**DATA ANALYSIS OF AMAZON PRIME**

ASSIGNMENT

*Submitted in fulfillment of the requirement of degree of*

Bachelor of technology

Computer engineering with specialization in data science

To

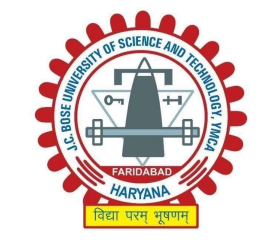
The Faculty of Informatics and computing

By

MANSI

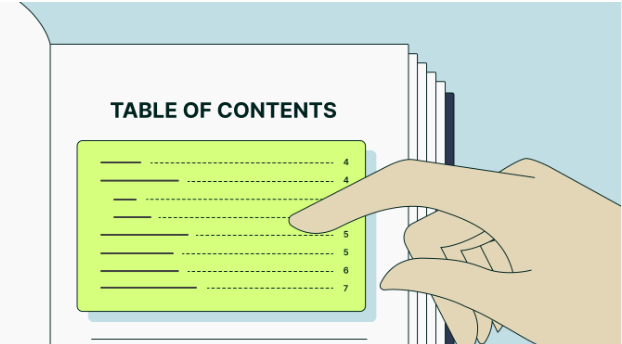
(Roll No.-23001016038)

Under Supervision of JYOTI MA’AM



Department of Computer Engineering J.C. Bose University of Science and Technology, YMCA, Faridabad Sector-6, Mathura Road, Faridabad-121006, Haryana, INDIA

OCTOBER 2024



* Introduction of Amazon Prime
* Steps of data analysis
* Need of Data Analysis
* Key Objectives
* Data sources
* Tools and libraries
* Methology
* Data Overview
* Data sample set
* Data analysis
* Challenges
* Outcomes
* Project Deliverables

**AMAZON PRIME**



Amazon Prime is a subscription service offered by Amazon that provides a range of benefits to its members.

Amazon Prime aims to provide value through convenience, entertainment, and exclusive offers, enhancing the overall shopping and entertainment experience for its members.

Key features include:

1. **Streaming Services**: Access to Prime Video, which offers a vast library of movies, TV shows, and original content. Prime members can also access Prime Music and Prime Gaming.
2. **Fast Shipping**: Free two-day shipping on eligible items, along with same-day or one-day delivery in select areas.
3. **Exclusive Deals**: Members receive early access to Lightning Deals and other discounts on products available on Amazon.
4. **Prime Reading**: Access to a rotating selection of books, magazines, and comics through Kindle.
5. **Amazon Photos**: Unlimited photo storage and enhanced storage for videos and documents.
6. **Twitch Prime**: Benefits for gamers, including free monthly channel subscriptions on Twitch and exclusive in-game loot.
7. **Amazon Family**: Discounts on family-oriented products, such as baby supplies and other essentials.

8**.Amazon First Reads**

* **Early Access to Books**: Members can read new books from popular authors before their official release.

9. **Amazon Music Unlimited**

* **Expanded Music Library**: Members can subscribe to Amazon Music Unlimited for access to millions of songs, with more advanced features.

10. **Global Availability**

* **International Shipping**: Prime members can enjoy benefits in many countries, with variations in available services.
* In a nutshell, Amazon Prime stands out as a multifaceted subscription service that combines convenience, entertainment, and exclusive benefits for its members. By offering a wide array of features—from fast shipping and streaming services to discounts and family-friendly perks—Amazon Prime enhances the shopping experience.
* Its ability to adapt to consumer needs and preferences, along with its innovative offerings, has solidified its position as a leader in both e-commerce and digital entertainment. As Amazon continues to evolve the service, prime not only drives customer loyalty but also shapes the competitive landscape, encouraging continuous innovation across industries



**Why data analysis of amazon prime needed?**

\

 **Customer Insights**: Understanding customer behaviour, preferences, and viewing habits helps Amazon tailor its content offerings and marketing strategies to enhance user engagement and satisfaction.

 **Content Optimization**: By analysing viewing trends, Amazon can determine which types of shows or movies resonate with viewers, guiding decisions on future content production and acquisition.

