## **Summary Report on COVID-19 Demographic Analysis**

### **Objectives**

The analysis aimed to assess the impact of COVID-19 vaccinations on hospitalization and mortality rates, with a particular focus on variations across age groups. The dataset included a variety of health metrics across different regions and demographics, enabling a detailed examination of how age and vaccination status influence COVID-19 outcomes.

### **Data Cleaning and Preparation Process**

1. **Handling Missing Values**:
   * **Critical Fields**: We observed high levels of missing values in fields critical to this analysis, such as ICU admissions, hospital admissions, and booster vaccination data.
   * **Imputation**: Median imputation was applied to hospitalization and mortality rates to preserve valuable data without distorting distributions. Less essential fields with substantial missing data were either excluded or noted as potential limitations.
2. **Outlier Detection and Treatment**:
   * **Outlier Identification**: Extreme values were found in hospitalization and mortality fields, which could skew analyses. For example, ICU patient counts reached as high as 28,891 while the mean was only 215.
   * **Outlier Treatment**: Values were capped at the 99th percentile for each hospitalization-related field, reducing the impact of extreme outliers.
3. **Data Validation**:
   * **Categorical Consistency**: To ensure consistent categorization, fields like age\_group (created based on the proportion of people aged 65+ as a proxy) and vaccination\_status were standardized.
   * **Categorization**: The population aged 65+ was segmented into groups (<5%, 5-10%, etc.) as a way to approximate age demographics. Vaccination status was categorized as “Unvaccinated,” “Partially Vaccinated,” or “Fully Vaccinated” based on the proportion of vaccinated individuals in each population.
4. **Rate Calculations**:
   * **Hospitalization and Mortality Rates**: Rates were calculated per 100,000 population for consistent comparisons across different regions and population sizes.

### **Analysis Results and Findings**

1. **Impact of Vaccination on Hospitalization and Mortality Rates**:
   * **Hospitalization Rates**: Fully vaccinated groups consistently had lower hospitalization rates across most age demographics, underscoring the effectiveness of vaccination in preventing severe cases. Partially vaccinated groups showed higher hospitalization rates than fully vaccinated groups.
   * **Mortality Rates**: Mortality rates also tended to be lower among fully vaccinated individuals, particularly in older populations, although in some cases, mortality rates in fully vaccinated groups were elevated, possibly reflecting underlying health conditions or regional factors.
2. **Age and Regional Differences**:
   * **Age Factor**: Mortality and hospitalization rates were notably higher in regions with larger populations aged 65+. Vaccination had a pronounced impact on reducing mortality rates in these age demographics, with fully vaccinated older populations experiencing the most significant reductions.
   * **Regional Variation**: North America and Europe exhibited higher mortality rates per 100,000 population compared to other regions. This may reflect differing levels of healthcare infrastructure, population density, and health policies across continents.
3. **Vaccination Status and Risk Reduction**:
   * The data indicates that full vaccination significantly reduces both hospitalization and mortality rates across age groups and regions, highlighting its role in mitigating severe COVID-19 outcomes. This trend is most visible in older populations, where the risk of severe outcomes is highest.

### **Conclusion**

This analysis demonstrates that COVID-19 vaccinations effectively reduce severe outcomes like hospitalization and death, with fully vaccinated groups consistently showing lower rates. Older populations and regions with higher mortality rates benefit most from vaccination, reinforcing public health efforts to prioritize vaccine access for vulnerable groups.

The data cleaning and preparation processes were essential in ensuring data integrity and reliability, allowing for more accurate insights into COVID-19 trends by age and vaccination status. This study supports the importance of vaccination campaigns, especially in high-risk demographics, to mitigate the impact of COVID-19.