Name: Roaa Mohamed Saad Mohamed

Round Code: DEPI_GIZ2_DAT2_S2

Job Profile: Microsoft Power BI Engineer

Data Cleaning:

To ensure data accuracy and consistency, several data cleaning steps were applied across different tables. These steps include handling duplicates, renaming tables, standardizing data types, and reordering columns.

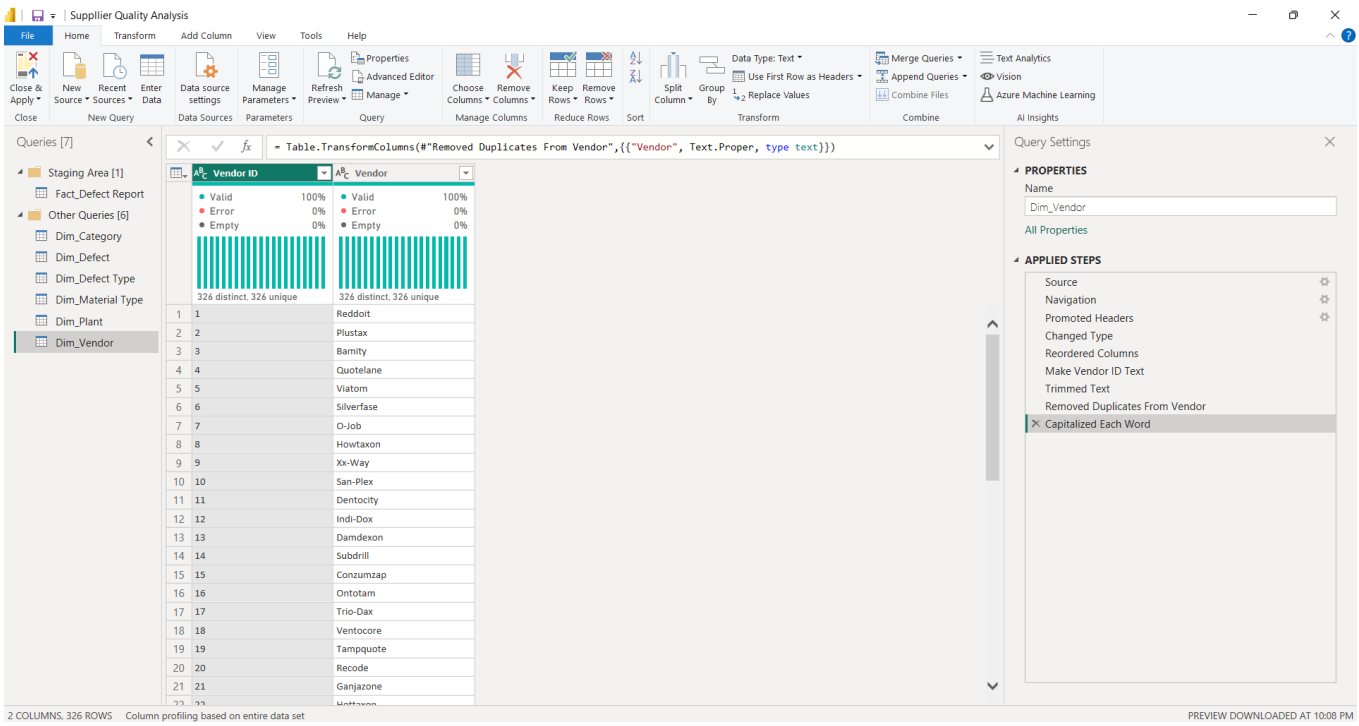
1. Dim_Vendor (Supplier Table)

Issue: Duplicate records in the Vendor column.

Inconsistent naming formats (uppercase, lowercase, extra spaces).