** Subscription Trends**: Analysing subscription data can reveal patterns in user sign-ups, cancellations, and renewals, helping Amazon optimize pricing strategies and promotional efforts.

** Operational Efficiency**: Data analysis can improve operational processes, from content delivery networks to customer service, ensuring a smoother user experience.

** Competitive Analysis**: Understanding how Amazon Prime compares to competitors in terms of content, pricing, and customer engagement can inform strategic decisions and positioning in the market.

** Marketing Effectiveness**: Evaluating the impact of marketing campaigns allows Amazon to refine its outreach strategies, ensuring better ROI on advertising spend.

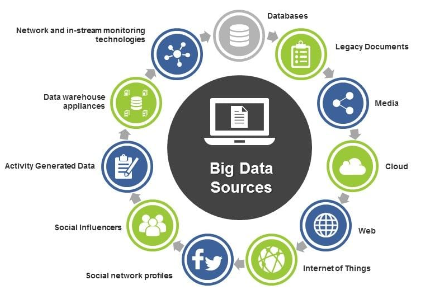
**Objectives**

**Amazon Prime Data Analysis**

* we will analyse a dataset from Amazon Prime, concentrating on visualizing key trends in the TV shows and movies available on the platform. The dataset encompasses various variables, including show type, country of origin, release year, rating, genre, and duration.
* Using Excel as our primary tool, we aim to uncover insights into the distribution of content on Amazon Prime and analyse trends in show production across different genres, countries, and years.
* This analysis will help us understand viewer preferences, identify popular content types, and assess how Amazon Prime's offerings have evolved over time. By leveraging visualizations, we will effectively communicate our findings and provide actionable insights that could inform future content strategies.

**Some key objectives of Amazon Prime:**

1. **Enhance Customer Experience**:
   * Provide a seamless and enjoyable user experience through an extensive library of content, personalized recommendations, and user-friendly navigation.
2. **Expand Content Library**:
   * Continuously increase the variety and quality of shows, movies, and original programming to attract and retain subscribers.
3. **Drive Subscriber Growth**:
   * Increase the number of Amazon Prime members globally through targeted marketing strategies, promotions, and exclusive content offerings.
4. **Increase Engagement**:
   * Encourage longer viewing sessions and higher engagement rates by offering compelling content, personalized suggestions, and interactive features.
5. **Leverage Data Analytics**:
   * Utilize viewer data and analytics to inform content acquisition and production decisions, ensuring alignment with viewer preferences and trends.
6. **Strengthen Brand Loyalty**:
   * Foster brand loyalty by providing additional benefits to Prime members, such as free shipping, exclusive deals, and access to Amazon Music and other services.
7. **Enhance Global Reach**:
   * Expand into new international markets by localizing content, catering to regional preferences, and forming partnerships with local content creators.
8. **Improve Original Programming**:
   * Invest in original content production to create exclusive titles that differentiate Amazon Prime from competitors and draw in new viewers.
9. **Monetization Strategies**:
   * Explore various monetization options, such as ads, partnerships, or tiered subscription models, to increase revenue while maintaining subscriber satisfaction.
10. **Integrate with Other Amazon Services**:
    * Promote synergy with other Amazon services, such as Alexa, Amazon Music, and Amazon Fresh, to enhance the overall value proposition for Prime members.



For data analysis of Amazon Prime, various specific data sources are utilized to gather insights and inform decisions. Here’s a breakdown of the key data sources:

**1. User Behaviour Data**

* **Viewing Logs**: Detailed records of what content users watch, including timestamps and durations.
* **Search History**: Data on what users search for on the platform, providing insights into content demand.

**2. Subscription and Billing Data**

* **Subscription Details**: Information on sign-up dates, payment methods, renewal patterns, and cancellation reasons.
* **Account Status**: Data on active versus inactive accounts to identify trends in membership engagement.

**3. Content Interaction Data**

* **Ratings and Reviews**: User feedback on shows and movies that can help assess content popularity and quality.
* **Watch lists**: Information about titles users add to their watch lists, indicating interest levels.

**4. Marketing and Promotion Data**

* **Campaign Performance Metrics**: Data on the effectiveness of marketing campaigns, including click-through rates and conversion rates.
* **Promotional Codes Usage**: Tracking the use of discounts and promotions to evaluate their impact on subscription growth.

