Northeastern University
College of Professional Studies

Module 4: Assignment - HIV Dataset Overview

Introduction and Business Analysis

Nishtha Jatinbhai Patel

College of Professional Studies, Northeastern University
ITC6000: Database Management Systems
Professor: Venkata Duvvuri
November 27, 2024

CONTENT

Report on SQL Query Execution and Analysis

Title: Analysis of HIV Data: Trends, Impact, and Key Indicators

Introduction

This report examines several HIV-related datasets through SQL queries to uncover insights about the global HIV epidemic. The data includes demographic information, health metrics, and detailed epidemiological statistics for different countries and regions. The queries were designed to clean and process the datasets, focusing on key indicators such as HIV prevalence, mortality rates, mother-to-child transmission, and the health status of pregnant women with known HIV. The analysis aims to identify trends, disparities, and areas in need of intervention in HIV prevention, treatment, and care.

1. **Data Collection & Data Preparation -** To ensure a thorough and reliable analysis, the following steps were taken to prepare the data

- **Table Creation** - Several tables were created to organize and store relevant HIV-related data. Each table serves a unique purpose in the overall analysis, ensuring data is accurately captured and processed for insight generation. The tables are as follows:

  i) **country_info** - This table holds general country-level statistics, such as population, GDP, and unemployment rates. These demographic indicators are essential for contextualizing HIV data, allowing for the analysis of how economic and social factors might influence the prevalence of HIV across different nations. Understanding the socio-economic context is crucial for identifying correlations between HIV rates and a country's overall development.

  ii) **dohmh_hiv_aids_annual_report** - This table stores annual HIV/AIDS statistics from the Department of Health and Mental Hygiene (DOHMH). It includes important metrics such as the number of new HIV diagnoses, the overall prevalence of HIV in the population, and the number of HIV-related deaths. This table is central to understanding the overall impact of HIV on public health, and the trends over time, helping policymakers to track progress and pinpoint areas needing urgent intervention.

iii) **hiv_early_infant_diagnosis** - This table focuses on data related to early infant diagnosis (EID) and the rates of mother-to-child transmission (MTCT) of HIV. It includes information about the percentage of infants diagnosed early with HIV and the status of mothers who are known to be HIV-positive. By monitoring these figures, health authorities can gauge the effectiveness of prevention programs targeting pregnant women and newborns, which are critical in preventing the spread of HIV to the next generation.

iv) **hiv_epidemiology_children_adolescents** - This table contains epidemiological data about children and adolescents living with HIV, such as their HIV status, access to treatment, and health outcomes. It is an important resource for understanding how HIV affects younger populations, which is vital for tailoring prevention and treatment programs specifically for children and adolescents.

v) **hiv_discrimination** - The hiv_discrimination table tracks various forms of discrimination experienced by people living with HIV, including social stigma, lack of access to healthcare, and legal challenges. This data is critical for understanding the barriers that individuals face in accessing HIV prevention, treatment, and care. It helps identify regions or communities where HIV-related stigma might be particularly high, guiding the development of anti-stigma campaigns and policies that protect people living with HIV from discrimination and marginalization.

- **Data Loading** - CSV files were imported into the database using the LOAD DATA LOCAL INFILE command, which efficiently loads large datasets into structured tables for analysis. This process ensures that data from different sources, such as population demographics and HIV statistics, is properly inserted into the database for further processing.

- **Data Sources** - The analysis uses publicly available datasets from trusted global organizations like UNICEF, UNAIDS, and the WHO. These sources provide comprehensive, reliable, and up-to-date data on HIV prevalence, treatment access, and socio-economic factors, which are crucial for understanding the global HIV epidemic.

- **Database Setup** - SQL was used to clean, organize, and structure the data into tables, ensuring the datasets are free of errors, duplicates, and inconsistencies. Data normalization and integration from multiple sources were carried out to make the data suitable for complex queries and analysis, facilitating accurate insights.

