Data Science

answer of assignments 2

Professor : Dr. Kherad Pishe

Assistant Professor : Mohammad Reza Khanmohammadi

Hasan Roknabady – 99222042

Colab Link :
https://colab.research.google.com/drive/1eGY0G8MIIcwg7zmugEsrq_EYMnoPpBLW?usp=sharing

COVID-19 Data Analysis Documentation

Introduction

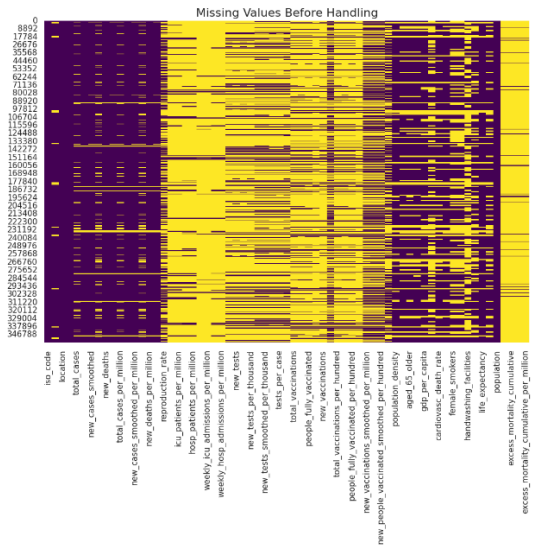
This documentation aims to present a detailed analysis of COVID-19 data, exploring the intricate relationships between demographic variables and COVID-19 outcomes. The dataset encompasses 67 columns, covering a range of factors such as total cases, deaths, and demographic metrics from various countries.

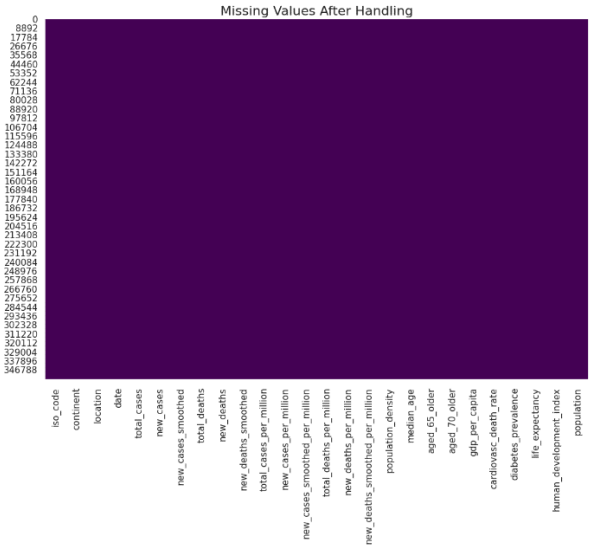
Data Overview

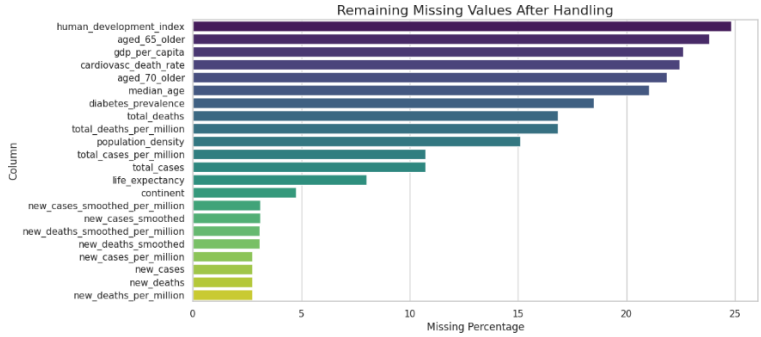
Upon initial examination, it was noted that several columns contained missing values, particularly those related to excess mortality and ICU admissions. Appropriate data imputation techniques were employed to handle these missing values and ensure the robustness of subsequent analyses.

Missing Data Handling

To address the missing data, we implemented methods to maintain the integrity of the dataset while minimizing the impact of missing information. And we handle missing values in fact







Demographic and COVID-19 Correlations

Correlation Heatmap

The correlation heatmap provided a visual representation of the relationships between 23 variables, offering insights into patterns and dependencies.

