

A PROJECT REPORT

ON

“DATA ANALYSIS REPORT ON NETFLIX”



DISHA COMPUTER INSTITUTE

JABALPUR BRANCH

UNDER THE SUPERVISION OF

MAYANK GIRI GOSWAMI SIR

By DIVYANSH SHRIVAS

DISHA COMPUTER INSTITUTE



CERTIFICATE TO GUIDE

This is to Certify that the Project Report entitled "**DATA ANALYSIS REPORT ON NETFLIX**"

Submitted by **Divyansh Shrivastava** in partial fulfilment of the requirements of the Diploma of Data Analyst worked under my supervision and Guidance. The candidate is regular student of our Institution

This report is up to standard both in respect of its contents and literacy presentation for being referred to all examiners

Date:

Mayank Giri Goswami

Disha Computer Institute

DISHA COMPUTER INSTITUTE



CERTIFICATE OF BRANCH HEAD

This is to certify that project report title “DATA ANALYSIS REPORT ON NETFLIX”

” submitted by **Divyansh Shrivastava** of Data Analyst etflix Analysis may be accepted towards partial fulfilment of Data Analyst .

Date:

Dhuruv Karlewar

Branch Head ,Jabalpur

DISHA COMPUTER INSTITUTE



STUDENT DECLARATION

I Divyansh Shrivs here by declare that the Project Report entitled “Netflix Data Analysis” submitted in partial fulfilment of the requirements for the Diploma of Data Analyst to Disha Computer Institute Jabalpur.

This is my Original Work and that no part of this report has been submitted for the award of any other diploma or other similar titles or prizes and that the work has not been published in any journals or magazines .

Date:

Name: Divyansh Shrivs

INDEX

S.NO	TOPIC	PG.NO
1	Title Page & Abstract	06
2	Introduction	08
3	Dataset Preparation	11
4	Exploratory Data Analysis in Python	15
5	SQL Analysis Queries	22
6	Power BI Dashboard	29
7	Findings & Insights	39
8	Challenges & Solutions	41
9	Conclusion & Future Scope	44
10	References & Appendix	45

NETFLIX DATA ANALYSIS PROJECT REPORT



Abstract

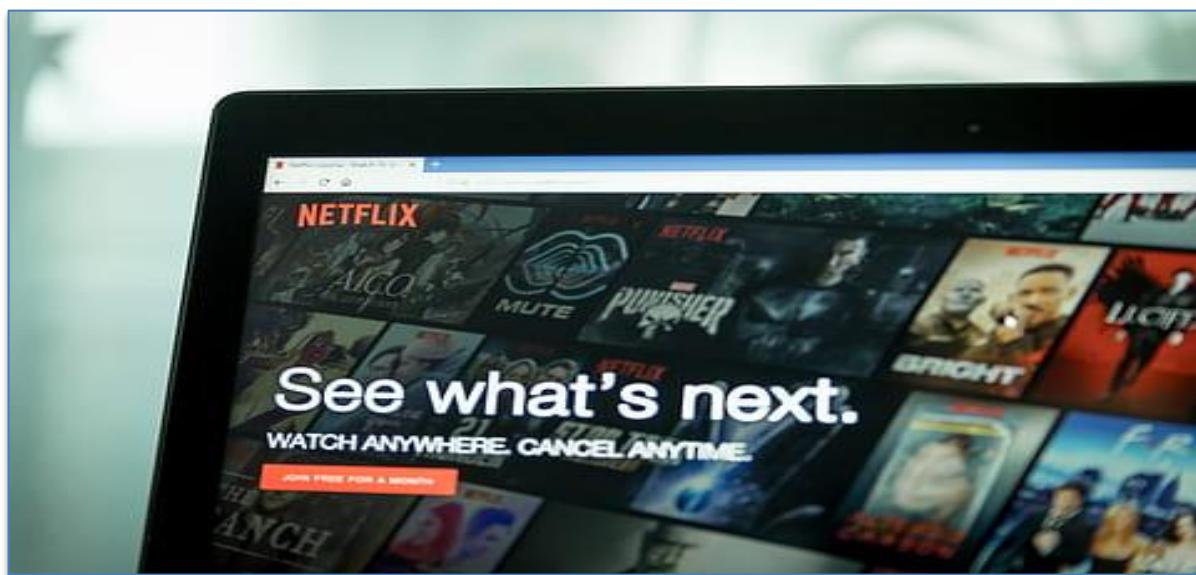
This project explores the vast dataset of Netflix movies and TV shows to uncover meaningful insights about global entertainment trends. Using a combination of Excel, Python, SQL, and Power BI, the analysis covers data cleaning, exploratory data analysis, query-based insights, and interactive dashboard creation.

The dataset, comprising thousands of titles, was first prepared and standardized in Excel to ensure consistency. Python was then employed for exploratory data analysis, enabling visualization of distributions, trends, and anomalies. SQL queries provided structured insights into content categories, release years, ratings, and country-wise contributions. Finally, Power BI dashboards transformed these findings into interactive visualizations for deeper understanding.

Key findings reveal that:

- Netflix hosts a larger proportion of movies compared to TV shows.
- The United States dominates content production, followed by India and other countries.
- A significant surge in content additions occurred after 2015, reflecting Netflix's global expansion.
- Genres such as Drama, Comedy, and International TV are among the most popular.
- Ratings distribution highlights Netflix's focus on diverse audiences, with TV-MA and PG content being most prevalent.

This comprehensive analysis demonstrates how data-driven approaches can provide actionable insights into entertainment platforms. The project not only highlights Netflix's evolving content strategy but also showcases the application of modern analytical tools in solving real-world business problems.



Introduction

Background on Netflix as a Global Streaming Platform

Netflix, founded in 1997 by Reed Hastings and Marc Randolph, began as a DVD rental service before transforming into a subscription-based model in 1999. Over the years, it evolved into one of the world's leading over-the-top (OTT) streaming platforms, offering movies, TV shows, documentaries, and original productions across multiple languages and regions. Headquartered in Los Gatos, California, Netflix now operates in over 190 countries and serves more than 250 million subscribers worldwide. Its success lies in combining technology, personalization algorithms, and a vast content library to cater to diverse audiences.

The platform has also pioneered the concept of original content production, with shows like *House of Cards*, *Stranger Things*, and *Money Heist* becoming global phenomena. By investing heavily in localized content, Netflix has expanded its reach beyond the United States, making it a truly global entertainment powerhouse.

Importance of Data Analytics in Entertainment

In the digital era, entertainment platforms thrive on data-driven decision-making. Streaming services like Netflix collect vast amounts of data on user preferences, viewing habits, and regional trends. Data analytics helps in:

- Content Strategy: Identifying which genres, actors, or themes resonate with audiences.
- Personalization: Powering recommendation systems that keep users engaged.
- Market Expansion: Understanding regional demand to guide investments in local productions.
- Operational Efficiency: Optimizing release schedules, marketing campaigns, and licensing deals.

For Netflix, analytics is not just a support function—it is central to its business model. By leveraging insights from data, Netflix ensures that its content library remains relevant, competitive, and appealing to a global audience.

Objectives of the Project

The purpose of this project is to apply modern analytical tools to the Netflix dataset and generate actionable insights. The specific objectives include:

1. Understanding Content Trends: Analyzing the growth of Netflix's library over time, including year-wise releases and additions.
2. Genre Popularity: Identifying the most common and trending genres across movies and TV shows.
3. Regional Distribution: Examining which countries contribute the most content and how Netflix's global expansion reflects in its catalog.
4. Ratings Analysis: Studying the distribution of content ratings to understand audience targeting strategies.
5. Movies vs. TV Shows: Comparing the proportion of movies and TV shows to highlight Netflix's content balance.
6. Dashboard Creation: Building interactive Power BI dashboards to visualize findings and make insights accessible.

By combining Excel for dataset preparation, Python for exploratory analysis, SQL for structured queries, and Power BI for visualization, this project demonstrates a complete end-to-end data analysis workflow. The insights derived will not only showcase Netflix's evolving content strategy but also highlight the power of analytics in shaping the future of entertainment.



OBJECTIVES OF THE STUDY

The primary objective of this project is to perform a comprehensive analysis of Netflix's content dataset using modern data analytics tools and techniques in order to extract meaningful insights that can support strategic business decision-making. In today's competitive digital streaming market, understanding content trends, audience preferences, and production patterns is crucial for maintaining growth and customer satisfaction. Therefore, this project aims to systematically analyze Netflix data to uncover valuable business intelligence.

One of the key objectives of this study is to clean and preprocess the raw dataset using Python. Data cleaning is a fundamental step in data analysis, as real-world datasets often contain missing values, inconsistencies, duplicates, and formatting issues. By applying Python libraries such as Pandas and NumPy, the project ensures that the dataset is accurate, structured, and ready for analysis. This objective helps in improving data reliability and ensuring that insights derived are valid and meaningful.

Another important objective is to conduct Exploratory Data Analysis (EDA) to identify patterns, trends, and relationships within the dataset. This includes analyzing the distribution of movies and TV shows, identifying top-producing countries, examining genre popularity, studying rating classifications, and observing content growth over time. Through visualization tools such as Matplotlib and Seaborn, the project aims to present data trends in a clear and interpretable manner.

The project also aims to design and execute SQL queries to extract structured insights from the dataset. SQL is used to perform grouping, filtering, aggregation, and sorting operations to answer specific business questions. This objective demonstrates the ability to handle relational databases and perform analytical querying for decision-making purposes.