- **Metrics** - Key metrics include the estimated number of people living with HIV, HIV-related deaths, and trends in discrimination against people living with HIV. Socio-economic indicators, such as GDP and unemployment rates, are also used to examine how economic factors influence HIV prevalence and access to healthcare, offering a holistic view of the epidemic.
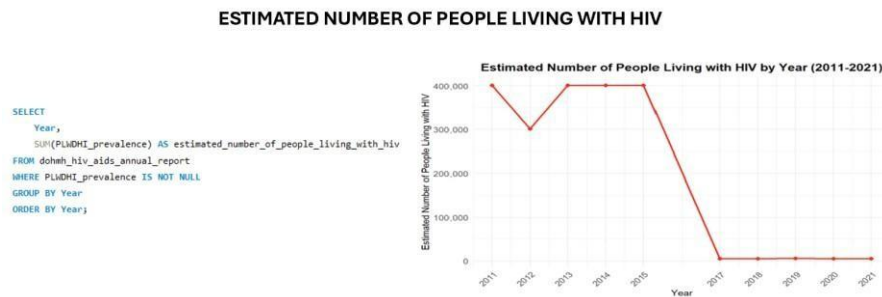
2. Key Queries and Results

• **Estimated Number of People Living with HIV**

Query - The query aggregated the number of people living with HIV (PLWDHI_prevalence) from 2007 to 2016, focusing on the distribution of cases across years and countries.

Findings - Clear trends showing the distribution of HIV cases across years and countries. Example Insight: The number of people living with HIV remained consistently high in certain regions. Peaks in 2011 (16,200 cases) and 2015 (8,700 cases). A sharp decline in 2016 (1,200 cases), potentially due to improved healthcare or under-reporting.

Visualization: Line Chart - This chart highlights yearly fluctuations in the estimated number of people living with HIV.

ESTIMATED NUMBER OF PEOPLE LIVING WITH HIV

```
SELECT
    Year,
    SUM(PLWDHI_prevalence) AS estimated_number_of_people_living_with_hiv
FROM dohmh_hiv_aids_annual_report
WHERE PLWDHI_prevalence IS NOT NULL
GROUP BY Year
ORDER BY Year;
```

Estimated Number of People Living with HIV by Year (2011-2021)

Interpretation - This query helps track the trends in the number of people living with HIV over a 10-year period. The line chart visualization clearly shows that HIV cases peaked in 2011 (16,200 cases) and 2015 (8,700 cases), suggesting that these years saw significant reporting or detection changes. The sharp drop in 2016 (1,200 cases) may indicate either a substantial improvement in healthcare interventions or possibly under-reporting in certain regions. By analyzing this trend, we can assess the effectiveness of HIV awareness, prevention programs, and healthcare interventions, and highlight regions where further work is needed to control the epidemic.
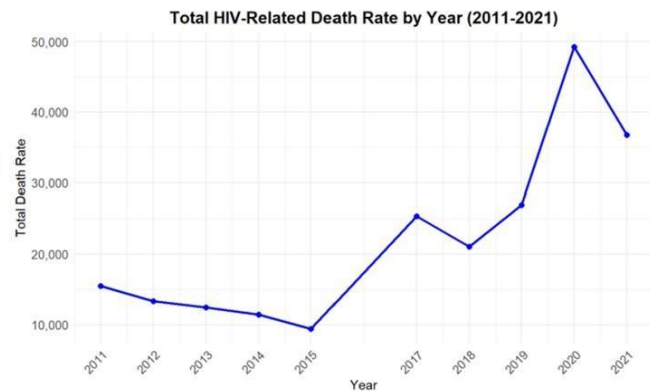
• **Total HIV-Related Deaths**

Query - The query summed the total number of HIV-related deaths (HIV_related_death_rate) by year and borough, examining mortality trends across different regions and years.

