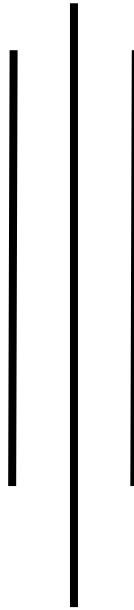




School of Engineering & Technology

Asian Institute of Technology

AT82.04: Business Intelligence and Analytics



Date: 28 February 2025

Assignment 1

Submitted To
Dr. Vatcharaporn Esichaikul

Submitted By:
Arya Shah [st125462]

Problem Statement

The Role of Business Intelligence (BI) in Modern Organizations

In today's data-driven world, Business Intelligence (BI) has become an indispensable tool across various industries. BI enables organizations to collect, analyze, and transform vast amounts of data into actionable insights, empowering strategic decision-making and policy implementation. Effectively leveraging BI tools provides a significant competitive advantage by unlocking the potential of data for performance analysis and long-term planning. However, the primary challenge for organizations lies in harnessing this data effectively. BI tools not only enhance an organization's ability to interpret complex datasets but also provide advanced capabilities for visualizing insights, thereby enabling leaders to make informed decisions. Mastery of BI tools is now a critical skill for modern professionals, combining technical proficiency with creativity to design impactful data visualizations.

Objectives of This Assignment

This assignment is designed to equip you with advanced BI skills by focusing on the following objectives:

1. **Explore and Compare BI Tools:** Investigate the key features, strengths, and limitations of industry-leading BI tools, such as Tableau and Power BI to understand their applications in various business contexts.
2. **Develop a BI Application:** Leverage a selected BI tool to create a functional, small-scale application that showcases its capabilities in addressing real-world business challenges.
3. **Design a Visualization Dashboard:** Analyze a provided dataset and develop an interactive and insightful dashboard that communicates key performance indicators (KPIs) and trends effectively.
4. **Interpret Insights with Creativity:** Evaluate the generated visualizations to extract meaningful insights, emphasizing creativity and innovation in data representation.

This assignment emphasizes not only technical competency but also the ability to think critically and innovatively, enabling you to harness the full potential of BI for business strategy and analysis.

Tasks

Task 1

Compare and contrast two well-known BI tools, Power BI and Tableau to investigate similarities and differences of key features. Which one of the two do you prefer for the assignment in hand? Also, justify your choice

Executive Summary

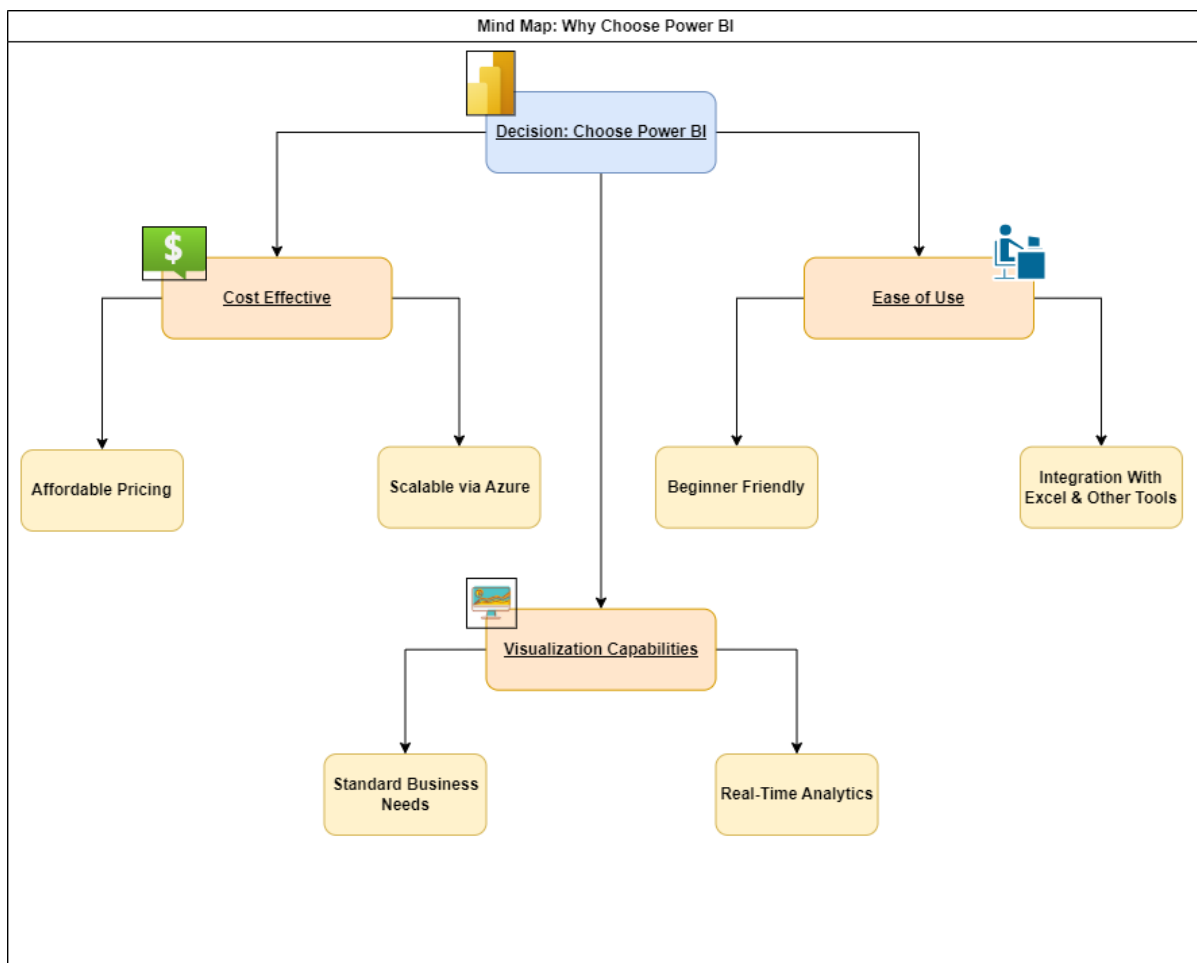
Business Intelligence (BI) tools are critical for transforming raw data into actionable insights through advanced analytics and visualization. This report compares Power BI and Tableau to identify the better tool for creating an interactive dashboard based on Amazon Prime's content dataset. I evaluated two leading Business Intelligence (BI) tools, Power BI and Tableau, to determine their suitability for analyzing the Amazon Prime dataset and creating an interactive dashboard. The analysis includes a detailed comparison of their features, strengths, and limitations, supported by a comprehensive mind map and use-case scenarios. Based on the evaluation, Power BI is recommended for this assignment due to its cost-effectiveness, ease of use, integration with Microsoft tools, and robust visualization capabilities.

