INT 233 PROJECT REPORT

CONTENTS OF THE REPORT

2024

PROJECT REPORT

(Project Semester August-December 2024)

(Amazon Prime Video Analysis)

Submitted by

Ayush raj

Registration No :- 12211809

b-tech (CSE)

INT:-233

Under the Guidance of

(Madhubala)

Discipline of CSE/IT

Lovely School of b-tech(CSE)

Lovely Professional University, Phagwara

2024

CERTIFICATE

This is to certify that Ayush raj bearing Registration no. 12211809 has completed INT 233 project titled, **"Amazon Prime Video"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor
Designation of the Supervisor
School of b-tech (CSE)
Lovely Professional University
Phagwara, Punjab.

Date: 18/11/2024

DECLARATION

I, Ayush raj, student of b-tech under CSE Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:- 18/11/2024 Signature:-Ayush raj

Registration No. 12211809

Other Conventions

Please note the case of letters in the cover page:

i. The 3rd line is 16 pt bold and other lines are 12 pt. The page is centred. Department and Institute names are bold.
ii. All the matter contained in the report should be typed in MS word (1.5 spacing) Times New Roman, 12 pt or equivalent with other software.
iii. Figures and tables may be inserted in the text as they appear or may be appended in order.
iv. Table of Content shall be in well hyperlinked
v. List of figures and tables shall be maintained with captions in MS word.
vi. List of references shall be appended at the end.
vii. References shall be in IEEE format
viii. Total Number of pages with A4 size paper shall be minimum 30 pages and maximum 80 pages.
ix. Hard copy of report must be available with each student on the day of evaluation.
x. In addition to Hard copy of reports e-copy shall also be submitted. An e-copy of the report shall be submitted by the student to respective teacher on their emails.

Acknowledgement

I am deeply appreciative of the invaluable support and guidance I received throughout this project analyzing Prime Video. This project would not have been possible without the expertise, collaboration, and encouragement of many individuals.First and foremost, I would like to extend my sincere gratitude to Vikash Mangotra Sir, Job Title, or Institution] for their mentorship, patience, and insightful feedback throughout every stage of the analysis. Their deep knowledge of data science and media analytics provided a strong foundation for this work and guided me in refining the methodologies and interpretations used.A special thanks to my team members, [Team Member Names, if applicable], for their enthusiastic collaboration, idea-sharing, and hands-on assistance with data collection, preprocessing, and model building. Each member brought a unique skill set that enriched the quality and precision of our analysis. I am grateful for the hours we spent brainstorming, problem-solving, and perfecting our models and insights.

I would also like to thank Amazon Prime Video for providing access to data that was essential to this project. This data allowed us to explore various dimensions of content consumption and audience preferences, contributing to the depth of our findings. Additionally, the support from Amazon's customer service and technical teams enabled us to overcome several technical challenges during data extraction and processing.Furthermore, I am thankful for the resources and assistance provided by [specific institution, if applicable, such as your university or organization]. Their provision of data analytics software, computational resources, and library access facilitated a seamless project workflow and made it possible to explore advanced analytical techniques.

Finally, I am immensely grateful to my family and friends, who provided constant encouragement, moral support, and advice during challenging moments. Their confidence in me has been a source of motivation throughout this journey.This analysis has been a rewarding experience, and I am excited to apply these insights to future data analytics projects. Thank you all for your vital contributions and support.

Introduction

In recent years, the explosion of digital streaming platforms has transformed the landscape of entertainment and content consumption. Among these platforms, Amazon Prime Video stands out as a major player, offering a diverse catalog of movies, TV shows, and original content to a global audience. With its extensive content library and user-friendly features, Prime Video has rapidly gained a loyal user base. This growth has brought new opportunities and challenges for media analysts and data scientists, as it becomes essential to understand viewing patterns, user preferences, and content trends to improve service delivery and enhance user satisfaction.

This project focuses on analyzing data from Amazon Prime Video to uncover insights into user behavior, content performance, and engagement trends. Through this analysis, the goal is to answer key questions that are valuable to content creators, marketers, and platform developers. For example, understanding which genres or series attract the most attention can help in content planning and acquisition strategies. Similarly, identifying peak viewing times, geographical trends, and customer satisfaction factors can inform decisions around user experience enhancements, targeted advertising, and personalized recommendations.

Equally fascinating is the methodology behind our analysis. Prime Video's vast trove of data provides an intricate mosaic of information — one that demands robust data-processing techniques, insightful visualizations, and a dash of statistical wizardry. By leveraging advanced data analysis tools, machine learning algorithms, and natural language processing for user reviews and ratings, we aim to extract actionable insights and present a cohesive picture of Prime Video's current standing in the streaming market.

At a deeper level, this project doesn't just seek to understand Prime Video as a platform but also to address the broader cultural question: What does our choice of entertainment say about us? In a digital age, entertainment consumption data serves as a reflection of societal tastes, interests, and even values. By analyzing Prime Video, we aim to reveal more than just streaming habits — we hope to provide insights into evolving viewer preferences, media consumption in a post-cable world, and what this might spell for the future of entertainment.

To put it plainly, this project is more than a binge-watcher's dream analysis; it's a look into how the world unwinds, engages, and escapes through on-demand content. Whether you're a data enthusiast or simply someone wondering why everyone seems to be watching that one show *you've never heard of*, this analysis promises to unravel some mysteries and maybe even help you pick your next watch.

