gmena

April 29, 2025

[15]:

	Missing Values	Percent Missing
continent	26525	6.176721
total_cases	17631	4.105627
new_cases	19276	4.488689
new_cases_smoothed	20506	4.775111
total_deaths	17631	4.105627
new_deaths	18827	4.384133
new_deaths_smoothed	20057	4.670555
total_cases_per_million	17631	4.105627
new_cases_per_million	19276	4.488689
new_cases_smoothed_per_million	20506	4.775111
total_deaths_per_million	17631	4.105627
new_deaths_per_million	18827	4.384133
new_deaths_smoothed_per_million	20057	4.670555
reproduction_rate	244618	56.962753
icu_patients	390319	90.891287
icu_patients_per_million	390319	90.891287
hosp_patients	388779	90.532677
hosp_patients_per_million	388779	90.532677
weekly_icu_admissions	418442	97.440125
weekly_icu_admissions_per_million	418442	97.440125
weekly_hosp_admissions	404938	94.295528
weekly_hosp_admissions_per_million	404938	94.295528
total_tests	350048	81.513617
new_tests	354032	82.441347
total_tests_per_thousand	350048	81.513617
new_tests_per_thousand	354032	82.441347
new_tests_smoothed	325470	75.790283
new_tests_smoothed_per_thousand	325470	75.790283
positive_rate	333508	77.662044
tests_per_case	335087	78.029737
tests_units	322647	75.132907
total_vaccinations	344018	80.109446
people_vaccinated	348303	81.107269
people_fully_vaccinated	351374	81.822395
total_boosters	375835	87.518484
new_vaccinations	358464	83.473401
new_vaccinations_smoothed	234406	54.584745

```
total_vaccinations_per_hundred                  344018    80.109446
people_vaccinated_per_hundred                   348303    81.107269
people_fully_vaccinated_per_hundred             351374    81.822395
total_boosters_per_hundred                      375835    87.518484
