

Netflix Dataset Insights & Metrics

1. Dataset Overview

The dataset was loaded from `netflix_titles.csv` into a Pandas DataFrame. It includes ~8,800 titles with attributes such as:

- **title**
 - **type** (Movie / TV Show)
 - **director**
 - **cast**
 - **country**
 - **date_added**
 - **release_year**
 - **rating**
 - **duration**
 - **listed_in** (genres)
 - **description**
-

2. Data Cleaning Steps (Pandas)

2.1 Duplicate Removal

- Checked duplicates → found and dropped them.
- Shape of dataset reduced accordingly.

2.2 Missing Value Handling

- Filled missing values in key fields:
 - `director` → *Unknown*
 - `cast` → *Not Available*
 - `country` → *Unknown*
 - `date_added` → *Not Available*
 - `rating` → *Not Rated*
 - `duration` → *Unknown*
- Dropped rows missing critical fields: `title`, `type`.

2.3 Standardization

- Trimmed whitespaces and normalized formatting:
 - `duration` cleaned (stripped, converted to Title Case).
 - `cast` standardized by stripping whitespace.

2.4 DataFrame Consolidation

- Saved the cleaned dataset as `df_clean`.
 - Exported to CSV as `netflix_cleaned.csv` for downstream use.
-

3. Key Insights

Content Distribution (Movies vs. TV Shows)

- Counted how many entries were **Movies** vs. **TV Shows**, showing Netflix's balance of formats.

Top Directors

- Extracted the **Top 10 directors** based on number of titles.

Geographical Spread

- Identified the **Top 10 countries** with the most Netflix content.

Ratings

- Listed the **Top 10 most common ratings**, highlighting Netflix's most frequent audience classifications.
-

4. Potential Applications

- **Content Strategy** → Use director and country-level insights to plan future acquisitions.
- **Genre & Rating Focus** → Explore dominant categories for personalized recommendations.
- **Regional Growth** → Understand high-content countries to strengthen global strategy.
- **Recommendation Systems** → Combine attributes like type, country, and rating to build content filters.