

✔ Summary of Changes – Task 1: Data Cleaning and Preprocessing

1. Handled Missing Values

- Identified and filtered blank cells in key columns like director, country, and cast
- Replaced missing values with "Unknown" or cleaned using Excel formulas

2. Removed Duplicates

- Applied **Remove Duplicates** using all columns to eliminate redundant rows

3. Standardized Text Columns

- Cleaned type, country, and rating using formulas like PROPER(), UPPER(), and TRIM()
- Converted inconsistent country names into proper case (e.g., "usa" → "Usa")

4. Converted Date Format

- Converted date_added from "September 25, 2021" to 25-09-2021 using formula-based parsing
- Ensured all dates are now in consistent dd-mm-yyyy format

5. Fixed Data Types

- Ensured numeric columns like release_year are stored as **numbers**, not text
- Formatted date columns properly and verified text/numeric consistency using ISNUMBER() and ISTEXT()

6. Structured the File Neatly

- Adjusted column widths and applied consistent alignment
- Added filters, borders, and bold headers
- Removed unnecessary blank rows and columns
- Renamed sheet tab for clarity