**DATA ANALYSIS ASSESSMENT SCENARIO**

## Background

Invictus Africa is currently evaluating awareness and barriers to the implementation of the Family Planning 2030 (FP2030) policy. Your task is to analyze survey data collected from respondents across different states in Nigeria, and derive meaningful insights to support policy recommendations and advocacy strategies.

## The Dataset

You have been provided with a CSV file containing responses from a recent survey. The dataset includes the following fields:

- Sex (Male/Female)
- Age group (18-24, 25-34, 36-45, 46-55, 56 and above)
- State of residence
- Area type (Urban/Rural)
- FP2030 policy awareness (Yes/No)
- Barriers to FP2030 implementation (Multiple selection):
    o Weak political will
    o Limited healthcare infrastructure
    o Lack of awareness
    o Resistance from traditional/religious leaders
    o Inadequate funding
    o No idea
    o Others (specified)

## Assessment Tasks

**Note:** Carry out **part 1 to 3** using either **R or Phyton** programming language.

**Part 1: Data Wrangling**
Produce a clean and transformed dataset by addressing the following:
- Only data from the South East and the North East states will be analysed.
- Age group "46-55" and "56 and above" should be merged into one group "Above 45"
- Some impossible combinations:
    o Some respondents selected barriers despite indicating they have not heard about the policy.
    o Some respondents chose "No Idea" along with other barrier options, which contradicts the uncertainty implied by "No Idea."

**Part 2: Analysis and Visualization**
Using your cleaned dataset:
1. Demographic Insights:
    a. Which states have the highest number of respondents?
    b. What is the distribution of respondents by sex, and by age group?
2. Policy Awareness:
    a. What percentage of respondents have heard of the FP2030 policy?
    b. Are there differences in awareness of FP2030 across the states?
    c. Which is a better predictor of awareness: sex or residential area (urban vs rural)?

3. Implementation Barriers:
    a. Identify the top three barriers to implementation.
    b. Does the perception of barriers differ based on urban vs. rural residence?
4. Visualization:
    a. Create at least two visualizations to communicate your key findings.

**Part 3: Reporting**
Based on your analysis:
1. Prepare a concise data interpretation report communicating the key insights (max 500 words).
2. Recommend 2-3 specific, data-driven recommendations based on the insights identified.

**Part 4: Interactive Dashboard**
Create a simple interactive dashboard using **your tool of choice** (e.g., Excel, Power BI, Tableau, R Shiny, Python Dash/Streamlit)
- Respondents' distribution across demographics
- Percentage distribution of FP2030 awareness levels
- FP2030 awareness disaggregated by sex and residential area
- Barriers to implementation ranked in descending order
- State Filters/Interactivity

**Deliverables**

- Cleaned dataset (Excel or CSV format)
- Word or PDF document containing data analysis methods, key findings, visualizations and recommendations.
- A published link to the interactive dashboard.