

**A SEMINAR REPORT**  
**ON**  
**NETFLIX DATA ANALYTICS**

**In partial fulfilment of requirements for the degree of  
Bachelor of Technology**

**In**

**Computer Science and Engineering**

Submitted by

**Bikash Chhatra (2201322038)**

**6<sup>th</sup> Semester**

**Academic Year : 2024-25**



**Department of Computer Science and Engineering**  
**Einstein Academy of Technology and Management**  
**Baniatangi, Khordha, Bhubaneswar, 752060**



# CERTIFICATE

This is to certify that the dissertation work entitled “NETFLIX DATA ANALYTICS ” is a Bonafide work done by Bikash Chhatra, Regd.No-2201322038 submitted in partial fulfilment for the award of Bachelor in Technology degree in Computer Science & Engineering from Einstein Academy of Technology and Management affiliated to Biju Patnaik University of Technology, Odisha in the academic session: 2024-25.

**Project Guide**  
**Prof. Nibedita Chhatoi**

**HOD**  
**Dr. Rati Ranjan Sahoo**

# Acknowledgement

I am deeply grateful for the opportunity to present this seminar on “NETFLIX DATA ANALYTICS “. This endeavour would not have been possible without the guidance, support, and encouragement of several individuals and institutions.

First and foremost, I express my sincere gratitude to Prof. Nibedita Chhatoi, whose insights and expertise have been invaluable in shaping this seminar. Their constructive feedback and constant motivation have played a pivotal role in the successful completion of this work.

I extend my heartfelt thanks to department of CSE, EATM, for providing the necessary resources and a conducive environment to explore this topic in depth. Special thanks to our HOD Dr. Rati Ranjan Sahoo, for their mentorship throughout this process.

I also wish to acknowledge the unwavering support of my family, friends, and colleagues. Whose encouragement kept me motivated during challenging times.

Finally, I am thankful to all the authors, researchers, and pioneers in the fields of Data Security On the Cloud, whose work served as the foundation for my research and presentation.

This seminar has been a significant learning experience, and I hope it adds value to the ongoing discourse in this critical domain.

**Name – Bikash Chhatria**

**Regd.No. - 2201322038**

**Department of CSE**

**EATM, Bhubaneswar**

## ABSTRACT

The **Netflix Data Analytics** project focuses on developing an interactive dashboard using Power BI to analyze Netflix's extensive content library and user engagement metrics. By leveraging a dataset encompassing details such as genre, release year, duration, ratings, and country availability, the project aims to uncover insights into content trends, distribution patterns, and audience preferences.

The project utilizes **Power BI** for building the dashboard, incorporating tools such as **Power Query** for data transformation and **DAX (Data Analysis Expressions)** for creating calculated metrics. Visual elements like bar charts, line graphs, pie charts, maps, and slicers enhance interactivity and user engagement.

Key findings from the analysis reveal that Movies constitute approximately 65% of the content, with Drama and Comedy being the most prevalent genres. The United States, India, and the United Kingdom emerge as top content contributors. A notable increase in content additions is observed post-2015, indicating a strategic shift towards original content production. Ratings analysis suggests a diverse target audience, with content catering to various age groups and regional standards.

# INDEX

Sl. No.	Topic	Page no.
1	Introduction & Dataset Overview	01
2	Tools and Technology used	02
3	Data Transformation & Modeling process	03
4	Dashboard design	05
5	Visual components	13
6	Future Enhancements	31
7	conclusion	32

# 1.0 Introduction & Dataset overview

In the age of digital streaming, data has become a cornerstone for shaping content strategies and improving user experiences. Netflix, as a global leader in on-demand streaming, possesses a vast and dynamic content catalog that spans numerous countries, genres, and audience demographics. This project, *Netflix Data Analytics*, aims to harness this data to uncover actionable insights using a visually engaging Power BI dashboard.