new_vaccinations_smoothed_per_million           234406    54.584745
new_people_vaccinated_smoothed                  237258    55.248874
new_people_vaccinated_smoothed_per_hundred      237258    55.248874
stringency_index                                233245    54.314390
population_density                               68943    16.054350
median_age                                       94772    22.068998
aged_65_older                                   106165    24.722018
aged_70_older                                    98120    22.848627
gdp_per_capita                                  101143    23.552575
extreme_poverty                                 217439    50.633740
cardiovasc_death_rate                           100570    23.419144
diabetes_prevalence                              83524    19.449742
female_smokers                                  182270    42.444142
male_smokers                                    185618    43.223771
handwashing_facilities                          267694    62.336326
hospital_beds_per_thousand                      138746    32.308964
life_expectancy                                  39136     9.113370
human_development_index                         110308    25.686774
excess_mortality_cumulative_absolute            416024    96.877059
excess_mortality_cumulative                     416024    96.877059
excess_mortality                                416024    96.877059
excess_mortality_cumulative_per_million         416024    96.877059
```
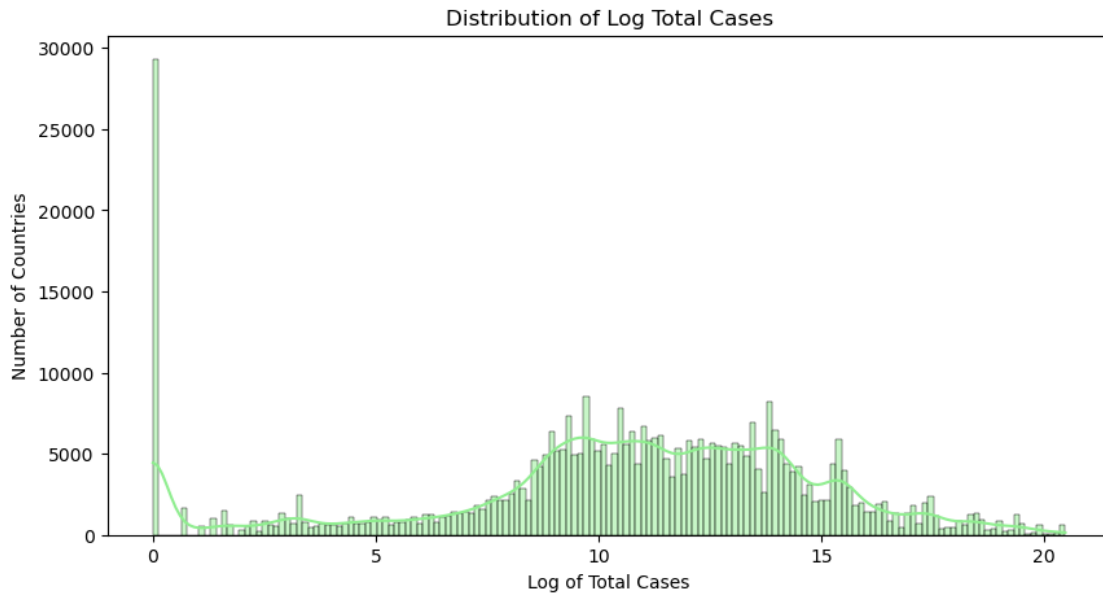
0.1 #### Data Preprocessing

```
[16]:                           Missing Values  Percent Missing