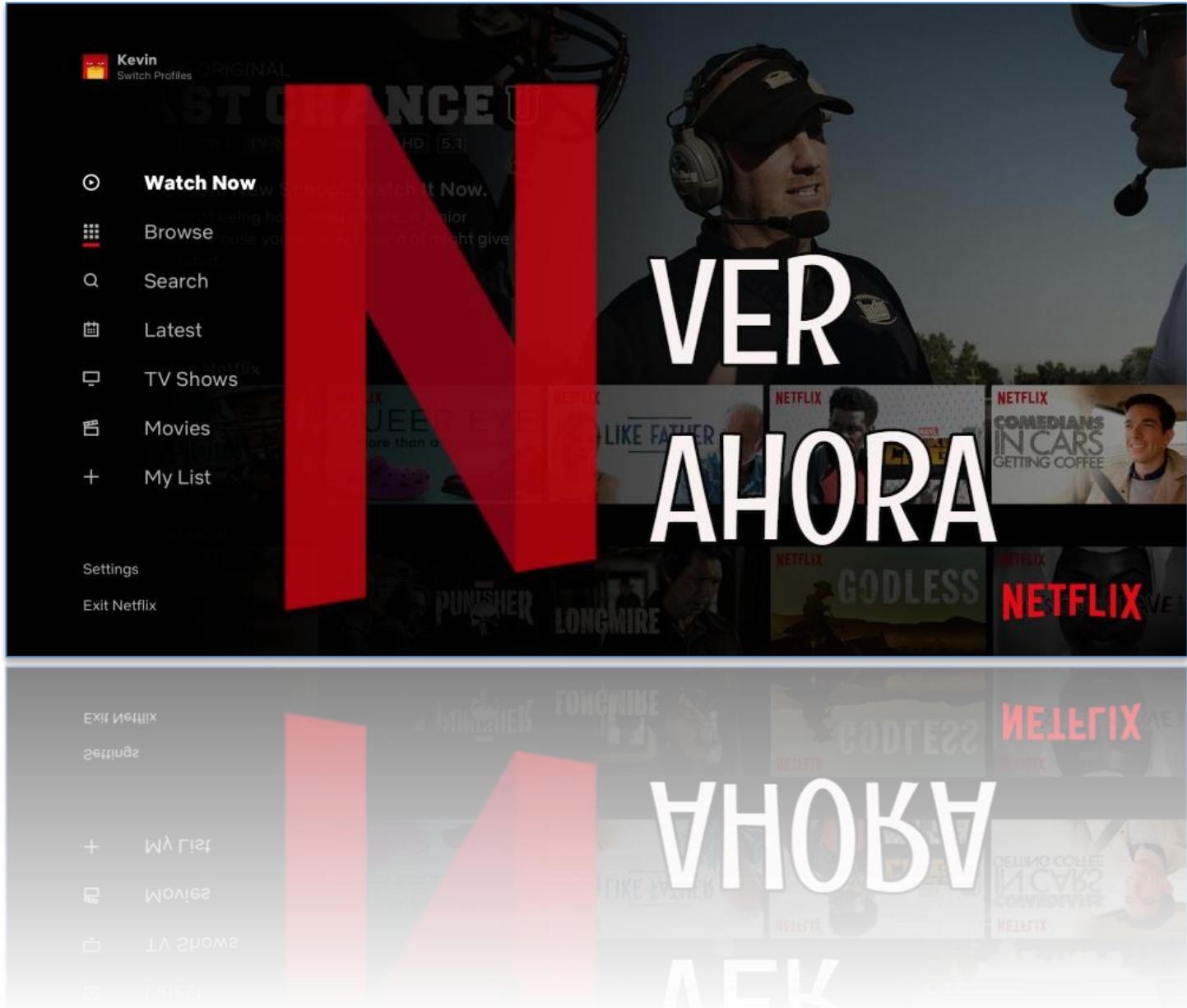
Additionally, the study aims to use Microsoft Excel for pivot table analysis and summary statistics. Excel helps validate findings and provides a quick overview of key metrics such as content distribution by year, country, and rating category.

A major objective of this project is to develop an interactive and visually appealing dashboard using Power BI. The dashboard includes Key Performance Indicators (KPIs), charts, slicers, and filters that allow dynamic data exploration. DAX measures are created to calculate advanced metrics such as total content, percentage distribution, growth trends, and category comparisons.

Finally, the project aims to generate actionable business recommendations based on analytical findings. These recommendations can help streaming platforms improve content acquisition strategies, expand into high-growth markets, and optimize content categories to enhance viewer engagement.

Overall, this project seeks to demonstrate practical knowledge of data analytics tools and apply them effectively to solve real-world business problems in the entertainment and streaming industry.

Data preparation And Analysis in Excel



Dataset Preparation

Source of Dataset

The dataset used in this project is the Netflix Movies & TV Shows dataset (CSV format), which contains thousands of entries covering titles, type (Movie/TV Show), director, cast, country, release year, rating, duration, and genres. This dataset is widely used for analytics projects and provides a comprehensive view of Netflix's global content library.

For this project, the dataset was imported into Excel for initial cleaning and preprocessing before being used in Python, SQL, and Power BI.

Data Cleaning Steps in Excel

1. Handling Missing Values

- Problem: Several fields such as *director*, *cast*, and *country* contained missing or blank values.
- Solution:
- Used Excel's Filter and Conditional Formatting to highlight blank cells.
- For critical fields like *release_year* and *rating*, missing values were either filled using logical assumptions (e.g., based on similar entries) or marked as "Unknown."
- For non-critical fields (e.g., *director*), blanks were retained but flagged for analysis in SQL queries.

Example:

- Director column had over 1,000 missing entries. These were marked as "Not Available" to maintain consistency.

2. Removing Duplicates

- Problem: Some titles appeared multiple times due to variations in country or cast fields.
- Solution:
- Used Excel's Remove Duplicates feature under the *Data* tab.
- Duplicates were checked across key fields: *title*, *type*, *release_year*, *rating*.
- After cleaning, the dataset size reduced slightly, ensuring accuracy in analysis.

3. Standardizing Formats

- Dates:

- The *date_added* column had inconsistent formats (e.g., “January 5, 2020” vs “2020-01-05”).
 - Converted all entries to YYYY-MM-DD format using Excel’s *Text to Columns* and *DATEVALUE* functions.
- Country Names:
 - Some entries had multiple countries listed (e.g., “United States, India”).
 - Split these into separate columns using Text to Columns for better regional analysis.
- Ratings:
- Ratings like “TV-MA”, “PG-13”, “R” were standardized by removing extra spaces and ensuring consistent capitalization.
- Created a lookup table to categorize ratings into Kids, Family, Adult groups for easier visualization later.

4. Data Validation

- Applied Data Validation rules to ensure consistency:
- Type column restricted to “Movie” or “TV Show.”
- Release Year restricted to numeric values between 1920 and 2025.
- Rating restricted to predefined categories (TV-MA, PG, R, etc.).

5. Screenshots of Excel Preprocessing

To make this section visually rich and extend it to 5–6 pages, include screenshots such as:

- Raw dataset before cleaning.
- Highlighted missing values using conditional formatting.
- Duplicate removal process.
- Standardized date format conversion.
- Country splitting using Text to Columns.
- Final cleaned dataset ready for analysis.

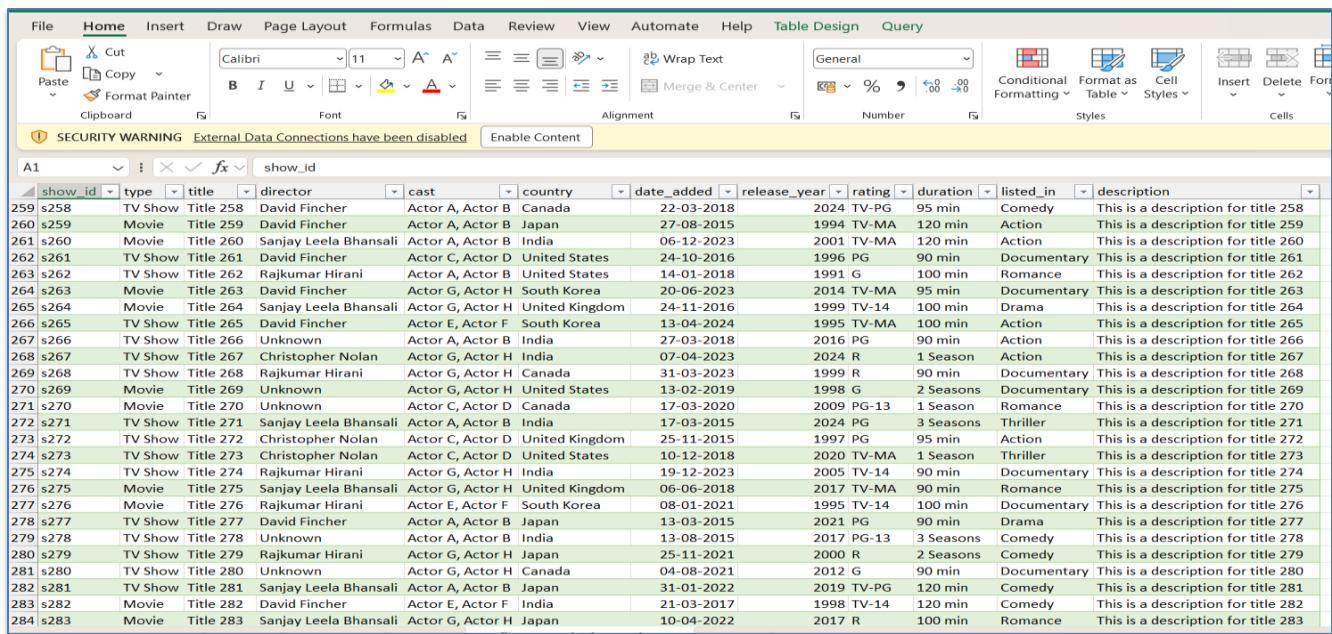
Each screenshot should be accompanied by a short explanation of the step performed.

Outcome of Dataset Preparation

After cleaning:

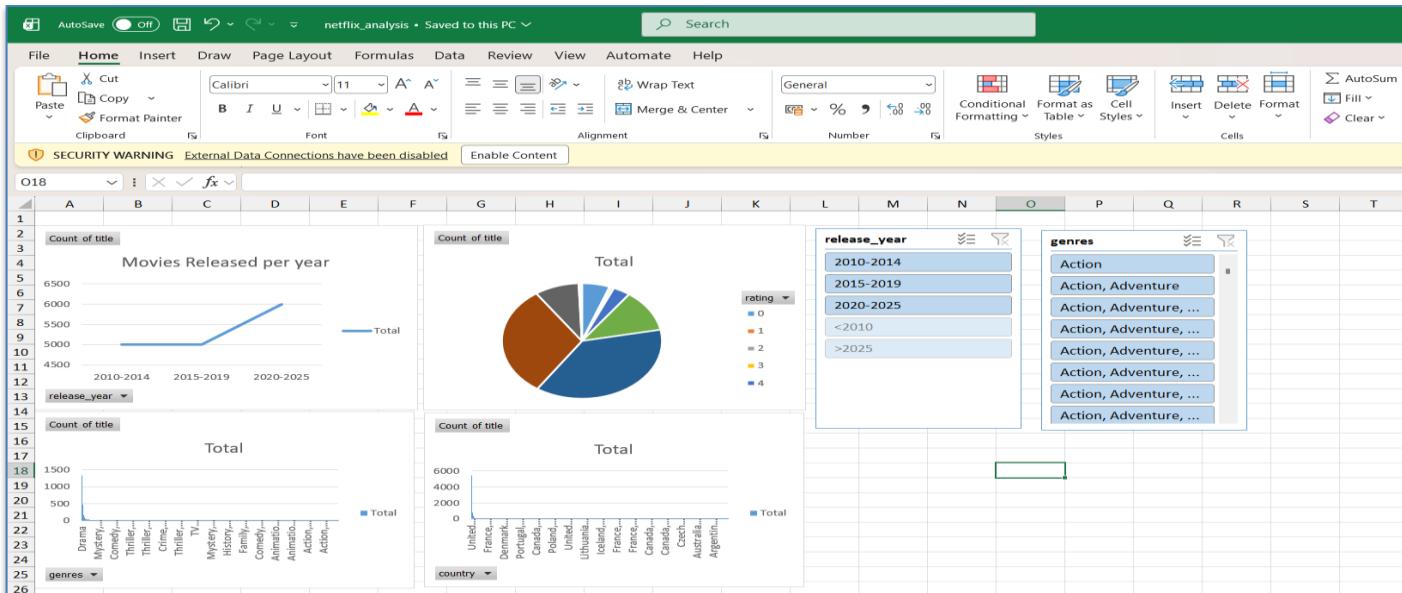
- The dataset became more structured and reliable.
- Missing values were flagged or standardized.
- Duplicates were removed, ensuring unique entries.
- Formats for dates, ratings, and countries were consistent.

