

# Netflix Dataset Cleaning Report

## Task 1: Data Cleaning and Preprocessing

### Objective:

Clean and prepare the Netflix dataset by handling nulls, duplicates, and inconsistent formats using Excel.

### Tools Used:

- Microsoft Excel

### Steps Performed:

1. Filled missing values in columns like 'director', 'cast', and 'country' with "Unknown". Filled 'rating' with "NR".
2. Removed rows with missing 'date\_added' as they could not be accurately imputed.
3. Removed all duplicate rows to ensure data accuracy and avoid redundancy.
4. Standardized text values:
  - Converted entries in columns like 'type', 'country', and 'rating' to lowercase.
  - Removed any trailing spaces in text fields.
5. Converted 'date\_added' column to a consistent format (dd-mm-yyyy) using Excel's date functions.
6. Renamed all column headers to a uniform format (lowercase with underscores), e.g., 'release\_year', 'date\_added'.
7. Ensured proper data types:
  - 'release\_year' as integer.
  - 'date\_added' as date format.

### Conclusion:

The dataset is now cleaned, consistent, and ready for further analysis or visualization.