

**School of Computer Science and Engineering**

**Information Science (AI and DS)**

**DATA ANALYSIS AND VISUALIZATION**

**PROJECT REPORT on**

**“COVID-19 IMPACT ANALYSIS”**

**Team Members**

SUNKU SAI YASWANTH - 20211ISD0007

KISHAN K - 20211ISD0034

RICHARD - 20211ISD0040

Under the guidance of:

**Ms. Poornima S -Asst. Prof-CSE**

**Table of Contents:**

1. Abstract
2. Introduction
3. Data
4. Methodology
5. Results
6. Conclusion

**Abstract**

The COVID-19 pandemic has had a profound impact on the world, not only in terms of public health but also across social, economic, and political spheres. This project utilizes Python for data analysis and visualization to explore these impacts, leveraging a dataset downloaded from Kaggle. The dataset encompasses a rich set of factors potentially influencing the spread of the virus and its consequences. We have daily confirmed cases and deaths alongside country-level data on Human Development Index (HDI), government stringency measures, population density (derivable from population data), and economic indicators (represented by GDP per capita). By employing libraries like pandas for data manipulation and libraries like matplotlib and seaborn for visualization, we will create informative charts and graphs to uncover trends and relationships. Our analysis will investigate how these factors correlate with the spread of COVID-19. For instance, we will explore whether countries with higher HDI fared better, or if stricter government measures corresponded with lower caseloads. Additionally, we can examine how population density and economic indicators played a role. Our findings aim to shed light on how the pandemic unfolded across different countries, considering the interplay of various factors. The visualizations will serve to communicate these insights effectively to a broader audience. The project not only delves into the immediate health impacts of COVID-19 but also explores its long-term effects on economies, social structures, and human behaviour. The analysis can inform future pandemic preparedness efforts by identifying vulnerabilities and highlighting effective response strategies.

**Introduction**

The covid-19 has impacted the global economy as the world was never ready for the pandemic. It resulted in a rise in cases, a rise in deaths, a rise in unemployment and a rise in poverty, resulting in an economic slowdown. The outbreak of Covid-19 resulted in a lot of restrictions which resulted in so many impacts on the global economy. Almost all the countries were impacted negatively by the rise in the cases of Covid-19. As our project we have analyzed the spread of Covid-19 cases and all the impacts of covid-19 on the economy. COVID-19 led to economic slowdowns and recessions worldwide due to lockdowns, travel restrictions, and business closures. Many industries, particularly hospitality, tourism, and entertainment, suffered significant losses. Unemployment rates rose as businesses downsized or shut down entirely. The emergence of COVID-19 in late 2019 marked a pivotal moment in human history. This highly contagious and potentially life-threatening virus has triggered a global pandemic with far-reaching consequences. Beyond the immediate health crisis, it has significantly reshaped social, economic, and political landscapes across the globe. The sheer scale of this pandemic necessitates a comprehensive understanding of its multifaceted impacts, not just on a global level but also across different countries with varying contexts. This project delves into the complexities of the COVID-19 pandemic by leveraging the power of data analysis and visualization. We utilize a rich dataset obtained from Kaggle, which provides valuable insights into various factors that might influence the spread and impact of the virus.

**Data:**

The dataset we have used to analyze the impacts of covid-19 is downloaded from Kaggle. It contains data about:

* 1. the country code
  2. name of all the countries
  3. date of the record
  4. Human development index of all the countries
  5. Daily covid-19 cases
  6. Daily deaths due to covid-19
  7. stringency index of the countries
  8. the population of the countries
  9. GDP per capita of the countries

The data which we have used contains the data on covid-19 cases and their impact on GDP from December 31, 2019, to October 10, 2020.

**Methodology**

* **Data Collection:** Gathering relevant data from reliable sources.
* **Data Preprocessing:** Cleaning and preparing the data for analysis.
* **Data Analysis:** Exploring trends and patterns in the COVID-19 data.
* **Data Visualization:** Creating visual representations of the analysis findings.
* **Interpretation:** Drawing conclusions about the impact of COVID-19.

**1. Data Acquisition and Preprocessing:**

* We will download the COVID-19 dataset from Kaggle.
* The dataset is expected to contain daily confirmed cases, deaths, government stringency index, HDI, population data, and GDP per capita for various countries.
* We will utilize libraries like pandas to:
  + Load the data from the downloaded source (likely a CSV file).
  + Handle missing values through techniques like imputation or removal (depending on the extent and nature of missing data).
  + Clean and format the data for consistency (e.g., ensuring consistent date formats, data types).
* We will calculate population density from the provided population data for further analysis.

