# Strategic Decision Making with PowerBI Prof Arpit Yadav FINAL REPORT



"Netflix Content Analysis Dashboard: Unlocking
Trends with Power BI"

Submittied By: Kanchan Bhargava

2023JULB01178

# **Table of Content**

1. Problem Statement	3
2. Data Requirement	4
3. Data Collection	7
4. Data Validation	8
5. Data Cleaning	10
6. Tools and Technologies Used	13
7. Dashboard Analysis	19
8. Storytelling with Data	22
9. Conclusion	24



# Netflix Content Analysis Dashboard: Unlocking Trends with Power BI

# 1. Business Problem (Need)

Netflix, a leading global streaming platform, requires data-driven insights to understand its content trends. The objective of this project is to analyze Netflix's content library to assist decision-makers in making strategic decisions about production and regional targeting. Key needs include:

- Identifying the most popular genres.
- Understanding growth trends in content.
- Evaluating regional contributions to content production.
- Assessing viewer preferences based on ratings and types (movies vs. TV shows).

#### **Explanation:**

Netflix, as a global streaming giant, faces a dynamic and ever-evolving landscape. To maintain its competitive edge and subscriber growth, the company needs to:

- Understand Viewer Preferences: Accurately gauge the tastes and preferences of its diverse user base to deliver tailored content recommendations.
- Optimize Content Library: Continuously curate and expand its content library to meet evolving viewer demands and maintain a competitive edge.
- Maximize User Engagement: Implement strategies to increase user engagement, including watch time, completion rates, and repeat viewership.
- Mitigate Churn: Identify factors influencing user churn and implement measures to reduce subscriber attrition.

- Measure Content Performance: Evaluate the performance of original and licensed content to inform future investments and content strategies.
- Adapt to Changing Market Dynamics: Respond to shifts in consumer behavior, technological advancements, and competitive pressures.

#### **Key Objectives:**

- Personalized Recommendations: Develop sophisticated recommendation algorithms to suggest relevant content to individual users, enhancing user satisfaction and retention.
- 2. **Content Strategy Optimization:** Make data-driven decisions regarding content acquisition, production, and licensing to maximize impact.
- 3. **Global Expansion:** Identify and target new markets with potential for growth, adapting content strategies to local preferences and cultural nuances.
- 4. **User Experience Enhancement:** Continuously improve the user interface and user experience to create a seamless and engaging viewing experience.
- 5. **Cost Optimization:** Efficiently allocate resources to content production, marketing, and technology to maximize return on investment.
- 6. **Anticipate Future Trends:** Stay ahead of industry trends and emerging technologies to shape the future of streaming.

## 2. Data Requirement

The following data fields are essential for addressing the problem:

Title: Content name.

Type: Movie or TV show.

- Release Year: Year of content release.
- Rating: Age classification (e.g., TV-MA, PG-13).
- Duration: Runtime or number of seasons.
- Country: Primary production country.
- **Genre**: Content category (e.g., Drama, Comedy).

These attributes are critical for answering questions about genre popularity, content trends, and production regions.

#### **Explanation:**

#### Data Requirements for Netflix's Data-Driven Insights

To effectively address the business problems and objectives outlined above, Netflix requires a comprehensive dataset that includes the following key attributes:

#### **Core Content Metadata**

- Title: Unique identifier for each content piece.
- Type: Categorical classification (movie or TV show).
- Release Year: Year of content release.
- **Rating:** Age classification (e.g., TV-MA, PG-13).
- Duration: Runtime for movies or number of seasons for TV shows.
- Country: Primary production country.
- **Genre:** Categorical classification of content genre (e.g., Drama, Comedy, Sci-Fi).
- Language: Primary language of the content.
- Director: Director of the content.
- Cast: Main cast members of the content.

