

COURSE GLOSSARY

Data Preparation in Power BI

1-10-100 rule: A data-quality heuristic that suggests it costs \$1 to verify data, \$10 to fix bad data, and \$100 if bad data goes unaddressed in downstream processes, highlighting the exponential cost of poor data quality

Anonymous connection: A type of data source authentication in Power Query that permits access without credentials, typically used for publicly available web content

Applied steps: The ordered list of transformations shown in Power Query that records each change to the data and can be reviewed, edited, or removed to reproduce or modify the data-preparation recipe

Binary data: A data type that stores raw byte-encoded content (often base64) such as images or files rather than readable text or numbers

Clean (transformation): A text transformation that removes non-printable control characters (like line breaks) from text values to prevent formatting issues

Clean data: A dataset that is free from errors, missing values, inconsistent formatting, unwanted duplicates, and incorrect data types, making it reliable for analysis

Column Distribution: A Power Query preview feature that shows a histogram and counts of unique/distinct values under each column to help identify duplicates and value spread

Frame: The explicit start and end boundaries (defined with ROWS BETWEEN or RANGE BETWEEN) that determine which rows are included as input to a window function for the current row

Generator: An iterator created either by a generator expression or a generator function that produces values lazily without constructing the entire sequence in memory

Data connector (New Source / Web connector): A built-in or custom interface in Power Query that allows you to connect to external data sources (files, databases, web pages, APIs) and import their data

Data preparation: The process of cleaning, transforming, and organizing raw data so it is accurate, consistent, and suitable for analysis

Data type: A classification of the kind of data a column holds (for example number, text, date/time, logical), which determines what operations and calculations are valid on that column

Duplicate entries: Repeated records in a dataset that represent the same real-world item and should often be removed to avoid skewed statistics

Logarithmic transformation: A specific numerical transformation that applies a log function to values to convert multiplicative or exponential relationships into additive (more linear) ones for analysis

Merge columns: A transformation that combines two or more columns into a single column using a specified separator, often used to create full names or composite keys

Missing values (nulls): Data cells that contain no value or a null indicator, which can bias analyses or break calculations if not addressed

Numerical transformations: Operations in Power Query (such as add, subtract, multiply, divide, absolute value, power, rounding) applied to numeric columns to correct, scale, or derive new numerical values

Outliers: Observations with values far outside the typical range of a variable, which may indicate data errors or important rare events and should be investigated before deciding how to treat them

Power BI: A Microsoft business analytics application for connecting to data, creating reports, and sharing visual insights

Power Query: The data preparation and transformation engine inside Power BI (and other Microsoft products) that lets you connect to sources, apply sequential transformations, and load cleaned data for analysis

Promote headers: A Power Query transformation that converts the first row of data into the table's column header names, commonly used when headers are included as a data row

Query: A saved sequence of data connection and transformation steps in Power Query that defines how raw data is loaded and shaped for analysis

Split column: A transformation that divides a single column into multiple columns based on a delimiter or fixed position so each column holds one piece of information

Trim (transformation): A text transformation that removes leading and trailing whitespace from values in a text column to standardize entries