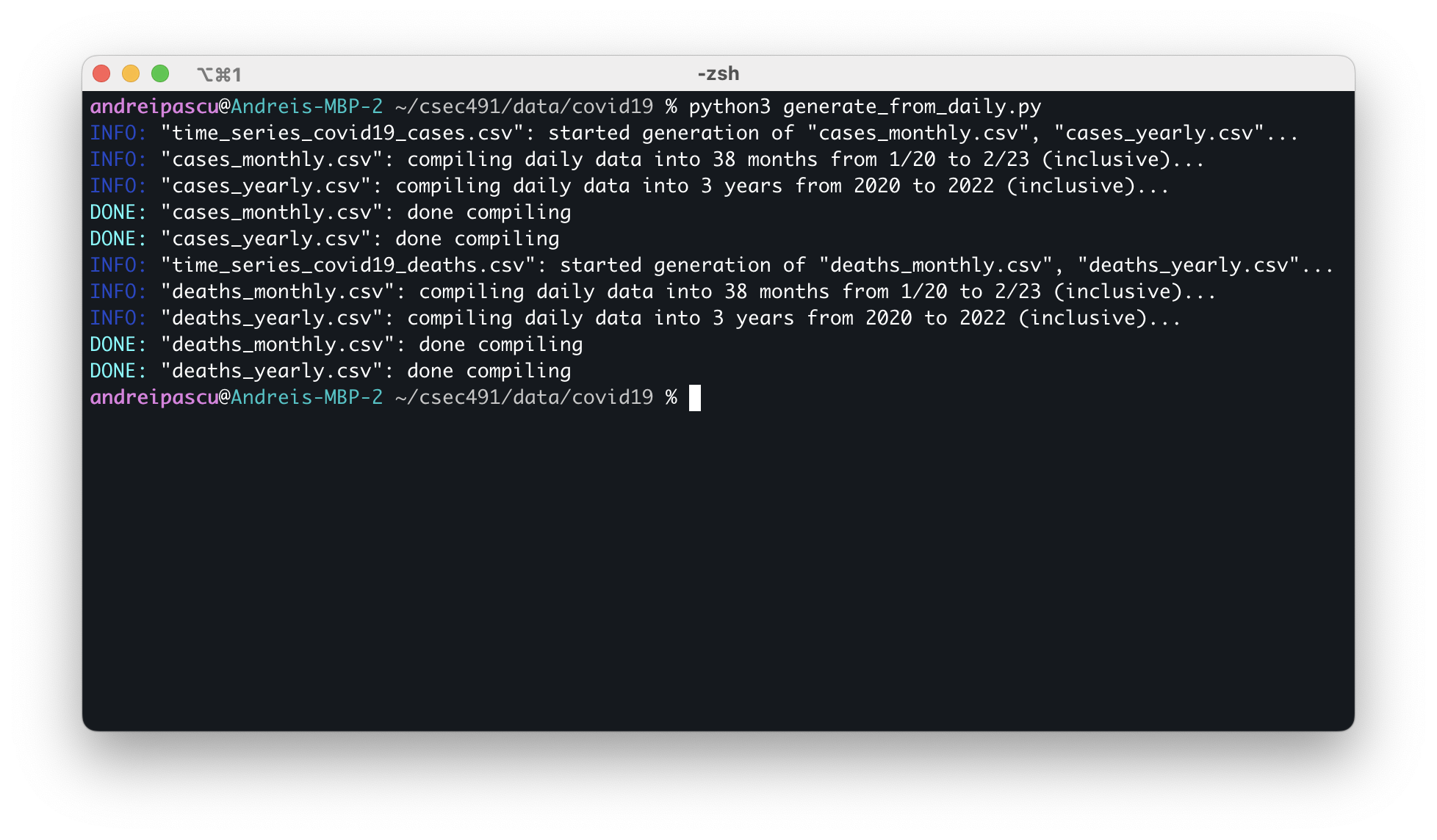


Figure 7: Sample run of the “generate\_from\_daily.py” Python file

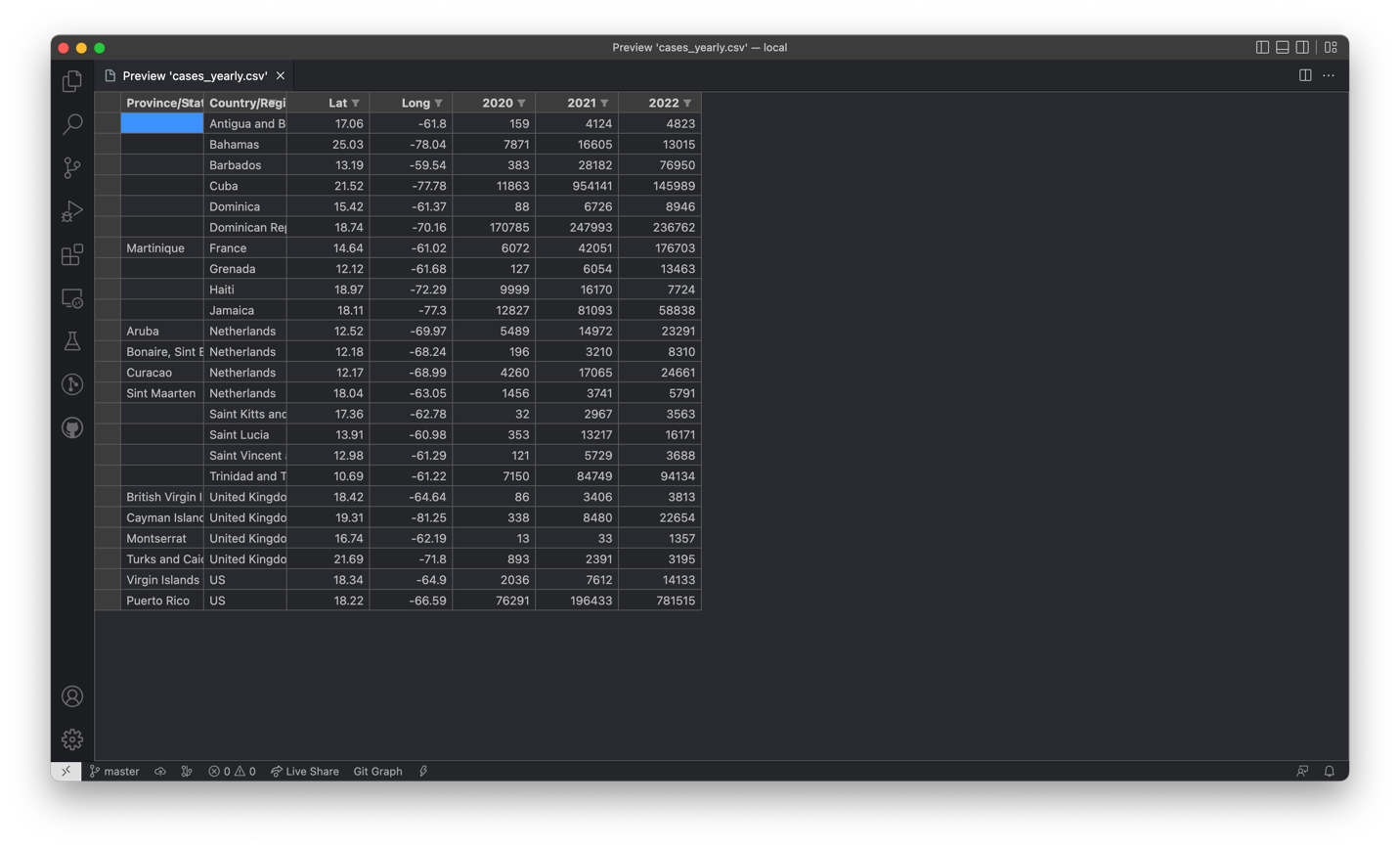


Figure : Generated COVID-19 yearly cases for Caribbean countries

Similar to the previous Python script, “generate\_from\_daily.py” outputs debug messages on the execution status of the program, informing the user of the start and completion of processing the data files. Figure 6 presents the output of the program, whereas Figure 7 illustrates the structure of one of the four generated files.

### Tourism Data

All of the Caribbean tourism data used for the scope of the statistical analysis for this project, such as number of tourists, expenditure and other economic variables of interest, have been compiled and manually curated from Statista[[10]](#footnote-10). The “data/sources” subdirectory contains a HTML table with online references to all of the various sources on the Caribbean tourism numbers; the data availability varies across each individual country with respect to economic variables and time periods. Some country-specific data may be missing for tourism volume, spending, overnight vs. same-day visitors, etc.; furthermore, data generally start from 2010 and frequently end in either 2020 and 2021 with some per-country exceptions.

