CENTRE FOR EPIDEMIOLOGICAL MODELLING AND ANALYSIS

STAGE 1 OF THE RECRUITMENT PROCESS FOR THE DATA SCIENCE INTERNSHIP.

DATA SCIENCE AND STATISTICAL ANALYSIS ON HIV CASES, MULTIDIMENSIONAL POVERTY, UNDER-FIVE MORTALITY RATE AND NEONATAL MORTALITY RATE


DANCAN MBAABU

Question 1 Solution:

I was provided with a dataset from the World Health Organization (WHO) Global Observatory, containing data on people living with HIV at the country level from 2000 to 2023.

i. Using this dataset, I Created a visualization that shows the trend of HIV cases in the countries that contribute to 75% of the global burden

Figure 1: HIV Trend in Countries Contributing 75% of Regional Burden
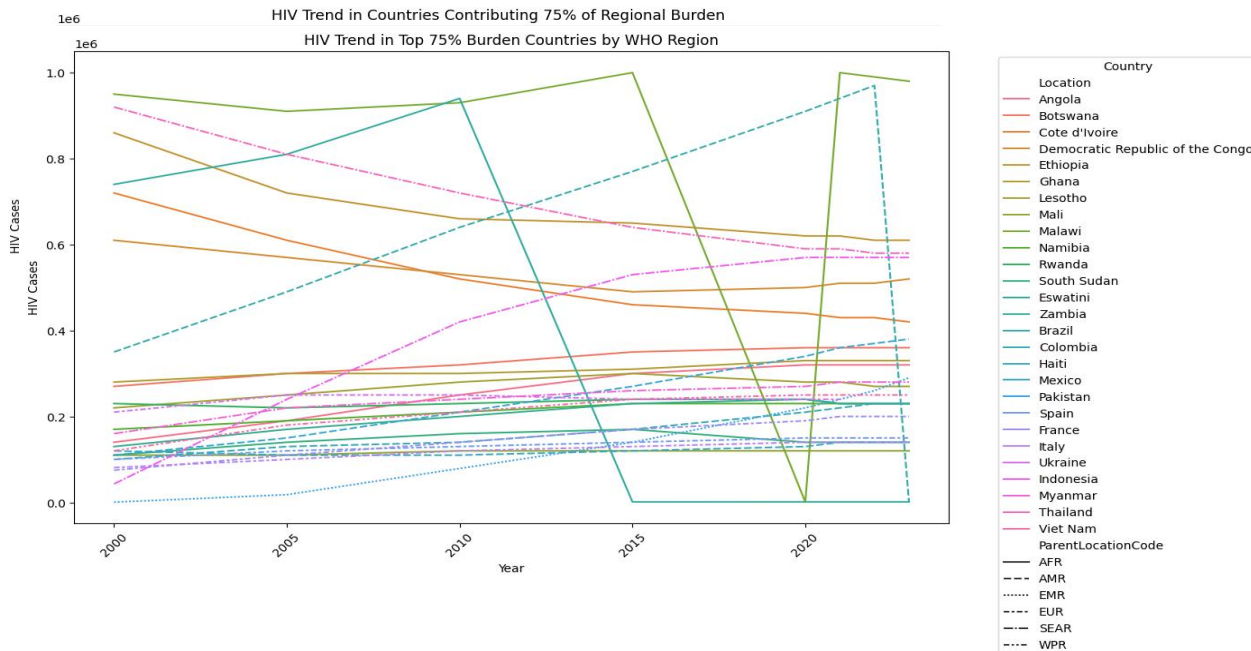
The graph shows HIV trends across countries contributing 75% of the global burden from 2000-2022. South Africa (likely the highest pink line) shows significant decline, while some nations like Brazil show substantial increases. Around 2015, unusual drop patterns appear in several countries, followed by recovery. Overall, there are contrasting trends with some countries showing improvement while others face worsening epidemics.

ii. A visualization that displays the trend of HIV cases in the countries contributing to 75% of the burden within each WHO region (column called ParentLocationCode contains the WHO regions)

