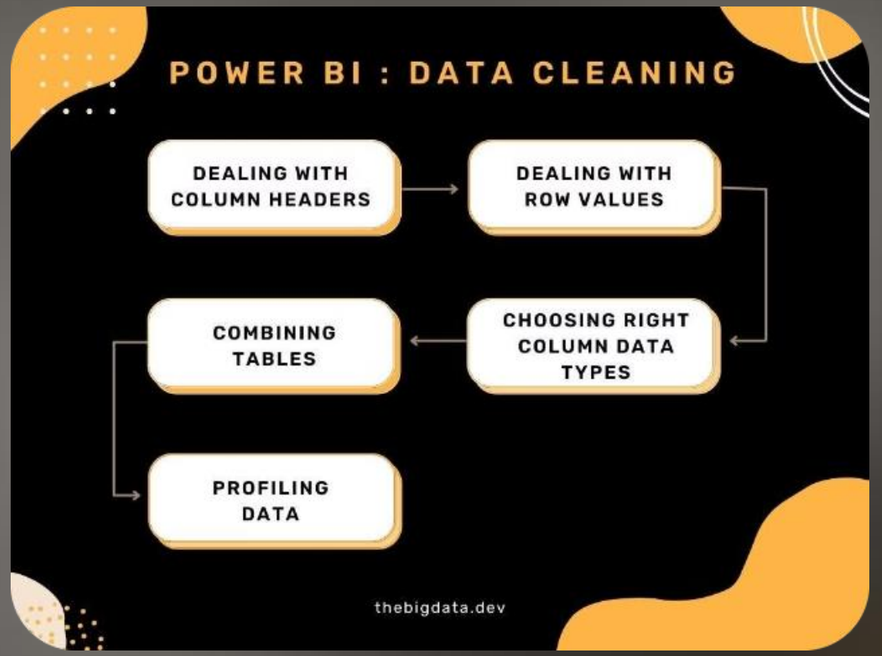
-Data Cleaning Cheat Sheet for Power BI

This cheat sheet provides essential steps and techniques for effectively cleaning and preparing your data in Power BI, empowering you to work with high-quality information and unlock valuable insights.

NB **by Naresh Boopathi**

-



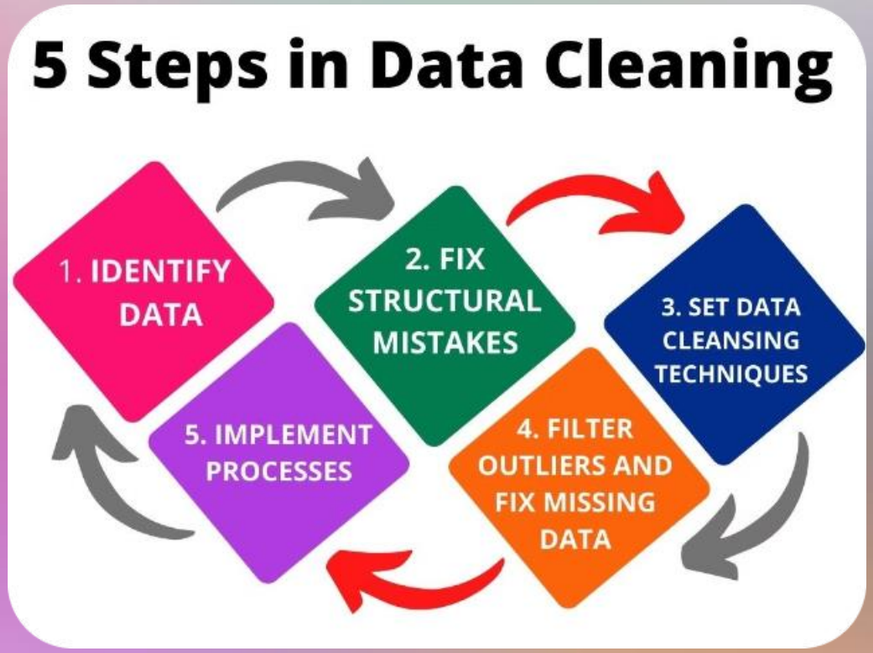
Introduction

1 Importance of Data Cleaning

Proper data cleaning is crucial for ensuring accurate and reliable analysis in Power BI.

2 Objectives of the Cheat Sheet

This cheat sheet aims to provide a concise guide to the essential data cleaning techniques in Power BI.



