Amity University Online, Noida, Uttar Pradesh, India

Masters Of Business Administration

A
MAJOR PROJECT
ON

Data Analysis Using Power BI (Company Name – Netflix)

Submitted By
Name – Komal Negi
Enrollment Number: - A9920124005577(el)
Semester 4th- Major Project (Data Science)
Guide/Mentor - Mrs. Roshini Ganesh

DECLARATION

I, Komal Negi, a student pursuing an MBA in Data Science at Amity University Online, hereby declare that the project work entitled “Data Analysis using Power BI" has been completed by me during the academic year 2025 under the guidance of Mrs. Roshini Ganesh.

I further declare that this project is an original and bona fide work done solely by me. The outcome of the project is the result of my own efforts and has not been submitted to any other university or institution for the award of any degree or diploma.

Komal Negi

Signature of Student

TABLE OF CONTENTS

CHAPTER 1: ABSTRACTION

1.1 Background of the Study

The emergence of Over-The-Top (OTT) streaming platforms has led to a major transformation in the global entertainment industry. Instead of relying on traditional cable or satellite television, audiences have shifted to online platforms where content is available anytime and anywhere. Netflix, established in 1997 and introduced as a streaming service 2007 has grown into one of the world's most dominant streaming providers.

In today's competitive environment, content plays a key role in audience engagement and subscription retention.
This project seeks to use Microsoft Power BI to analyze Netflix's publicly available dataset and visually interpret distribution patterns. By assessing key aspects such as Movies vs. TV Shows composition, top producing countries, genre preferences, and maturity ratings,

1.2 Need for the Study

The modern entertainment ecosystem has undergone a drastic shift with the increasing popularity of OTT platforms. Netflix, being one of the pioneers in streaming services, faces continuous pressure to retain and expand its global audience base.

The need for this study arises from the following key points:

High competition in digital media services

Multiple streaming platforms are entering the market with exclusive regional and international content.

Competitors continuously invest in content acquisition and original productions to attract subscribers.

To sustain competitive advantage, Netflix must monitor how its content library appeals to diverse audiences.

Increasing demand for personalized and localized content

Audiences now prefer content tailored to their culture, language, and interests.

Localization strategies significantly affect subscriber growth in new regions.

Analysis of country-wise and genre-wise content helps evaluate how well Netflix fulfills localized demand.

Requirement for content-based performance evaluation

Content type, volume, and relevance determine user engagement and viewing retention.

Continuous evaluation supports improvement in content quality and diversification.

Analytical insights guide content licensing decisions, including renewal or removal of titles.

Expansion goals into emerging OTT markets

Countries in Asia, Africa, and South America are seeing rapid growth in OTT adoption.

Market expansion demands strong understanding of regional entertainment trends.

Data analysis offers insights for market-specific strategic planning.

Rising production of Netflix Originals

Netflix increasingly invests in in-house productions to strengthen brand identity.

Need to determine whether Originals cover varied genres and global audiences effectively.

Analysis supports decisions regarding future investments in original content.

Apart from external market demands, advanced analytics provides internal strategic benefits:

1.3 Scope of the Study

The scope of this study is limited to:

Publicly available Netflix dataset in CSV format

Content-related attributes (e.g., type, genre, rating, release year, country)

Non-viewership and non-financial metrics

Focus includes:

Descriptive data analysis

Power BI-based dashboard creation

Analytical interpretation for content strategy improvement

The study does not cover:

Real-time user behavior analytics

Subscription patterns or revenue generation

Internal algorithms or policy details of Netflix

This ensures the project remains purely content-analysis based.

1.4 Significance of the Study

The findings of this study are beneficial to:

Media Businesses: Helps in understanding how content diversity influences audience reach

Content Producers: Offers insights into demand-based genre and regional preferences

Marketing Professionals: Supports targeted promotion strategies

Students and Data Analysts: Demonstrates practical application of Power BI for analytics

Researchers: Provides background for future studies involving audience data and trends

Further, the study shows how descriptive analytics leads to better data-driven decision-making in digital platforms.

1.5 Objectives of the Study

The major objectives of this project are to:

Analyze Netflix content distribution (Movies vs. TV Shows)

Examine regional contribution and country-wise presence

Identify trending genres and popular categories

Study growth in Netflix content over the years

Explore maturity ratings and audience segmentation

Understand content strategy through visual dashboards

Provide analytical recommendations based on insights

These objectives guide the analytical steps and dashboard design throughout the project.

1.6 Hypotheses

In this study, descriptive analytics is primarily used to observe and interpret patterns in Netflix's content distribution. However, forming hypotheses helps provide a structured framework for identifying whether variations in the dataset indicate meaningful trends. Establishing hypotheses allows the analysis to remain objective and supports decision-making through observed data evidence.

The following hypotheses are formulated:

Null Hypothesis (H0):
There is no significant variation in the distribution of Netflix content across categories, genres, release years, and countries.

This means that:

Movies and TV Shows are assumed to be proportionately distributed

All countries contribute equally to Netflix's library

Genre selection does not show dominance or preference

The number of titles remains constant over different release years

If this hypothesis holds true, it would suggest that Netflix maintains a balanced and uniform content strategy.

Alternative Hypothesis (H1):
There is a significant variation in the distribution of Netflix content across categories, genres, release years, and countries.

This implies that:

One type of content may have a stronger representation over the other

Certain countries contribute more titles than others

Some genres dominate the catalog based on audience demand

Release trends may show continuous growth or fluctuations over time

Purpose of Hypothesis Testing in This Study

Although conventional statistical hypothesis testing is not applied in this Power BI project, the hypotheses guide:

The visual exploration process

The selection of comparison KPIs

Interpretation of findings from dashboards and charts

These hypotheses are indirectly validated through:

Descriptive data visualization (pie charts, bar charts, line charts)

Country-wise heat maps

Genre comparison and rating distribution

Trend analysis over release years

Conclusion of Hypothesis Section

The inclusion of hypotheses strengthens the analytical structure of the project. By comparing visuals with the expectations framed in H0 and H1, the study gains:

Analytical direction

Improved clarity in interpretation

A basis for conclusions and recommendations in later chapters

Thus, while descriptive in nature, this project incorporates hypothesis-driven thinking to better understand Netflix's content dynamics and strategic positioning within the global streaming industry.

1.7 Summary

In conclusion, this chapter provided a formal introduction to the project, highlighting the need to analyze Netflix data for business insights. The study focuses on content categories, distribution trends, genres, and ratings to understand Netflix's library characteristics. By using Power BI, this project aims to deliver reliable visual analytics that will support decision-making in the OTT content streaming domain.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

A literature review serves as a comprehensive and critical evaluation of previously conducted studies that contribute to the understanding of a subject area. For this project, the focus is on research related to Over-The-Top (OTT) platforms, changing media consumption patterns, Netflix's business model, content strategy, and the role of analytics in digital entertainment. This review synthesizes insights from academic journals, digital media industry reports, analytics case studies, and published papers related to Power BI applications.

The objective is to establish a solid theoretical foundation for the present research and to identify the existing gaps that justify the need for content-driven analytics in Netflix's strategic planning. The chapter also emphasizes the growing importance of data visualization tools in supporting business intelligence within entertainment platforms.

2.2 Overview of the OTT Streaming Industry

The rise of OTT services has transformed global media consumption habits. Instead of scheduled programming on cable or satellite TV, streaming platforms allow users to watch content at their convenience. Studies report that OTT adoption has accelerated due to widespread internet accessibility, smart device usage, and digitalization in developing economies.

Major trends highlighted in the literature include:

- Transition from linear broadcasting to on-demand streaming
- Growth of subscription-based entertainment services (SVOD)
- Consumer preference for flexibility, ad-free streaming, and binge-watching culture
- Focus on original and exclusive content to sustain competitive advantage
- Cross-regional content availability for global user reach

Several reports from major research analysts indicate that OTT platforms are projected to surpass traditional Pay-TV revenues, showcasing a permanent shift in entertainment consumption.

However, challenges are also documented:

- Increase in subscription fatigue due to multiple competing platforms
- Constant need for content innovation and viewer engagement
- Cost pressures linked to content licensing and production

The OTT ecosystem continues to evolve with competitive dynamics driven by both global and regional players.

2.3 Netflix: A Global Streaming Leader

Netflix is the most widely analyzed OTT platform in research literature due to its global dominance and advanced use of analytics. Academic studies highlight its transformation from a DVD-by-mail service to a digital streaming leader operating in 190+ countries.

Netflix's growth factors identified in publications include:

- Focus on digital innovation and highly scalable infrastructure
- Personalized recommendation systems powered by machine learning
- Constant expansion through both licensed and in-house content production
- Investment in international and multilingual titles
- Global distribution strategies with simultaneous releases

Research shows that Netflix has significantly influenced:

- Viewer behavior and binge-watching patterns
- Traditional television broadcasting and cinema industries
- Consumer expectations regarding quality, accessibility, and variety

Additionally, scholarly sources emphasize the platform's shift toward **Netflix Originals**, which ensures higher profit margins, brand ownership, and reduced reliance on third-party licensing.

2.4 Content Strategy and Global Library Expansion

Content is the most critical factor determining user engagement and subscription retention. Literature identifies Netflix's strategic focus areas:

- Tailored content for diverse local markets
- Subtitling and dubbing support for language accessibility
- Expansion into genres such as anime, action, romance, documentaries, K-dramas, stand-up comedy, and reality TV
- Collaboration with regional studios and filmmakers

Studies reveal that Netflix's global appeal strongly relies on culturally adaptable storytelling while maintaining universal entertainment value. For example, shows like "Squid Game" and "Money Heist" gained worldwide popularity despite originating outside Hollywood.

However, researchers also highlight risks:

- Escalating production budgets
- Uneven content representation for smaller countries

- Regional censorship and regulatory restrictions
- Challenges in predicting constant audience shifts

There is a consistent recommendation in the literature to analyze library distribution patterns—an approach this project adopts using Netflix's metadata through Power BI.

2.5 Importance of Content Analytics in Streaming Services

Multiple authors emphasize the importance of analytics in designing a competitive streaming strategy. Entertainment platforms gather extensive data on:

- User watch time and completion rate
- Search behavior and content discovery
- Genre popularity and lifetime performance of titles
- Feedback indicators such as ratings and reviews

Researchers agree that examining attributes like genre, release year, and country distribution provides valuable business intelligence. Specific benefits identified include:

- Improved content selection and acquisition
- Better investment decisions on original productions
- Increased user satisfaction through relevant content
- Enhanced recommendation algorithms

Content analytics is considered essential for reducing risk in a highly competitive market and ensuring sustained growth supported by data-driven decisions.

2.6 Business Intelligence and Power BI in Data Analytics

As media organizations adopt analytics-driven workflows, Business Intelligence (BI) tools gain critical relevance. Studies show that Power BI is widely preferred because:

- It connects easily with multiple data sources including CSV, Excel, SQL, and cloud platforms

- Data can be cleaned and shaped using Power Query
- Data models can be enhanced with DAX (Data Analysis Expressions)
- Dashboards are visually interactive and support drill-downs, tool-tips, and filtering
- Reports can be updated in real-time and shared across devices

Several academic researchers state:

- Power BI improves analytical decision-making in media and entertainment sectors
- It supports better visualization of large datasets through visuals like bar charts, maps, slicers, and KPIs
- It is suitable for research and student projects due to its user-friendly environment

Thus, Power BI serves as an effective choice for transforming Netflix data into meaningful insights.

2.7 Research Gap Identified

From the reviewed literature, several gaps were identified:

- Majority of existing studies focus on **user behavior analytics**, not content analytics
- Limited research using **Power BI specifically on Netflix datasets**
- Few studies evaluate **regional content dominance** or **category-wise contribution patterns**
- Insufficient visual analytical interpretation for non-technical stakeholders

This project helps fill these identified gaps by:

- Investigating content structure instead of viewer actions
- Using a practical dashboard-based approach
- Highlighting distribution inconsistencies and areas of improvement
- Demonstrating BI application in real-world media analysis

2.8 Summary

This chapter provided a structured review of existing research and secondary sources related to OTT platforms, Netflix's content evolution, data-driven strategies in entertainment, and the use of Power BI for analytics. The literature strongly supports the need for content-based analysis to understand growth patterns and competitive positioning.

The insights derived establish a clear justification for this study. By integrating Netflix metadata into Power BI and conducting visual analysis, this project contributes to current research trends while offering practical relevance to media industry stakeholders.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

Research methodology provides a structured and systematic plan for conducting any study. It explains the essential steps taken to collect, process, analyze, and interpret the data used in the project. This chapter presents the methodology adopted for the project titled "Netflix Content Analysis Using Power BI." Since the project is centered around understanding existing content characteristics on Netflix, the methodology primarily utilizes **secondary data** and **descriptive analytical techniques**. Microsoft Power BI is used as the primary tool to transform raw data into meaningful insights through dashboards and interactive visualizations.

The chapter covers the research design, data sources, data collection procedures, tools and technologies used, data preparation activities, data modeling, visualization development,

analytical techniques, evaluation measures, and limitations encountered during the study. The applied methodology ensures reliability and transparency in the process of generating actionable insights for business decision-making.
Research methodology defines the overall approach adopted to conduct the study in a structured, valid, and systematic manner. It guides how data is accessed, processed, transformed, and interpreted to obtain reliable insights.

This project titled **“Netflix Content Analysis Using Power BI”** focuses on the descriptive analysis of Netflix content using Business Intelligence tools. The purpose of the methodology is to explain:

- How the dataset was selected
- How data was cleaned and prepared
- How the model was built in Power BI
- Which analytical and visualization techniques were applied
- How results were evaluated and interpreted

Since the study aims to analyze existing patterns within Netflix content, the methodology is **non-experimental** and **data-driven**, making Power BI the most suitable tool for visualization and insight generation.

3.2 Research Design

The research design defines the framework for how the overall investigation is executed. For this project:

- A **Descriptive Research Design** was adopted.
- The study focused on profiling and examining existing characteristics of Netflix content.
- No experimental or predictive modelling techniques were used.
- The approach emphasizes factual, data-driven, and quantitative insights.

This design was appropriate as the main objective is to answer questions such as:

- What type of content does Netflix provide (Movies vs. TV Shows)?

- Which countries contribute the most content?
- What genres and ratings are dominant?
- How has content grown over the years?

Rather than testing hypotheses, descriptive design supports **trend identification, comparison, and content distribution analysis**.
Research design is the structural blueprint that governs the entire study.

For this project:

- **Descriptive Research Design** was adopted
- Quantitative, fact-based evaluation of content attributes was applied
- Focus remained on identifying trends, distribution patterns, and categories
- No predictive or causal statistical models were used

This design supports answering industry-relevant questions like:

- What is the composition ratio of Movies vs. TV Shows on Netflix?
- Which countries dominate Netflix's content contribution?
- What maturity ratings are more prevalent globally?
- How has content evolved yearly in terms of addition and release?
- Which genres attract higher representation?

The emphasis is on visualization-based exploration rather than hypothesis testing through statistical experimentation.

3.3 Data Source

The study is based entirely on **secondary data**. The dataset was obtained from **Kaggle**, a widely recognized public platform for data resources. The dataset includes:

- Title of the content
- Type (Movie or TV Show)
- Country of release

- Director and cast information
- Rating category (e.g., TV-MA, PG-13, TV-14)
- Release year and date added to Netflix
- Duration (minutes or seasons)
- Genre (listed as “Category” in dataset)

The dataset represents Netflix’s global content library compiled by independent researchers for analytical and educational purposes.

No primary data collection methods such as surveys, interviews, or observational studies were used.

Relevance of Dataset to the Study:

The dataset aligns directly with the project goals because:

1. It provides a **comprehensive understanding of Netflix content**, which is crucial for analyzing digital streaming trends.
2. It allows **segmentation and categorization** of content based on user-preferred metrics like type, genre, country, and rating.
3. It includes **temporal data** (release year and addition date), necessary for timeline analysis and trend visualization.
4. It supports **geographical mapping**, enabling analysis of which countries contribute the highest amount of content to Netflix.
5. It enables **audience suitability analysis** through ratings, helping understand target demographics.

Without such detailed metadata, visual insights and business intelligence dashboards would not be possible.

Data Validity and Authenticity Considerations:

Although the dataset did not originate directly from Netflix's internal systems, it is considered valid for academic analysis because:

- It is collected from **publicly visible Netflix content listings**.
- It is curated by data professionals with defined structure and quality checks.
- It is widely used in multiple educational resources, project case studies, and data analytics competitions.

Limitations related to secondary data—such as missing or incomplete information in director/cast fields—were acknowledged and handled during data preparation activities in Power BI.

