## Introduction

The purpose of this report is to present the findings of an analysis of HIV/AIDS data obtained from Kaggle. The report provides an overview of the data source, the data content, the data cleaning process, and the schema design used in the analysis. Additionally, it will cover the steps taken to create the schema and tables used in this analysis.

## Data Source

The data used in this analysis was obtained from Kaggle, a popular data science community platform. The dataset contains information on the number of people living with HIV/AIDS, the number of deaths due to HIV/AIDS, the number of cases among adults (15-45), prevention of mother-to-child transmission estimates, ART coverage among people living with HIV estimates, and ART coverage among children estimates.

Data Source

## Data Content

The data set contains the following variables:

No. of people living with HIV/AIDS.

No. of deaths due to HIV/AIDS.

No. of cases among adults (15-45).

Prevention of mother-to-child transmission estimates.

ART coverage among people living with HIV estimates.

ART coverage among children estimates.

## Data Cleaning

Prior to conducting the analysis, the data was cleaned to ensure that it was consistent and free from errors. The following data cleaning steps were taken:

- Renamed the required columns to make them more meaningful.
- Removed unnecessary data that was not relevant to the analysis.

## Schema Design and Table Creation

To analyse the data, a schema was designed that accurately represented the relationships between the variables. The schema was designed in such a way that the tables were normalized to reduce redundancy and ensure consistency.

To create the tables in the schema, the following SQL code was used:

1. Antiretroviral Therapy by country among people with HIV (by country)

```
CREATE TABLE Antiretrovirals_Coverage
    (
        Country varchar(64)  NOT NULL,
        People_recieving_ART int DEFAULT NULL,
        People_living_with_HIV int DEFAULT NULL,
        People_living_with_HIV_min int DEFAULT NULL,
        People_living_with_HIV_max int DEFAULT NULL,
        ART_coverage_among_people_living_with_HIV__Med int DEFAULT NULL,
        ART_coverage_among_people_living_with_HIV__Min int DEFAULT NULL,
        ART_coverage_among_people_living_with_HIV__Max int DEFAULT NULL,
        WHO_Region varchar(50) DEFAULT NULL,

        PRIMARY KEY(Country)
    )
```

```
-- 2. Overall people who are living with HIV
CREATE TABLE People_living_with_HIV
    (
        Country varchar(64) NOT NULL,
        Record_Year int DEFAULT NULL,
```

```sql
        Count_Med int DEFAULT NULL,
        Count_Min int DEFAULT NULL,
        Count_Max int DEFAULT NULL,
        WHO_Region varchar(50) DEFAULT NULL,

        FOREIGN KEY(Country) REFERENCES Antiretrovirals_Coverage(Country)

    )


-- 3. HIV cases in Adults by country Age group 15-49
CREATE TABLE Cases_in_Adults
    (
        Country varchar(64) NOT NULL,
        Record_Year int DEFAULT NULL,
        Count_Med double DEFAULT NULL,
        Count_Min double DEFAULT NULL,
        Count_Max double DEFAULT NULL,
        WHO_Region varchar(50) DEFAULT NULL,

        FOREIGN KEY(Country) REFERENCES Antiretrovirals_Coverage(Country)
    )



-- 4. Overall deaths of HIV patients
CREATE TABLE Overall_Deaths
    (
        Country varchar(64) NOT NULL,
        Record_Year int DEFAULT NULL,
        Count_Med int DEFAULT NULL,
        Count_Min int DEFAULT NULL,
        Count_Max int DEFAULT NULL,
        WHO_Region varchar(50) DEFAULT NULL,

        FOREIGN KEY(Country) REFERENCES Antiretrovirals_Coverage(Country)


-- 5. Perinatal Prevention (Mother to Child transmission)
CREATE TABLE Perinatal_Prevention
    (
        Country varchar(64) NOT NULL,
        Received_Antiretrovirals varchar(50) DEFAULT NULL,
        Needing_antiretrovirals_Med int DEFAULT NULL,
        Needing_antiretrovirals_Min int DEFAULT NULL,
        Needing_antiretrovirals_Max int DEFAULT NULL,
        Percentage_Recieved_Med int DEFAULT NULL,
        Percentage_Recieved_Min int DEFAULT NULL,
        Percentage_Recieved_Max int DEFAULT NULL,
        WHO_Region varchar(50) DEFAULT NULL,

        PRIMARY KEY(Country)

)

-- 6. Antiretroviral (ART) therapy schemes for the initial therapy of HIV infection in children
CREATE TABLE ART_Pediatric_Coverage
```

```sql
(
    Country varchar(64) NOT NULL,
    No_of_children_receiving_ART int DEFAULT NULL,
    No_of_children_needing_ART_Med int DEFAULT NULL,
    No_of_children_needing_ART_Min int DEFAULT NULL,
    No_of_children_needing_ART_Max int DEFAULT NULL,
    ART_coverage_among_children_with_HIV__Med int DEFAULT NULL,
    ART_coverage_among_children_with_HIV__Min int DEFAULT NULL,
    ART_coverage_among_children_with_HIV__Max int DEFAULT NULL,
    WHO_Region varchar(60) DEFAULT NULL,

    PRIMARY KEY (Country)
)
```

