# Femanalytica – Interview Exercise (Round 1)

Position: Data Science and Programs Specialist

Dataset: ds\_dataset.csv

Submission Deadline: **1st May 2025, 23:59**

## Dataset Description

You are provided with two datasets from adult individuals across different regions Malawi with responses about their health seeking behavior and presence of chronic conditions. You are provided with two separate files containing related information. You may use them as needed to complete the task.

The datasets include the following variables:

1. Demographics.csv

|  |  |
| --- | --- |
| Variable | Description |
| id | Unique respondent identifier |
| gender | 1 = Female, 2 = Male |
| age | Respondent’s age |
| region | 1 = North, 2 = Central, 3 = South |
| education\_level | 1 = None, 2 = Primary, 3 = Secondary, 4 = Tertiary |

1. Health\_data.csv

|  |  |
| --- | --- |
| Variable | Description |
| id | Unique respondent identifier |
| recent\_health\_check | Number of times the respondent visited a health facility in the past 12 months |
| reported\_barriers\_to\_care | Reported barrier to accessing care (e.g., distance, cost, time) |
| chronic\_condition | Self-reported chronic condition: None, Hypertension, Diabetes |

## Task

Using a software of your choice, do the following:

1. Data Preparation  
Prepare the datasets for analysis. You are expected to identify and address any quality issues that would affect interpretation or reporting.

2. Descriptive Summary  
Provide a summary of statistics of the population. This should include appropriate breakdowns or stratifications that you consider relevant to the data.

3. Visualizations  
Create three visualizations:  
- One examining gender differences in health check behavior  
- One examining gender differences in prevalence of chronic disease

- One examining gender differences in chronic conditions, disaggregated by age group

You are free to choose the most appropriate visual format for each.

4. Summary   
Write a short summary (maximum two paragraphs) that (1) summarizes the steps you took to prepare the data for analysis (2) communicates one key finding to a non-technical audience. Suggest one potential implication for programmatic or policy work.

5. Annotated Code  
Your script should be well-commented, with explanations for each step and decision made. This includes how you cleaned the data, any grouping or recoding.

## What to Submit

- Cleaned dataset (CSV)  
- Annotated analysis code of the software you used  
- The three visualizations you produced (as image files or include in your summary)  
- Summary (PDF or Word)

Please submit the above by the deadline to admin@femanalytica.org Late submissions will not be considered. Any queries regarding the task can be sent to the same email.