**5. Demographic and Psychographic Data**

* **User Profiles**: Information on user demographics (age, gender, location) and interests to tailor content recommendations.
* **Surveys and Feedback Forms**: Insights from user surveys regarding preferences and satisfaction levels.

**6. Sales and Purchase Data**

* **Cross-Selling Data**: Insights from purchases made alongside streaming subscriptions, helping to understand buying patterns.
* **Amazon Marketplace Data**: Analysing related product sales to identify trends that can inform content strategies.

**7. Social Media and Online Sentiment**

* **Social Media Monitoring**: Data from platforms like Twitter and Instagram to gauge public sentiment about content and services.
* **Online Reviews**: Aggregated reviews from sites like IMDb and Rotten Tomatoes for additional feedback on content.

**8. Operational Data**

* **System Performance Metrics**: Data on streaming quality, load times, and user interface interactions to identify technical issues.
* **A/B Testing Results**: Data from experiments testing different features or layouts on the platform to determine user preferences.

**THE MAJOR DATA SOURCES USED ARE:**

* Amazon Advertising API
* Amazon Associates Program
* IMDb
* Statista
* Brand watch
* Hoot suite
* Survey Monkey
* Google Forms
* Nielsen Ratings
* Kaggle
* Alexa Internet
* eMarketer

**LIBRARIES FOR DATA ANALYSIS:**

1. **Pandas (Python)**: Data manipulation library for structured data.
2. **NumPy (Python)**: Fundamental library for numerical computing and arrays.
3. **Matplotlib (Python)**: Library for creating static and animated visualizations.
4. **Seaborn (Python)**: High-level interface for attractive statistical graphics.
5. **Scikit-learn (Python)**: Machine learning library for data mining and analysis.
6. **Statsmodels (Python)**: Library for estimating and testing statistical models.
7. **Dplyr (R)**: Functions for easy data manipulation in R.
8. **Tidyverse (R)**: Ecosystem of R packages for data science.
9. **ggplot2 (R)**: Visualization package for creating complex plots in R.
10. **Plotly (Python and R)**: Library for interactive plots and dashboards.

**TOOLS FOR DATA ANALYSIS**

 **Python**: Versatile programming language for data analysis.

 **R**: Statistical language ideal for data visualization and analysis.

 **Jupyter Notebook**: Interactive web app for live coding and data exploration.

 **Tableau**: Powerful tool for creating interactive data visualizations.

 **Power BI**: Microsoft tool for business analytics and data visualization.

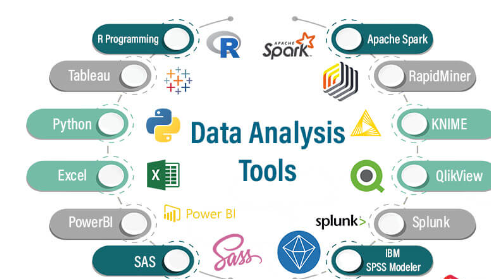
 **Apache Spark**: Distributed computing system for big data processing.

 **Microsoft Excel**: Widely used spreadsheet software for data manipulation.

 **Google Data Studio**: Free tool for customizable data reports and dashboards.

 **SAS**: Comprehensive software suite for advanced analytics.

 **Rapid Miner**: Data science platform for data preparation and machine learning.



### Methodology for Data Analysis on Amazon Prime

#### 1. Define Objectives

* **Clarify Goals**: Establish clear, specific objectives for the analysis, such as evaluating user engagement, assessing content performance, or analysing subscription trends.
* **Identify Key Performance Indicators (KPIs)**: Determine relevant KPIs to measure success, such as average watch time, churn rates, and acquisition costs.

#### 2. Data Collection

* **Identify Data Sources**: Gather data from diverse sources, including:
  + **User Activity Logs**: Records of viewing history, search behaviour, and content interaction.
  + **Subscription Data**: Information on sign-ups, cancellations, and payment histories.
  + **Customer Feedback**: User reviews, ratings, and survey responses to assess satisfaction levels.
  + **Market Research**: External industry reports and competitor analysis data.
* **Utilize Data Extraction Tools**: Employ APIs, web scraping techniques, or database queries to collect necessary data.