The number of incoming tourists may be available in five categories: total volume of tourists, overnight or same-day visitors, and quantity of visitors arriving by cruise or air. Depending on the availability of this data, the total number of tourists may be derived. Additionally, the expenditure of incoming tourists is also segmented into total spending and monetary investment in travel or transportation to the region. Table 1 outlines the data availability on tourism for each Caribbean country in this project analysis.

The tourism data collected at the individual Caribbean country level presented in Table 1 is appended with its corresponding incidence of COVID-19 cases and deaths. In the subdirectory “data/countries,” all of the economic and pandemic data have been manually compiled on a per-country basis into Excel tables. Each of these contains an overview of the data with appropriate citations and online sources. Figure 8 and Figure 9 illustrate an example of a curated Excel file.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | Tourism Volume | | | | | | Tourism Expenditure | | | |
| Time-  frame | Total | Over-night | Same-day | Cruise | Air | Time-  frame | Total | Travel | Trans-port |
| Antigua & Barbuda | 2010–2021 | x |  |  |  |  | 2010–2021 | x |  |  |
| Aruba | 2010–2021 |  | x |  |  |  | 2010–2020 | x |  |  |
| Bahamas | 2010–2021 | x | x | x |  |  | 2010–2021 | x | x | x |
| Barbados | 2010–2020 | x | x | x |  |  | 2010–2020 |  | x |  |
| Bonaire, St. Eustatius & Saba | 2012–2021 |  |  |  |  | x | N/A |  |  |  |
| British Virgin Islands | 2010–2020 | x | x | x |  |  | 2010–2019 | x |  |  |
| Cayman Islands | 2010–2020 | x | x | x |  |  | 2010–2019 | x |  |  |
| Cuba | 2013–2021 | x |  |  |  |  | 2010–2021 | x | x | x |
| Curaçao | 2010–2020 | x | x\* | x |  |  | 2010–2021 | x | x | x |
| Dominica | 2010–2021 | x |  |  |  |  | 2010–2021 | x |  |  |
| Dominican Republic | 2010–2021 | x |  |  | x | x | 2010–2021 | x |  |  |
| Grenada | 2010–2022 | x |  |  |  |  | N/A |  |  |  |
| Haiti | 2010–2020 | x | x | x |  |  | N/A |  |  |  |
| Jamaica | 2010–2020 | x |  |  |  |  | 2010–2020 | x |  |  |
| Martinique | 2010–2021 | x | x | x |  |  | N/A |  |  |  |
| Montserrat | 2010–2021 | x |  |  |  |  | 2010–2021 | x |  |  |
| Puerto Rico | 2014–2020 |  | x |  |  |  | 2010–2019 | x |  |  |
| St. Kitts & Nevis | 2010–2021 | x |  |  |  |  | 2010–2021 | x |  |  |
| St. Lucia | 2010–2021 | x |  |  |  |  | 2010–2021 | x |  |  |
| St. Vincent & Grenadines | 2010–2021 | x |  |  |  |  | 2010–2021 | x |  |  |
| St. Maarten | 2012–2021 |  |  |  | x |  | N/A |  |  |  |
| Trinidad & Tobago | 2010–2020 | x | x | x |  |  | 2010–2021 | x | x | x |
| Turks & Caicos | 2010–2021 | x |  |  |  |  | N/A |  |  |  |
| U.S. Virgin Islands | 2010–2020 | x |  |  |  |  | N/A |  |  |  |
|  | Pct. % | 83.3 | 41.7 | 33.3 | 8.3 | 8.3 | Pct. % | 66.7 | 20.8 | 16.7 |

Table : Per-country data availability

(\* 2010–2021)

Furthermore, all of the COVID-19 and tourism data collected in the Excel files created for each Caribbean country is accumulated into the “\_CARIBBEAN.xlsx” file as a panel data for econometric analysis of the entire available Caribbean region, in order to identify averaging trends among the list of selected countries and utilize a larger dimension of data for more precise and informative regressions.

### Analysis Code

As previously described in the outline of the structure of the project files, the “code” folder contains all pertinent programs used to analyze the Caribbean data using Python, statistical regressions libraries in JupyterLab and display intuitive plots using Matplotlib. In order to access the generated data, the Bash script “sync\_data.sh” in the base directory of the project copies all information in the Excel files in the “data/countries” folder and synchronizes it into folder names matching the origin filenames, creating “Data.csv” files containing all of the data for that specific country—for example, the file “data/countries/bahamas.xlsx” is copied over into the “code/analysis/bahamas/Data.csv” file. The same principle applies for the “\_CARIBBEAN.xlsx” file containing the Caribbean panel data.



