

Excel

Data Preparation

Alish Bista - 08/20/2025

Basic Steps

Basic Steps

Doing The Data Laundry

- Collect/Import the raw data
- Explore the data. Sort and filter as you see fit.
- Format the data
- Eliminate duplicates and errors
- Enrich or Summarize if necessary

SubTotal

SubTotal

Checking All Levels

- Automatically groups and calculates data
- Works with sorted data - groups similar items together first
- Calculation options: Sum, Average, Count, Max, Min
- **Examples:**
 - Sales by region
 - Expenses by department

Lookups

Lookups

Tell Me More

- Use relationships between tables to enrich data
- Using a common key, one can go from one table and look for associated records in another
- ***Functions:***
 - VLOOKUP
 - HLOOKUP
 - XLOOKUP

VLOOKUP

The Most Looked Up Lookup Function

- Syntax: *VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])*
- **Arguments:**
 - lookup_value - value to search for
 - table_array - table or dataset to be searched
 - col_index_num - column number with return value
 - range_lookup - FALSE for exact match or TRUE for approximate match

HLOOKUP

The Unsung One

- Syntax: *HLOOKUP(lookup_value, table_array, **row_index_num**, [range_lookup])*
- **Arguments:**
 - **row_index_num** - row number with return value
- Searches data horizontally

XLOOKUP

The Rising Star!

- Syntax: *XLOOKUP(lookup_value, lookup_array, return_array, [if_not_found], [match_mode], [search_mode])*
- **Arguments:**
 - lookup_value - value to search for
 - lookup_array - array or range to search
 - return_array - array or range to return
 - [if_not_found] - Value to return when valid match is not defined

INDEX/MATCH

INDEX/MATCH

The Matchmaking Power Couple

- **INDEX**(table_array, [row_num], [column_num]) - Returns value at specific position
- **MATCH**(lookup_value, lookup_array, [match_type]) - Finds position of value
- **Together:** **INDEX**(table_array, **MATCH**(lookup_value, lookup_array, FALSE), [column_num])

Pivot Tables

Pivot Tables

Pivot. Pivot. Pivot!!!

- Helps in summarizing data
- Aggregates and populates data in dynamic summary tables
- Easily produces basic aggregations like Average, Sum, and Count

Points of Pivot

Getting Ready To Pivot

- Generally, fields in a table have two types of information:
 - **Data:** Contains numerical data to be summarized
 - **Category:** Describe the data
- In a **Pivot Table**,
 - One or more data fields get summarized
 - Category fields appear in Rows, Columns, or Filters

Parts of Pivot

Parts that Pivot Pack

- **Column Labels:** Field in the column orientation (appearing horizontally)
- **Row Labels:** Field in the row orientation (appearing vertically)
- **Values:** Cells that contain summary data
- **Filter:** Field used to limit the data being summarized Category fields
- **Subtotals:** Row or Column that displays subtotals for detail cells
- **Grand Total:** Row or Column that displays total for all detail cells
- **Group:** A collection of items treated as a single entity
- **Item:** Elements in a field that appears as row or column header

Q&A