

DESCRIPTION

This code loads a Netflix dataset using pandas, explores its structure, and performs essential data cleaning tasks. It starts by displaying summary information like column types, missing values, and duplicates. Then, it standardizes the column names by making them lowercase and removing spaces or special characters. The `date_added` column is converted to a consistent dd-mm-yyyy format. The code also focuses on cleaning and standardizing the country column by lowercasing the values, mapping common variations (e.g., "USA", "U.S." → "united states"), and handling entries with multiple countries by splitting, standardizing, deduplicating, and rejoining them. Finally, it prints all unique, cleaned country names found in the dataset.