Key Correlations

1. Positive Correlations:

- *Total cases and new cases (0.94)*
- *New cases and new cases smoothed (0.84)*
- *New deaths and new deaths smoothed (0.91)*

2. Demographic Correlations:

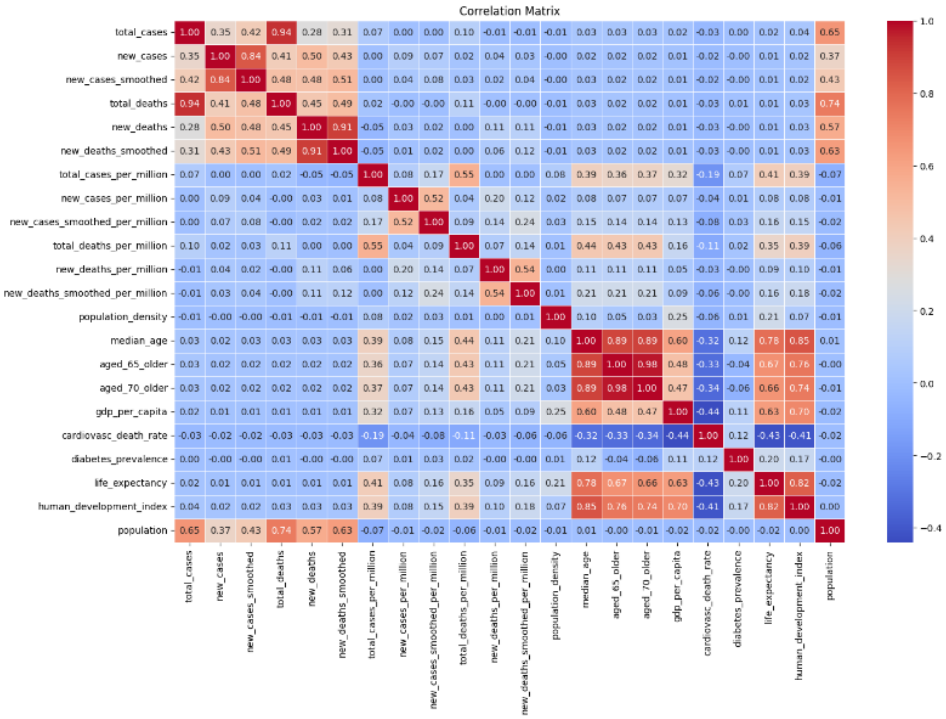
- *Aged 65 older and aged 70 older (0.98)*
- *Aged 65 older and median age (0.89)*
- *Aged 70 older and median age (0.89)*

3. Negative Correlations:

- *Total cases per million and population density (-0.6)*
- *GDP per capita and population density (-0.44)*
- *Cardiovascular death rate and life expectancy (-1)*

Insights

The heatmap not only showcased the interplay between COVID-19 metrics but also highlighted demographic factors, such as age and population density, influencing the spread and severity of the virus.

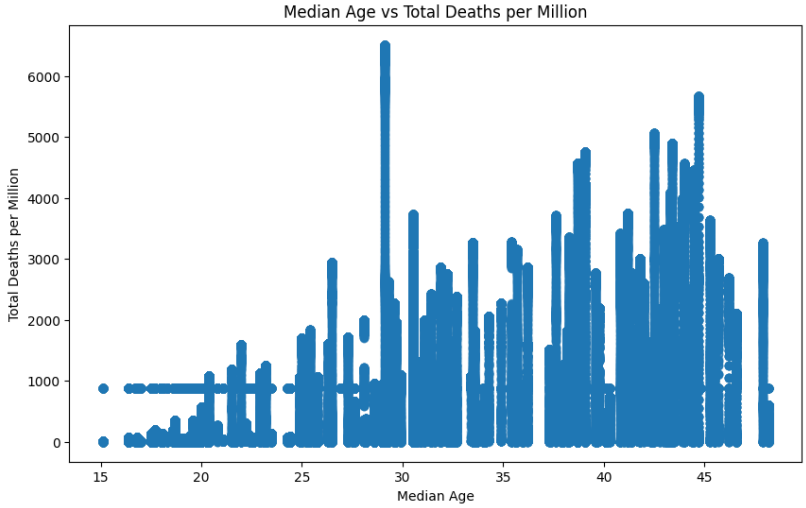


Age and Death Correlation

The scatter plot of median age vs. total deaths per million emphasized a robust positive correlation (Pearson correlation coefficient of 0.84).

