

**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.

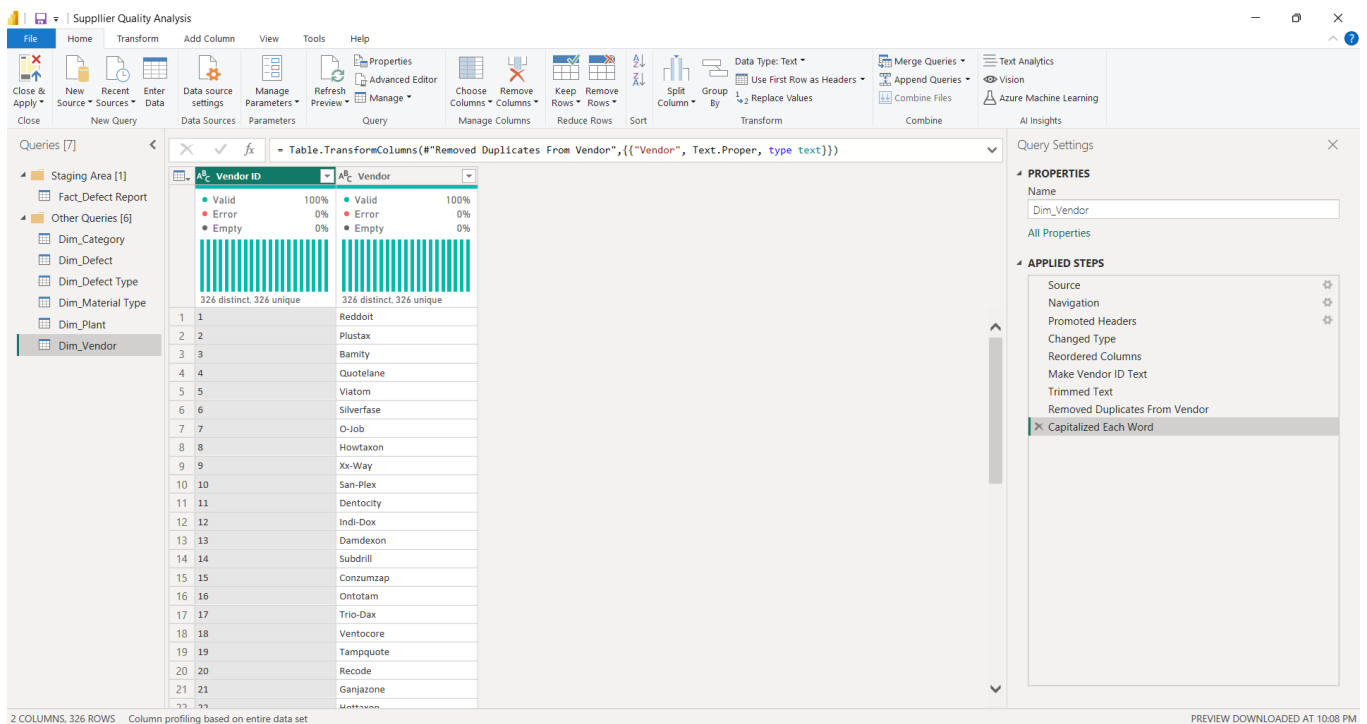
## 1. Dim\_Vendor (Supplier Table)

**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.

The screenshot displays the Microsoft Power BI Desktop interface for a project named 'Supplier Quality Analysis'. The main view shows a table with two columns: 'Sub Category ID' and 'Category'. Both columns have a data type of 'Text' and are marked as 'Valid' with 100% accuracy. The table contains 6 rows of data, with 6 distinct values in each column. The 'Sub Category ID' column has values 1 through 6, and the 'Category' column has corresponding values: Electrical, Logistics, Materials & Components, Mechanicals, Packaging, and Goods & Services.

The 'Query Settings' pane on the right shows the 'APPLIED STEPS' for the query. The steps listed are: Source, Navigation, Promoted Headers, Changed Type, Reordered Columns, Make Sub category ID Text, and Removed sub Category Column. The 'Removed sub Category Column' step is currently selected.

At the bottom of the window, a status bar indicates '2 COLUMNS, 6 ROWS' and 'Column profiling based on entire data set'. A timestamp in the bottom right corner reads 'PREVIEW DOWNLOADED AT 10:08 PM'.