Action Taken:

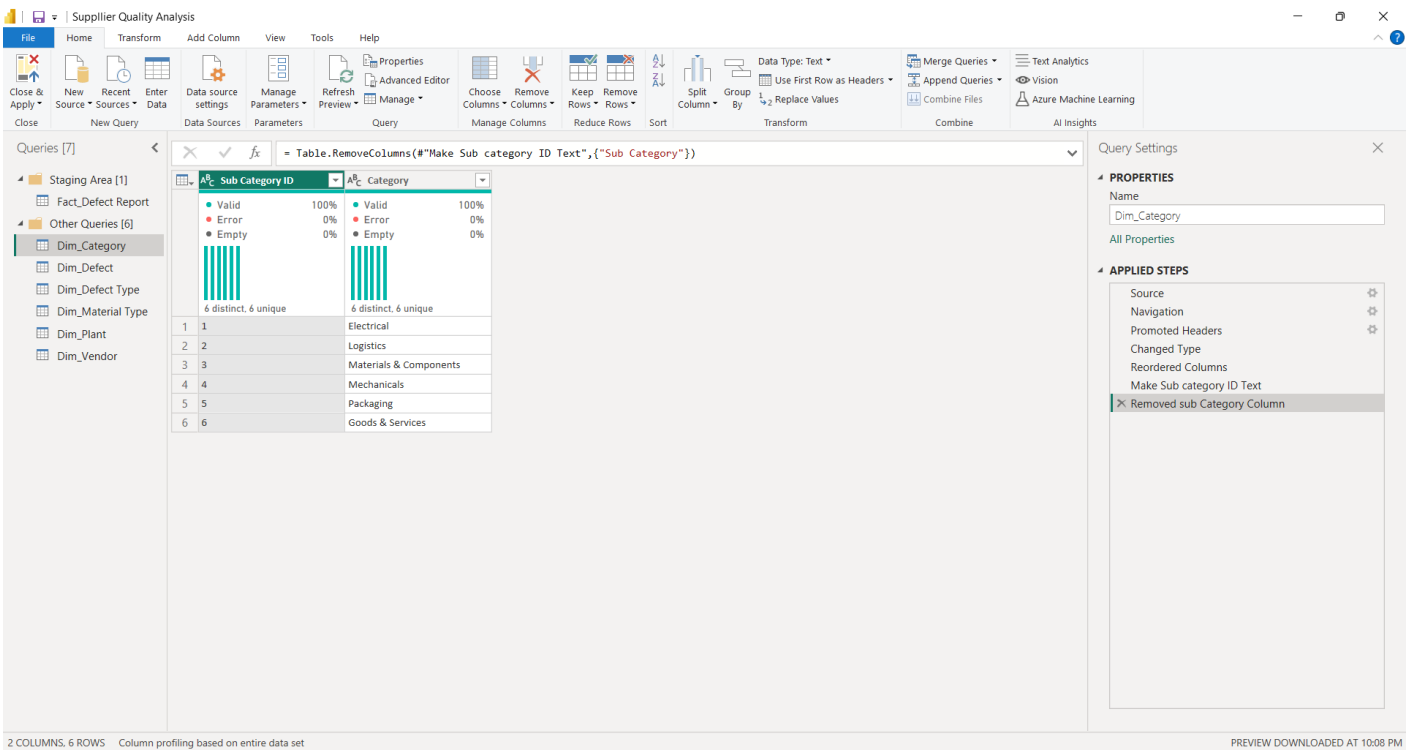
- Trimmed spaces.
- Applied Remove Duplicates in Power Query based on Vendor ID to ensure each supplier appears only once.
- Ensured all Vendor names were correctly filled to avoid missing values.
- Capitalized Each Word in the Vendor column for consistency.
- Renamed table from Vendor to Dim_Vendor for consistency.
- Converted Vendor ID to Text format.
- Moved Vendor ID to be the first column.



2. Dim_Category

Issue: Category and Sub Category had identical values , creating redundancy.
Action Taken:

- Kept only one column (Category) and removed Sub Category since it was redundant.
- Ensured unique categories were correctly classified.
- Trimmed spaces.
- Renamed table from Category to Dim_Category.
- Converted Category ID to Text format.
- Moved Category ID to be the first column.



3. Dim_Defect

Issue: Duplicate defect records found.