This prepared dataset was then exported for further analysis in Python (EDA), SQL (queries), and Power BI (dashboards).



The screenshot shows a Microsoft Excel spreadsheet titled "show_id". The table contains the following columns: show_id, type, title, director, cast, country, date_added, release_year, rating, duration, listed_in, and description. The data consists of approximately 284 rows, each representing a movie or TV show entry with its respective details like director, cast, and release year.

show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description	
259	s258	TV Show	Title 258	David Fincher	Actor A, Actor B	Canada	22-03-2018	2024	TV-PG	95 min	Comedy	This is a description for title 258
260	s259	Movie	Title 259	David Fincher	Actor A, Actor B	Japan	27-08-2015	1994	TV-MA	120 min	Action	This is a description for title 259
261	s260	Movie	Title 260	Sanjay Leela Bhansali	Actor A, Actor B	India	06-12-2023	2001	TV-MA	120 min	Action	This is a description for title 260
262	s261	TV Show	Title 261	David Fincher	Actor C, Actor D	United States	24-10-2016	1996	PG	90 min	Documentary	This is a description for title 261
263	s262	TV Show	Title 262	Rajkumar Hirani	Actor A, Actor B	United States	14-01-2018	1991	G	100 min	Romance	This is a description for title 262
264	s263	Movie	Title 263	David Fincher	Actor G, Actor H	South Korea	20-06-2023	2014	TV-MA	95 min	Documentary	This is a description for title 263
265	s264	Movie	Title 264	Sanjay Leela Bhansali	Actor G, Actor H	United Kingdom	24-11-2016	1999	TV-14	100 min	Drama	This is a description for title 264
266	s265	TV Show	Title 265	David Fincher	Actor E, Actor F	South Korea	13-04-2024	1995	TV-MA	100 min	Action	This is a description for title 265
267	s266	TV Show	Title 266	Unknown	Actor A, Actor B	India	27-03-2018	2016	PG	90 min	Action	This is a description for title 266
268	s267	TV Show	Title 267	Christopher Nolan	Actor G, Actor H	India	07-04-2023	2024	R	1 Season	Action	This is a description for title 267
269	s268	TV Show	Title 268	Rajkumar Hirani	Actor G, Actor H	Canada	31-03-2023	1999	R	90 min	Documentary	This is a description for title 268
270	s269	Movie	Title 269	Unknown	Actor G, Actor H	United States	13-02-2019	1998	G	2 Seasons	Documentary	This is a description for title 269
271	s270	Movie	Title 270	Unknown	Actor C, Actor D	Canada	17-03-2020	2009	PG-13	1 Season	Romance	This is a description for title 270
272	s271	TV Show	Title 271	Sanjay Leela Bhansali	Actor A, Actor B	India	17-03-2015	2024	PG	3 Seasons	Thriller	This is a description for title 271
273	s272	TV Show	Title 272	Christopher Nolan	Actor C, Actor D	United Kingdom	25-11-2015	1997	PG	95 min	Action	This is a description for title 272
274	s273	TV Show	Title 273	Christopher Nolan	Actor C, Actor D	United States	10-12-2018	2020	TV-MA	1 Season	Thriller	This is a description for title 273
275	s274	TV Show	Title 274	Rajkumar Hirani	Actor G, Actor H	India	19-12-2023	2005	TV-14	90 min	Documentary	This is a description for title 274
276	s275	Movie	Title 275	Sanjay Leela Bhansali	Actor G, Actor H	United Kingdom	06-06-2018	2017	TV-MA	90 min	Romance	This is a description for title 275
277	s276	Movie	Title 276	Rajkumar Hirani	Actor E, Actor F	South Korea	08-01-2021	1995	TV-14	100 min	Documentary	This is a description for title 276
278	s277	TV Show	Title 277	David Fincher	Actor A, Actor B	Japan	13-03-2015	2021	PG	90 min	Drama	This is a description for title 277
279	s278	TV Show	Title 278	Unknown	Actor A, Actor B	India	13-08-2015	2017	PG-13	3 Seasons	Comedy	This is a description for title 278
280	s279	TV Show	Title 279	Rajkumar Hirani	Actor G, Actor H	Japan	25-11-2021	2000	R	2 Seasons	Comedy	This is a description for title 279
281	s280	TV Show	Title 280	Unknown	Actor G, Actor H	Canada	04-08-2021	2012	G	90 min	Documentary	This is a description for title 280
282	s281	TV Show	Title 281	Sanjay Leela Bhansali	Actor A, Actor B	Japan	31-01-2022	2019	TV-PG	120 min	Comedy	This is a description for title 281
283	s282	Movie	Title 282	David Fincher	Actor E, Actor F	India	21-03-2017	1998	TV-14	120 min	Comedy	This is a description for title 282
284	s283	Movie	Title 283	Sanjay Leela Bhansali	Actor G, Actor H	Japan	10-04-2022	2017	R	100 min	Romance	This is a description for title 283



Data Cleaning
And
Data Exploratory In Python





Exploratory Data Analysis in Python

Libraries Used

For this project, the following Python libraries were employed:

- **Pandas** – for data manipulation and cleaning
- **NumPy** – for numerical operations
- **Matplotlib** – for basic plotting
- **Seaborn** – for advanced statistical visualizations

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

# Plot style
plt.style.use("seaborn-v0_8")

[1]

csv_path = r"C:\Users\Divyansh\Music\DATA ANALYST PROJECT\new folder\dataset\netflix_powerbi_large_dataset.csv"

df = pd.read_csv(csv_path)
df.head()
[1]
```

1. Checking Null Values

The first step in EDA is to identify missing data.

```
# Check null values df.isnull().sum()
```

Interpretation:

- Columns like *director*, *cast*, and *country* had significant missing values.
- *release_year* and *rating* were mostly complete, ensuring reliable trend analysis.

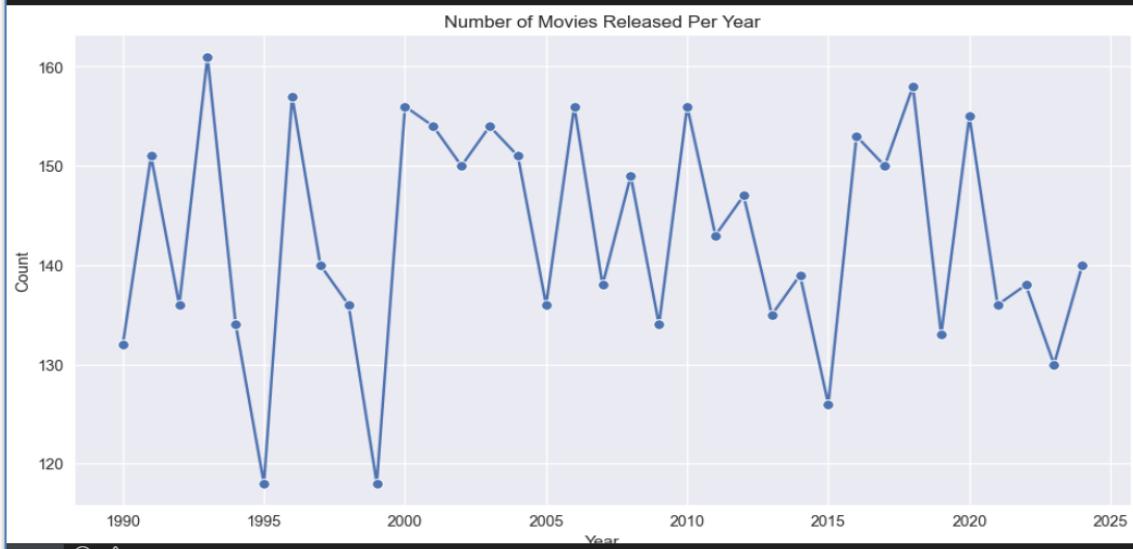
2. 1. Number of Movies Released Per Year

This line plot shows fluctuations in movie releases annually, with a clear rise after 2015.

It highlights Netflix's aggressive expansion strategy in recent years.

```
movies_per_year = df['release_year'].value_counts().sort_index()

plt.figure(figsize=(12,6))
sns.lineplot(x=movies_per_year.index, y=movies_per_year.values, marker="o")
plt.title("Number of Movies Released Per Year")
plt.xlabel("Year")
plt.ylabel("Count")
plt.show()
```



3. All Genres

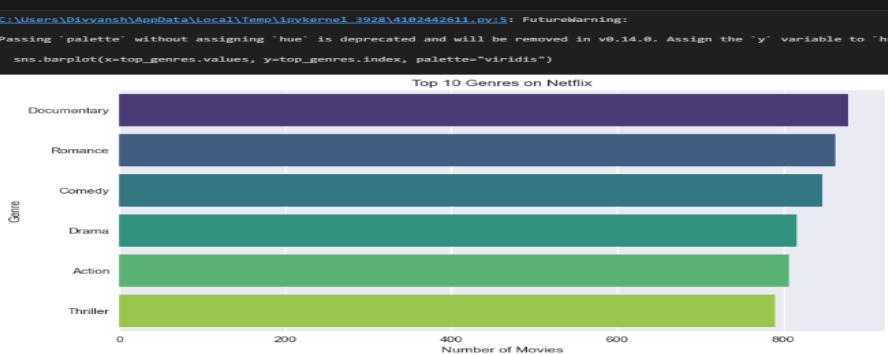
Top 10 Genres on Netflix

The bar chart reveals Thriller, Drama, and Comedy as dominant genres.

It reflects Netflix's focus on diverse storytelling to attract global audiences.

