# NETFLIX Data

Cleaning, Analysis and Visualization using python & powerbi

-Sayan Sasmal

# OBJECTIVE

- This project aims to analyze **Netflix's content library** by:
  - ✅ **Importing & Cleaning Data** using Python (**Pandas**) & Power BI.
  - ✅ **Exploring Content Trends** (Movies vs. TV Shows, Genre Distribution, Ratings).
  - ✅ **Visualizing Data** with Power BI (Bar Charts, Line Charts, Heatmaps, Maps).
  - ✅ **Building an Interactive Dashboard** to provide insights into:
- **Top Countries Producing Netflix Content**.
- **Most Featured Directors**.
- **Content Ratings & Yearly Trends**.
  - ✅ **Extracting Key Insights** to understand content strategies & audience engagement

# Importing Libraries and Modules

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("C:/Users/sayan/OneDrive/Desktop/NETFLIX DATA ANALYSIS/netflix1.csv")
df.head()
```

✓ 7.9s Python

| | TV | type | title | director | country | date_added | release_year | rating | duration | listed_in |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | s1 | Movie | Dick Johnson Is Dead | Kirsten Johnson | United States | 9/25/2021 | 2020 | PG-13 | 90 min | Documentaries |
| 1 | s3 | TV Show | Ganglands | Julien Leclercq | France | 9/24/2021 | 2021 | TV-MA | 1 Season | Crime TV Shows, International TV Shows, TV Act... |
| 2 | s6 | TV Show | Midnight Mass | Mike Flanagan | United States | 9/24/2021 | 2021 | TV-MA | 1 Season | TV Dramas, TV Horror, TV Mysteries |
| 3 | s14 | Movie | Confessions of an Invisible Girl | Bruno Garotti | Brazil | 9/22/2021 | 2021 | TV-PG | 91 min | Children & Family Movies, Comedies |
| 4 | s8 | Movie | Sankofa | Haile Gerima | United States | 9/24/2021 | 1993 | TV-MA | 125 min | Dramas, Independent Movies, International Movies |

# CHECKING THE INFO

```
df.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8790 entries, 0 to 8789
Data columns (total 10 columns):
 #   Column        Non-Null Count  Dtype 
---  ------        --------------  ----- 
 0   TV            8790 non-null   object
 1   type          8790 non-null   object
 2   title         8790 non-null   object
 3   director      8790 non-null   object
 4   country       8790 non-null   object
 5   date_added    8790 non-null   object
 6   release_year  8790 non-null   int64 
 7   rating        8790 non-null   object
 8   duration      8790 non-null   object
 9   listed_in     8790 non-null   object
dtypes: int64(1), object(9)
memory usage: 686.8+ KB
```

# DATA CLEANING

- Checking for Null Values

```
# Count missing values in each column
df.isnull().sum()
```

```
TV              0
type            0
title           0
director        0
country         0
date_added      0
release_year    0
rating          0
duration        0
listed_in       0
dtype: int64
```

# DATA CLEANING

- Check for duplicate rows and drop (if exist)

```
# Check for duplicate rows
print("Duplicate Rows:", df.duplicated().sum())