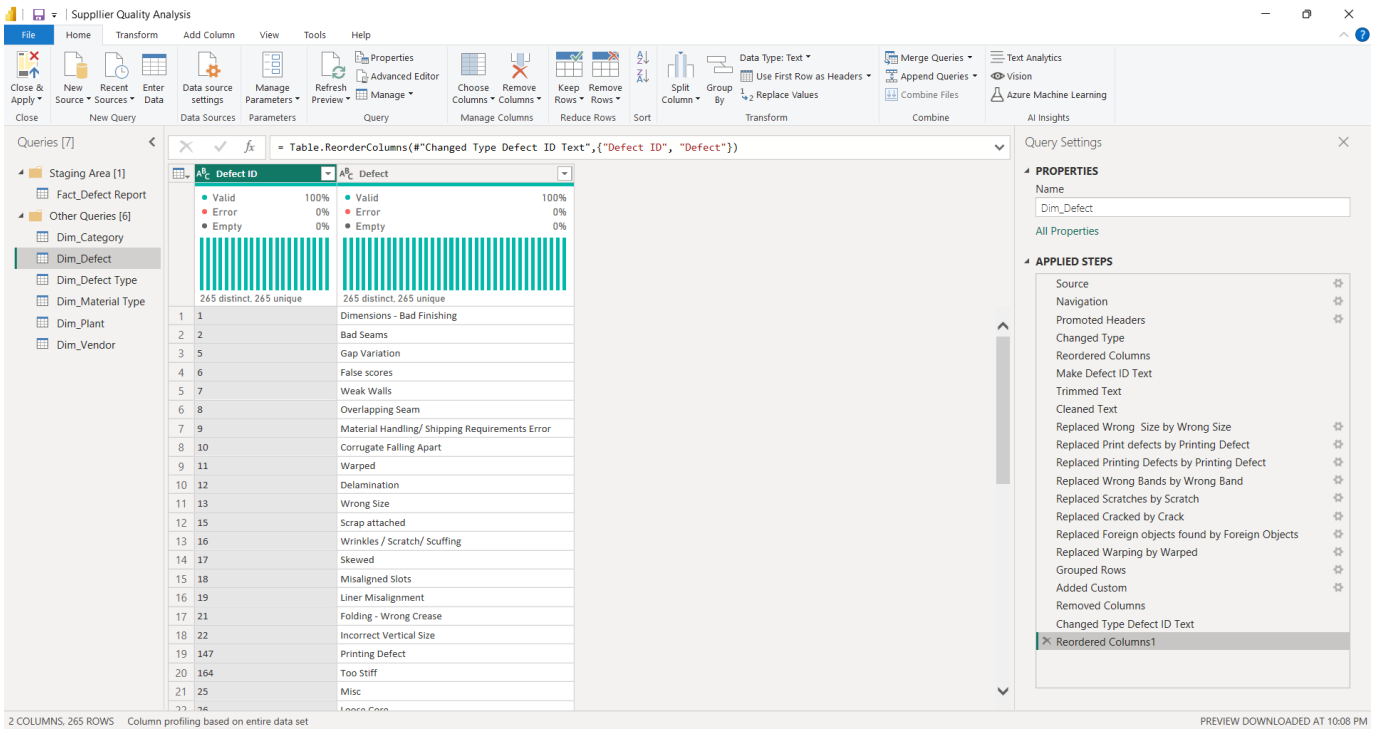
Some defect names had slight variations (e.g., "Print defects" vs. "Printing Defect").

Action Taken:

- Trimmed spaces and cleaned text to remove inconsistencies.
- Converted Defect ID to Text format for consistency.
- Renamed table from Defect to Dim_Defect for clarity.
- **Replace Inconsistent Values:**
    - "Wrong Size" → "Wrong Size"
    - "Print defects" → "Printing Defect"
    - "Printing Defects" → "Printing Defect"
    - "Wrong Bands" → "Wrong Band"
    - "Scratches" → "Scratch"
    - "Cracked" → "Crack"
    - "Foreign objects found" → "Foreign Objects"
    - "Warping" → "Warped"

- Reordered columns for better readability.
- Grouped defect records to consolidate similar issues.
- Removed unnecessary columns to optimize data size.
- Removed duplicates based on Defect ID to ensure uniqueness.