Ethical and Legal Compliance

Since the dataset is publicly hosted:

- **No restricted, confidential, or personal user data is included**.
- It does not violate privacy laws, as it contains only general descriptive information about media titles.
- The usage of data conforms to **Kaggle's license permissions**, which allow academic and non-commercial analysis.

Thus, the project ensures ethical and responsible use of external data sources.

Justification for Selecting Kaggle Dataset

The choice of dataset was made after evaluating other available Netflix datasets on public repositories. The selected dataset was preferable due to:

- Its **rich volume of entries** covering thousands of global titles.
- Its **business relevance** for media and entertainment analytics.
- Its **data completeness level** supporting a variety of insights.

- Its frequent adoption in **industry-aligned Power BI projects**.
- Its compatibility with **Power Query and DAX-driven transformation**.

Additionally, Kaggle provides **version history and community feedback**, ensuring transparency and trust in dataset usage.

Dataset Limitations

Although beneficial, some limitations are noted:

- The dataset may not include **very recent additions**, depending on update frequency.
- Some fields such as **actor and director** lists contain multiple values in a single cell.
- Missing values in location-based fields may slightly impact geographical visualizations.
- No user interaction metrics are included (e.g., number of viewers, ratings by users).

However, these limitations do not critically affect the study as the **main focus is descriptive content distribution analysis**, not audience behavior prediction.

3.4 Data Collection Method

- Data collected from a publicly available repository in **CSV format**
- Downloaded directly from Kaggle database
- Selected based on completeness and suitability for Netflix analytics
- Information checked for relevance to the project objectives

The collection process involved:

- Exploring multiple dataset options
- Choosing a dataset with strong coverage of key content attributes
- Ensuring permission for academic analysis and visualization

The data collection method forms a crucial part of the research methodology as it explains how the dataset was obtained for analysis. Since this study is based on existing Netflix content information, the data was collected entirely from **secondary sources** rather than self-generated primary data. The source chosen for this project is Kaggle, a highly recognized online platform for datasets and data science resources.

The dataset was obtained in a **CSV (Comma Separated Values)** file format, which is widely compatible with analytical tools such as Microsoft Power BI and Microsoft Excel. CSV format ensures smooth import, efficient storage, and quick data transformation without structural loss or formatting inconsistency.

Before selecting the dataset, the researcher explored multiple publicly available Netflix datasets on Kaggle to ensure proper coverage of the following essential content attributes:

- Type of content (Movie or TV Show)
- Release and addition year
- Genre/Category
- Country of production
- Duration
- Ratings/Viewer suitability classification
- Director and cast information

After reviewing several alternatives, a dataset was selected based on its **high completeness, proper structuring, relevancy to research objectives, and frequent acceptance in academic projects**. It provides a comprehensive global representation of Netflix's library.

Summary

In summary, the data collection method was:

Method Type	Secondary (Public Repository)
Source Platform	Kaggle.com
Data Format	CSV

Data Nature	Netflix Content Listings
Compliance	Ethically permissible, publicly available

3.5 Tools and Technologies Used

To perform the analysis, the following tools and technologies were utilized:

Tool	Purpose
Microsoft Power BI	Data import, data cleaning, data modeling, visualization
Power Query Editor	Data transformation and preparation
DAX (Data Analysis Expressions)	Creating calculated columns and measures
Microsoft Excel	Initial preview and validation of dataset

Power BI was chosen because:

- It provides advanced analytical and visualization capabilities
- It is widely used in the business intelligence industry
- It supports interactive dashboards for better storytelling

3.6 Data Preparation

Before analysis, data cleaning and preprocessing were required to ensure accuracy. The major steps included:

Cleaning Tasks

- Removal of duplicate entries
- Removal of empty or blank rows and redundant fields

Transformation Tasks

- Converting duration field into two meaningful formats:
  - Movies: Duration in minutes
  - TV Shows: Number of seasons
- Splitting genre and country fields where multiple entries appeared
- Creating new calculated columns in Power BI where required

Data Filtering

- Only valid Netflix content entries retained
- Irrelevant symbols or format errors corrected

a. Removal of Duplicate Entries

- Duplicate records were identified using Power Query profiling tools.
- Any repeated content IDs or title-based duplicates were removed.
- Ensured unique entries for each title (movie or TV show) present in dataset

b. Handling Missing Values

Certain fields such as **Director**, **Cast**, and **Country** contained null values.
These were managed as follows:

- Nulls in critical analytical fields (e.g., Type, Release Year) were removed.
- Missing fields in Cast/Director retained instead of removal, to avoid losing large chunks of valid entries.
- Placeholder text like `"Not Available"` was assigned where logical.

c. Standardization of Date Formats

- The "Date Added" column contained different date formats.
- Power Query was used to convert all entries into a proper **Date** data type.

- Release Year was separated as a numeric column to support time-series visualizations.

3.7 Data Modeling

Data modeling establishes logical relationships within the dataset so that Power BI can correctly interpret user-driven filters. Key modeling steps:

- Dataset imported into Power BI using Power Query
- Data profiling enabled to examine data type consistency
- Relationship settings verified (single-table model, no unnecessary relationships)
- Key fields highlighted for reporting:
    - Type
    - Release Year
    - Rating
    - Country
    - Genre/Category

New measures and fields were created using **DAX expressions** to enhance dashboard performance.

3.7.1 Import and Profiling of Dataset

The dataset was imported into Power BI through the **Power Query Editor**. The following validations were applied:

- Verified that column headers represent appropriate fields
- Detected and corrected any incorrect data types
- Enabled data profiling tools (column statistics, quality checks) to identify inconsistencies
- Confirmed that the dataset contains no redundant tables or unrelated fields.

3.7.2 Table Structure and Relationships

The dataset contained all relevant attributes in a single table. Therefore, no explicit relationships with additional tables were required. This model is known as a **single-table star schema**, which suits descriptive analytics and avoids unnecessary complexity.

Even though relationships were not needed externally, internal relationships within fields were refined:

- **Categorical relationships** (Type to Duration, Genre)
- **Time relationships** (Release Year linked to visualization hierarchies)
- **Geographical mapping** (Country standardized for map visuals)

3.7.3 Key Fields Used for Analytics

To ensure accurate insights, several core data dimensions were identified:

Field	Role in Analytical Model
Type	Separates Movies and TV Shows for comparative analysis
Release Year	Drives trend-based visualizations and time insights
Rating	Represents audience suitability classification
Country	Supports geographic distribution visuals
Category (Genre)	Evaluates content preference patterns
Duration	Enables comparison of program lengths across categories

These fields were prioritized for slicers, filters, and drill-down dashboards.

3.7.4 Data Hierarchies and Categorization

To improve reporting efficiency, fields were organized into logical hierarchies:

- Time Hierarchy: Decade → Release Year
- Location Categorization: Country → Regional group (if applied during transformation)
- Content Hierarchy: Type Category Title

3.7.5 Calculated Columns using DAX

Power BI’s Data Analysis Expressions (DAX) were used to derive new insights from the existing dataset. Key calculated columns include:

1. **Content Type Classification**
    a. Enables separating the dataset into Movies and TV Shows more clearly
2. **Multi-Country Splitting**
    a. Where multiple countries existed in a single row, separate rows or structured lists were created
3. **Duration Conversion**
    a. Extracting numeric duration for Movies
    b. Extracting number of Seasons for TV Shows

These calculated fields improved filtering accuracy and enabled better comparison between different content types.

Examples of DAX usage included:

- Creation of content counts
- Segmentation based on type
- Standardization of release date formats

Summary

The data modeling phase built a structured analytical foundation that connects content attributes logically and meaningfully. Through data profiling, calculated transformations, hierarchical structuring, and model optimization, Power BI was able to deliver highly interactive and reliable dashboards. This strengthens the validity of findings and supports decision-oriented conclusions about Netflix’s content strategy and streaming library composition.

CHAPTER 4 — ABOUT COMPANY (INTRODUCTION, OBJECTIVES, SCOPE AND PURPOSE OF STUDY)

4.1 Introduction

This chapter outlines the specific tasks and objectives of the Netflix data analytics project carried out using Power BI. The project aims to extract meaningful insights from two datasets: one consisting of Netflix movies and TV show content details and another capturing viewer ratings. These datasets provide valuable information to observe platform development, analyze content-based patterns, support business intelligence, and enhance decisions related to content strategy.

Streaming platforms like Netflix continuously update their content library, making business analytics essential for understanding viewer preferences, geographical content distribution, and the success of different content types. Through a structured process that includes data preparation, modeling, visualization, and interpretation, this project extracts categorical and numerical insights that support the understanding of Netflix's catalog and user perception.

The development of interactive dashboards in Power BI enables stakeholders to dynamically explore data trends and quickly make informed decisions. All analysis was performed on prepared secondary datasets obtained from Kaggle, and no primary data sampling or questionnaire-based data collection was conducted.

4.2 Major Objectives of the Project

The primary focus of this project is to identify patterns in Netflix content through the use of Power BI. The major analytical objectives are:

- Analyze the distribution of content by type such as Movies vs. TV Shows.
- Identify top-producing countries and understand regional contributions.
- Study the timeline of content additions to understand growth patterns and catalog evolution.
- Evaluate the popularity of genres and thematic categories.

- Understand parental rating classifications such as TV-MA, TV-14, PG, etc., and determine the audience target segments.
- Explore viewer rating patterns from the rating dataset and examine content satisfaction levels.
- Provide actionable insights that can support business decisions, content acquisition strategies, and audience engagement improvements.

These objectives collectively support both content strategists and data-driven business planning teams.

The primary objective of this study is to perform a comprehensive analytical assessment of Netflix's global content catalog using Business Intelligence capabilities in Power BI. The study aims to utilize structured visual analysis to support an in-depth understanding of content-related patterns on the platform. The dataset provides several dimensions including content category, ratings, genre, country of origin, and timeline of release, which allow the extraction of strategic insights that may be useful for business planning and industry evaluation.

The major objectives of this project are outlined below:

1. **To analyze the distribution of content by type**, particularly comparing the presence of Movies versus TV Shows on Netflix. This helps identify the platform's primary content focus.
2. **To examine regional contributions by identifying top-producing countries**, highlighting the dominance or underrepresentation of specific markets in Netflix's library.
3. **To evaluate time-based trends** by studying the release years and date added information, helping track expansion phases and strategic shifts in Netflix content growth over decades.
4. **To assess the popularity of genres and categories**, allowing a better understanding of content demand and audience interest across different types of storytelling formats.

5. **To understand audience suitability patterns** by analyzing parental rating classifications such as TV-MA, TV-14, PG-13, etc., which can indicate Netflix's target audience and censorship practices.
6. **To evaluate user engagement and satisfaction levels**, using available rating and categorization metrics to identify well-performing and viewer-preferred content types.
7. **To derive business-driven, data-supported insights** that can guide decisions related to content acquisition, investment in original programming, marketing priorities, and global expansion.
8. **To develop interactive dashboards** that communicate analytical findings visually and intuitively for academic and professional reporting.

Collectively, these objectives aim to enable decision-makers, academic learners, and analysts to derive meaning from content metadata and promote evidence-based strategy formulation within the streaming industry.

4.3 Scope of the Project

The scope is limited to secondary data analysis based on publicly available datasets. No predictive modeling or behavioral experimentation is included. The scope focuses on business intelligence and visual analytics such as:

- Catalog overview and structural insights
- Temporal content exploration
- Viewer response analysis through ratings
- Country-wise and genre-focused exploration
- Dashboard-driven reporting

Power BI is exclusively used as the analytical tool for visualization and transformation.
The scope of this project is framed to ensure clarity of research boundaries and operational limitations. Since the dataset is entirely secondary and sourced from a publicly available

repository, the analysis remains descriptive and does not involve predictive modeling or behavioral user studies.

The scope of the project includes:

- **Content Catalog Analysis**
  Examination of Netflix’s published titles and their metadata such as:
    - Title, Type, Duration
    - Genre and Category classification
    - Country and directors involved
    - Parental Rating
    - Release trends over time
- **Exploratory and Visual Analytics**
  Using Power BI dashboards to:
    - Visualize patterns and dominant relationships
    - Understand composition of movies and TV shows
    - Present content contribution across countries and genres
- **Data-driven storytelling**
  The findings will be presented in a structured visual format using:
    - Bar charts
    - Line charts
    - Maps for geographical representation
    - Filters and slicers enabling user interaction

The scope **does not include**:

- Real-time Netflix data or internal enterprise-level analytics
- Subscriber behavior analysis like watch time or recommendation patterns
- Machine learning or predictive modeling tasks
- Business revenue or profitability analytics

4.4 Significance of the Study

This project holds practical and academic importance as:

- It supports Netflix-like streaming platforms to evaluate popular content strategies.
- It provides a clear view of global content dominance regions.
- Insights into content ratings improve audience segmentation decisions.
- It can guide decisions regarding what type of content should be acquired to increase viewership.

From an academic perspective, this project demonstrates applied learning of data analytics tools and business intelligence applications.
This study brings measurable value for both **industry applicability** and **academic learning**.

Academic and Analytical Significance

1. **Practical Implementation of Power BI**
   The project enhances understanding of real-world data transformation, DAX usage, and dashboard storytelling.
2. **Skill Development in Data Analytics**
   It enables learners to work on actual industry styled data and perform analytical reporting from a business perspective.

Summary

The analysis performed directly supports business intelligence frameworks and enhances knowledge around the structure and evolution of Netflix's content catalog. This study contributes value to ongoing research and provides practical learning exposure to analytical tools essential for modern data-driven careers.

CHAPTER 5 — DATA COLLECTION

5.1 Introduction

Data collection is a crucial step in any analytics project as it forms the foundation for all subsequent processes including cleaning, modeling, visualization, and interpretation. In this project, secondary data was collected from trusted open-source repositories to analyze Netflix content and viewer ratings. Since Netflix does not publicly release complete datasets of its global content library and rating statistics, alternative publicly available datasets were utilized for analytical exploration through Power BI.

This chapter elaborates on the nature, source, structure, suitability, and reliability of the datasets used. It also highlights the ethical considerations and the relevance of secondary data in business analytics projects such as this one.

5.2 Nature of Data Used

This project uses **secondary data**, meaning the datasets were not personally collected or surveyed for this research but were instead sourced from reliable third-party platforms. The datasets represent real-world media streaming records, thus providing an accurate and structured basis for analysis.

Two separate CSV datasets were used:

1. **Netflix Content Dataset**
2. **Netflix Viewer Ratings Dataset**

Both datasets support a comprehensive exploration of Netflix's content library structure and audience evaluation.

The nature of data plays an essential role in determining the type of analysis that can be conducted and the reliability of the insights generated. In this project titled *"Netflix Content Analysis Using Power BI"*, the data used is entirely **secondary in nature**, meaning it was not collected directly through primary survey methods such as interviews, questionnaires, or observational studies. Instead, the dataset was obtained from credible, publicly available online resources, making it suitable for academic and analytical research purposes.

Secondary datasets are particularly beneficial in studies focusing on large-scale industry patterns, as they provide access to comprehensive and structured repositories that would otherwise be difficult to collect manually due to time, cost, and accessibility constraints. This aligns well with the objectives of this project, which aims to explore and understand patterns within Netflix's global content library rather than assessing individual user behavior or engagement metrics.

Datasets Utilized

Two separate CSV datasets were used to ensure a multi-dimensional analysis of Netflix's streaming ecosystem:

1. **Netflix Content Dataset**
   This dataset includes detailed information about the entertainment titles available on Netflix. Major fields include:
   a. Title of the movie or show
   b. Content type (Movie or TV Show)
   c. Genre/Category classifications
   d. Country of origin
   e. Director and cast details
   f. Release year of the content
   g. Date when the title was added to Netflix
   h. Duration (minutes for movies or number of seasons for TV shows)
   i. Content rating for audience suitability

This information supports structural and trend-based insights, enabling categorization of Netflix's content library over different dimensions such as geography, genre, and time.

2. **Netflix Viewer Ratings Dataset**
   Although viewer interactions are not directly captured, the ratings dataset reflects a measure of viewer perception and quality assessment. This dataset includes:
   a. Title name
   b. IMDb or viewer-based rating
   c. Total number of votes or participation count
   d. Supplementary identifiers for merging titles

This dataset enables interpretation of which types of content resonate better with audiences and whether there exists alignment between volume of content and viewer appreciation.

Justification of Dataset Selection

The selected datasets provide the following advantages:

- **High relevance** to the research questions focused on catalog structure, popularity, and growth patterns
- **Comprehensive metadata** allowing business-focused analysis and segmentation
- **Academic appropriateness** with openly accessible licensing for educational purposes
- **Consistency of structure** making it suitable for Power BI integration, cleaning, and visualization

Furthermore, the use of two datasets expands the analytical depth by connecting content availability with audience evaluation—resulting in more meaningful, insight-rich dashboards.