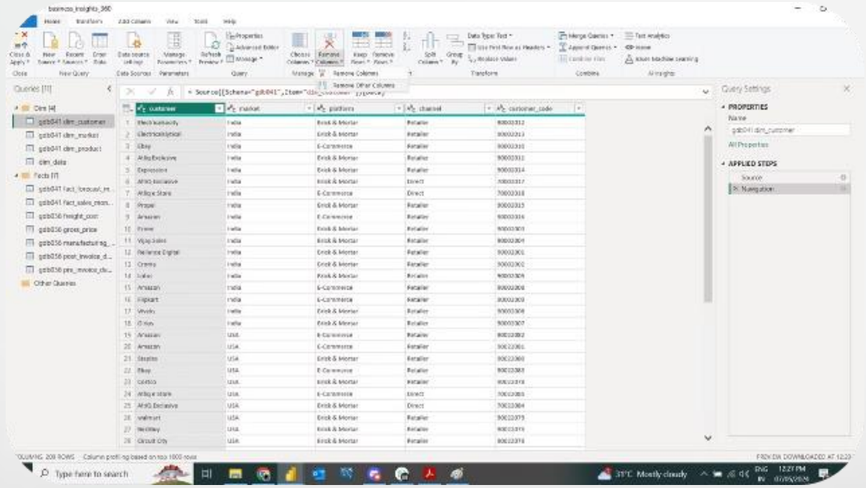
Remove Unnecessary Columns

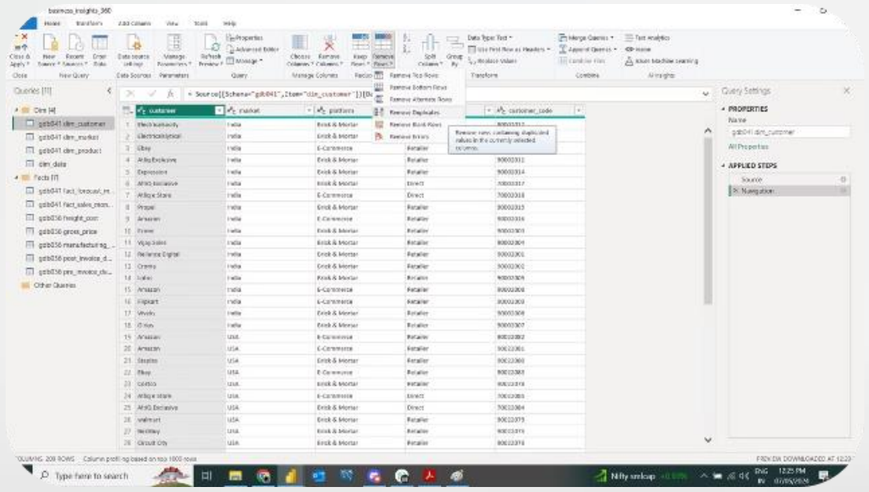
Explanation

Removing unnecessary columns helps streamline your dataset and improve performance.

Steps

Go to the Home tab, select Remove Columns, and choose Remove Columns or Remove Other Columns.





Remove Duplicate Rows

1

Select Columns

Choose the columns you want to check for duplicates.

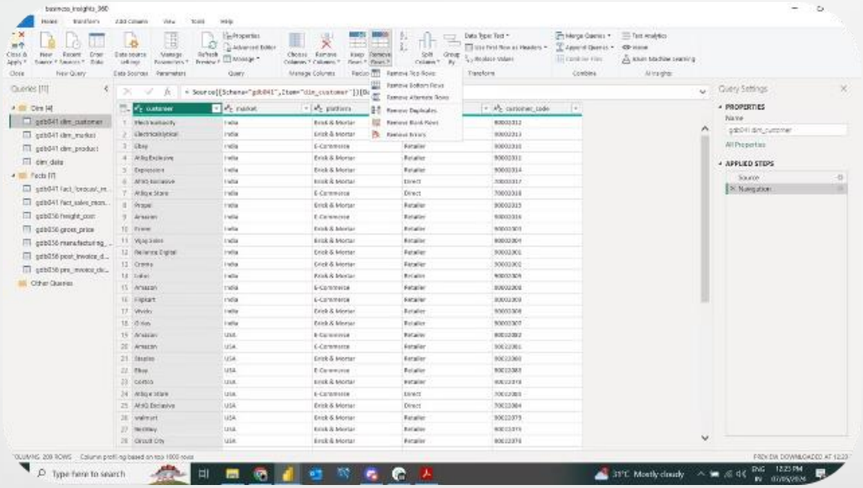
2

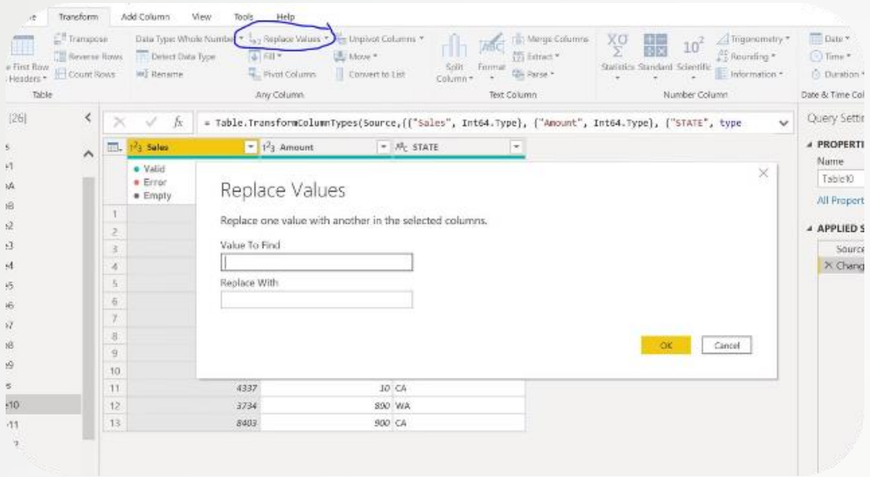
Remove Duplicates

Go to the Home tab, select Remove Rows, and choose Remove Duplicates.