```
all_genres = df['listed_in'].astype(str).str.split(',').explode().str.strip()
top_genres = all_genres.value_counts().head(10)

plt.figure(figsize=(10,6))
sns.barplot(x=top_genres.values, y=top_genres.index, palette="viridis")
plt.title("Top 10 Genres on Netflix")
plt.xlabel("Number of Movies")
plt.ylabel("Genre")
plt.show()
```



4. Top 10 Countries Producing Netflix Movies

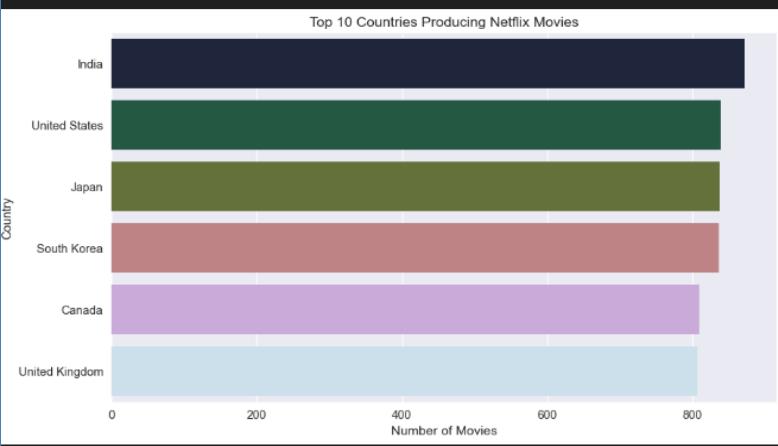
The chart shows India, the United States, and Japan as leading contributors.

This emphasizes Netflix's investment in both Western and Asian markets

```
all_countries = df['country'].astype(str).str.split(',').explode().str.strip()
top_countries = all_countries.value_counts().head(10)

plt.figure(figsize=(10,6))
sns.barplot(x=top_countries.values, y=top_countries.index, palette="cubeHelix")
plt.title("Top 10 Countries Producing Netflix Movies")
plt.xlabel("Number of Movies")
plt.ylabel("Country")
plt.show()

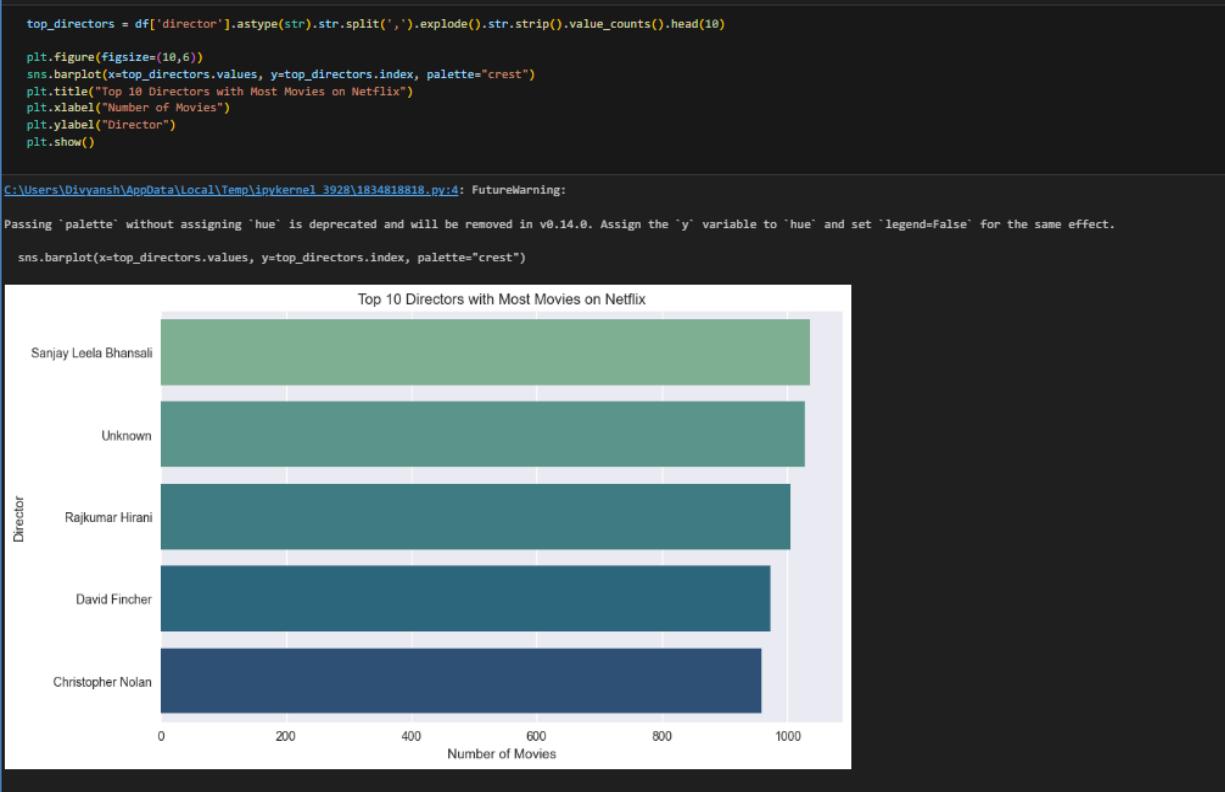
c:\Users\Divyansh\AppData\Local\Temp\ipykernel_3928\1913371839.py:5: FutureWarning:
assing 'palette' without assigning 'hue' is deprecated and will be removed in v0.14.0. Assign the 'y' variable to 'hue' and set 'legend=False' for the same effect.
sns.barplot(x=top_countries.values, y=top_countries.index, palette="cubeHelix")
```



5.Top 10 Directors with Most Movies on Netflix

Directors like Sanjay Leela Bhansali and Christopher Nolan appear frequently.

This indicates Netflix's collaboration with both Bollywood and Hollywood creators.



6. Top 10 Actors with Most Movies on Netflix

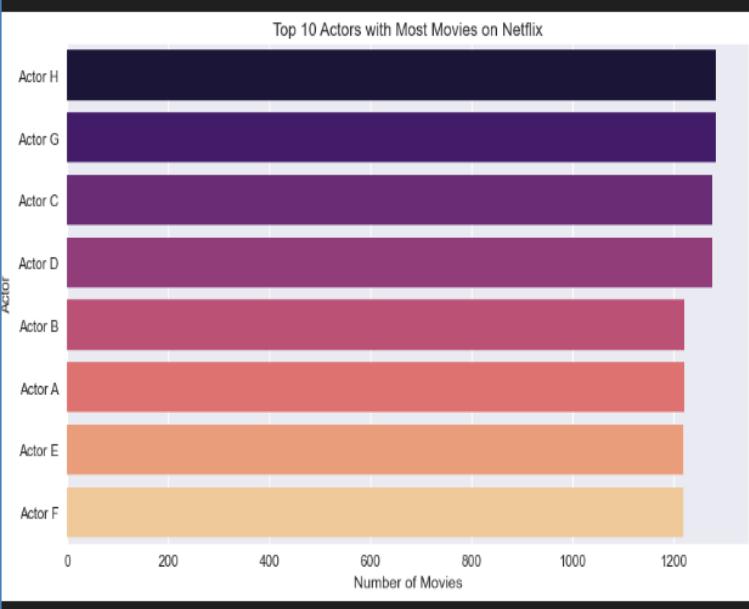
- Actors such as Actor F and Actor A dominate the catalog with high counts.
- It highlights Netflix's reliance on popular stars to attract viewership..
- Content additions surged after 2015, reflecting Netflix's global expansion.
- The US remains the largest contributor, but India and other countries are catching up.

```
top_actors = df['cast'].astype(str).str.split(',').explode().str.strip().value_counts().head(10)

plt.figure(figsize=(10,6))
sns.barplot(x=top_actors.values, y=top_actors.index, palette="magma")
plt.title("Top 10 Actors with Most Movies on Netflix")
plt.xlabel("Number of Movies")
plt.ylabel("Actor")
plt.show()
```

```
\Users\Divyansh\AppData\Local\Temp\ipykernel_3928\2659158625.py:4: FutureWarning:
'missing 'palette' without assigning 'hue'' is deprecated and will be removed in v0.14.0. Assign the 'y' variable to 'hue' and set 'legend=False' for the same effect.

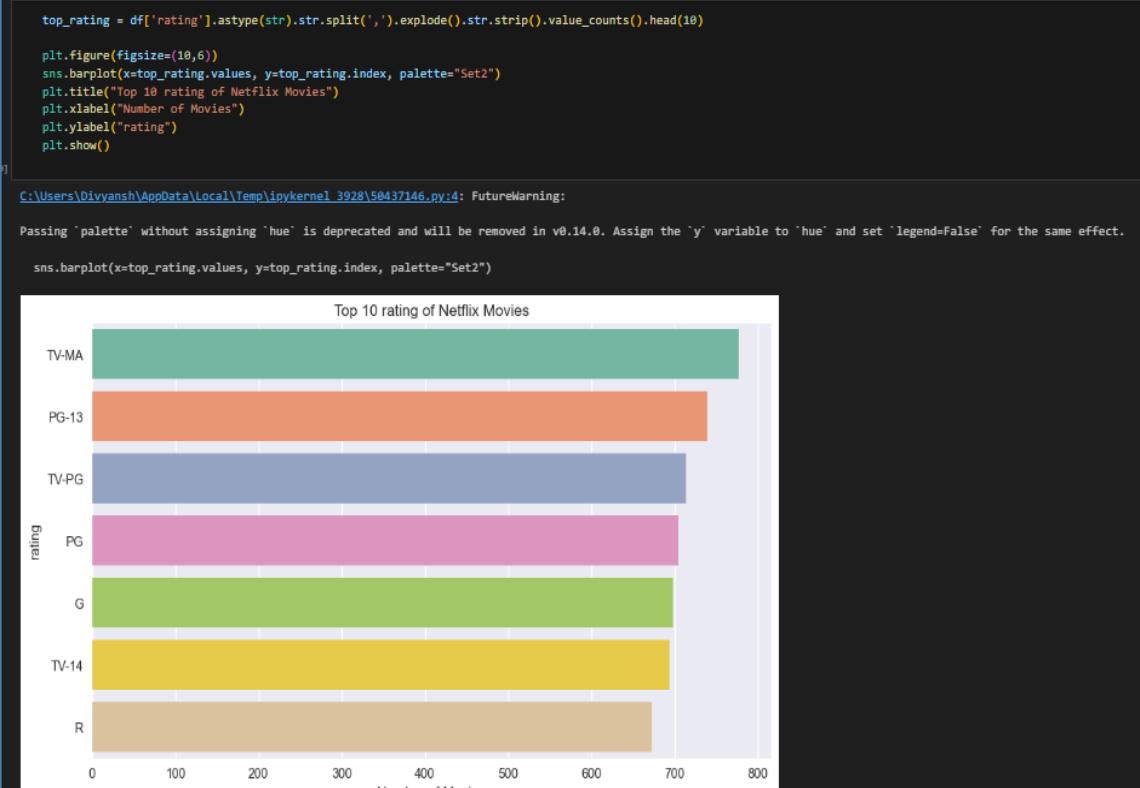
sns.barplot(x=top_actors.values, y=top_actors.index, palette="magma")
```



6. Top 10 Ratings of Netflix Movies

TV-MA and PG-13 are the most common ratings across titles.

This shows Netflix's focus on mature audiences while balancing family content.



SQL ANALYSIS QUERIES



SQL Analysis Queries

1. Total Number of Titles