       continent                         26525         6.176721
       total_cases                       17631         4.105627
       new_cases                         19276         4.488689
       new_cases_smoothed                20506         4.775111
       total_deaths                      17631         4.105627
       new_deaths                        18827         4.384133
       new_deaths_smoothed               20057         4.670555
       total_cases_per_million           17631         4.105627
       new_cases_per_million             19276         4.488689
       new_cases_smoothed_per_million    20506         4.775111
       total_deaths_per_million          17631         4.105627
       new_deaths_per_million            18827         4.384133
       new_deaths_smoothed_per_million   20057         4.670555
       population_density                68943        16.054350
       median_age                        94772        22.068998
       aged_65_older                    106165        24.722018
       aged_70_older                     98120        22.848627
```

```
gdp_per_capita                   101143      23.552575
cardiovasc_death_rate            100570      23.419144
diabetes_prevalence               83524      19.449742
female_smokers                   182270      42.444142
male_smokers                     185618      43.223771
hospital_beds_per_thousand       138746      32.308964
life_expectancy                   39136       9.113370
human_development_index          110308      25.686774
```
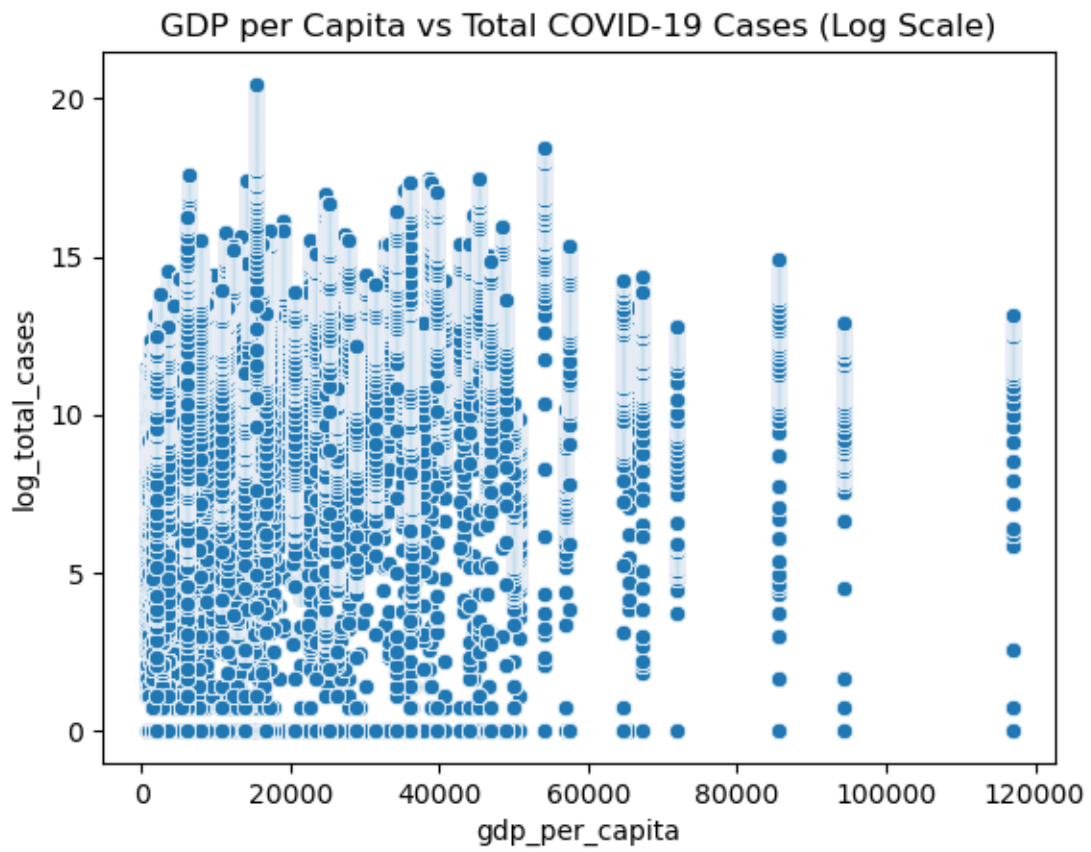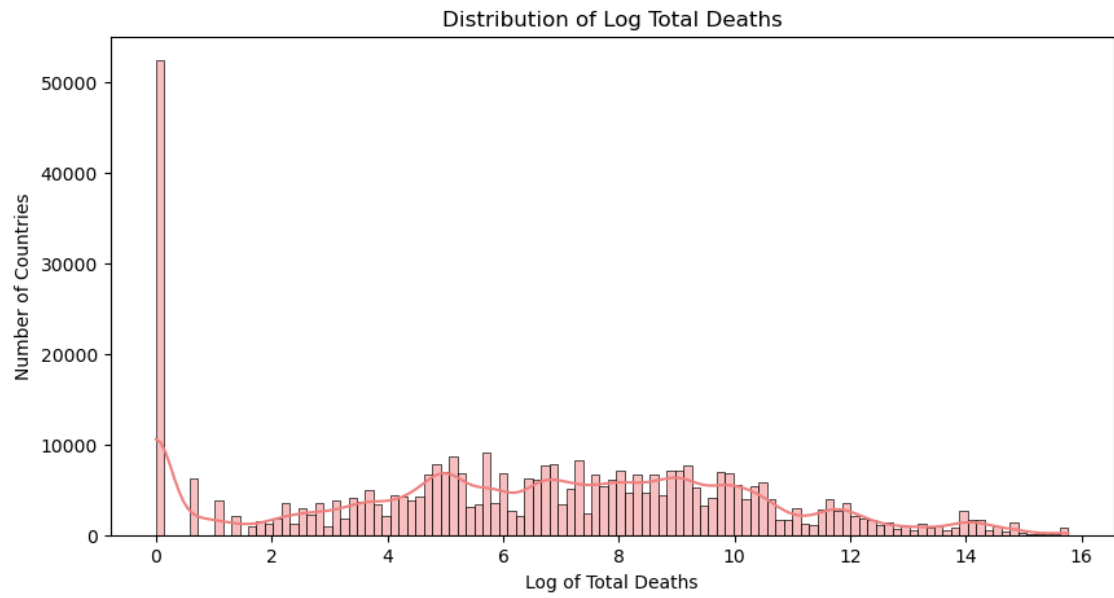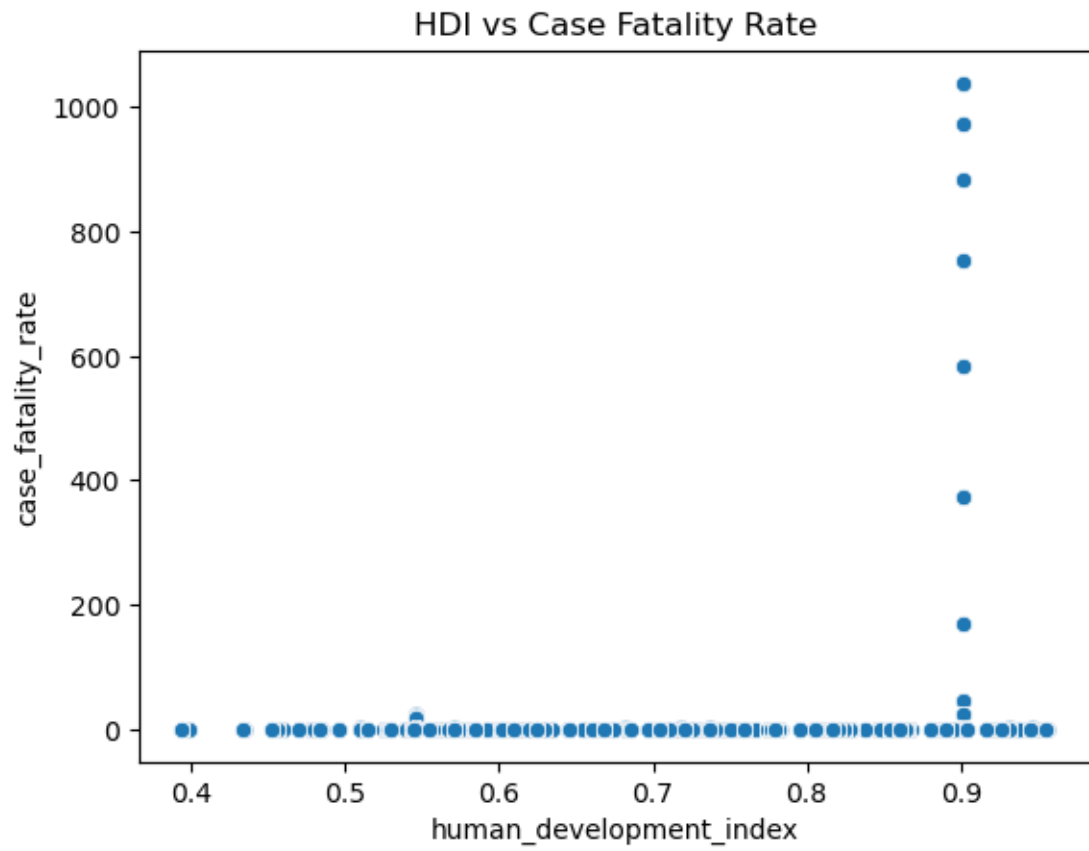
Feature Selection and Derivation