Figure : Sample Excel overview page (e.g., Bahamas)

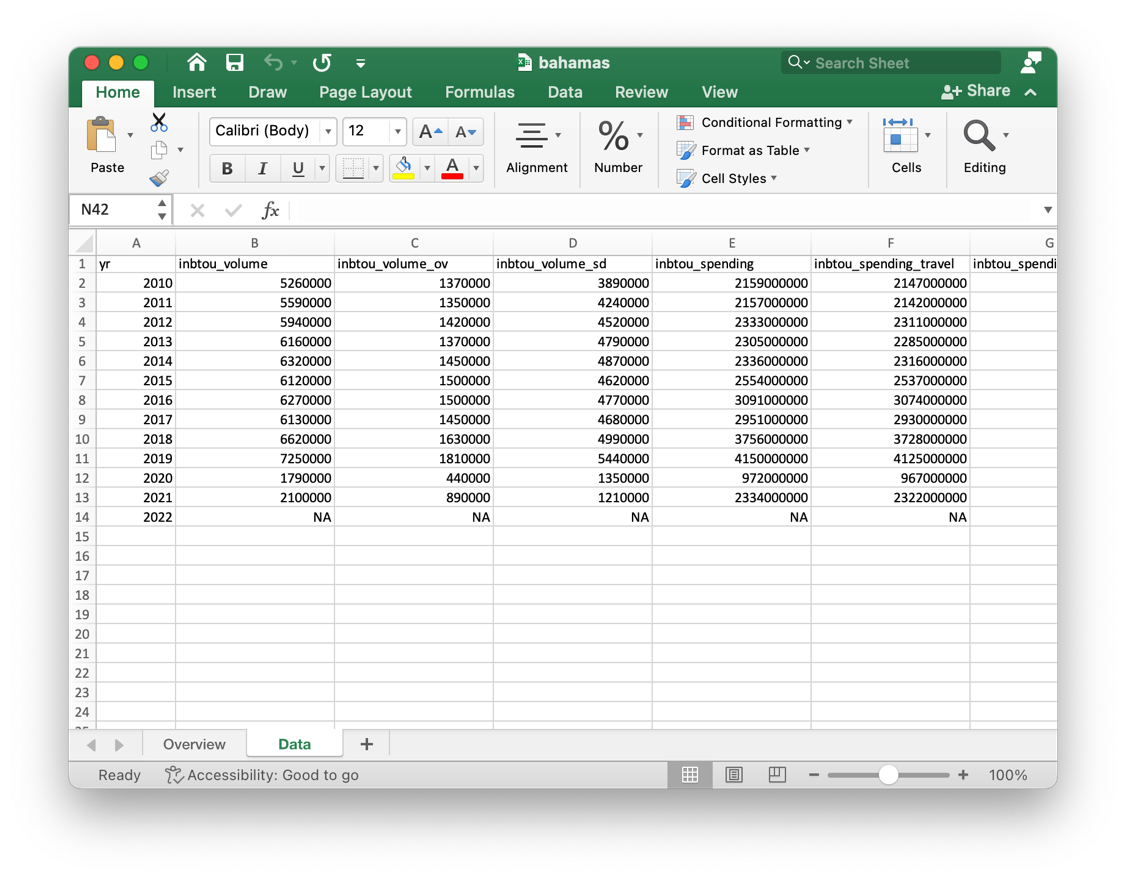


Figure : Sample Excel data sheet (e.g., Bahamas)

Figure 10 shows the output of executing the “sync\_data.sh” script to synchronize the Caribbean economic and COVID-19 data from the “data” folder to the “code” directory. In the case of an error about “gnumeric” or “ssconvert,” run the following command:

|  |
| --- |
| **$ brew install gnumeric** |

Then retry running “sync\_data.sh.”