#### 3. Data Cleaning and Pre-processing

* **Handle Missing Values**: Address missing data through imputation methods or exclusion of incomplete records.
* **Data Transformation**: Convert data into appropriate formats, ensuring consistency across datasets.
* **Remove Duplicates**: Identify and eliminate duplicate entries to enhance data integrity.
* **Normalization**: Standardize data as needed to facilitate analysis across varying scales.

#### 4. Exploratory Data Analysis (EDA)

* **Visualization Techniques**: Implement visualizations such as histograms, scatter plots, and box plots to uncover trends and relationships.
* **Statistical Summaries**: Calculate descriptive statistics (mean, median, standard deviation) to gain insights into data distributions.
* **Segment Analysis**: Analyse user segments based on demographics or behaviour to identify distinct patterns.

#### 5. Data Modeling

* **Model Selection**: Choose appropriate analytical techniques based on objectives:
  + **Descriptive Analytics**: Summarize historical data for insights into user behaviour.
  + **Predictive Analytics**: Utilize regression models or machine learning algorithms to forecast future trends.
  + **Prescriptive Analytics**: Recommend actions based on analysis outcomes, optimizing strategies for content and marketing.
* **Feature Engineering**: Create new variables to enhance model accuracy, such as calculating average watch time per user.

#### 6. Analysis and Interpretation

* **Execute Analyses**: Apply selected models using statistical software or programming languages (e.g., Python, R).
* **Interpret Results**: Analyse outputs to draw actionable insights, identifying patterns or anomalies relevant to defined objectives.
* **Cross-Validation**: Validate findings through cross-validation techniques or subset testing to ensure robustness.

#### 7. Validation

* **Benchmark Comparison**: Cross-check results against industry standards or historical data to confirm validity.
* **A/B Testing**: Implement A/B tests where applicable to compare strategies and validate findings through user interaction metrics.

#### 8. Reporting and Visualization

* **Dashboard Creation**: Develop interactive dashboards using tools like Tableau or Power BI to visualize key insights effectively.
* **Comprehensive Reporting**: Prepare detailed reports summarizing methodology, findings, and strategic recommendations tailored for various stakeholders.

#### 9. Actionable Recommendations

* **Strategic Insights**: Provide actionable recommendations based on analysis, such as:
  + Content acquisition or production strategies informed by user preferences.
  + Targeted marketing strategies for specific user segments.
  + User experience enhancements based on interaction data.

#### 10. Continuous Monitoring

* **Establish KPIs for Ongoing Analysis**: Set up KPIs to monitor user engagement and content performance continuously.
* **Feedback Loop Implementation**: Create a mechanism for incorporating new data and user feedback into the analysis framework.
* **Iterative Improvement**: Treat data analysis as an iterative process, refining models and strategies based on evolving insights and market dynamics.

Data Overview

The dataset consists of the following key columns:

• Show ID: A unique identifier for each show.

• Show Type: Specifies whether the entry is a TV show or a movie.

• Country: The country where the show was produced.

• Release Date: The year when the show was first released

. • Description: A brief summary or tagline of the show.

• Genre: The genre(s) the show belongs to, such as Drama, Comedy, Action, etc.

• Rating: Audience rating, which reflects the overall reception of the show.