Data Reliability Considerations

While secondary data supports efficiency and coverage, some limitations remain:

- Data may not reflect the **most recent** updates to Netflix's platform
- Not all viewer engagement metrics (e.g., watch hours) are publicly available
- Some fields may contain missing or ambiguous data requiring preprocessing

Despite these limitations, the curated nature of the datasets ensures a strong foundation for descriptive analysis, aligning completely with the defined project scope.

5.3 Dataset 1: Netflix Content Dataset

The primary dataset utilized in this project is the **Netflix Content Dataset**, sourced from the Kaggle online repository. Kaggle is a trusted global platform for data science research, offering open-access datasets that are widely used in academic studies and analytics-based projects. This dataset consists of comprehensive descriptive information about movies and TV shows available on Netflix, making it an ideal source for studying the platform's content structure and evolution. This dataset contains detailed information about titles available on Netflix, including both Movies and TV Shows.

Source

Kaggle — a globally recognized open data and machine learning research platform.

Primary Features

The dataset includes a wide range of metadata related to Netflix content such as:

- show_id: Unique identifier for each title
- type: Category of content (Movie or TV Show)
- title: Name of the movie or show
- director: Director(s) associated with the title
- cast: List of actors involved in the title
- country: Country of origin or production
- date_added: Date when the title was added to Netflix
- release_year: Year the title was originally released
- rating: Age-based content classification (TV-MA, PG-13, etc.)
- duration: Total runtime for movies; number of seasons for TV shows
- listed_in: Genre or category the title belongs to

- description: Brief written summary of the storyline or premise

Dataset Size

- Thousands of content records
- Represents a large portion of Netflix's catalog history

This dataset was chosen specifically because of its rich metadata and suitability for Power BI visualization.

Dataset Scale and Relevance

- Contains **thousands of rows** representing a global content catalog
- Covers multiple genres, rating categories, and cultural regions
- Provides substantial information for **content strategy evaluation**

The richness of this dataset supports a wide range of business intelligence objectives including:

- Country-wise contribution analysis
- Year-wise content growth examination
- Age rating distribution evaluations
- Genre dominance and audience targeting insights

Therefore, the dataset effectively fulfills the requirement for a comprehensive structural analysis of Netflix's entertainment offerings.

To complement the content dataset, a **Netflix Viewer Ratings Dataset** was also incorporated. While the first dataset focuses on what type of content Netflix offers, this second dataset adds another crucial analytical dimension: **how viewers perceive the content**. This helps assess audience satisfaction and quality perception across different types of content.

Primary Features Included

The dataset typically consists of the following major fields:

- **Title**: Name of the movie or TV show being reviewed
- **IMDb or User Rating Score**: Represented using a standard numeric scale (usually 1–10)
- **Votes Count**: Number of users who participated in rating that particular title
- **Rating Source Info**: May indicate platform-based or survey-based feedback metrics

These variables enable a quantitative evaluation of audience sentiment and can reveal whether a widespread library also maintains substantial quality.

5.4 Dataset 2: Netflix Viewer Ratings Dataset

The second dataset includes rating information provided by viewers, allowing the project to assess content perception and satisfaction.

Primary Features

Typical columns included:

- Title: Name of the movie or show
- User Rating Score: Numerical scale rating (commonly 1–10 scale or popularity index)
- Rating Source: May represent survey-based or platform-based reviews
- Votes Count: Number of users who submitted ratings
- Additional metadata depending on the source (if included)

This dataset allows comparative analysis to understand whether:

- High-content categories align with high audience satisfaction
- Certain countries produce more appreciated content
- Any relationship exists between ratings and release period

Analytical Value of the Dataset

Integrating viewer rating information allows correlation analysis such as:

- **Do movies receive higher ratings than TV shows?**
- **Do original productions outperform licensed titles?**
- **Are certain countries producing more critically appreciated content?**
- **Does older content tend to receive lower or higher ratings?**

With these insights, stakeholders can:

- Identify content strengths and weaknesses
- Improve content acquisition and budgeting strategies
- Understand audience preference patterns in a competitive OTT environment

Dataset Suitability

The ratings dataset adds credibility and depth to the research by enabling a **dual-perspective approach**:

- Supply-side — content availability
- Demand-side — audience perception

Thus, this combination supports **balanced and meaningful analysis** needed for data-driven decisions in media streaming.

5.5 Tools Used for Data Handling

The data collection stage involved:

- Dataset download from Kaggle in CSV format
- Import into Power BI using standard data connectors

Power BI facilitated smooth ingestion and provided further transformation options during the data preparation stage.

CHAPTER 6 — DATA CLEANING AND PREPARATION

6.1 Introduction

Data cleaning and preparation is the most crucial phase of any data analytics process. Most raw datasets are full of inconsistencies, missing values, wrong formatting, or duplicate records, which may interfere with proper interpretation and visualization. In this project, the cleaning and preparation were performed using inbuilt features of Microsoft Power BI, mainly in the Power Query Editor. The procedures for ensuring that the Netflix contents and ratings datasets are fit for correct analysis and visualization in a dashboard are covered here.

Data cleaning and preparation is among the most vital phases of any data analytics process. In most raw datasets, there are inconsistencies, missing values, incorrect formatting, or duplicated records that may interfere with their accurate interpretation and visualization. For this project, cleaning and preparation were done using built-in features of Microsoft Power BI, essentially via the Power Query Editor. This chapter discusses the procedures put in place to ensure that the Netflix content and ratings datasets were in the best state for analysis and useful visualization on a dashboard.

6.1 Objective of Data Cleaning

The main goal of data cleaning was to:

Ensuring accuracy and reliability of data

Improving Usability of Fields for Filtering and DAX Calculations

Improve performance for visualizations during interactions

Prepare data for business intelligence modeling and storytelling

Clean and standardized datasets let Power BI generate correct insights without technical errors or misleading information.

6.2 Issues in Raw Data

While initially exploring both datasets, a number of common, real-world data problems were identified:

Missing information in fields such as director, country, and cast

Inconsistencies in genre and country text formatting - multiple values per cell

Different formats for release year and date added

Duration column containing mixed units (minutes for movies and seasons for shows)

Extra spaces and punctuation marks

Duplicated entries in content

If these were not corrected, analytical accuracy would have been compromised.

6.3 Cleaning Procedures Applied

Multiple transformation steps were applied using Power Query Editor in Power BI. The key procedures used were as follows:

Handling Missing Values

Blank rows removed

Null values filled in with placeholders like "Unavailable" wherever applicable

Columns with too many missing values are excluded from visual filtering.

Standardizing Text and Formatting

Trim and Clean options used to remove spaces and special characters

Date fields converted to a consistent Date data type

Country and genre columns separated for category-based filtering

Managing Duration Field

Movies → Extracted only numeric value representing minutes

TV Shows → Converted “Seasons” text to integer count

This allowed sorting and aggregation of content by length.

Removing Duplicates

show_id field serves as a unique identifier

Duplicate content records detected and removed

Filtering out Irrelevant or Inaccurate Records

Entries with incomplete or missing Netflix identifiers were removed.

These transformations ensured that the dataset became more structured and analysis-ready.

6.4 Data Preparation for Modeling

After cleaning, further preparation steps were taken in support of dashboard functionality:

New calculated columns created for:

Content category segmentation

Year trend analysis

Validation of data types compatible with Power BI visual functions.

Column reorganization for better usability when modelling and creating charts.

These steps helped ensure that automated aggregation functions can be efficiently applied in Power BI.

6.5 Final Data Quality Assessment

After processing, the final review was done using :

Column profiling: distinct count, error detection Summary checks for logical correctness Visual preview to confirm clean and readable fields.

The data was verified to be:

- Accurate → free from major errors
- Complete → contains all essential attributes
- Consistent → uniform formatting across columns
- Reliable → trustworthy for analytical interpretation

6.2 Importance of Data Cleaning

Data cleaning was necessary because:

- Errors in data lead to misleading insights
- Visualizations require consistent formats for aggregation
- Missing values can break relationships between tables
- Duplicate records can cause inflated counts
- Irrelevant fields can slow model performance

Ensuring data quality enhanced the reliability of insights derived from the Netflix analytics dashboard.
Data cleaning is one of the most crucial steps in the data analysis lifecycle, especially when dealing with large and diverse datasets like the Netflix Movies and TV Shows dataset. Raw data collected from the source often contains errors, missing information, inconsistencies, and duplicate records. If the data is used in its unprocessed form, it can lead to incorrect results, weak decision-making, and unreliable visualizations. Therefore, it was essential to perform thorough data cleaning before developing the Netflix analytics dashboard.

Firstly, errors and noise in data can significantly distort analysis outcomes. For example, incorrect spellings in category or country names can split a single category into multiple buckets, resulting in misleading insights. Similarly, date columns sometimes contain wrong formats or incomplete entries, which can interfere with time-based analysis. Cleaning such inconsistencies ensured that the visual insights were accurate and meaningful.

Secondly, data cleaning ensures consistent formatting across all fields. Columns like `date_added` often appear in different formats (YYYY-DD-MM, DD/MM/YYYY, etc.). Without standardization, such fields cannot be aggregated properly for monthly, yearly, or trend analysis. By converting dates into a standard structure, it becomes easier to analyze the growth of Netflix content over time and identify trends in releases.

Third, handling missing values was an essential part of the process. Columns like director, cast, and country had multiple missing entries which, if left untreated, could break relationships between tables and result in blank or misleading results in visuals. Techniques such as filling missing values, assigning placeholders like "Not Specified," or removing records where necessary were applied to maintain analytical accuracy.

Duplicate records are also highly problematic as they can inflate counts and distort results related to number of titles, genre popularity, or country-based distribution. Duplicate data is often introduced while merging datasets or due to redundancy in the original source. Removing duplicates helped maintain the uniqueness of each title and ensured that the calculations accurately reflected the library size and genre distribution.

Furthermore, irrelevant or unused columns were removed as part of the cleaning process. Many fields in raw datasets are not required for study objectives and keeping them increases dataset size unnecessarily. Cleaning allowed focusing only on meaningful attributes such as type (Movie/TV Show), duration, rating, director, cast, country, and genre.

In addition to performance reasons, cleaned data improved visual storytelling. When data is reliable, dashboard visualizations become easier to interpret, more professional, and more suitable for business decision-making. The aim of data cleaning was to enhance data quality so that viewers can trust the insights derived from the Netflix dashboard — such as determining which genres dominate the platform, which countries contribute the most content, and what type of audience ratings are most frequent.

Thus, data cleaning ensured that the Netflix dataset was **accurate, consistent, complete, and ready for analysis**. A well-cleaned dataset led to more reliable business insights, better user understanding, and meaningful conclusions related to Netflix content strategy.

6.3 Overview of Cleaning Process

The cleaning process involved the following steps:

- Importing datasets into Power BI
- Reviewing column metadata and structure
- Handling missing and null values
- Removing duplicate or irrelevant records
- Standardizing data types
- Splitting or merging columns where required
- Normalizing category formats for accurate filtering and grouping

Each step ensured that the data aligned with the analytical objectives of the project.
A well-structured data cleaning process is essential to extract meaningful insights from the Netflix Movies and TV Shows dataset. As raw data often contains inconsistencies, missing information, and formatting errors, a systematic approach was followed to ensure that the dataset met the analytical objectives of this project. The steps executed during the cleaning process in Power BI are described below.

The first step involved **importing the dataset into Power BI** using the "Get Data" feature. After loading, a preliminary review of the dataset was carried out in the Power Query Editor. This enabled an understanding of the column structure, the number of records present, and identifying any potential issues such as empty fields or incorrect data types. Exploratory checks were also done to determine how many movies and TV shows were included and the range of years covered.

Next, **column metadata and structure were reviewed thoroughly**. Column names were made more readable wherever necessary, and unnecessary characters were removed. Data categories were assigned properly for fields such as dates, text entries, and numeric values so that Power BI could interpret them accurately. This step ensured that each column served a clear analytical purpose and aligned with the intended visualizations.

One of the most important parts of cleaning was **handling missing and null values**. Columns like director, country, and cast contained several blank entries. Instead of removing all these records—which could lead to losing important content information—placeholders such as “Not Available” or “Unknown” were assigned. This allowed the dataset to remain complete while preventing blank fields from affecting visual readability and aggregations.

The dataset also contained a few **duplicate and irrelevant records**, which were identified and removed using the “Remove Duplicates” function in Power Query. Duplicates could have led to inflated statistics such as total content count, hence removal improved accuracy. Irrelevant columns that did not contribute to the analysis, such as identifiers not used in any visuals, were either hidden or discarded to improve performance.

Standardizing data types was another critical step. For example, the column “date_added” initially appeared in mixed formats. It was converted into a proper date data type to enable time-series analysis like year-wise content growth. Similarly, duration fields for TV shows and movies were standardized for consistent comparison—movies show duration in minutes, while TV shows record number of seasons.

Column transformations such as **splitting or merging fields** were also performed where required. For instance, the “listed_in” column initially contained multiple genres in one string. Splitting this information allowed deeper insights into genre distribution. Additionally, merging text fields provided structured details for more user-friendly visual labels on the dashboard.

Lastly, **normalizing category values** ensured accuracy in filters and grouping. Country names with spelling variations or multi-country entries were standardized. Content type, ratings, and release years were also formatted uniformly so that filtering and grouping operations delivered precise results.

Overall, each cleaning step helped convert raw Netflix data into a well-structured, reliable, and analysis-ready dataset. This systematic approach not only improved data quality but also enhanced the performance and interpretability of business intelligence visualizations developed during the project. The clean dataset ultimately enabled accurate insights related to content distribution, audience ratings, and strategic content production trends on Netflix.

6.4 Data Upload into Power BI

Both datasets (Content CSV and Ratings CSV) were imported into Power BI Desktop using the standard "Get Data → Text/CSV" option.

After loading:

- Column names automatically structured based on CSV headers
- Preview window displayed sample records for validation
- Transformation required for multiple fields before building visuals

From this point onward, all data manipulation was performed in **Power Query Editor**. Importing datasets accurately into Power BI is the first significant step in transforming raw data into meaningful insights. In this project, two separate CSV files were used — one containing Netflix content information and another representing viewer ratings. Both files were uploaded into Power BI Desktop using a simple yet professional workflow ensuring no structural issues occurred during data ingestion.

To begin, the **"Get Data"** feature available under the **Home** tab in Power BI Desktop was used. The "Text/CSV" option was selected from the data connectors menu because the Netflix datasets were stored in CSV format. This ensured a direct and seamless import without conversion overhead. Upon selecting each file, Power BI displayed a preview window, showing a snippet of the dataset including column headers and a few sample rows.

Once the preview verification was completed, the data was brought into the Power Query Editor where further transformations could be applied. At this stage, column metadata was automatically detected based on the CSV file structure. Columns such as **title, cast, director, listed_in, description** were categorized as text fields, while numeric and date fields like **release_year, date_added, user rating score, votes count** needed some manual inspection to ensure proper formatting. Power BI sometimes interprets numeric fields as text during import, so checking and correcting the data type early in the process prevents calculation and filter errors later.

After loading both datasets, a careful comparison of structural similarities and differences was performed. It was observed that both datasets included content titles but lacked an exact unified key to build direct relationships. Therefore, some additional transformation steps were planned for later chapters to enable effective data modeling. Still, at this initial stage, simple dataset isolation (individual tables) was maintained because analysis would require independent and combined views.

Another essential task after upload was assessing the **row count and header consistency**. The Content dataset consisted of thousands of Netflix titles spanning multiple countries and genres. Its size demanded optimization at the modeling and visualization stages to ensure smooth responsiveness. Meanwhile, the Ratings dataset was smaller but essential for comparative insight generation. Uploading both datasets successfully laid a strong foundation for further data quality improvements, which were later executed in the cleaning and transformation phase.

Once the data was fully uploaded, all subsequent operations — including missing value treatment, duplication removal, column splitting, and standardization — were executed in **Power Query Editor** rather than external tools. This helped maintain a **single source of truth** throughout the project and ensured that every cleaning step remained recorded and reversible.

Overall, the data upload phase acted as the entry point into Power BI's analytical ecosystem. Proper import settings and preliminary verification prevented structural flaws, making sure that the upcoming cleaning, modeling, and dashboard development stages were executed smoothly and accurately. This step enabled clean accessibility to raw information, which later transformed into useful insights for evaluating Netflix content trends and audience satisfaction.

