

DIME Analytics

Peer Code Review - Cleaning Checklist

v1.0

Reviewer:

Coder:

NOTE: Make sure that fill out this checklist **ONLY IF** your partner's submission includes **cleaning** tasks.

Data cleaning tasks

This checklist lists important factors to consider while reviewing your code review partner's **cleaning** scripts/code. Please fill this checklist, and submit it as an attachment when you submit [this detailed form](#).

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 are justifiable.

Check duplicates:

The code checks for duplicates in ID variables and other important variables

All identified duplicates are resolved in code, or explanation is provided for why they cannot be resolved

Check merges:

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

Any mismatches between datasets are explained in the code

Check labels:

All variables are correctly labelled

All value labels are consistent (example of inconsistency: varA 1 = yes, 0 = no; varB 1 = yes, 2 = no)

Check missing values:

Missing values are coded consistently

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