**ALY6010: Probability Theory and Introductory Statistics**

**Analysis of Adidas US Sales Dataset** Module5ROutput\_ZihanMa\_03.26.2023

Logo

Description automatically generated

**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Boston, MA, 02215.**

**Submitted by:**

**Group 2**

Bhavishya Koluguri

Lakshmi Nikita Kruthiventi

Rachana Maguluri

Zihan Ma

**Instructor**

 Prof. Na Yu

**Analysis of HIV in Southern and Eastern Asia from 2008 to 2017**

**Introduction:**

Human immunodeficiency virus (HIV) is a viral infection that compromises the immune system and can lead to acquired immunodeficiency syndrome (AIDS). HIV poses a significant public health concern in various parts of Asia, with approximately 5 million people living with HIV in the region. Countries like Thailand, Cambodia, and India have some of the highest HIV prevalence rates globally. Furthermore, the HIV epidemic in Asia is complex and diverse, with different sub-populations and geographic areas affected differently.

Effective data management is essential for understanding the HIV epidemic in Asia and developing successful interventions. This involves collecting accurate and timely data on HIV prevalence, Incidence, and risk factors and monitoring the implementation and effectiveness of HIV prevention and treatment programs.

**Purpose of the project and the assigned region**:

This report aims to analyze HIV trends in the Southern and Eastern regions of Asia between 2008 and 2017. Due to a large number of countries in Asia, the analysis will focus specifically on these regions where reliable and comprehensive data is available. Our team's investment in understanding the HIV epidemic in this part of the world stems from our members' origins in these regions.

**Data Preparation**:

The dataset for our project was sourced from the World Health Organization (WHO), Our World in Data, and the United Nations Children's Fund (UNICEF). After obtaining the dataset, we conducted thorough data cleaning and filtering to extract relevant data pertaining to the Southern and Eastern regions of Asia from the entire dataset. Our approach ensured that the resulting data was accurate, reliable, and representative of the specific geographic area of interest.

**Data Cleaning performed:**

We utilized Microsoft Excel for data cleaning and preparation, identifying and correcting inconsistencies in the dataset, and filling in any missing values. In some cases, column headings contained invalid symbols, such as hyphens, which were removed. Additionally, we ensured that all column names were valid and descriptive, making them more easily understood and accessible for data analysis. Using Excel allowed for a systematic and efficient approach to data cleaning and preparation, ensuring that the resulting dataset was accurate, consistent, and ready for analysis. The table below displays the first five rows of the dataset.

Graphical user interface, table

Description automatically generated

**Data Analysis & SQL Queries**:

This query selects all columns from the "HIV\_Dataset" table and the "HIV\_Asia\_TIP" table, and joins them on the "Country" and "Year" columns. It then filters the results to only include rows where the year in the "HIV\_Asia\_TIP" table is either "2008" or "2017". Which allows for a comparison of data between the "HIV\_Dataset" and "HIV\_Asia\_TIP" tables for the years 2008 and 2017. This comparison could be useful for identifying trends and changes in HIV-related data over time, as well as for analyzing the accuracy and consistency of data between the two tables.Table

Description automatically generated

1. **Deaths of all ages in each country (total of 14 countries)**

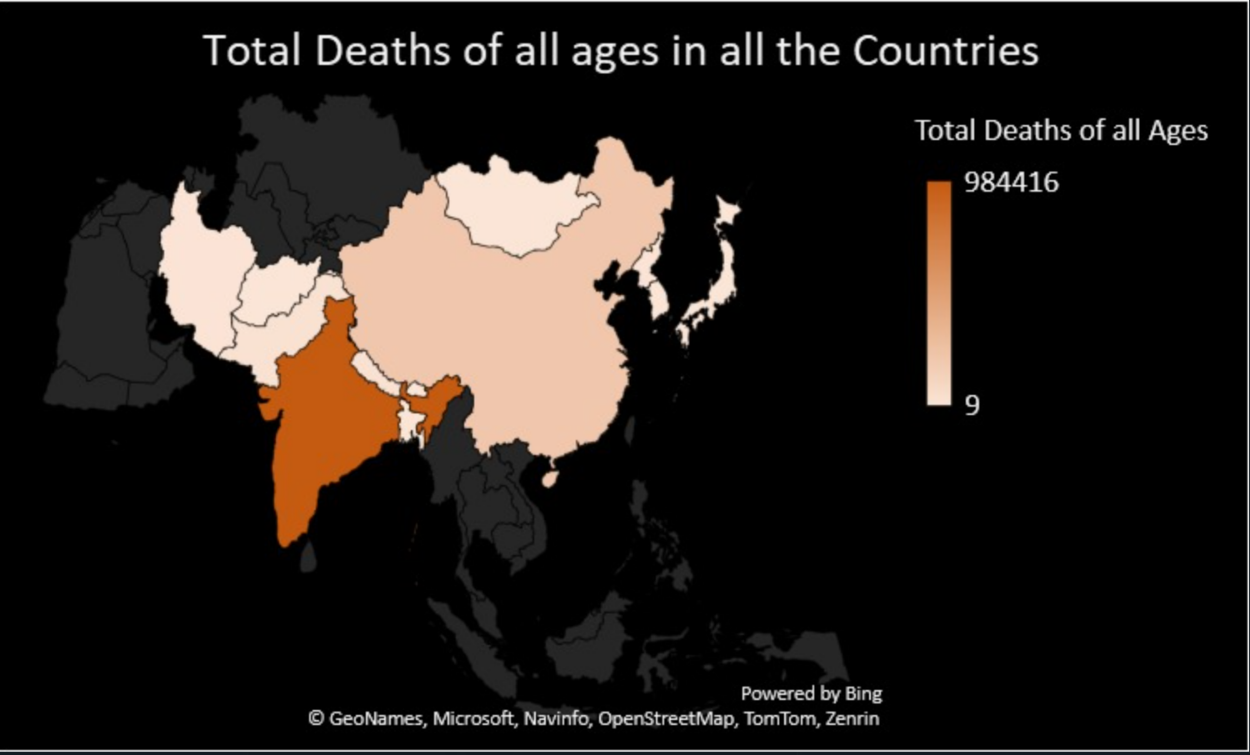
This query selects the "Country" column and the sum of the "Total\_Deaths" column from the "HIV\_ASIA" table. It then groups the results by "Country," calculating the total number of deaths for each country listed in the "HIV\_ASIA" table.