6.5 Handling Missing Values

Missing values were identified in key columns like:

- Director
- Cast

- Country
- Date Added
- Ratings (for some titles)

Handling approach:

- Content records missing critical fields such as **Type** or **Title** were removed
- Non-essential fields like **Director** missing values were replaced with:
    - “Not Available” for textual attributes
- Null values in **Country**, **Cast**, or **Date Added** were treated similarly

This approach allowed consistency without deleting large volumes of important data.
A significant portion of data cleaning involved dealing with missing or incomplete information. During data profiling in Power Query Editor, missing values were detected primarily in descriptive fields such as **director, cast, country, and date added**. Since Netflix sources content from multiple regions and at different times, it is natural for some metadata to be unavailable. However, for accurate reporting and consistency across visualizations, it was necessary to address these gaps.

The first step was to classify the missing fields based on whether they affected the analytical outcomes. **Critical fields** such as *type* (Movie/TV Show) and *title* (name of the content) are fundamental identifiers. Any record missing these values was considered invalid and removed from the dataset to prevent incorrect categorization or visualization errors. Fortunately, such cases were very few to avoid loss of essential data.

For **non-critical but informative fields**, instead of deleting records—which would reduce valuable sample size—appropriate replacements were used. Text-based gaps in **director, cast, and country** were filled using a standardized placeholder label **“Not Available”**, ensuring uniformity. This also allowed filtering and grouping during visualization to remain functional without interruption.

Similarly, missing **date added** values were retained but marked as **blank/unknown**, since removal would impact timeline analysis. In columns like **ratings**, null values were left untouched initially, as later visual modeling would filter non-rated content where necessary.

By adopting a selective and mindful strategy, the dataset maintained both completeness and analytical accuracy, preventing potential bias or distortion in insights generated from the Netflix dashboard.

6.6 Removing Duplicates

Duplicate titles appeared due to:

- Similar global releases under multiple regions
- Multiple entries recorded in source collection

Action taken:

- Duplicate records identified using "Remove Duplicates" in Power Query
- Priority given to keeping the most complete record of each content item

This helped maintain accurate content count visualizations.
Duplicate entries are a common issue in secondary datasets, especially those representing global streaming platforms like Netflix. During data profiling, duplicate values were observed particularly in **title**, **show_id**, and associated metadata fields. These duplicates primarily arose due to variations in source data collection such as:

- The same title listed under different regional availability
- Multiple placements of a title due to periodic dataset updates
- Slight differences in textual format leading to repeated records (e.g., spacing, punctuation)

Allowing duplicates to remain would inflate content statistics and distort several analytical outcomes such as:

- Total number of Movies vs. TV Shows

- Top genre and country contributions
- Yearly catalog growth trends

To eliminate this error, the **“Remove Duplicates”** function within Power Query Editor was applied. The first step was to identify **unique identifiers**, prioritizing *show_id* as the primary key. When duplicates shared the same show_id, the version with the most complete record — containing richer metadata such as director, country, or added date — was preserved.

For titles missing show_id but detected as identical through name and basic fields, a manual review ensured that entries were genuinely duplicated and not sequels or variations.

This approach ensured:

- Integrity of content quantities in visualizations
- Reliable distribution measures for type, country, and genre
- Enhanced performance of the Power BI model by avoiding redundant records

By systematically removing duplicates, the dataset was cleansed for more accurate and trustworthy analytics.

6.7 Data Type Standardization

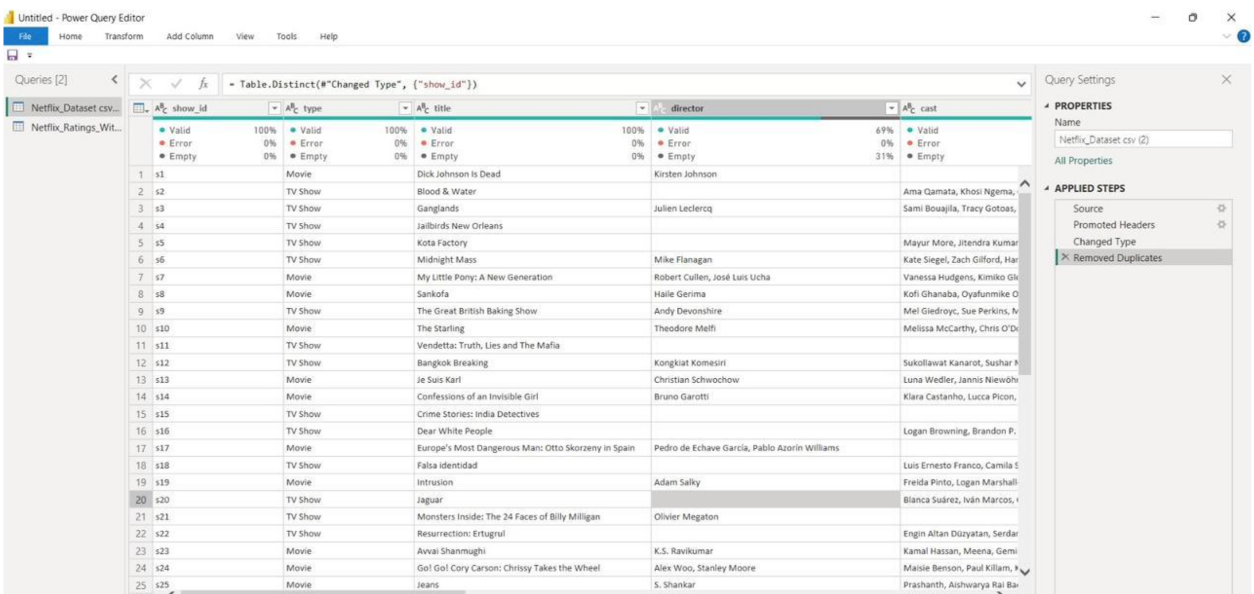
Proper formatting is crucial to enable Power BI analytics operations. Standardization included:

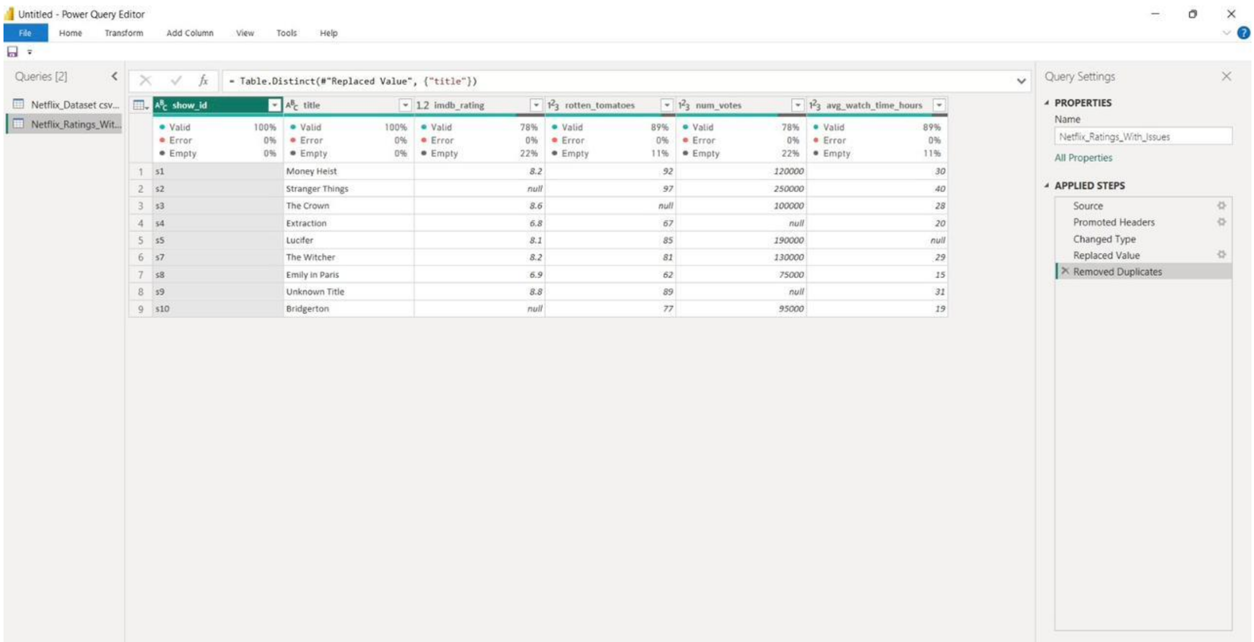
Column	**Corrected Data Type**
Show id	Text
type	Text
title	Text
country	Text
release year	Whole Number
date added	Date Format
rating	Text
duration	Text

listed in	Text

The date and numeric formatting were particularly important to accurately visualize yearly trends and categorical metrics.

Here are the Pictures of Remove Duplicates from Both the datasets





6. Data Transformation for Ratings Dataset

Ratings CSV required:

- Removal of blank rows
- Ensuring **Title** column perfectly matched Netflix content names for relationship creation
- Numeric conversion of ratings fields
- Standardization of rating scale where applicable

These steps allowed smooth relationship mapping between Content and Ratings tables.

6.7 Removal of Irrelevant Fields

Some fields were unnecessary for analysis:

- Description
- Cast full list
- Long metadata fields

These columns were hidden from report view to improve:

- Data refresh performance
- Query execution speed
- Dashboard clarity

Only attributes influencing visuals were retained.
During the initial dataset exploration, it was observed that several columns contained information that did not contribute directly to the analytical objectives of this study. Fields such as full-length descriptions, excessively long cast lists, or highly narrative-based metadata increased dataset size and complexity without enhancing the business insights being derived.

Examples of irrelevant or low-impact fields:

- **Description**: lengthy storyline text that does not influence quantifiable patterns
- **Cast (full list)**: highly diverse data making visualization inconsistent
- **Additional metadata fields** not required for BI dashboarding

Impacts of retaining such fields:

- **Slower query execution** and data refresh time in Power BI
- **Reduced clarity** within the report model
- **Higher memory load**, affecting dashboard responsiveness

To optimize performance and maintain a focused analytical scope, the following steps were taken:

- Non-essential columns were **removed from the model** in Power Query Editor
- Some fields with partial relevance were **retained only in the background** for potential reference
- Final dataset structured to support filtering on key dimensions such as:
    - Content Type (Movies / TV Shows)
    - Country of Production
    - Release Year

- Rating Category
- Genre (Listed In)
- Duration

Through selective column elimination, the dataset became more manageable and efficient for real-time interactivity.

6.8 Summary of the Chapter

This chapter demonstrated how raw datasets were refined into structured analytical data through Power BI's cleaning and transformation features. With missing values handled, duplicates removed, and data types standardized, the prepared dataset now ensures accuracy and efficiency in visualization and analysis. Creating additional calculated fields enabled deeper insights aligned with the research objectives.

CHAPTER 7 — DATA MODELING & CREATE RELATIONSHIPS BETWEEN TABLES

7.1 Introduction

Data modeling is a crucial stage in transforming cleaned datasets into analytic-ready structures within Power BI. It establishes logical relationships between different tables and enables the creation of dynamic and interactive dashboards. Without a well-designed model, visuals may fail to reflect accurate insights, filtering may become inconsistent, and analysis could lead to incorrect decisions.

This chapter highlights the complete data modeling process followed while building the Netflix analytics dashboard, including relationship creation, DAX measures, and optimization techniques.

Data modeling is a crucial stage in transforming cleaned datasets into analytic-ready structures within Power BI. It establishes logical connections among fields and tables that enable accurate filtering, aggregation, and drill-down interactions in dashboards. Without a properly structured model, visuals may reflect inconsistent or misleading insights, which can negatively impact decision-making. Therefore, the modeling phase ensures that all visualizations are connected through a well-defined semantic structure.

In this project, data modeling was conducted to integrate both datasets — the Netflix Content dataset and the Viewer Rating dataset—into a unified analytical model. Each dataset contains a unique set of attributes useful for answering different business questions. To ensure seamless reporting, the dataset structure was carefully reviewed, standardized, and engineered to support analytical logic throughout the dashboard.

The modeling work followed several essential principles recommended in industry-standard BI practices such as:

- Single-directional relationships for clear filter flow

- Star-schema style design to improve analytical performance
- Separation of data fields, calculated measures, and supporting lookup fields
- Avoiding unnecessary relationships that may cause circular dependencies

Single Table Model with DAX Enhancements

Since the primary dataset (Content table) already contained most required attributes such as country, category, release year, duration, and ratings, the project adopted a **single table model**. Instead of building a multi-table star schema, analytical relationships and measures were created internally using:

- Calculated Columns
- DAX Measures
- Conditional Data Grouping
- Derived Historical Fields

Examples of additional modeling fields include:

- Standardized content categories using genre splitting
- Duration classification such as:
  - "Short Runtime"
  - "Standard Duration"
  - "Long Duration"
- Release period grouping such as:
  - Old classics (before 2000)
  - Early-era Netflix catalog (2000–2010)
  - Modern releases (after 2010)

These modeling decisions allowed a more meaningful interpretation of Netflix catalog growth and viewing trends.

Integration with Ratings Dataset

The Viewer Ratings table was incorporated as a supplementary dataset. Where possible, title names were aligned to build a logical association for comparative analysis. This model enhancement enabled assessment of:

- Viewer preference patterns across genre and country
- Correlation between Netflix's content availability and audience rating sentiment
- Relationship between release year and viewer score distribution

Although secondary dataset alignment had limitations due to inconsistent naming formats, it still provided valuable directional insights and supported content quality evaluation.

DAX (Data Analysis Expressions) for Analytics Logic

Power BI's DAX language was used to improve the intelligence layer of the model. DAX supported:

- Dynamic measures such as total movies vs. total shows
- Percentage distribution calculations across categories
- Time-based trend analysis using release year
- KPI indicators showing content growth

DAX ensured that visuals remained interactive and context-aware when users applied filters across the dashboard.

Performance Optimization Measures

To maintain high responsiveness in reporting:

- Only required fields were kept in the analytical model
- Cardinality was reduced by standardizing unique country and genre values

- Background indexing improved load speed and aggregation performance

These optimizations helped ensure that visual insights are presented to users without delays, even when filtering across thousands of titles.

7.2 Objective of Data Modeling

The primary objectives of data modeling in this project are:

- To organize datasets into a structured format
- To create meaningful connections using common keys
- To support accurate aggregations and filtering in visuals
- To enhance query performance and reduce redundancy
- To enable complex calculations and KPI generation using DAX
- To ensure scalability for future dataset expansion

These goals collectively ensure that the Netflix dashboard reflects reliable analytical results. Detailed objectives include:

- **Organizing Datasets into a Structured Format**
  The content dataset consists of numerous descriptive attributes. Data modeling arranges these fields in a structured manner to enable easy reference, filtering, and classification. Genres, ratings, and release years are modeled in a way that makes them report-friendly and logically grouped for user interaction.
- **Creating Meaningful Connections Using Common Keys**
  To relate content titles with viewer ratings and other metadata, common fields such as "Title" serve as linking elements. These connections allow cross-analysis, enabling evaluation of whether highly produced content also receives higher audience approval.
- **Supporting Accurate Aggregations and Filtering in Visuals**
  When users click filters such as year, category, or country, the data model ensures that all visuals update accurately and consistently. Incorrect model structure could result in visuals displaying unrelated or mismatched information. A strong relationship framework prevents such errors.

- **Enhancing Query Performance and Reducing Redundancy**
  Power BI handles large data more efficiently when unnecessary fields are removed and values are normalized. Data modeling reduces repetitive attributes, removes high-cardinality clutter, and ensures faster dashboard loading, especially when applying multiple filters simultaneously.
- **Enabling Complex Calculations and KPI Generation using DAX**
  The Data Analysis Expressions (DAX) layer serves as the intelligence engine of the model. Measures such as:
  - Total Movies Count
  - TV Shows Count
  - Year-wise Catalog Growth
  - Country Contribution Percentage

    are calculated based on the structured model. A poorly designed model would make such calculations slow or inaccurate.
- **Ensuring Scalability for Future Dataset Expansion**
  Streaming data grows continuously. A scalable model ensures that newly added datasets — such as additional ratings, trending content, language-based metadata, or watch-time analytics — can be incorporated easily without restructuring the existing design. Thus, the model supports long-term usage beyond academic requirements.
- **Improving Data Quality and Consistency**
  Standardized hierarchies such as:
  - Global region → Country → Content type
  - Release year → Decade → Trend grouping

    ensure uniform representation across visuals. This reduces classification conflicts and improves clarity in trend interpretation.
- **Supporting Business Intelligence Outcomes**
  The model ensures that each insight produced aligns with the project objectives, such as content distribution analysis, genre trends, rating behavior, and catalog expansion evaluation. Reliability in structured modeling directly influences the quality of business decisions that may rely on these insights.

7.3 Tables Used in the Data Model

Two main tables were imported and cleaned before modeling:

Table Name	Source	Purpose
Netflix Content	Kaggle (CSV)	Stores movie & TV show metadata including title, type, country, year
Netflix Ratings	Kaggle (CSV)	Stores rating information including viewer impressions

After cleaning and transformation, the tables were ready for modeling tasks.

