***Task 1: Data Cleaning and Preprocessing – Netflix Dataset***

Dataset Used:

Netflix Movies and TV Shows – downloaded from Kaggle.

Tools Used:

Microsoft Excel

Steps Performed:

1. Removed Duplicates:

Used Excel’s “Remove Duplicates” feature to clean repeated records.

2. Handled Missing Values:

Filled missing values in director, cast, and country with "Unknown".

Forward filled some missing date\_added values using previous row data or left them blank if not available.

3. Standardized Column Headers:

Renamed all headers to lowercase and replaced spaces with underscores.

For example, Date Added → date\_added.

4. Converted Date Format:

Formatted the date\_added column to dd-mm-yyyy format using Excel’s Custom Date Format.

5. Standardized Text Values:

Cleaned type, country, and other text columns using LOWER() and TRIM() functions to ensure consistency.

6. Final Check:

Scanned the data to ensure no nulls, formatting issues, or duplicates remain.

Outcome:

A clean and standardized dataset ready for analysis or visualization.