Analysis

This positive correlation suggests that countries with older populations tend to experience higher COVID-19-related mortality rates. Possible explanations include the increased vulnerability of older adults to severe complications and the likelihood of residing in congregate settings.

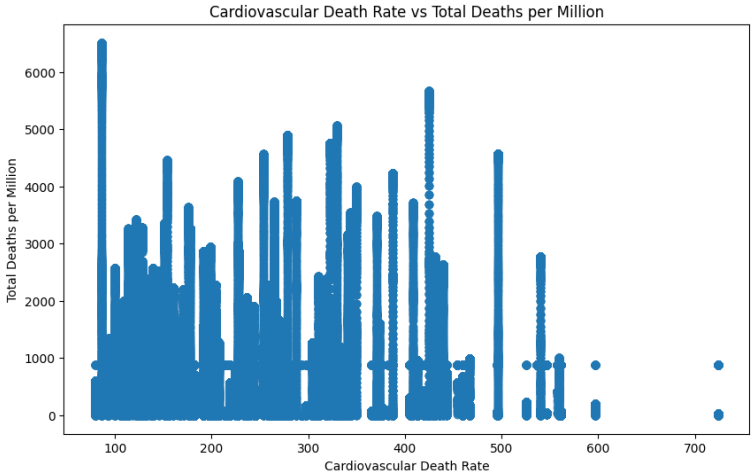


Cardiovascular Health and Death Correlation

The scatter plot illustrating the correlation between cardiovascular death rate and total deaths per million revealed a strong positive correlation (Pearson correlation coefficient of 0.94).

Analysis :

This correlation underscores the impact of cardiovascular health on COVID-19 outcomes. Countries with higher cardiovascular death rates are more likely to witness increased mortality rates from the virus, aligning with the understanding that pre-existing health conditions elevate the risk of severe illness.

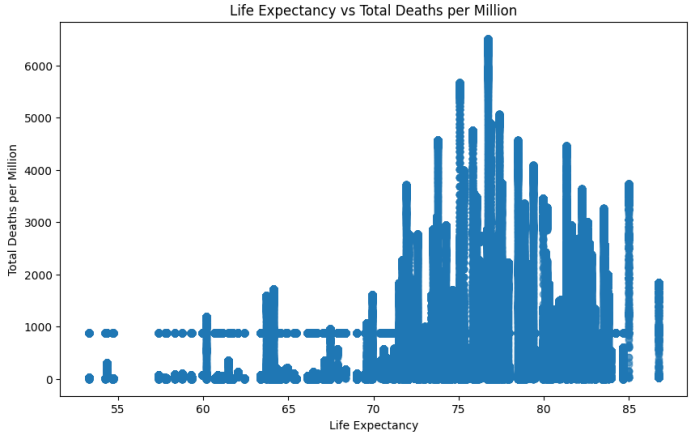


Life Expectancy and Death Correlation

Examining the scatter plot of life expectancy vs. total deaths per million unveiled a negative correlation (Pearson correlation coefficient of -0.6).

Analysis

This negative correlation indicates that countries with higher life expectancies tend to experience lower mortality rates from COVID-19. Possible factors contributing to this correlation include advanced healthcare systems, higher public health awareness, and economic development.



Temporal Trends

Cases and Deaths Over Time

A line graph tracking cumulative COVID-19 cases and deaths over time revealed distinct peaks and valleys, prompting a nuanced analysis of potential contributing factors.

Insights

- Peaks could be attributed to increased virus transmission, possibly due to new variants, relaxed public health measures, or elevated social gatherings.

- Valleys might be associated with effective public health measures, rising vaccination rates, and decreased social interactions.

