

Data Preparation in Excel

AND: A logical function that returns TRUE only if all provided conditions are true, commonly used to create composite tests for filters or flags

CONCAT: An Excel function that joins text from multiple cells or strings into a single text string without automatically inserting delimiters

CSV (Comma-Separated Values): A plain-text file format where each row is a record and fields are separated by commas (or other delimiters), commonly used to exchange tabular data between systems

Data cleaning: The subset of data preparation that identifies and fixes errors, removes duplicates, handles missing values, and corrects inconsistencies to improve dataset quality

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

Data type: A classification of a cell's content such as text, number, date, or boolean that determines how Excel interprets, stores, and processes the value

Excel Table: A structured range formatted as a table in Excel that enables automatic header recognition, structured references, and automatic propagation of formulas and formatting

HLOOKUP: A lookup function that searches for a value in the first row of a horizontal table array and returns a value from a specified row number in the same column, with an option for exact or approximate matches

IF: A conditional function that evaluates a single logical test and returns one value if the test is TRUE and another if the test is FALSE

Import (From Text/CSV): The Excel feature and workflow for bringing external text-based files (such as .csv or .txt) into a workbook, including selecting file origin, delimiter, and data type detection

LEN: An Excel text function that returns the number of characters in a string, including spaces and special characters, useful for spotting trailing spaces or length anomalies

Nested IF: A construction where multiple IF functions are placed inside one another so that additional conditions are tested when previous conditions are FALSE, enabling more than two outcomes

NOT: A logical function that returns the inverse of a single logical value, turning TRUE to FALSE and vice versa

OR: A logical function that returns TRUE if any one or more of the provided conditions are true, useful for inclusive conditional checks

PivotTable: An interactive Excel feature that summarizes, groups, and aggregates large datasets into a dynamic table allowing quick exploration of patterns and metrics

Remove Duplicates: An Excel tool that identifies and deletes duplicate rows based on selected columns to prevent double-counting and maintain unique records

TEXTJOIN: An Excel function that concatenates multiple text values using a specified delimiter and optionally ignores empty cells

TODAY: An Excel date function that returns the serial number (displayed as a date) of the current date and is often used for dynamic age or recency calculations

Unique identifier: A field or column (such as Car ID) that holds values unique to each record, enabling reliable joins, lookups, and de-duplication in data preparation

VLOOKUP: A lookup function that searches for a value in the first column of a vertical table array and returns a value from a specified column number in the same row, with an option for exact or approximate matches

WEEKDAY: An Excel function that converts a date to an integer representing the day of the week, with an optional parameter to set the week's starting day

WORKDAY: An Excel function that returns the date a specified number of working days before or after a start date, with an optional list of holidays to exclude

Worksheet protection: A worksheet-level security setting in Excel that restricts user actions on cells, ranges, or features (such as sorting or filtering) and can be optionally password-protected