**Figure 1: HIV Trend in Top 75% Burden Countries by WHO Region**
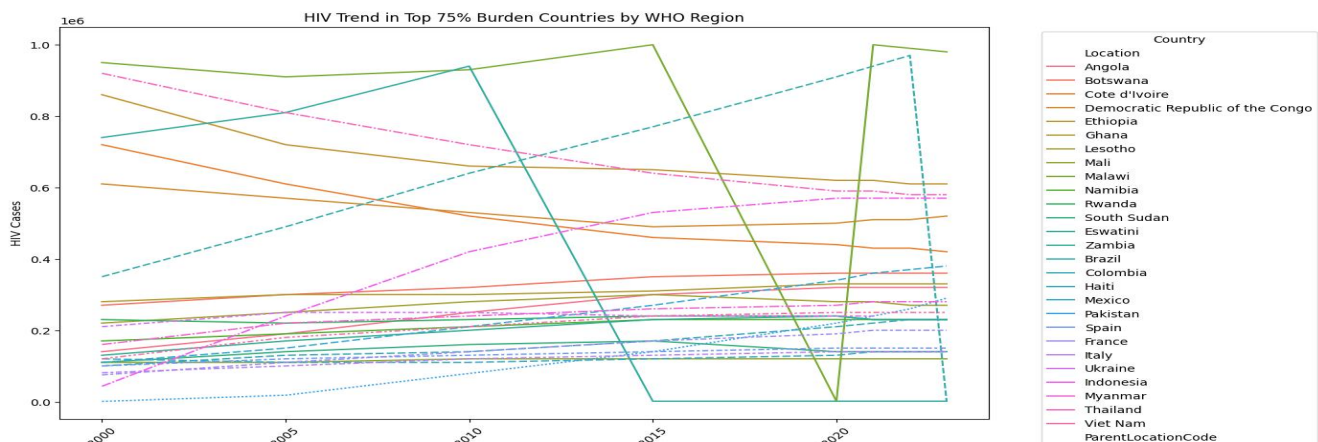
I Merged this dataset with the HIV data and analysed the relationship between people living with HIV and multidimensional poverty, and the individual factors that contribute to the ratio. I accounted for the random effects (country, year).

Figure 3: Relative Importance of Factors Contributing to HIV Prevalence
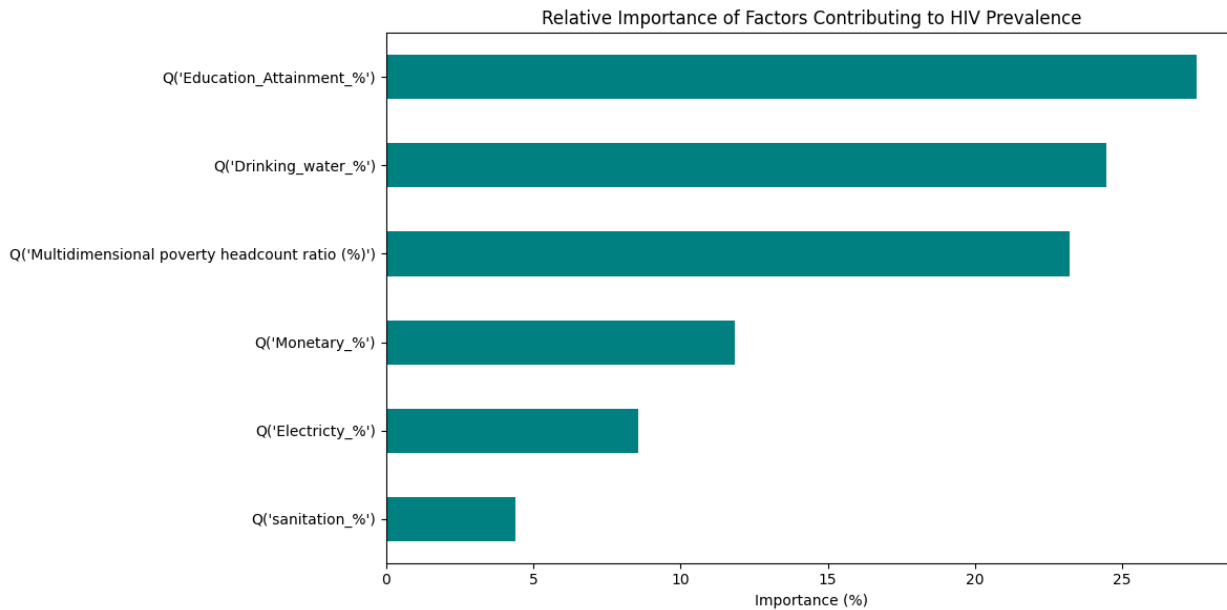
From the Figure 3 above, education attainment appears to be the most significant factor contributing to HIV prevalence (~27%), followed by access to drinking water (~24%) and multidimensional poverty headcount ratio (~22%). Monetary factors (~12%), electricity access (~9%), and sanitation (~5%) have comparatively less importance. This suggests that interventions focusing on education and clean water access might be most effective in addressing HIV prevalence when accounting for country and year random effects.