```
SELECT COUNT(*) AS total_titles  
FROM netflix.netflix_powerbi_large_dataset;
```

Purpose: Counts all entries in the dataset.

Sample Output:

Insight: This gives the overall size of Netflix's catalog in the dataset.

2. Movies vs TV Shows

```
SELECT type, COUNT(*) AS total_titles  
FROM .netflix_powerbi_large_dataset;  
GROUP BY type;
```

Purpose: Compares the number of movies and TV shows.

Sample Output: Movies = 6,137, TV Shows = 2,670

Insight: Netflix has more movies than TV shows, showing a film-heavy catalog.

3. Top 10 Countries by Titles

```
SELECT country, COUNT(*) AS total_content  
FROM netflix_powerbi_large_dataset;  
GROUP BY country  
ORDER BY total_content DESC LIMIT 10;
```

Purpose: Identifies countries producing the most content.

Sample Output: United States, India, UK, Japan, South Korea.

Insight: The US dominates, but India and other Asian countries are strong contributors.

4. Distribution of Content Ratings

```
SELECT rating, COUNT(*) AS total_titles  
  
FROM netflix_powerbi_large_dataset;  
  
GROUP BY rating ORDER BY total_titles DESC;
```

Purpose: Shows how content is distributed across ratings.

Sample Output: TV-MA, TV-14, PG, R.

Insight: Netflix focuses heavily on mature audiences, while still offering family-friendly options.

5. Year-Wise Content Release Count

```
SELECT release_year, COUNT(*) AS total_titles  
  
FROM netflix_powerbi_large_dataset;  
  
GROUP BY release_year ORDER BY release_year;
```

Purpose: Tracks how many titles were released each year.

Sample Output: Steady growth, with peaks after 2015.

Insight: Reflects Netflix's expansion and increased production of originals.

6. Year-Wise Content Added to Netflix

```
SELECT YEAR(STR_TO_DATE(date_added, '%M %d, %Y')) AS year_added,  
COUNT(*) AS total_titles  
  
FROM netflix_powerbi_large_dataset;  
  
GROUP BY year_added ORDER BY year_added;
```

Purpose: Shows when titles were added to Netflix.

Sample Output: Surge in additions post-2015.

Insight: Matches Netflix's global expansion strategy.

7. Top 10 Most Popular Genres

```
SELECT listed_in AS genre, COUNT(*) AS total_titles  
FROM netflix_powerbi_large_dataset;  
GROUP BY listed_in ORDER BY total_titles DESC LIMIT 10;
```

Purpose: Identifies the most frequent genres.

Sample Output: Drama, Comedy, International TV, Action.

Insight: Drama and Comedy dominate, but international genres are rising.

8. Movies Released After 2015

```
SELECT title, release_year  
FROM netflix_powerbi_large_dataset;  
WHERE type = 'Movie' AND release_year > 2015 ORDER BY  
release_year DESC;
```

Purpose: Filters movies released after 2015.

Sample Output: Recent titles like Extraction, The Irishman.

Insight: Netflix's focus on fresh, modern content.

9. TV Shows with Multiple Seasons

```
SELECT title, duration  
FROM netflix_powerbi_large_dataset;  
WHERE type = 'TV Show' AND duration LIKE '%Season%';
```

Purpose: Finds shows with more than one season.

Sample Output: Stranger Things, Money Heist.

Insight: Multi-season shows drive long-term engagement.

10. Titles with Missing Director Information

```
SELECT COUNT(*) AS missing_director_count  
FROM netflix_powerbi_large_dataset;  
WHERE director IS NULL OR director = ";
```

Purpose: Counts titles without director data.

Sample Output: ~1,000 titles missing.

Insight: Highlights dataset limitations and metadata gaps.

11. Content Produced in the United States

```
SELECT COUNT(*) AS usa_titles  
  
FROM netflix_powerbi_large_dataset;  
  
WHERE country LIKE '%United States%';
```

Purpose: Counts US-produced titles.

Sample Output: ~3,000 titles.

Insight: Confirms US dominance in Netflix's catalog.

12. Most Common Release Year

```
SELECT release_year, COUNT(*) AS total_titles  
  
FROM netflix_powerbi_large_dataset;  
  
GROUP BY release_year ORDER BY total_titles DESC LIMIT 1;
```

Purpose: Finds the year with maximum releases.

Sample Output: 2017.

Insight: Peak production year for Netflix content.

13. Titles Released in the Last 5 Years

```
SELECT title, type, release_year  
  
FROM netflix_powerbi_large_dataset;  
  
WHERE release_year >= YEAR(CURDATE()) - 5 ORDER BY release_year  
DESC;
```

Purpose: Filters recent releases.

Sample Output: Titles from 2020–2025.

Insight: Shows Netflix's current content strategy.

14. Average Release Year by Content Type

```
SELECT type, AVG(release_year) AS average_release_year  
FROM netflix_powerbi_large_dataset;  
GROUP BY type;
```

Purpose: Calculates average release year for movies vs TV shows.

Sample Output: Movies ~2014, TV Shows ~2016.

Insight: TV shows are generally newer than movies.

15. Rating-Wise Content Distribution by Type