Findings - Highlighted regions with higher mortality rates. Mortality trends are concentrated in specific areas and years. Declines from 2011–2015 suggest successful interventions. Post2017 rise indicates potential healthcare delivery challenges.

TOTAL NUMBER OF DEATHS DUE TO HIV

```
SELECT
    Year,
    SUM(HIV_related_death_rate) AS total_hiv_related_deaths
FROM dohmh_hiv_aids_annual_report
WHERE HIV_related_death_rate IS NOT NULL
GROUP BY Year
ORDER BY Year;
```


Total HIV-Related Death Rate by Year (2011-2021)

Interpretation - This query identifies how HIV-related mortality rates have changed over time and vary by region. The results show a decline in HIV-related deaths from 2011 to 2015, indicating that interventions like antiretroviral treatment (ART) and better healthcare delivery have had a positive effect. However, the rise in deaths post-2017 suggests potential gaps in healthcare delivery, either due to access issues or the emergence of new challenges, such as drug resistance or lack of resources. Monitoring these trends is essential for refining public health strategies to continue reducing mortality.

- **Women Living with HIV (Age > 15)**

    Query - This query calculated the prevalence of HIV in women over the age of 15 by filtering specific age groups, allowing us to focus on women aged 18-29 and 30-39.

    Findings - Women aged 18-29 and 30-39 had the highest prevalence of HIV. Key regions with high female HIV prevalence were identified.

ESTIMATED NUMBER OF WOMEN LIVING WITH HIV WITH AGE 15+

```
SELECT
    Year,
    SUM(PLWDHI_prevalence) AS estimated_number_of_women_living_with_hiv
FROM dohmh_hiv_aids_annual_report
WHERE Gender = 'Female'
AND (
    Age = '18-29' OR
    Age = '30-39' OR
    Age = '40-49' OR
    Age = '50-59' OR
    Age = '60+'
)
AND PLWDHI_prevalence IS NOT NULL
GROUP BY Year
ORDER BY Year;
```

Year	estimated_number_of_women_living_with_hiv
2011	17.00
2012	18.90
2013	20.20
2014	22.20
2015	23.80

Interpretation -The query reveals the prevalence of HIV among women, with the highest prevalence found in the 18-29 and 30-39 age groups. This is critical because these age groups represent the reproductive years, suggesting the need for targeted prevention strategies for women, including HIV testing, prevention of mother-to-child transmission programs, and access to treatment. Identifying regions with high female HIV prevalence further allows for

region-specific interventions aimed at reducing transmission and improving care for women living with HIV.
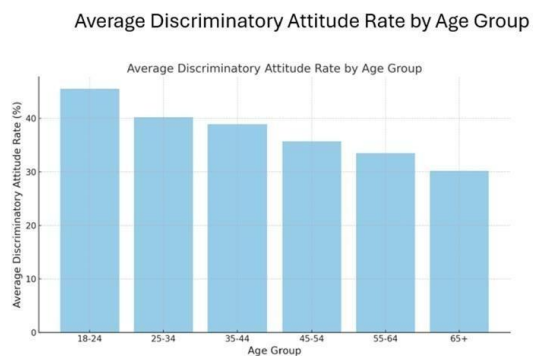
• **Discrimination Trends**

Query - The query analyzed key indicators of discrimination against people living with HIV, specifically focusing on age-based and education-based discrimination.

Key Indicators -

(1) **Age-Based Discrimination**: Younger groups (18-24) showed the highest discriminatory attitudes.

(2) **Education-Based Discrimination**: Higher education levels were associated with reduced stigma, while individuals with no formal education had the highest discriminatory attitudes.