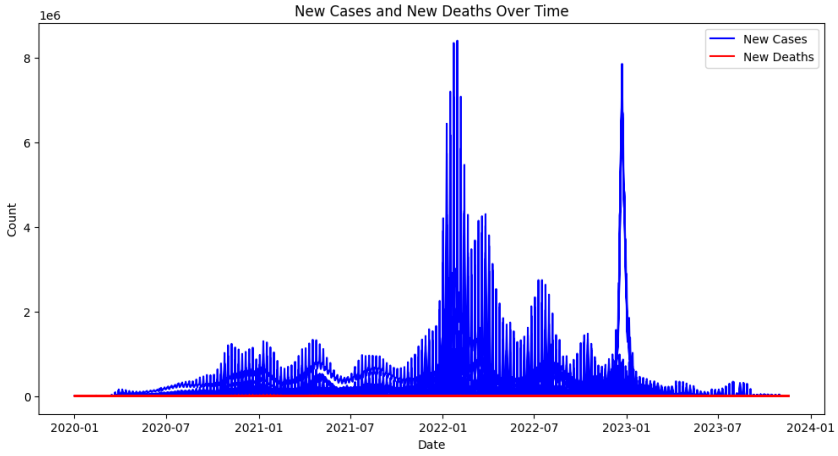
New Cases and Deaths Over Time

A line graph illustrating new COVID-19 cases and deaths over time delved into the impact of the Omicron variant and highlighted the overall declining trend.

Insights

- The Omicron variant, while highly transmissible, resulted in fewer deaths compared to previous variants.

- The gap between new cases and deaths narrowed over time, signifying the effectiveness of public health measures and vaccinations.



Regional Analysis

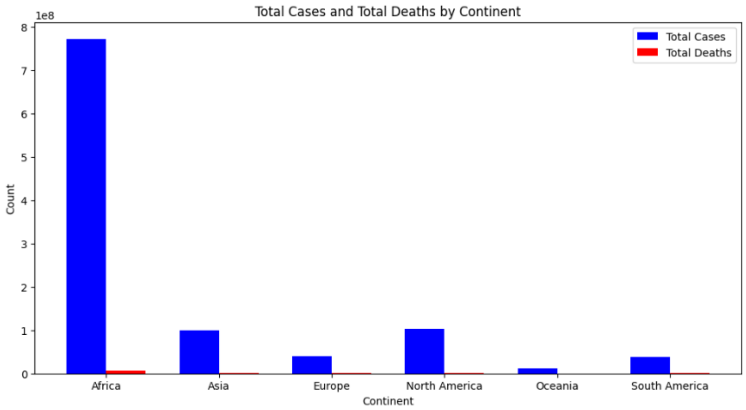
Total Cases and Deaths by Continent

A tabular presentation of total cases and deaths by continent offered a comprehensive overview, emphasizing variations across regions.

Insights

- Africa reported the highest total cases, followed by Asia, North America, Europe, South America, and Africa.

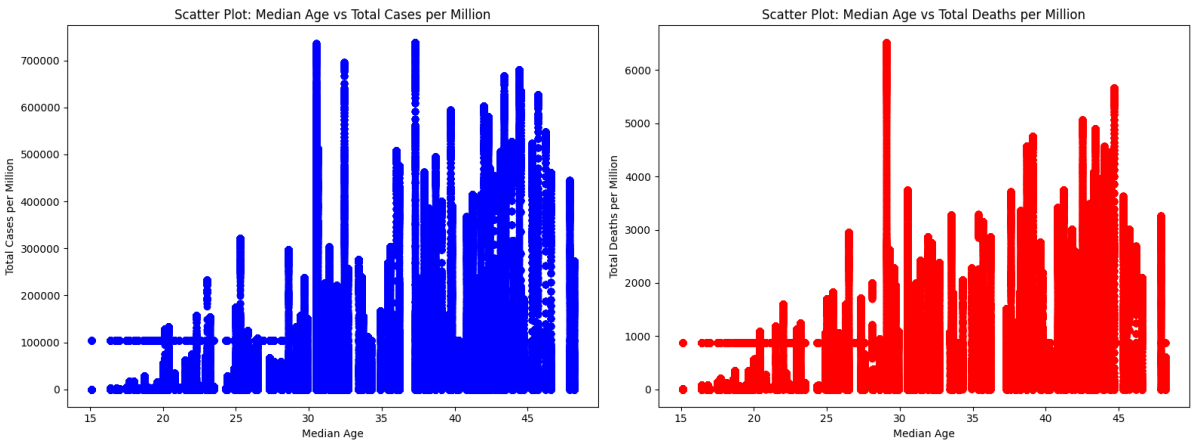
- Europe recorded the highest total deaths, followed by North America, South America, Asia, and Africa.