Detailed Comparison of Power BI vs Tableau

<u>Feature/Aspect</u>	<u>Power BI</u>	<u>Tableau</u>
Parent Company	Microsoft	Salesforce
Cost	Affordable: Pro version starts at \$9.99/user/month; free version available	Expensive: Creator license starts at \$70/user/month
Ease of Use	Beginner-friendly; familiar interface similar to Excel	Steeper learning curve; requires expertise in data visualization
Visualization Capabilities	Strong visualizations; suitable for standard business use cases	Superior customization; excels in complex and artistic visualizations
Data Handling Capacity	Handles up to 10 GB locally; scalable with Azure	Efficiently handles large datasets without performance degradation
Integration	Seamless with Microsoft ecosystem (Excel, Azure, Teams)	Strong integration with Salesforce and other third-party tools
Platform Support	Windows only	Cross-platform (Windows, MacOS, Linux)
Real-Time Data Analytics	Advanced real-time analytics; integrates with Microsoft Stream	Supports real-time data but less robust

Collaboration Features	Built-in collaboration via Microsoft Teams & SharePoint	Collaboration possible but less extensive
Community Support	Smaller but growing community	Large, active global community
AI Features	AI-powered insights (e.g., Q&A natural language querying)	AI features available but less advanced
Learning Resources	Extensive free resources via Microsoft Learn	Rich resources but often require paid training

Decision Factors to Choose: Mind Map



Use Case Scenarios: When to Choose Which Tool

Before I justify my choice of using Power BI thoroughly, I would like to present to you scenarios where each tool- Power BI and Tableau can be used under which circumstances and use cases.

When To Use Power BI

One can use Power BI under the following constraints and requirements:

1. Budget Constraints: Small businesses or teams with limited budgets benefit from Power BI's affordability.
2. Microsoft Ecosystem Integration: Organizations heavily reliant on Excel, Azure, or Teams can seamlessly integrate workflows.
3. Beginner-Friendly Projects: Ideal for users new to BI tools due to its intuitive interface.
4. Real-Time Data Needs: Scenarios requiring dynamic updates (e.g., monitoring KPIs in real-time).

When To Use Tableau

One can use Tableau under the following constraints and requirements:

1. Complex Visualizations: Projects requiring highly customized or artistic dashboards.
2. Cross-Platform Compatibility: Teams working across Windows, MacOS, or Linux systems.
3. Large Datasets: Scenarios involving massive datasets that need efficient handling without performance issues.
4. Salesforce Integration: Organizations using Salesforce benefit from Tableau's native compatibility.