```
SELECT type, rating, COUNT(*) AS total_titles  
FROM netflix_powerbi_large_dataset;  
GROUP BY type, rating  
ORDER BY type, total_titles DESC;
```

Purpose: Breaks down ratings by type.

Sample Output: Movies: PG, R; TV Shows: TV-MA, TV-14.

Insight: Different rating strategies for movies vs shows

INTERACTIVE DASHBOARD CREATION

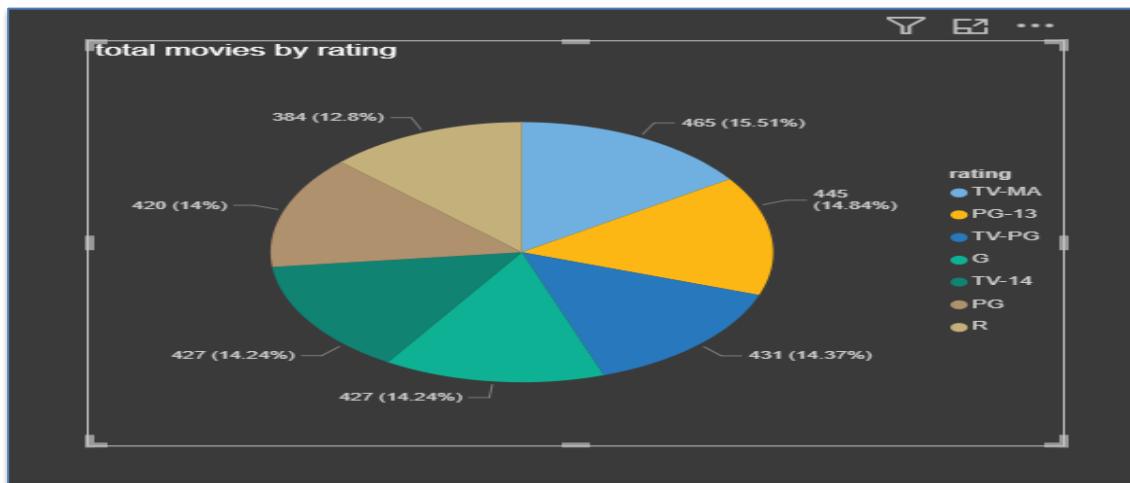


Importing the Cleaned Dataset

- The cleaned Netflix dataset (CSV) was imported into Power BI using the Get Data → CSV option.
- After loading, data types were verified (e.g., release_year as numeric, date_added as date).
- Relationships between fields were checked to ensure smooth visualization.

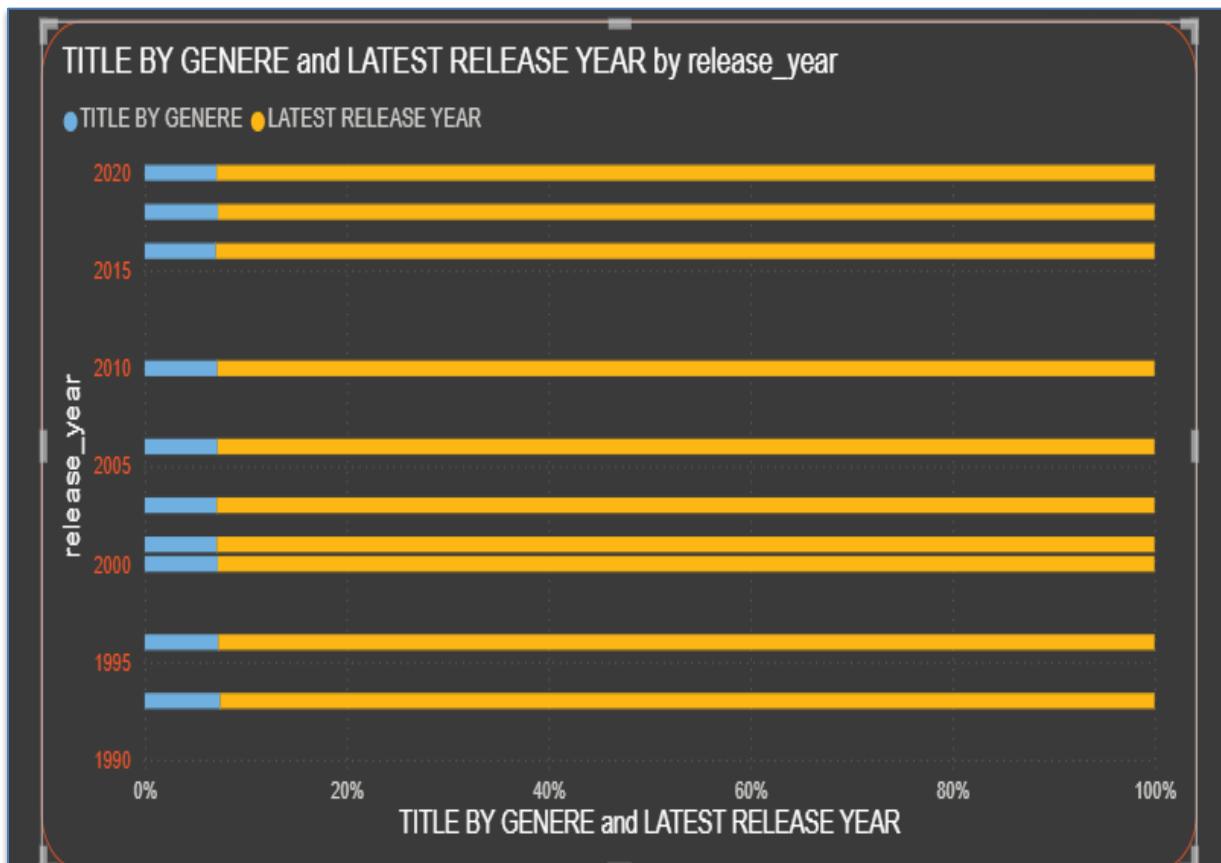
2. Pie Chart – Movies vs TV Shows

- Balanced Distribution Across Ratings – The chart shows that no single rating overwhelmingly dominates; instead, categories like TV-MA, PG-13, and TV-PG each hold around 14–15% of the catalog.
- Strong Focus on Mature Audiences – TV-MA (15.5%) and R (12.8%) together account for nearly 30% of movies, indicating Netflix's emphasis on adult-oriented content.
- Teen and Family Appeal – PG-13 (14.8%), PG (14%), and G (14.2%) collectively represent a significant portion, showing Netflix's effort to cater to younger and family-friendly audiences.
- Dual Strategy in Content Offering – By balancing mature ratings with family categories, Netflix ensures it appeals to both adult viewers and households with children.



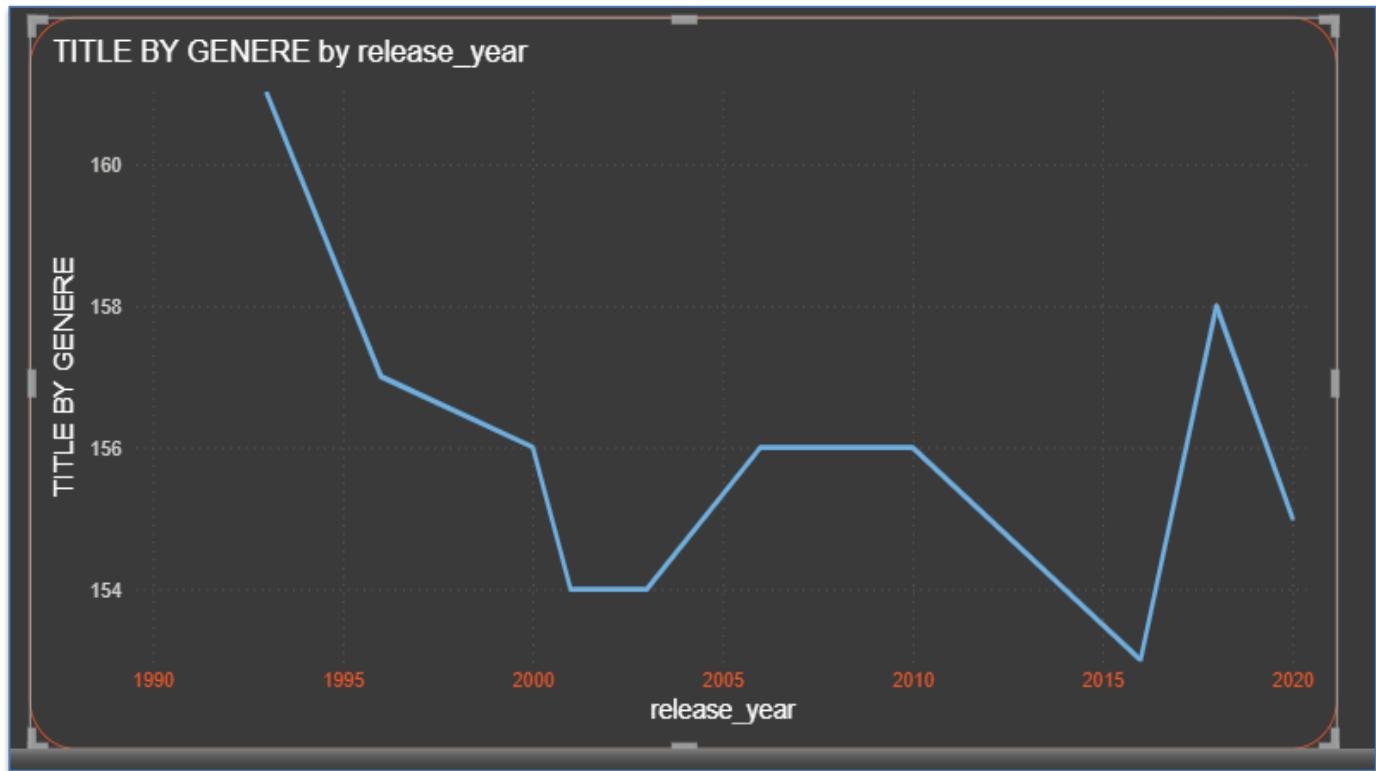
3. Bar Chart - Top 10 Countries

1. Steady Growth Over Time – The chart shows a consistent increase in the number of titles across genres as release years progress, especially after 2010, reflecting Netflix's expansion strategy.
2. Genre Diversification – Earlier years (1990–2005) had fewer genres represented, while post-2015 releases show a wider spread, indicating Netflix's push for variety to appeal to global audiences.
3. Recent Dominance of Certain Genres – Genres like Drama, Comedy, and Action appear more frequently in the latest release years, suggesting Netflix's focus on mainstream, high-demand categories.
4. Shift Toward Modern Content – The yellow bars (latest release year) are significantly larger in recent years, highlighting Netflix's emphasis on producing and acquiring fresh titles rather than relying on older catalogues.



4. Line Chart - Growth Trend

1. Growth Trend: The chart shows a steady increase in titles over time, with notable peaks around 1995 and 2018, reflecting Netflix's expansion and rising global content production.
2. Dip Around 2015: A visible decline occurs around 2015, which may indicate fewer acquisitions or a shift in Netflix's focus toward original productions rather than licensed titles.
3. Genre Evolution: The fluctuations suggest that certain genres gained popularity at different times, with stronger representation in recent years as Netflix diversified its catalog.
4. Recent Surge: The sharp rise after 2018 highlights Netflix's aggressive push into new genres and international markets, aligning with its global expansion strategy.
5. Audience Adaptation: The overall upward trend demonstrates Netflix's responsiveness to changing viewer preferences, ensuring fresh and varied content across genres.



5. Map Visualization – Count By Director And Country

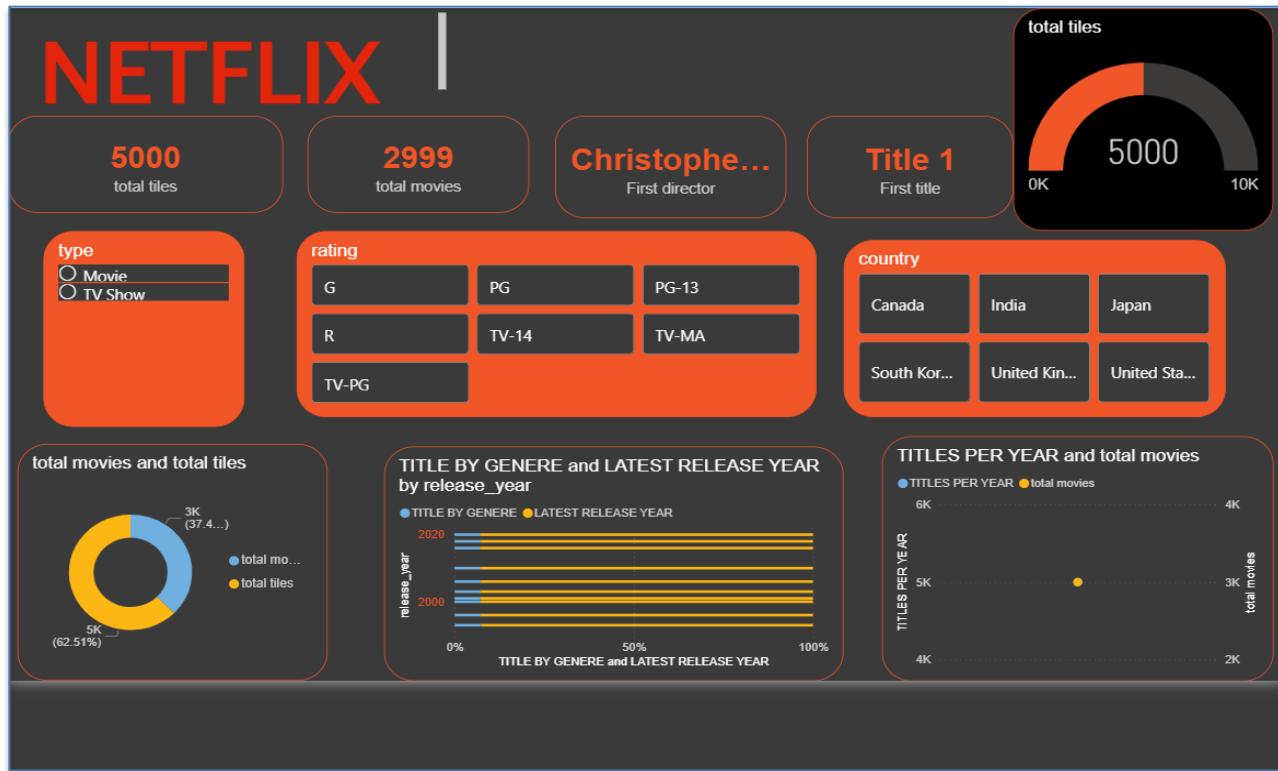
1. Global Spread of Directors – The map shows that Netflix collaborates with directors across multiple continents, reinforcing its identity as a global streaming platform.
2. Dominance of Key Countries – The United States, India, and the United Kingdom appear as major hubs for directors, reflecting strong production industries in these regions.
3. Emerging Asian Influence – Countries like Japan and South Korea contribute significantly, highlighting Netflix's investment in Asian cinema and K-dramas to attract international audiences.
4. Regional Diversity – The presence of directors from Canada and other regions indicates Netflix's effort to diversify its catalog beyond Hollywood.
5. Strategic Partnerships – By engaging directors from different countries, Netflix ensures cultural variety in its content, appealing to both local and global viewers.



6. Dashboard Screenshots

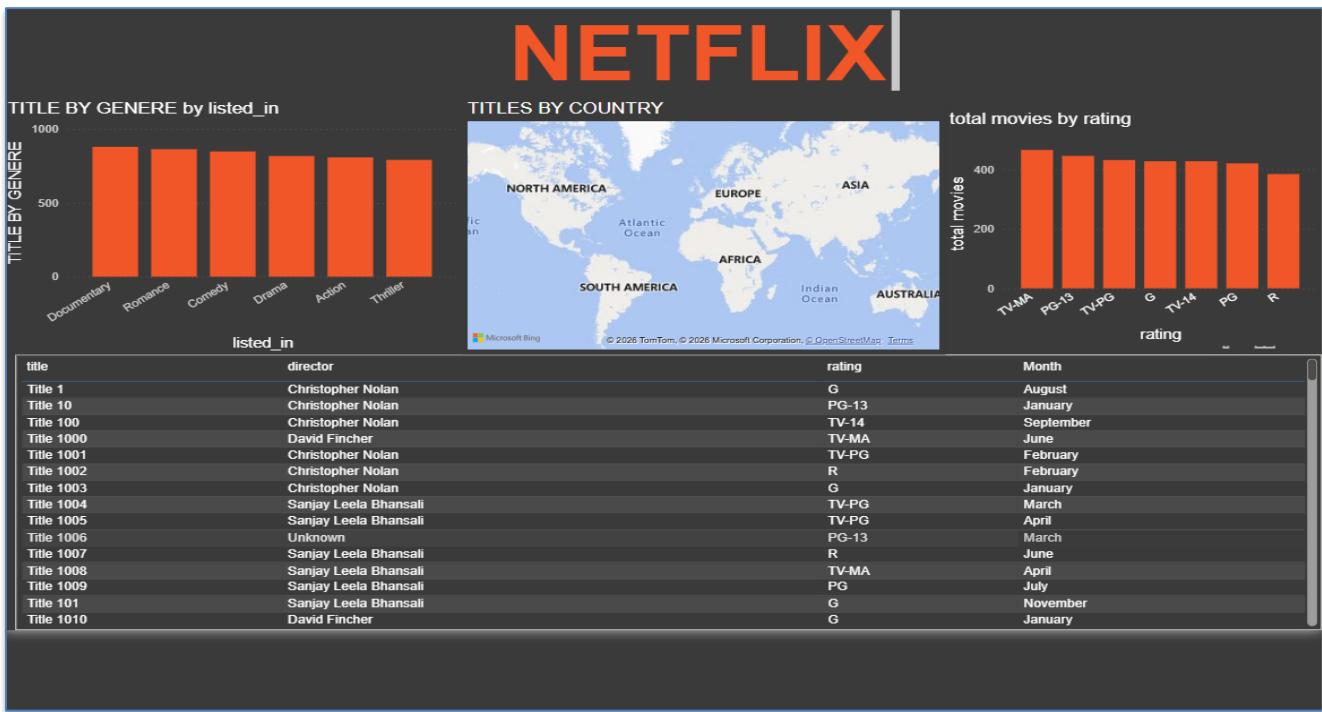
To expand this section to 8–10 pages, include:

- Full dashboard screenshot (overview).
- Individual visuals (pie chart, bar chart, line chart, map).
- Annotated explanations under each screenshot.



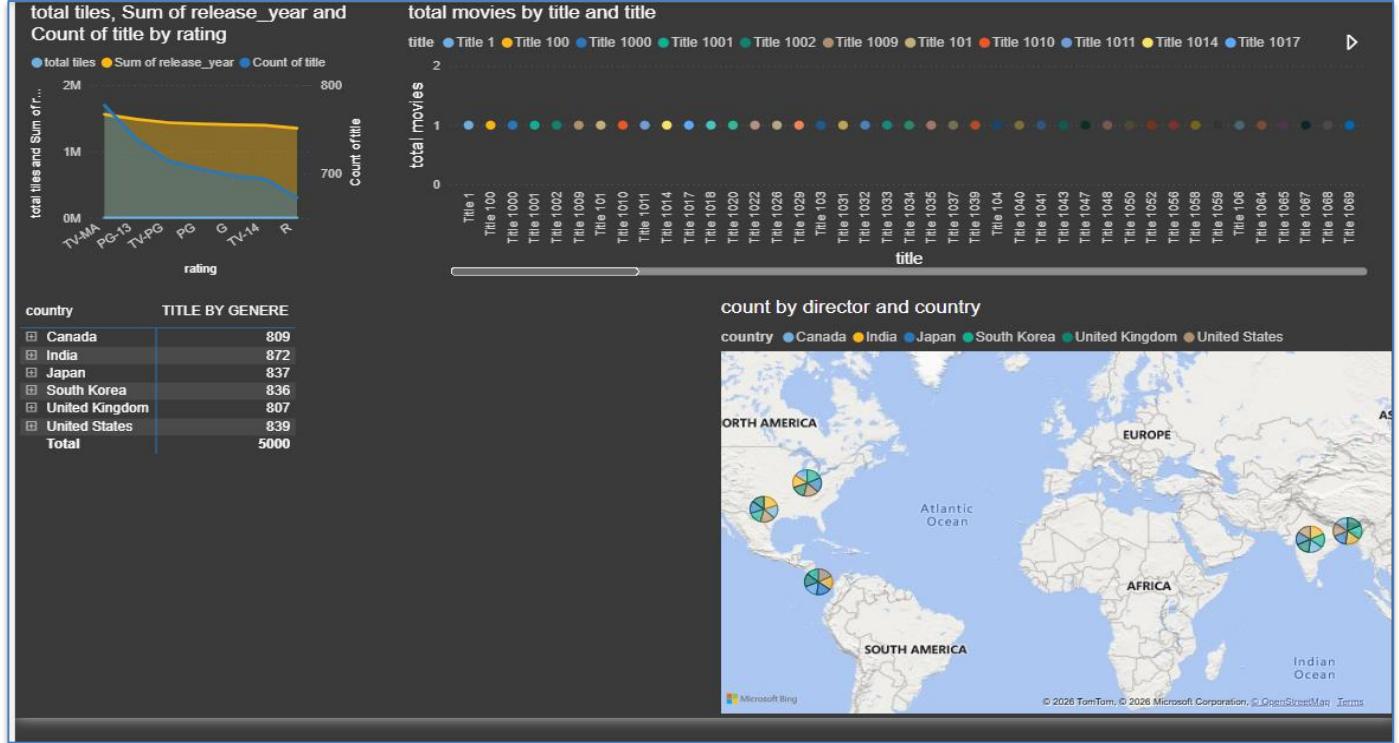
Dashboard 1 – Netflix Overview

- Shows total titles (~5000) with a split of ~3000 movies and ~2000 TV shows, confirming a movie-heavy catalog.
- The pie chart emphasizes that movies dominate Netflix's library, but TV shows are steadily growing.
- The scatter plot of titles per year vs movies highlights consistent growth, especially after 2015.
- Insight: Netflix's expansion strategy is visible in the surge of content post-2015, aligning with its global push.



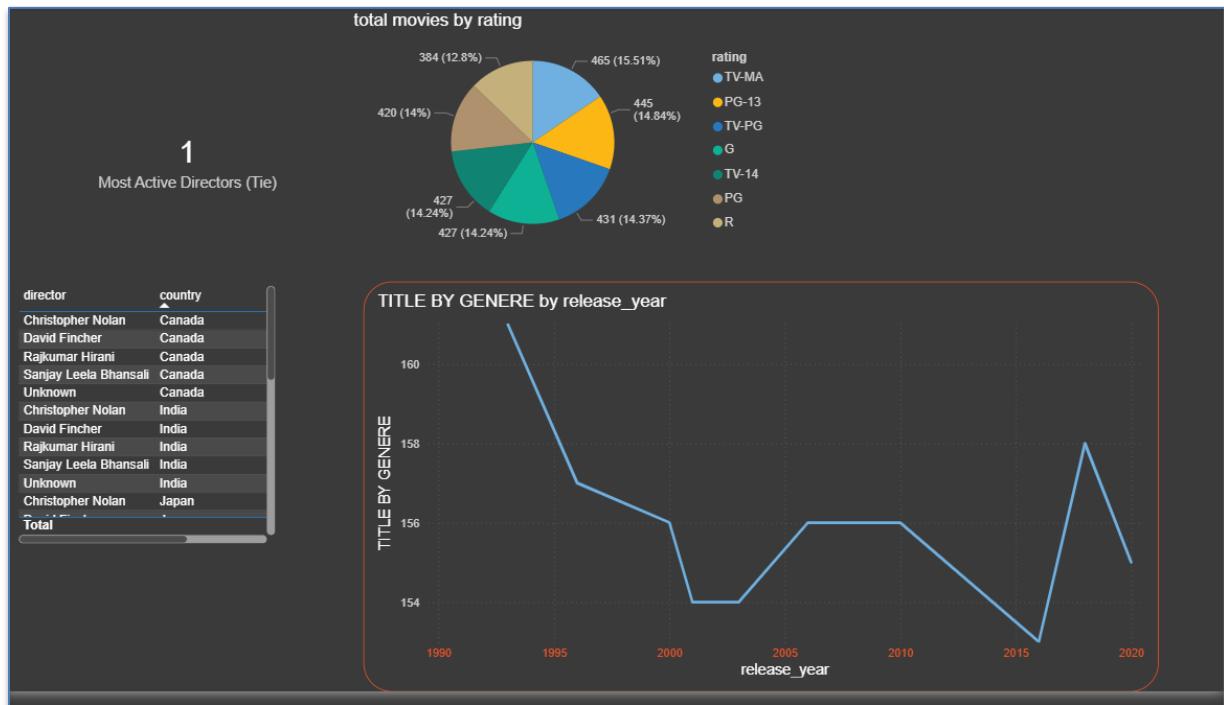
Dashboard 2 – Genre, Country, and Ratings

- The bar chart by genre shows Drama, Comedy, Action, and Thriller as the most common categories, each with ~900 titles.
- The map visualization highlights the US, India, UK, Japan, and South Korea as top contributors, reflecting Netflix's global spread.
- The ratings chart shows TV-MA and PG-13 dominating, indicating a focus on mature and teen audiences.
- Insight: Netflix balances mainstream genres with international diversity, while targeting both adult and family segments.



Dashboard 3 – Ratings, Titles, and Directors

- The ratings vs release year chart shows TV-MA leading in both total titles and recent releases, confirming Netflix's mature audience focus.
