# **Netflix Movies and Shows Data Analysis Report**

### **By: Shorya Panwar (23125032)**

This report presents the Exploratory Data Analysis (EDA) conducted on the Netflix Movies and Shows dataset. The objective was to clean the dataset and derive insights about Netflix content, including trends in content type, duration, genres, and regional distributions.

### Dataset Overview

The dataset contains the attributes:

1. Show\_ID (Unique ID given to the show or movie)
2. Type (Movie or TV Show)
3. Title (Name of the show/movie)
4. Director
5. Cast
6. Country (Country of production)
7. Date Added (When was the content added to Netflix)
8. Release Year
9. Rating (Age classification)
10. Duration (Movie length or number of seasons for TV shows)
11. Listed\_in (Genres)
12. Description (Short summary of the content)

### Data Cleaning

We performed the following cleaning steps:

#### Handling Missing Values

Missing values in the 'director' column were replaced with "Not given".

Missing values in the 'cast' column were replaced with "Not given".

Missing values in the 'country' column were replaced with "Not given".

#### Removing Duplicates

No duplicate entries were found in the dataset.

#### Handling Outliers

We found that the 'release\_year' column had values all the way from 1925 to 2021 but the majority of values were from 2000. So, we used the Interquartile Range (IQR) method to detect and remove outliers in 'release\_year' column. We also used box plots and the describe() method before and after this removal to verify our claim.

#### Converting Genres to a List

We found that the genre of a show/movie was a comma separated string. So, we split the string to create a list of strings representing the genres of the show/TV.

### Exploratory Data Analysis (EDA)

We created a variety of visualizations to analyse trends in Netflix content. We used both univariate and bivariate/multivariate analysis.

## Univariate Analysis

#### 1. Movies vs TV Shows

Plots: Bar chart showing the number of Movies vs TV Shows and Pie chart showing the percentage of content that is a Movie or TV show

Insight: The dataset contains significantly more Movies (67.8%) than TV Shows (32.2%) on Netflix.

#### 2. Movie Duration Distribution

Plot: Histogram showing the distribution of movie duration

Insight: Majority of movies are centred around the 100 minute mark.

#### 3. Top 10 Directors

We identified the top 10 directors who have produced the most number of movies and TV shows.

#### 4. Top 5 Countries

Plot: Bar chart showing the top 5 countries with the most Netflix content produced.

Insight: The United States, India, UK, Japan and South Korea are the top 5 countries in Netflix content production.

#### 5. Rating Distribution

Plot: Bar chart showing the frequency of different age ratings.

Insight: Most Netflix content is rated TV-MA and TV-14, indicating a focus on mature and teenage audience.

#### 6. Content Trends over time

Plot: Histogram and KDE plot showing the distribution of content release year.

Insight: Release of Netflix content has been steadily increasing over the years with a fall from 2020 possibly due to the COVID-19 pandemic.

#### 7. Distribution of TV Show Seasons

Plot: Histogram showing the frequency of TV shows having certain number of seasons.

Insight: Majority of TV shows have 1 or 2 seasons.

#### 8. Top 10 Cast Members

Plot: Bar plot showing the top 10 cast members on Netflix TV shows and Movies.

## Bivariate/Multivariate Analysis

#### 1. Movies vs TV Shows over the years

Plots: Line plots showing the number of Movies vs TV Shows released over the years on Netflix.

Insight: The number of both Movies and TV shows rose until 2017. Also, until that time movies dominated TV shows. From 2017 the number of movies started falling as compared to TV shows.

#### 2. Content added by month

Plot: Histogram showing the distribution of content released in each month.

Insight: We tried to find a relation between amount of content released and the release month but we found no significant correlation.

#### 3. Relationship between Release Year and Content Type

Plots: Bar plots showing the number of Movies vs TV Shows released over the years on Netflix.

Insight: The number of both Movies and TV shows rose until 2017. Also, until that time movies dominated TV shows. From 2017 the number of movies started falling as compared to TV shows.

#### 4. Rating vs Type of Content

Plot: Bar chart showing the number of releases in each rating category divided by content type into movies and TV shows.

Insight: The content with rating TV-7 and TV-Y (Content for children) were dominated by TV shows while other rating categories were dominated by Movies.

#### 5. Country vs Type of Content

Plot: Bar chart showing the number of releases in each country divided by content type into movies and TV shows.

Insight: The content in UK, Japan and South Korea were dominated by TV shows while other countries were dominated by Movies. The domination by Movies in India was very high.

#### 6. Genre vs Type of Content

Plot: Bar chart showing the number of releases in each genre divided by content type into movies and TV shows.

Insight: The content in the Kids genre was dominated by TV shows while other genres were dominated by Movies. We also find that Documentaries, Dramas, International Movies and Stand-Up Comedy are the most popular genres.

### Key Findings & Conclusion

From this analysis, we discovered:

* Netflix primarily offers Movies over TV Shows.
* The number of releases has grown upto 2019 and then decreased possibly due to the COVID-19 pandemic.
* TV-MA and TV-14 ratings dominate, indicating a focus on mature and teenage audience.
* The United States, India, and the UK are the top content producers.
* Dramas, Stand-Up Comedy, and Documentaries are the most common genres.
* Movie duration is centred around 100 minutes.