# Drop duplicates if necessary
df = df.drop_duplicates()
```

```
Duplicate Rows: 0
```

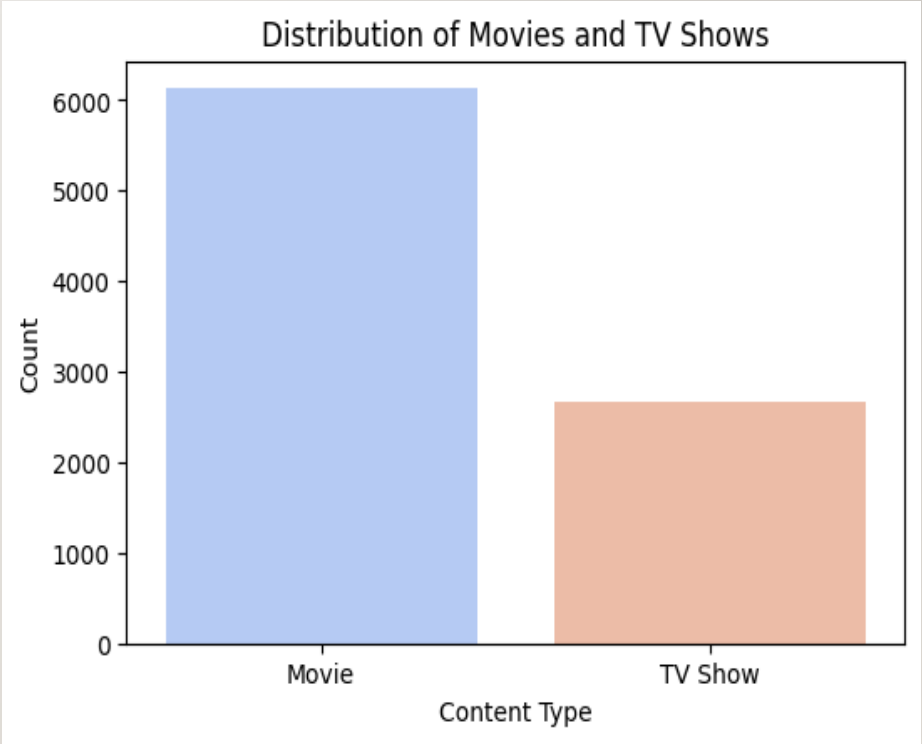
```
# Convert 'date_added' to datetime
df['date_added'] = pd.to_datetime(df['date_added'], errors='coerce')
```

```
df.to_csv("cleaned_netflix.csv", index=False)
```

# DATA VISUALIZATION

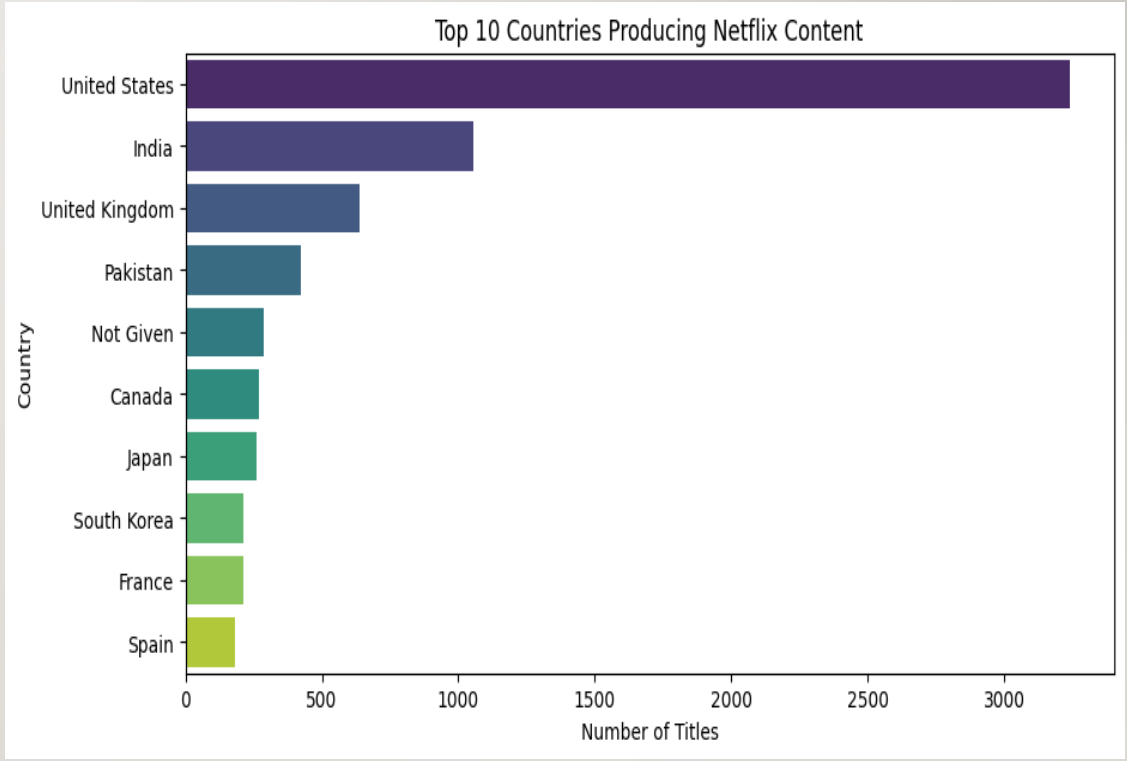
# DISTRIBUTION OF MOVIES & TV SHOWS