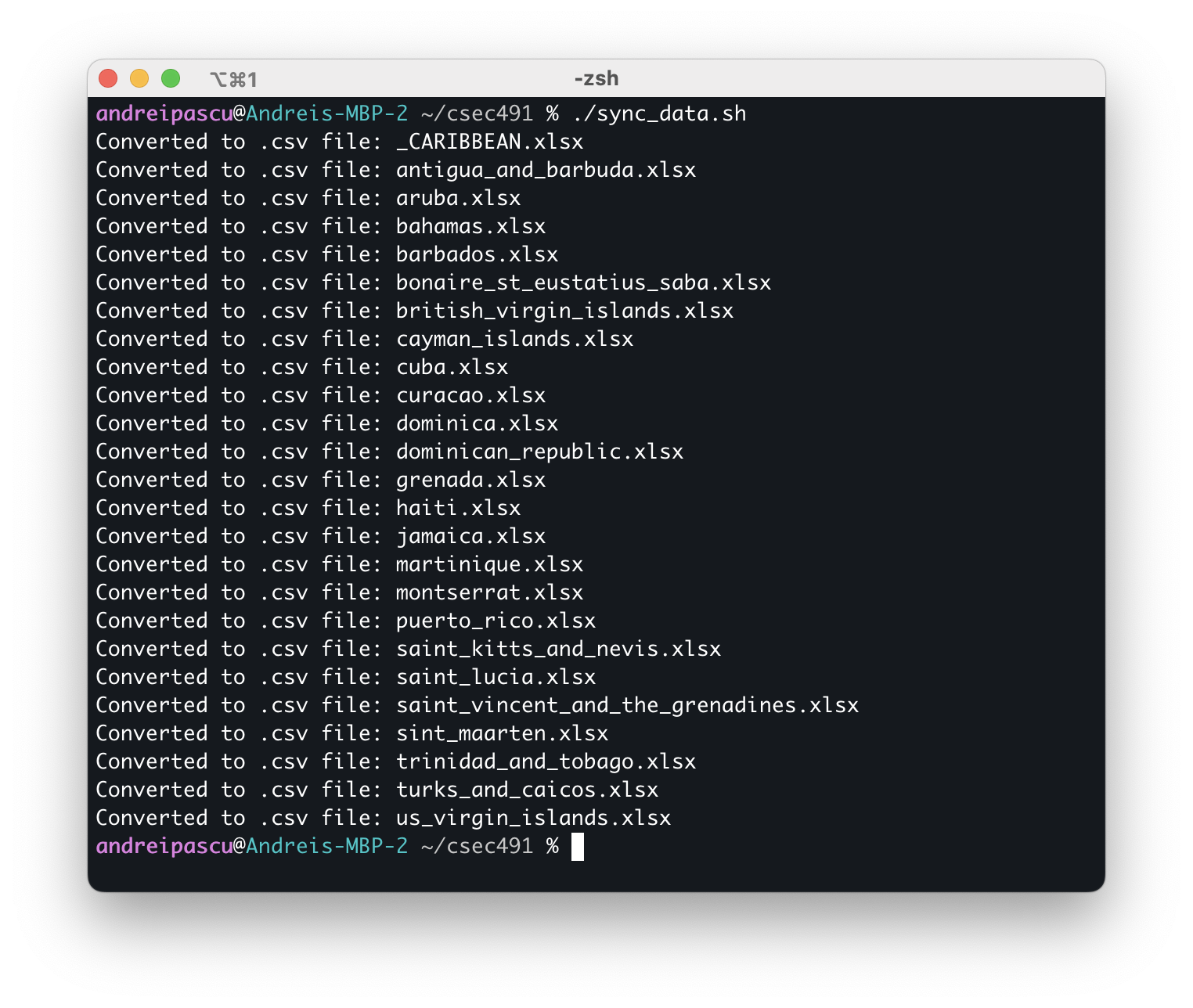


Figure : Sample run of the “sync\_data.sh” Bash script

After the synchronization step outline above is completed, data are ready to be analyzed. Jupyter notebook files titled “main.ipynb” are contained within each Caribbean panel and per-country folder. Many of these files import from the “code/common” folder, which includes a “common.py” Python module that contains function definitions for frequently used code, such as generating difference-in-differences plots and performing OLS regression.

In order to access these files, it is recommended to use JupyterLab for viewing, editing and execution of the notebook files. In a terminal window, change directories to the “code/analysis” folder and run the following command:

|  |
| --- |
| **$ jupyter-lab** |

After a few seconds of executing the command, an internet browser window should appear with the JupyterLab interface. To investigate a country’s analysis or the panel regression, open the corresponding folder and the “main.ipynb” Python notebooks. Sections III and IV present the results that were obtained from running the econometric analyses contained within these files.

# Section III. Individual Caribbean Country Analysis

This section focuses on the per-country analysis and is dependent on the data availability for the respective region. Most countries have trending economic data starting in 2010 through 2019 and continues through the onset of the pandemic. The first analysis technique considers the pre-2020 data as a representative trendline that projects the hypothetical economic values in the absence of COVID-19. We use the slope of the line of best-fit to estimate the impact of the pandemic during 2020 and following years by computing the intercept difference between the predicted—i.e., under the hypothesis that a health crisis never occurred in the year 2020—and recorded post-pandemic data.

