**Guide for basic data cleaning in Power Query**

This document outlines how to perform various data cleaning and transformation tasks on a sample dataset using Power Query Editor in Excel. Each section details the task, steps to achieve it, the importance of the function, and a real-life example of its application.  
  
**Power Query Operations Used**

* **Add Column**
* **Conditional Column**
* **Group By**
* **Filter**
* **Sort**
* **Transform**
* **Remove Columns**

**Dataset**[**:** SampleData.xlsx](https://scottgardner368-my.sharepoint.com/:x:/g/personal/ashwini_patole_scottandgardner_com/EfCZ8ZfXTcZPpucMyvkthUMBBHlMT0nIHDi9RMow62L63A?e=l1nWKN&nav=MTVfe0FGRDY2N0RCLUQ4MEEtNEUzNS1BOEYyLTZCNjgzMTdCNkE5QX0)

**1. Filter Data by a Specific Category**

**Task:** Filter the dataset to show only "Electronics" category products.

**Steps:**

1. Go to the Power Query Editor by clicking "Get Data" -> "From File" -> "From Workbook" and selecting "SampleData.xlsx".
2. In the "Transform" tab, click on the dropdown arrow in the header of the "Category" column.
3. Select "Filter Items" -> "Equals" and choose "Electronics" from the dropdown menu.

**Importance:** Filtering helps focus on specific subsets of data, allowing for targeted analysis. It removes irrelevant information, improving data clarity and efficiency.

**Real-Life Example:** A retail store can filter sales data by product category to analyze specific product performance and identify areas for improvement.

**2. Sort Data by Sales**

**Task:** Sort the dataset by "Sales" in descending order (highest to lowest).

**Steps:**

1. Click on the header of the "Sales" column.
2. In the "Sort" section of the "Home" tab, select "Descending".

**Importance:** Sorting data by a specific column helps identify trends and outliers easily. It facilitates quick identification of top-selling products or regions.

**Real-Life Example:** An e-commerce platform can sort orders by sales value to prioritize high-value customers and optimize marketing strategies.

**3. Group Data by Region and Calculate Total Sales**

**Task:** Group data by "Region" and calculate the total sales for each region.

**Steps:**

1. Go to the "Home" tab and click on "Group By".
2. In the "Group By" window, select "Region" as the grouping column.
3. Click on the dropdown arrow in the header of the "Sales" column.
4. Select "Sum" to calculate the total sales for each region.

**Importance:** Grouping data summarizes information by categories, making it easier to identify regional sales trends and resource allocation.

**Real-Life Example:** A sales manager can group sales data by region to analyze regional performance and identify areas needing additional support or marketing efforts.

**4. Filter Sales Above a Certain Value**

**Task:** Filter the dataset to show products with sales greater than 800.

**Steps:**

1. Click on the header of the "Sales" column.
2. Go to the "Home" tab and click on the filter icon.
3. Select "Number Filters" -> "Greater Than" and enter "800" in the value box.

**Importance:** Filtering by a specific value range allows for focused analysis on high-performing products, valuable customers, or significant trends within the data.

**Real-Life Example:** A company can filter customer data by purchase amount to identify high-value customers for targeted marketing campaigns and loyalty programs.

**5. Add a Custom Column for Month of Sale**

**Task:** Create a new column that extracts the month (e.g., January, February) from the "Date" column.

**Steps:**

1. Go to the "Add Column" tab and click on "Date".
2. Select "Month From Date". A new column named "Month" will be added.

**Importance:** Adding a custom column based on existing data facilitates further analysis by category or trend. Extracting the month allows for analyzing seasonal variations or sales patterns.

**Real-Life Example:** A clothing retailer can extract the month from sales data to analyze seasonal clothing trends and adjust inventory or marketing strategies accordingly.

**6. Remove a Column from the Dataset**

**Task:** Remove the "Product ID" column.

**Steps:**

1. Click on the header of the "Product ID" column.
2. Right-click and choose "Remove Columns".

**Importance:** Removing unnecessary columns streamlines the data view and reduces file size. It improves data clarity and eliminates irrelevant information.

**Real-Life Example:** When analyzing product sales trends, the "Product ID" might not be necessary, keeping only product category and sales data for a focused analysis.

**7. Replace Values in a Column**

**Task:** Replace "West" in the "Region" column with "Western".

**Steps:**

1. Click on the header of the "Region" column.
2. Right-click and choose "Replace Values".
3. In the "Find" box, enter "West".
4. In the "Replace With" box, enter "Western".
5. Click "OK".

**8. Filter Data Based on a Date Range**

**Task:** Filter the dataset to show rows where "Date" is after 01/05/2024.

**Steps:**

1. Click on the header of the "Date" column.
2. Go to the "Home" tab and click on the filter icon.
3. Select "Date Filters" -> "After" and enter "01/05/2024" in the value box.

**Importance:** Filtering by a date range allows for analyzing specific time periods, such as quarterly or annual performance.

**Real-Life Example:** A marketing team can filter sales data by date range to analyze the impact of recent marketing campaigns or seasonal trends.

**9. Change Data Type of a Column**

**Task:** Change the "Sales" column data type from text to a whole number.

**Steps:**

1. Click on the header of the "Sales" column.
2. Go to the "Home" tab and click on the "Transform" button.
3. Select "Change Type" -> "Whole Number".

**Importance:** Ensuring correct data types is crucial for accurate calculations and analysis. Converting text values to numbers allows for mathematical operations and statistical analysis.

**Real-Life Example:** A financial analyst needs to convert sales data from text to numbers to calculate total revenue, average sales, and other financial metrics.

**10. Add a Conditional Column**

**Task:** Add a new column that labels sales as "High" if "Sales" are greater than 1000, otherwise "Low".

**Steps:**

1. Go to the "Add Column" tab and click on "Conditional Column".
2. In the "New Column Name" box, enter "Sales Category".
3. In the "Condition" dropdown, select "If".
4. In the "Value" box, enter "Sales".
5. In the "Operator" dropdown, select "Greater Than".
6. In the "Value" box, enter "1000".
7. In the "Then" box, enter "High".
8. In the "Otherwise" box, enter "Low".
9. Click "OK".

**Importance:** Conditional columns allow for categorization of data based on specific criteria, making it easier to visualize and analyse trends.

**Real-Life Example:** A customer support team can add a conditional column to categorize customer satisfaction ratings as "Satisfied", "Neutral", or "Dissatisfied" for easier analysis and identification of areas for improvement.