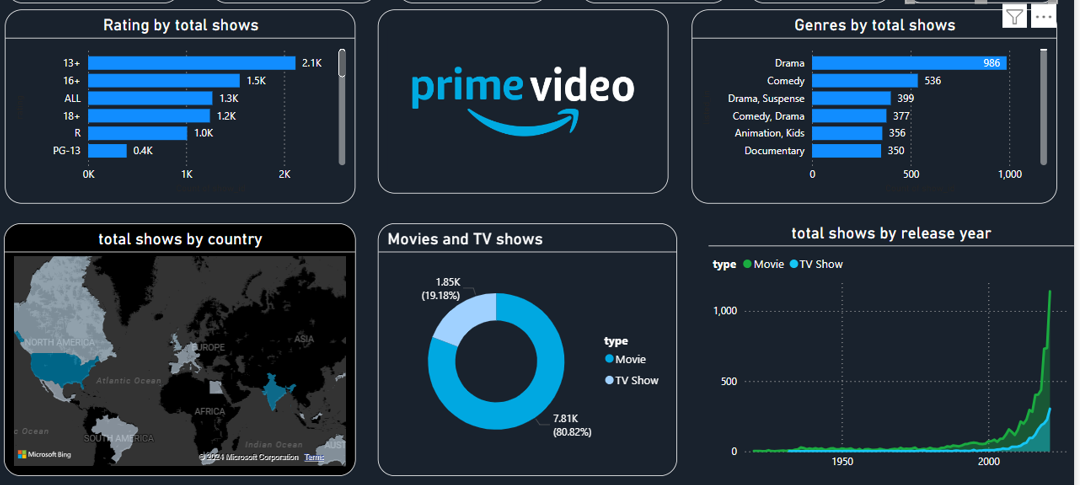
• Duration: The length of the movie or the number of seasons for TV shows.

DATA SAMPLE SET



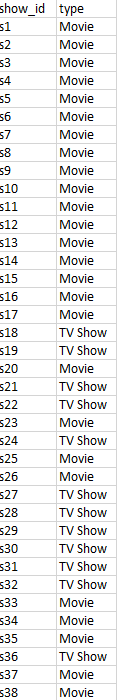


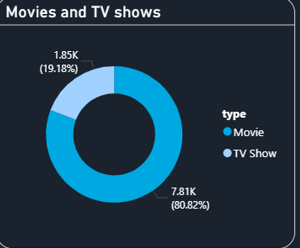


DATA ANALYSIS

To extract meaningful insights from the Netflix dataset, I utilized Tableau and Power BI, a powerful data visualization tool. The analysis focused on answering specific questions, shedding light on the distribution, rating trends, content type breakdown, popular genres, and temporal patterns.

## **Content Type on Amazon Prime**





\col = "type"

grouped = df[col].value\_counts (). reset\_index ()

grouped = grouped. Rename (columns = {col : "count", "index" : col})

*## plot*

trace = go. Pie(labels=grouped[col], values=grouped['count'], pull=[0.05, 0], marker=dict(colors=["#6ad49b", "#a678de"]))

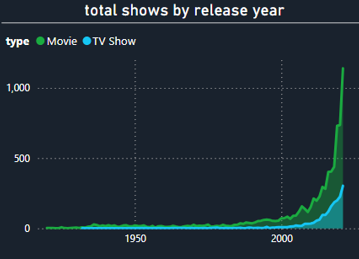
layout = go. Layout (title="", height=400, legend=dict (x=0.1, y=1.1))

fig = go. Figure (data = [trace], layout = layout)

iplot(fig)

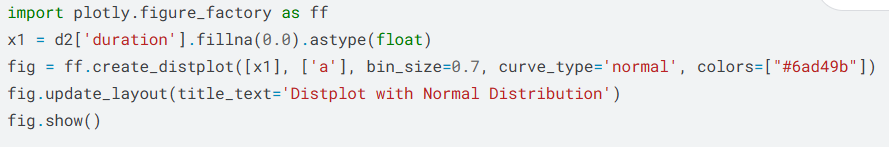
## **Original Release Year of the movies**

