**Insights into Global Vaccination Efforts Against Covid-19: An Exploratory Analysis**

The dataset selected for the exploratory data analysis is titled "Covid-19 Global Vaccination." The data is stored in a CSV file named country\_vaccinations.csv.

**Outcome of EDA**

In the initial data cleaning phase, it was found that the United Kingdom had duplicate entries one for the entire UK and separate entries for Scotland, England, Wales, and Northern Ireland. To obtain a unified entry for the United Kingdom, the data for these individual regions was summed, and the separate entries were removed.

After plotting a histogram based on the total number of vaccinated individuals, China initially appeared to lead. However, in subsequent analyses following data cleaning, China no longer appeared, as it was found that there was no data available on the total number of people vaccinated in China. The countries leading in vaccination efforts consistently across all histograms are the United States, India, the United Kingdom, Brazil, and Germany. In terms of vaccine types, Pfizer is widely used in the United States, France, and Germany, while AstraZeneca is more commonly administered in India compared to France or Germany.

**What do you feel was missed during the analysis?**

An error during the data cleaning process led to discrepancies in certain plots, like scatter plots, which did not align with expected outcomes. This oversight highlights the importance of careful data preparation to ensure accurate analysis results.

**Were there any variables you felt could have helped in the analysis?**

I couldn’t identify any additional variables that would have enhanced the analysis, as all necessary columns are already included in the dataset. This comprehensive data coverage has streamlined the analytical process, ensuring a focus on meaningful insights.

**Were there any assumptions made you felt were incorrect?**

While displaying histograms based on the total number of people immunized and other variables, it was initially assumed that there were no outliers in the dataset. However, some outliers became apparent upon examining the scatter plots. This discovery emphasizes the value of using multiple visualization techniques to accurately detect anomalies in data.

**What challenges did you face, what did you not fully understand?**

I encountered difficulties when creating scatter plots due to an inability to scale the X and Y axes with the correct units, which might have helped in accurately representing the data. Additionally, I struggled to present the scatter plots effectively, as I couldn’t remove the jitter caused by improper unit conversion of the variables. This experience highlighted the importance of precise scaling and unit management in data visualization for clarity.