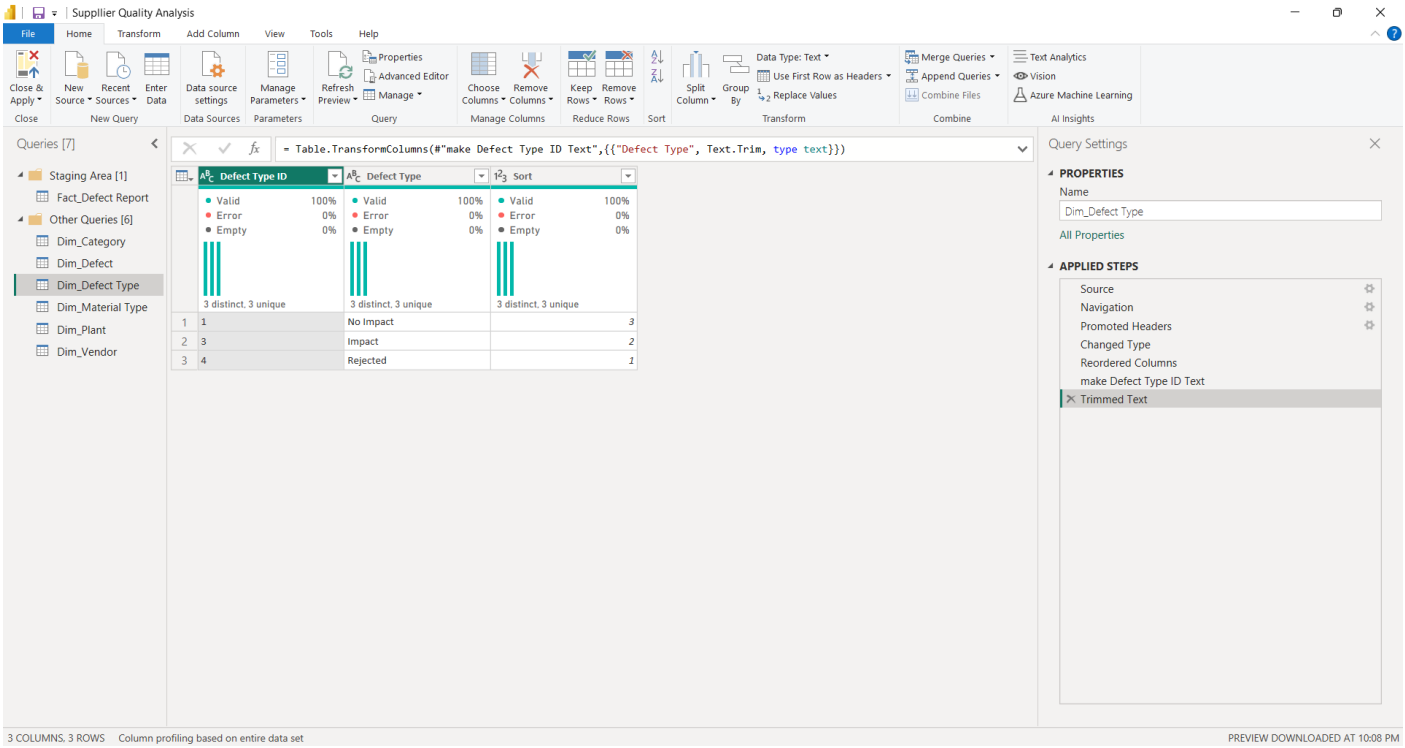
Dim_Defect


Supplier Quality Analysis
= Table.ReorderColumns(#"Changed Type Defect ID Text",{"Defect ID", "Defect"})
Queries [7]
Staging Area [1]
Fact_Defect Report
Other Queries [6]
Dim_Category
Dim_Defect
Dim_Defect Type
Dim_Material Type
Dim_Plant
Dim_Vendor
Defect ID
Defect
Valid 100%
Error 0%
Empty 0%
265 distinct, 265 unique
1 1 Dimensions - Bad Finishing
2 2 Bad Seams
3 5 Gap Variation
4 6 False scores
5 7 Weak Walls
6 8 Overlapping Seam
7 9 Material Handling/ Shipping Requirements Error
8 10 Corrugate Falling Apart
9 11 Warped
10 12 Delamination
11 13 Wrong Size
12 15 Scrap attached
13 16 Wrinkles / Scratch/ Scuffing
14 17 Skewed
15 18 Misaligned Slots
16 19 Liner Misalignment
17 21 Folding - Wrong Crease
18 22 Incorrect Vertical Size
19 147 Printing Defect
20 164 Too Stiff
21 25 Misc
Query Settings
PROPERTIES
Name
Dim_Defect