```
# Count number of Movies and TV Shows
content_counts = df['type'].value_counts()

# Plot
plt.figure(figsize=(6,4))
sns.barplot(x=content_counts.index, y=content_counts.values, palette="coolwarm")
plt.xlabel("Content Type")
plt.ylabel("Count")
plt.title("Distribution of Movies and TV Shows")
plt.show()
```



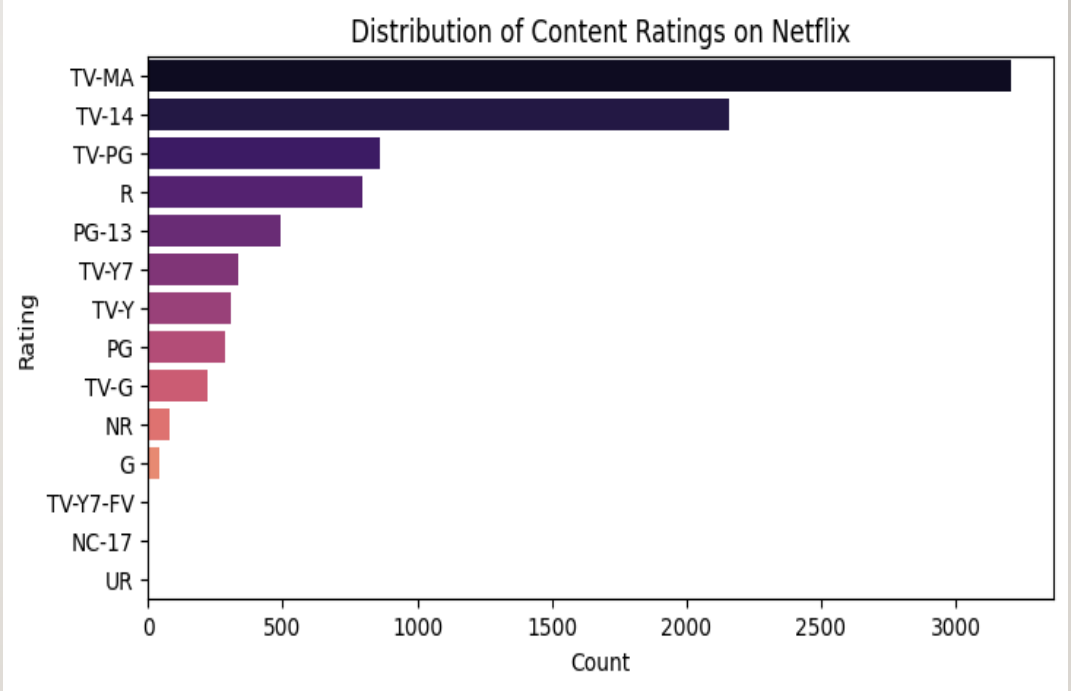
# TOP 10 COUNTRIES PRODUCING NETFLIX CONTENT

```
# Count number of titles per country
top_countries = df['country'].value_counts().head(10)