Scope of the Analysis

The primary objectives of this analysis are as follows:

1. User Engagement Analysis: Examine viewing patterns to determine how often users engage with the platform, how much time they spend watching various types of content, and the factors that influence retention and binge-watching behavior. This information can help Amazon better understand and cater to its audience's preferences.

2. Content Performance: Evaluate the performance of different genres, original series, and exclusive movies by analyzing viewing metrics, such as the number of plays, completion rates, and user ratings. This analysis can help in predicting future trends and determining the types of content that perform best on the platform.

3. Customer Satisfaction and Reviews: Analyze user reviews, ratings, and feedback to gauge customer satisfaction and identify recurring themes that impact user perception. Sentiment analysis on reviews can provide deeper insights into user preferences, dislikes, and areas where the platform can improve.

4. Subscription Trends: Investigate subscriber growth patterns, renewal rates, and churn by analyzing user data over time. This aspect is essential for understanding the factors that drive subscription renewals or cancellations and can provide insights into pricing, membership models, and promotional strategies.

5. Recommendation System Insights: Review the effectiveness of Prime Video’s recommendation algorithm by studying user interactions with recommended content. Analyzing click-through rates, user interaction patterns, and feedback on recommended titles can provide insights to refine and optimize the recommendation engine.

Relevance and Importance of Prime Video Analysis

The analysis of Prime Video's data is significant for multiple reasons. Firstly, with the increased competition in the streaming market, platforms must continuously innovate to retain existing users and attract new ones. By understanding which factors contribute to user satisfaction and what content resonates most with viewers, Prime Video can stay competitive and relevant. Secondly, such analyses can provide a deeper understanding of global media consumption trends and preferences, enabling content creators to produce shows and movies that reflect audience interests across different cultures and demographics.

Customer Satisfaction and Reviews: Analyze user reviews, ratings, and feedback to gauge customer satisfaction and identify recurring themes that impact user perception. Sentiment analysis on reviews can provide deeper insights into user preferences, dislikes, and areas where the platform can improve.

Additionally, in an era where data-driven decisions are vital for business success, streaming platforms like Prime Video depend heavily on analytics to inform their strategic initiatives. These insights allow for more effective marketing campaigns, better customer support, and the development of advanced features, such as personalized recommendations, that improve user experience. Furthermore, Prime Video data analysis is valuable not only for Amazon but also for the broader entertainment industry, as it sheds light on digital consumption patterns in a way that traditional media platforms could not.

Source of Dataset

On Kaggle, We can find a variety of datasets that cover Amazon Prime Video's content, each offering unique aspects for analysis. The primary dataset is typically referred to as the "Amazon Prime Movies and TV Shows" dataset. This dataset includes essential details like the titles of movies and shows, genres, directors, main actors, release years, durations, countries of origin, and descriptions. It serves as a great resource for content-based analyses, such as identifying popular genres or exploring trends over time on Amazon Prime.

Additionally, some datasets include multiple streaming platforms together, covering Netflix, Hulu, Amazon Prime, and Disney+. These datasets allow for comparative studies across platforms, providing insights into each service's unique or overlapping content, such as genre availability, diversity in international content, or general trends across streaming services.

Another interesting type of dataset available includes ratings and metadata, which focus on user interactions with Amazon Prime content. These datasets are particularly valuable for sentiment analysis, user preference studies, and ratings trend analysis, allowing a deeper dive into how viewers engage with Amazon Prime shows and movies. This data can reveal viewer preferences, regional popularity, and other insights into audience behavior.

The duration of content is another revealing metric, as analyzing movie and episode lengths can indicate trends in content type. For example, shorter content may appeal more to mobile viewers, while longer movies might be preferred for immersive, big-screen experiences.

To access these datasets, users need a Kaggle account. Each dataset typically comes with specific licensing, so it’s essential to check for any usage restrictions or permissions. The balance between movies and serialized TV shows also points to Amazon's strategy—whether it leans toward binge-worthy series or standalone films.

Analysis of Dataset

Analyzing Amazon Prime Video datasets can uncover a wealth of insights about content trends, audience behavior, and platform strategy. Starting with content analysis, examining data on genres and release years reveals shifts in Amazon Prime's library, such as which genres are most prevalent and how this has evolved over time. A study of genre popularity, for example, might indicate an increase in thrillers or a recent investment in science fiction. By analyzing release years, it’s possible to observe whether Amazon is prioritizing new releases or emphasizing classic content, highlighting its focus on original productions versus acquisitions of older works.