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
Reordered Columns
Make Defect ID Text
Trimmed Text
Cleaned Text
Replaced Wrong Size by Wrong Size
Replaced Print defects by Printing Defect
Replaced Printing Defects by Printing Defect
Replaced Wrong Bands by Wrong Band
Replaced Scratches by Scratch
Replaced Cracked by Crack
Replaced Foreign objects found by Foreign Objects
Replaced Warping by Warped
Grouped Rows
Added Custom
Removed Columns
Changed Type Defect ID Text
Reordered Columns1
2 COLUMNS, 265 ROWS
Column profiling based on entire data set
PREVIEW DOWNLOADED AT 10:08 PM

4. Dim_Defect Type

Action Taken:

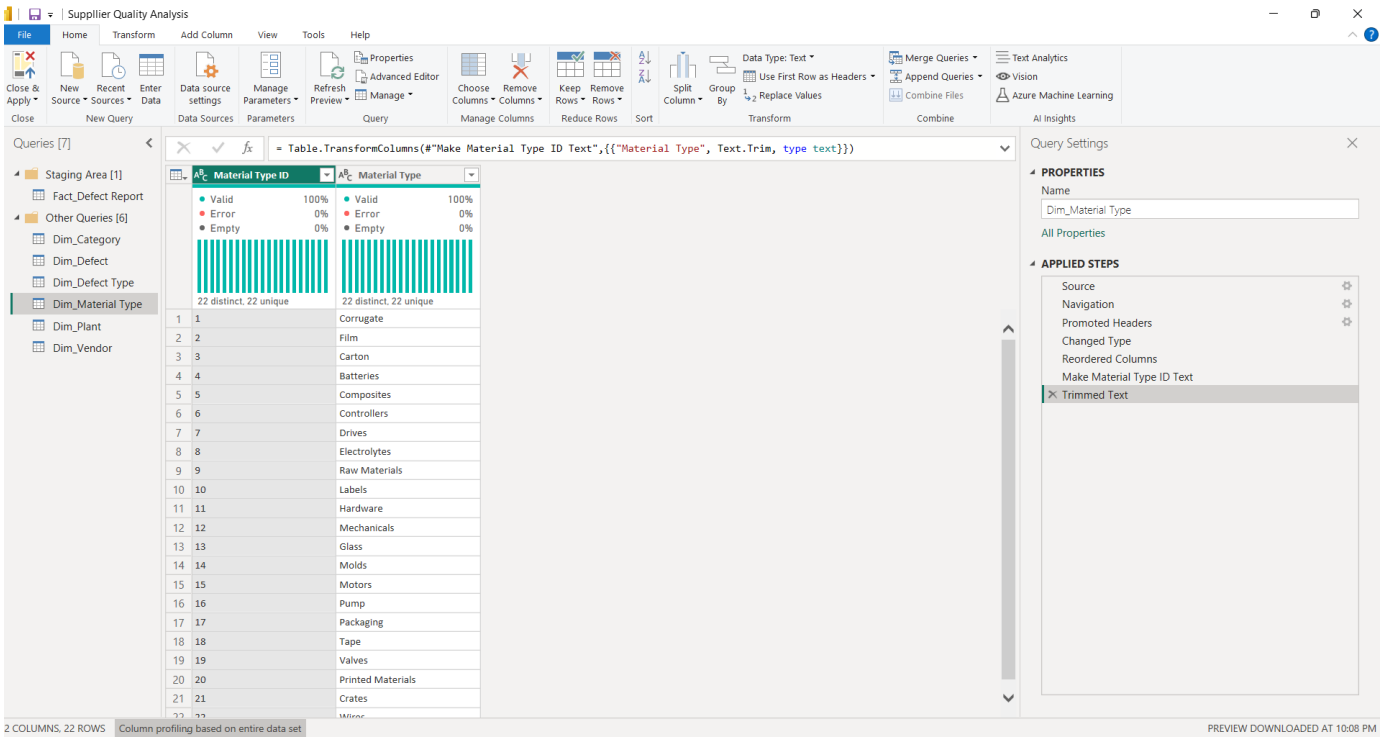
- Trimmed spaces.
- Ensured all Defect Type were correctly filled.
- Renamed table from Defect Type to Dim_Defect Type.
- Converted Defect Type ID to Text format.
- Moved Defect Type ID to be the first column.