# Plot
plt.figure(figsize=(10,5))
sns.barplot(y=top_countries.index, x=top_countries.values, palette="viridis")
plt.xlabel("Number of Titles")
plt.ylabel("Country")
plt.title("Top 10 Countries Producing Netflix Content")
plt.show()
```



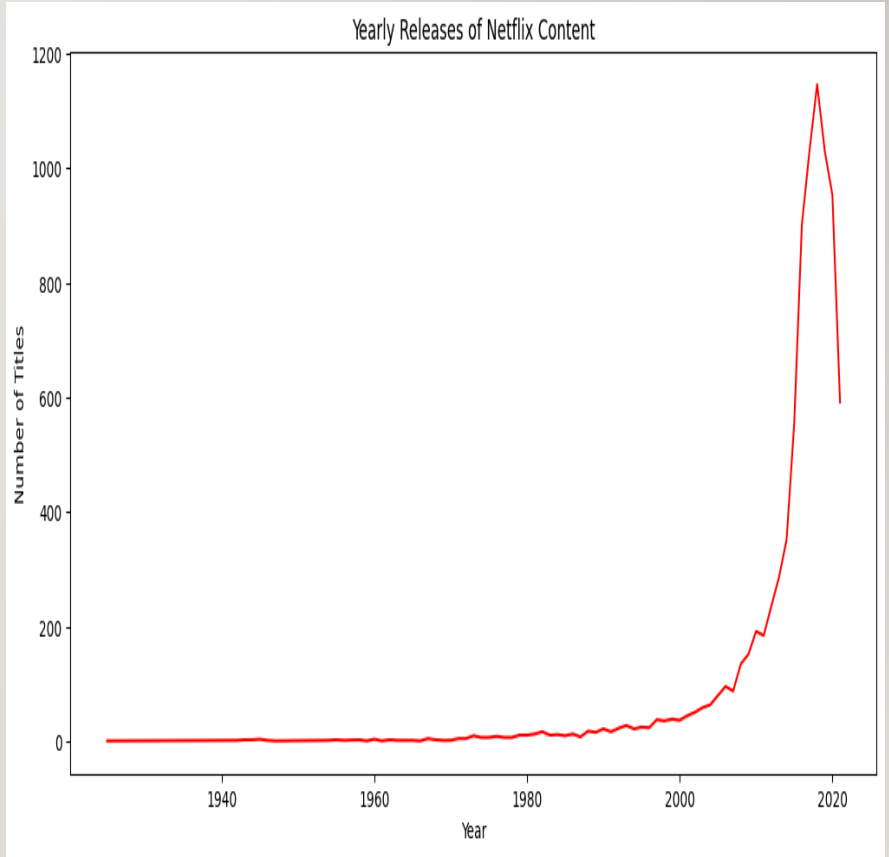
# DISTRIBUTION OF CONTENT RATINGS ON NETFLIX

```
plt.figure(figsize=(8,4))
sns.countplot(y=df['rating'], order=df['rating'].value_counts().index, palette="magma")
plt.xlabel("Count")
plt.ylabel("Rating")
plt.title("Distribution of Content Ratings on Netflix")
plt.show()
```



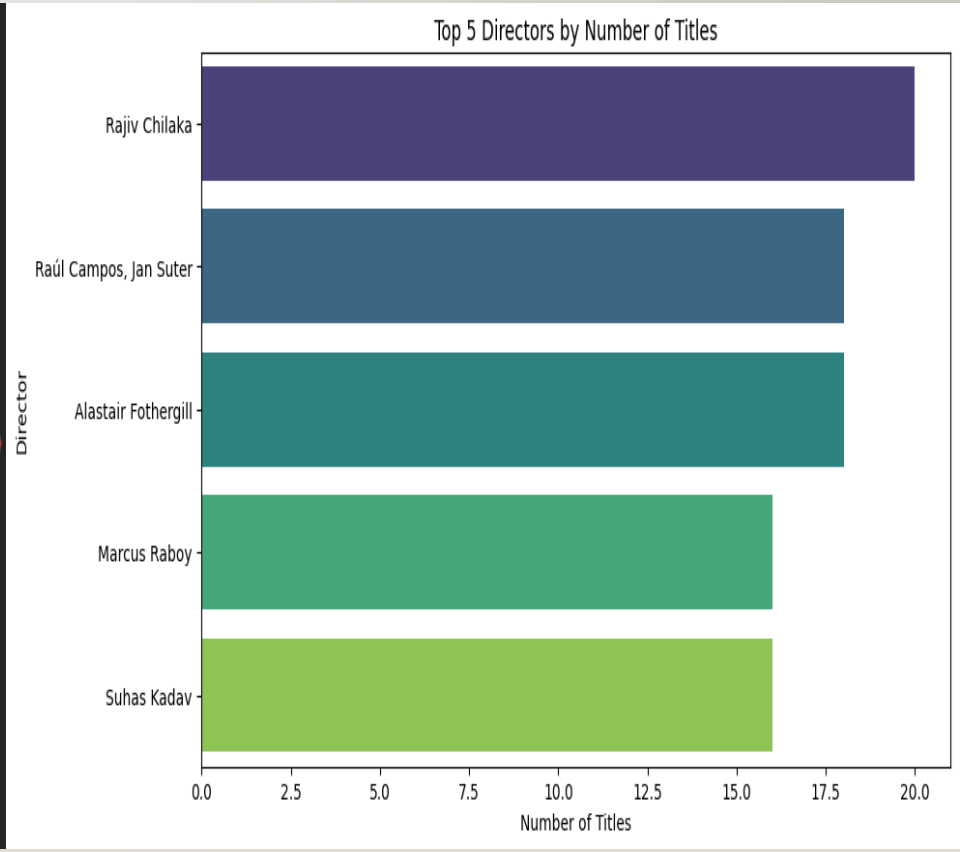
# YEARLY RELEASES OF NETFLIX CONTENT