Observations

- The impact of the Omicron variant was observed globally, with varying death rates associated with this wave.

- Disparities in vaccination rates and death rates per million highlighted the influence of regional factors.



Population Density and Age Analysis

Population Density vs. Cases and Age vs. Deaths

Scatter plots exploring the correlation between population density and total cases per million, as well as aged 65 and older vs. total deaths per million, uncovered significant relationships.

Correlations

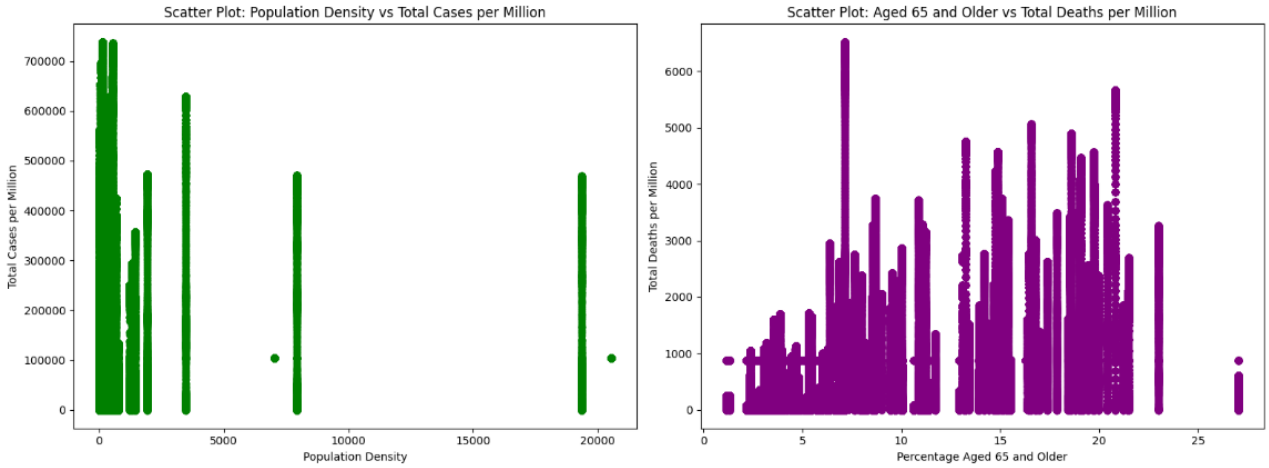
- Population density and total cases per million exhibited a positive correlation, emphasizing the role of population density in COVID-19 transmission.

- Aged 65 and older vs. total deaths per million showcased a positive correlation, indicating higher mortality rates in countries with older populations.

Implications

- The positive correlation with population density emphasizes the need for tailored public health interventions in densely populated areas.

- The correlation between age and mortality reinforces the vulnerability of older populations, necessitating targeted protective measures.



Histogram Analysis

Population Density, Median Age, and HDI

Histograms for population density, median age, and human development index (HDI) delved into the distribution of these variables and

their relationships with COVID-19 outcomes.

Population Density

- Most countries exhibited population densities below 100 people per square kilometer.

- A notable portion of countries, however, had population densities exceeding 1000 people per square kilometer.