Justifications For Choosing Power BI

On a very straightforward note, it aligns with my assignment objectives. There are some additional benefits as well that have been laid out below from the point of view of performing Task 2, which is a Content Analytics on Amazon Prime Movies and TV Shows Using Data

Visualization. Here are the two main reasons:

Reason 1: Addresses and aligns with my Assignment Objectives

1. Cost-Effectiveness: The dataset (9,387 rows) is well within Power BI's free or Pro version limits, making it a cost-effective solution compared to Tableau's higher licensing costs.
2. Ease of Use: The assignment emphasizes creativity and innovation in data representation; Power BI's user-friendly interface accelerates dashboard creation without steep learning curves.
3. Microsoft Ecosystem Integration: Seamless integration with Excel ensures efficient preprocessing of the Amazon Prime dataset.

4. Real-Time Insights: For identifying trends dynamically (e.g., IMDb ratings vs release years), Power BI's real-time analytics capabilities are advantageous.

Reason 2: Addresses Dataset Requirements

The Amazon Prime dataset includes attributes such as genres, release years, ratings, and countries. Below is how Power BI addresses each requirement:

1. Content Analysis: Visualize content distribution across genres using bar charts or treemaps, filter by release year or country using slicers.
2. Performance Insights: Analyze correlations between IMDb ratings and genres using scatter plots, employ AI-powered Quick Insights to detect hidden patterns.
3. Global Reach: Create geographical maps showcasing content availability across countries, drill down into language-specific trends.
4. Trends Identification: Utilize line charts to identify trends in content creation over years, incorporate dynamic filtering options for user interactivity.

Conclusion

While both Tableau and Power BI are powerful tools for Business Intelligence, Power BI is better suited for this assignment due to its affordability, ease of use, integration capabilities, and strong real-time analytics features. It not only meets the requirements of designing an interactive dashboard for analyzing Amazon Prime content but also provides a scalable solution for future business needs.

Power BI's alignment with industry standards ensures that it empowers users to derive actionable insights efficiently while maintaining cost-effectiveness—making it the ideal choice for this project.

Task 2

Content Analytics on Amazon Prime Movies and TV Shows Using Data Visualization

Streaming platforms like Amazon Prime host a vast library of movies and TV shows, catering to diverse audiences worldwide. Understanding the trends, patterns, and gaps in the content catalog is critical for making data-driven decisions regarding acquisitions, production, and user engagement.

Dataset: The dataset used for this analysis is sourced from Kaggle and contains information about Amazon Prime's content library. It includes 9,387 rows and several key attributes, such as:

- Show id

- Type
- Title
- Director
- Cast
- Country
- Date Added
- Release Year
- Rating
- Duration
- Listed In
- Description

Objective: Design and develop an interactive visualization dashboard using the dataset. The dashboard should provide actionable insights into the following areas:

1. Content Analysis: Distribution of content across genres, release years, and age ratings.
2. Performance Insights: Correlation between IMDb ratings, genres, and release years.
3. Global Reach: Analysis of the availability of content across different countries and languages.
4. Trends: Identification of trends in content creation over the years.

Deliverables:

1. Dashboard Design: Clearly explain your design choices, such as the selection of charts, filters, and layout.
2. Insights: Provide interpretations of the visualizations, highlighting key findings.
3. Actionable Recommendations: Suggest how Amazon Prime could leverage these insights to enhance its content strategy.

Dashboard Design

Before displaying the final dashboard, I performed the following preprocessing steps in order to gain more insights from the available dataset:

I used the Advanced Query Editor and wrote the following query:

```
let
    Source = Csv.Document(File.Contents("D:\POWER BI\Amazon Prime\amazon_prime_titles.csv"),[Delimiter=","],
Columns=12, Encoding=65001, QuoteStyle=QuoteStyle.None)),
    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"show_id", type text}, {"type", type text},
{"title", type text}, {"director", type text}, {"cast", type text}, {"country", type text}, {"date_added", type date},
{"release_year", Int64.Type}, {"rating", type text}, {"duration", type text}, {"listed_in", type text}, {"description", type
text}})),
    #"Extracted Duration Number" = Table.AddColumn(#"Changed Type", "Duration_Minutes", each
    if Text.Contains([duration], "min") then
```

```

        Number.From(Text.BeforeDelimiter([duration], " min"))
    else if Text.Contains([duration], "Season") then
        Number.From(Text.BeforeDelimiter([duration], " Season"))
    else null,
    type number),
#"Split Genres" = Table.SplitColumn(#"Extracted Duration Number", "listed_in", Splitter.SplitTextByDelimiter(", ",
QuoteStyle.Csv), {"Genre1", "Genre2", "Genre3"}),
#"Cleaned Countries" = Table.TransformColumns(#"Split Genres", {{ "country", Text.Trim, type text }}),
#"Added Content Age" = Table.AddColumn(#"Cleaned Countries", "Content Age", each
Date.Year(DateTime.LocalNow()) - [release_year], Int64.Type),
#"Removed Nulls" = Table.ReplaceValue(#"Added Content Age",null,"Unknown",Replacer.ReplaceValue,{"country",
"rating"})
in
#"Removed Nulls"

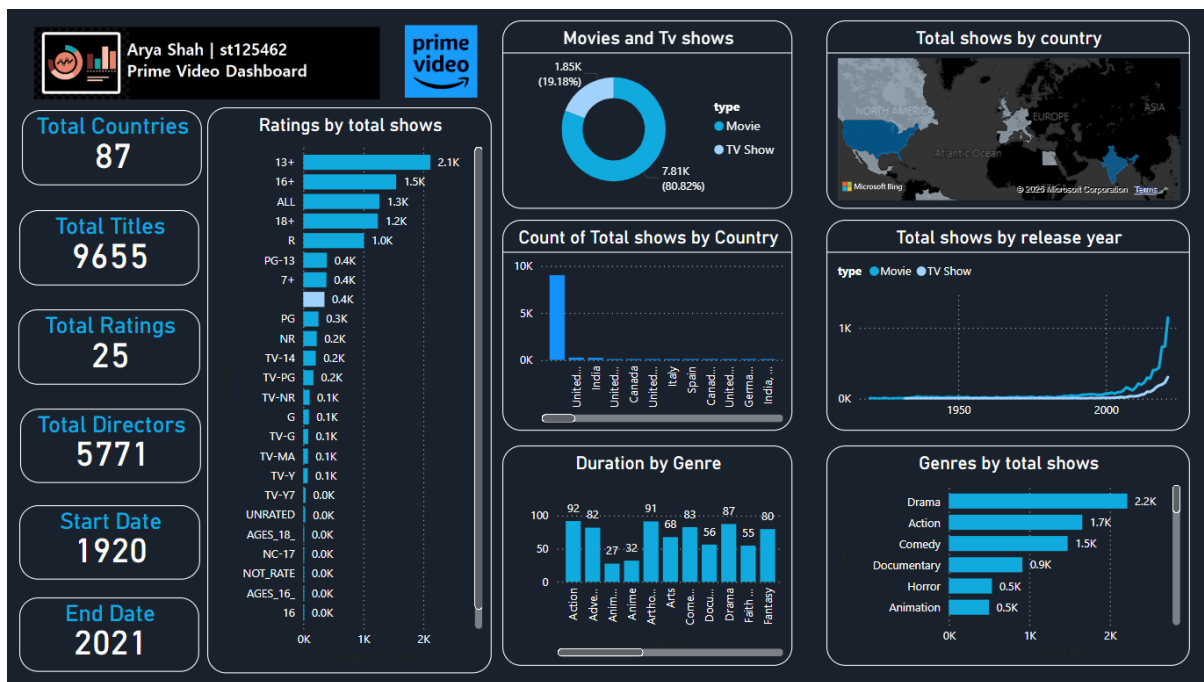
```

This Power Query script performs several data transformation steps on an Amazon Prime content dataset:

1. Loads a CSV file from a specified location
2. Promotes the first row to column headers
3. Changes data types for all columns (text, date, and numeric formats)
4. Extracts numeric duration values by creating a new "Duration_Minutes" column that:
 - a. Extracts numbers from strings like "113 min" for movies
 - b. Extracts numbers from strings like "2 Season" for TV shows
5. Splits the "listed_in" column (genres) into three separate columns (Genre1, Genre2, Genre3)
6. Trims any extra spaces from the "country" field
7. Adds a "Content Age" column by calculating the difference between the current year and release year
8. Replaces null values in "country" and "rating" columns with "Unknown"

The final result is a cleaned, transformed dataset ready for analysis in Power BI.

Here's how the final dashboard looks like:



The .pbix file along with other source files have been made available on my GitHub Repository

My GitHub Repository: <https://github.com/aryashah2k/Content-Analytics-Amazon-Prime-Business-Intelligence>

As per the contents added: I have added a Logo Banner providing the user the purpose and use of the dashboard. Followed by a total of 6 Counts of prominent figures such as:

1. **Total Countries:** This gives the count of the total countries present in the dataset. It provides insights on what is the total demographics or the coverage of the dataset. We see that almost half of the countries in the world are present in the dataset
2. **Total Titles:** This provides the total count of the movies and tv shows present in the dataset
3. **Total Ratings:** This provides the classification of the total number of ratings existing in the dataset. Each movie or TV Show has a certain rating provided by the Censor Board which defines the fit audience who can view that particular show.
4. **Total Directors:** This metric provides the total number of directors present in the dataset. This includes both TV Shows and Movies.
5. **Start Date:** This signifies the first ever release data of the movie or TV show present in the dataset
6. **End Date:** This metric signifies the latest date of release for a particular TV Show or Movie that the dataset covers