The primary objective is to analyze Netflix's content library by exploring key factors such as genre distribution, content type (Movies vs. TV Shows), regional availability, release trends, content duration, and ratings. Through comprehensive data modeling and visualization, the project provides stakeholders with the ability to explore patterns, assess performance, and support data-driven decisions for content expansion and audience targeting.

By transforming raw Netflix data into meaningful insights, the project not only demonstrates technical proficiency in data visualization but also underscores the critical role of analytics in driving strategic planning in the media and entertainment sector.

The primary objective is to analyze Netflix's content library by exploring key factors such as genre distribution, content type (Movies vs. TV Shows), regional availability, release trends, content duration, and ratings. Through comprehensive data modeling and visualization, the project provides stakeholders with the ability to explore patterns, assess performance, and support data-driven decisions for content expansion and audience targeting.

Ultimately, this project serves not only as a practical exercise in business intelligence and data visualization, but also as a demonstration of how companies like Netflix can leverage analytics to drive creative and strategic content initiatives on a global scale.

## 1.1 Overview

The dataset used for this project comprises publicly available data on Netflix titles, originally sourced from repositories like **Kaggle**. It includes over 6,000 entries and reflects a diverse array of content available on the platform.

Key Dataset Characteristics:

- **Source:** Kaggle – Netflix Movies and TV Shows Dataset
- **Format:** CSV (Comma-Separated Values)
- **Total Records:** Approximately 6,000+ titles

### 1.2 Fields and columns:

- Title – **Netflix Data analytics**
- Type – Movie or TV Show
- Genre – Primary and secondary genres
- Country – Country or countries of availability
- Duration – Length in minutes (Movies) or number of seasons (TV Shows)
- Release Year – Year the title was released
- Date Added – Date the title was added to Netflix
- Rating – Content rating (e.g., TV-MA, PG, R)
- Description – Brief synopsis of the content
- **Date of Last Update:** [03-04-2025]

This dataset serves as the foundation for all analysis performed in Power BI. Prior to visualization, the data underwent a thorough cleaning and transformation process to address inconsistencies and optimize it for modeling and reporting.

## 2 :Tools & Technology used

### 2.1 • Power BI

Power BI is the primary platform used for this project. It provides robust capabilities for data modeling, interactive dashboard creation, and real-time visual analytics. With its intuitive interface and strong integration features, Power BI enables users to build insightful visual reports that are both user-friendly and dynamic.



Power BI

Figure: 2.1 powerbi

### 2.2 • Excel/CSV

The Netflix dataset was sourced in CSV (Comma-Separated Values) format and reviewed initially in Excel. These formats are commonly used for storing structured data and are compatible with Power BI, making it easy to import, inspect, and manage the raw data before analysis.

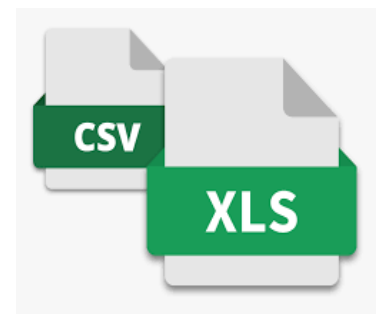


Figure: 2.2 Excel/CSV File

### 2.3 • Power Query

Power Query is a powerful data connection and transformation tool within Power BI. It is used for performing ETL (Extract, Transform, Load) operations. In this project, Power Query helped clean the dataset by removing duplicates, correcting inconsistencies, and reshaping the data for effective modeling.



Figure: 2.3 power Query



## 2.4 • DAX (Data Analysis Expressions)

DAX is a formula language in Power BI used for creating custom calculations and KPIs (Key Performance Indicators). It was utilized to calculate metrics such as the total number of titles, percentage of Movies vs. TV Shows, average duration, and top countries by title count—enhancing the analytical depth of the dashboard.