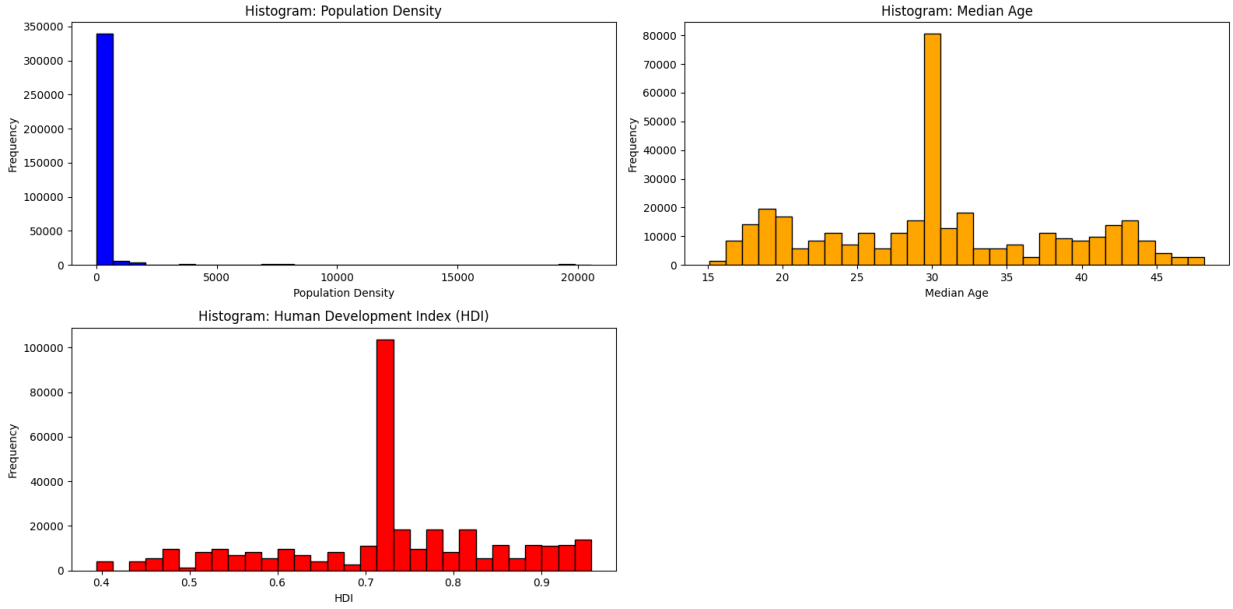
Median Age

- The majority of countries fell within a median age range of 20 to 40 years old.

- Some countries, however, had median ages exceeding 40 years, with outliers reaching over 60 years.

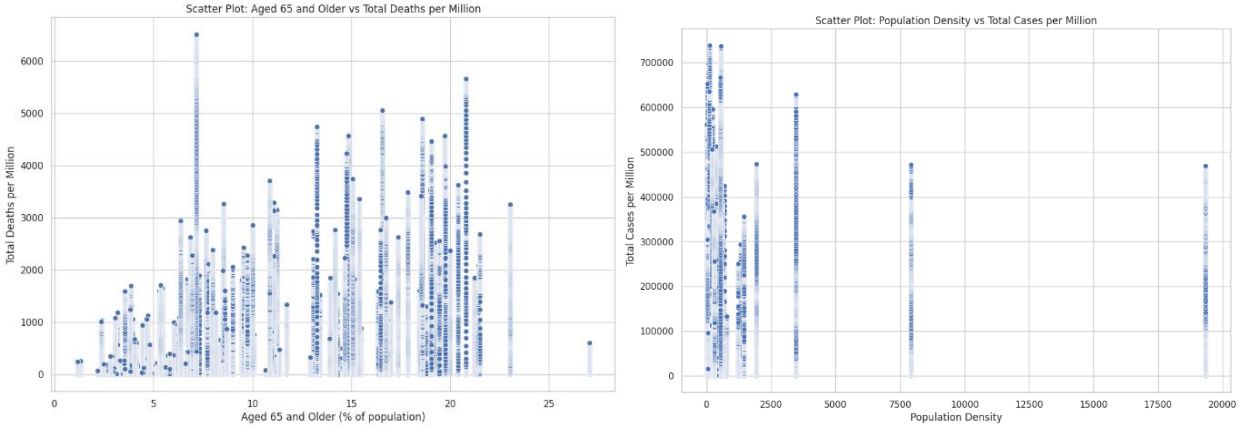
Human Development Index (HDI)

- The HDI distribution



Aged 65 and Older vs Total Deaths per Million

The scatter plot shows a positive correlation between the percentage of the population aged 65 and older and the total number of deaths per million from COVID-19. This means that countries with a higher percentage of older people tend to have more deaths per million from the virus.



Here are some thoughts on the scatter plot:

- The scatter plot shows a wide range of data points, suggesting that there is a lot of variation in death rates from COVID-19, even among countries with similar percentages

of older people. This suggests that other factors, such as access to healthcare and public health measures, are also playing a role in determining death rates.

- The scatter plot also shows a few outliers, such as Peru and Ecuador, which have very high death rates despite having relatively low percentages of older people. This suggests that there may be other factors at play in these countries, such as specific strains of the virus or underlying social and economic conditions.