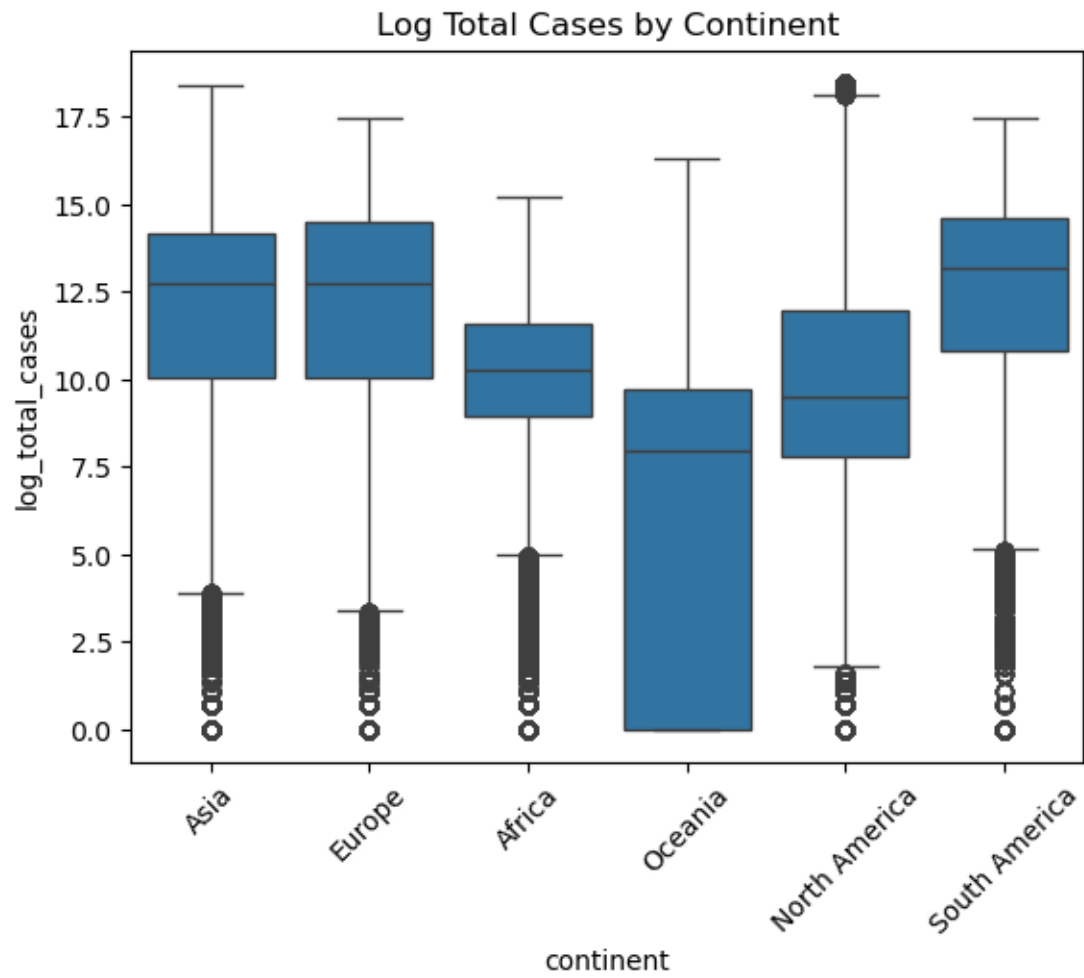## 0.2 ### Exploratory Data Analysis

Visualizations

```
        total_cases   total_deaths   gdp_per_capita   life_expectancy
count   4.118040e+05   4.118040e+05   328292.000000    390299.000000
mean    7.365292e+06   8.125957e+04    18904.182986        73.702098
std     4.477582e+07   4.411901e+05    19829.578099         7.387914
min     0.000000e+00   0.000000e+00      661.240000        53.280000
25%     6.280750e+03   4.300000e+01     4227.630000        69.500000