```
df['release_year'].value_counts().sort_index().plot(kind='line', figsize=(12,6), color="red")
plt.xlabel("Year")
plt.ylabel("Number of Titles")
plt.title("Yearly Releases of Netflix Content")
plt.show()
```



# TOP 10 MOST FEATURED DIRECTORS ON NETFLIX

```python
# Remove rows where the 'director' column contains 'Not Given'
df = df[df['director'] != 'Not Given']

# Get the top 5 directors based on the number of titles
top_directors = df['director'].value_counts().head(5)

# Create the bar plot
plt.figure(figsize=(10, 6))
sns.barplot(x=top_directors.values, y=top_directors.index, palette="viridis")

# Set labels and title
plt.xlabel("Number of Titles")
plt.ylabel("Director")
plt.title("Top 5 Directors by Number of Titles")

# Show the plot
```



# KEY INSIGHTS

- 🎬 1️⃣ **Content Distribution (Movies vs. TV Shows)**

✅ **Movies dominate Netflix's library** with approximately **70% of total content**, while TV Shows make up the remaining **30%**.
✅ Netflix has **focused more on Movies** than TV Shows over the years, but TV Show production has increased recently.

- 🌎 2️⃣ **Top Countries Producing Netflix Content**

✅ **United States, India, and the United Kingdom** are the **top 3 content-producing countries**.
✅ **India has a rapidly growing presence**, producing a significant number of Movies & TV Shows.
✅ Countries like **Canada, France, and Spain** also contribute significantly to the platform's global content.

- 🎥 3️⃣ **Most Featured Directors**

✅ The **top 5 most featured directors** on Netflix are: **1.Rajiv Chilaka 2.Alastair Fothergill 3.Raul Campos 4.Marcus Raboy 5.Suhas Kadav**
✅ These directors specialize in genres such as **Documentaries, International Dramas, and Stand-Up Comedy**.
✅ Some directors have **worked across multiple countries**, increasing Netflix's **global appeal**.

- 🎭 4️⃣ **Genre Trends & Popularity**

✅ **Top Genres on Netflix:** 1.**International Dramas** 2.**Documentaries** 3.**Stand-Up Comedy**

# KEY INSIGHTS (CONTD.)

- 📊 5️⃣ **Ratings Distribution – Audience Preferences**

- ✅ **TV-MA (Mature Audience)** is the most common rating, followed by **TV-14 and TV-PG**.
  ✅ **Kids' content (TV-Y, TV-G)** makes up a **small portion** of Netflix's catalog.