Filter Rows

You can filter rows based on text, number, or date criteria. Go to the Home tab, select Remove Rows, and choose Keep Rows, Keep Top Rows, Keep Range of Rows, or Keep Duplicates. Use Text Filters, Number Filters, or Date Filters to apply specific criteria.





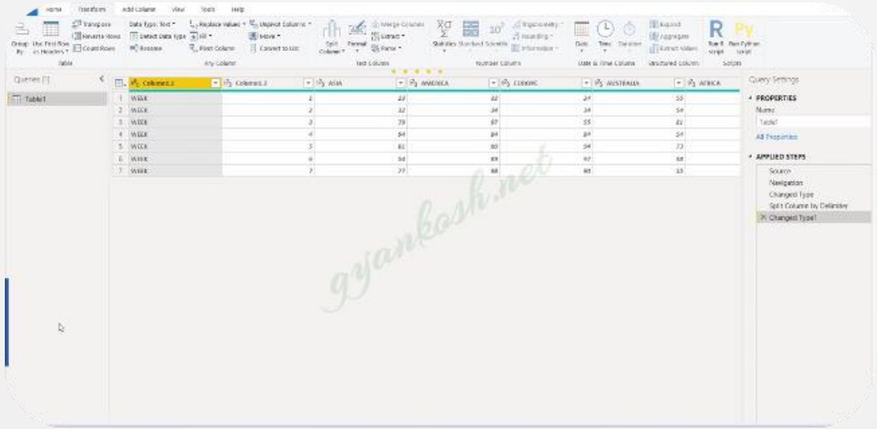
Replace Values

Explanation

Value replacement is useful for correcting inconsistencies or errors in your data.

Replacement Steps

Go to the Home tab, select Replace Values, and enter the value to find and the replacement value.



Split Columns

By Delimiter

Split columns based on a specific delimiter (e.g., comma, semicolon).

By Number of Characters

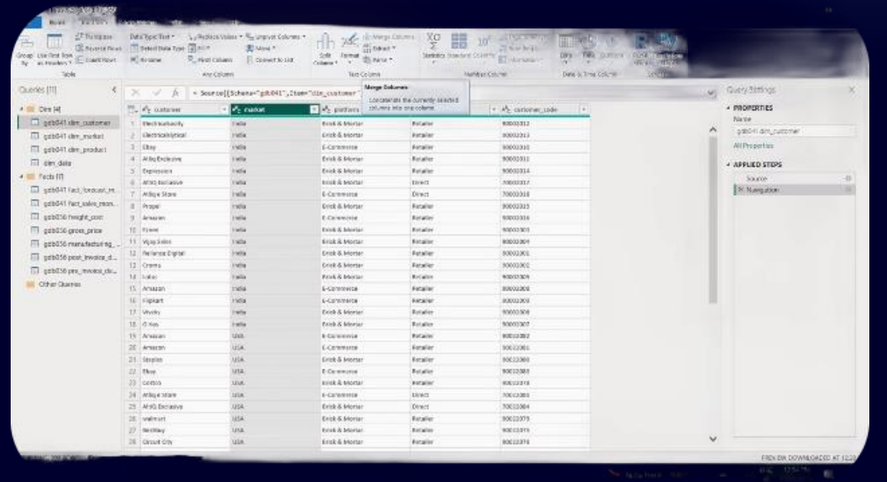
Split columns based on a specified number of characters.

By Positions

Split columns based on character positions within the data.

Merge Columns

Explanation	Merging columns can help consolidate related data into a single column.
Steps	Go to the Transform tab, select Merge Columns, and choose the columns to merge.



Trim and Clean

1

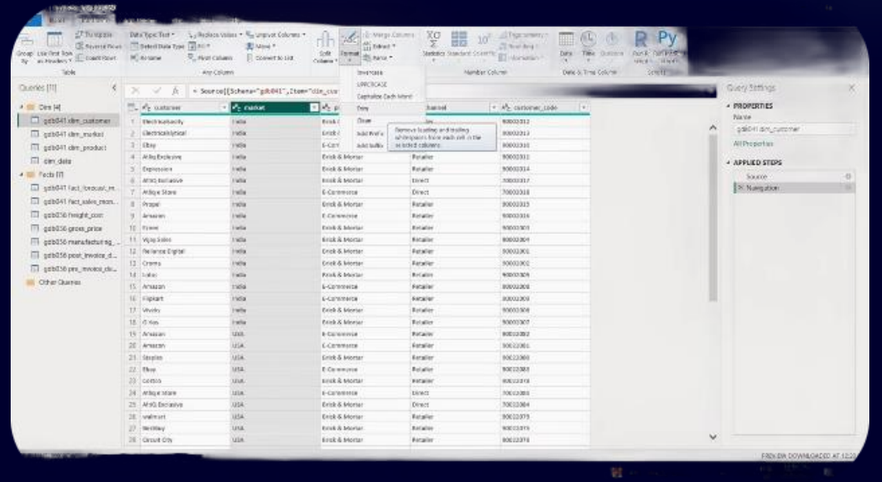
Trim

Remove leading and trailing spaces from your data.