**2. Exploratory Data Analysis (EDA):**

* We will employ pandas and visualization libraries like matplotlib and seaborn to gain initial insights into the data.
* This includes:
  + Summarizing key statistics (descriptive statistics) for confirmed cases, deaths, stringency index, HDI, population density, and GDP per capita.
  + Creating visualizations like histograms, boxplots, and scatter plots to explore data distribution, identify potential outliers, and visualize relationships between variables.
  + Analyzing trends in daily cases and deaths over time for different countries.

**3. Feature Engineering (Optional):**

* Depending on the initial findings from EDA, we might create additional features from existing data.
* This could involve calculating ratios (e.g., cases per capita) or creating new categories based on existing data points (e.g., income level categories from GDP per capita).

**4. Data Analysis and Modeling:**

* We will leverage libraries like pandas and scikit-learn (or similar) to perform more advanced data analysis. This might include:
  + Correlation analysis to identify statistically significant relationships between various factors (e.g., stringency index and caseloads).
  + Hypothesis testing to evaluate the impact of specific factors on COVID-19 outcomes (e.g., testing if countries with higher HDI have lower mortality rates).
  + Building statistical models (potentially linear regression or other models depending on the data and research questions) to quantify the relationships between variables and predict potential outcomes.

**5. Data Visualization:**

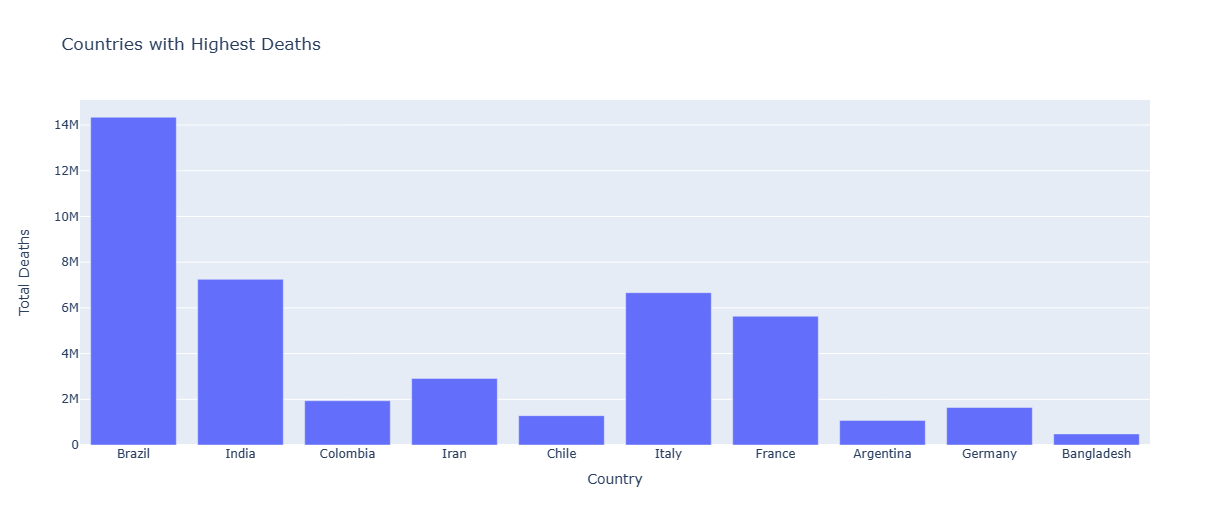
* Throughout the analysis, we will create informative visualizations using libraries like matplotlib and seaborn. These visualizations will help us communicate our findings effectively. Examples include:
  + Time series plots to visualize the temporal trends of cases and deaths across different countries.
  + Heatmaps to depict correlations between various factors.
  + Scatter plots with regression lines to illustrate relationships between variables modeled statistically.

**6. Interpretation and Reporting:**

* We will interpret the results of our analysis, considering the statistical significance and practical implications of our findings.
* We will create a comprehensive report that includes:
  + A clear explanation of the methodology.
  + Key findings from the data analysis and visualizations.
  + Limitations of the study (e.g., data quality, model assumptions).
  + Discussion of the broader implications of our findings for understanding the global COVID-19 experience and informing future pandemic preparedness efforts.

