Hello,

I hope this email finds you well. I’ve completed the cleaning and standardization of the datasets (Mwanza\_r, Pwani\_r, and DSM (Ilala)\_r) as per the agreed schema. Below is a summary of the changes made:

### ****Schema Updates:****

1. **Removed SRA CODE**: This field was only used in the Mwanza\_r dataset and has been removed for consistency.
2. **Health Facility Types**: Standardized to three types:
   * Hospital (HOSP)
   * Health Centre (HC)
   * Dispensary (DISP)
3. **Ownership Types**: Simplified to two categories:
   * **Public**: Assumed for anything not explicitly stated as private.
   * **Private**: Explicitly labelled as such.
4. **Star/Rating Levels**: Confirmed to be consistent across all datasets (0, 1, and 2).

### ****Cleaned Data Updates:****

#### ****Mwanza\_r:****

1. Conducted a spell check.
2. Copied **HF Name**, **HF Type**, **Owner**, and **Star Rating** into the clean dataset.
3. Added **Mwanza** to the region column.
4. Standardized **HF Type**:
   * HC = Health Centre
   * HOSP = Hospital
   * DISP = Dispensary
5. Removed redundant terms (**Hospital**, **Dispensary**, **Health Centre**) from the **HF Name** column.
6. Standardized **Ownership Type**:
   * 1 = Public
   * 2 = Private
   * 0 and ?? = N/a (not available).

#### ****Pwani\_r:****

1. Conducted a spell check.
2. Removed breaks in data (blanks, SN, NAME OF FACILITIES, STAR RATING).
3. Removed the **Ref.** column.
4. Inserted **HF Type** column using the formula:

plaintext

Copy

=IF(ISNUMBER(SEARCH("disp", A1)), "Dispensary", IF(ISNUMBER(SEARCH("hosp", A1)), "Hospital", "N/A"))

1. Removed duplicates.
2. Copied cleaned data to the main dataset.
3. Added **Pwani** to the region column and marked **Ownership** as N/a (data not available).

#### ****DSM (Ilala)\_r:****

1. Conducted a spell check.
2. Unmerged all cells.
3. Copied **Name**, **HC Type**, **Star Rating**, and **Ownership Type** to the clean dataset.
4. Standardized **Ownership Type** using the formula:

plaintext

Copy

=IF(ISNUMBER(SEARCH("private", F653)), "Private", "Public")

1. Added **Ilala** to the region column.

### ****Additional Changes:****

* Removed **‘-Star’** from the **Star Rating** column.
* Standardized formatting:
  + Uniform font and text size applied.
  + Removed unnecessary capitalization.
* Conducted a duplicate check across all datasets to ensure data integrity.

### ****Next Steps:****

Please review the cleaned dataset and let me know if any further adjustments are required. I’m happy to address any questions or concerns.

Thank you for your time, and I look forward to your feedback.