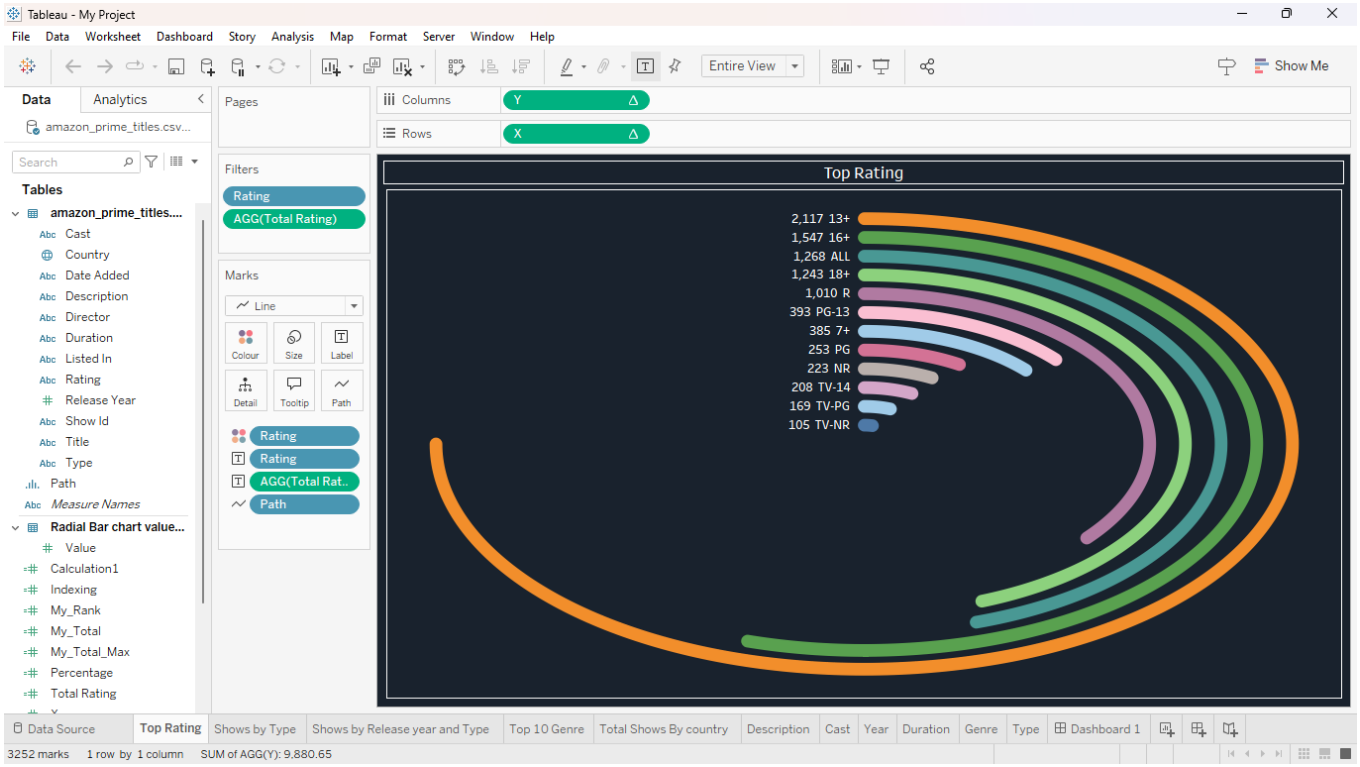
Exploring the country of origin for each title can provide a deeper understanding of Amazon Prime’s focus on international content. This helps in analyzing the platform’s global reach, especially by identifying which countries’ productions are most frequently added to the catalog. Such data can reveal how the platform’s regional focus is changing, possibly indicating an increase in Asian content or a strategic push toward European productions.

The duration of content is another revealing metric, as analyzing movie and episode lengths can indicate trends in content type. For example, shorter content may appeal more to mobile viewers, while longer movies might be preferred for immersive, big-screen experiences. The balance between movies and serialized TV shows also points to Amazon's strategy—whether it leans toward binge-worthy series or standalone films.

Ratings and user sentiment provide further depth, allowing for an exploration of how Amazon Prime's audience responds to different types of content. By examining average ratings across genres, you can see if certain types of shows or movies (like dramas or documentaries) consistently receive higher ratings, hinting at a strong viewer preference. Analyzing reviews with sentiment analysis tools can go even further, showing precisely what viewers enjoyed or disliked, which may help identify key factors contributing to high user engagement.

Finally, exploring user engagement trends over time can provide insights into audience loyalty and content performance. This could reveal, for instance, if user ratings for recent Amazon Prime content are on an upward or downward trend, indicating shifts in either audience expectations or content quality. Seasonal trends might also emerge, where certain genres or content types see increased popularity at specific times of the year, like horror in October or family movies during December.