2

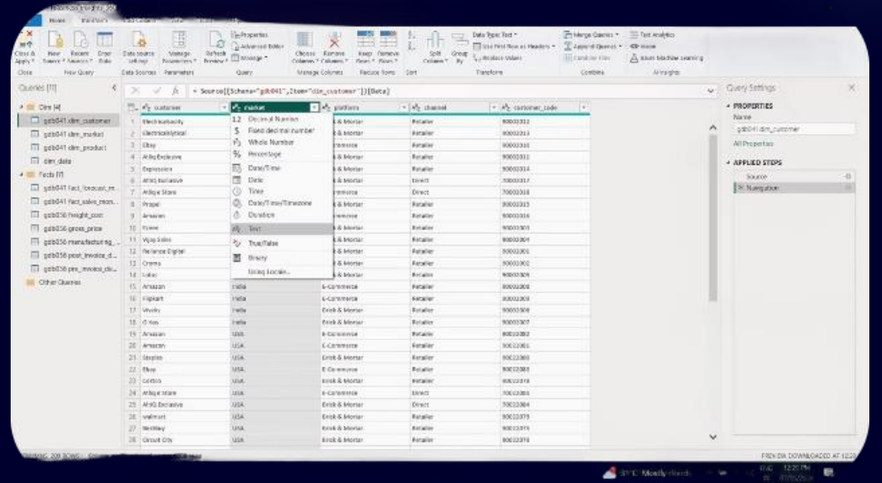
Clean

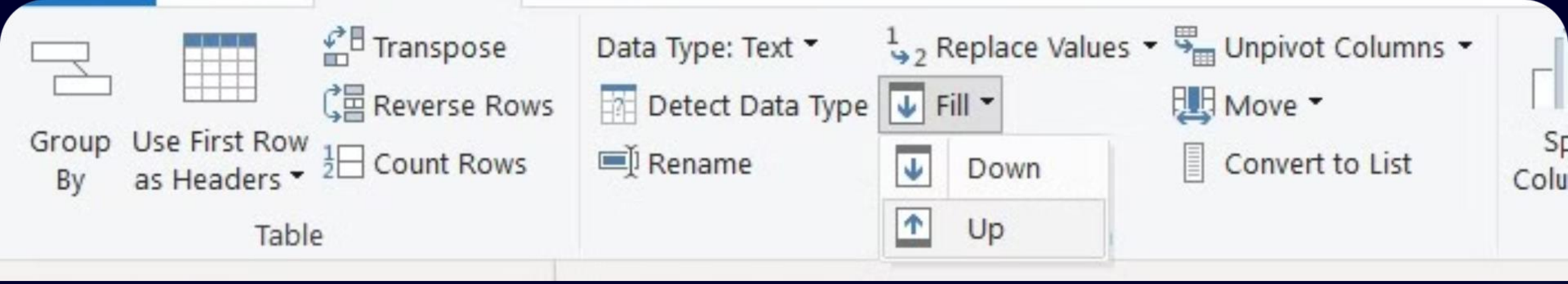
Remove non-printable characters from your data.



Change Data Types

- Explanation of data types (Text, Number, Date).
- Steps to change data types.
- Go to Transform tab > Data Type.
- Choose the appropriate data type.