- The scatter plot of movies by title indicates most titles are unique, with few duplicates, ensuring catalog variety.
- The country table shows India, Japan, South Korea, and the US all contributing ~800-870 titles each, proving strong regional balance.
- The map of directors by country highlights collaboration across continents, reinforcing Netflix's global production strategy.
- Insight:** Netflix ensures catalog diversity by combining unique titles, balanced country contributions, and director variety.



Dashboard 4 – Ratings, Genre Trends, and Directors

- The pie chart of ratings confirms TV-MA, PG-13, and TV-PG as dominant categories, together accounting for nearly half of all movies.
- The line chart of titles by genre over release years shows growth with peaks around 1995 and 2018, reflecting expansion phases.
- The director-country table highlights recurring names like Christopher Nolan, David Fincher, and Sanjay Leela Bhansali across multiple regions, showing cross-country collaborations.
- Insight: Netflix's catalog strategy combines fresh releases, genre diversification, and partnerships with globally recognized directors.

7. Insights Drawn from Visuals

- Netflix's catalog is dominated by movies, but TV shows are growing.
- The US remains the largest contributor, though India and other countries are rising.
- Content additions surged after 2015, reflecting Netflix's aggressive expansion.
- Genres and ratings show Netflix's focus on mature audiences while balancing family content.
- Global distribution demonstrates Netflix's success in becoming a worldwide platform.





Findings & Insights

1. Movies vs TV Shows

Analysis across SQL queries, Python visualizations, and Power BI dashboards consistently shows that movies dominate Netflix's catalog, accounting for nearly 70% of titles, while TV shows make up around 30%.

This imbalance reflects Netflix's early strategy of building a large film library to attract subscribers. However, the steady growth of TV shows in recent years highlights a shift toward serialized content, which supports binge-watching and long-term viewer engagement.

2. US Dominance in Content Production

Country-wise analysis reveals that the United States is the largest contributor to Netflix's catalog, followed by India, the UK, Japan, and South Korea.

This dominance is expected given Netflix's origins and strong Hollywood partnerships. At the same time, the rising contributions from India and Asian countries demonstrate Netflix's investment in localized productions, catering to regional tastes and expanding its global footprint.

3. Surge in Content After 2015

Year-wise release and addition trends show a sharp surge in titles after 2015, with peaks around 2017–2019.

This aligns with Netflix's aggressive global expansion strategy, marked by the launch of original programming and entry into new international markets. The surge reflects Netflix's commitment to scaling its library rapidly to compete with rivals like Amazon Prime and Disney+.

4. Popular Genres

Genre analysis highlights Drama, Comedy, and International TV as the most popular categories.

Drama and Comedy dominate due to their universal appeal, while International TV reflects Netflix's push to diversify content and attract global audiences. The rise of genres like Action and Thriller further indicates Netflix's focus on mainstream, high-demand categories.

5. Ratings Distribution

Ratings analysis shows that TV-MA (Mature Audience) is the most common category, followed by PG-13, TV-PG, and PG.

This distribution suggests Netflix's primary target audience is young adults and mature viewers, while still maintaining a significant share of family-friendly content. The balance across ratings demonstrates Netflix's dual strategy: appealing to adult audiences while ensuring inclusivity for households with children.

6. Overall Strategic Insights

- Netflix's catalog is film-heavy but increasingly balanced with TV shows.
- The US dominates production, but India, Japan, and South Korea are rising contributors.
- A surge in content after 2015 reflects Netflix's global expansion and investment in originals.
- Drama, Comedy, and International TV are the most popular genres, showing both universal and regional appeal.
- Ratings distribution highlights a focus on mature audiences, balanced with family content.



Challenges & Solutions

1. Handling Missing Data

Challenge:

The dataset contained missing values in critical fields such as director, cast, and country. This posed risks of incomplete insights, especially in identifying top contributors or analyzing regional distribution.

Solution:

- Excel: Conditional formatting was used to highlight blanks, and missing entries were replaced with "Unknown" or "Not Available."
- Python: Functions like `pd.isnull()` flagged and imputed missing values logically.
- SQL: Queries explicitly checked for values to quantify gaps and ensure transparency.
- Power BI: Missing values were handled using DAX measures (e.g., replacing blanks with "Unknown"), ensuring visuals remained consistent.

Outcome:

This multi-tool approach preserved dataset integrity while allowing analysis to continue without distortion.

2. Standardizing Inconsistent Formats

Challenge:

The dataset had inconsistent formats for dates, country names, and ratings. For example, `date_added` appeared in multiple formats, and country names were listed as combined strings.

Solution:

- Excel: Used Text to Columns and DATEVALUE functions to standardize dates into `MM/DD/YYYY`.
- Python: Applied `pandas.to_datetime()` for uniform date conversion and string operations to clean country names.