This information can help analyze the impact of HIV on various countries in Asia and identify areas where interventions may be needed to reduce the number of deaths due to HIV.

Table

Description automatically generated

**Visualization**:



**Observations:**

The visualizations reveal significant variations in HIV-related deaths among different countries within the Southern and Eastern regions of Asia. India and China have the highest numbers among all Southeastern Asian countries.

**2. Deaths of age 70+ by country**

This query selects the "Country" column and the sum of the "Deaths\_Age\_70plus\_years" column from the "HIV\_Asia" table. It then renames the sum of "Deaths\_Age\_70plus\_years" as "Deaths\_70plus" using the "AS" keyword. The query groups the results by "Country," calculating the total number of deaths of individuals aged 70 or older for each country listed in the "HIV\_Asia" table.

The query result can help analyze HIV's impact on elderly populations in Asian countries and identify areas for interventions to reduce HIV-related deaths.

Table

Description automatically generated

**Visualization:**

Chart, funnel chart

Description automatically generated

**Observations:**

From the graph above, we can see India and China have the highest two numbers in all Southeastern Asia countries for 70+ ages as well. For both query results above, we will focus on both countries in this report.

3. **Deaths of Age 70+ in China**

This query selects the "Country", "Year", and "Deaths\_Age\_70plus\_years" columns from the "HIV\_Asia" table, but only for rows where the "Country" column contains the exact string "CHINA" and the "Deaths\_Age\_70plus\_years" column is not null. It then renames the "Deaths\_Age\_ 70plus\_years" column as "Deaths\_70plus" using the "AS" keyword.

The query result can aid in analyzing trends of HIV-related deaths among older individuals in China and identifying intervention areas to reduce such deaths.

Table

Description automatically generated

**Visualization:**

Chart, waterfall chart

Description automatically generated

**Observations:**

The table above shows that the number of people aged 70+ dying from HIV is increasing. However, we cannot determine whether the main factor is the increase in the population, the deterioration of medical conditions, or the rise in the HIV infection rate.

**4. Deaths of Age: 15 to 49 in India**

This query selects the "Country," "Year", and "Deaths\_Age\_15to49\_years" columns from the "HIV\_Asia" table, but only for rows where the "Country" column contains the exact string "INDIA". It then groups the results by "Year".

The query result can assist in analyzing trends of HIV-related deaths among younger individuals in India and identifying intervention areas to reduce such deaths.

Table

Description automatically generated

**Visualization:**

Chart

Description automatically generated

**Observations:**

This graph reflects that the number of HIV-infected individuals who died between 15 and 49 years old has decreased over these ten years. Factors such as an increase in the population, improvement in medical conditions, and a reduction in the HIV infection rate could apply here.

**5. Number of deaths in all the countries from the year 2008 to 2017**

This query selects the "Country" column and the sum of "Deaths\_Age\_5to14\_years", "Deaths\_Age\_15to49\_years", "Deaths\_Age\_50to69\_years", and "Deaths\_Age\_70plus\_years" columns from the "HIV\_Asia" table. It then groups the results by "Country", calculating the total number of deaths for each age group in each country listed in the "HIV\_Asia" table.

Table

Description automatically generated

The query result can be valuable in analyzing HIV's impact on different age groups in Asian countries and identifying intervention areas to reduce HIV-related deaths across all age groups.

