Kevin Galvan

Hiram Rodriguez

SI 206 Final Report

Github Repo Link : <https://github.com/kevingalvan86/SI-206-Data-Analysis->

**Goals**

Considering everything that is happening right now in society with the COVID-19 pandemic, we wanted to create a code that is relevant to the issue. Our goals for this project was to extract important data about the virus to further our understanding and severity of the issue. Originally we wanted to collect data of the most affected counties in the United States, but after not finding any COVID-19 APIs that return information on a per county basis we decided on collecting data per country. This data includes country region, population, COVID-19 cases and deaths and google trends data for the term “COVID-19”. We then want to create visualizations that show the differences in region population and region COVID-19 data.

**Achievements**

With our code we were able to successfully extract information regarding regions and their total number of cases, deaths, and recoveries within different countries. Additionally, we were able to successfully create two visualizations. One which was a bar graph of different regions around the world and their population number compared to one another. This was an important visualization to have because it allows us to see the total number of population each region has so we can later compare their numbers on cases and deaths. The second visualization is another bar graph that consists of the number of cases and deaths per region. This visualization is used to calculate the ratio of deaths and cases in comparison to the total population of their respective region. With these two graphs we were able to accomplish our goal of providing data on the status of COVID-19 worldwide.

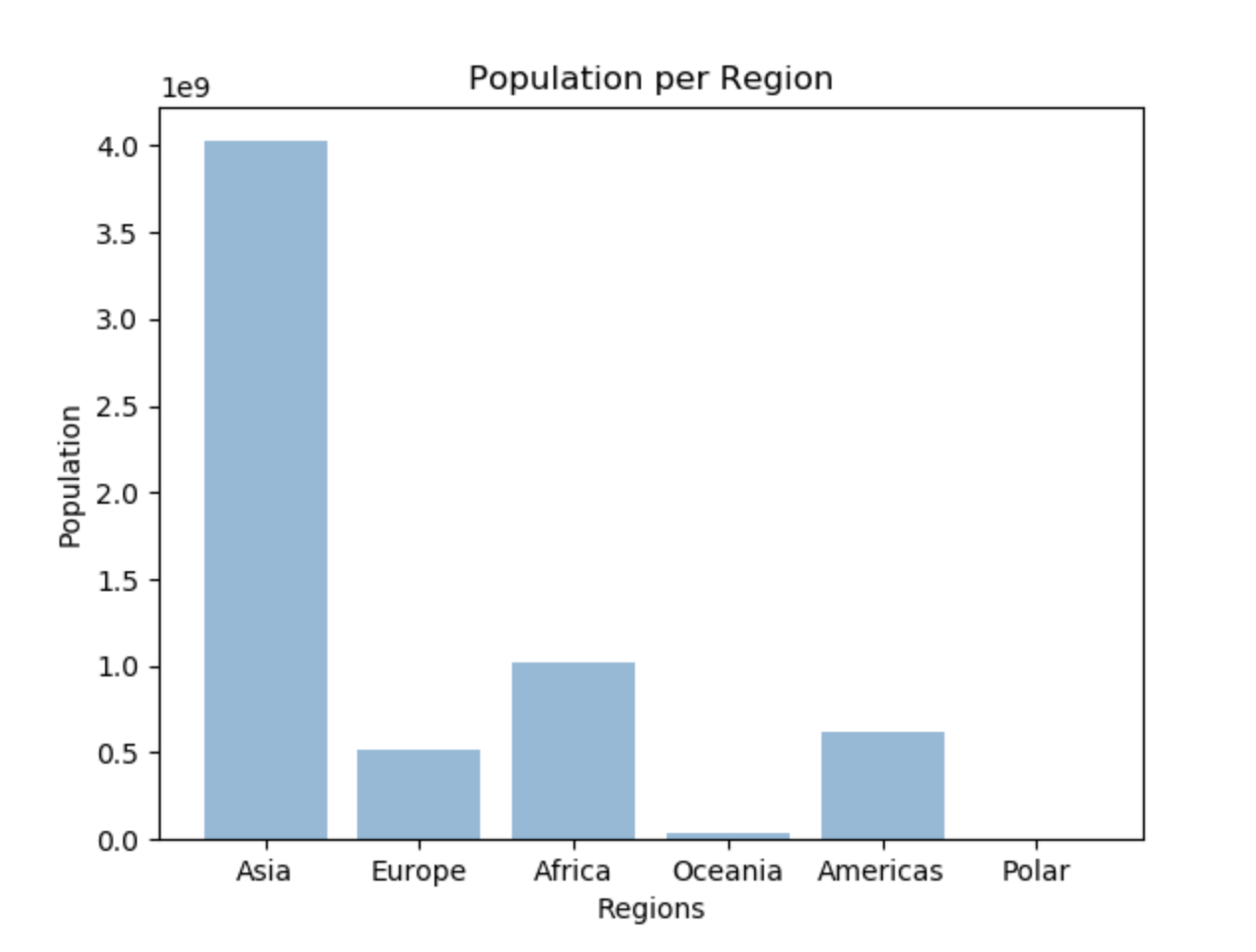
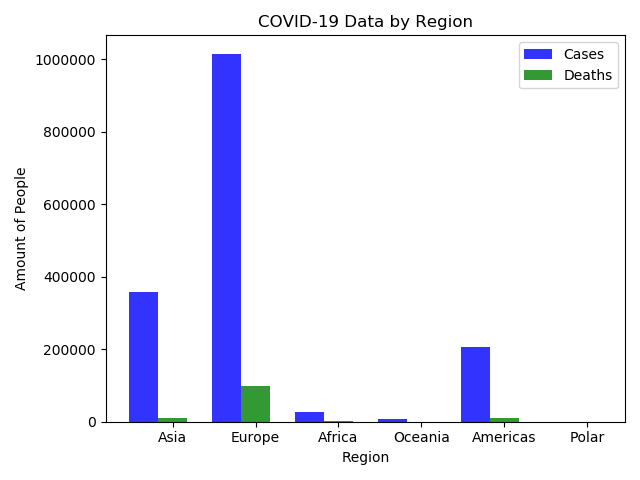
**Problems**