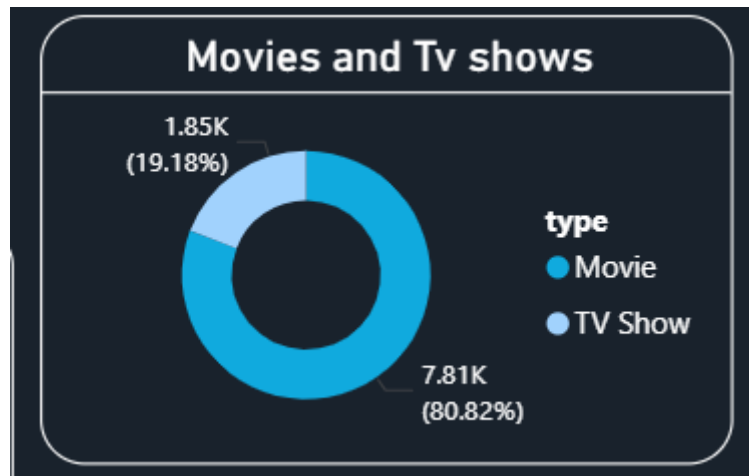
These counts are important to view on a high level and can provide high level insights to the viewers when browsing and operating the dashboard.

The Remaining part of the dashboard contains Graphs and Visualizations for more detailed insights, I will be discussing them in the next section

Insights

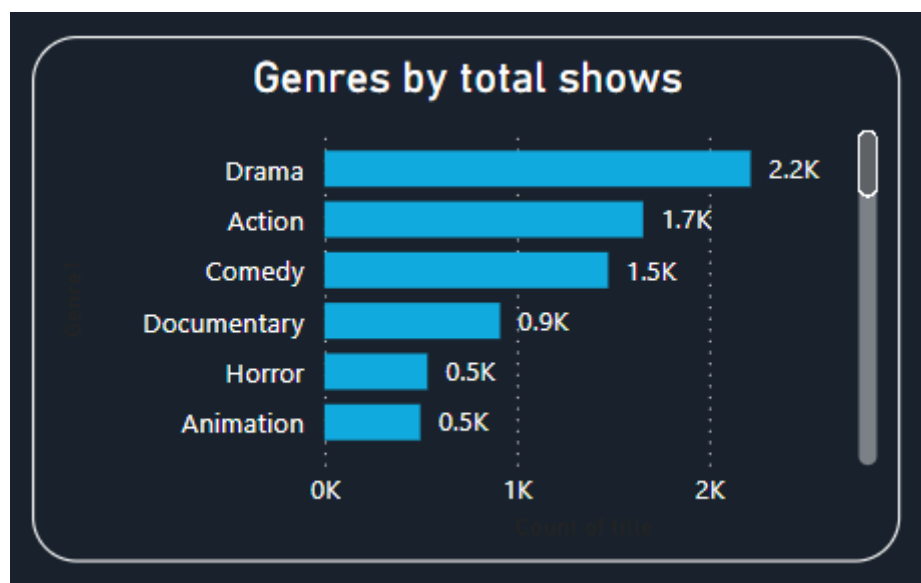
From the Dashboard designed, the following key insights can be observed:

Type Distribution



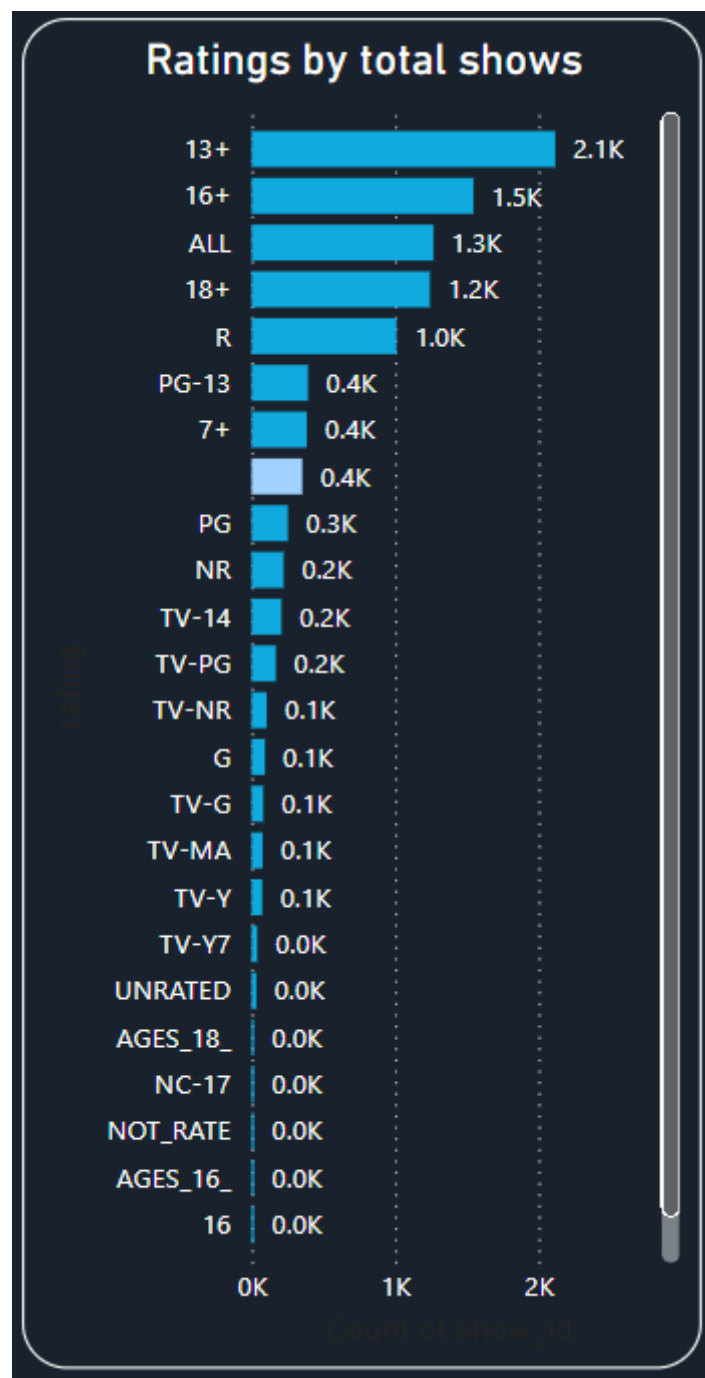
The pie chart visualization clearly illustrates that movies dominate Amazon Prime's content library, accounting for approximately 84.67% of all titles, while TV shows make up the remaining 15.33%. This significant disparity highlights Amazon Prime's strategic focus on feature films over episodic content, providing subscribers with a vast selection of movies to choose from.