7. How many Pregnant mothers (Perinatal transmission) received ART?

```sql
SELECT
    Country,
    WHO_Region,
    Percentage_Recieved_Med
FROM
    Perinatal_Prevention pp
WHERE
    Percentage_Recieved_Med IS NOT NULL
```

## Queries:

1.What is the estimated number of people worldwide who are HIV-positive and living with the virus?

```sql
SELECT
    year_2000.*,
    year_2005.Patient_2005,
    year_2010.Patient_2010,
    year_2018.Patient_2018
FROM
    (
        SELECT
            Country, SUM(Count_Med) AS Patient_2000
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2000
            AND Count_Med IS NOT NULL
        GROUP BY
            Country
    ) AS year_2000 ,
    (
        SELECT
            Country, SUM(Count_Med) AS Patient_2005
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2005
            AND Count_Med IS NOT NULL
        GROUP BY
            Country
    ) AS year_2005 ,
```

```sql
    (
        SELECT
            Country, SUM(Count_Med) AS Patient_2010
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2010
            AND Count_Med IS NOT NULL
        GROUP BY
            Country
    ) AS year_2010 ,
    (
        SELECT
            Country, SUM(Count_Med) AS Patient_2018
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2018
            AND Count_Med IS NOT NULL
        GROUP BY
            Country
    ) AS year_2018
WHERE
    year_2000.Country = year_2005.Country
    AND year_2005.Country = year_2010.Country
    AND year_2010.Country = year_2018.Country
GROUP BY
    year_2000.Country
ORDER BY
    year_2000.Country
    ASC
```
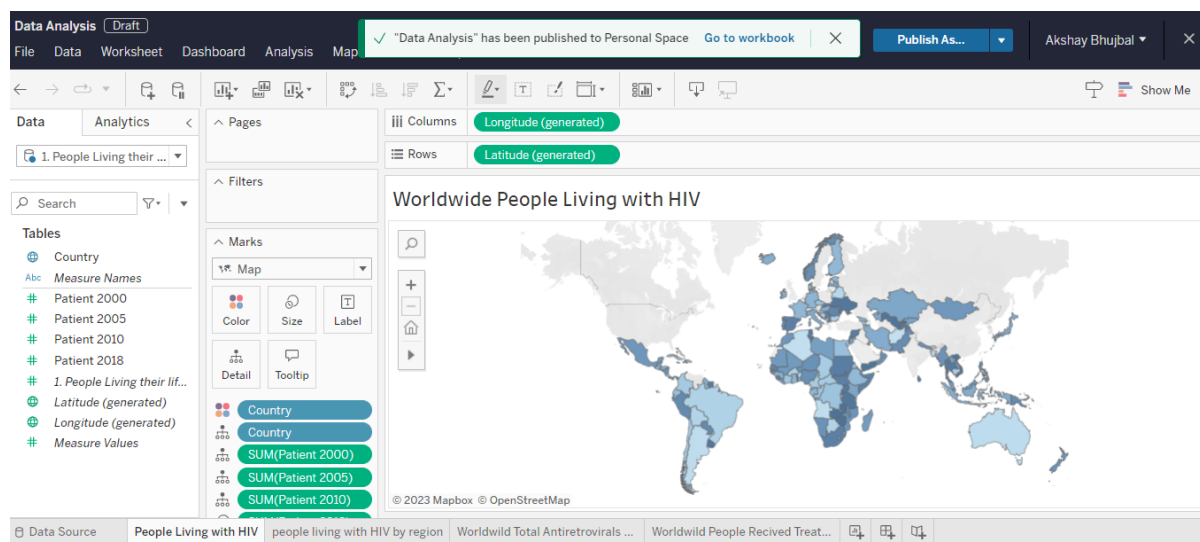