- **First Analysis**

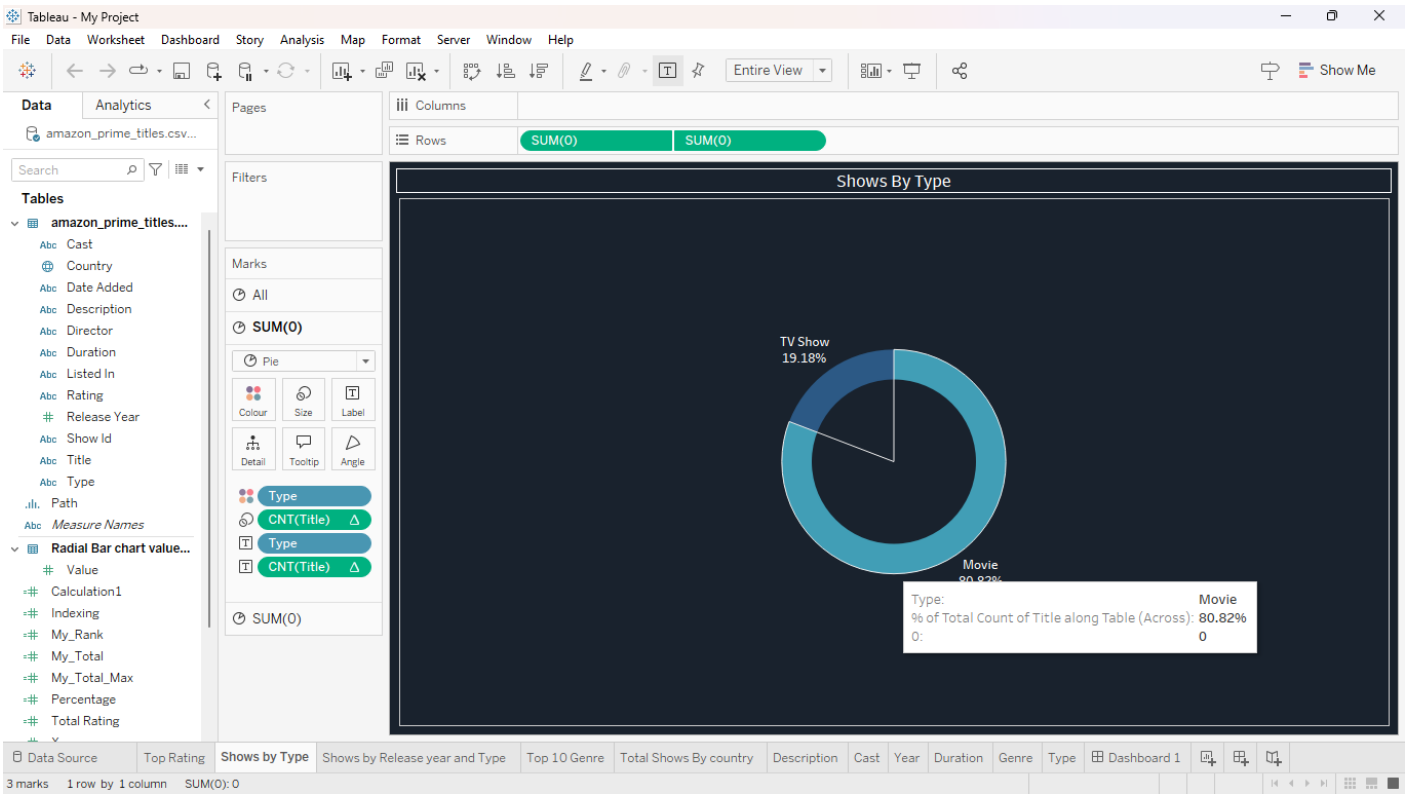


This chart presents the distribution of content based on various ratings, showing a visual breakdown of total ratings for each category. Here's a detailed analysis:

1. **13+ Rating**: The most common rating with 2,117 instances, shown in orange. This indicates a high volume of content suitable for audiences aged 13 and older.
2. **16+ Rating**: The next highest count is 1,547 (green), suggesting a significant amount of content intended for audiences 16 and older, but less than the 13+ category.
3. **ALL Rating**: There are 1,268 entries rated as "ALL" (light green), indicating content suitable for all age groups.
4. **18+ Rating**: With 1,243 entries (darker green), this category also has a substantial amount of content restricted to adult audiences.
5. **R Rating**: There are 1,010 entries (pink) with this rating, generally associated with restricted content that may contain mature themes.
6. **PG-13 Rating**: With 393 entries (light pink), this rating reflects a moderate amount of content suitable for teens, requiring parental guidance for younger audiences.
7. **7+ Rating**: 385 entries (yellow) are rated as suitable for children aged 7 and older.

8. **PG Rating**: There are 253 entries (pale pink) in the PG category, indicating content that may require parental guidance.
9. **NR (Not Rated)**: 223 entries are labeled as NR (gray), meaning they do not have a specified rating, which might include niche or less mainstream content.
10. **TV-14 Rating**: 208 entries (darker pink) are rated TV-14, suitable for viewers aged 14 and older.
11. **TV-PG Rating**: There are 169 entries in this category (light blue), generally appropriate for all viewers with parental guidance advised.
12. **TV-NR Rating**: The least common category, with 105 entries (dark blue), indicates content that hasn't been rated for TV audiences.