### 3. Dim\_Defect

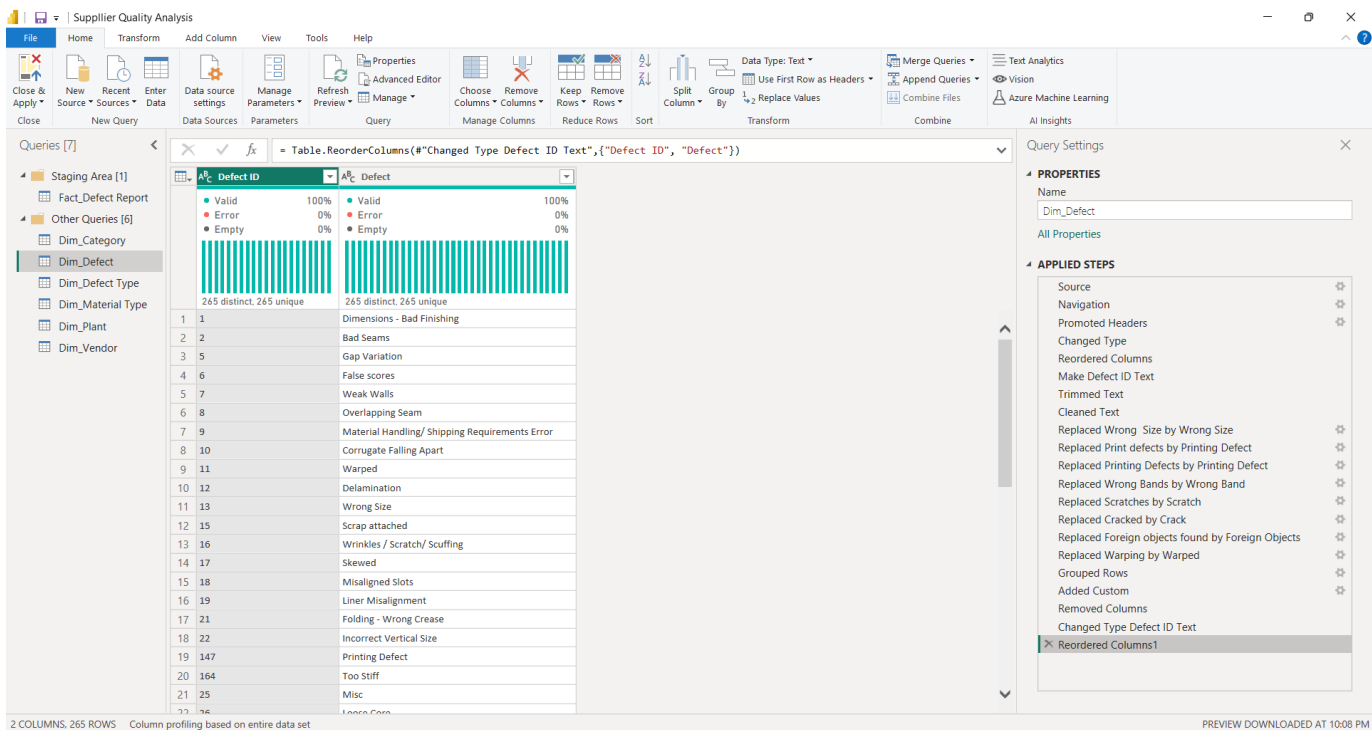
**Issue:** Duplicate defect records found.

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



## 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.

Supplier Quality Analysis

File Home Transform Add Column View Tools Help

Close & Apply New Recent Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Data Type: Text Use First Row as Headers Replace Values Merge Queries Append Queries Combine Files Combine Text Analytics Vision Azure Machine Learning

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

Table.TransformColumns(#"make Defect Type ID Text",{{"Defect Type", Text.Trim, type text}})

| Defect Type ID | Defect Type | Sort |
|----------------|-------------|------|
| 1              | No Impact   | 3    |
| 2              | Impact      | 2    |
| 3              | Rejected    | 1    |

3 COLUMNS, 3 ROWS Column profiling based on entire data set

PREVIEW DOWNLOADED AT 10:08 PM

## 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.

Supplier Quality Analysis

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Query Choose Remove Keep Remove Split Group Data Type: Text Merge Queries Text Analytics Append Queries Vision Azure Machine Learning Combine Files AI Insights

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

Table.TransformColumns(#"Make Material Type ID Text",{("Material Type", Text.Trim, type text)})

| Material Type ID | Material Type     |
|------------------|-------------------|
| 1                | Corrugate         |
| 2                | Film              |
| 3                | Carton            |
| 4                | Batteries         |
| 5                | Composites        |
| 6                | Controllers       |
| 7                | Drives            |
| 8                | Electrolytes      |
| 9                | Raw Materials     |
| 10               | Labels            |
| 11               | Hardware          |
| 12               | Mechanicals       |
| 13               | Glass             |
| 14               | Molds             |
| 15               | Motors            |
| 16               | Pump              |
| 17               | Packaging         |
| 18               | Tape              |
| 19               | Valves            |
| 20               | Printed Materials |
| 21               | Crates            |
| 22               | Misc              |

22 distinct, 22 unique

22 distinct, 22 unique

Column profiling based on entire data set

PREVIEW DOWNLOADED AT 10:08 PM

Query Settings

PROPERTIES

Name  
Dim\_Material Type

All Properties

APPLIED STEPS

Source  
Navigation  
Promoted Headers  
Changed Type  
Reordered Columns  
Make Material Type ID Text  
Trimmed Text

## 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.

Supplier Quality Analysis

File Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Enter Data Data source settings Data Sources Manage Parameters Refresh Preview Advanced Editor Choose Columns Remove Columns Keep Rows Remove Rows Sort Split Column Group By Data Type: Text Merge Queries Append Queries Combine Files Text Analytics Vision Azure Machine Learning

Queries [7]

Staging Area [1]

Other Queries [6]

Dim\_Category

Dim\_Defect

Dim\_Defect Type

Dim\_Material Type

Dim\_Plant

Dim\_Vendor

Table.TransformColumns(#"Make Plant ID Text",{"Plant", Text.Trim, type text})

Plant ID Plant

Valid 100% Valid 100%

Error 0% Error 0%

Empty 0% Empty 0%

24 distinct, 24 unique 24 distinct, 24 unique

1 1 Grand Rapids, MI

2 2 Milwaukee, WI

3 3 Springfield, IL

4 4 Chicago, IL

5 5 Indianapolis, IN

6 6 Northbrook, IL

7 7 Detroit, MI

8 8 Gary, IN

9 9 Joliet, IL

10 10 Monro, IN

11 11 Madison, WI

12 12 Green Bay, WI

13 13 Naperville, IL

14 14 Rockford, IL

15 15 Appleton, WI

16 16 Cincinnati OH

17 17 Dayton, OH

18 18 Toledo, OH

19 19 Lansing, MI

20 20 Bangor, MI

21 21 Elgin, IL

22 22 Elmhurst, IL

Query Settings

PROPERTIES

Name

Dim\_Plant

All Properties

APPLIED STEPS

Source

Navigation

Promoted Headers

Changed Type

Reordered Columns

Make Plant ID Text

Trimmed Text

2 COLUMNS, 24 ROWS Column profiling based on entire data set

PREVIEW DOWNLOADED AT 10:08 PM

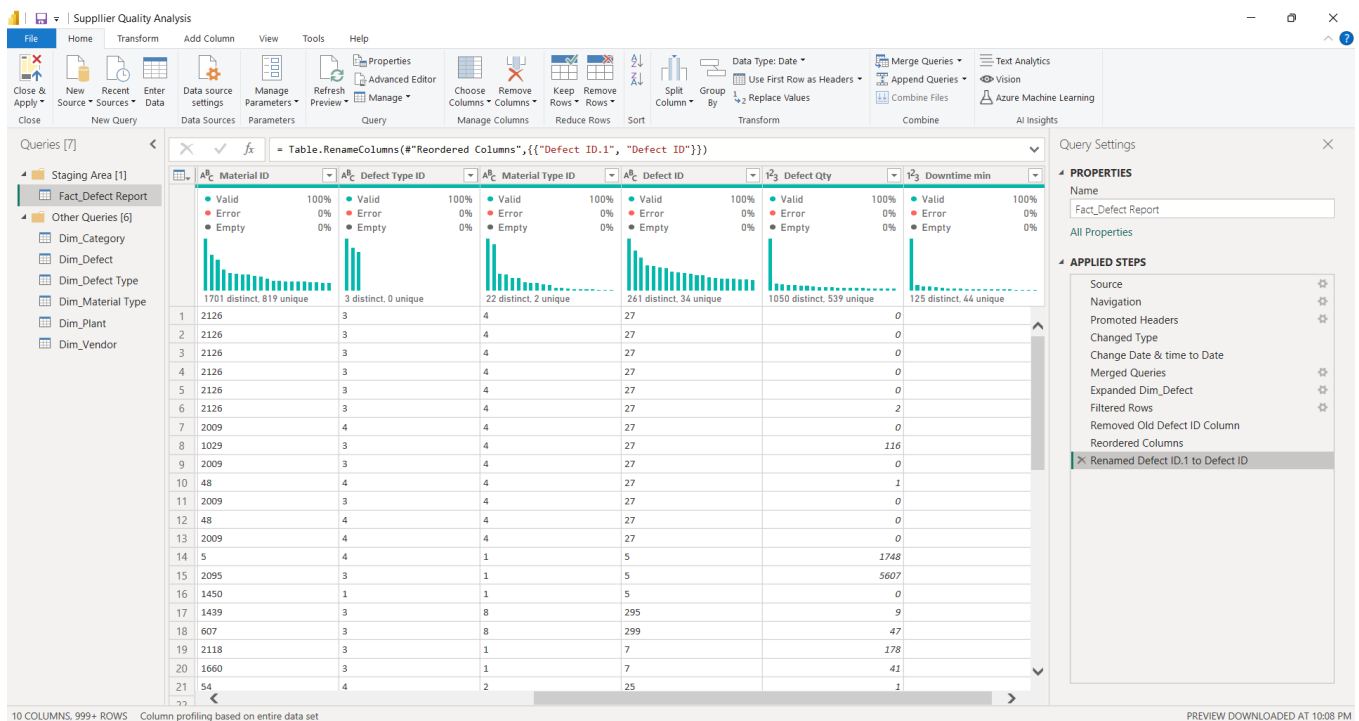


## 7. Fact\_Defect Report

**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.