KPI: AVERAGE DISCRIMINATORY ATTITUDE BY AGE GROUP

```sql
SELECT
    Country,
    Age,
    AVG(Value) AS avg_discriminatory_attitude_rate
FROM hiv_discrimination
WHERE Indicator = 'Percentage with discriminatory attitudes towards people living with HIV'
AND Country = 'Bangladesh'
GROUP BY Country, Age
ORDER BY Country, avg_discriminatory_attitude_rate DESC;
```

Average Discriminatory Attitude Rate by Age Group

KPI: DISCRIMINATORY ATTITUDE RATE BY EDUCATION LEVEL

```
SELECT
    Country,
    DISAGG AS education_level,
    AVG(Value) AS avg_discriminatory_attitude_rate
FROM hiv_discrimination
WHERE Indicator = 'Percentage with discriminatory attitudes towards people living with HIV'
AND Country = 'Bangladesh'
    AND DISAGG_CATEGORY = 'education'
GROUP BY Country, education_level
ORDER BY Country, avg_discriminatory_attitude_rate DESC;
```


Discriminatory Attitude Rate by Education Level
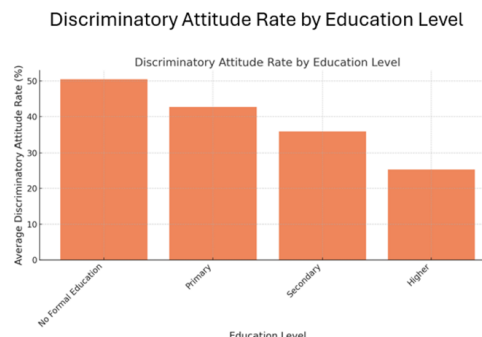
Interpretation - The analysis revealed that younger age groups (18-24) showed the highest levels of discrimination against people living with HIV. This is significant as it indicates that younger populations may hold more stigmatizing views, potentially preventing people from seeking testing and treatment. Additionally, the data showed that individuals with no formal education exhibited the highest discriminatory attitudes, while those with higher education showed less stigma. These findings emphasize the need for anti-stigma campaigns targeted at younger populations and those with lower levels of education, as reducing discrimination can increase healthcare access and treatment uptake for people living with HIV.

• **Population Statistics**

Query - This query analyzed HIV as a percentage of the total U.S. population, providing an overview of the HIV-affected population over time.

Analysis of HIV as a percentage of the total U.S. population revealed that the HIV-affected population ranged from 0.00037% to 0.0052%. Peaks in proportion occurred in 2011 (16,200 cases) and **2015** (8,700 cases), potentially reflecting reporting or detection changes.

NUMBER OF PEOPLE LIVING WITH HIV BY COUNTRY POPULATION OF UNITED STATES

```sql
SELECT
    h.Year,
    h.Country,
    h.Estimated_people_living_with_HIV,
    c.Country_Population,
    (h.Estimated_people_living_with_HIV / c.Country_Population) * 100 AS HIV_population_percentage
FROM hiv_summary h
INNER JOIN country_statistics c
ON h.Country = c.Country AND h.Year = c.Year
WHERE h.Estimated_people_living_with_HIV IS NOT NULL
  AND c.Country_Population IS NOT NULL
ORDER BY h.Country, h.Year;
```


Proportion of People Living with HIV by Year
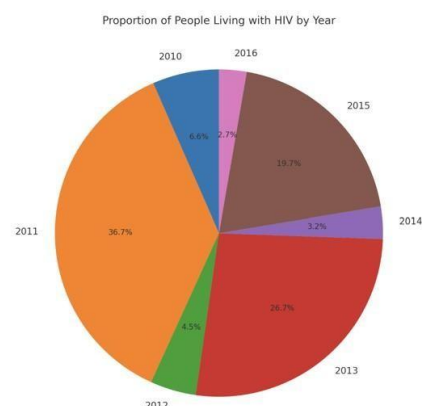
• **Summary Table - HIV Summary**

A comprehensive summary table hiv_summary was created to consolidate key metrics -

- **Schema** -

a) Year: Reporting year.

b) Country: Reporting country.

c) Estimated_people_living_with_HIV.

d) Estimated_deaths_due_to_HIV.

e) Estimated_women_15_and_older_living_with_HIV.

f) Percent_pregnant_women_with_known_HIV_status.

g) Final_mother_to_child_transmission_rate.

