**Data Preparation Explained: Power BI (Population\_2010-2019)**

**------------------------------------------------**

**📊 Data Cleaning for Power BI**

Data cleaning, also known as **data cleansing** or **data scrubbing**, is the process of detecting and correcting (or removing) corrupt or inaccurate records from a dataset. In the context of **Power BI**, clean data is absolutely crucial for generating reliable and accurate reports, dashboards, and analyses. Without proper data cleaning, insights derived from your data can be misleading, leading to poor decision-making.

Power BI leverages **Power Query Editor** as its primary tool for data transformation and cleaning. Power Query provides a user-friendly interface and a powerful backend (M language) to perform a wide array of cleaning operations.

**🚀 Accessing Power Query Editor**

To begin data cleaning in Power BI, you first need to load your data and access the Power Query Editor:

1. **Open Power BI Desktop**.
2. Go to the **Home** tab.
3. Click **"Get Data"** to import your data source (e.g., Excel, SQL Server, Web).
4. After selecting your data, instead of clicking "Load", click **"Transform Data"**. This will open the Power Query Editor.

In the Power Query Editor, you'll see your data as a table, and on the right-hand side, there's an **"Applied Steps"** pane. Every cleaning or transformation action you perform will be recorded here, allowing you to review, modify, or delete steps.

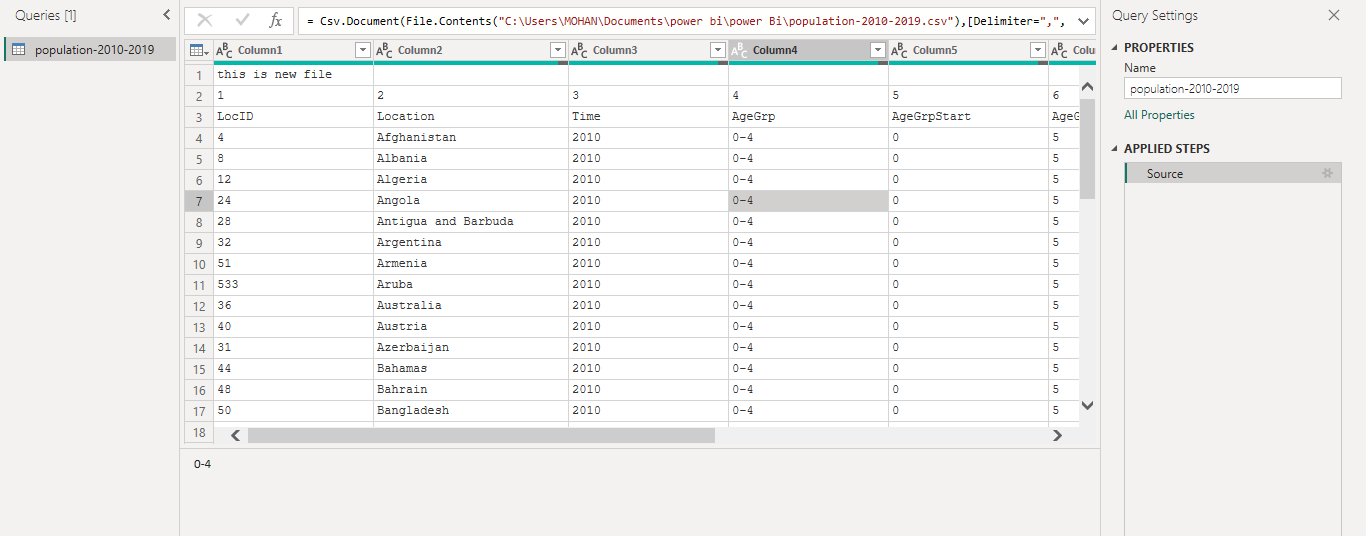
**✨ Common Data Cleaning Steps in Power Query**

Here are some of the most common data cleaning operations you'll perform in Power Query:

**1. Handling Missing Values (Nulls)**

Missing values (often represented as null or blank) can skew aggregations and lead to errors.

* **Identify Missing Values**: Visually inspect columns, or use "View" > "Column quality" to see the percentage of valid, error, and empty values.
* **Remove Rows with Nulls**: Right-click on the column header, select **"Remove Empty"** or **"Remove Errors"**. You can also choose **"Remove Rows"** from the Home tab and specify criteria.
* **Replace Values**: Right-click on the column header, select **"Replace Values..."**. Enter the value to find (e.g., null or (Blank)) and the value to replace it with (e.g., 0, N/A, or the average/median for numerical columns).
* **Fill Down/Up**: For columns where null values logically represent the value from the row above or below (common in poorly structured reports), right-click the column, select **"Fill"**, then **"Down"** or **"Up"**.