Fill Down/Up

1. **Explanation of filling missing values**

2. **Steps for fill down**

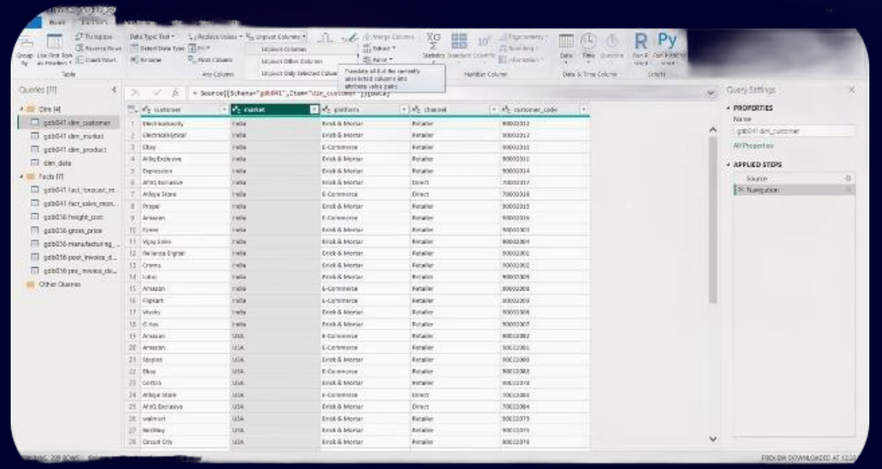
   Go to Transform tab > Fill > Down

3. **Steps for fill up**

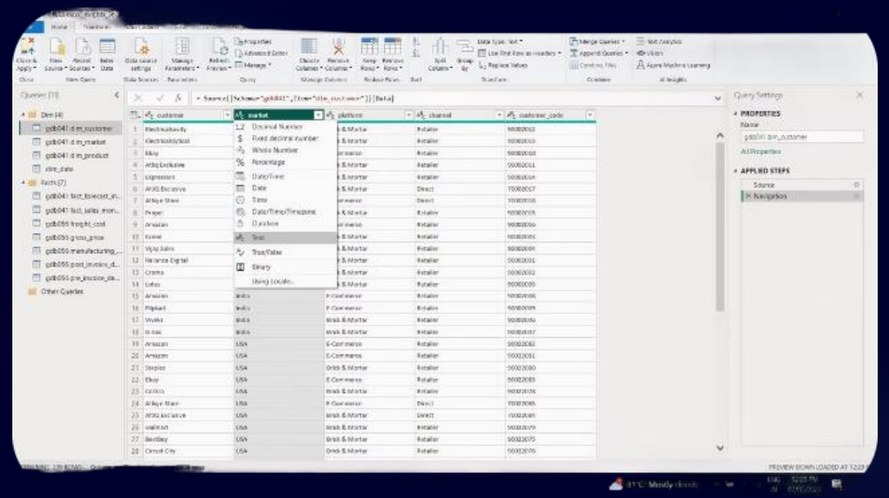
   Go to Transform tab > Fill > Up

Pivot/Unpivot Columns

- Explanation of pivoting and unpivoting.
- Steps for each process.
- For Pivot: Go to Transform tab > Pivot Column.
- Select the column to pivot and the aggregation method.
- For Unpivot: Go to Transform tab > Unpivot Columns or Unpivot Other Columns.



Change Data Types



1 Explanation of data types

Power BI supports various data types such as Text, Number, and Date. Properly setting the data type is crucial for accurate analysis and calculations.

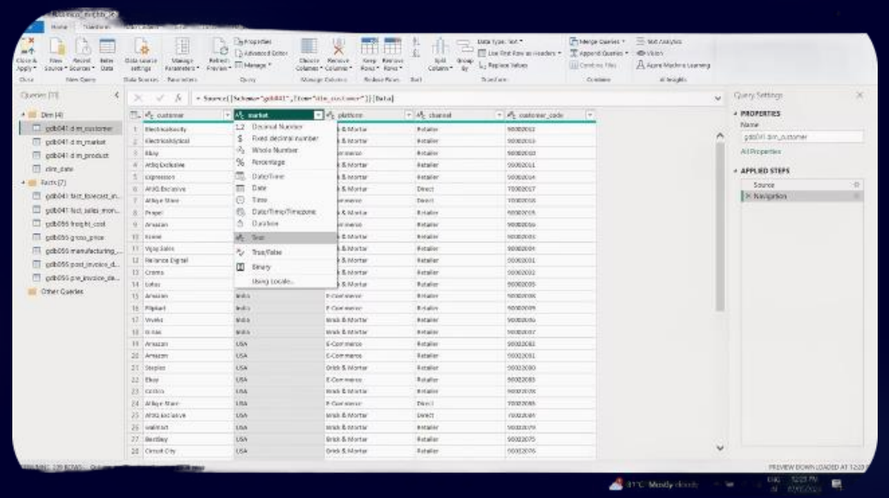
2 Steps to change data types

To change data types, go to the Transform tab and select the Data Type option. Then choose the appropriate data type for each column.

3 Visual example

For example, if a column contains text data that should be treated as numbers, you can change the data type from Text to Number to enable numeric calculations.

Change Data Types



1 Explanation of data types

Power BI supports various data types such as Text, Number, and Date. Properly setting the data type is crucial for accurate analysis and calculations.

2 Steps to change data types

To change data types, go to the Transform tab and select the Data Type option. Then choose the appropriate data type for each column.

3 Visual example

For example, if a column contains text data that should be treated as numbers, you can change the data type from Text to Number to enable numeric calculations.

