**Data Cleaning Report – Uganda Financial Instability Project**

**1. Missing Values Check**

* Verified both datasets using Python for any missing values.
* Result: **No missing values** were found in either dataset.

**2. Duplicate Records**

* Checked for duplicate rows across both datasets.
* Result: **No duplicates** were present.

**3. Column Name Cleaning**

* Applied the strip() method to remove any leading or trailing whitespaces from column names.

**4. Currency Standardization**

* Converted all relevant monetary columns to **UGX (Ugandan Shillings)** for consistency.

**5. Time Series Preparation**

* Created a new column Year\_Month using the Excel formula:  
  =EDATE(DATE(2023,0,1), B2)  
  This was necessary because the original Month column ranged from 1 to 24.  
  The new column enables accurate **time series analysis** across two years (2023 and 2024).

**6. Total Expenses Calculation**

* Introduced a new column Total\_Expenses by summing the following:
  + Food\_Expenses
  + Healthcare\_Expenses
  + Education\_Expenses
  + Discretionary\_Expenses

**7. Post-Expense Income**

* Added a new column Income\_After\_Expense, calculated as:  
  Total\_Income - Total\_Expenses  
  This measures net income after essential and non-essential spending.

**8. Wealth Categorization**

* Created a derived column Wealth\_Status using the following logic:

=IFS(

Income\_After\_Expense <= 0, "Bankrupt",

Income\_After\_Expense < 25000, "Low Income",

Income\_After\_Expense < 60000, "Middle Income",

Income\_After\_Expense > 60000, "High Income"

)

This categorizes households based on their net disposable income.