**2. Correcting Data Types**

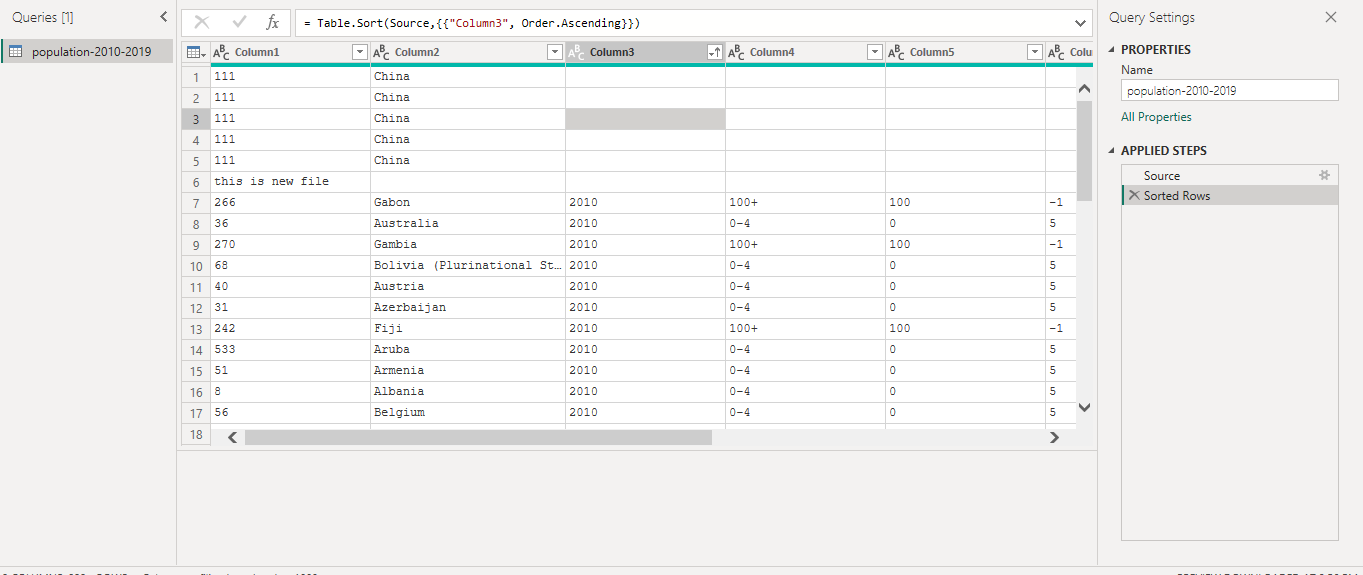
Incorrect data types are a frequent source of errors and performance issues. For example, numbers stored as text cannot be used in calculations.

* **Change Data Type**: Right-click the column header, select **"Change Type"**, and choose the appropriate type (e.g., Whole Number, Decimal Number, Date, Text, True/False).
* **Locale Specific Types**: For dates, numbers, or currencies, consider using "Using Locale..." to ensure correct interpretation based on regional settings.

**3. Removing Duplicates**

Duplicate rows can inflate counts and distort analysis.

* **Remove Duplicates**: Select one or more columns that uniquely identify a row (e.g., an ID column, or all columns if an entire row should be unique). Right-click the header(s) and select **"Remove Duplicates"**. Power Query keeps the first occurrence.