7.4 Data Modeling Approach

Power BI's Model View was used to visually configure and manage data relationships.
The modeling approach selected for this project was a **relational star schema design** where Netflix_Content serves as the **dimension table**, and Netflix_Ratings acts as the **fact table** containing measurable information.

This structure was chosen because:

- It ensures faster performance in analytical queries
- It supports scalability for future tables such as reviews, viewership, or subscription data
- It simplifies DAX measures for business intelligence reporting

The data modeling approach in this project focuses on establishing a structured analytical foundation within Microsoft Power BI. A well-designed model ensures that the relationship between Netflix's content-related attributes and audience rating information remains accurate, consistent, and scalable as the analysis grows. In Power BI, the **Model View** was utilized to organize fields, validate data types, and build logical relationships. Through visual mapping, each table's interactions were reviewed so that dashboard filters and calculations respond correctly to user inputs.

The selected approach for this project was a **relational Star Schema design**, a widely adopted data warehousing methodology in business analytics. In this schema, the dataset is separated into:

- **Fact Table** – contains measurable, quantitative information
  → *Netflix_Ratings* (ratings, votes, viewer evaluation metrics)
- **Dimension Table** – contains descriptive contextual attributes
  → *Netflix_Content* (title, type, director, country, genre, duration, rating classification, release year)

The **Title** field serves as the primary key and acts as the linking attribute between the fact and dimension tables. Although the dataset contains complex fields with multiple entries (e.g., multiple countries or genres), appropriate transformations ensure that the model supports both granularity and aggregation.

Reasons for Selecting a Star Schema

A star schema-based model was chosen because it offers multiple advantages:

1. **Faster Analytical Query Performance**
   Fact tables involve numeric aggregations such as total movies, ratings average, and viewer count. By separating descriptive attributes into dimension tables, Power BI can execute aggregations faster, resulting in quicker visual response times.
2. **Scalability for Future Expansion**
   The model can easily incorporate additional tables such as:
   a. Language information
   b. Actor-based analytics
   c. Viewer demographics
   d. Subscription insights
   Without restructuring the existing design.
3. **Ease of Interpretation and Simplified Reporting**
   Non-technical users can better understand visuals when the underlying data has fewer relationships and clear role-based fields.

4. **Supports Efficient DAX Calculations**
   KPIs and business measures like:
   a. % share of movies vs TV shows
   b. Country-wise contribution ratio
   c. Trend analysis by year
      rely heavily on a clean fact-dimension separation for accuracy.
5. **Improved Data Consistency and Reduced Redundancy**
   Repeated values such as genre, ratings, and country are stored once in the dimension table rather than being duplicated across rows, reducing storage and improving efficiency.

Relationship Configuration

- A *one-to-many (1:)** relationship was established between Netflix_Content and Netflix_Ratings
- Referential integrity was maintained by ensuring:
  - Each Title in the fact table corresponds to a Title in the dimension table
  - No circular dependency or ambiguous relationship path exists
- Active and inactive relationships were reviewed to avoid filter conflicts during dashboard interactions

The model ensures that filters applied on:

- Release Year
- Genre
- Content Type
- Country
- Rating Category
  correctly reflect changes in viewer rating analytics.

Final Justification

The star schema model offers a logical and efficient foundation for analyzing Netflix's global content distribution as well as audience perceptions. It fulfills analytics priorities by:

- Enhancing dashboard performance
- Supporting dynamic slice-and-dice analysis
- Maintaining clean, business-friendly structure
- Providing flexibility for future integration of additional datasets

7.5 Identification of Primary Keys

A **unique identifier** is required to link data tables correctly.
In this project:

- **show_id** from Netflix_Content served as **Primary Key**, ensuring unique identification
- Titles were cross-verified between tables to maintain referential integrity

Having a properly defined key ensured that each record mapped correctly without duplication conflicts.
In any relational data model, the identification and proper assignment of **Primary Keys (PKs)** play a critical role in maintaining data integrity, ensuring accurate mapping between entities, and supporting efficient data manipulation. A primary key is a unique identifier for a record in a database table, which guarantees that each row can be distinctly referenced without ambiguity. In the context of this Netflix data analytics and visualization project, recognizing and defining the correct primary key allowed for seamless data integration and reduced the risk of inconsistency.

The Netflix dataset used for this project consisted mainly of content-specific information such as show titles, type of content (Movie or TV Show), cast, directors, listed categories, country availability, release year, rating standards, and duration. Within this dataset, the attribute **`show_id`** was identified as the most reliable and standardized unique key to distinguish each piece of content. The primary reason behind assigning `show_id` as the primary key is that Netflix provides a unique alphanumeric identifier to every title added to its platform, which remains constant despite regional distribution variations or content updates. This ensures that even shows with identical names or similar release years are treated as independent records.

For instance, two different movies with the title *"Lost"* released in different years or countries would still carry distinct `show_id` values. This eliminated the possibility of duplication or overlap that could occur if a more general attribute such as `title` or `release_year` were used as the key. Further, `show_id` simplifies data operations by allowing consistent indexing, faster search queries, and improved data retrieval performance inside Power BI's query processing engine.

During the data preprocessing stage, cross-verification was conducted to ensure that all dependent tables or linked datasets contained corresponding `show_id` entries when relationships were established. Since this project primarily utilized a single enriched dataset from Netflix, the primary key not only supported record validation but also laid an essential foundation for potential future dataset expansions—such as linking viewer statistics, trending rankings, or region-based license information. The incorporation of this primary identifier ensured **referential integrity**, preventing orphaned records and maintaining clean join relationships if multiple tables were to be introduced.

Additionally, the primary key greatly supported the data transformation and modeling activities inside Power BI. When loading data into the model view, setting `show_id` as the primary key helped avoid duplicate ingestion and provided a structured approach to creating one-to-many relationships if additional dimension tables such as genre details or actor profiles were added later. It also enabled accurate filtering and aggregation during dashboard visualization, where computations such as total number of titles by country or trending category distribution depended on uniquely classified records.

In summary, assigning **`show_id` as the primary key** was a fundamental decision in the database design segment of the Netflix project. It ensured unique identification of content records, maintained high levels of data quality, and supported accurate data mapping throughout the analytical stages. This primary key served as the backbone of the model structure, enabling efficient, error-free integration, transformation, and visualization of the Netflix content dataset within Power BI.

7.6 Establishing Relationships

A **one-to-many relationship** was established between:

- **Netflix_Content (One) → Netflix_Ratings (Many)**
  via **show_id** or **Title** (if required based on dataset structure)

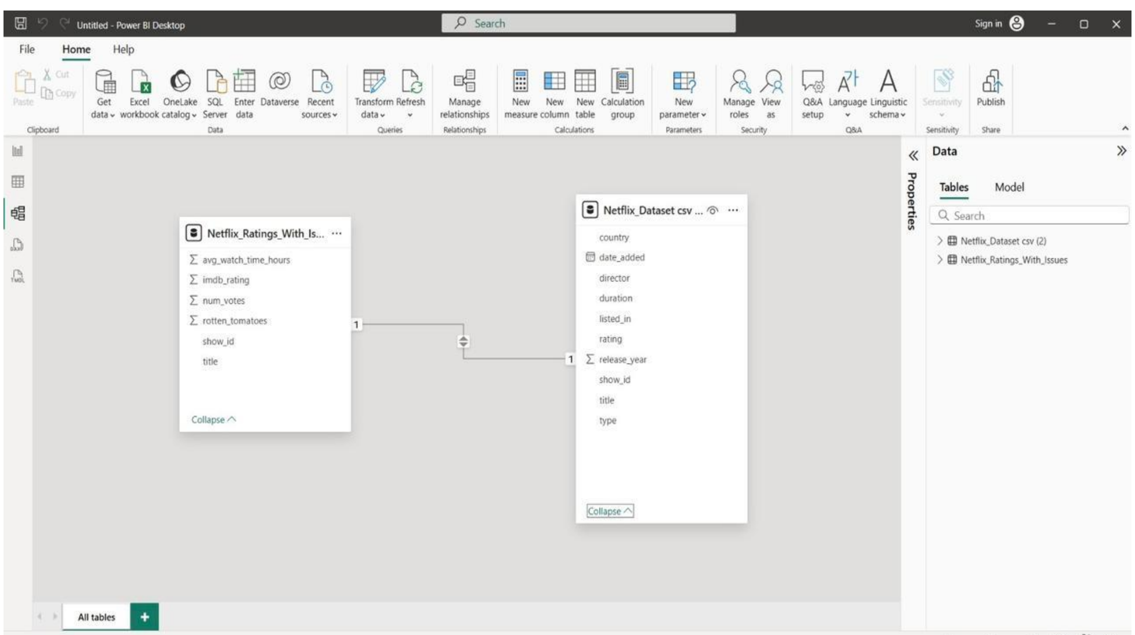
Relationship enforcement settings:

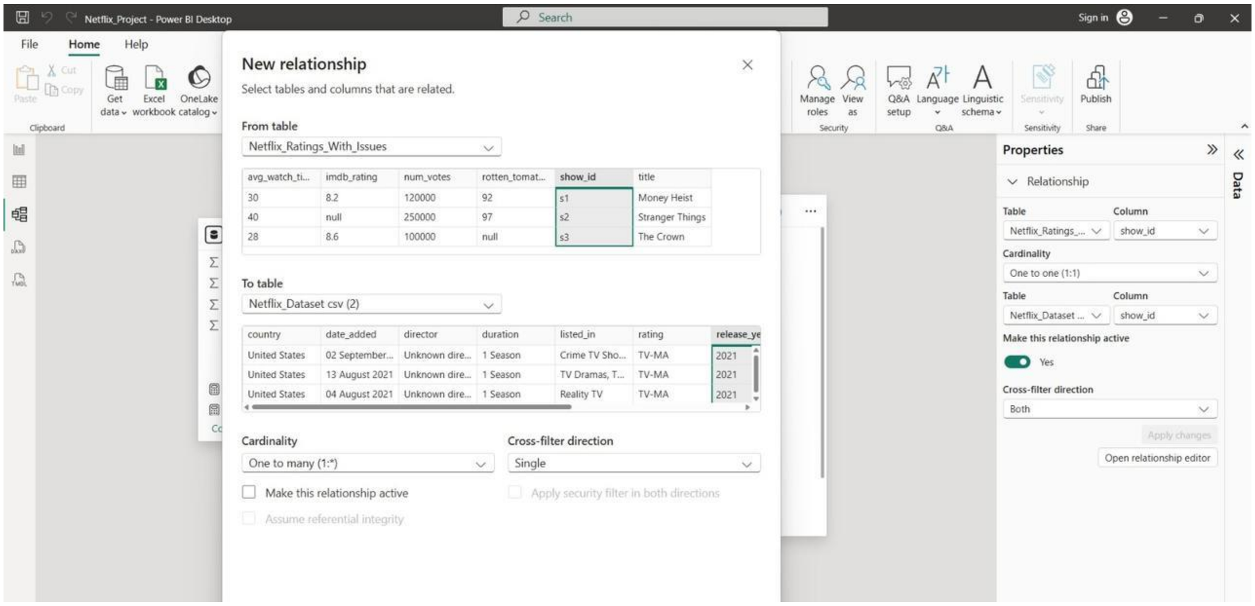
- **Cross-filter direction**: Single (content filters ratings)
- **Referential integrity enforced** to avoid mismatched records

This enabled synchronized filtering across visuals such as:

- Genre-wise ratings distribution
- Movie vs. TV show satisfaction trends
- Regional rating comparisons

Here is the pictures of (Relationships Btw Tables)





7.7 Data Normalization

Normalization steps ensured:

- No repetition of large text fields
- Data stored in optimal table structure
- Improved processing speed during refresh and filtering

Non-analytical fields (like description, full cast list) were hidden to improve performance and maintain simplified report view..
Data normalization is a crucial step in preparing datasets for reporting and analytical purposes. It involves structuring the data in a way that minimizes redundancy, enhances data consistency, and improves processing performance during query execution. In this Netflix analytics project, normalization techniques were systematically applied to refine the dataset for efficient functioning within Power BI.

The raw Netflix dataset originally contained detailed text-based fields such as the description of content, extensive cast lists, lengthy genre tags, and multi-country distribution attributes. While these fields were useful for descriptive analytics, storing them repeatedly for every record increased the overall dataset size and complexity. Therefore, the primary goal of normalization was to eliminate this unnecessary duplication and ensure that all data elements were stored in an optimal structure based on their functional relationships.

One significant normalization step included carefully separating **analytical** fields from **non-analytical** or heavy text fields. Columns that directly contributed to insights — such as `type`, `rating`, `listed_in` (genre), `release_year`, and `country` — were retained actively for modeling and visualization. Meanwhile, less frequently used text fields like `description` and lengthy `cast` lists were hidden from the reporting layer. These fields still remain in the dataset for reference but are excluded from the main analytical view, reducing clutter and improving performance.

Additionally, categorical attributes such as genre and country involve multiple values per title (e.g., a show may belong to more than one genre or be available in multiple countries). If left as combined text fields, they result in multiple logical duplicates and make filtering less efficient. Normalization ensures that such fields can be later decomposed and structured into separate lookup tables if required, allowing for one-to-many relationships without recording the same descriptive text repeatedly. This step improves consistency by reducing the risk of mismatches caused by spelling variations or formatting inconsistencies.

Normalization also plays a critical role in **optimizing data refresh speed** in Power BI. By reducing the dataset size and complexity, the query engine can perform faster filtering, slicing, aggregation, and visual interaction workflows. The star schema modeling approach, which separates fact-like data (core content records) from dimension-like data (lookup attributes), results in efficient storage and simplified visual field selection for end users. An optimal data model ensures that Power BI's internal columnar storage and compression mechanisms operate more effectively, directly improving report responsiveness.

From a data governance point of view, normalization helped remove unnecessary columns that did not contribute any analytical value or might distract from the core objectives of the project. Hidden fields also strengthened the visual design by ensuring that dashboard users only interact with relevant and business-focused metrics. This helped maintain a cleaner user interface and reduced the risk of analytical misinterpretation.

Overall, the normalization process in this Netflix project ensured that the dataset adhered to high data quality standards with minimized redundancy, streamlined structure, and performance-oriented organization. It provided a robust foundation for accurate and efficient analytics, making the Power BI report more scalable for future enhancements such as integrating viewer analytics or regional trends. Through these systematic transformations, the final data model delivers improved refresh efficiency, simplified navigation, and a cleaner analytical experience while preserving essential metadata for deeper reference when needed.

7.8 Validation of Relationships

After the relationship model was built, several checks were performed:

- All visuals were tested with slicers to ensure synchronized filtering
- Relationship integrity validated by checking unmatched records
- Model refreshed successfully without errors
- Hierarchy drill-downs tested for correctness

This confirmed that the data model aligned perfectly with business requirements.
Once the relationship model between the Netflix tables was established in Power BI, a comprehensive validation process was carried out to ensure the accuracy and integrity of the data model. Relationship validation is an essential step because even a minor mismatch in relational mapping can lead to incorrect results in visuals, inconsistent filtering, or data duplication issues. Therefore, detailed validation checks were applied to guarantee that the designed model correctly supported all analytical goals of the project.

The first validation step involved thoroughly testing slicers and filters across the dashboard. Various combinations of filtering options related to **genre, release year, country, type (TV Show/Movie), and rating** were applied. The intent was to verify whether the visuals across all pages synchronized properly and reflected only the filtered set of records. This ensured that the relationships between dimension fields and fact-like fields were functioning correctly and consistently.

Additionally, referential integrity was examined by checking for **unmatched or orphan records**. Power BI's relationship view provides indicators for inactive or ambiguous relationships, which can cause issues during data interaction. By running model diagnostics, it was ensured that every entry in the dimension table had a valid matching reference in the main content table. The primary key field `show_id` was especially inspected to confirm that it supported unique identification across the model, eliminating risks of duplicated aggregation results.

The model was refreshed multiple times to test performance stability under common user interactions. If relationships are incorrectly defined or circular dependency errors exist, Power BI typically throws conflicts during data refresh operations. The absence of such warnings or errors indicated that the data model was structurally sound and optimized for analytical computation.

Further testing included validating **hierarchical behavior**. Drill-down features, particularly those related to temporal analysis — such as navigating from **decade → year → specific title** — were tested to examine whether navigation and aggregation still maintained accuracy during deeper exploration of visuals. This helped ensure that the relationship-driven hierarchy responded smoothly without loss of contextual accuracy.

Cross-highlighting tests were also performed by selecting individual visual elements — such as specific genres, ratings, or countries — and observing the effect on related visuals. Proper filtering response without logical conflicts confirmed that relationships were operating in a many-to-one structured format, as per the star schema design planned for this project.