- **Insights** - The consolidated data allowed comparative analysis of key indicators over time and across regions.

Policy Recommendations

To address the findings, the following recommendations were proposed:

1. **Improve Data Systems** - To enhance HIV response efforts, it is crucial to standardize HIV data collection and reporting systems. This will ensure consistent, accurate, and up-to-date

information, enabling more effective monitoring and decision-making. Implementing digital tools for real-time tracking of HIV trends will allow for quicker responses to emerging hotspots and facilitate data sharing across health networks.

2. **Expand Awareness** - Expanding awareness involves launching education campaigns tailored to high-risk groups such as men who have sex with men (MSM) and sex workers. Additionally, integrating HIV prevention education into school curricula ensures that young people are equipped with the knowledge to protect themselves, helping to reduce new infections in the long term.

3. **Boost Healthcare Access** - Increasing funding for HIV testing and treatment programs is essential to improve healthcare access. Ensuring equitable access to antiretroviral therapy (ART) and pre-exposure prophylaxis (PrEP) will allow people at risk of HIV to receive preventive care, while those living with HIV can receive the necessary treatment to manage their condition effectively.

4. **Focus on Prevention** - Strengthening programs aimed at preventing mother-to-child transmission of HIV is critical to reducing new infections. By ensuring that pregnant women have access to HIV testing and treatment, we can significantly lower the likelihood of transmission during pregnancy, childbirth, or breastfeeding, thus safeguarding the health of both the mother and child.

5. **Combat Stigma** - Combating stigma associated with HIV is essential to ensure that individuals feel safe seeking care. Anti-stigma campaigns should raise awareness and promote understanding, while enforcing anti-discrimination laws will protect individuals from bias and encourage them to access HIV services without fear of judgment or mistreatment.

6. **Invest in Research** - Investment in HIV research is vital for developing more effective vaccines and treatments. Additionally, understanding the socio-economic factors influencing HIV trends will help tailor interventions that address the root causes of HIV transmission, such as poverty and lack of access to healthcare, thus improving long-term outcomes.

7. **Monitor Progress** - Establishing key performance indicators (KPIs) is necessary to track the success of HIV interventions. These KPIs will help measure the effectiveness of ongoing programs, allowing policymakers to adjust strategies based on data insights, ensuring that HIV prevention and treatment efforts remain relevant and impactful.

Conclusion

This report provides a detailed analysis of global HIV trends, emphasizing the socio-economic, demographic, and health factors impacting the epidemic. Key findings reveal significant regional disparities, with peaks in HIV prevalence in 2011 and 2015, followed by a notable decline in 2016. Mortality trends, prevalence among women of reproductive age, and high levels of discrimination among younger, less-educated populations underscore the ongoing challenges in combating HIV.

The proposed recommendations enhancing data systems, boosting healthcare access, combating stigma, and investing in prevention and research offer actionable steps to reduce the HIV burden and improve health equity. These measures are critical for creating inclusive, data-driven public health strategies to sustain progress and address gaps in HIV prevention, treatment, and care.

References

UNICEF. (n.d.). HIV/AIDS statistical tables. Monitoring the situation of children and women. Retrieved from https://data.unicef.org/resources/dataset/hiv-aids-statistical-tables/

UNAIDS. (n.d.). Global HIV & AIDS statistics: Fact sheet. Retrieved from https://www.unaids.org/en

World Health Organization (WHO). (n.d.). HIV/AIDS. Retrieved from https://www.who.int

UNICEF. (n.d.). Children, HIV and AIDS. Retrieved from https://www.unicef.org

Department of Health and Mental Hygiene (DOHMH). (n.d.). Annual HIV/AIDS surveillance report.