# Netflix Insights

## Data Cleaning Steps (using Pandas)

1. Loaded the raw Netflix dataset into a Pandas DataFrame.

2. Removed duplicate records to ensure uniqueness of shows.

3. Handled missing values by either filling or removing them depending on the column.

4. Dropped unnecessary columns (e.g., 'Unnamed: 0' index column).

5. Standardized date formats in the 'date\_added' column.

6. Converted 'release\_year' to integer type for proper analysis.

7. Ensured 'duration' field consistency (minutes for Movies, seasons for TV Shows).

8. Cleaned and standardized text fields such as 'title', 'director', 'cast', and 'country'.

## Metrics and Insights

• Movies vs TV Shows distribution:

- Movie: 5185

- TV Show: 147

• Top 5 most common ratings:

- TV-MA: 1822

- TV-14: 1214

- R: 778

- PG-13: 470

- TV-PG: 431

• Top 5 countries producing content:

- United States: 1846

- India: 875

- United Kingdom: 183

- Canada: 107

- Spain: 91

• Content release trend (last 5 years):

- 2017: 657

- 2018: 648

- 2019: 519

- 2020: 442

- 2021: 161

• Top 5 most popular genres:

- International Movies: 2369

- Dramas: 2293

- Comedies: 1553

- Action & Adventure: 806

- Independent Movies: 740