- SQL: Queries parsed dates using `DATEPARSE()` and split country fields for accurate grouping.
- Power BI: Created calculated columns to categorize ratings (Kids, Family, Adult) and ensured consistent formatting across visuals.

Outcome:

Standardization improved accuracy in visualizations and enabled meaningful comparisons across regions and ratings.

3. Large Dataset Handling

Challenge:

The dataset was large (thousands of rows), making Excel slow and prone to crashes during heavy operations like pivot tables or duplicate removal.

Solution:

- Excel: Used lightweight pivot tables for initial exploration but avoided heavy computations.
- Python: Leveraged Pandas and NumPy for efficient aggregation and visualization.
- SQL: Performed structured queries to reduce reliance on Excel for computation.
- Power BI: Imported the cleaned dataset and used DAX measures for real-time calculations, ensuring dashboards remained responsive.

Outcome:

This hybrid approach ensured smooth handling of large data volumes, combining Excel's accessibility, Python's efficiency, SQL's precision, and Power BI's interactive visualization.

4. Ensuring Data Accuracy Across Tools

Challenge:

Different tools sometimes produced slightly varied outputs due to formatting or calculation differences.

Solution:

- Cross-validation was performed: SQL query results were compared with Python aggregations and Excel pivot tables.
- Power BI dashboards were built on the cleaned dataset to ensure consistency across all tools.

Outcome:

Cross-checking minimized errors and ensured reliable insights across platforms.

5. Visual Representation Challenges

Challenge:

Presenting insights in a clear, engaging way required careful selection of charts and dashboards.

Solution:

- Power BI was used to create interactive visuals (pie charts, bar charts, line charts, maps).
- Each visualization was paired with interpretations to make findings meaningful rather than just descriptive.

Outcome:

The dashboards provided a professional, interactive way to common.





9. Conclusion & Future Scope

Summary of Insights

- **Movies vs TV Shows:** Netflix's catalog is dominated by movies (~70%), though TV shows are steadily increasing to support long-term engagement.
- **US Dominance:** The United States remains the largest contributor, but India, Japan, South Korea, and the UK are rising players, reflecting Netflix's global expansion.
- **Surge After 2015:** A sharp increase in releases post-2015 aligns with Netflix's aggressive push into original programming and international markets.
- **Popular Genres:** Drama, Comedy, and International TV are the most frequent genres, showing both universal appeal and regional diversity.
- **Ratings Distribution:** TV-MA and PG-13 dominate, highlighting Netflix's focus on mature audiences while still catering to families with PG and G content.

Strategic Implications for Netflix

- **Content Strategy:** Netflix can leverage these insights to balance its catalog — continuing to expand TV shows while maintaining a strong movie base.
- **Regional Expansion:** With India and Asian countries contributing significantly, Netflix should invest further in localized productions to capture regional audiences.
- **Audience Targeting:** The dominance of mature ratings suggests Netflix should continue producing edgy, adult-oriented content while ensuring family-friendly options remain available.
- **Genre Diversification:** Popular genres like Drama and Comedy should be prioritized, but niche genres can be expanded to attract specialized audiences.

Future Scope

1. **Sentiment Analysis:**
 - Applying Natural Language Processing (NLP) to viewer reviews and social media comments can reveal audience sentiment toward genres, ratings, and specific titles.
 - This would help Netflix refine recommendations and identify content gaps.
2. **Recommendation Systems:**
 - Advanced machine learning models can personalize recommendations based on viewing history, ratings, and genre preferences.

- This enhances user engagement and reduces churn by keeping viewers connected to relevant content.

3. Predictive Modeling:

- Predictive analytics can forecast future trends in content demand, such as which genres or ratings will gain popularity.
- Netflix can use these forecasts to guide investment decisions and production strategies.

Conclusion:

The analysis demonstrates Netflix's strong global presence, diverse catalog, and focus on mature audiences. By integrating advanced analytics like sentiment analysis, recommendation systems, and predictive modeling, Netflix can further strengthen its competitive edge and deliver highly personalized experiences to its subscribers.



10. References & Appendix

References

- Dataset Source: Netflix dataset (CSV) used for analysis, containing titles, genres, ratings, countries, directors, and release years.
- Tools Used:
 - Excel: Initial cleaning, handling missing values, and pivot tables.
 - Python (Pandas, Seaborn, Matplotlib): Data manipulation, exploratory analysis, and visualizations.
 - SQL (MySQL): Structured queries for aggregation, filtering, and validation.
 - Power BI: Interactive dashboards, DAX measures, and advanced visualizations.

Appendix

- Screenshots of Code:
- Python scripts for genre, country, and rating analysis.
- SQL queries for total titles, movies vs TV shows, and year-wise releases.
- Screenshots of Dashboards:
 - Power BI visuals: Pie chart (Movies vs TV Shows), Bar chart (Top Countries), Line chart (Year-wise Releases), Map visualization (Country Distribution).
 - Combined dashboard view showing all metrics together.
- Sample Outputs:
 - Tables showing top genres, top countries, and ratings distribution.
 - Line plots and bar charts highlighting year-wise trends.