## 2.5 • Visualization Tools

The project employs various data visualization elements:

- **Bar Charts:** For comparing categorical data like genre and country.
- **Line Graphs:** For showing trends over time (e.g., yearly content additions).
- **Pie Charts:** For proportional comparisons (e.g., content type distribution).
- **Maps:** To visualize country-wise content availability.
- **Slicers:** For interactive filtering by genre, country, year, or rating—making the dashboard more user-friendly and customizable.

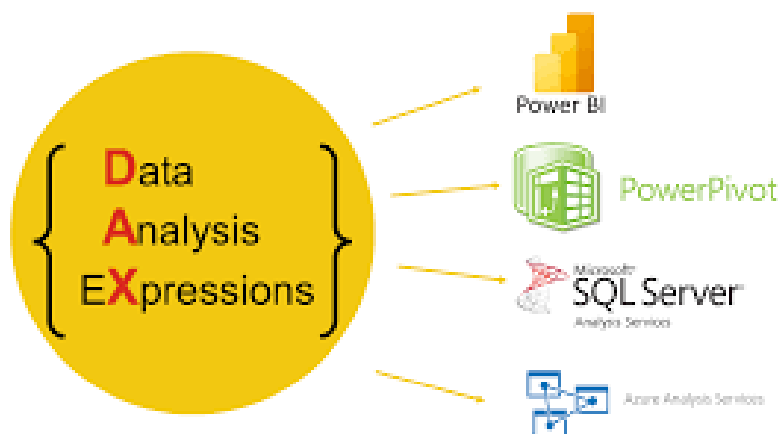


Figure: 2.4/2.5 DAX&Tools

## 3 : Data Cleaning And Modeling Process

### 3.1 Data Cleaning Process :

Before performing any meaningful analysis, it was essential to clean and prepare the dataset to ensure accuracy, consistency, and usability. The following steps were undertaken during the data cleaning phase:

#### 3.1.1 Removed Nulls and Duplicates

Incomplete data entries and duplicate records were identified and eliminated to maintain data integrity. This step ensured that the final visualizations reflected accurate and unique content insights without distortion.

#### 3.1.2 Standardized Country and Genre Entries

Country names and genre labels were often inconsistent due to formatting or variations (e.g., "USA" vs. "United States"). These entries were standardized for uniformity, enabling reliable grouping and filtering during analysis.

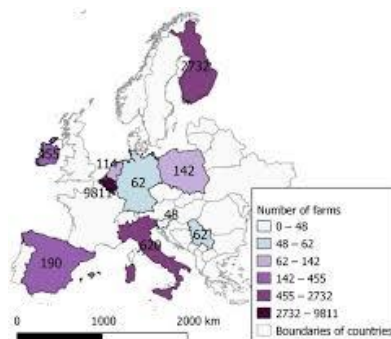


Figure 3.1.2 Country & Genre Entries

#### 3.1.3 Parsed Dates and Durations

Dates from fields like Date Added and Release Year were converted into consistent formats for temporal analysis. Similarly, content durations (minutes for movies, seasons for TV shows) were parsed and normalized for comparison.

#### 3.1.4 Grouped Ambiguous Genres for Consistency

Titles often included multiple genres or variations (e.g., "Sci-Fi & Fantasy", "Science Fiction"). These were grouped under consistent categories to simplify genre analysis and avoid fragmented insights.

## 3.2 Data Transformation (Modeling)

Data transformation is a crucial step in the data preparation process, as it reshapes raw and cleaned data into a structured and analysis-ready format. For the Netflix Data Analytics project, the transformation process played a central role in enabling powerful visualizations and meaningful insights through Power BI. This phase involved several key operations to enhance the usability, granularity, and interactivity of the dataset.