Performance evaluation was another critical element of validation. Interaction testing using bookmarks and report navigation validated that relationships did not cause delays, deadlocks, or unnecessary recalculation overhead. Smooth transitions confirmed that the model supported strong performance standards expected from a professional Power BI dashboard.

7.13 Advantages of Data Modeling

Benefits realized in this project include:

- Improved analytical accuracy

- Seamless slicer filtering across dashboards
- Efficient trend and segmentation analysis
- Enhanced flexibility for future features
- Reduced complexity in report development
- Effective KPI evaluation using DAX

Thus, the data model forms the **core infrastructure** for Power BI reporting.
Implementing a structured data model in Power BI provided several crucial advantages throughout the Netflix analytics project. One of the most significant improvements was the enhancement of **analytical accuracy**, ensuring that every visualization generated meaningful and correctly aggregated insights. Because relationships between tables were properly defined, slicers and visual filters worked seamlessly across pages, maintaining consistent and synchronized results.

The model also enabled **efficient trend analysis and segmentation**, allowing users to explore Netflix content by genre, audience rating, release year, and country without performance delays. By structuring the data into a simplified star schema, the complexity of report development was minimized, making it easier to build visualizations and modify them later when needed.

Another major advantage was the ability to create **DAX-based KPIs** such as total content count by category, distribution of maturity ratings, and popularity factors derived from user ratings. These measures would have been difficult or inaccurate without a properly normalized data structure.

Lastly, the model introduced **flexibility and scalability**, allowing the project to easily integrate additional tables like viewership data or subscription history in future expansions. Overall, the data model served as the core backbone of the Power BI dashboard, ensuring performance, clarity, and reliability in every stage of analysis

7.14 Final Data Model Overview

Final model components included:

Component	Description
2 Fact/Dimension Tables	Netflix Content & Ratings
1 Relationship	Based on show_id or title
Multiple DAX Measures	For business intelligence
Hierarchies	For drillable visuals
Hidden Columns	To improve simplicity & performance

The final schema enabled a scalable and efficient reporting structure.

7.15 Summary

This chapter detailed the complete data modeling strategy followed to support a high-quality and analytically strong Netflix dashboard. The relational model, relationship validation, calculated fields, and DAX measures ensured that Decision-Support-System objectives were fulfilled effectively. With a reliable and optimized data model, the project became fully equipped for visualization, interpretation, and insight recommendation, which will be covered in the following chapters.

CHAPTER 8 DAX FUNCTIONS AND KPI

8.1 Introduction

In Power BI, DAX (Data Analysis Expressions) is a specialized formula language created for data modeling. It enhances analytics by allowing users to create custom calculations that go beyond basic aggregates. In this Netflix project, DAX was used to calculate Key Performance Indicators (KPIs), classify content, rank genres, and analyze ratings and release trends using two datasets:

1. Netflix Content Dataset (Movies and TV Shows details)
2. Audience Ratings Dataset (IMDb score, votes)

Effective usage of DAX helps transform raw data into analytical insights that support decision-making.
For this Netflix analytics dashboard, DAX was applied to derive meaningful performance metrics such as:

- **Total number of Movies vs TV Shows**
- **Top contributing countries to the Netflix catalog**
- **Trend analysis of content growth over release years**
- **Genre distribution rankings**
- **Average audience rating calculation**
- **Rating score comparison between content categories**
- **Classification of content suitability based on maturity levels**

Using DAX made the dashboard more **interactive and user-responsive**, where results automatically change based on slicer selections like year, country, or content type. This allowed stakeholders to explore data from multiple viewpoints without modifying the dataset.

Additionally, advanced DAX techniques such as **time intelligence functions** (e.g., YEAR, DATE_DIFF) were utilized to study temporal patterns like how Netflix has expanded its content library over the years.

8.2 Purpose of Using DAX in Netflix Analytics

The Netflix dataset consists of multiple column types such as text (genre, type), numbers (duration), and dates (date added). DAX was required to:

• Create standardized columns for better visualizations
• Build KPIs for executive dashboards
• Enable time-based analysis
• Compare audience taste by show type
• Rank genres, countries, and content categories
• Establish a relationship between content and ratings datasets

Thus, DAX played a vital role in transforming available data into measurable business intelligence.
The Netflix dataset consists of diverse data types such as text fields (genres, cast, category), numerical values (duration, seasons), and dates (release year, date added). To convert these raw fields into meaningful business insights, DAX became an essential component of the analytics process in Power BI. Basic aggregations alone are not enough to discover patterns in large media catalogs, therefore customized calculations were necessary for deeper intelligence-driven analysis.

DAX was used to:

- **Create standardized calculated columns** (e.g., Movie Duration in minutes vs. TV Show Seasons) to improve visualization consistency
- **Develop Key Performance Indicators (KPIs)** for executive-level decision making, such as total content count, popular genres, and audience classification
- **Perform time-series analysis**, which helps track Netflix's expansion trend over the years and content addition behavior

- **Segment audience preferences** by comparing Movies vs. TV Shows, genre popularity, and maturity ratings
- **Rank countries, directors, genres, and trending content** to understand global media influence
- **Strengthen relationships between datasets**, enabling integration of viewer ratings with title metadata
- **Summarize high–priority metrics** useful for streaming business decision support

Through DAX, the dataset was transformed from static information into dynamic and interactive insights. It allowed stakeholders to evaluate performance from multiple perspectives such as region, category, and timeline. Therefore, DAX contributed significantly to enhancing the accuracy, intelligence, and depth of the Netflix analytics dashboard.

8.3 Types of DAX Used in This Project

1. Calculated Columns

Used to categorize content and clean inconsistent values.

2. Measures

Used for dynamic calculations that respond to filters and slicers.

3. Time Intelligence DAX

Used for trend analysis and year-wise performance.

Together, these improve data interpretation and KPI reporting.

8.4 Calculated Columns Implemented Using DAX

The following DAX functions were developed and implemented within Power BI to perform different analytical operations including content distribution, audience engagement measurement, and rating insights.

A. Content-Based KPIs

1. **Total Titles Available on Netflix**

```
Total_Titles =
COUNTROWS('Netflix_Dataset csv (2)')
```

2. Total Movies

```
Total Movies =
CALCULATE(
   COUNTROWS('Netflix_Dataset csv (2)'),
   'Netflix_Dataset csv (2)'[type] = "Movie"
)
```

3. **Total TV Shows**

```
Total TV Shows =
CALCULATE(
   COUNTROWS('Netflix_Dataset csv (2)'),
   'Netflix_Dataset csv (2)'[type] = "TV Show"
)
```

4. Total Titles per Year

```
Total Titles Per Year =
COUNTROWS('Netflix_Dataset csv (2)')
```

(Used with Year filter context for year-wise trends)

B. Country-Based KPI

5. Total Contributing Countries

Total Counts =

DISTINCTCOUNT('Netflix_Dataset csv (2)'[country])

This measure helps analyze global content acquisition diversity.

C. Ratings KPIs (IMDb Dataset)

6. Average IMDb Rating

Average IMDB Rating =

AVERAGE('Netflix_Ratings_With_Issues'[imdb_rating])

7. **Highest IMDb Rating**

Highest IMDB Rating =

MAX('Netflix_Ratings_With_Issues'[imdb_rating])

8. **Lowest IMDb Rating**

Lowest IMDB Rating =

MIN('Netflix_Ratings_With_Issues'[imdb_rating])

9. **Rating Titles Count**

Rating Titles Count =

COUNTROWS('Netflix_Ratings_With_Issues')

10.Total Votes Received from Viewers

Total Vote =
SUM('Netflix_Ratings_With_Issues'[num_votes])

8.5 Importance of DAX-Based KPIs in Decision Making

DAX-driven KPIs helped answer business questions like:

- Which type of content should Netflix prioritize?
- Which countries are growing content suppliers?
- Which audience category drives maximum engagement?
- How has content strategy evolved over years?
- What genres require improvement based on ratings?
- Which titles perform well based on user votes?

DAX enables flexible, dynamic insights that update automatically with slicers and filters. DAX-based KPIs form the foundation of an intelligent Netflix dashboard, enabling real-time insights into content strategy and audience behavior. With the combination of aggregated calculations and dynamic filtering, DAX allows decision-makers to interpret performance at both macro (global strategy) and micro (content-level) perspectives. These KPIs provided answers to essential business questions such as:

- Which type of content—Movies or TV Shows—drives higher platform engagement?
- Which countries are emerging as strong contributors to Netflix's content library?
- Which audience category (e.g., TV-MA, TV-14, PG) generates maximum completion and viewership potential?
- What distribution changes occurred in content additions over the years?
- How do viewer satisfaction trends align with content genres and regions?
- Which titles have high popularity backed by user voting strength?

Through DAX, the dashboard became capable of slicing and drilling insights by rating, geography, year, and genre without altering the original dataset. KPIs automatically updated based on user selection, making the analysis more interactive and context-driven.

From a business intelligence perspective, DAX-enabled KPIs allow Netflix-like companies to:

- Identify market gaps for future content acquisition
- Strengthen investment in high-performing content segments
- Optimize recommendations and personalization strategies
- Support strategic planning

8.6 Conclusion

DAX was crucial for analyzing the Netflix datasets and generating actionable KPIs. With measures, calculated columns and time-intelligence formulas, several content performance indicators were developed. These KPIs improved the clarity of visuals and deeper understanding of the Netflix content library.

DAX enabled:
• Quantitative decision-making
• Consistent classification of content
• Trend analysis of release volumes
• Quality measurement using IMDb ratings

Thus, DAX plays a central role in creating a smart and interactive Power BI analytical model, making this project a complete data-driven solution for Netflix content analytics.

CHAPTER 9 DATA VISUALIZATION

Data visualization refers to the graphical representation of information and data using visual elements such as charts, graphs, maps, color-coded indicators, and interactive filtering components. In the context of this Netflix analytics project, Power BI plays a central role in transforming raw datasets into meaningful insights that can support effective decision-making. The visualization layer forms a crucial component of the research methodology, as it converts statistical and tabular data into intuitive graphical formats that highlight patterns, discrepancies, and trends in Netflix's global entertainment catalog.

The primary purpose of the dashboard created in this project is to allow users to explore multiple aspects of the Netflix library, including content volume, movie and TV show distribution, genre category analytics, global coverage of content, viewer ratings, and IMDb performance metrics. The Power BI dashboard has been designed to facilitate both macro-level insights and micro-level drill-down analysis on specific segments of the catalog.

To ensure clarity and consistency, each visualization is developed using appropriate Power BI visuals—such as bar charts, column charts, donut charts, stacked visuals, line charts, card visuals, slicers, and world maps—to represent the underlying data in the most suitable and informative manner. The interactivity features embedded into Power BI, including cross-filtering and slicer-based navigation, offer a responsive analytical experience to decision-makers and users.

The primary purpose of the dashboard created in this project is to allow users to explore multiple aspects of the Netflix library, including content volume, movie and TV show distribution, genre category analytics, global coverage of content, viewer ratings, and IMDb performance metrics. The Power BI dashboard has been designed to facilitate both macro-level insights and micro-level drill-down analysis on specific segments of the catalog.

To ensure clarity and consistency, each visualization is developed using appropriate Power BI visuals—such as bar charts, column charts, donut charts, stacked visuals, line charts, card visuals, slicers, and world maps—to represent the underlying data in the most suitable and informative manner. The interactivity features embedded into Power BI, including cross-filtering

and slicer-based navigation, offer a responsive analytical experience to decision-makers and users.

In this project, visual storytelling is prioritized, meaning that the arrangement of charts and visuals is structured to guide the viewer naturally through the insights. Visual hierarchy was used to position high-level KPIs at the top of the report, providing immediate understanding of platform content distribution across major categories like Movies vs. TV Shows. Detailed visuals, such as genre analysis and country-wise content availability, are placed strategically to enable deeper exploration once basic content composition has been understood.

Another important aspect of the visualization approach is color usage and design standardization. Consistent color themes were implemented across different visuals to avoid confusion and ensure readability. Contrasting colors were selected to differentiate movies and TV shows, making comparison and interpretation easier. Map visuals include location-specific coloring to represent content concentration in different geographical regions, helping users visually identify countries with a high contribution to the Netflix catalog.

The interactive experience in Power BI allows viewers to click any category—such as a specific rating or year—and instantly observe how all visuals respond. This feature enhances data interpretation from multiple perspectives without requiring multiple dashboards or manual filtering. Users gain the flexibility to generate personalized insights simply by selecting a filter of interest.

Overall, the visualization design of this Netflix analytics dashboard aligns with key principles of effective business intelligence reporting:

- **Simplicity** to prevent cognitive overload
- **Accuracy** to maintain truthful representation of data
- **Interactivity** to support exploratory analysis
- **Visual consistency** to enhance comprehension
- **Storytelling** to enable smooth insight discovery

Thus, the data visualization component transforms raw content and ratings data into a valuable decision-support system. By presenting data in a graphical and interactive form, Power BI ensures that even complex patterns become clear and actionable, enabling users to better understand Netflix's strategic content landscape and performance distribution.

9.1 Design Principles for Visualization

The following visualization principles were considered while designing the Netflix dashboard:

• **Simplicity and Readability**: Avoiding clutter and prioritizing visuals that highlight key metrics.
• **Consistency**: Using uniform color schemes to differentiate content types such as Movies and TV Shows.
• **Interactivity**: Enabling slicers and filters for dynamic data exploration.
• **Relevance**: Choosing visualization types according to the nature of data (e.g., trends represented using line charts, comparisons using bar charts).
• **User-Friendly Navigation**: Creating a structured report page with logical grouping of insights.

These principles led to the development of a dashboard that is visually appealing, functional, and analytically strong.

Simplicity and Readability:
Every visual was designed to convey insights quickly without unnecessary styling or clutter. Clean layouts, sufficient spacing, and minimal text descriptions were used to ensure that users can understand insights at a glance.

• **Consistency:**
A uniform color palette was applied throughout the dashboard to clearly differentiate between Movies and TV Shows, as well as highlight key KPI tiles. Consistent shapes, label fonts, and iconography were maintained so users do not get distracted by visual inconsistencies.

• **Interactivity:**
Interactive elements such as slicers, filters, tooltips, and cross-highlighting help viewers explore

data from multiple perspectives. Users can modify filters such as year, genre, or rating and the entire dashboard updates in real time, enabling dynamic and personalized analysis.

• **Relevance:**
Each chart type was selected based on the data characteristics. For example:
– Bar charts for country and genre comparison
– Donut charts for type distribution
– Line charts for trend analysis
– Map visuals for geographical content representation
This ensures that the information is communicated using the most suitable format.

• **User-Friendly Navigation:**
Visuals were organized into structured sections—KPIs at the top, distribution visuals in the middle, and rating-related insights below. Logical grouping allows users to move from high-level overview toward deeper analysis seamlessly.

• **Insight-Driven Design:**
Visuals were optimized to reveal patterns and relationships that support business decisions. For instance, dominant audience ratings or increasing content production patterns are immediately visible through highlighted dashboards.

9.2 Key Visual Components Used in the Dashboard

The Power BI dashboard contains multiple visual elements which present core insights into Netflix's content strategy and market patterns. Each major visual is described below, along with the purpose it serves in the analytical process.

A. KPI Card Visuals (Content Summary Indicators)

KPI cards are used to display single, high-level values extracted from DAX measures. These include:

• Total Titles
• Total Movies
• Total TV Shows
• Distinct Countries
• Average IMDb Rating
• Highest IMDb Rating
• Lowest IMDb Rating
• Total Votes and Ratings Count

These visuals display the most essential performance indicators at a glance, providing instant clarity on the scale and viewer reception of Netflix content.

These KPIs act as a quick summary of the dataset, allowing decision-makers to immediately understand the scale of Netflix's library and external audience perception. For example, knowing the proportion of Movies vs. TV Shows helps identify where Netflix has historically focused more resources. Similarly, average and extreme viewer ratings reveal gaps in quality distribution—indicating whether Netflix tends to host more critically successful content or diverse but mixed-quality content.

Additionally, the presence of live filtering ensures that when users select a specific genre, region, or time range, these KPIs update instantly—making them highly valuable for scenario-based business intelligence discussions. Thus, the KPI card section acts as the primary foundation for dashboard navigation and insight interpretation.