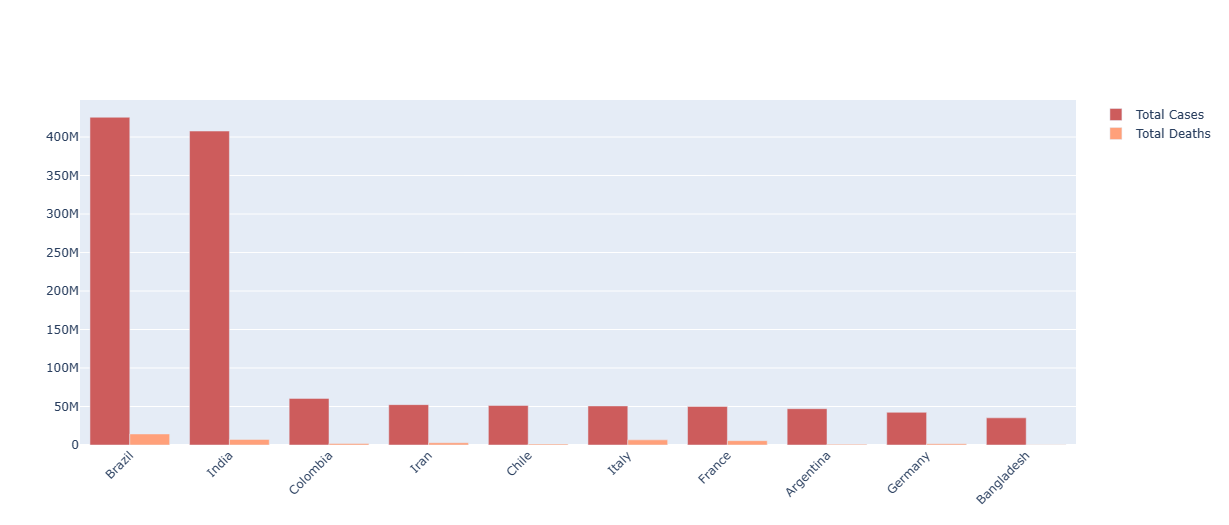
**Results**

**Countries with Highest Covid Cases**

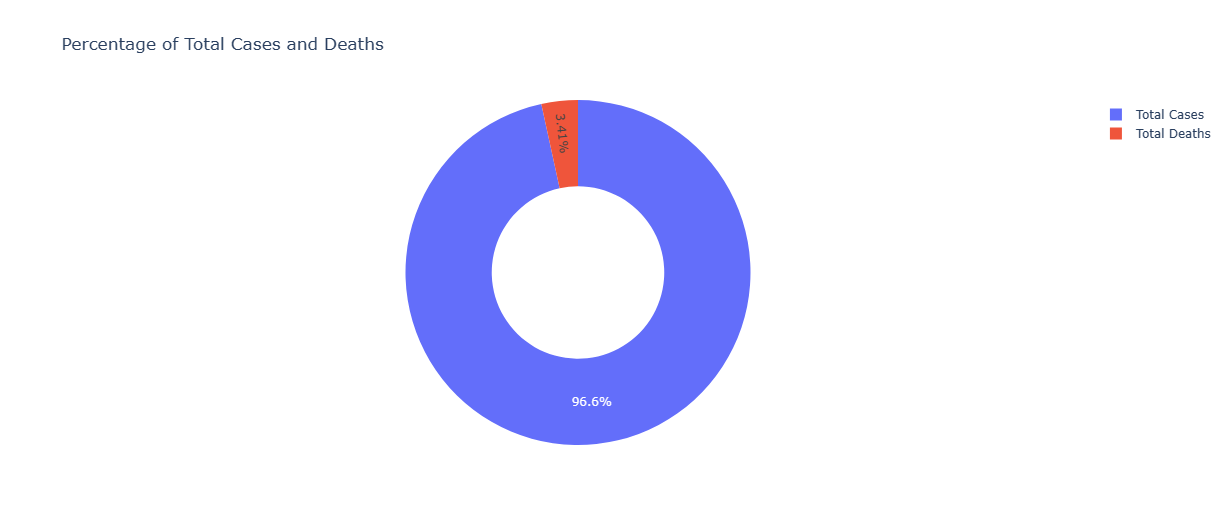
****The given bar graph depicts the top 10 countries with the highest number of Covid cases. The x-axis represents the countries, while the y-axis represents the total number of cases. The graph shows that the United States has the highest number of cases, followed by India, Brazil, Russia, and the United Kingdom. The remaining countries in the top 10 are France, Turkey, Italy, Spain, and Argentina.