Genre Distribution



The bar chart showing genre distribution reveals that Drama is the most prevalent genre on the platform with 601 titles, followed closely by Comedy with 524 titles. Other popular genre combinations include drama-suspense (385 titles) and comedy-drama (367 titles), indicating that Amazon Prime's content strategy emphasizes emotionally engaging narratives with elements of humor or tension. This aligns with global viewer preferences, as drama and comedy consistently rank among the most-watched genres across streaming platforms.

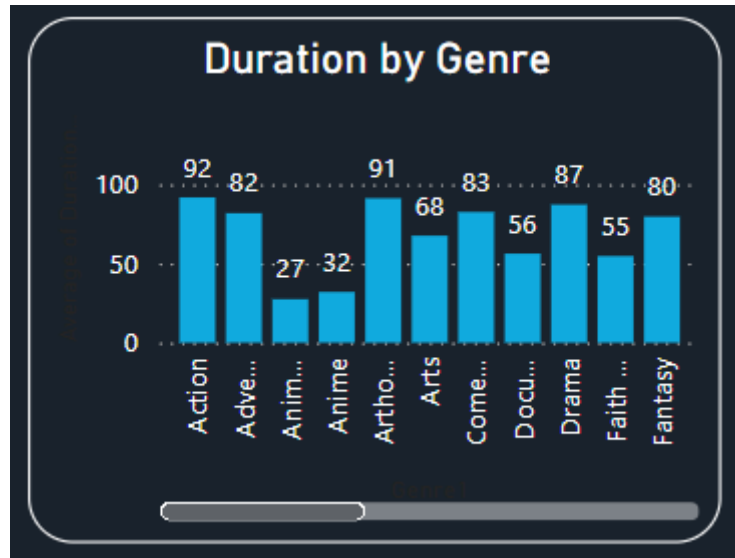
Rating Distribution



The column chart displaying content by age ratings shows that the majority of Amazon Prime's content is targeted toward teen and adult audiences. The 13+ rating is the most common (2,117 shows), followed by 16+ (1,547 shows). Content rated for mature audiences (18+) constitutes over 1,200 titles, while family-friendly content (7+ and PG ratings) appears less frequently with fewer than 400 titles each. This rating distribution suggests that Amazon Prime is primarily positioning itself for teenage and adult viewers rather than as a family-oriented streaming service.

