Covid-19 Project Summary

**Project Introduction**

The purpose of this project was to analyze Covid-19 data, including case and death counts by country, in conjunction with the Human Development Index (HDI). The HDI is an indicator that attempts to measure a country’s standard of living by considering three different components: Gross National Income per Capita, Life Expectancy and years of schooling (measured as mean years of schooling for adults and expected years of schooling for children). By analyzing Covid-19 data alongside HDI scores, it is possible to not only inspect how the pandemic impacted standards of living but also the relationship between a country’s HDI score and Covid-19 experience. Specifically, this project sought to address the following questions:

* *What is the mean/median confirmed case count? What does the distribution of confirmed cases look like?*
* *What is the mean/median confirmed death count? What does the distribution of confirmed deaths look like?*
* *How did COVID-19 impact different regions of the world?*
* *Is there a correlation between HDI scores and case/death counts*

**Methodology and Data Sources**

This project employed Python in the Jupyter Notebooks environment to carry out the importation, cleaning and analysis of the relevant data. Various packages were utilized, including Pandas, NumPy, Matplotlib, Seaborn, Requests and Beautifulsoup4.

The data was sourced from two different locations. The Covid-19 data was web scraped from worldometers.info, leveraging the package Beautifulsoup4 to parse the HTML code and extract the desired data. The HDI data was obtained from ourworldindata.org and downloaded directly from the site in csv format.

**Findings and Future Work**

Upon employing the relevant data analysis, several conclusions were made. The mean case count differed drastically from the median case count (3,736,089 vs. 463,210), indicative of several countries with extremely high case counts. The same could be said for total deaths. By region, Europe experienced the highest case count with South America and Asia following close behind. Oceania fared the best. Givin the strong positive correlation between total cases and total deaths, the same conclusions were found for total deaths by region. Comparing HDI scores and total cases and deaths both yielded weak positive correlations, perhaps undermining the idea that more well to do countries would be more resilient to a pandemic.

In conclusion, this project raised more questions than it did answer. It would be interesting to include more explanatory variables alongside the Covid-19 data to better understand why some countries fared worse during the Pandemic than others. For example, Stringency Index data(excluded due to time constraints) , population density and urbanization rates are all interesting variables that may help to explain the distribution of Covid-19 case counts and deaths.