**Hi everyone!**

My capstone project focused on **human migration**, and one of the biggest challenges was working with a complex dataset. The data came in a CSV format that required extensive cleaning

Before diving into the cleaning process, I had to first **understand the structure of the data**—what information it contained, what was relevant to my analysis, and what needed to be removed. So I had to spend a lot of time in cleaning/ removing unnecessary rows / col/ missing val. Reformatting it so it was readable in python  
Once the data was cleaned, the analysis itself wasn’t straightforward. I had to explore trends, visualize patterns, and make sense of migration flows across different regions.

**Hi everyone!**

My capstone project focused on **human migration**, and one of the biggest challenges was working with a complex dataset. The data came in a CSV format that required extensive cleaning

new column showing the probability of mass killing as a **percentage chance** rather than "1 in X" format.

imr\_per\_1000 that shows the **actual infant mortality rate** (IMR) after undoing the square root transformation. If the **IMR per 1000** is **56**, it means that in that country, **56 infants die for every 1,000 live births**.