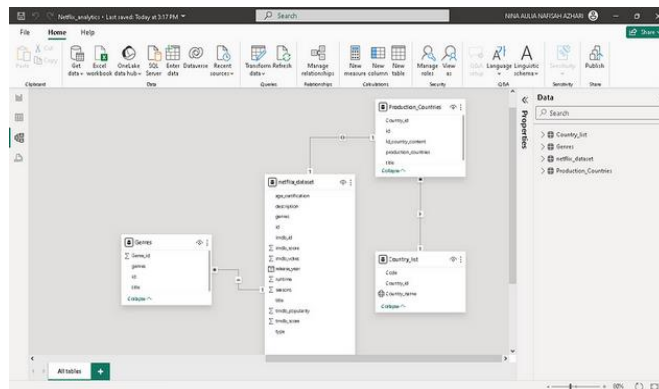


Figure 3.3.1 Data cleaning and relation making process

### 3.2.1. Extracting the Year from “Date Added”

The original dataset contained a “Date Added” field, which included complete date values (e.g., "March 5, 2020"). While this was helpful for time-specific insights, most analysis in this project focused on broader trends over the years. To support this, a new column was created by extracting just the **year** from the “Date Added” field. This allowed for clearer visualizations showing how Netflix’s content library has evolved over time and helped stakeholders identify peak years of content growth or shifts in strategy.

### 3.2.2. Converting Durations to a Uniform Format

Netflix titles include both Movies and TV Shows, each using different metrics to describe duration. Movies typically list duration in minutes (e.g., "120 min"), while TV Shows use the number of seasons (e.g., "2 Seasons").

To allow for meaningful comparisons and filtering, the data was normalized. Movies retained a numeric minute value, while TV Shows were encoded with a consistent numerical representation of seasons. This step was essential in creating calculated metrics and visual elements that compare content lengths across types.

### 3.2.3 Categorizing Multi-Genre Entries

Many titles in the Netflix catalog are tagged with multiple genres (e.g., "Action & Adventure, Sci-Fi & Fantasy"). These multi-label entries could fragment genre-based analysis if left unchecked. To maintain consistency and clarity, a process was applied to categorize each title under a **primary genre** while optionally allowing users to explore additional genre tags through filters. This helped avoid duplication in visuals and ensured a clean, interpretable genre breakdown, while still retaining the richness of the content.

### 3.2.4 Creating Calculated Columns for Type Breakdown and Region Tagging

To support deeper insights, several **calculated columns** were introduced using DAX (Data Analysis Expressions) in Power BI. One of these columns broke down content type clearly, enabling quick comparisons between Movies and TV Shows through KPIs and visuals. Another set of calculated fields was used for **region tagging**, allowing for country-based filtering and mapping.

For instance, titles were tagged with a primary country of origin or availability, making it possible to analyze Netflix's regional content strategy. In cases where a title was associated with multiple countries, a consistent tagging logic was used (e.g., the first-listed country or a predefined priority list).

## 4: Dashboard Design

The design of the dashboard is one of the most crucial components in delivering insights effectively. A well-structured and visually appealing dashboard ensures that users can understand complex data quickly and interact with it effortlessly. For this project, Power BI was used to build an intuitive and engaging dashboard that supports both exploration and decision-making.

- The design incorporates the following key elements:

### 4.1 Clean Layout with Easy Navigation

The dashboard layout was designed with simplicity and clarity in mind. Visual elements were arranged logically, grouping similar information together—such as content type comparisons, genre distributions, and temporal trends. The design avoids clutter, uses consistent fonts and spacing, and ensures that every chart or visual serves a clear purpose. This clean structure allows users to focus on the insights without being distracted by unnecessary design elements.