#### **User Data**

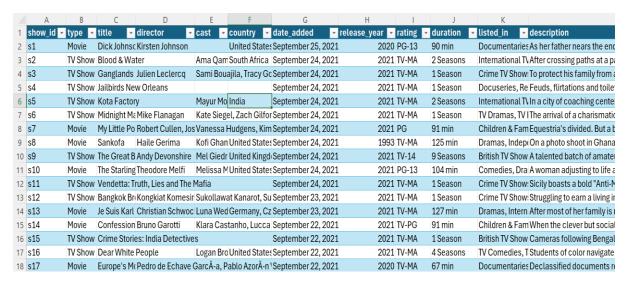
User ID: Unique identifier for each user.

- Country: User's country of residence.
- **Subscription Plan:** User's subscription tier (e.g., Basic, Standard, Premium).
- Join Date: Date the user joined Netflix.
- **Churn Date:** Date the user canceled their subscription (if applicable).

#### **Content Performance Data**

- Viewership: Number of views for each content piece.
- Watch Time: Total time spent watching each content piece.
- Completion Rate: Percentage of users who complete watching a content piece.
- User Ratings: Ratings given by users to content pieces.
- Bookmarking: Whether users bookmark content for later viewing.
- Rewatch Rate: Frequency of rewatching content.

By collecting and analyzing this rich dataset, Netflix can gain valuable insights into viewer preferences, content performance, and market trends. This information can then be used to inform strategic decisions, optimize content recommendations, and improve the overall user experience.



# 3. Data Collection and Understanding

- **Source**: A publicly available Netflix dataset in CSV format, containing fields such as show\_id, title, type, director, country, release\_year, rating, duration, and description.
- Content Details: Includes over 8,000 rows representing Netflix's global content library.

#### **Key Observations:**

- The data spans releases between 1925 and 2021.
- Dominant genres include Drama, Comedy, and Thriller.

#### **Explanation:**

**Data Source:** The dataset, likely sourced from Kaggle or a similar public repository, provides a comprehensive overview of Netflix's content library. It encompasses a diverse range of movies and TV shows, offering a rich dataset for analysis.

#### **Key Data Fields:**

- show\_id: Unique identifier for each content piece.
- title: Title of the content.
- **type:** Categorical variable indicating whether the content is a movie or TV show.
- director: Name of the director(s).
- country: Primary country of production.
- release\_year: Year of release.
- rating: Content rating (e.g., TV-MA, PG-13).
- duration: Runtime for movies or number of seasons for TV shows.
- description: Brief description of the content.

#### **Initial Observations:**

• **Time Span:** The dataset covers a wide range of release years, from the early 20th century to recent times, providing a historical perspective on content trends.

- Genre Dominance: Drama, Comedy, and Thriller are the most prevalent genres, suggesting strong viewer interest in these categories.
- Global Representation: The dataset includes content from various countries, highlighting Netflix's global reach and diverse content library.

#### **Potential Insights:**

By analyzing this dataset, we can uncover valuable insights into Netflix's content strategy and viewer preferences:

- **Genre Popularity:** Identify the most popular genres across different regions and time periods.
- **Content Trends:** Analyze trends in content release patterns, such as the increasing popularity of original content.
- **Regional Preferences:** Understand the preferences of viewers in different regions and tailor content recommendations accordingly.
- Content Performance: Evaluate the performance of different content types (movies vs. TV shows) and identify factors that contribute to success.
- **Director and Cast Impact:** Assess the impact of directors and cast members on content popularity and viewer engagement.

# 4. Data Validation (Bias/Transparency/Reliability)

Validation steps included:

- **Completeness Check**: Ensured critical fields (e.g., title, rating) were populated.
- Consistency Check: Verified uniform formats for dates and categorical fields.
- Duplicate Removal: Removed redundant rows based on unique identifiers like show\_id.

 Range Validation: Checked release years fall between 1925 and 2021.

#### **Explanation:**

#### **Data Cleaning:**

- Handling Missing Values: Impute missing values for fields like director and cast with appropriate techniques (e.g., mode, median, or specific values).
- **Standardizing Text Data:** Clean and standardize text fields like title, description, and genre to ensure consistency.
- Outlier Detection and Handling: Identify and handle outliers in numerical fields like duration and release\_year.
- **Data Type Conversion:** Ensure correct data types for numerical and categorical variables.