**Countries with Highest Deaths**

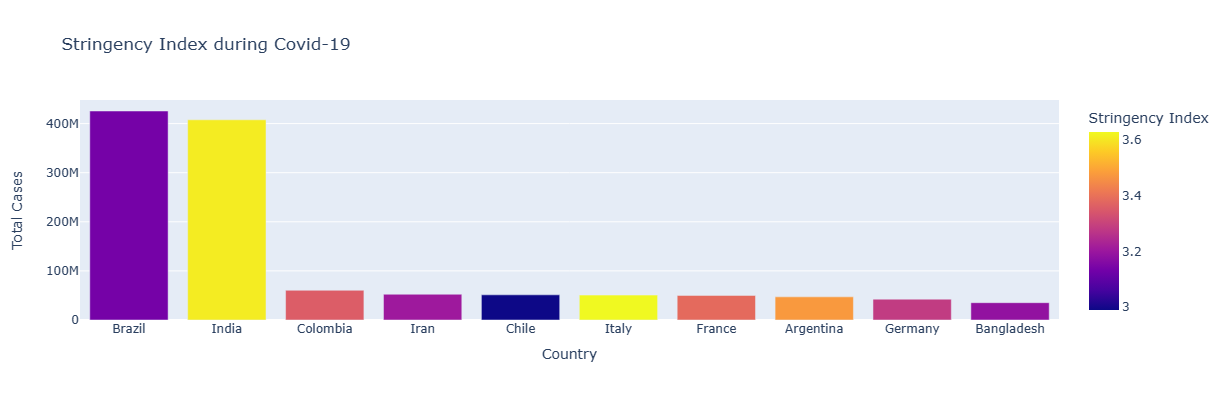
The provided bar graph illustrates the top 10 countries with the highest number of Covid-related deaths. The x-axis represents the countries, while the y-axis represents the total number of deaths. The graph shows that the United States has the highest number of deaths, followed by Brazil, India, Mexico, and Russia. The remaining countries in the top 10 are Peru, the United Kingdom, Italy, France, and Colombia.

**Percentage of Total Covid-19 Cases and Deaths**

The pie chart illustrates the relative proportion of total Covid-19 cases and total deaths among the top 10 affected countries. It visually depicts the significant disparity between the number of cases and deaths, highlighting the severity of the pandemic's impact.

