

Data Cleaning Explanation (Excel-Only Format)

This document describes how the Netflix Movies and TV Shows dataset can be cleaned using Excel tools only. These steps are general and can be applied to the full dataset without limiting to 1000 rows.

1. Handling Missing Values

- Apply Filters to each column to identify blank cells.
- Use Go To Special → Blanks to highlight all empty cells.
- For text fields like director, cast, country, and rating, fill blanks with “Unknown”.
- For date_added, either leave blanks or fill with “Unknown” depending on requirements.

2. Removing Duplicate Records

- Navigate to the Data tab.
- Select Remove Duplicates.
- Choose all relevant columns to ensure entire duplicate rows are removed.
- Confirm results when Excel reports how many duplicates were removed.

3. Standardizing Text Formats

- Use =PROPER() for names or titles (e.g., “movie title” → “Movie Title”).
- Use =UPPER() for values like rating (e.g., “pg-13” → “PG-13”).
- Use =LOWER() where needed for consistency.
- Use =TRIM() to remove extra spaces.

4. Standardizing Date Formats

- Select the date_added column.
- Use Format Cells → Date.
- Choose a consistent format such as DD-MM-YYYY.
- If dates are text, convert using =DATEVALUE(cell).

5. Renaming Column Headers

- Manually rename headers to clean, consistent labels.

- Example transformations:

Show Id → show_id

Date Added → date_added

Listed In → listed_in

- Use lowercase and underscores for consistency.

6. Correcting Data Types

- Ensure release_year is formatted as Number.
- Ensure date_added is formatted as Date.
- Duration should remain as Text because it contains both numbers and text (e.g., "90 min").
- Rating and type should remain as Text.

7. Final Review and Save

- Re-check filters to ensure no blank or inconsistent values remain.
- Sort fields to ensure formatting is consistent.
- Save the cleaned dataset as an Excel file.

Conclusion

These Excel-only steps ensure the dataset becomes clean, standardized, and ready for analysis or reporting without the use of any programming tools.