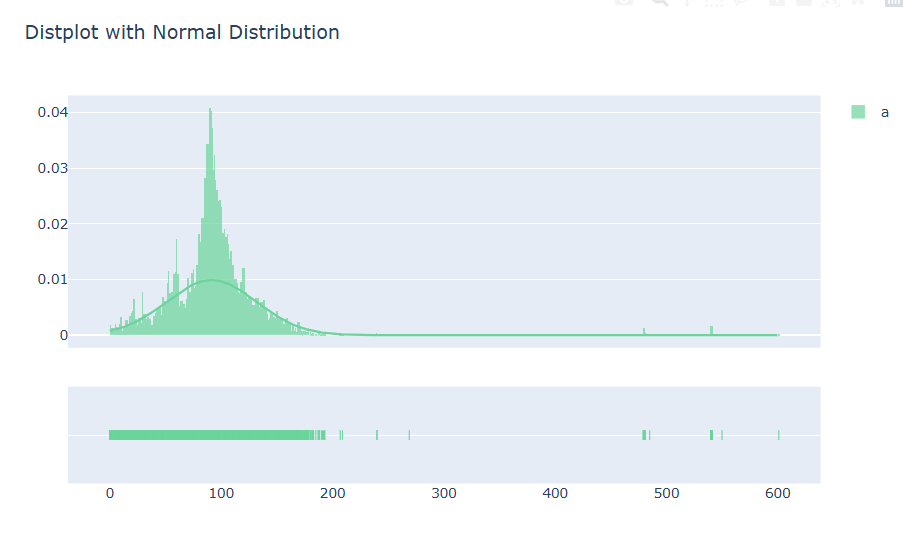


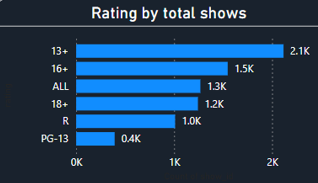
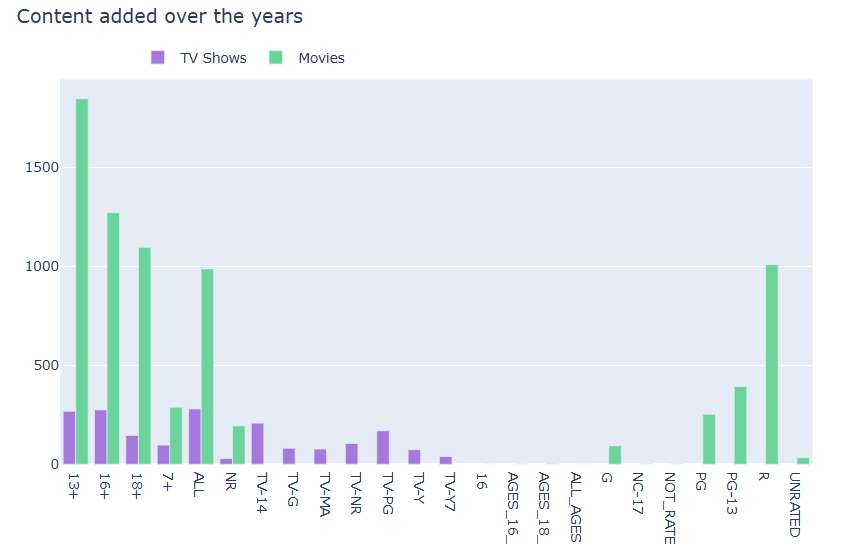