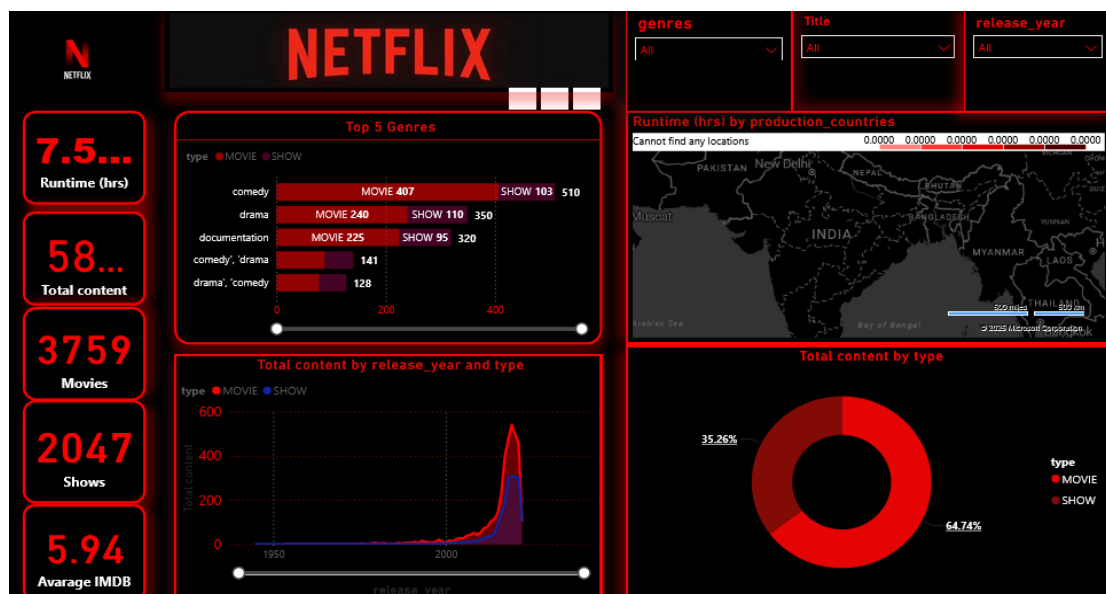


Figure 4.1: Netflix Dashboard

## 4.2 Slicers for Filtering by Country,Year,Type,Genre, and Rating

To enhance interactivity, slicers were implemented throughout the dashboard.



Figure 4.2.1 Slicers for country

These allow users to filter data dynamically based on key dimensions such as:

- **Country** – Explore content availability by region
- **Year** – Focus on specific timeframes
- **Type** – Toggle between Movies and TV Shows
- **Genre** – Drill down into preferred genres
- **Rating** – Analyze content targeting various age groups

These slicers enable users to tailor the dashboard view to their specific interests or questions, turning the static data into an exploratory tool.

## 4.3 Color-Coded Visuals for Clarity

Each visual element on the dashboard uses a thoughtful color scheme to enhance readability. For example, Movies and TV Shows are displayed in distinct colors, and genre charts use a palette that's both vibrant and accessible. This helps users differentiate categories at a glance and improves the visual appeal of the dashboard.

### 4.4 Optimized for Quick Insights

The dashboard was optimized for performance and usability. Visuals load quickly, filters respond in real time, and tooltips provide additional context without overwhelming the view. Key metrics and KPIs are placed at the top or in focal areas to ensure stakeholders can get essential insights within seconds

## 5: Visual Components

### 5.1 Bar Charts – Type Comparison

Bar charts were used to visually compare the distribution of content types—**Movies vs. TV Shows**. These visuals help identify which format dominates Netflix’s library and how this has changed over time or across different countries and genres. The bars are easy to interpret and give a clear visual of content balance.

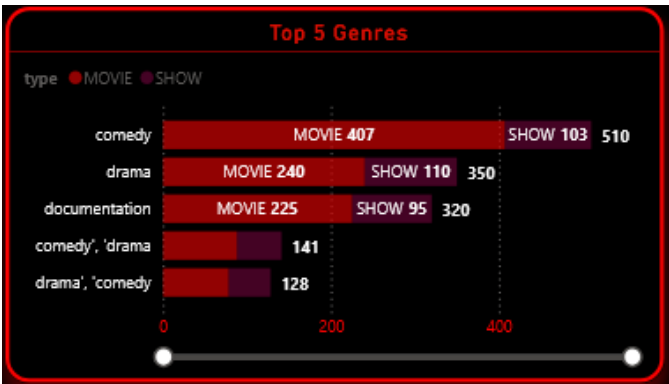


Figure 5.1 Top 5 genes using Bar graph

### 5.2 Line Graphs – Yearly Trends

Line graphs were employed to represent **year-wise trends** in content addition. By plotting the number of new titles added each year, these visuals highlight Netflix’s growth, peak production years, and strategic shifts. Line graphs are ideal for spotting patterns and long-term trends.