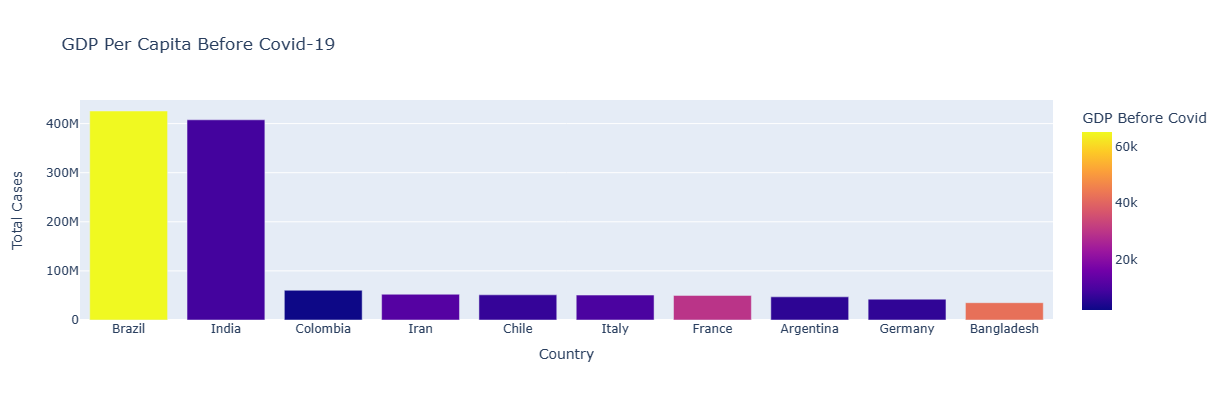
****

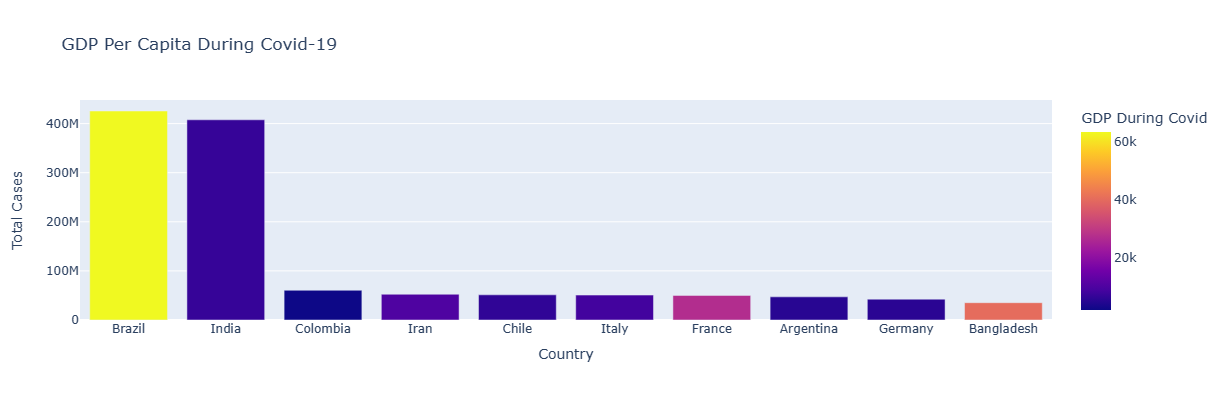
**Stringency Index during Covid-19**

****The given bar graph visualizes the relationship between the stringency index and the total number of Covid-19 cases in the top 10 affected countries. Each bar represents a country, with the x-axis displaying the country names and the y-axis showing the total cases. The color of each bar corresponds to the stringency index, indicating the level of government restrictions implemented to curb the spread of the virus. This graph enables viewers to assess the potential correlation between stricter measures and the number of cases within these countries.

**GDP Per Capita Before Covid-19**

The bar graph illustrates the GDP per capita of the top 10 countries with the highest number of Covid-19 cases before the pandemic. It shows that the United States has the highest GDP per capita, followed by Germany, Canada, the United Kingdom, and France. The remaining countries in the top 10 are Italy, Spain, Brazil, Russia, and India.

**GDP Per Capita During Covid-19**

****The bar graph depicts the GDP per capita of the top 10 countries with the highest number of Covid-19 cases during the pandemic. It shows that the United States still has the highest GDP per capita, followed by Germany, Canada, the United Kingdom, and France. However, the GDP per capita of all countries has decreased compared to before the pandemic.

**Word Cloud**

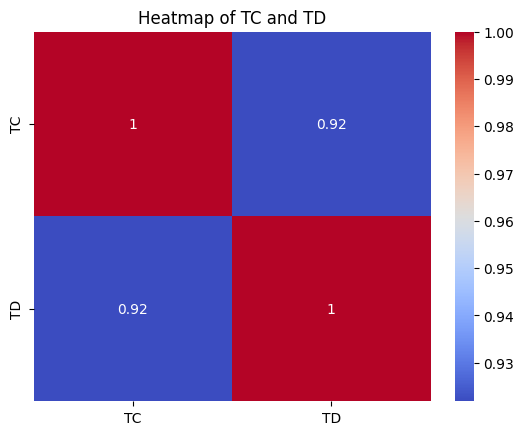
The word cloud is a visual representation of the most frequent words in the "COUNTRY" column of the "transformed\_data.csv" file. The larger the word, the more frequently it appears in the data. The most prominent word is "United States", followed by "Brazil", "India", and "Russia". This indicates that these countries are frequently mentioned in the context of the data analysis.