The map graph presents global data on the number of people living with HIV in 2000, 2005, 2010, and 2018. The highest prevalence of HIV is observed in African countries, whereas the Republic of North Macedonia has a lower number of cases.

2.How does the number of HIV-positive individuals living with the virus vary by WHO region/continent?

```sql
SELECT
    year_2000.*,
    year_2005.Patient_2005,
    year_2010.Patient_2010,
    year_2018.Patient_2018
FROM
    (
        SELECT
            WHO_Region, SUM(Count_Med) AS Patient_2000
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2000
            AND Count_Med IS NOT NULL
        GROUP BY
            WHO_Region
    ) AS year_2000 ,
    (
        SELECT
            WHO_Region, SUM(Count_Med) AS Patient_2005
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2005
            AND Count_Med IS NOT NULL
        GROUP BY
            WHO_Region
    ) AS year_2005 ,
    (
        SELECT
            WHO_Region, SUM(Count_Med) AS Patient_2010
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2010
            AND Count_Med IS NOT NULL
        GROUP BY
            WHO_Region
    ) AS year_2010 ,
    (
        SELECT
            WHO_Region, SUM(Count_Med) AS Patient_2018
        FROM
            People_living_with_HIV PLWH
        WHERE
            Record_Year = 2018
            AND Count_Med IS NOT NULL
        GROUP BY
            WHO_Region
    ) AS year_2018
WHERE
    year_2000.WHO_Region = year_2005.WHO_Region
    AND year_2005.WHO_Region = year_2010.WHO_Region
    AND year_2010.WHO_Region = year_2018.WHO_Region
GROUP BY
```

```
    year_2000.WHO_Region
ORDER BY
    year_2000.WHO_Region
    ASC
```

## worldwide People living with HIV by WHO Region
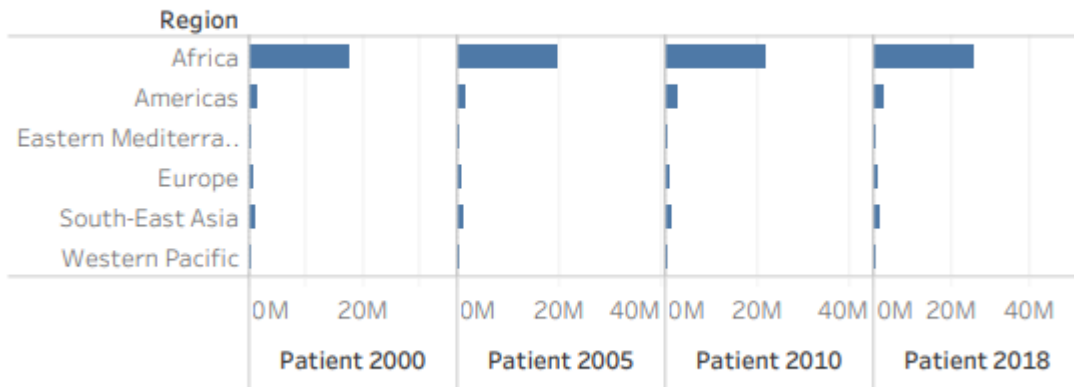


This graph illustrates the number of HIV-infected individuals by WHO region. Africa has the highest number of infections, while the Eastern Mediterranean region has the lowest.

3.What is the number of people receiving antiretroviral therapy (ART) in each country?

```
SELECT
    Country,
    WHO_Region,
    SUM(People_recieving_ART) AS Total_Antiretrovirals_Coverage
FROM
    Antiretrovirals_Coverage AC
WHERE
    People_recieving_ART IS NOT NULL
GROUP BY
    Country
ORDER BY
    Total_Antiretrovirals_Coverage DESC
```

## Worldwild Total Antiretrovirals Coverage

The graph shows the total number of individuals who have received antiretroviral therapy (ART) for HIV treatment. South Africa has the highest number of individuals receiving ART, while Middle Eastern countries such as Qatar and Syria have a lower number.