50%     6.365300e+04   7.990000e+02    12294.876000        75.050000
75%     7.582720e+05   9.574000e+03    27216.445000        79.460000
max     7.758668e+08   7.057132e+06   116935.600000        86.750000
```
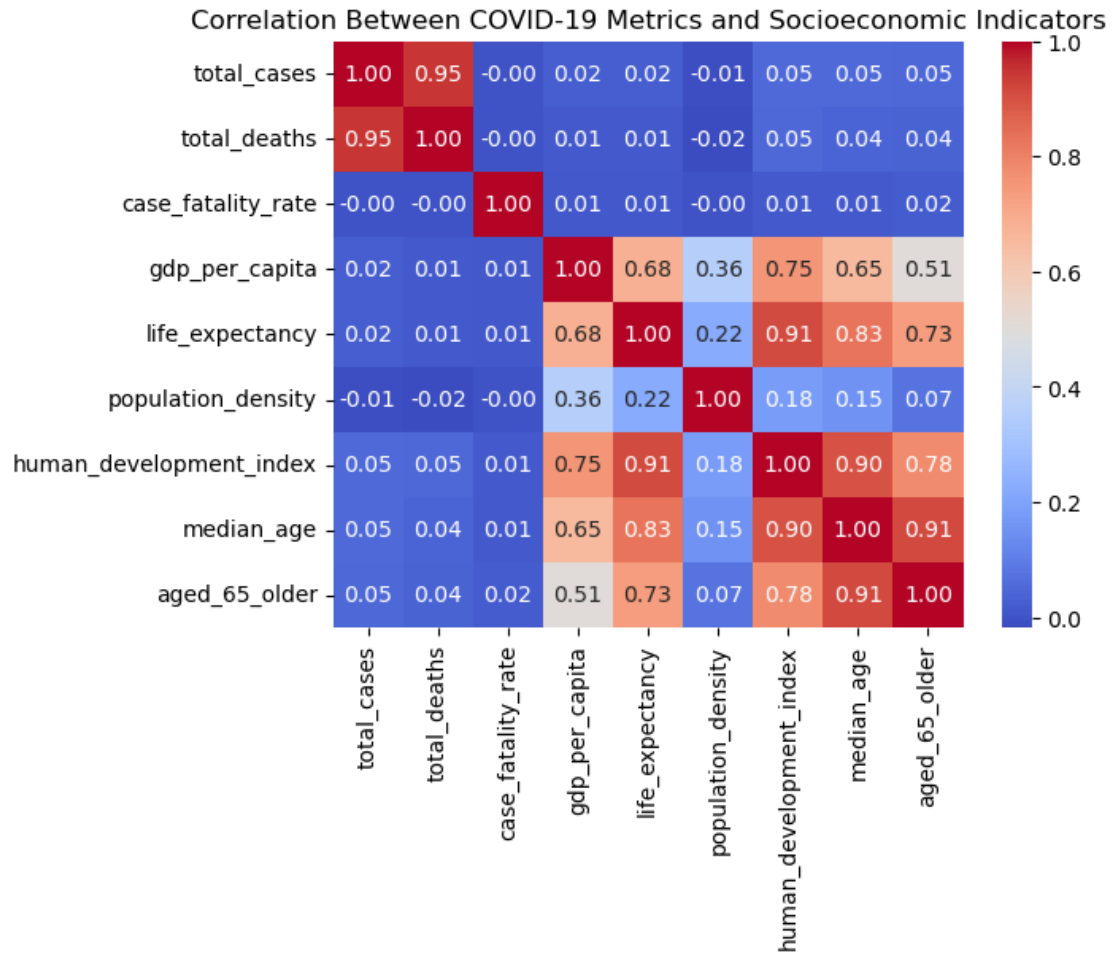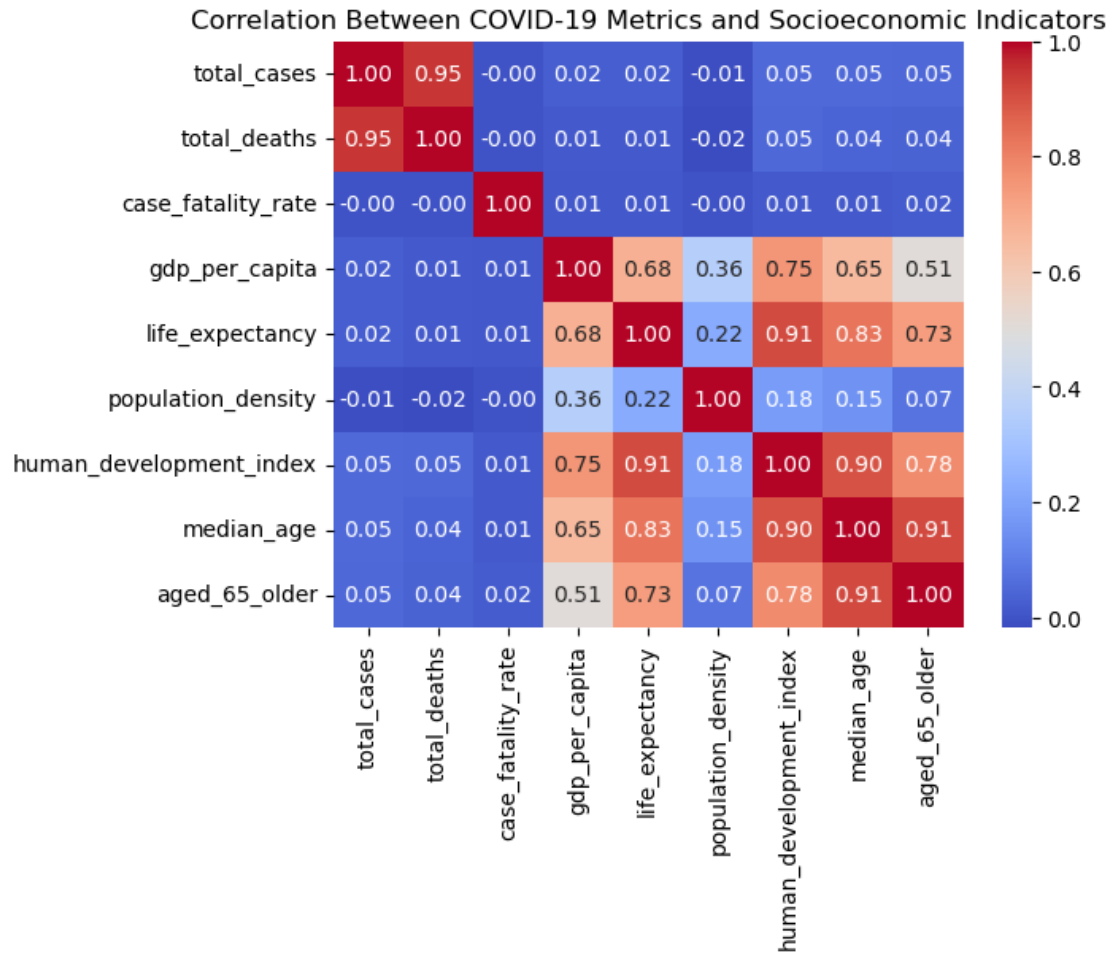

Distribution of Log Total Cases

Distribution of Log Total Deaths

GDP per Capita vs Total COVID-19 Cases (Log Scale)

HDI vs Case Fatality Rate

Log Total Cases by Continent

Correlation Analysis

Correlation Between COVID-19 Metrics and Socioeconomic Indicators

Correlation Between COVID-19 Metrics and Socioeconomic Indicators

1 Uncovering Response Patterns: Clustering Analysis of Global COVID-19 Data

1.1 Clustering Methodology: K-means, DBSCAN & Hierarchical Approaches

[ ]:

1.2 Cluster Evaluation Metrics: Purity and Sum of Squared Errors

[ ]:

1.3 Socioeconomic Profiling of Clusters

[ ]:

1.4 Parameter Tuning for Optimal Clustering

[ ]: