Cleaning and Modeling Data in PBI

Contents

[Part 1: Cleaning Data in Power Query 1](#_Toc172170040)

[1. Import Data into Power Query 1](#_Toc172170041)

[2. Clean the Customers Table 1](#_Toc172170042)

[3. Clean the Products Table 2](#_Toc172170043)

[4. Clean the Orders Table 2](#_Toc172170044)

[5. Load Cleaned Data Back to Power BI 2](#_Toc172170045)

[Part 2: Setting Up Relationships in Power BI 3](#_Toc172170046)

[1. Open the Relationships View 3](#_Toc172170047)

[2. Create Relationships Manually 3](#_Toc172170048)

[Customers and Orders Relationship: 3](#_Toc172170049)

[Products and Orders Relationship: 3](#_Toc172170050)

[3. Verify Relationships 3](#_Toc172170051)

Here are some suggestions on how to clean the customers, products, and orders data tables in Power Query and then set up the relationships between the tables after you have loaded the data into Power BI:

# Part 1: Cleaning Data in Power Query

## 1. Import Data into Power Query

1. Open Power BI Desktop.
2. Click on Get Data and select the data source (e.g., Excel, CSV, etc.).
3. Load each dataset (Customers, Products, and Orders) into Power BI.

## 2. Clean the Customers Table

1. Go to the Home tab and click on Transform Data to open Power Query Editor.
2. **Split Customer Names**:
   * Rightclick the CustomerName column.
   * Select Split Column > By Delimiter > Lowercase to Uppercase.
   * Hold down Ctrl while clicking on the column headers for all the new columns to select them (customerName.1, customerName2, customerName3)
   * Right click on one of the selected column headers and select Merge Columns > Separator: Space and New Column Name: Customer Name.
3. **Look at the Column Profiles to Identify Problems:**
   * Click on the View tab at the top of the screen and check Column profile.
   * Click on each column header to see how many empty/null values, how many distinct values, etc. on the left side of the column profile.
   * The frequency chart on the right shows how many of each value in the column. If you sort the column first, this is useful to see if there are misspellings or alternative spellings that need to be remedied.
4. **Fix City Misspellings**:
   * Select the City column.
   * Use the Replace Values option in the Transform tab to correct misspellings and alternate spellings (e.g., replace Marakesh with Marrakesh, NYC with New York).
   * Do the same for country (e.g. Replace United States with USA)
5. **Remove Leading/Trailing Spaces**:
   * Select all columns.
   * Click on Transform > Format > Trim.
6. **Fill Missing Values in Country Column**:
   * Identify the rows with missing Country values.
   * Manually infer the correct Country based on the City column.
   * Select the Country column and use the Replace Values option to fill in the correct country values for those rows.
7. **Remove Duplicates:**
   * Hold down Ctrl while selecting all columns except the CustomerID column.
   * Right Click on any of the selected column headers and select Remove Duplicates.
8. **Recheck Column Profiles:**
   * Click on the View tab at the top of the screen and check Column profile.
   * Click on each column header to see how many empty/null values, how many distinct values, etc. on the left side of the column profile.
9. **Remove Unnecessary Columns**:
   * Remove the Gender column or any other columns if they are not needed.

## 3. Clean the Products Table

1. **Remove Duplicates**:
   * Select the ProductID column.
   * Click on Remove Duplicates.
2. **Remove Currency Symbols**:
   * Select the Price column.
   * Use Replace Values to remove GBP.
   * Change the column data type to Decimal Number.

## 4. Clean the Orders Table

1. **Remove Duplicates**:
   * Select the OrderID column.
   * Click on Remove Duplicates.
2. **Fix Date Formats**:
   * Ensure the OrderDate column is in Date format.
   * If not, change the data type to Date.

## 5. Load Cleaned Data Back to Power BI

Click Close & Apply in the Home tab to load the cleaned data back into Power BI.

# Part 2: Setting Up Relationships in Power BI

## 1. Open the Relationships View

1. Click on the Model icon on the left-hand side to open the Relationships view.

## 2. Create Relationships Manually

### Customers and Orders Relationship:

1. Drag the CustomerID column from the Customers table to the CustID column in the Orders table.
2. Ensure the relationship is set to Many to One (\*:1) with Customers as the one side.

### Products and Orders Relationship:

1. Drag the ProductID column from the Products table to the ProdID column in the Orders table.
2. Ensure the relationship is set to Many to One (\*:1) with Products as the one side.

## 3. Verify Relationships

1. Click on each relationship line to open the relationship editor.
2. Ensure the following settings:
   * Cardinality: Many to One (\*:1)
   * Cross filter direction: Single
   * Active: Checked