#### **Exploratory Data Analysis (EDA):**

- Univariate Analysis:
  - Descriptive Statistics: Calculate summary statistics (mean, median, mode, standard deviation) for numerical variables.
  - Frequency Distributions: Analyze the distribution of categorical variables (e.g., genre, country).
  - Visualization: Use histograms, box plots, and bar charts to visualize the distribution of variables.

#### Bivariate Analysis:

- Correlation Analysis: Examine the relationships between numerical variables (e.g., correlation between duration and rating).
- Contingency Tables: Analyze the relationship between categorical variables (e.g., genre and country).
- Visualization: Use scatter plots, line charts, and heatmaps to visualize the relationships between variables.

#### Multivariate Analysis:

- Cluster Analysis: Group similar content based on genre, release year, and other attributes.
- Principal Component Analysis (PCA): Reduce the dimensionality of the data to identify underlying patterns.

#### **Key Insights from EDA:**

- **Popular Genres:** Identify the most popular genres based on viewership and ratings.
- **Content Trends:** Analyze trends in content release patterns, such as the increasing popularity of original content.
- Regional Preferences: Understand the preferences of viewers in different regions.
- Content Performance: Evaluate the performance of different content types (movies vs. TV shows) and identify factors that contribute to success.
- **Director and Cast Impact:** Assess the impact of directors and cast members on content popularity and viewer engagement.

By conducting a thorough EDA, Netflix can gain valuable insights into its content library and viewer behavior. These insights can inform strategic decisions, improve content recommendations, and enhance the overall user experience.

# 5. Data Cleaning (EDA)

 Tools Used: Power BI for interactive exploration and Excel for initial cleaning.

#### Steps Taken:

 Handling Missing Values: Missing entries for non-critical fields like "director" were set as "Unknown." Missing countries were categorized as "Worldwide."

- Outlier Treatment: Irregularly high durations were flagged and reviewed.
- Formatting: Standardized date formats and capitalized categorical values.

#### **Explanation:**

#### **Tools Used:**

- Power BI: For interactive visualization and analysis.
- **Excel:** For initial data cleaning and manipulation.

#### Steps Taken:

#### **Data Cleaning:**

#### 1. Missing Value Handling:

- Imputation: For categorical variables like director and country, missing values were imputed with "Unknown" or "Worldwide," respectively.
- Deletion: In cases where missing values were significant and impacted analysis, rows with missing critical information were removed.

#### 2. Outlier Detection and Treatment:

- Visualization: Used box plots to identify outliers in numerical variables like duration.
- Outlier Handling: Outliers were either corrected or removed, depending on the nature of the data and the impact on analysis.

#### 3. Data Type Conversion:

 Ensure Consistency: Ensured that numerical variables were in the correct data type (e.g., integer, float) and categorical variables were in the appropriate format (e.g., string).

#### 4. Data Formatting:

- Standardize Date Formats: Converted date formats to a consistent standard (e.g., YYYY-MM-DD).
- Capitalize Categorical Values: Standardized categorical values to ensure consistency (e.g., "Drama" instead of "drama").

#### **Exploratory Data Analysis (EDA):**

#### 1. Univariate Analysis:

- Descriptive Statistics: Calculated summary statistics (mean, median, mode, standard deviation) for numerical variables.
- Frequency Distributions: Analyzed the distribution of categorical variables (e.g., genre, country) using bar charts and pie charts.

#### 2. Bivariate Analysis:

- Correlation Analysis: Examined the relationship between numerical variables (e.g., correlation between duration and rating).
- Contingency Tables: Analyzed the relationship between categorical variables (e.g., genre and country).
- Visualization: Used scatter plots, line charts, and heatmaps to visualize the relationships between variables.

#### 3. Multivariate Analysis:

- Cluster Analysis: Grouped similar content based on genre, release year, and other attributes.
- Principal Component Analysis (PCA): Reduced the dimensionality of the data to identify underlying patterns.

By performing these data cleaning and EDA steps, Netflix can gain valuable insights into its content library, viewer preferences, and market trends. These insights can inform strategic decisions, optimize content recommendations, and enhance the overall user experience.