4.How many people have received ART and how does this correlate with the number of deaths due to HIV/AIDS?

```
SELECT
    AC.Country,
    SUM(AC.People_recieving_ART) AS Recieving_treatment,
    deaths.Total_deaths

FROM
    Antiretrovirals_Coverage AC
    LEFT JOIN (SELECT Country, SUM(Count_Med) AS Total_deaths  FROM Overall_Deaths GROUP
BY Country) AS deaths ON deaths.Country = AC.Country
WHERE
    AC.ART_coverage_among_people_living_with_HIV__Med IS NOT NULL
GROUP BY
    AC.Country
```
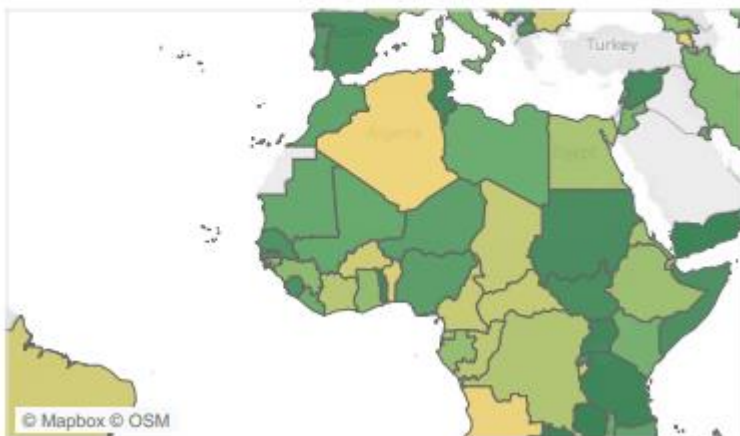


Worldwild People Recived Treatment and Deaths

This graph compares the number of individuals who received ART treatment and the total deaths recorded. Kenya has the highest number of individuals receiving treatment and recorded deaths, followed by Mozambique and Nigeria.