**Visualization:**

Chart

Description automatically generated

**Observations:**

This table shows the total death number of people by age across the countries, which shows China’s HIV death case is more distribute to the elder side (0.87%/62.14%/28.51%/8.36%) compare to India (2.57%/ 77.68%/ 18.68%/ 0.37%).

6. **Prevalence rate by country**   
This query selects the "Country" column and the "Prevalence\_Age\_All\_Ages" column from the "HIV\_Asia" table. It then groups the results by "Country", displaying the prevalence of HIV for all ages for each country listed in the "HIV\_Asia" table.

The result here can aid in analyzing HIV spread across Asian countries and identifying intervention areas to decrease its prevalence.

Table

Description automatically generated

**Visualization:**

Chart

Description automatically generated

**Observations:**

A considerable disparity in prevalence among different countries can be seen, suggesting that the impact of HIV varies widely across the regions.

**More Visualizations:**

**1. Prevalence(Living with HIV) Vs Incidence(HIV Occurrence)**

Chart

Description automatically generated

**Observations:**

From the graph above, we can observe that the death rate for HIV is relatively high in China/India/south Korea/Pakistan because both Prevalence and Incidence are high at the same time. Also, Pakistan and South Korea have higher HIV death, which is not normal because South Korea is a developed country. Usually, a developed country should have a better medical system compared to some least developed countries.

**2. Population Vs. Total Deaths in all countries**

Chart, bar chart

Description automatically generated

**Observations:**

The table above shows the overall HIV death rate for each country, and we can observe significant differences between our two focus countries. India has a much higher death rate compared to China. Besides this, Nepal also has a high death rate because we can see noticeable blue under its bar.

**Conclusion**:

**Based on the analysis, the following results were observed:**

1. In southeast Asia, China and India have the highest HIV incidence and prevalence rate across all ages from 2008 to 2017.
2. The number of 70+ years old HIV incidence are increasing every year from 2008 to 2017 in China.
3. The number of 15 to 49 years old HIV incidence decreases yearly from 2008 to 2017 in India.
4. Countries with the highest Incidence, such as China and India, are also usually high in prevalence.

Overall, from the findings above, both China and India’s population are still heavily suffered at HIV. But by seeing the HIV death numbers increasing for Chinese elders, and HIV death numbers decreasing at India, we could conclude that both countries’ HIV spread have been retarded. There are still have some countries have high Prevalence Incidence ratio. Those countries should put more effort to reduce the HIV death ratio. And make more commutation to the countries China and India to learn how to treat HIV effetely.

**Policy Recommendations**:

**Based on the analysis and findings, suggest policy recommendations for future HIV interventions:**

1. Countries with the highest prevalence rates, such as China and India, require more focused and effective interventions to reduce the spread of HIV and mitigate its impact. These countries should Implement targeted prevention and awareness programs and strengthen healthcare systems and infrastructure.
2. Strengthen regional cooperation among Southern and Eastern Asian countries to facilitate sharing of best practices, resources, and expertise in addressing the HIV epidemic. This includes promoting cross-border collaboration in surveillance, research, and policy development.

**References:**

1. World Economic Forum. (2019, July 11). India has the world's third-largest HIV epidemic. Here's what they're doing to tackle it. <https://www.weforum.org/agenda/2019/07/india-hiv-epidemic-tackle/>
2. Global Times. (2020, October 14). China's battle against HIV/AIDS. <https://www.globaltimes.cn/page/202010/1203044.shtml>
3. AVERT. (n.d.). HIV/AIDS in India: A country profile. <https://www.avert.org/professionals/hiv-around-world/asia-pacific/india>
4. CNN. (2019, December 1). China's HIV epidemic is growing, but stigma remains a barrier to care. <https://www.cnn.com/2019/12/01/health/china-hiv-stigma-intl-hnk/index.html>
5. Arora, P., Nagelkerke, N. J. D., & Schechter, M. T. (2020). HIV prevention and treatment in India: Achievements and challenges. Journal of Global Health, 10(1), 010401. <https://doi.org/10.7189/jogh.10.010401>
6. UNAIDS. (2019). HIV and AIDS in Latin America and the Caribbean. Retrieved from <https://www.unaids.org/en/regionscountries/regions/lac>
7. UNICEF. (2018). HIV and AIDS in Brazil. Retrieved from <https://www.unicef.org/brazil/hiv-and-aids>
8. World Health Organization. (2017). HIV/AIDS in Latin America and the Caribbean. Retrieved from <https://www.who.int/hiv/regions/lac/en/>
9. Ministry of Health, Brazil. (2016). National AIDS/STD Program. Retrieved from <http://www.aids.gov.br/en/>
10. Joint United Nations Programme on HIV/AIDS. (2016). Global AIDS Monitoring 2016: Country Progress Report – Brazil. Retrieved from <https://www.unaids.org/sites/default/files/country_progress_report/2016/GlobalAIDSMonitoring_Brazil_2016.pdf>