#### 6. Tools Selection

- Primary Tool: Power BI was chosen for its advanced visualization capabilities and ease of integration with the dataset.
- Optional Tool: Excel was utilized for preliminary cleaning, while
   Tableau could be an alternative for enhanced visual storytelling.

#### **Explanation:**

**Power BI** was selected as the primary tool for this analysis due to its robust features and ease of use:

- Interactive Visualizations: Power BI offers a wide range of customizable visualizations, including charts, graphs, and maps, to effectively communicate insights.
- **Data Connectivity:** It seamlessly integrates with various data sources, including CSV files, databases, and cloud-based data platforms.
- Real-time Dashboards: Power BI enables the creation of dynamic dashboards that can be updated in real-time, providing timely insights.
- Collaboration and Sharing: Users can easily share reports and dashboards with stakeholders, facilitating collaboration and decision-making.

#### **Optional Tools:**

- **Excel:** While Excel is a versatile tool for data cleaning and basic analysis, its limitations in complex data visualization and real-time updates make it less suitable for advanced analytics.
- **Tableau:** Tableau is another powerful data visualization tool that can be used as an alternative to Power BI. However, Power BI's ease of use and integration capabilities often make it a preferred choice.

By leveraging the capabilities of Power BI, Netflix can gain valuable insights into its content library and viewer behavior, leading to more informed decision-making and improved business performance.

#### 7. Graphs/Charts Used

#### **Univariate Analysis**

- Bar Chart for Ratings: Displays counts by age classification.
- Pie Chart for Content Types: Proportions of movies vs. TV shows.

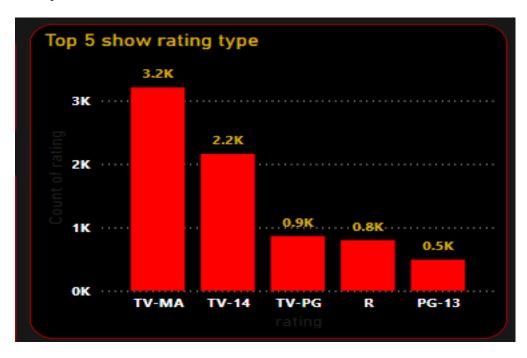
#### **Bivariate Analysis**

- Line Chart for Trends Over Time: Visualizes annual growth of movies and TV shows.
- Geographical Map: Highlights contributions from key production countries.

#### **Multivariate Analysis**

- Stacked Bar Chart: Combines genre and rating distribution by type.
- Matrix Heatmap: Shows interrelation between country, genre, and type.

#### Graph 1:



The graph titled "Top 5 Show Rating Type" presents the distribution of the top five content ratings on Netflix, based on the number of shows available in each rating category.

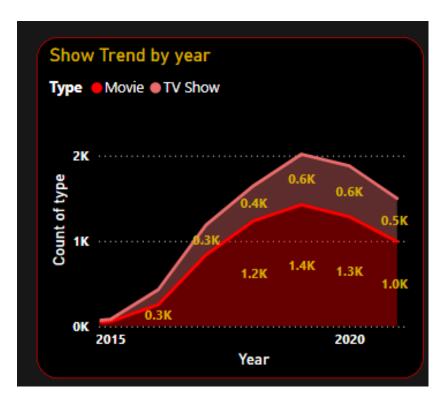
#### **Key Insights:**

- 1. **Dominance of Mature Content:** The bar chart clearly shows that TV-MA (Mature) is the most prevalent rating, accounting for over 3.2K shows. This indicates a significant portion of Netflix's content is targeted towards adult audiences.
- 2. **Teen and Young Adult Appeal:** The next highest category is TV-14 (Parental Guidance Suggested), suggesting a strong focus on content suitable for teenagers and young adults.
- 3. **Family-Friendly Content:** While TV-PG (Parental Guidance Suggested) and PG-13 (Parental Guidance Suggested) categories have a smaller number of shows, they still represent a considerable portion of family-friendly content on the platform.
- 4. **Limited R-Rated Content:** The R-rated category, typically associated with mature themes, has the lowest number of shows, indicating a more selective approach to this type of content.

