**Data Analytics Framework for Analyzing Annual HIV deaths and Government Health Spending of total government spending (%)**

**1. Introduction**

This framework provides a structured approach for analyzing the incidence of HIV death and Government Health Spending of total government spending (%) in Countries with the in Sub-Sahara region between 2002 - 2022. The objective is to assess if there is a relationship between the budget allocated to health and the proportion of HIV related deaths

**2. Objectives**

* What has been the trend of the annual HIV deaths within the countries of interest?
* How has the Gov’t Health Spending evolved from 2002 – 2022?
* How does he Gov’t Health Spending behave in relation to the annual HIV deaths?

**3. Data Sources**

The analysis will use multiple data sources, including:

* [Gapminder](https://www.gapminder.org/data/):
  + Annual HIV deaths
  + Population
  + Government health spending of total government spending (%)

**4. Data Collection & Processing**

* **Data Extraction**: Gathering structured and unstructured data from different sources.
* **Data Cleaning**: Removing duplicates, handling missing values, and standardizing formats.
* **Data Integration**: Merging datasets for TB and HIV cases to facilitate comparative analysis.
* **Data Anonymization**: Ensuring patient confidentiality in compliance with ethical guidelines.

**5. Methodology**

**5.1 Descriptive Analysis**

* **Incidence Rate Calculation**:
  + TB Incidence Rate = (New TB Cases / Population at Risk) \* 100,000
  + HIV Incidence Rate = (New HIV Cases / Population at Risk) \* 100,000
* **Demographic Analysis**:
  + Age and sex distribution of TB and HIV cases.
  + Regional disparities in incidence rates.
* **Trend Analysis**:
  + Time series visualization of TB and HIV cases over the years.

**5.2 Inferential Analysis**

* **Regression Models**:
  + Logistic regression for predicting the likelihood of co-infection.
  + Poisson regression for analyzing incidence rate determinants.
* **Multilevel Modeling**:
  + Assessing the influence of geographic and socioeconomic factors.
* **Spatial Analysis**:
  + Mapping TB and HIV hotspots using GIS technology.
* **Survival Analysis**:
  + Estimating time to co-infection and mortality risk factors.

**6. Co-Infection Analysis**

* **Prevalence of TB-HIV Co-Infection**:
  + Percentage of TB patients who are HIV-positive.
* **Risk Factor Assessment**:
  + Immunosuppression, socio-economic status, and healthcare access.
* **Treatment Outcomes**:
  + Evaluating the effectiveness of TB and HIV treatment programs.

**7. Data Visualization & Reporting**

* **Dashboards**:
  + Interactive dashboards displaying incidence trends, geographical distribution, and patient demographics.
* **Infographics & Reports**:
  + Policy briefs for stakeholders.
  + Research papers for academic dissemination.

**8. Ethical Considerations & Data Governance**

* **Data Privacy Compliance**:
  + Adhering to GDPR and national data protection laws.
* **Informed Consent**:
  + Ensuring patient data is used ethically.
* **Bias Mitigation**:
  + Addressing data gaps and ensuring representation in the analysis.

**9. Limitations & Challenges**

* **Data Gaps**: Missing values in surveillance datasets.
* **Underreporting**: Challenges in capturing all incident cases.
* **Data Harmonization**: Variability in reporting standards across regions.

**10. Conclusion & Recommendations**

* Strengthen data integration across health information systems.
* Enhance surveillance to improve case detection.
* Develop targeted interventions for high-risk populations.
* Use predictive analytics to forecast future incidence trends.

This framework serves as a guide for conducting robust analyses on TB and HIV incident cases, ultimately contributing to improved disease control efforts and public health policies.