Covid-19 Report and Analysis

Introduction

This report analyzes data collected on 204 countries, from approximately February 24, 2020 through approximately October 03, 2021. The client has requested that the data be separated by country, and that an average of new cases per day in that timeframe be reported for each country.

Body

Data

This report was written using MyJupyterNotebooks. First I imported Pandas, MatPlotLib, Seaborn, and Geopandas to analyze and write reports on the data. Next I imported the Covid-19 data from a .csv file. Then I requested the number of rows and columns (121034 and 65 respectively), the data types of each column (all were float or string/"Object"), and the column names.
Because I wanted to create a heat map of the data, I also downloaded a generalized world countries file that contained the GPS data for all world countries.

Method

First I merged the country data with the Covid-19 data to create a master dataset which could be used to create a heatmap of Covid-19 cases. I dropped the columns that the client did not request for this user story, and then requested a list of all unique country names in the spreadsheet. From there, I separated the data by country, so that each country had its own table. This was not strictly necessary for the heat map, but the smaller datasets were used to check my work. Next, I calculated the sum and average of the new_cases column for each country. This was done by adding all rows (number of cases reported per day), then dividing by total number of entries. Finally, I merged the country data into my new dataframe of sums and averages.
The cleaned sum and average + country dataframe was used to generate a table of the results, as well as a chloropleth heat map of the averages for each country.

Results

For a better readability, I have published a table containing the sums and averages of Covid-19 data for each country here: New_Cases_Data

Analysis

An example of the separated data for one country, after the Covid-19 data had been combined with the Country data:

```
data_Congo #cleaned data for the country of Congo in the continent of Africa
```

	COUNTRY	ISO	geometry	continent	date	total_cases	new_cases
22030	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2020-03-15	1.0	1.0
22031	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2020-03-16	1.0	0.0
22032	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2020-03-17	1.0	0.0
22033	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2020-03-18	1.0	0.0
22034	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2020-03-19	3.0	2.0
...	...	...	...	...	...	...	...
22593	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2021-09-29	14244.0	0.0
22594	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2021-09-30	14244.0	0.0
22595	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2021-10-01	14359.0	115.0
22596	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2021-10-02	14359.0	0.0
22597	Congo	CG	POLYGON ((1768922.388 -439280.619, 1767784.921...	Africa	2021-10-03	14359.0	0.0

568 rows × 7 columns

The Sum and Mean of new cases table, grouped by country:

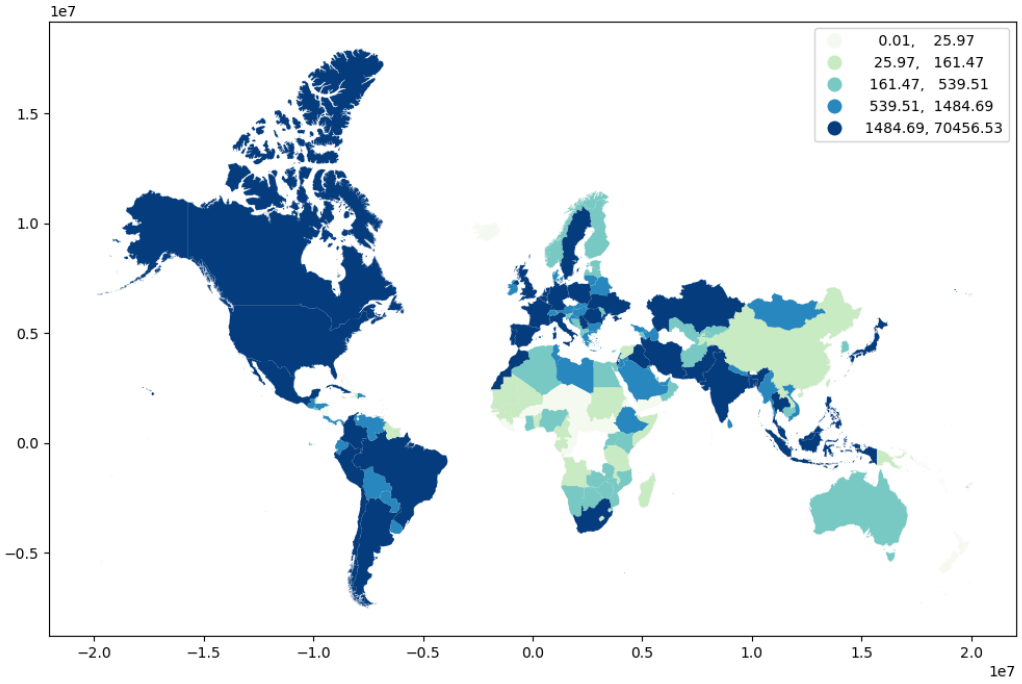
```
#Sum and average of cases per country
sumTotal=map_and_stats.groupby('COUNTRY')['new_cases'].agg(['sum', 'mean'])
sumTotal
```

COUNTRY	sum	mean
Afghanistan	155191.0	263.930272
Albania	171794.0	299.292683
Algeria	203789.0	347.170358
Andorra	15222.0	26.199656
Angola	58943.0	104.694494
...	...	...
Vietnam	808578.0	1304.158065
Wallis and Futuna	0.0	NaN
Yemen	9139.0	16.861624
Zambia	209163.0	370.200000
Zimbabwe	131129.0	232.911190

205 rows × 2 columns

The heatmap of average new cases:

```
fig, ax=plt.subplots(figsize=(12,10), subplot_kw=dict(aspect='equal'))
worldData.plot(column='New Cases Mean', scheme='Quantiles', k=5, cmap='GnBu', legend=True, ax=ax)
fig.savefig('CovidAverageMap.png')
```



Conclusion

Out of the 204 countries for which we have data in the approximate range of February 24, 2020-October 03, 2021, these five countries had the highest average reported new cases of Covid-19:

United States	70456.52742
India	55195.27243
Brazil	36635.01877
United	12970.27778

Kingdom	
France	12085.8479

These five countries had the lowest average reported new cases of Covid-19:

Solomon Islands	0.05602240896
Kiribati	0.01438848921
Vanuatu	0.01219512195
Marshall Islands	0.01173020528
Samoa	0.009375

And the following countries were excluded as outliers, since they had no data:

Anguilla
Aruba
Bermuda
British Virgin Islands
Cayman Islands
Cook Islands
Curacao
Falkland Islands
French Polynesia
Gibraltar
Greenland
Guernsey
Isle of Man
Jersey
Montserrat
Nauru
New Caledonia

Niue
Pitcairn
Saint Helena
Tokelau
Tonga
Turkmenistan
Turks and Caicos Islands
Tuvalu
Wallis and Futuna