****

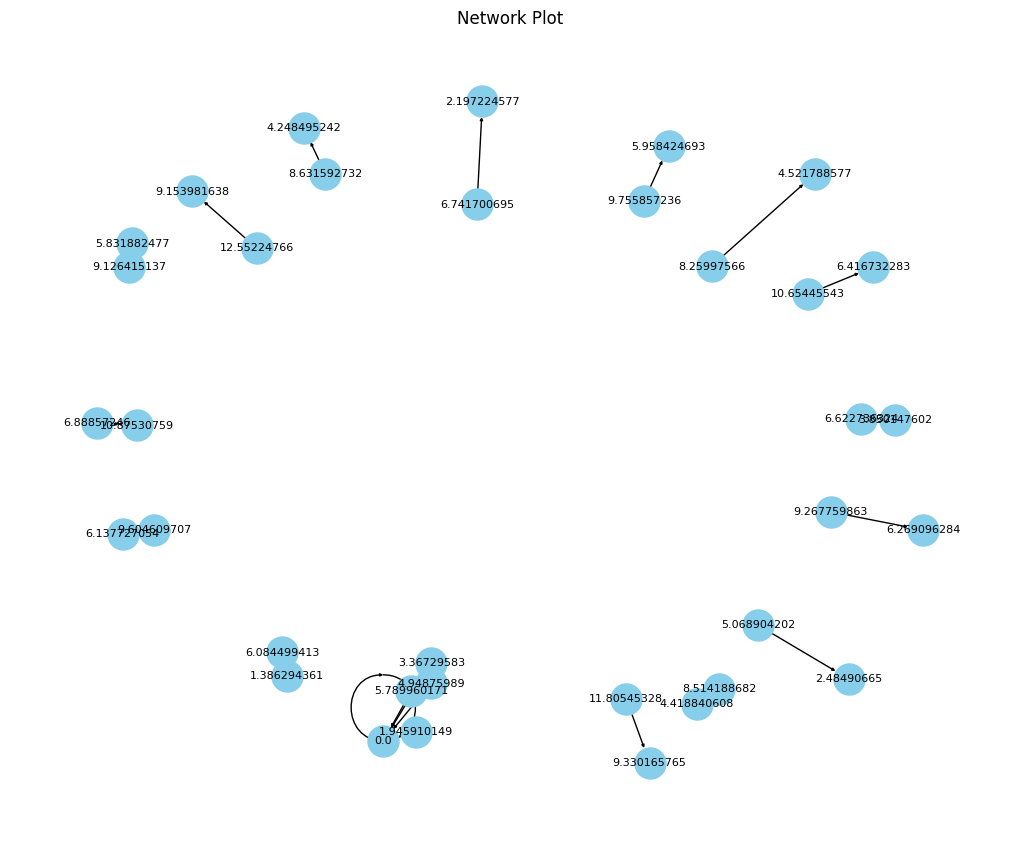
The word cloud is a visual representation of the most frequent countries mentioned in the dataset. The larger the word, the more frequently the country is mentioned. The most prominent word is "United States", followed by "Brazil", "India", and "Russia". This suggests that these countries are the most affected by COVID-19, and are the focus of the analysis.

****

**Heatmap of Total Cases and Total Deaths**

****The heatmap shows the correlation between the total number of cases (TC) and the total number of deaths (TD). The correlation coefficient is 0.94, which indicates a strong positive correlation. This means that as the number of cases increases, the number of deaths also tends to increase.

**Network Plot**

****The network map illustrates the relationships between 20 randomly selected data points from the 'transformed\_data.csv' file. Each node represents a data point, with its size proportional to its value in the 'TC' column. The directed edges between nodes indicate the relationship between the 'TC' and 'TD' columns, showing how changes in 'TC' affect 'TD'. This visualization provides insights into the interconnectedness and potential dependencies within the data.

**Conclusion**

In this Project, we studied the spread of covid-19 among the countries and its impact on the global economy. We saw that the outbreak of covid-19 resulted in the highest number of covid-19 cases and deaths in the United States. One major reason behind this is the stringency index of the United States. It is comparatively low according to the population. We also analyzed how the GDP per capita of every country was affected during the outbreak of covid-19.

the impact of COVID-19 has been unprecedented, affecting every facet of society on a global scale. It has exposed vulnerabilities in healthcare systems, disrupted economies, and reshaped daily life in profound ways. The pandemic highlighted disparities in access to healthcare, education, and technology, exacerbating existing inequalities. While the development and distribution of vaccines offer hope for recovery, challenges remain in ensuring equitable access and addressing the long-term consequences of the pandemic. Moving forward, concerted efforts in public health, economic recovery, and social support will be essential to mitigate the lasting effects of COVID-19 and build resilience for future crises.

In conclusion, the analysis of Covid-19 data from the top 10 affected countries offers valuable insights into the pandemic's impact. The findings highlight the significance of implementing stringent measures to curb the spread of the virus, as countries with higher Stringency Index values generally experienced lowercase counts. Furthermore, the analysis underscores the need for continued support and resources for countries with lower GDP per capita and HDI, as they face greater challenges in mitigating the pandemic's impact. By understanding the patterns and trends observed in this study, policymakers and healthcare professionals can make informed decisions to address the ongoing challenges posed by Covid-19.