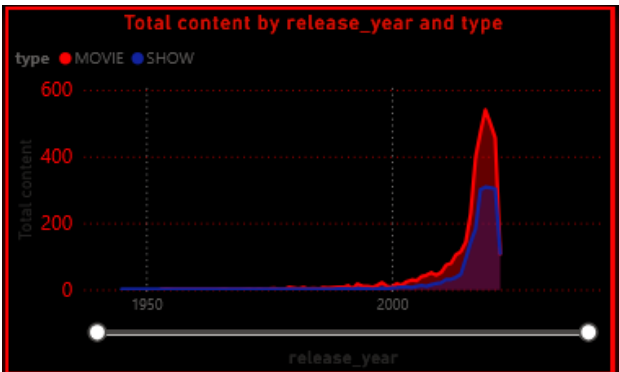


Figure 5.2 Lines Graph



### 5.3 Maps – Country-wise Availability

Geographic maps were used to illustrate **content availability by country**. These visuals highlight which countries have the most titles available and help identify regional diversity in Netflix’s offerings. This is useful for analyzing global content strategy and localization trends.

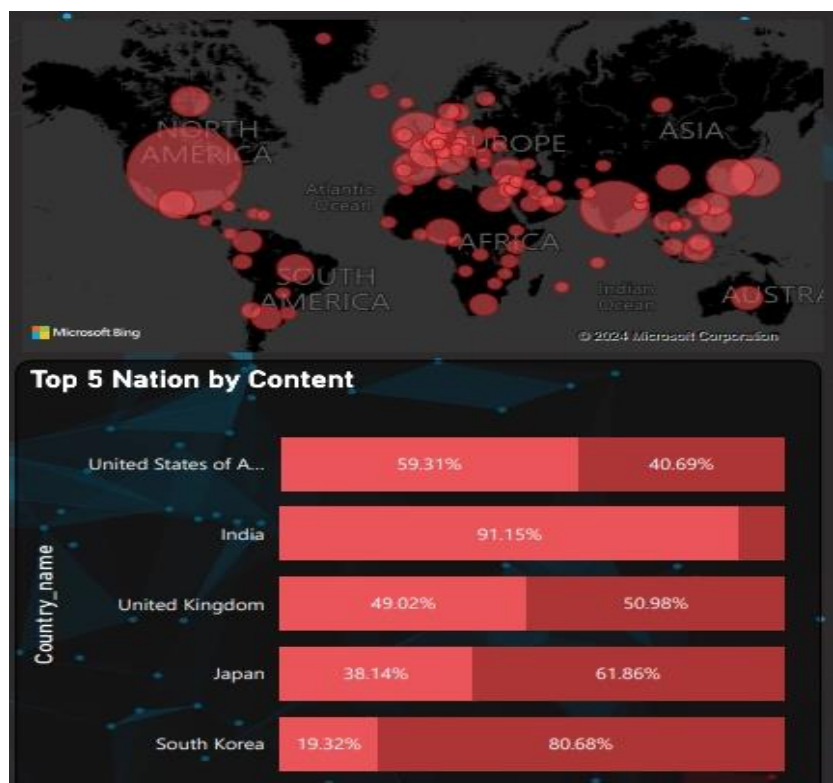


Figure 5.3 Geographical map

### 5.4 Slicers – Interactive Filtering

Slicers are dynamic filters that let users **interactively explore data**. In the dashboard, slicers were added for **country, genre, type, year, and rating**, allowing users to personalize the view and drill down into specific subsets of data. They enhance the dashboard's usability and engagement



