DIME Analytics
Peer Code Review - Cleaning Checklist

Reviewer Details

Reviewer Name: Coder Name:

Note: Please complete this checklist **only if** the submission includes **cleaning** tasks.

Data Cleaning Tasks

This checklist highlights key aspects to review in your partner's **cleaning scripts/code**. Once completed, please submit it as an attachment along with this form.

General Cleaning Checks

Cleaning scripts do not create indicators (this should be done in a separate construction script).

Any changes made to specific values are well-documented and justified.

Duplicate Checks

The code checks for duplicates in ID variables and other key variables.

All identified duplicates are either resolved in the code or justified with an explanation.

The method used to resolve duplicates is stable, meaning it does not randomly drop duplicates but instead systematically identifies and tags specific observations.

Data Type Checks

The dataset contains string variables only where necessary, such as open-ended responses, proper nouns (not used as categories), or alphanumeric IDs.

All categorical string variables are converted into labeled categorical variables or factors.

Labeling Checks

All variables have clear and correct labels.

Value labels are consistent (e.g., avoiding cases where varA: 1 = yes, 0 = no, but varB: 1 = yes, 2 = no).

Missing Value Checks

Missing values are coded consistently.

Extended missing values are used where applicable (e.g., `.d` for **Do not know**, `.r` for **Refuse to answer**, etc.).

Merge Checks

If any observations are dropped, a clear justification is provided in the code.

Any mismatches between datasets are explicitly explained in the code.

m:m merge is NOT perfomed

Clean Dataset Checks

The clean dataset is saved only once throughout the script and is not overwritten multiple times.

The clean dataset has unique values for the designated ID variable.