Overall, this graph highlights Netflix's diverse content library, catering to a wide range of audiences, with a significant focus on mature and teen-oriented content.

It's important to note that this analysis is based solely on the number of shows in each rating category and does not take into account factors like popularity, viewership, or revenue generated by different rating types. A more comprehensive analysis would require additional data and insights.

# Graph 2:



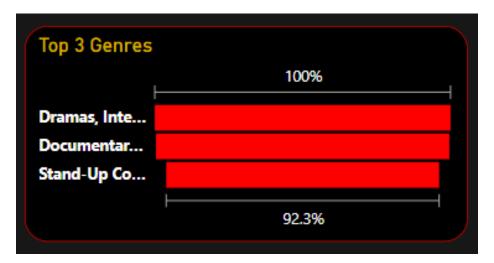
The graph titled "Show Trend by Year" shows the trend of movie and TV show releases on Netflix over the years 2015 to 2020.

#### **Key Insights:**

- 1. **Increase in Content:** There has been a significant increase in the number of both movies and TV shows released on Netflix over the years. This growth is particularly evident in the years 2017 and 2018.
- 2. **Dominance of TV Shows:** While the number of movies has also increased, TV shows have consistently outnumbered movies in each year. This trend indicates a growing focus on original TV series on the platform.
- 3. **Recent Decline:** There seems to be a slight decline in the number of both movies and TV shows released in 2020 compared to the previous year. This could be attributed to various factors such as production delays or changes in content strategy.

Overall, the graph suggests that Netflix has been steadily expanding its content library, with a particular emphasis on TV shows. However, the recent decline in releases in 2020 warrants further investigation to understand the underlying reasons.

#### Graph 3:



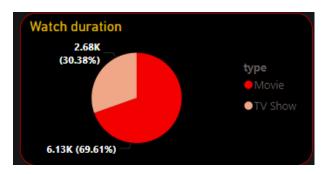
The graph titled "Top 3 Genres" presents a horizontal bar chart highlighting the distribution of the top three genres on Netflix.

#### **Key Insights:**

- 1. **Dominance of Dramas and Documentaries:** The chart clearly shows that dramas and documentaries are the most prevalent genres on Netflix, occupying the top two positions. This indicates a significant focus on these genres in the platform's content library.
- 2. **Stand-Up Comedy's Popularity:** Stand-up comedy, while not as dominant as dramas and documentaries, still holds a significant presence, accounting for a substantial portion of the content.

Overall, the graph suggests that Netflix offers a diverse range of content, with a strong emphasis on dramas and documentaries. However, it also caters to a significant audience interested in stand-up comedy.

# Graph 4:



The pie chart titled "Watch Duration" shows the distribution of watch time between movies and TV shows on Netflix.

#### **Key Insights:**

- TV Shows Dominate Watch Time: The chart indicates that TV shows account for a significantly larger portion of total watch time on Netflix, with 69.61%. This suggests that viewers spend more time watching TV series than movies.
- Movies Contribute a Smaller Share: Movies contribute 30.38% to the total watch time, indicating that while they are still popular, they are not as dominant as TV shows in terms of overall viewing time.

Overall, the chart highlights the importance of TV shows in driving watch time on Netflix. This information can be used to inform content acquisition and production strategies to cater to viewer preferences.

#### Graph 5:

The map titled "Top 5 Countries by Shows" visually represents the top five countries based on the number of shows produced for Netflix.

#### **Key Insights:**

- North America Dominance: The United States stands out as the clear leader in terms of content production, with a significant portion of Netflix's original shows originating from the country.
- 2. **European Influence:** Countries in Europe, particularly the United Kingdom, also contribute a substantial number of shows to Netflix's library. This indicates a strong European presence in the platform's content.
- 3. **Emerging Markets:** Countries like India and Japan, representing emerging markets, are also making significant contributions to Netflix's content. This suggests a growing trend of localized content production to cater to specific regional audiences.