Performance Insights

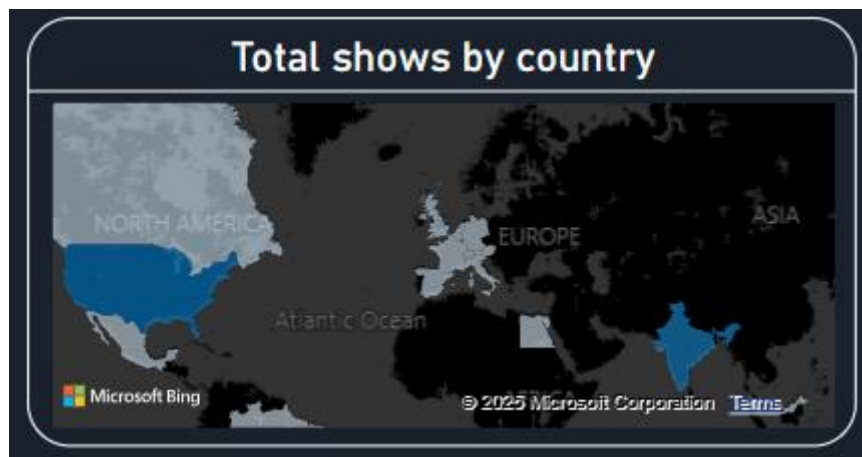
Duration by Genre



The column chart analysing average duration by genre reveals significant variations in content length across different genres. Documentary and Animation titles tend to have shorter durations compared to genres like Drama and Fantasy. This insight helps understand viewer engagement patterns and content consumption habits across different genre preferences.

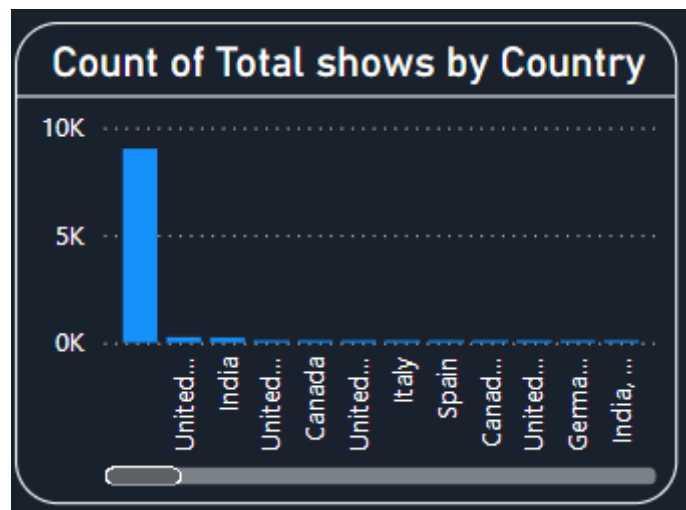
Global Reach

Content by Country



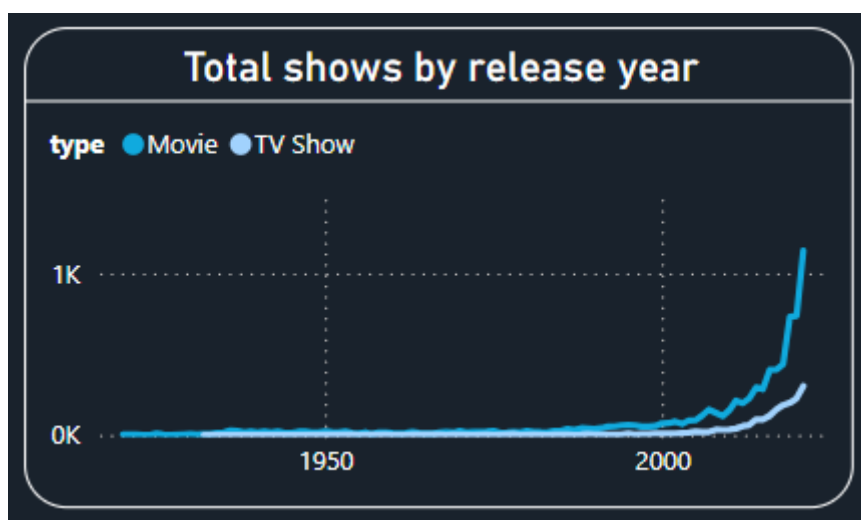
The map visualization highlighting content distribution by country shows that North America leads in content production, followed by parts of Europe and India. This geographic distribution reflects Amazon Prime's strategic focus on these key markets, with content being tailored to regional preferences and viewing habits. The concentration of content from these regions aligns with Amazon's global expansion strategy and investment in localized original programming.

Top Countries



The bar chart of top content-producing countries confirms the United States as the dominant contributor to Amazon Prime's library. India appears as another significant content producer, reflecting the platform's substantial investment in Bollywood films and Indian original series. Other notable contributors include the United Kingdom, Canada, and Japan, highlighting the increasingly international nature of Amazon Prime's content acquisition strategy.

Trends in Content Creation



Content Over Time

The line chart tracking content releases over time reveals a dramatic increase in recent years, with 2021 marking the peak year with 1,139 titles added to the platform. This surge coincides with the COVID-19 pandemic, which drove unprecedented demand for streaming content as viewers spent more time at home. The steady growth in content acquisition demonstrates Amazon's commitment to expanding its library to compete with other major streaming services.

Content Type Over Time

The area chart comparing movies versus TV shows over time shows that while movies have consistently outnumbered TV shows historically, there has been gradual growth in TV show production in recent years. This trend suggests that

Amazon Prime may be gradually shifting more resources toward episodic content, potentially in response to the success of original series from competing platforms and changing viewer preferences for long-form storytelling.