Pivot/Unpivot Columns

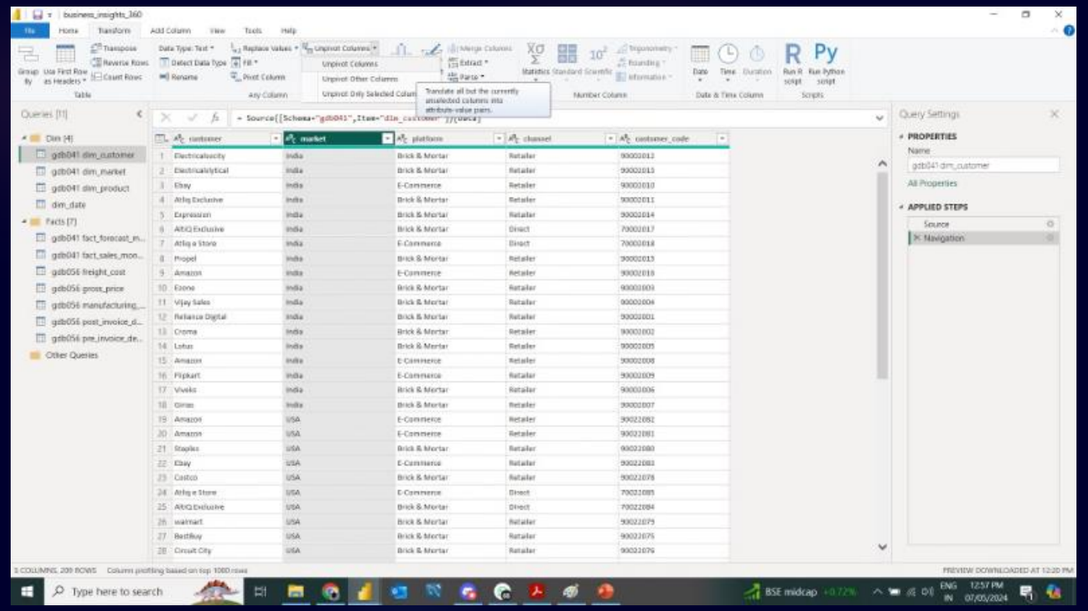
Explanation of pivoting and unpivoting

Steps for each process:

For Pivot: Go to Transform tab > Pivot Column.

Select the column to pivot and the aggregation method.

For Unpivot: Go to Transform tab > Unpivot Columns or Unpivot Other Columns.



Group By

Explanation of grouping data

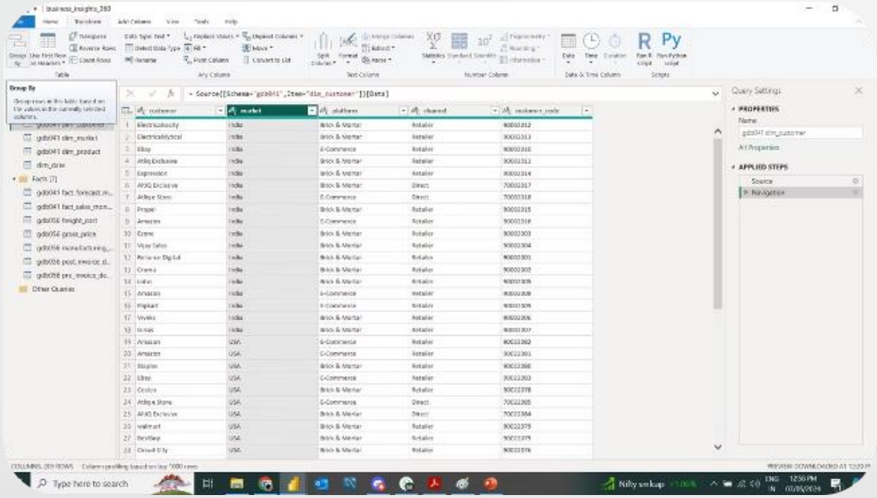
Grouping data is a common data transformation technique that allows you to aggregate and summarize information based on one or more columns.

