What I Did – Step-by-Step Summary

◆ **Step 1: Chose the Dataset**

- Selected `netflix_titles.csv` as the dataset for cleaning (contains Netflix shows and movie metadata).

◆ **Step 2: Opened CSV File in Excel**

- Opened the `.csv` file in **Microsoft Excel** using the **Delimited** (Comma) format.

- Verified that the data was correctly split into columns.

◆ **Step 3: Applied Filters**

- Enabled **filters** using `Ctrl + Shift + L` to explore missing and inconsistent values in each column.

◆ **Step 4: Checked for Duplicates**

- Used Excel's **"Remove Duplicates"** tool.

- Result: No duplicate rows were found.

◆ **Step 5: Handled Missing Values**

- Identified missing data in columns like:

  - `director`, `cast`, `country`, `date_added`, and `rating`

- Filled all missing values with the placeholder text `"Unknown"` to maintain consistency.

◆ Step 6: Standardized Text Data

- Cleaned up categorical values in columns like `type`, `country`, and `rating`.

- Converted all text to **lowercase** and removed extra spaces for consistency.

◆ Step 7: Fixed Date Format

- Reformatted the `date_added` column to a consistent **DD-MM-YYYY** format using Excel's **Format Cells** feature.

◆ Step 8: Renamed Column Headers

- Made all column headers **lowercase**, removed spaces, and replaced them with **underscores**.

  - Example: `Release Year` → `release_year`

◆ Step 9: Verified Data Types

- Checked and ensured:

  - `release_year` is stored as **Number**

  - `date_added` is stored as **Date**

◆ Step 10: Saved Cleaned Dataset

- Saved the final cleaned data as:
  ✅ `netflix_titles_cleaned_with_summary.xlsx`

◆ Step 11: Created Cleaning Summary Sheet

- Added a new sheet to the Excel file titled **"Cleaning_Summary"**

- Listed all steps taken to clean and format the data.

◆ Step 12: Created Supporting Files

- Created the following for submission:

    - `netflix_titles_original.xlsx` (original data)

    - `netflix_titles_cleaned.xlsx` (shows what changed)

    - `netflix_cleaning_summary.pdf` (task summary in PDF)

    - `README.md` file for GitHub repo