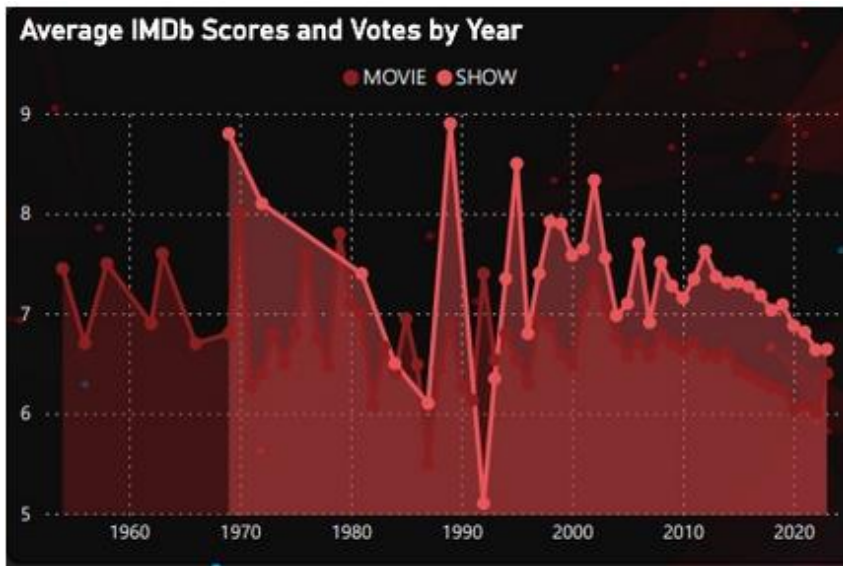


Figure 5.4 slicer

## 5.5 Ratings & Audience Targeting –

Netflix categorizes its content using various **ratings** such as **TV-MA, PG, R, G, and TV-Y7**, which indicate the appropriate audience age group for each title. These ratings are essential for content compliance, regional censorship, and guiding viewers. By analyzing the distribution of these ratings, the dashboard reveals how Netflix targets different **demographics**—from children and families to mature audiences. For example, a high proportion of **TV-MA** content suggests a focus on adult-oriented programming, while the presence of **TV-Y** or **PG** titles highlights Netflix's investment in family-friendly or youth content. This analysis helps stakeholders understand how well Netflix caters to various audience segments and complies with **regional content standards** and viewer expectations. It also informs content strategy decisions—whether to produce more kid-friendly series, expand mature-themed content, or maintain a balanced mix.