5. Dim_Material Type

Action Taken:

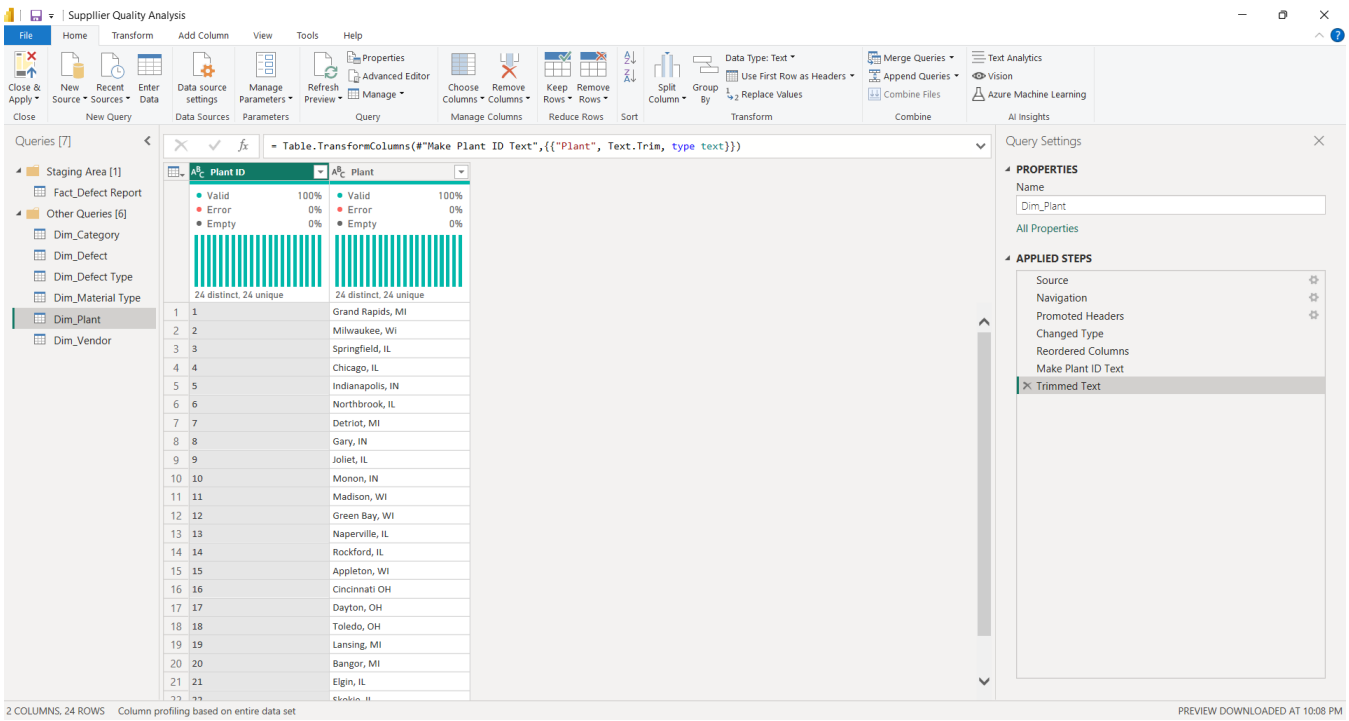
- Trimmed spaces.
- Ensured all Material Type was correctly filled.
- Renamed table from Material Type to Dim_Material Type.
- Converted Material Type ID to Text format.
- Moved Material Type ID to be the first column.