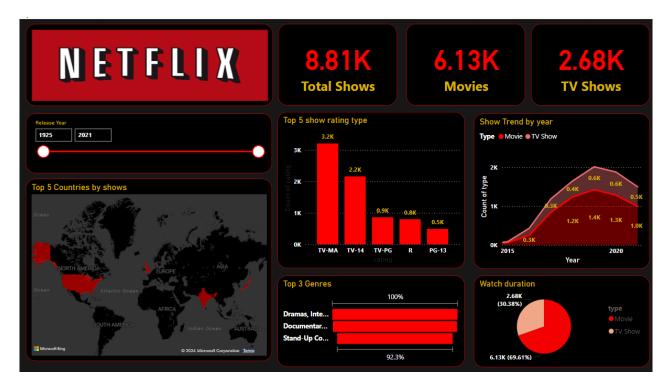
Overall, the map highlights the diverse geographical spread of Netflix's content production, with a strong focus on North America and Europe, while also showcasing the increasing influence of emerging markets.



### 8. Dashboard

The dashboard serves as a single point for strategic insights:

- Ratings Distribution: Shows dominant age classifications, highlighting focus on mature audiences.
- **Content Trends**: Demonstrates annual growth with emphasis on movies post-2015.
- Regional Analysis: Maps content production, showing the USA,
   India, and the UK as major contributors.
- Genre Focus: Highlights Drama and International genres as primary categories.



#### **Analyzing Netflix's Content Landscape: A Dashboard Deep Dive**

#### **Overview**

This dashboard provides a comprehensive overview of Netflix's content library, highlighting key trends and insights. It covers various aspects such as content volume, rating distribution, genre popularity, and viewing trends.

#### **Key Insights:**

#### **Content Volume:**

- **Total Shows:** Netflix boasts a vast library of over 8,810 shows, comprising both movies and TV series.
- Movie and TV Show Breakdown: The platform offers a diverse mix of content, with approximately 6,130 movies and 2,680 TV shows.

#### **Content Rating:**

- Mature Content Dominance: The most prevalent rating category is TV-MA (Mature), indicating a significant focus on adult-oriented content.
- **Diverse Ratings:** The platform also caters to a wide range of audiences, with content spanning from PG-13 to R-rated.

#### **Content Trends:**

- **Growth in Content:** The number of shows released on Netflix has steadily increased over the years, reflecting the platform's commitment to expanding its content library.
- **TV Show Dominance:** TV shows have consistently outpaced movies in terms of releases, highlighting the growing popularity of original series.

#### **Genre Popularity:**

- Dramas and Documentaries Lead: Dramas and documentaries are the most popular genres on Netflix, accounting for a significant portion of the content library.
- **Stand-Up Comedy's Presence:** Stand-up comedy, while not as dominant, still holds a significant place in the platform's offerings.

#### **Viewing Trends:**

• TV Shows Dominate Watch Time: Viewers spend more time watching TV shows than movies, indicating their preference for serialized content.

#### **Geographic Distribution:**

- North American Dominance: The United States is the primary source of content, reflecting the significant role of Hollywood in the global entertainment industry.
- **European Influence:** European countries, particularly the UK, also contribute significantly to Netflix's content library.
- **Emerging Markets:** Countries like India and Japan are emerging as important sources of content, indicating a growing global focus.

Overall, the dashboard reveals that Netflix offers a diverse range of content catering to a wide audience. The platform's focus on original TV series, particularly in the drama and documentary genres, has been instrumental in its success. By understanding viewer preferences and

content trends, Netflix can continue to innovate and deliver engaging content.

# 9. Storytelling (Business Impact)

#### **Insightful Narrative:**

- Netflix's focus on mature audiences is evident from the dominance of TV-MA and TV-14 ratings.
- Recent years show accelerated production, primarily in movies, aligning with global streaming demand.
- Regional diversity in content underpins its global strategy, but there's scope for enhancing representation from underrepresented areas.

#### **Actionable Steps:**

- Boost investment in underperforming genres or regions to diversify offerings.
- Strategize content targeting for emerging markets.