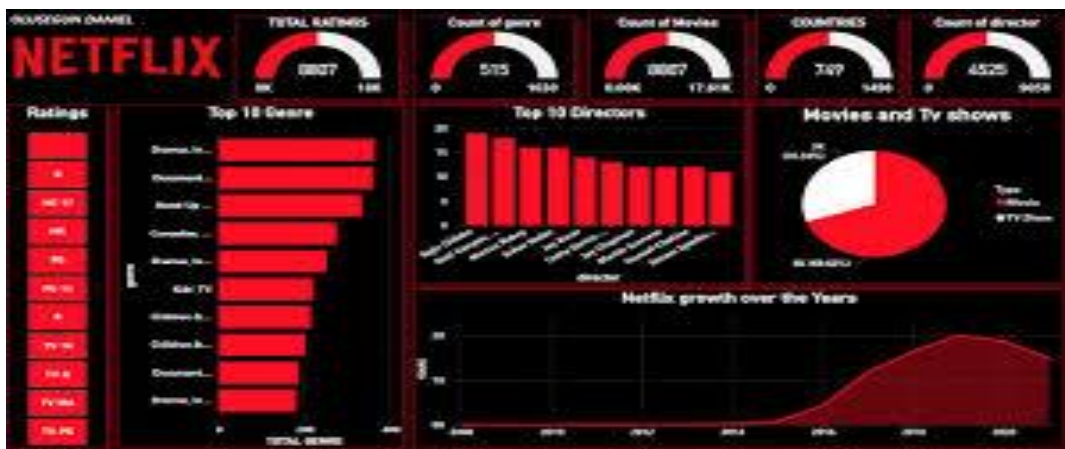


Figure 5.5.1 customer ratings

## 6. Future Enhancements

To further improve the depth and impact of the Netflix Data Analytics dashboard, several future enhancements are proposed. By integrating external rating sources like IMDb or Rotten Tomatoes, the platform can offer richer insights into content reception and quality. Incorporating subscription or viewership statistics would add valuable context to content performance, enabling a more strategic evaluation of what resonates with audiences. Differentiating between Netflix Originals and licensed content could provide clarity on production investments and content sourcing trends. Additionally, leveraging AI for genre prediction or personalized content recommendations would introduce a layer of intelligent forecasting, supporting smarter decision-making for content planning and user engagement.

## 6.1 Add IMDb or Rotten Tomatoes Scores

Integrating external review data from IMDb or Rotten Tomatoes would allow users to assess content quality and popularity, providing a more holistic view of viewer satisfaction and critical reception.

## 6.2 Integrate Subscription/Viewership Stats

Combining content data with subscription or viewership metrics would help measure real audience engagement. This could highlight which genres or regions drive the most views, supporting more informed content strategy decisions.

## 6.3 Compare Originals vs. Licensed Titles

Introducing a filter or metric to differentiate **Netflix Originals** from **licensed content** would offer insights into the company's investment strategy, production trends, and success of original programming over time.

## 6.4 Use AI for Genre Prediction or Content Suggestion

Implementing AI/ML algorithms could enable predictive analytics, such as forecasting emerging genres or suggesting content bundles based on user behavior, enhancing personalization and content planning.



figure 6.1 Netflix's Evolving competitive landscape.

## 7 . Conclusion

The Netflix Data Analytics project illustrates how data-driven insights can be harnessed to better understand and enhance a global content strategy. By using Power BI to transform a complex dataset into an interactive dashboard, this project enables a detailed exploration of Netflix's vast content library. From genre distribution and content type comparisons to regional availability and audience targeting through ratings, the dashboard serves as a comprehensive analytical tool for stakeholders.

The data cleaning and transformation process ensured accuracy and consistency, while thoughtful data modeling using a star schema supported efficient performance and user-friendly navigation. The visual design—featuring bar charts, line graphs, pie charts, and interactive slicers—was developed to allow users to filter and drill down into specific content segments easily. This empowers users to uncover content trends, track Netflix's growth over time, and explore differences in offerings across countries and genres.

Looking ahead, the project lays a strong foundation for future enhancements, such as integrating external review scores (IMDb, Rotten Tomatoes), viewership statistics, and predictive analytics using AI. These additions could deepen the strategic value of the dashboard and offer even greater clarity on audience engagement and content success.

In conclusion, this project not only showcases the value of visual analytics in media strategy but also highlights how data tools can support smarter, more informed decisions in the fast-evolving world of digital entertainment

## 8. References

- 1. Netflix Titles Dataset – Kaggle**  
*Shivam Bansal. "Netflix Movies and TV Shows Dataset". Kaggle. Available at:*  
<https://www.kaggle.com/datasets/shivamb/netflix-shows>
- 2. Microsoft Power BI Documentation**  
*Microsoft Learn. "Power BI Documentation."*  
<https://learn.microsoft.com/en-us/power-bi/>
- 3. DAX Guide – Data Analysis Expressions**  
*SQLBI. "DAX Guide: The definitive reference for DAX."*  
<https://dax.guide/>
- 4. Power Query Documentation**  
*Microsoft Learn. "Power Query in Power BI."*  
<https://learn.microsoft.com/en-us/power-query/>
- 5. Data Visualization Best Practices**  
*Tableau Public Resources and Microsoft Power BI Blog*