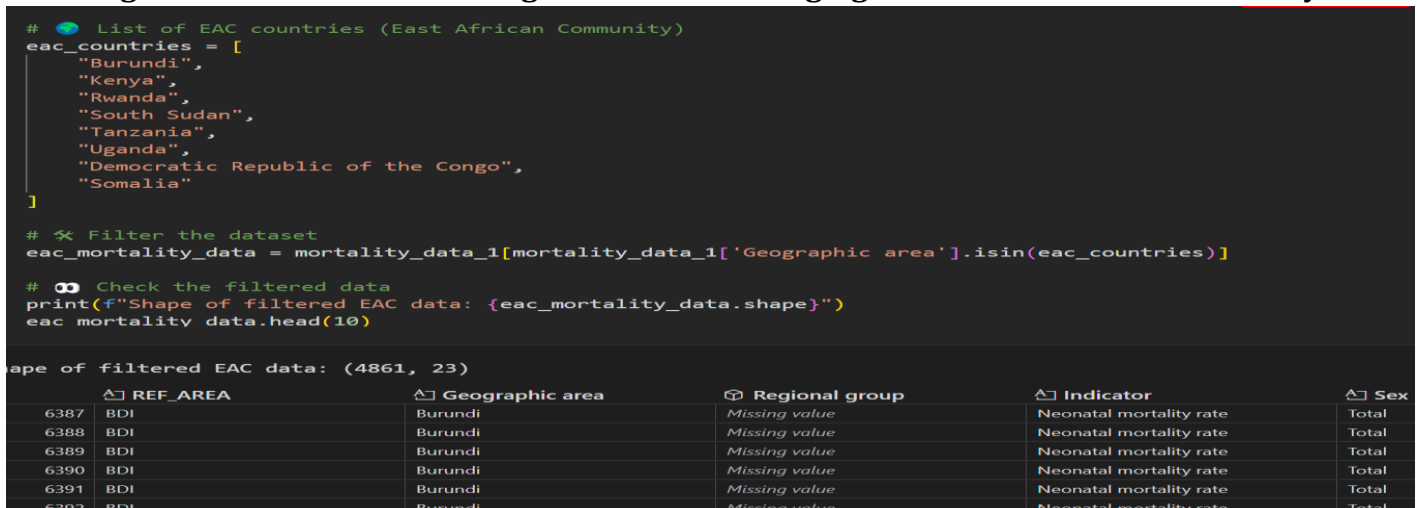**Figure 4: Correlations Between HIV and Poverty indicators**
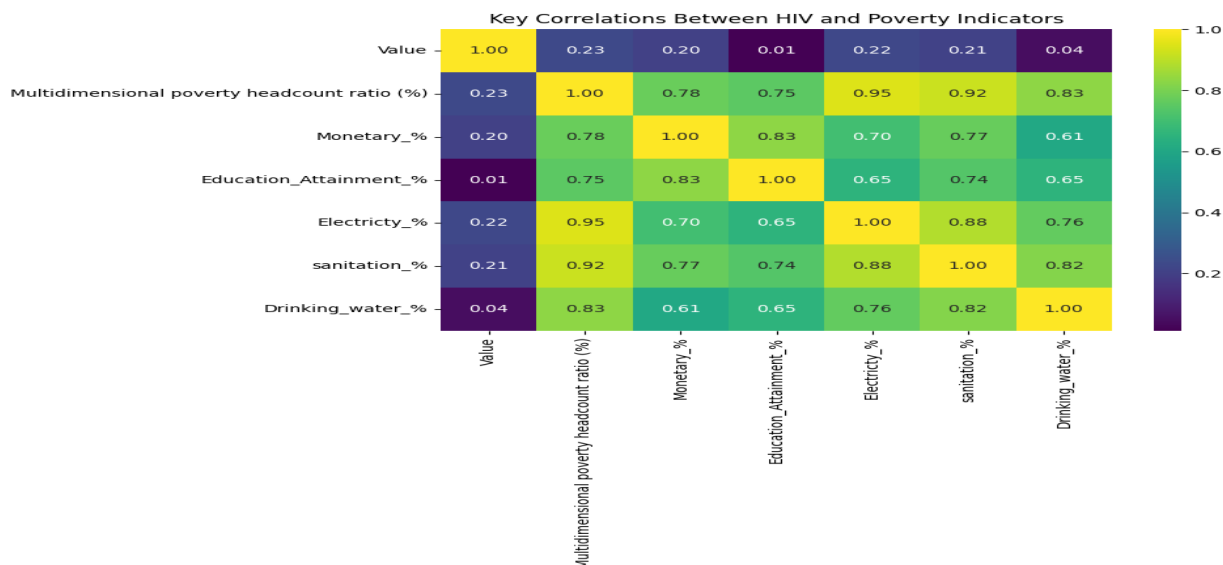
The heatmap shows moderate correlations between HIV prevalence and poverty, electricity, and sanitation. Education and water access, though previously ranked important, have weak direct correlations. Strong intercorrelations exist among poverty indicators. This suggests that while education and water access are key in multivariate models, their direct effects are limited.