This method of analysis resembles a pseudo-difference-in-differences technique used in statistical models with the notable exception that there is no control group; the pandemic affected all countries around the world and the best one can do is hypothesize about the economy had COVID-19 not been present. To this extent, we assume a pseudo-control group constructed through a projection based on the time series trend observed in the years prior to 2020, in order to give a quantitative evaluation into the impact of the pandemic on the Caribbean tourism sector. We estimate the per-country average economic effect during the pandemic timeframe using the following statistical model:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

The dependent variable in Equation 1 corresponds to the economic data of choice—e.g., the volume and expenditure of incoming tourists—whereas represents the yearly trend from the pre-2020 data available and constitutes the average effect of the COVID-19 pandemic during 2020 and further years: is an indicator variable for years succeeding the onset of the pandemic.

Second, we perform a linear OLS regression on the selected econometric variables on the number of COVID-19 cases, deaths and mortality rate. The model that is used is as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Equation 2 represents a log-level statistical analysis and therefore alters the interpretation of each of the estimated coefficients. Distinct from the prior regression, corresponds to the economic variable which has been transformed into logarithmic value, whereas is the main regressor in the model—here chosen to represent the number of COVID-19 cases, deaths or mortality rates across different regression runs.

Importantly, denotes the percent increase in the original dependent variable which results from a marginal increase in the regressor ; in other words, a one-unit increase in results in a percent increase in . In the case of COVID-19 cases and deaths, we can interpret the coefficients directly. However, mortality rates are represented in the dataset in percentage values, rendering a one-unit increase in the regressor as a 100 percent mortality growth. The proportional death rate of the COVID-19 virus is in the 1 to 10 percent range; thus, when interpreting the coefficient estimate for mortality rates as the regressor, we divide the numerical value by 1,000 to be interpreted as follows: a 0.1 percentage point increase in the mortality rate results in a percent increase in . The remaining term in Equation 2 consists of indicator variables after and on the year 2021 that contextualize the significance of the estimate: it is the impact on the economic variable from a marginal increase in the regressor conditional on the year 2020. Consequently, it acts as the effect specific for the onset year of the COVID-19 pandemic and as the baseline point-of-reference for the remaining coefficient estimates.

The variable represents the end year for the available data and is generally one of 2020, 2021 or 2022. Note that if is equal to 2020, the summation term is nullified and is thus not included in the regression. For larger than 2020, each represents the intersection between the COVID-19 regressor variable and the indicator variable for year , for starting at year 2021 and ending in year . The purpose of these intersected variables is to observe baseline changes to the impact of the regressor in the year 2020: , and so forth are the effects of a marginal increase in the regressor on the dependent variable relative to the baseline estimate, conditional on the year 2021, 2022 and so forth, respectively. More precisely, if the regressor is the number of COVID-19 cases and is equal to 2022, then is the baseline impact of case incidence on the economic dependent variable in the year 2020, whereas is the effect of COVID-19 cases in the year 2021 relative to that of 2020: therefore, the actual effect of cases on the dependent variable for the year 2021 is , and for 2022, it is .

The rest of this section presents the results using the analyses outlined above in Equation 1 and Equation 2. It is important to note the caveat that limited data points in the regressions exist and present a small sample size problem; thus, the coefficient estimates should be treated with the scope of the data availability in mind.

## Antigua and Barbuda

1. <https://www.imf.org/en/Publications/fandd/issues/2020/12/impact-of-the-pandemic-on-tourism-behsudi> [↑](#footnote-ref-1)
2. [https://jupyter.org](https://jupyter.org/) [↑](#footnote-ref-2)
3. [https://brew.sh](https://brew.sh/) [↑](#footnote-ref-3)
4. <https://matplotlib.org> [↑](#footnote-ref-4)
5. [https://numpy.org](https://numpy.org/) [↑](#footnote-ref-5)
6. [https://pandas.pydata.org](https://pandas.pydata.org/) [↑](#footnote-ref-6)
7. <https://www.statsmodels.org/stable/index.html> [↑](#footnote-ref-7)
8. [https://bashtage.github.io/linearmodels](https://bashtage.github.io/linearmodels/) [↑](#footnote-ref-8)
9. [https://seaborn.pydata.org](https://seaborn.pydata.org/) [↑](#footnote-ref-9)
10. [https://www.statista.com](https://www.statista.com/) [↑](#footnote-ref-10)