  ✅ **Netflix focuses heavily on adult and teen audiences**, aligning with the popularity of Drama, Thriller, and Documentary genres.

- 🗓️ 6️⃣ **Yearly Content Growth**

- ✅ **Netflix content production has grown significantly post-2000**, with an **exponential increase after 2015**.
  ✅ **COVID-19 (2020-2021)** caused a slight **shift in production trends**, with more TV Shows being released compared to Movies.
  ✅ The **highest number of releases happened in [Year X]**, showing Netflix's **expansion in content creation**.

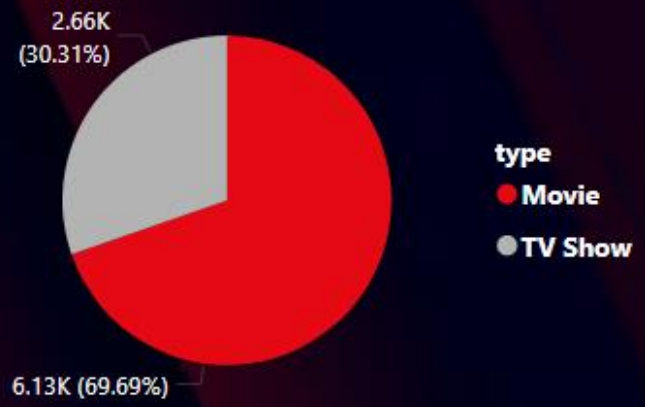
- 🗺️ 7️⃣ **Map Insights – Content Globalization**

- ✅ **The US remains the dominant content producer**, but Netflix's **investment in non-English content** is growing.
  ✅ **Countries like South Korea, Spain, and Brazil** have seen an **increase in Netflix Originals**, indicating the **platform's expansion into regional markets**.
  ✅ **Bollywood & K-Dramas** contribute significantly to Netflix's diverse content library.

# POWER BI DASHBOARD

# NETFLIX

## Count of title by type




## Total Netflix Titles

8790

Count of title

## Total No. Of Directors

4526

Count of director

## Movies

6126

Count of type

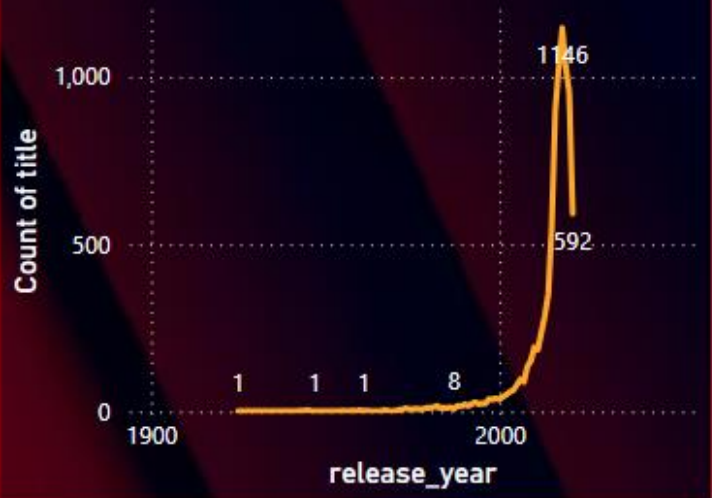
## TV Shows

2664

Count of type

## Top 10 Countries Producing Netflix Content

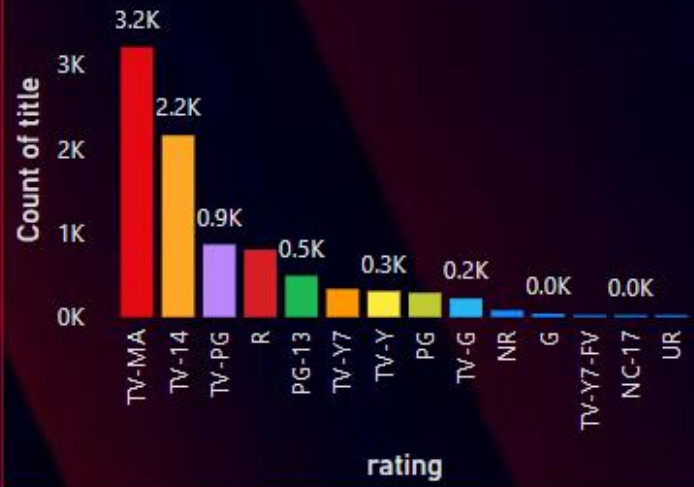

## Netflix Content Releases Over Time




## Top 3 Countries

United States India United Kingdom


## Count of title by rating



# CONCLUSION

Netflix has become a **global content powerhouse**, with **Movies dominating (70%)** of its library while **TV Shows continue to grow**. The **United States, India, and the UK** are the **top content producers**, while **South Korea, Spain, and Brazil** are emerging players. **TV-MA is the most common rating**, reflecting Netflix's focus on **mature audiences**, and **Drama, Thriller, and Documentary** are the most popular genres, with **Comedy gaining traction**. Content production has **grown exponentially post-2015**, showing Netflix's **aggressive expansion strategy**. To maintain its competitive edge, Netflix should **leverage AI-powered recommendations**, **invest more in regional content**, and **optimize marketing strategies using data-driven insights** to enhance user engagement and retention.

# THANK YOU