5.How many children are in need of ART, how many have received treatment among top 30 countries?

```
SELECT
    Country,
    WHO_Region,
    SUM(No_of_children_needing_ART_Med) AS Needed_ART,
    SUM(No_of_children_receiving_ART) AS Recieving_ART ,
    SUM(ART_coverage_among_children_with_HIV__Med) AS ART_Covergar_Percentage
FROM
    ART_Pediatric_Coverage apc
    WHERE  (No_of_children_needing_ART_Med IS NOT NULL) AND (No_of_children_receiving_ART IS  NOT
NULL) AND (ART_coverage_among_children_with_HIV__Med IS NOT NULL)
GROUP BY
    Country
```
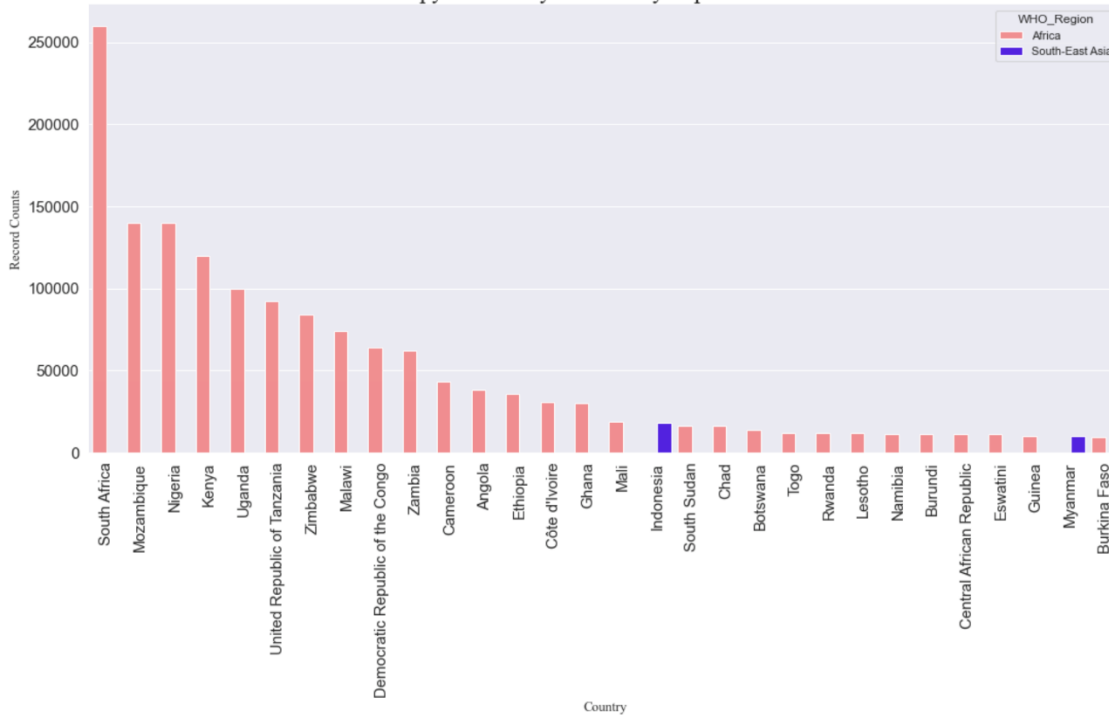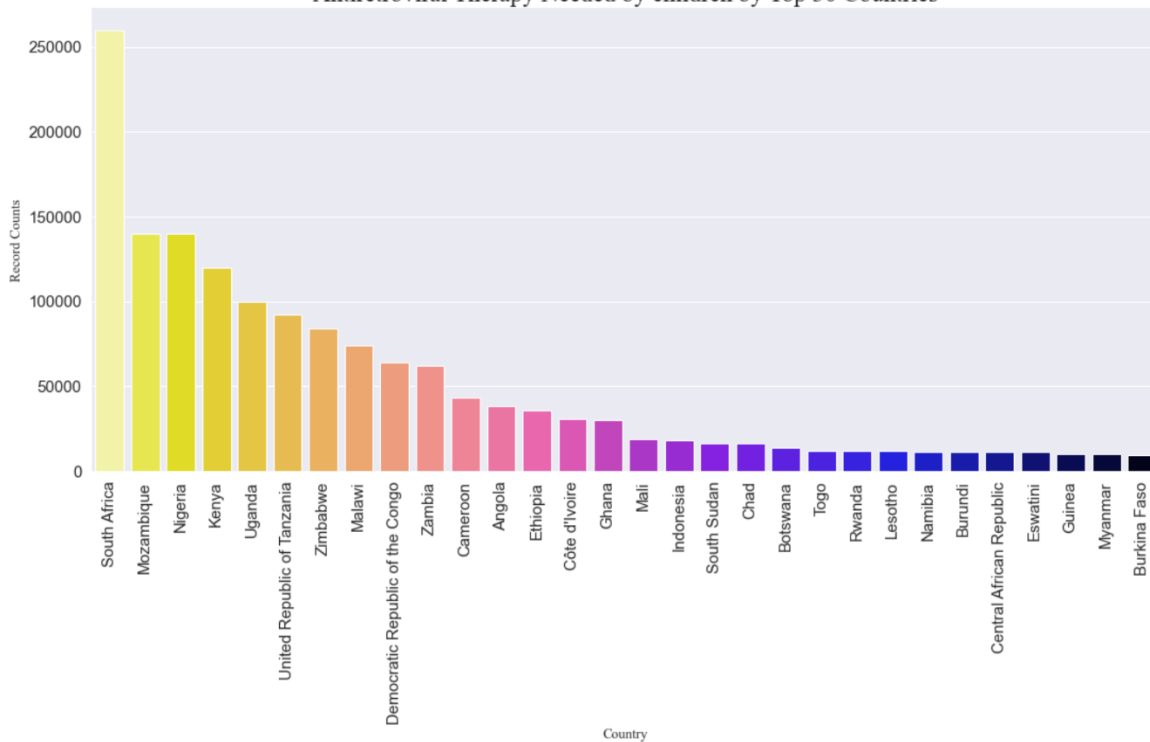
Antiretroviral Therapy Needed by children by Top 30 Countries and Continents