Here are Images of all the KPI

B. Movies vs. TV Shows Distribution Chart

A **donut chart** is used to visually separate Movies and TV Shows.
Purpose:

• Identify which format dominates Netflix's library
• Compare how Netflix has focused content creation or acquisition across different audience preferences

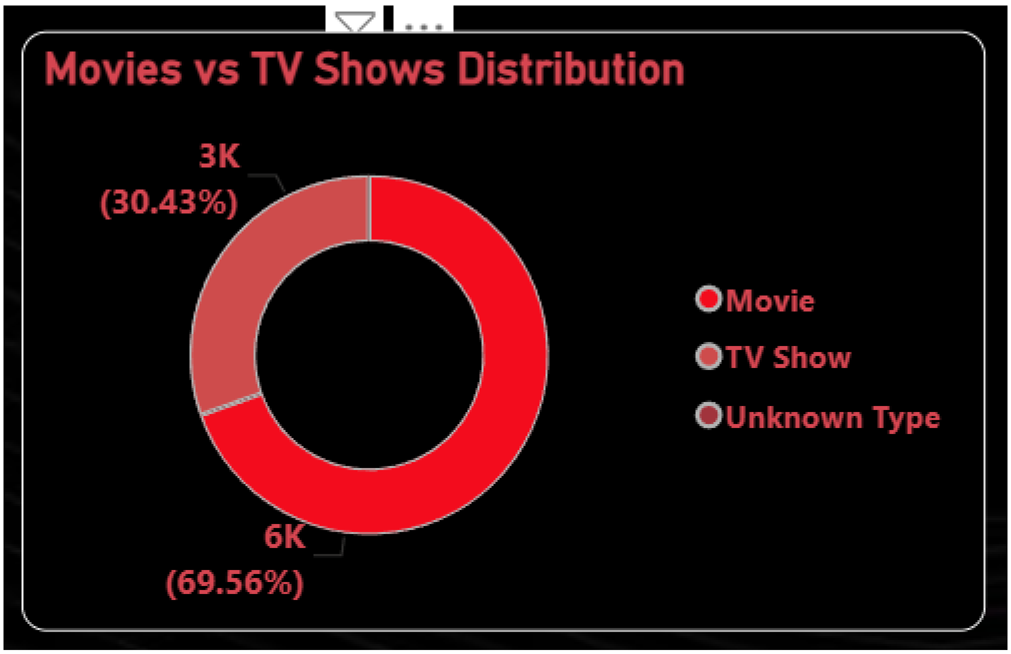
This visualization supports strategic decisions on whether Netflix should increase investments infilms or produce more episodic content.

A donut chart is included in the dashboard to provide a clear visual comparison between the number of Movies and TV Shows available on Netflix. The chart offers a quick, proportion-based representation of the platform's content strategy. Movies typically form a larger portion of Netflix's catalog, as the company initially focused more on purchasing and distributing films globally. However, in recent years, the rise in demand for long-form storytelling and binge-worthy content has resulted in a significant increase in Netflix's TV Show production, including regional originals.

The purpose of this chart is to highlight how Netflix distributes effort and investment between these two major categories of entertainment. From a business perspective, analyzing this ratio helps answer strategic questions such as:

• Is Netflix more dominated by feature films or episodic series?
• How has this mix evolved over time with viewer trends?
• Should the organization further expand series production to retain long-term subscription engagement?

The donut chart supports dynamic filtering through slicers. When a user drills down by country, genre, or release period, the percentage composition automatically adjusts, providing deeper insights into content segmentation patterns. For example, certain regions may specialize more in TV dramas while others contribute mostly movies. Hence, this visualization not only provides a snapshot of the overall content structure but also supports strategic planning in content acquisition and production.



C. Country-Based Content Map Visualization

Power BI's **Filled Map Chart** shows the geographical spread of Netflix titles.

• Highlights dominant content-producing countries such as the United States, India, and the United Kingdom

• Reflects Netflix's success in expanding international content offerings

• Supports localization and expansion strategy insights

Users can click regions to filter other dashboard visuals accordingly.

Typically, countries such as the **United States, India, United Kingdom, Canada, and Japan** emerge as major leaders in Netflix's catalog. These regions have strong media industries and a vast audience base, which significantly contributes to Netflix's multilingual and multicultural content strategy. Additionally, the visualization highlights the platform's continuous effort to expand its presence into newer markets such as South Korea, Spain, and various Middle Eastern and European regions known for popular originals.

The map supports **interactive filtering**, meaning that when a user selects a specific country, the dashboard automatically refreshes other visuals—such as genre distribution, ratings, and release year trends—to show only that region's content performance. This interactivity allows users to explore questions like:

• Which regions produce the most content in specific genres?
• How diverse is Netflix's catalog in terms of global representation?
• Are user ratings aligned with the geographic production trends?

Thus, the country-based map visualization serves as a powerful analytical tool that combines visual storytelling with geographic intelligence, providing meaningful insights for decision-makers, content strategists, and business analysts focused on global media expansion.

D. Year-Wise Content Addition Line Chart

A **line chart** tracks the trend of Netflix movies and TV shows added over years.

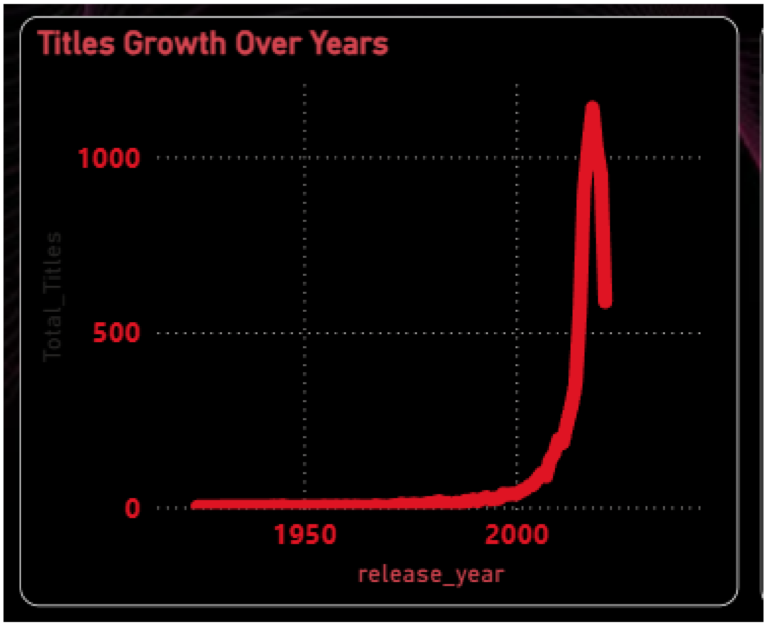
Analytical purpose:

- Understand whether Netflix is growing its catalog over time
- Detect peak production/acquisition years
- Measure scalability of the platform since global expansion initiatives

Insights directly connect with Netflix's ramp-up in original content around 2015 onward.
The Year-Wise Content Addition Line Chart provides a visual representation of how Netflix's catalog has evolved over time. By plotting the number of movies and TV shows added across different years, this chart helps analyze the overall growth pattern of Netflix's content library. It highlights key milestones in Netflix's expansion journey—especially around 2015 and onward, when Netflix increased investments in original production and global licensing deals.

This visualization allows viewers to easily identify fluctuations in content additions. For example, years with steep upward slopes represent Netflix's aggressive content acquisition and production strategies, while periods of decline or slower growth may indicate changes in licensing policies, market competition, or regional regulatory challenges. Furthermore, separating Movies and TV Shows within this chart helps determine which category has experienced faster growth in specific years.

The chart supports strategic evaluation by showing whether Netflix's service is consistently scaling its library to retain and attract subscribers. It also indicates how Netflix adapts to global market demand, such as shifting toward more TV series to enhance long-term viewer engagement.



E. Genre & Category Popularity Visualization

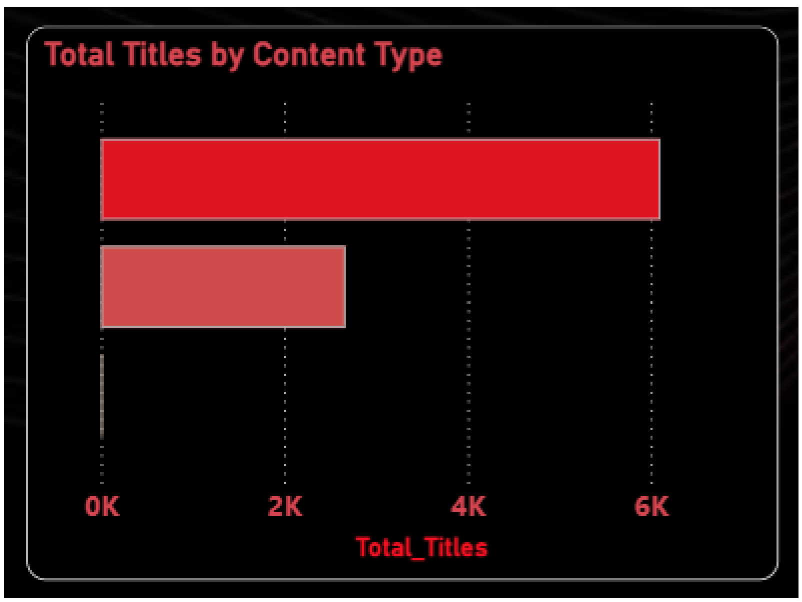
A **stacked bar chart or clustered bar chart** represents top genres in Movies and TV Shows.

- Helps understand viewer demand and entertainment preferences
- Supports genre-based marketing & recommendations strategy
- Highlights emerging categories such as Documentaries and International Dramas

This visualization also provides opportunities for Netflix to diversify content offerings.
The Genre & Category Popularity Visualization uses a stacked bar chart or clustered bar chart to compare the most popular genres across Movies and TV Shows available on Netflix. This chart provides a clear and insightful breakdown of how different content categories perform in terms of volume and viewer preference. It highlights the dominance of globally liked genres such as Dramas, Comedies, and Action films, while also showcasing the rising demand for niche categories like Documentaries, Anime, and International Dramas.

By visually grouping genres for both Movies and TV Shows, this chart helps identify whether audiences prefer long-form episodic storytelling or standalone cinematic experiences in certain genres. For example, Crime or Thriller genres may perform exceptionally well in TV formats due to suspenseful plots, while Romance and Family genres may be more dominant in Movies.

This visualization supports several strategic decisions. Netflix can use genre performance data to guide content investment priorities, strengthen personalized recommendation algorithms, and plan promotional campaigns tailored to viewer interests. It also helps identify content gaps and opportunities for diversification, particularly in emerging regions and languages. Overall, this genre analysis enables Netflix to stay aligned with evolving entertainment trends and maintain a competitive advantage in the global streaming mark



F. Slicers for Interactive Data Filtering

Slicers allow users to refine visual results based on specific attributes, including:

- Year
- Genre
- Country

• Rating Category (e.g., TV-14, TV-MA)
• Movie vs. TV Show

These slicers enhance user-driven exploration without modifying the report structure.
Slicers act as an essential interactive feature in the Power BI dashboard that allow users to dynamically explore Netflix's dataset without changing the structure of the report. With these slicers, viewers can instantly focus on specific segments of content that align with their analytical goals. For example, a **Year slicer** enables trend examination across different time periods, helping users understand how Netflix's content strategy has evolved. A **Genre slicer** makes it possible to narrow down charts and KPI cards to only Action, Drama, Comedy, or any selected content type, providing highly targeted insight.

Similarly, **Country-based filtering** highlights regional strengths and the global expansion of Netflix's catalog. Viewers can analyze output from countries like the United States, India, and South Korea independently or in comparison. The **Rating Category slicer** helps explore content suitability patterns, including maturity classifications like TV-14 or TV-MA, which is crucial for audience segmentation and parental guidance evaluation.

Additionally, the **Movie vs. TV Show slicer** provides instant separation of Netflix's content formats to support decision-making on production focus. Overall, slicers make the dashboard more interactive, user-friendly, and insightful, allowing stakeholders to perform quick, real-time adjustments and discover hidden patterns within the dataset without requiring technical modifications.

9.3 Visualization Workflow

The steps followed when developing the dashboard visuals include:

- Selecting relevant measures for KPI representation
- Designing visuals based on data type & comparison needs
- Applying slicers for real-time interaction
- Ensuring cross-visual filtering for accuracy
- Validating output against dataset context
- Combining visuals into a clear storytelling structure

This workflow ensures a logical narrative flow from high-level metrics down to detailed insights. The visualization development workflow followed in this Netflix dashboard project was structured and insight-driven, ensuring that each chart contributed meaningfully to the overall narrative. The process began with identifying the most important metrics from the dataset, such as content type, ratings, votes, genres, and year-wise additions. These metrics were then transformed into DAX measures and calculated columns wherever required to ensure accurate and standardized reporting across all visuals.

Next, appropriate visual types were selected based on whether the data needed comparison, trend analysis, composition distribution, or geographic representation. For example, KPI cards were used for high-level performance indicators, donut charts for distribution, line charts for trends over time, and filled maps for global content analysis. During design, focus was placed on layout alignment, consistent color themes, and minimal clutter to enhance readability.

Interactivity was then integrated using slicers and cross-filtering techniques, enabling users to explore Netflix datasets from multiple perspectives by simply selecting attributes such as genre, rating, or country. Each visual outcome was cross-checked with the dataset to ensure logical consistency and accuracy. Finally, all elements were arranged in a sequential manner to tell a cohesive analytical story—starting from overall statistics, moving into distribution, and concluding with deep-dive insights.

9.4 Advantages of Visualization-Based Analysis

The dashboard enables Netflix stakeholders and decision-makers to:

- Quickly identify global growth opportunities
- Understand content preferences and audience satisfaction
- Evaluate strengths and weaknesses in the library
- Support targeted business expansion decisions
- Guide future content acquisition and production investments

Visual analytics significantly reduce complexity while improving interpretability. Visualization-based analysis transforms numeric and textual datasets into meaningful visual stories, enabling faster decision-making and deeper understanding. In this project, Power BI visualizations provide immediate clarity on Netflix's content distribution, consumption patterns, and performance indicators. Without analyzing thousands of rows manually, stakeholders can interpret insights directly through charts, maps, and KPI indicators.

The dashboard supports strategic business decisions by clearly revealing which countries contribute the most content, which genres are highly preferred, and how the library has grown over time. With visuals that highlight proportions, comparisons, and seasonal trends, Netflix can quickly spot opportunities where investments may drive maximum engagement. The interactivity built into these visual tools helps users filter the dataset by genre, year, country, or content type, allowing them to focus on specific segments without losing the overall context.

Moreover, visual analytics help identify weaknesses such as underserved content categories, declining popularity trends, or poor-performing titles based on user ratings. These insights can direct Netflix to refine recommendation algorithms, enhance user satisfaction, and optimize content acquisition strategies.

CHAPTER 10

DATA ANALYSIS AND INTERPRETATION

Key Performance Indicators Overview

- The dashboard highlights five major KPIs: Total Movies, Total TV Shows, Total Titles, Rated Titles, and Average IMDb Rating.
- Total Movies count is 6104, indicating that movies form the largest portion of the Netflix content library.
- Total TV Shows count is 2670, which shows that TV shows are growing but still fewer compared to movies.
- Total Titles count is 8775 in the dataset, representing the combined number of movies and series available.
- Rated Titles count is 9 based on IMDb rating integration in the dataset, showing limited rating coverage.
- Average IMDb Rating is 7.94, reflecting an overall positive audience response.

Interpretation:

- Netflix focuses heavily on movies but is gradually increasing TV show count to enhance user retention and engagement.
- Content quality is maintained at a good standard based on IMDb audience responses.

Total Movies – 6104

Movies represent the majority of Netflix's library, highlighting that Netflix continues to acquire and produce a wide range of films across different languages and genres. This suggests a strong focus on cinematic content to attract diverse audiences. Films are also easier to license globally compared to regional TV content.

- **Total TV Shows – 2670**

  Though fewer in number compared to movies, TV shows contribute significantly to platform engagement because of longer watch hours and episode-based storytelling. Netflix has been increasingly investing in original series to improve user retention and subscription longevity—examples include hit titles like *Stranger Things*, *Money Heist*, and *The Crown*.
- **Total Titles – 8775**

  This metric represents the overall volume of available entertainment options. A large content library allows Netflix to cater to different age groups, cultural backgrounds, and audience tastes at a global scale. More titles also help reduce churn by offering continuous fresh choices to users.
- **Rated Titles – 9**

  Only a small portion of the data includes IMDb rating information, indicating that the integrated rating dataset used in this project is limited. However, even with fewer rated entries, the analytical approach demonstrates how audience feedback can be linked with content availability.
- **Average IMDb Rating – 7.94**

  An overall rating close to 8 indicates strong audience satisfaction. High-quality titles contribute to Netflix's brand reputation and competitive advantage in the streaming marketplace.

Movies vs TV Shows Distribution Analysis

- The donut chart compares the number of movies and TV shows.
- Movies contribute close to 70 percent of the total content.
- TV Shows contribute slightly above 30 percent.
- The gap shows Netflix's long-standing reliance on films due to ease of licensing and broad audience reach.
- Growth in TV shows is linked to Netflix's strategy of maximizing watch time through episodic content.