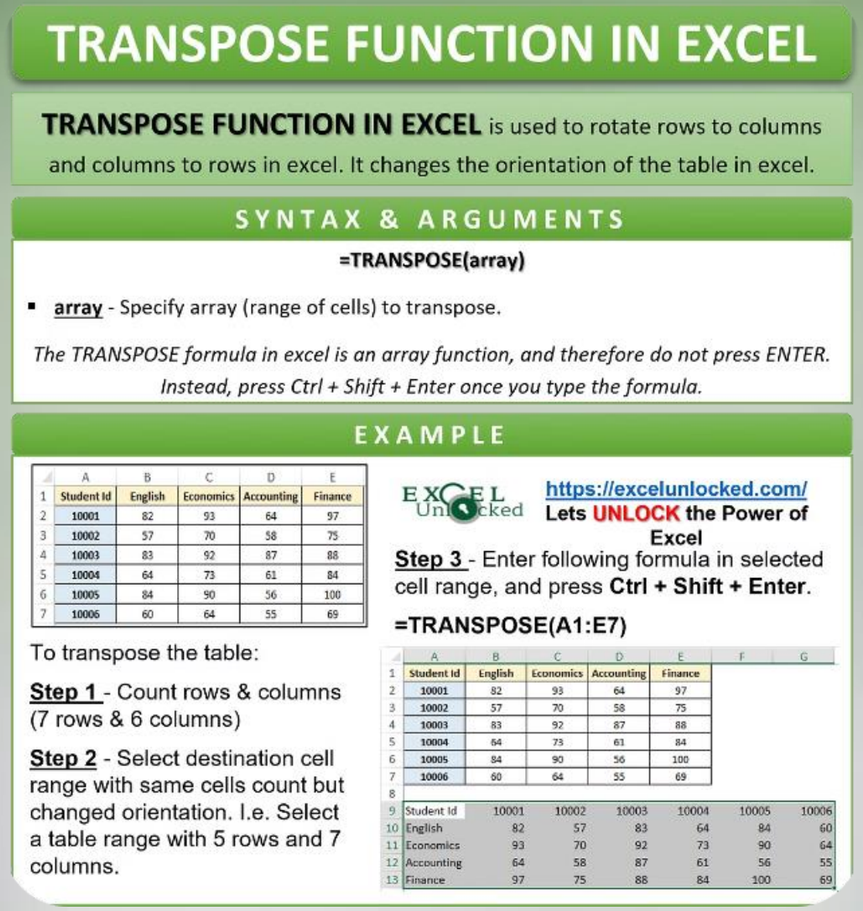
Steps to apply Group By

In Power BI, you can use the Group By feature on the Transform tab. You can choose between Basic and Advanced grouping options, depending on your analysis needs.

Example with screenshot

For instance, you can group sales data by product category and calculate the total sales for each category, providing valuable insights into your business performance.





	A	B	C	D	E
1	Student Id	English	Economics	Accounting	Finance
2	10001	82	93	64	97
3	10002	57	70	58	75
4	10003	83	92	87	88
5	10004	64	73	61	84
6	10005	84	90	56	100
7	10006	60	64	55	69

	A	B	C	D	E	F	G
1	Student Id	English	Economics	Accounting	Finance		
2	10001	82	93	64	97		
3	10002	57	70	58	75		
4	10003	83	92	87	88		
5	10004	64	73	61	84		
6	10005	84	90	56	100		
7	10006	60	64	55	69		
8							
9	Student Id	10001	10002	10003	10004	10005	10006
10	English	82	57	83	64	84	60
11	Economics	93	70	92	73	90	64
12	Accounting	64	58	87	61	56	55
13	Finance	97	75	88	84	100	69

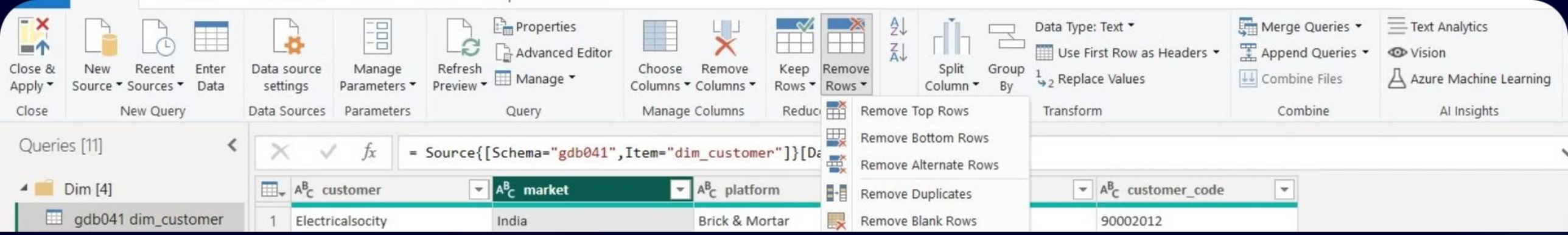
Transpose Table

Explanation of transposing data

Transposing a table in Power BI involves switching the rows and columns, effectively rotating the data by 90 degrees.

Steps to transpose table

To transpose a table, go to the Transform tab and select Transpose. This will convert the rows into columns and the columns into rows.



Handling Missing Values

1

Explanation of handling missing data

Missing data can be a common challenge in data analysis. Power BI offers several methods to handle missing values, including removing, replacing, and filling them.

2

Steps for removing blank rows

To remove blank rows, go to the Home tab and select Remove Rows > Remove Blank Rows.

3

Steps for replacing errors

To replace missing values, go to the Transform tab and select Replace Values. You can choose to replace errors, blanks, or specific values.

