**1️⃣ Master Processing Logic**

The script will:

1. **Read all .docx files** from:  
   D:\AAAA\_Data\GENDER\Questionnaires
2. **Identify the County name** from the filename.
3. **Define a master template** with fixed themes and questions (same for all counties).
4. **Scan document content**:
   * Extract free-text responses under each question.
   * Extract tables (row by row).
   * Detect ticks (✔, ✓, Yes/No) and normalize values.
   * Handle missing answers (empty cells).
5. **Export one Excel per county** with identical structure.

**2️⃣ Master Template (applied to all)**

| County | Theme | Question | Sub-question | Enterprise/Value Chain | Gender Group | Response | Barriers | Decision Control | Climate Impact | Resources Access | Market & Profitability | Policy Gaps | Support Needed | Key Quotes |

*(Fixed row order, empty if not provided)*

**3️⃣ Key Features of the Script**

* **Handles Mixed Formats:**
  + Tables are transformed to “long format” (one response per row).
  + Bullets, ticks, free text all captured.
* **Standardized Output:**
  + Same columns, same row order for all 38 counties.
* **Automated Naming:**
  + Saves files as  
    D:\AAAA\_Data\GENDER\Outputs\<County>\_gendered\_enterprise.xlsx
* **Merge Ready:**
  + Later, you can simply pd.concat() all Excel outputs for analysis across counties.

**4️⃣ Example Output (for Kakamega)**

| **County** | **Theme** | **Question** | **Enterprise** | **Gender** | **Response** |
| --- | --- | --- | --- | --- | --- |
| Kakamega | Enterprise Sel. | How was selection done? |  |  | PRA, Barazas, Farmer groups |
| Kakamega | Identification | Which enterprises women dominate? | Poultry, Vegetables | Women | Low capital, easy management |
| Kakamega | Gender Roles | Who controls decision/income for poultry? | Poultry | Women | Low-income enterprise for men |
| Kakamega | Climate Change | New enterprises due to drought? | Sweet potatoes | Women | Climate-resilient food security |
| … | … | … | … | … | … |

**5️⃣ Next Steps**

* I can prepare a **Python script (v1)** that:
  + Reads **all 38 questionnaires** in your folder.
  + Applies the above logic.
  + Outputs 38 uniform Excel files.

You will only need to:

* Install Python packages (python-docx, pandas, openpyxl).
* Update the script with your folder path if needed.
* Run once, process all questionnaires automatically.