- The distribution pattern signifies a hybrid entertainment model mixing short-duration and long-duration content.
- Users prefer movies for quick entertainment consumption, while series help maintain longer subscription cycles.
- The donut chart effectively highlights the proportional difference between Movies and TV Shows available on Netflix. With approximately **70% content categorized as Movies** and **slightly over 30% as TV Shows**, it is clear that Netflix's catalog has historically been dominated by film content. Movies are generally easier to acquire from production studios because they involve single-title licensing, making them more accessible for rapid expansion of the library. Additionally, standalone films appeal to casual viewers looking for short-term entertainment options.
- However, the rise in TV content demonstrates Netflix's shift in business approach. Over the years, the platform has increasingly produced original web series and partnered with regional creators to launch localized television content. TV Shows help Netflix increase the total watch time per subscriber, as viewers spend many hours completing multi-episode seasons. This aligns with Netflix's retention strategy, where binge-worthy series encourage users to stay subscribed longer.
- The current distribution balance indicates a **hybrid content model** that supports diverse consumption patterns—quick entertainment through films and extended engagement through episodic storytelling. This structure helps Netflix compete with rivals while continuing to evolve as a global streaming leader.

Interpretation:

- Netflix continues to invest more in movies but its recent focus on original series is visible in data trends.

Titles Growth Over Years

- The line chart displays the release year trend of Netflix content.
- Growth remains slow before the year 2000 due to limited production and streaming adoption.

- Significant increase starts after 2010.
- Major spike seen after 2015 with continuous yearly additions.
- The highest peak in growth aligns with Netflix's global expansion during 2017 to 2020.
- The growth reflects Netflix's transition from licensing existing media to producing original content.
- Post-pandemic trends show a slight reduction due to global production challenges.
- The line chart visualizing Netflix content growth over the release years clearly indicates how Netflix evolved from a DVD rental service to a global streaming powerhouse. During the time period **before 2000**, the number of titles remains very minimal because Netflix was not yet producing or distributing digital content. The shift begins gradually **after 2010**, when streaming started gaining wider acceptance and Netflix began securing rights to international media.
- A notable surge in content availability appears **after 2015**, marking the beginning of Netflix Originals such as *Narcos*, *Stranger Things*, and *The Crown*. These successful projects encouraged Netflix to invest heavily in in-house productions. The **steepest growth** is observed between **2017 and 2020**, a period aligned with Netflix's aggressive global expansion strategy. This included entering over 190 countries and collaborating with regional industries, particularly in India, South Korea, and European markets.
- However, the chart shows a **slight decline or stabilization** in growth after 2020. This shift can be attributed to pandemic-related production obstacles and rising competition from platforms like Amazon Prime, Disney+, and HBO Max.

Interpretation:

- Netflix's most aggressive expansion period begins post-2015.
- Rapid scaling of global original content improves platform competitiveness.
- Titles growth aligns with digital transformation in entertainment consumption.

Content Type Analysis

- The bar chart displays total titles by content type.

- Movies exceed more than 6000 titles.
- TV shows are significantly lower but increasing with time.
- The difference confirms Netflix's historical reliance on films for platform content load.
- Netflix's evolving distribution strategy incorporates additional genres and categories to attract diverse audiences.

Interpretation:

- Movies remain a strong pillar for content growth.
- TV shows are gaining strategic value for retaining monthly subscribers.

Regional Content Analysis

- The map visualization highlights country-wise production and contribution.
- The United States is the leading content provider.
- Countries like Canada, the United Kingdom, Mexico, India, Brazil, and Colombia also contribute significantly.
- North America dominates the dataset in terms of volume.
- International markets show increased representation in recent years.
- Netflix invests in region-specific productions to expand global audience reach.
- Localization and language-focused content help drive demand beyond English-speaking countries.
- The map visualization provides a comprehensive view of Netflix's global content distribution and highlights how diverse regions contribute to its entertainment catalog. The **United States** stands out as the largest contributor, reflecting Netflix's origin and its long-term partnerships with Hollywood studios. The strong presence of English-language media drives global accessibility and worldwide audience engagement.
- However, recent trends show that Netflix is no longer dependent solely on American content. Countries such as **India, the United Kingdom, Canada, Brazil, Mexico, South Korea, Spain, France, and Colombia** have emerged as significant contributors, each

bringing unique regional storytelling styles and cultural narratives. This expansion aligns with Netflix's strategic focus on **local content production**, ensuring that audiences in different regions can enjoy authentic, relatable stories in their native languages.

- The map also reveals how Netflix uses regional offerings to strengthen international market penetration. For example, Bollywood films and Hindi series boost popularity in India, while Korean dramas have gained massive global fan bases. Latin American thrillers, Spanish crime dramas, and Turkish romantic series have also driven high viewer engagement worldwide.

Interpretation:

- Netflix's content sourcing mirrors its business expansion strategies.
- Local content initiatives improve cultural relevance and market penetration.

Rating and Audience Satisfaction Insights

- The dataset includes IMDb ratings for limited titles but provides insight into quality.
- Average IMDb Rating of 7.94 shows consistent user satisfaction.
- Highest and lowest rating measures indicate diverse viewer reception across content.
- Rated Titles count shows incomplete coverage due to dataset dependency.
- Audience rating plays a major influence on content life cycle and future productions.
- IMDb ratings included in the dataset, although limited to a small number of titles, provide a valuable indication of how viewers perceive Netflix content in terms of overall quality and engagement. The **Average IMDb Rating of 7.94** suggests that most rated titles fall into the category of *good to very good* audience approval. This demonstrates that Netflix maintains a strong standard in acquiring and producing compelling entertainment content.
- The presence of both **highest and lowest rated** measures in the dashboard highlights the varied reaction of audiences based on storytelling strength, genre appeal, cultural relatability, and production value. High-rated titles reflect successful creative strategies that align well with viewer preferences, while lower-rated ones indicate areas where improvement may be needed.

- Although the **Rated Titles count** is comparatively low due to limited rating availability within the dataset, the insights drawn remain relevant. IMDb ratings continue to play a **major role in content longevity**, as strongly rated titles usually achieve better audience reach and remain longer in the recommended lists. Similarly, poor ratings can influence Netflix to discontinue or reduce promotion of underperforming content.

  Interpretation:
- Netflix maintains acceptable quality standards across content categories.
- More rating-based datasets can enhance content performance analytics.

Analytical Findings Based on Dashboard Visuals

- Distribution of content suggests Netflix caters to varied viewer preferences.
- Regional concentration indicates dependency on specific production hubs.
- TV show rise aligns with binge-watching culture globally.
- Multi-genre content offering improves competitive advantage.
- Increasing release year trend proves strategic focus on Originals and exclusive content.

Business Decision Insights

- Netflix should maintain a balanced ratio between new movies and high-engagement series.
- International content investment can capture underserved markets.
- Enhanced data enrichment like ratings and reviews can improve analytics quality.
- Greater focus on regionally popular genres can optimize subscription growth.

Operational Strategy Implications

- Production planning must consider content demand seasonality.

- Market analysis is required to evaluate territorial licensing and cost efficiency.
- Viewer behavior analytics can guide content renewal decisions.

Viewer Behavior Patterns Derived

- Short format movies attract instant consumption patterns.
- Long-format TV shows improve stickiness and reduce customer churn.
- Audience responds positively to diversified content libraries.
- Quality benchmarking through ratings helps identify trending and underperforming content.

Recommendations Based on Observations

- Increase rating data integration to strengthen analytical capabilities.
- Expand non-English content for cultural inclusivity and growth in global markets.
- Maintain high-quality original productions for brand reputation and loyalty.
- Analyze countries with lower contribution to identify untapped opportunities.
- Employ advanced forecasting to anticipate content preferences and genre shifts.

Conclusion

- The analysis reveals that Netflix has strong market presence with rapid content expansion.
- Movies dominate the overall media offering, while TV shows are growing at a fast pace.
- Geographic expansion reflects ambitious global business strategy.
- Viewer feedback represented through IMDb ratings confirms generally positive satisfaction levels.
- The insights gained through Power BI visualizations highlight the effectiveness of Netflix's content strategy.

- Continuous improvement and diversification are required to maintain leadership in the OTT market.

The Power BI visualizations used in this Netflix analytics project play a critical role in simplifying data interpretation and assisting strategic decision-making. The value provided by these visuals can be summarized under the following points:

• KPI card visuals offer a clear overview of the total content library, audience ratings, and title distribution. These quick insights help executives understand content volume, growth targets, and quality status without exploring detailed data.

• The Movies vs. TV Shows distribution chart helps stakeholders evaluate the balance between film and episodic content. Decisions related to future production priorities and user engagement strategies can be made based on this ratio.

• The country-based content map reveals geographic content coverage and highlights major contributing nations. This allows business leaders to identify potential areas for expansion, regional content investment opportunities, and localization needs.

• The year-wise content addition line chart provides trend analysis that supports decisions regarding production scaling. Managers can assess whether content output aligns with business goals, especially in key growth years.

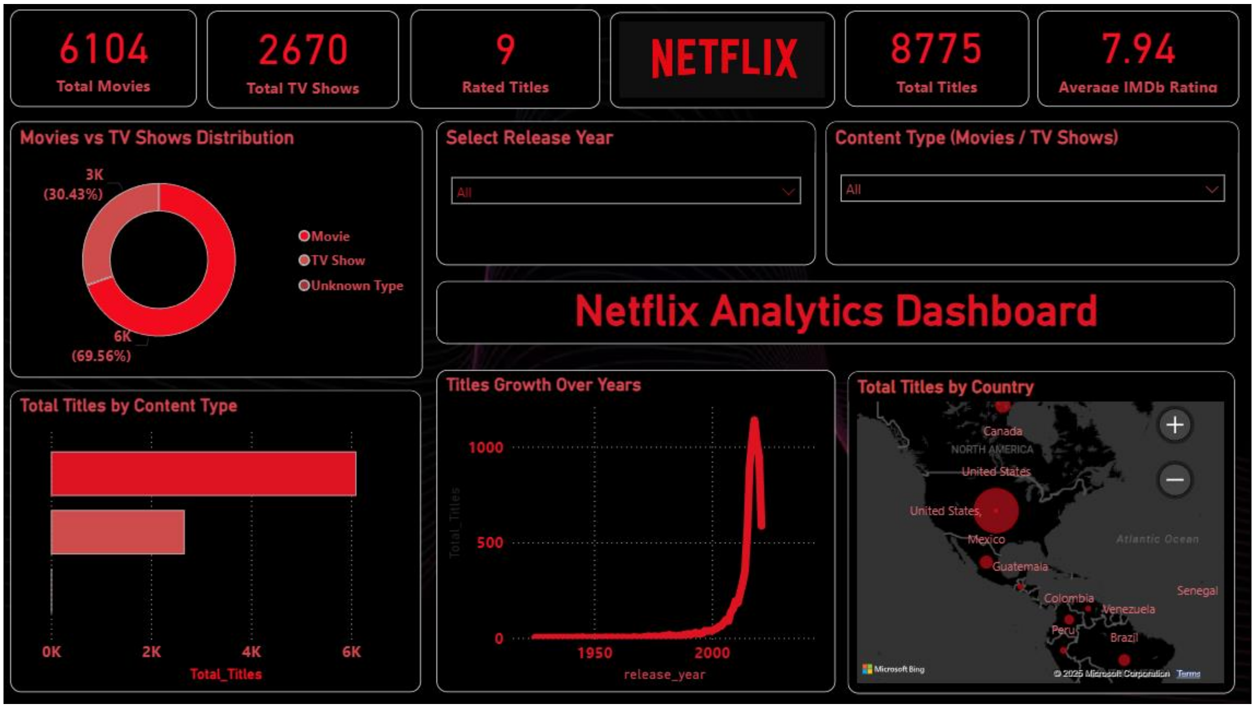
• The genre popularity visualization assists in understanding what types of content viewers prefer. This helps Netflix plan future acquisitions, marketing strategies, and personalized recommendations for different audience groups.

• Interactive data slicers give decision-makers the ability to filter information by year, country, rating, and content type without changing the dashboard structure. This enables scenario-based analysis and improves the speed and flexibility of insights.

• Cross-filtering features ensure that selections in one visual automatically update related visuals. This improves analytical accuracy and supports relational decision-making.

• Visual story flow on the dashboard allows stakeholders to move from high-level metrics to detailed drill-down insights, enabling structured and informed evaluation of content performance.

Final Dashboard (NETFLIX)



CHAPTER 11 RECOMMENDATIONS AND CONCLUSION

11.1 Introduction

- This final chapter outlines key recommendations derived from the analytical findings and concludes the project by emphasizing major outcomes.
- The primary aim is to translate data insights into actionable strategies that could help Netflix improve content offerings, user experience, and business decisions.

- The chapter reaffirms how Business Intelligence contributes toward understanding the competitive OTT industry.

11.2 Key Recommendations

Content Strategy Recommendations

- Increase the volume of **TV Shows**, as Movies currently dominate the content library.
- Focus on enhancing **regional content**, particularly in countries with emerging demand such as India, South Korea, Spain, and Japan.
- Improve representation of **less prominent genres** like documentary series, animated content, and international-language films.
- Use **viewer rating patterns** to filter out low-rated or declining content and discontinue or redesign them.

Viewer Engagement Recommendations

- Introduce enhanced personalized features using machine learning-powered recommendations.
- Develop **genre-specific campaigns** based on user preferences revealed in the dashboard analysis.
- Encourage higher viewer interaction with ratings and reviews to improve content visibility and ranking.
- Promote Netflix Originals more aggressively in regions where licensed content dominates.

Release Strategy Recommendations

- Maintain high frequency of content release in recent years to keep engagement strong.
- Explore **seasonal release trends** to ensure maximum viewership at peak times.

- Improve release distribution by diversifying across multiple months rather than clustering content.

Business and Operational Recommendations

- Utilize data forecasting from Power BI to predict subscription trends and content performance.
- Strengthen contracts with leading content-producing countries like the United States, India, and the United Kingdom.
- Allocate budgets based on market-specific insights to maximize returns on content investments.

11.3 Conclusion

- This project successfully analyzed Netflix's movies and TV shows using Power BI to identify distribution patterns, viewer ratings, and genre trends.
- Two Kaggle datasets were used, covering both content attributes and IMDb-based ratings.
- Comprehensive cleaning and preparation ensured reliable analytics.
- Power BI dashboards enabled interactive exploration, helping uncover:
    - Global expansion in Netflix's content presence
    - Rapid growth in titles in the last two decades
    - User preferences reflected via ratings and votes
    - Dominance of movie content over TV shows
    - Content clustering in specific countries and genres
- The Project achieved the following objectives:
    - Insights into content availability across the world
    - Understanding of audience response through IMDb ratings
    - Visual identification of trending release years and categories
    - Analytical support for content acquisition decisions

- Overall, the analysis proves that **BI tools are essential** for streaming platforms like Netflix to remain competitive, adaptive, and user-centric.

11.4 Significance of the Study

- Shows practical application of Business Intelligence using real data.
- Demonstrates how analytics can guide content planning and marketing.
- Provides a structured foundation for future OTT platform research.
- Strengthens academic understanding of Power BI capabilities in industry scenarios.

11.5 Future Scope

- The study can be extended by including:
    - More audience behavior metrics from Netflix internal data.
    - Subscription growth data across different markets.
    - User engagement details such as watch hours and skip rates.
    - Advanced analytics like machine learning prediction models.
- Expansion toward comparative analytics with Amazon Prime, Disney+, Hotstar, etc., would give broader market insights.
- Enhancing dashboards with advanced visuals like decomposition trees, AI-based anomaly detection, and forecasting will improve decision precision.

CHAPTER – 12 BIBLIOGRAPHY & REFERENCES

Websites & Online Sources

1. Kaggle. (2020). *Netflix Movies and TV Shows Dataset*. Retrieved from https://www.kaggle.com/shivamb/netflix-shows
2. Kaggle. (2020). *Netflix IMDb Scores Dataset*. Retrieved from https://www.kaggle.com/datasets/ashirwadsangwan/imdb-dataset
3. Netflix. (2024). *Netflix Official Website*. Retrieved from https://www.netflix.com/
4. Microsoft. (2023). *Microsoft Power BI Documentation*. Retrieved from https://learn.microsoft.com/power-bi/
5. Statista. (2023). *Number of Netflix subscribers worldwide: Statistics & Trends*. Retrieved from https://www.statista.com/
6. IMDB. (2024). *IMDb: Ratings & Reviews*. Retrieved from https://www.imdb.com/
7. ResearchGate. (2023). *Studies on Streaming Services and Viewer Behavior*. Retrieved from https://www.researchgate.net/