While there were many APIs available right now for COVID-19 data per country, it was difficult to find one that returned the type of data we needed. Most of the APIs we were finding did not return the region data that matched with the region data our other api had. We needed them to match in order to get per region data in our code. Another problem we faced was trying to organize our database tables and keys in a way that would make it possible to extract the information we needed in our later select statement.

**Data**

Calculations from the data in the database can be found [here](https://docs.google.com/spreadsheets/d/1q-AcsWH9h_DtaHjTjnwr3VyRzux8jI-FIj-PARKPTUw/edit#gid=0) on an excel sheet and [here](https://github.com/kevingalvan86/SI-206-Data-Analysis-/blob/master/Calculation_File) on github.

**Visualizations**



**Instructions**

In order to get all the data needed for the visualizations in the database, you must run the final\_project\_loads.py file multiple times. Each time you run it, it will only add 20 items to the database. You must keep running it until the last line of outputted text is “Country Already in Google Table 198” which means all the data is already in the database. After that you must run the final\_project\_dumps.py file in order to create the visualizations from the data in the database.

|  |  |  |  |
| --- | --- | --- | --- |
| **Functions** | **Input** | **Output** | **Purpose** |
| get\_covid\_info() | N/A | Dictionaries of countries with their new confirmed cases, total confirmed cases, new deaths, total deaths, new recoveries, and total recoveries. | We created this function to neatly organize our data that we’ll be extracting information from. It uses the api at api.covidapi.com to get the amount of COVID-19 cases and deaths per country |
| get\_glonal\_info() | N/A | List of dictionaries that consists of global information. Information such as a country’s region, timezone, population, currency, and capital. | We created this function to neatly organize our data that we’ll be extracting information from. It utilizes the restcountries.eu API to get data summaries from each country |
| get\_google\_trends() | N/A | Returns a csv file of google trend data for each country for the term “COVID-19” | We created this function to neatly organize our data that we’ll be extracting information from. It utilizes the pytrends module in order to get google trend data for each country. |

**Resources**

**Rest Countries API**

URL: https://<https://restcountries.eu/>

**COVID-19 API**

URL: <https://covid19api.com/>

**Pytrends Google Trends API**

<https://github.com/GeneralMills/pytrends>

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Issue Description** | **Location of Resource** | **Result** |
| 04/10/20 | Could not find API that returned Region information | <https://covid19api.com/> | Issue solved |
| 04/11/20 | Could not find country API that returns region information with the names in the same format as the region names from the other API | <https://restcountries.eu/> | Issue solved |