Question 2 Solution:

Provided with data on the under-five mortality rate and neonatal mortality rate for the African region, which has been downloaded from the UN Inter-agency Group for Child Mortality Estimation. The task was to

a) Filter data for the eight countries belonging to the East African Community. Filtered The data According to the instruction.

Figure 5: Filter data for the eight countries belonging to the East African Community.

```
#  🌍 List of EAC countries (East African Community)
eac_countries = [
    "Burundi",
    "Kenya",
    "Rwanda",
    "South Sudan",
    "Tanzania",
    "Uganda",
    "Democratic Republic of the Congo",
    "Somalia"
]

#  🔧 Filter the dataset
eac_mortality_data = mortality_data_1[mortality_data_1['Geographic area'].isin(eac_countries)]

#  👓 Check the filtered data
print(f"Shape of filtered EAC data: {eac_mortality_data.shape}")
eac_mortality_data.head(10)
```

Shape of filtered EAC data: (4861, 23)

|	REF_AREA	Geographic area	Regional group	Indicator	Sex
6387	BDI	Burundi	Missing value	Neonatal mortality rate	Total
6388	BDI	Burundi	Missing value	Neonatal mortality rate	Total
6389	BDI	Burundi	Missing value	Neonatal mortality rate	Total
6390	BDI	Burundi	Missing value	Neonatal mortality rate	Total
6391	BDI	Burundi	Missing value	Neonatal mortality rate	Total
6392	BDI	Burundi	Missing value	Neonatal mortality rate	Total

b) Visualize the latest estimate of each indicator at the country level using shapefiles,

Figure 6: Visualization of the latest estimate of each indicator at the country level
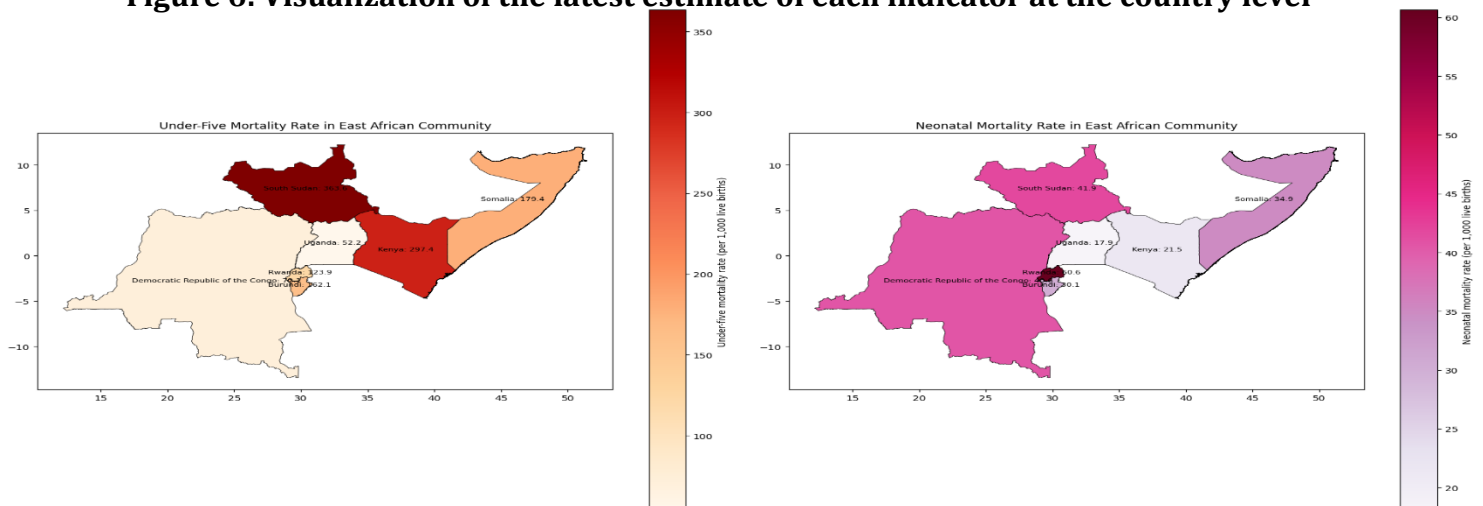
The average trends in the mortality rates over time (plot the average trend line and add the points in the graphic for the country level estimates for each indicator. Expected two plots).