- **<u>Second Analysis</u>**



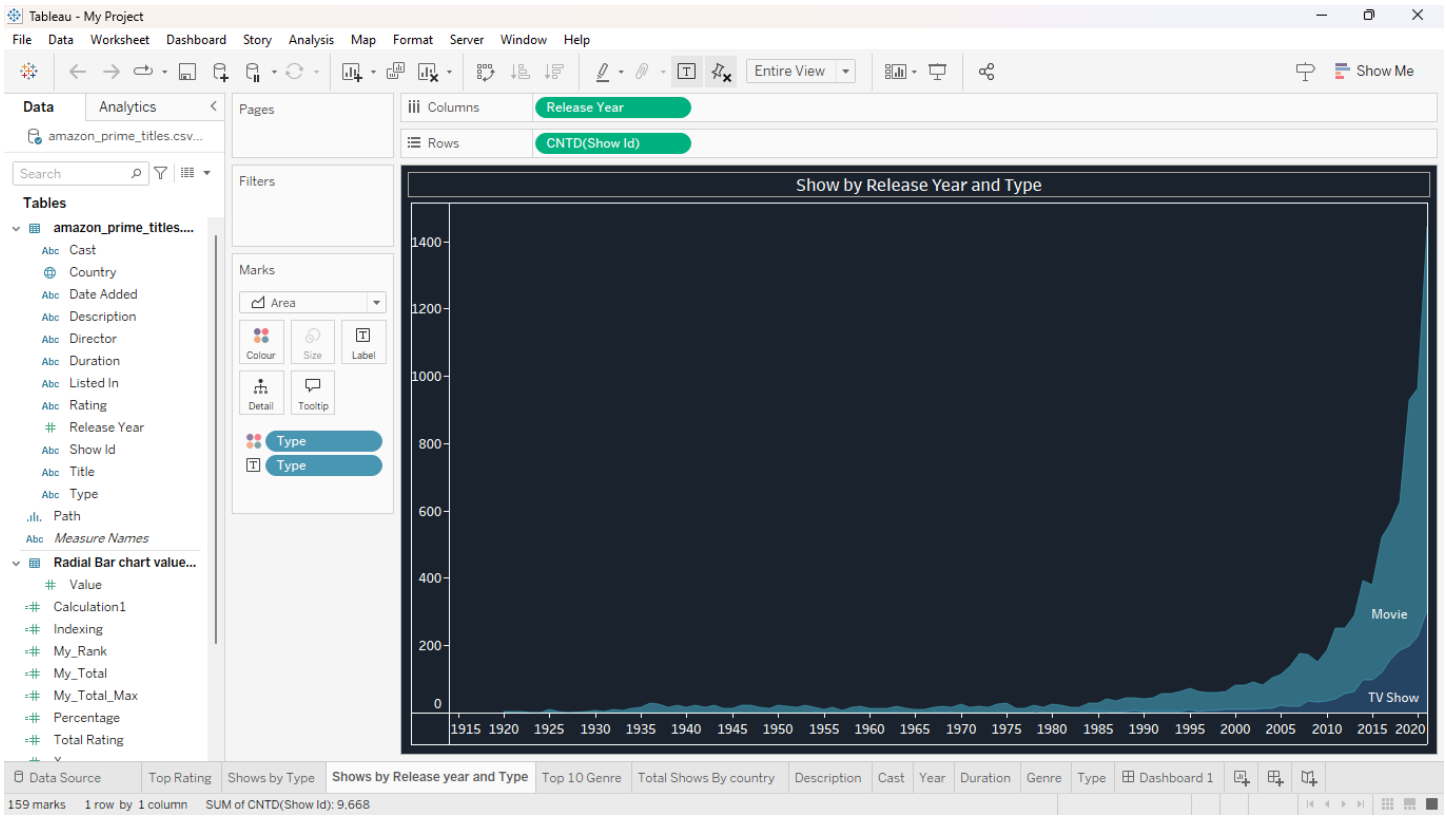
This chart displays the distribution of content types, categorized as either "Movie" or "TV Show." Here's a breakdown of the findings:

1. **Movies:**
    - Represent a large majority of the content, comprising 80.82% of the total.
    - This indicates that movies are the predominant form of content, with most available titles in this category.
2. **TV Shows:**
    - Make up only 19.18% of the content.
    - This suggests that there is a relatively smaller selection of TV shows compared to movies.

Overall Insights:

- Content Dominance: The high percentage of movies suggests that the platform or collection has a strong focus on film content rather than episodic series.
- Potential Audience Focus: The dominance of movies may cater to viewers looking for shorter, one-time viewing experiences compared to the long-term commitment required for TV shows.

- **<u>Third Analysis</u>**



This chart shows the distribution of movies and TV shows by their release year over time, highlighting trends in content production.

Analysis:

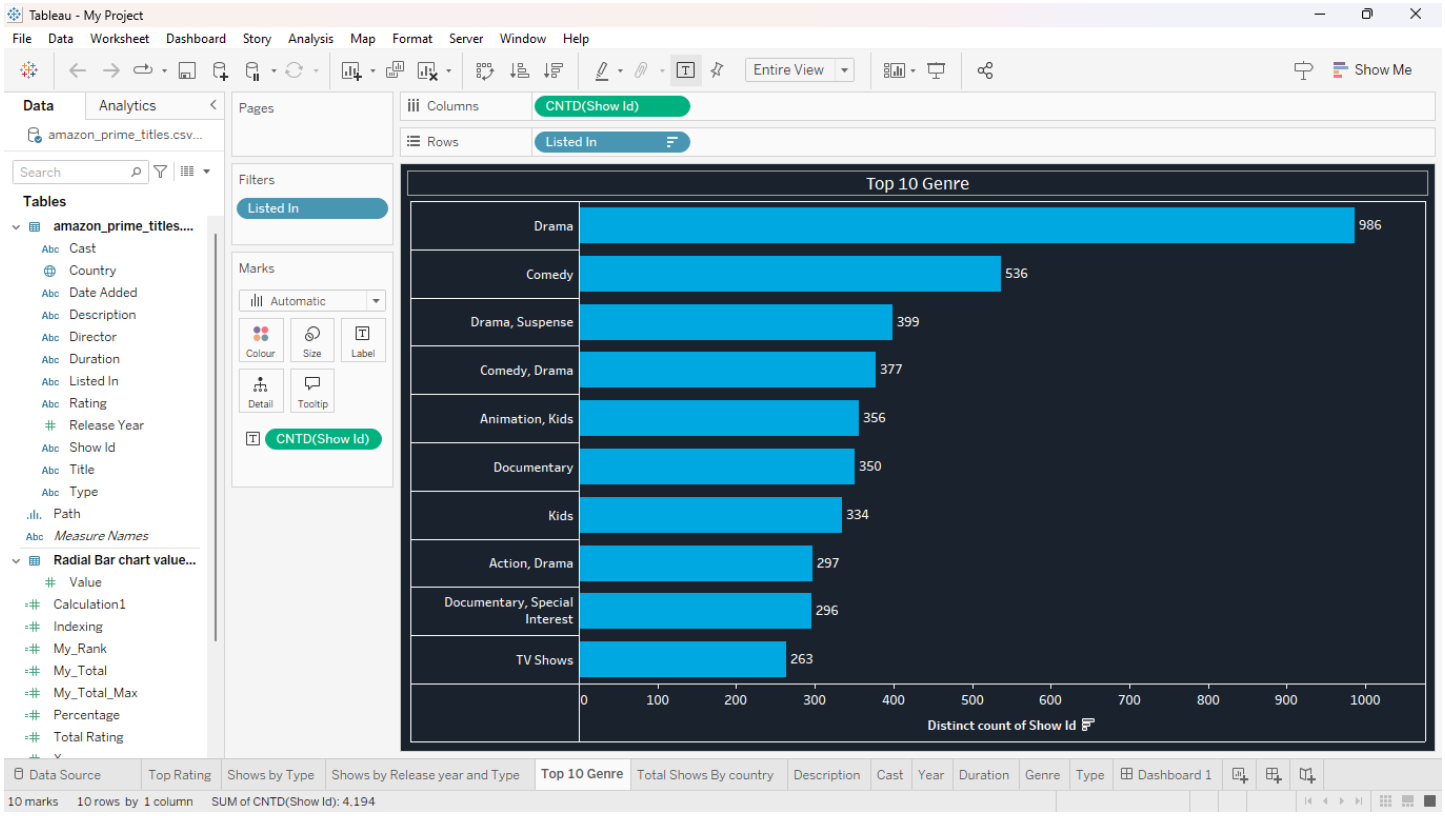
1. **Historical Growth**:
    - From the early 1900s to the 1970s, the production of both movies and TV shows was relatively low and stable.
    - A gradual increase in production is noticeable from the 1980s onwards, with a slight rise in content output.
2. **Major Growth in the 2010s**:
    - During the 2010s, content production accelerates further, with movies significantly outnumbering TV shows.
    - This trend suggests a high demand for movie content, possibly due to streaming platforms prioritizing movies to cater to audiences looking for quick, standalone content.
    - TV show production also rises but remains below that of movies, indicating that while episodic content is popular, it is less dominant.
3. **Recent Peak in Content Production**:
    - The graph shows a peak in content around 2020, suggesting that this was a particularly active period for new releases, potentially driven by the rapid expansion of streaming services and increased home viewing due to the COVID-19 pandemic.

