

## **Data Cleaning and Preprocessing**

Objective: Clean and prepare a raw dataset (with nulls, duplicates, inconsistent formats).

Tools: Python (Pandas)

Identify and handle missing values using `.isnull()` in Python or filters in Excel.

Remove duplicate rows using `.drop_duplicates()` or Excel's "Remove Duplicates".

Standardize text values like gender, country names, etc.

Convert date formats to a consistent type (e.g., dd-mm-yyyy).

Rename column headers to be clean and uniform (e.g., lowercase, no spaces).

Check and fix data types (e.g., age should be int, date as datetime).

### **Dataset names from Kaggle**

Mall Customer Segmentation Data