Figure 7: The average trends in the mortality rates over time
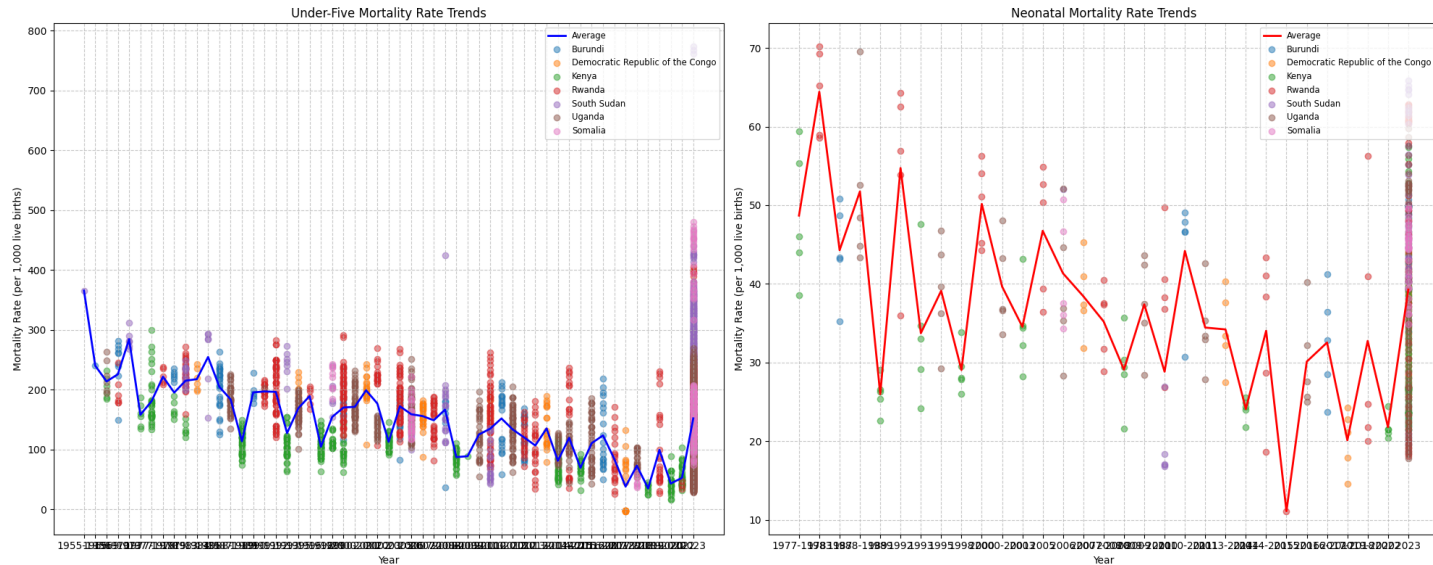
Based on thevvisualizations, the Country with highest under-five mortality rate is South Sudan (363.6 per 1,000 live births, 2023) while; the Country with highest neonatal mortality rate is Rwanda (60.6 per 1,000 live births, 2023)