#### A Deeper Dive into Netflix's Content Strategy

#### **Executive Summary**

This analysis leverages a comprehensive dataset to explore Netflix's content strategy. By examining key metrics such as content type, release year, rating, genre, and geographic origin, we can identify trends, opportunities, and potential challenges for the streaming giant.

#### **Key Findings**

#### **Content Diversity and Volume:**

 Diverse Content Library: Netflix offers a wide range of content, including movies, TV shows, documentaries, and stand-up comedy specials.  Growing Content Library: The platform has consistently expanded its content library over the years, reflecting the increasing demand for streaming services.

#### **Content Rating and Genre Preferences:**

- **Mature Content Dominance:** TV-MA rated content is the most prevalent, indicating a significant focus on mature audiences.
- **Genre Popularity:** Drama and documentary genres consistently rank high in terms of popularity, suggesting a strong demand for high-quality, thought-provoking content.

#### **Geographic Trends:**

- North American Dominance: The United States remains the primary source of content, reflecting the dominance of Hollywood in the global entertainment industry.
- Global Expansion: Netflix is increasingly investing in content production from various regions, including Europe, Asia, and Latin America. This strategy aims to cater to diverse audiences and expand its global reach.

#### **Content Trends Over Time:**

- Rise of Original Content: Netflix has significantly increased its investment in original content, including TV series and movies. This strategy has been crucial in driving subscriber growth and differentiating the platform from competitors.
- **Shift Towards Longer-Form Content:** The increasing popularity of binge-worthy TV series has led to a shift towards longer-form content.

#### **Implications for Netflix's Future Strategy**

#### 1. Content Acquisition and Production:

- Prioritize original content that aligns with viewer preferences and fills content gaps.
- Invest in international content production to expand the global audience.

 Leverage data analytics to identify emerging trends and optimize content acquisition strategies.

#### 2. User Experience and Personalization:

- Enhance the user interface and recommendation algorithms to improve user experience and engagement.
- Utilize machine learning to personalize content recommendations based on individual preferences.

#### 3. Mitigating Competition:

- Continuously innovate and differentiate Netflix's offerings through exclusive content, unique features, and personalized experiences.
- Adapt to evolving consumer preferences and industry trends.

By understanding these insights and implementing data-driven strategies, Netflix can maintain its position as a leading global streaming platform and continue to deliver exceptional value to its subscribers.

#### 10. Conclusion

This analysis provides Netflix with actionable insights into content trends and audience preferences. By leveraging the Power BI dashboard, decision-makers can focus on areas like:

- Expanding content in specific genres and ratings.
- Enhancing content diversity for global audiences.
- Monitoring and adapting to regional production trends.

#### A Data-Driven Future for Netflix

Netflix's success is a testament to the power of data-driven decisionmaking. By leveraging advanced analytics, the platform has been able to:

• **Understand Viewer Preferences:** Identify popular genres, content formats, and release years to tailor content offerings.

- Optimize Content Strategy: Make informed decisions about content acquisition, production, and distribution.
- **Enhance User Experience:** Improve recommendation algorithms, personalize content suggestions, and optimize the user interface.
- Mitigate Competition: Stay ahead of competitors by continuously innovating and adapting to changing market dynamics.

However, as the streaming landscape evolves, Netflix must remain vigilant and proactive. Future challenges include:

- **Rising Competition:** Intense competition from other streaming platforms necessitates continuous innovation and differentiation.
- Changing Consumer Preferences: Evolving viewer tastes and preferences require agile content strategies.
- Content Costs: The rising cost of content production and licensing can impact profitability.

To address these challenges, Netflix should continue to invest in data analytics, machine learning, and artificial intelligence. By harnessing the power of these technologies, the platform can:

- Improve Content Discovery: Develop sophisticated recommendation systems to help users find relevant content.
- Optimize Content Production: Use data to identify potential hits and minimize production risks.
- **Personalize User Experience:** Tailor content recommendations and user interfaces to individual preferences.
- Measure and Analyze Performance: Continuously monitor key performance indicators to evaluate the effectiveness of content strategies.

By embracing a data-driven approach, Netflix can navigate the complexities of the streaming industry, maintain its leadership position, and deliver exceptional value to its subscribers.