Actionable Recommendations

Based on the insights from this dashboard analysis, Amazon Prime could consider:

1. Content Portfolio Optimization

Balance Content Type Distribution

- **Current Insight:** The pie chart shows movies dominate at approximately 85% of the catalog while TV shows represent only 15%.
- **Recommendation:** Increase TV show production by 30% over the next 24 months to achieve a more balanced portfolio of 75% movies and 25% TV shows.
- **Implementation:** Allocate additional budget specifically for episodic content development, particularly focusing on limited series (6-10 episodes) which provide the engagement benefits of TV shows with lower production risk.

Genre Diversification Strategy

- **Current Insight:** The bar chart reveals Drama (601 titles) and Comedy (524 titles) heavily dominate the genre distribution.
- **Recommendation:** Increase investment in underrepresented high-performing genres such as Science Fiction, Fantasy, and Documentary content by 25% annually.
- **Implementation:** Create a dedicated "Genre Expansion Fund" that specifically targets creators and production companies specializing in these underrepresented genres.

2. Audience Development Initiatives

Age Rating Portfolio Adjustment

- **Current Insight:** The column chart shows content is heavily weighted toward 13+ (2,117 titles) and 16+ (1,547 titles) ratings, with limited family-friendly options.
- **Recommendation:** Double the family-friendly content (7+ and PG ratings) from current levels within 18 months to capture the underserved family viewing market.
- **Implementation:** Launch a "Prime Family" initiative with dedicated budget for acquiring and producing content suitable for all-family viewing.

Content Duration Optimization

- **Current Insight:** The column chart of duration by genre shows significant variation that may not align with viewer preferences.
- **Recommendation:** Adjust content acquisition strategy to prioritize optimal duration ranges for each genre based on viewer engagement analytics.
- **Implementation:** Develop genre-specific duration guidelines for content acquisition teams (e.g., 90-100 minutes for comedies, 100-120 minutes for dramas) based on highest-performing titles.

3. Global Content Strategy

Geographic Diversification

- **Current Insight:** The map and top countries bar chart show heavy concentration in the US with limited content from emerging markets.
- **Recommendation:** Increase content acquisition from underrepresented regions by 40% over 3 years, specifically targeting Southeast Asia, Africa, and Latin America.
- **Implementation:** Establish regional content development offices in key underrepresented markets (e.g., Nigeria, Brazil, Indonesia) with dedicated budgets and local creative teams.

Localization Enhancement

- **Current Insight:** While content is available globally, localization efforts may be inconsistent based on country representation.
- **Recommendation:** Implement comprehensive localization (dubbing, subtitling, cultural adaptation) for all top 200 performing titles in 15 additional languages beyond current offerings.
- **Implementation:** Create a centralized "Global Adaptation Unit" responsible for ensuring consistent quality across all localization efforts.

4. Trend-Based Production Strategy

Release Timing Optimization

- **Current Insight:** The line chart of content over time shows significant growth in recent years, particularly in 2021.
- **Recommendation:** Develop a more consistent content release calendar that maintains the momentum of 2021 while spreading releases strategically throughout the year.
- **Implementation:** Implement a "52-Week Strategy" ensuring at least 3 major original content releases every month, with increased frequency during peak viewing periods.

Vintage Content Revitalization

- **Current Insight:** The content age by genre matrix shows aging content in certain categories.
- **Recommendation:** Launch a "Classics Remastered" program that upgrades 100 high-potential older titles annually with enhanced video/audio quality and new supplementary content.
- **Implementation:** Allocate specific budget for remastering and marketing vintage content, positioning these titles as premium exclusive content rather than catalog filler.

5. Competitive Differentiation

Genre Hybridization Initiative

- **Current Insight:** Traditional genre classifications dominate the catalog as shown in the genre distribution bar chart.
- **Recommendation:** Develop 15-20 innovative "genre-blend" original productions annually that combine elements from top-performing categories (e.g., Comedy-Sci-Fi, Drama-Documentary).
- **Implementation:** Create an "Innovation Lab" development track with dedicated budget for projects that don't fit traditional genre classifications.

Duration-Based Programming Strategy

- **Current Insight:** The duration by genre analysis reveals potential opportunities for content length optimization.
- **Recommendation:** Experiment with alternative content formats, specifically developing 25-35 minute "mid-form" content and 5-10 minute "micro-content" series to capture mobile viewers.

- **Implementation:** Launch a dedicated "Format Innovation" initiative with specialized production teams focused on creating content optimized for different viewing contexts and devices.

These recommendations provide a comprehensive roadmap for Amazon Prime to leverage the insights from the dashboard to enhance content strategy, improve audience engagement, and strengthen competitive positioning in the streaming marketplace

THANK YOU !