6. Dim_Plant

Action Taken:

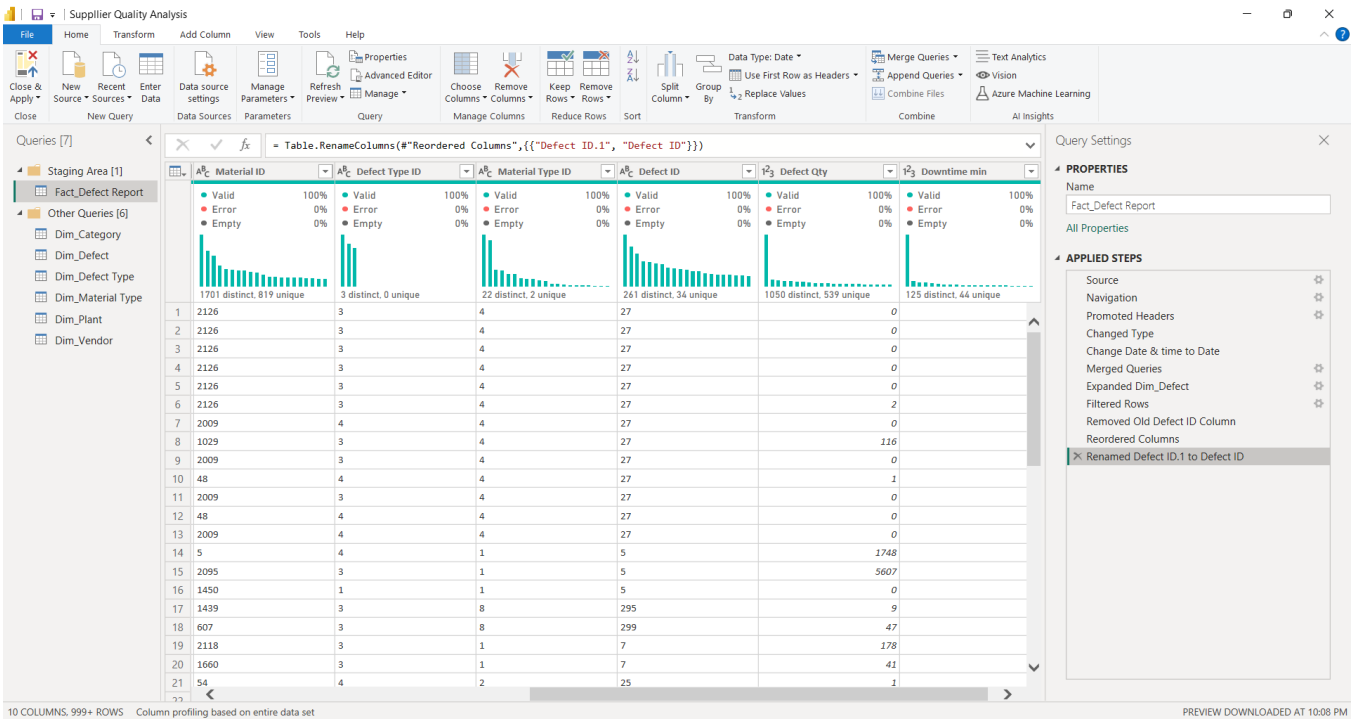
- Trimmed spaces.
- Ensured all plant locations were correctly filled.
- Renamed table from Plant to Dim_Plant.
- Converted Plant ID to Text format.
- Moved Plant ID to be the first column.



7. Fact_Defect Report

Issue: Data consistency issues in defect tracking.
Action Taken:

- Ensured all foreign keys (e.g., Vendor ID, Plant ID, Defect ID) exist in corresponding dimension tables.
- Standardized date format for Date column.
- Renamed table from Defect Report to Fact_Defect Report.
- Converted all IDs to Text format for consistency.
- Moved Date to be the first column for chronological analysis.
- Trimmed spaces.



Summary of Key Data Cleaning Steps

1. **Trimmed & Cleaned Text**
   - Removed extra spaces and inconsistencies in defect descriptions.
2. **Standardized Defect Names**
   - Unified defect names (e.g., "Print defects" → "Printing Defect").
3. **Handled Duplicates**
   - Removed duplicate records based on **ID** to ensure uniqueness.
4. **Formatted Data**
   - Converted **ID** to **Text** format for consistency.
   - Reordered columns for better structure.
5. **Filtered & Refined Data**
   - Removed unnecessary columns.
   - Kept only valid defect records.