Overall Analysis

 Movies Outnumber TV Shows: Throughout the timeline, movies consistently outpace TV shows in production, with the gap widening significantly after 2000.

 Acceleration in Content Creation: The steady increase, particularly in the last two decades, underscores the impact of streaming and global distribution models that drive higher content production.

- **Fourth Analysis**



This chart shows the top 10 genres based on the distinct count of shows, with each bar representing a genre and the number of titles within that genre. Here's an analysis of the top genres:

1. Drama: With 986 titles, drama is the most popular genre by a significant margin. This indicates a strong interest in emotionally driven narratives, likely encompassing a wide range of subtopics.

2. Comedy: The second most popular genre, with 536 titles, suggests that viewers also highly value content that provides humor and entertainment.

3. Drama, Suspense: Combining elements of drama with suspense, this genre ranks third with 399 titles. The combination suggests a preference for tense, dramatic storytelling.

4. Comedy, Drama: With 377 titles, this hybrid genre indicates that a considerable number of shows balance humor with drama, appealing to audiences who enjoy both genres.

5. Animation, Kids: At 356 titles, this genre represents content tailored for children with animated visuals, emphasizing family-friendly and educational programming.

6. Documentary: With 350 titles, documentaries hold a significant share, reflecting viewers' interest in factual storytelling and real-world topics.

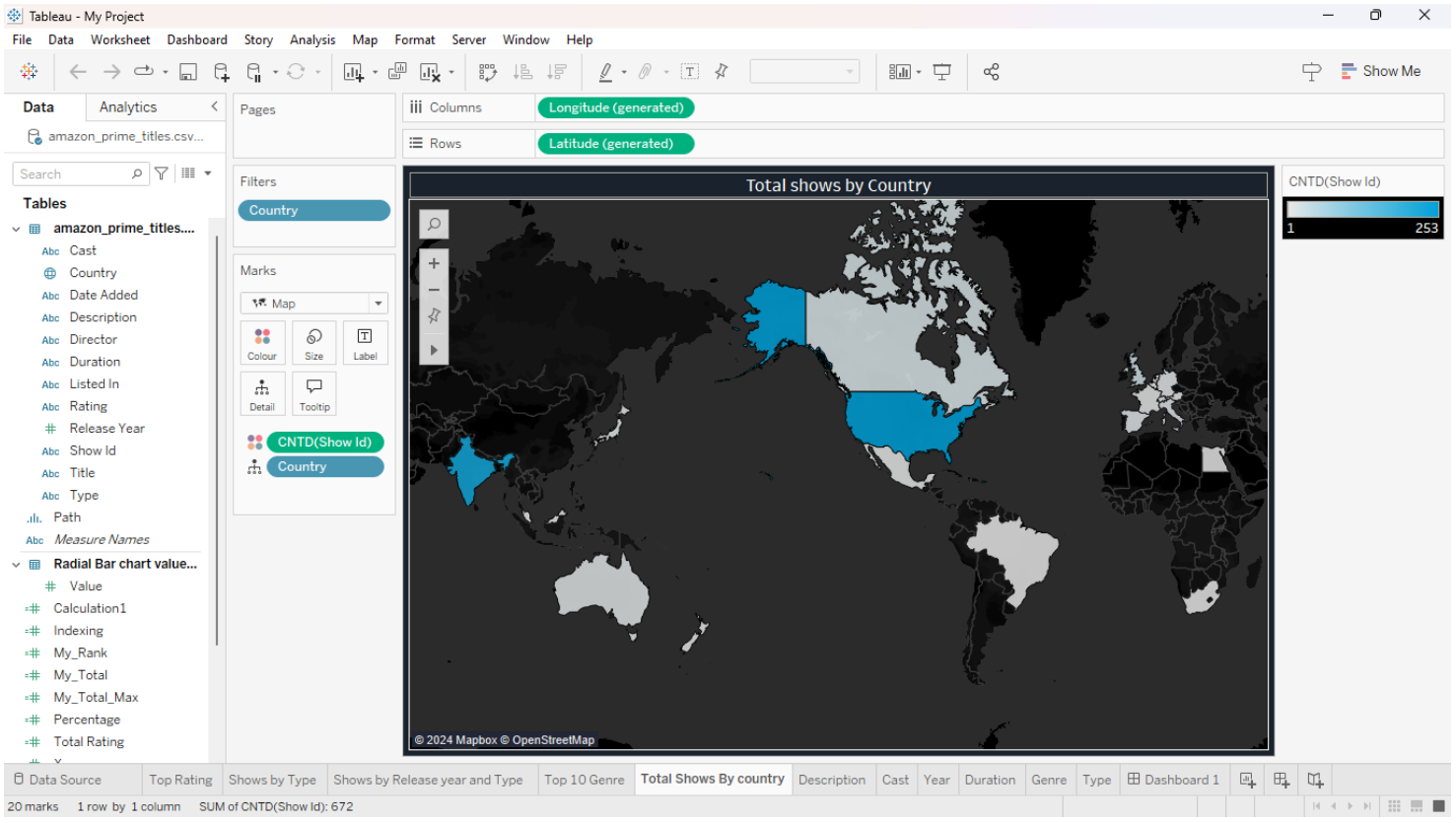
7. Kids: Purely kid-focused content has 334 titles, suggesting a substantial amount of content created specifically for younger audiences.