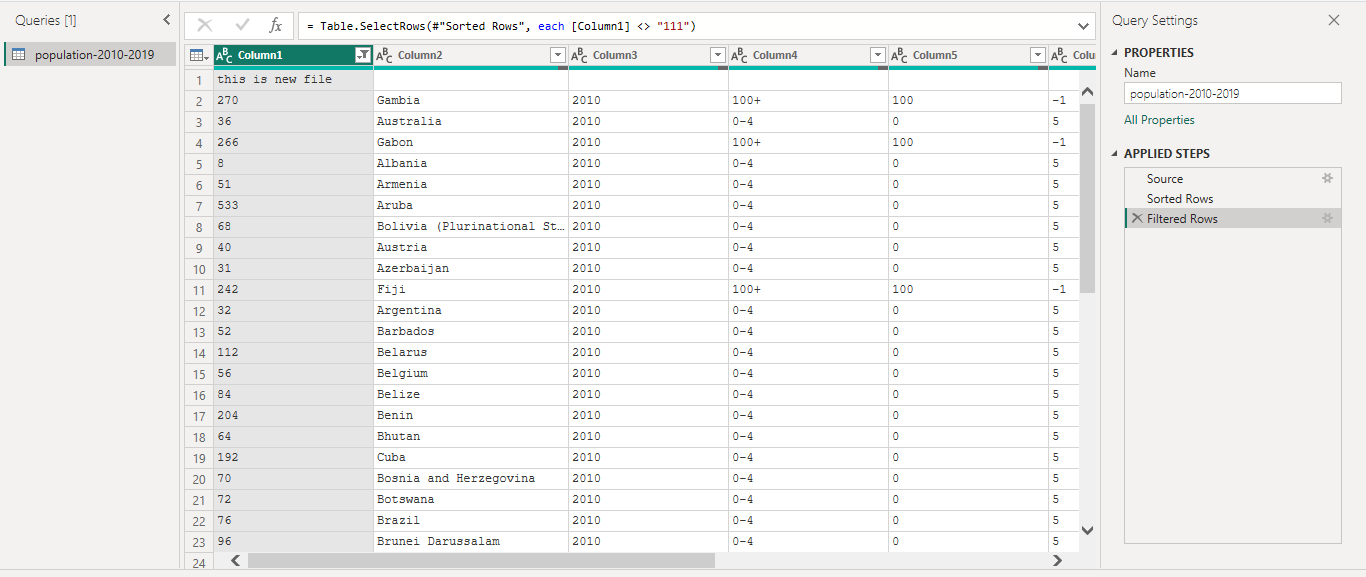


**4. Standardizing Text (Trim, Clean, Case)**

Inconsistent text formatting can prevent accurate matching, grouping, and filtering.

* **Trim**: Removes leading and trailing whitespace. Select the column, go to **Transform** tab > **"Format"** > **"Trim"**.
* **Clean**: Removes non-printable characters. Select the column, go to **Transform** tab > **"Format"** > **"Clean"**.
* **Change Case**: Convert text to lowercase, uppercase, capitalize each word, etc. Select the column, go to **Transform** tab > **"Format"** > **"Lowercase"** / **"Uppercase"** /

**"Capitalize Each Word"**.



**5. Splitting and Merging Columns**

Often, a single column contains multiple pieces of information (e.g., "First Name Last Name") or you need to combine related information.

* **Split Column**: Select the column, go to **Transform** tab > **"Split Column"**. You can split by delimiter (e.g., space, comma), number of characters, or by positions.
* **Merge Columns**: Select two or more columns (Ctrl+Click), right-click and choose **"Merge Columns"**. Specify a separator (e.g., space, comma) if needed.

**6. Unpivoting/Pivoting Data**

These are powerful transformations for reshaping your data from wide to tall (unpivot) or tall to wide (pivot), which is essential for analytical purposes.

* **Unpivot Columns**: When you have multiple columns that are really different categories of the same measure (e.g., "Sales 2021", "Sales 2022"), unpivoting transforms them into two columns: "Attribute" (for the year) and "Value" (for sales). Select the measure columns, go to **Transform** tab > **"Unpivot Columns"** (or "Unpivot Other Columns").
* **Pivot Column**: The opposite of unpivot. Used when you have distinct values in one column that you want to become new column headers (e.g., turning product names into columns with their corresponding sales). Select the column containing the values to become new headers, go to **Transform** tab > **"Pivot Column"**.

**7. Filtering Rows**

Remove rows that don't meet your criteria or are irrelevant.

* **Filter Rows**: Click the filter icon (down arrow) on the column header. You can select specific values, or use text filters (Begins With, Contains, Does Not Equal), number filters (Greater Than, Less Than), or date filters.

**8. Replacing Values**

This is a versatile tool for correcting typos, standardizing inconsistent entries, or removing unwanted characters.

* **Replace Values**: Select the column, right-click, and choose **"Replace Values..."**. Enter the value to find and the value to replace with. This is excellent for fixing common misspellings (e.g., replacing "Calif." with "California").

**9. Handling Errors**

Errors in data can arise from incorrect data types, division by zero, or other issues.

* **Remove Errors**: Right-click the column header and select **"Remove Errors"**. This will delete rows containing errors in that specific column.
* **Replace Errors**: Right-click the column header and select **"Replace Errors..."**. You can replace error values with a specified value (e.g., 0, N/A).