4

Steps for filling missing values

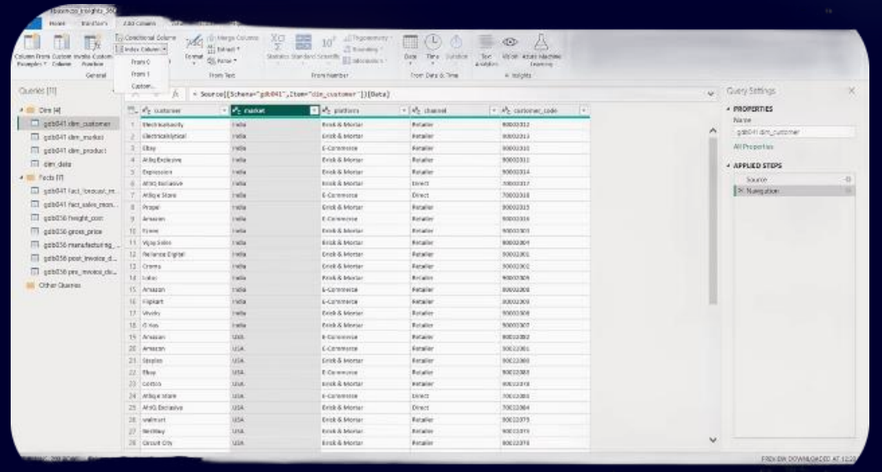
Adding Index Column

Explanation of index columns

Index columns provide a sequential numbering system for rows, which can be useful for tracking and referencing data.

Steps to add an index column

In Power BI, you can add an index column by going to the Add Column tab and selecting Index Column. You can choose to start the index from 0, 1, or a custom value.



Extracting Values

Explanation of extracting values

Power BI's Extract feature allows you to split or extract specific parts of text data, such as the first or last characters, or a range of characters.

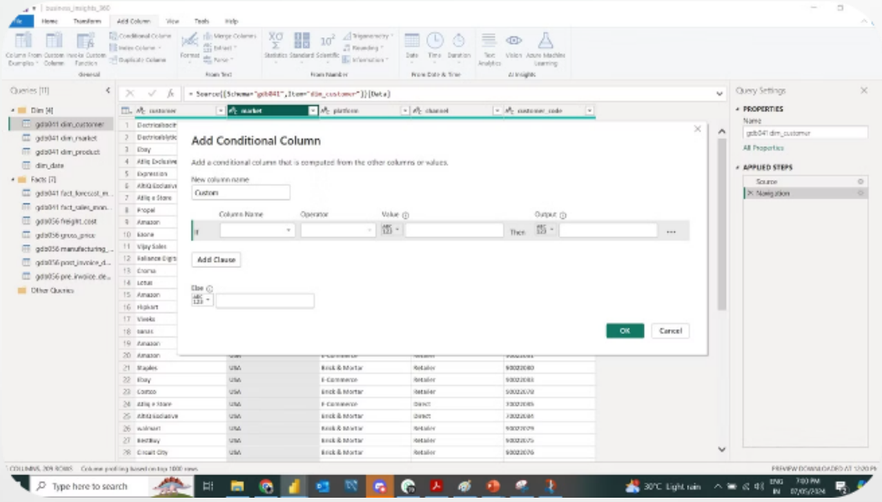
Steps for extraction

To extract values, go to the Add Column tab and select Extract. You can choose options like First Characters, Last Characters, or Range to isolate the desired information.

Example with screenshot

For example, if you have a column with full names, you can extract the first or last name into a new column for easier analysis.

Conditional Columns



1 Explanation of conditional columns

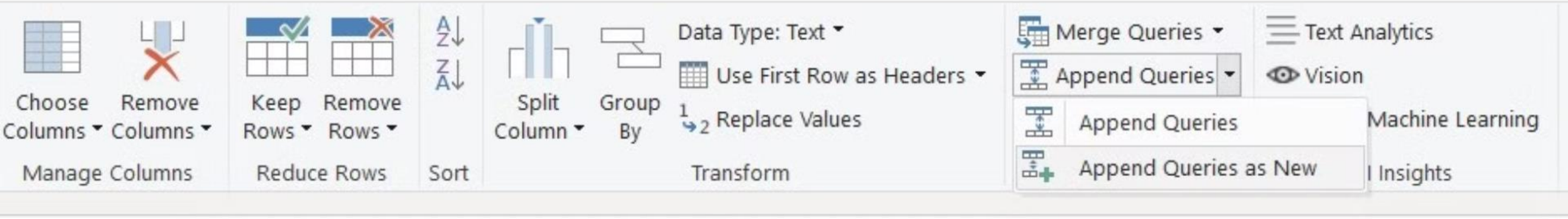
Conditional columns in Power BI allow you to create new columns based on specific conditions or rules. This can be useful for categorizing, flagging, or transforming data.

2 Steps to create conditional columns

To create a conditional column, go to the Add Column tab and select Conditional Column. Define the conditions and the corresponding output values.

3 Example with screenshot

For instance, you can create a column that categorizes sales data as "High", "Medium", or "Low" based on the sales amount, providing valuable insights at a glance.



Appending Queries

Explanation of appending queries

Appending queries in Power BI allows you to combine multiple tables or datasets into a single, unified table. This is useful when you have data from different sources that need to be consolidated.

Steps to append queries

To append queries, go to the Home tab and select Append Queries. You can then choose the tables you want to combine, and Power BI will create a new table with all the data.

Example with screenshot

For example, if you have sales data for different regions stored in separate tables, you can append them to create a comprehensive sales report.