Antiretroviral Therapy Needed by children by Top 30 Countries

The following graph depicts the top 30 countries with the highest demand for ART treatment in children. The graph highlights that African countries have the highest need for this therapy, with 80% of the countries being in Africa.
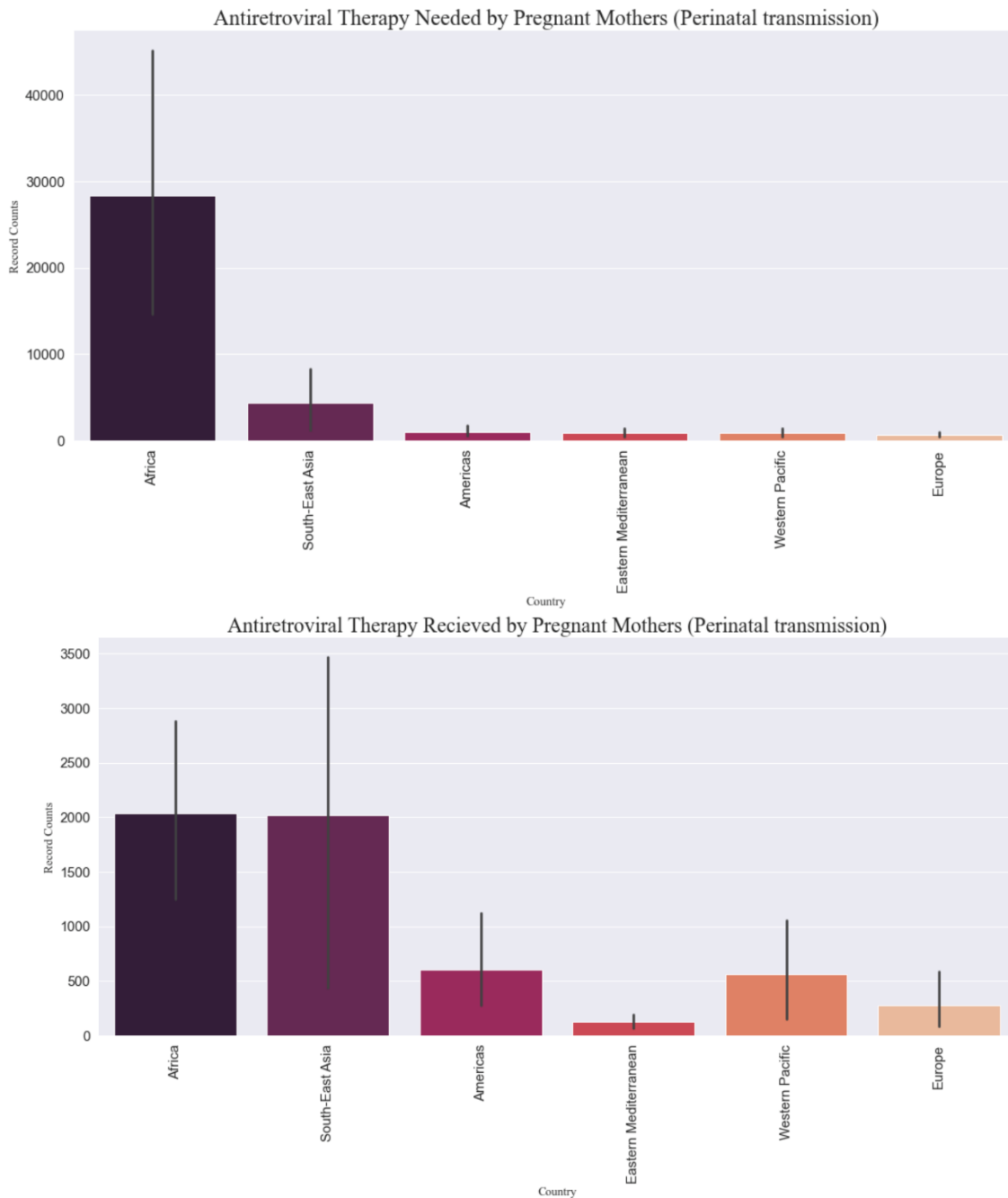
6.What is the gap between the number of ART treatments needed and the actual number of treatments received for perinatal transmission cases among top 30 countries and continent?

```sql
SELECT
    Country,
    WHO_Region,
    SUM(Needing_antiretrovirals_Med) AS Needed_ART,
    SUM(Received_Antiretrovirals) AS Recieving_ART
FROM
    Perinatal_Prevention pp
    WHERE  (Needing_antiretrovirals_Med IS NOT NULL) AND (Received_Antiretrovirals IS  NOT NULL)
```