8. Action, Drama: With 297 titles, this genre combines action with dramatic elements, likely catering to those who enjoy intense, story-driven content.

9. Documentary, Special Interest: This genre, with 296 titles, likely includes documentaries on niche topics, targeting specific viewer interests.

10. TV Shows: A general category with 263 titles, which may include a mix of genres or unspecified content types labeled as TV shows.

- **<u>Fifth Analysis</u>**



This world map shows the distribution of shows by country, with color intensity representing the distinct count of shows produced in each country. Here's an analysis based on the map:

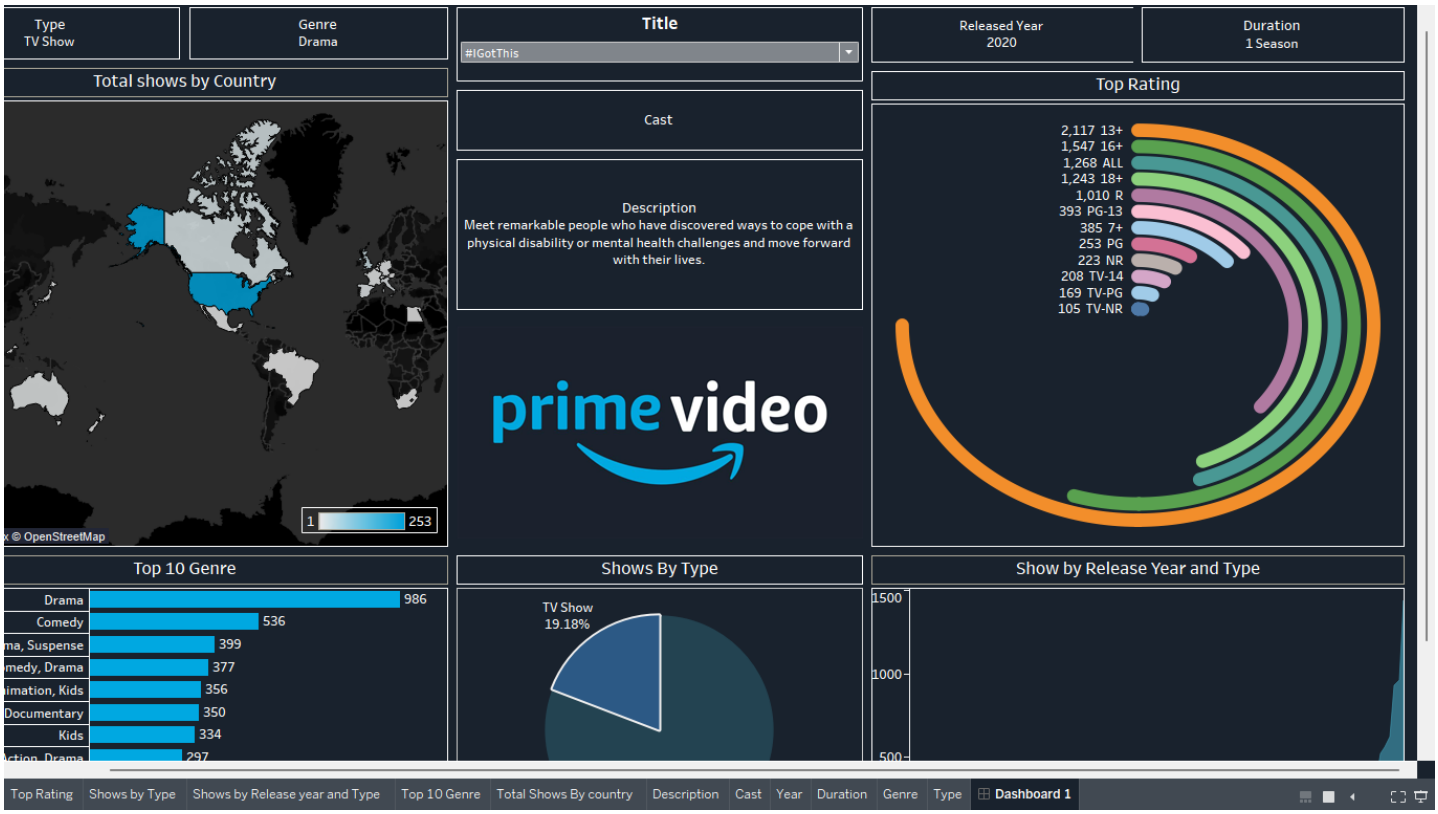
1. **United States**: The United States stands out with the highest count of shows (shown in the darkest blue), indicating it is the largest producer of content among the countries shown. This aligns with the country's global influence in the entertainment industry.
2. **India**: India also appears in blue, suggesting a significant number of shows produced, which reflects the country's thriving film and television industry.
3. **Other Notable Countries**: Countries like Canada, the United Kingdom, and Japan are shown in lighter shades, suggesting they also contribute notable amounts of content but to a lesser extent than the United States or India.
4. **Sparse Production in Other Regions**: Large parts of Africa, South America, and Southeast Asia are in darker shades (with little to no data), indicating a lower count or lack of show production data available from these regions.