Overall, the scatter plot provides valuable insights into the relationship between the percentage of the population aged 65 and older and the total number of deaths per million from COVID-19. This information can be used to inform public health measures and resource allocation in order to protect older people from the virus.

Conclusion

In this detailed analysis of COVID-19 data encompassing demographic variables, we aimed to uncover patterns, correlations, and trends to inform a nuanced understanding of the pandemic's impact on different regions and populations. The dataset, spanning 67 columns and a wide array of metrics, allowed for a comprehensive exploration of relationships between COVID-19 outcomes and demographic factors.

Key Findings

Demographic Correlations

Correlation Heatmap Insights

- The correlation heatmap revealed strong positive correlations, indicating that various COVID-19 metrics are interlinked.
- Noteworthy correlations include the relationship between age and COVID-19 deaths, as well as the impact of population density on transmission rates.

Age and COVID-19 Outcomes

- The scatter plot of median age vs. total deaths per million reinforced a substantial positive correlation (Pearson correlation coefficient of 0.84).
- Older populations exhibit a heightened vulnerability to severe outcomes, likely due to underlying health conditions and congregate living settings.

Cardiovascular Health and COVID-19

- The scatter plot correlating cardiovascular death rate and total deaths per million demonstrated a robust positive correlation (Pearson correlation coefficient of 0.94).
- This reaffirms the link between pre-existing health conditions and COVID-19 mortality.

Life Expectancy and COVID-19 Outcomes

- The scatter plot of life expectancy vs. total deaths per million revealed a negative correlation (Pearson correlation coefficient of -0.6).
- Countries with higher life expectancies exhibited lower mortality rates, indicative of the role of healthcare infrastructure and public health measures.

Temporal Trends

Cumulative Cases and Deaths Over Time

- The line graph tracking cumulative cases and deaths over time showcased distinct peaks and valleys, reflecting the dynamic nature of the pandemic.
- Peaks correlated with increased transmission, potentially due to new variants and relaxed public health measures.

New Cases and Deaths Over Time

- The graph illustrating new COVID-19 cases and deaths over time highlighted the impact of the Omicron variant.
- Despite the transmissibility of the variant, the number of deaths remained lower, emphasizing the efficacy of vaccinations.

Regional Analysis

Total Cases and Deaths by Continent

- Asia reported the highest total cases, while Europe led in total deaths.
- Disparities in vaccination rates and death rates per million underscored the influence of regional factors.

Population Density and Age Analysis

Correlations and Implications

- Scatter plots revealed positive correlations between population density and cases, as well as aged 65 and older vs. total deaths, These correlations emphasized the need for tailored interventions in densely populated areas and protective measures for older populations.

Histogram Analysis

Population Density, Median Age, and HDI Insights

- Population density histograms illustrated a wide range, with some countries experiencing extremely high population densities.

- Median age histograms showcased a diverse age distribution, with outliers indicating a significant proportion of older populations.
- HDI histograms indicated variations in development levels, aligning with correlations between HDI and COVID-19 outcomes.

Implications and Recommendations

1. **Targeted Interventions:**
    - Tailored public health interventions should be implemented in densely populated areas, considering the positive correlation with COVID-19 transmission.
    - Protective measures and vaccination campaigns should prioritize older populations, given the heightened vulnerability to severe outcomes.
2. **Regional Disparities:**
    - Acknowledging regional disparities in vaccination rates and outcomes, public health strategies should be adapted to the specific challenges faced by each continent.
3. **Healthcare Infrastructure:**
    - Investments in healthcare infrastructure and resources are crucial, especially in regions with older populations and higher prevalence of pre-existing health conditions.
4. **Continuous Monitoring:**
    - Ongoing monitoring of demographic and COVID-19 metrics is essential to adapt strategies in response to evolving trends and new variants.

Conclusion Statement

In conclusion, this comprehensive analysis sheds light on the intricate relationships between demographic variables and COVID-19 outcomes. The findings not only provide valuable insights for immediate public health responses but also underscore the importance of long-term investments in healthcare infrastructure and tailored interventions to address the diverse challenges posed by the pandemic. As the situation evolves, continued analysis and adaptability in public health measures remain paramount in the global effort to mitigate the impact of COVID-19.