### Antiretroviral Therapy Needed by Pregnant Mothers (Perinatal transmission)



### Antiretroviral Therapy Recieved by Pregnant Mothers (Perinatal transmission)



The graph provides data on the demand for antiretroviral therapy among pregnant women. The highest demand is in African and Southeast Asian countries, while Western Pacific and European countries have lower demand.

7.How many pregnant mothers with perinatal transmission have needed and received ART for HIV prevention?

SELECT
    Country,
    WHO_Region,
    Percentage_Recieved_Med
FROM
    Perinatal_Prevention pp
WHERE
    Percentage_Recieved_Med IS NOT NULL

Pregnant mothers (Perinatal transmission) recived Antiretroviral Therapy by Continent
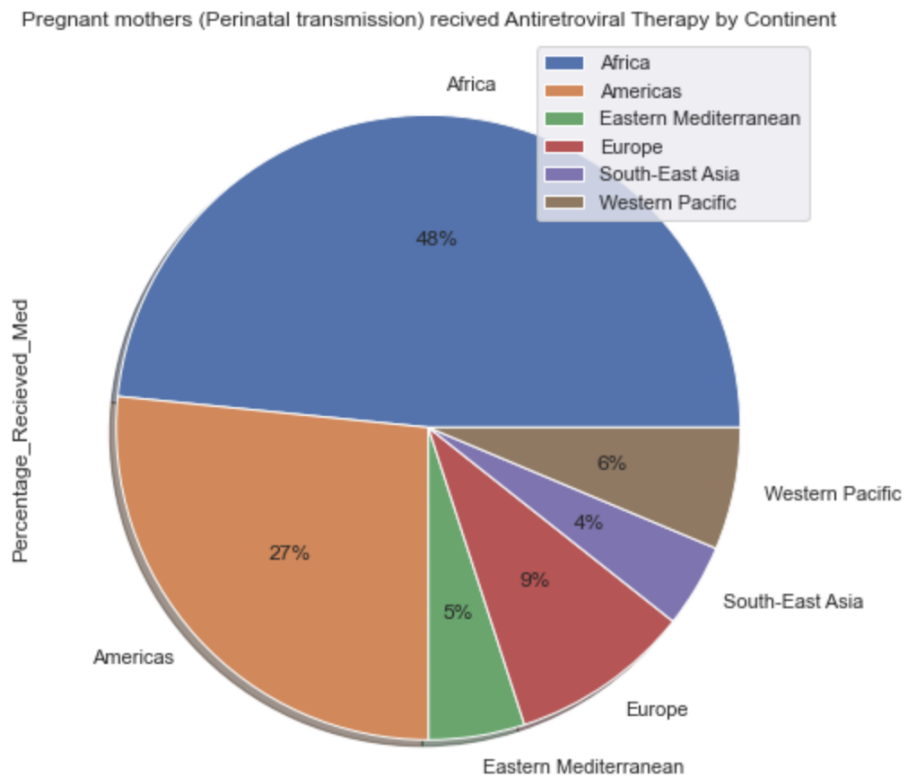
The graph shows the transmission of HIV from pregnant mothers to their infants, with the highest transmission rates observed in Africa, followed by America and the Western Pacific. Southeast Asia and the Eastern Mediterranean regions have comparatively lower transmission rates.

Conclusion

The global data presented in these graphs highlights the continued prevalence of HIV/AIDS in many regions of the world. African countries, in particular, have a higher prevalence of HIV infections and a greater need for antiretroviral therapy, especially in children and pregnant women. Despite the availability of ART treatment, the number of deaths due to HIV/AIDS remains high in certain countries, highlighting the need for continued efforts to improve access to treatment and prevention measures. Overall, these graphs demonstrate the ongoing challenges in addressing the HIV/AIDS epidemic globally and the need for sustained efforts to improve access to care and reduce transmission rates.