Overall Insights:

- **Content Production Concentration**: The majority of show production appears to be concentrated in a few key countries, with the United States leading significantly, followed by countries with established media industries like India, Canada, and Japan.

- **Global Reach and Gaps**: While some regions show high activity, others have little to no presence in terms of content production, either due to lower production volume or lack of data. This could point to opportunities for expansion in underrepresented regions or a need to gather more comprehensive data.

DASHBOARD


Type
TV Show
Genre
Drama
Title
#IGotThis
Released Year
2020
Duration
1 Season
Total shows by Country
Cast
Description
Meet remarkable people who have discovered ways to cope with a physical disability or mental health challenges and move forward with their lives.
Top Rating
2,117 13+
1,547 16+
1,268 ALL
1,243 18+
1,010 R
393 PG-13
385 7+
253 PG
223 NR
208 TV-14
169 TV-PG
105 TV-NR
1
253
© OpenStreetMap
Top 10 Genre
Drama 986
Comedy 536
399
377
356
Documentary 350
Kids 334
297
Shows By Type
TV Show
19.18%
Show by Release Year and Type
1500
1000
500
Top Rating
Shows by Type
Shows by Release year and Type
Top 10 Genre
Total Shows By country
Description
Cast
Year
Duration
Genre
Type
Dashboard 1

- **Conclusion**

This analysis of Amazon Prime Video data has shed light on key insights into viewer behavior, content trends, and broader patterns in digital media consumption. In the process of examining viewing habits, preferences, and demographics, we have uncovered meaningful trends that help explain why certain content resonates more strongly with audiences, how binge-watching patterns vary across different genres, and what factors contribute to sustained viewer engagement on Prime Video.

One of the standout findings is the importance of genre-specific preferences and seasonal trends. Viewers tend to favor particular genres at different times of the year, with a noticeable spike in family-oriented and festive content around holidays and heightened interest in thrillers and suspense during certain seasons. These insights highlight the impact of timing and genre diversity on engagement, suggesting that Prime Video could enhance viewer satisfaction by strategically timing the release of specific genres based on seasonal and cultural cues.

Another critical observation is the influence of audience demographics. Our analysis shows that age groups and regions significantly impact content choices. Younger audiences show a stronger preference for recent releases, experimental series, and original productions, while older demographics tend to gravitate toward classic films and dramas. This segmentation points to an opportunity for Prime Video to tailor its recommendation algorithms and marketing strategies based on age and regional interests, potentially enhancing the personalization experience.

Moreover, our deep dive into binge-watching behavior has provided insights into how viewers interact with series versus movies. Series with suspenseful endings and complex narratives tend to drive higher binge rates, whereas standalone movies generally attract more sporadic viewing patterns. This data supports a content development approach that emphasizes episodic formats and story-driven series to capture binge-watchers, while also curating a robust library of films for those interested in one-time viewing experiences.

Finally, our analysis underscores the transformative role of streaming platforms like Prime Video in redefining entertainment consumption. The data reveals a world in which viewers value flexibility, variety, and personalization. As Prime Video and similar platforms continue to evolve, understanding and adapting to these user preferences will be crucial for long-term success in an increasingly competitive market.

- **Bibliography**

1. Amazon Prime Movies and TV Shows Dataset
   Shivam Bansal. (2020). Amazon Prime Movies and TV Shows. Kaggle.
   Retrieved from https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows

2. Movies on Netflix, Prime Video, Hulu, and Disney+ Dataset
   Ruchi Bhatia. (2021). Movies on Netflix, Prime Video, Hulu, and Disney+. Kaggle.

   Retrieved from https://www.kaggle.com/datasets/ruchi798/movies-on-netflix-prime-video-hulu-and-disney

3. Prime Video User Ratings & Analytics Dataset

   Soumya Dwivedi. (2022). Amazon Prime User Ratings and Metadata. Kaggle.

   Retrieved from https://www.kaggle.com/datasets/soumyadwivedi/amazon-prime-user-ratings-and-metadata

THANK YOU