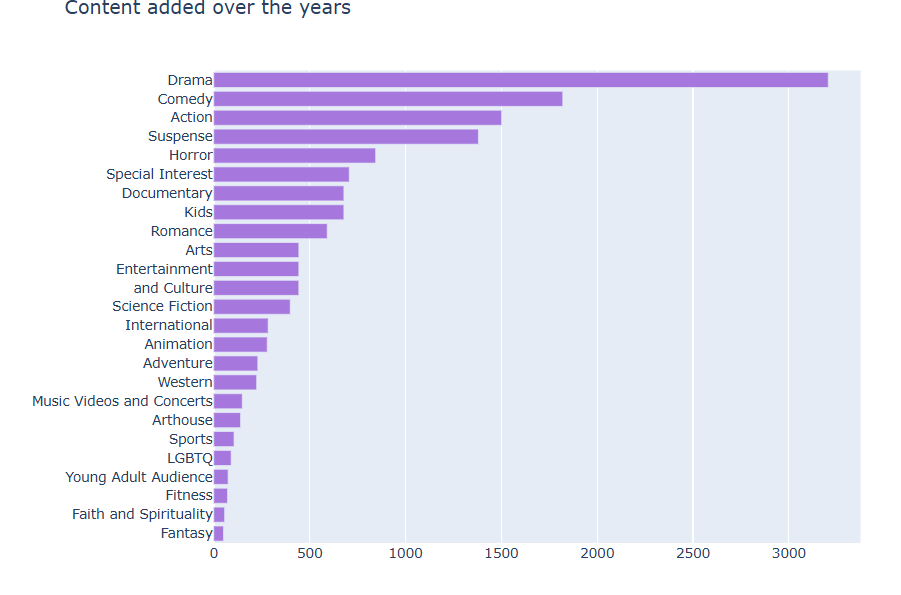
## **Distribution of Movie Duration**

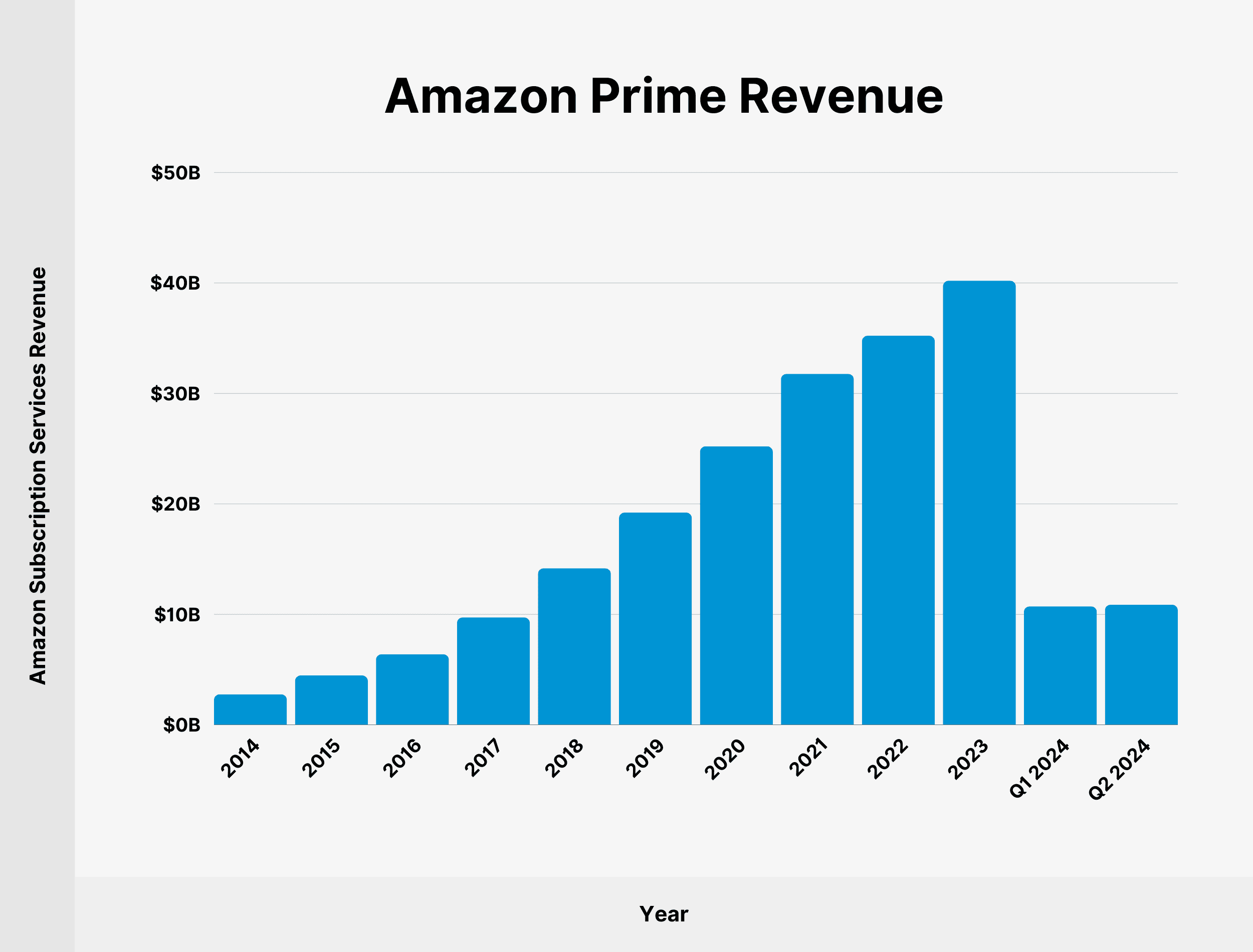
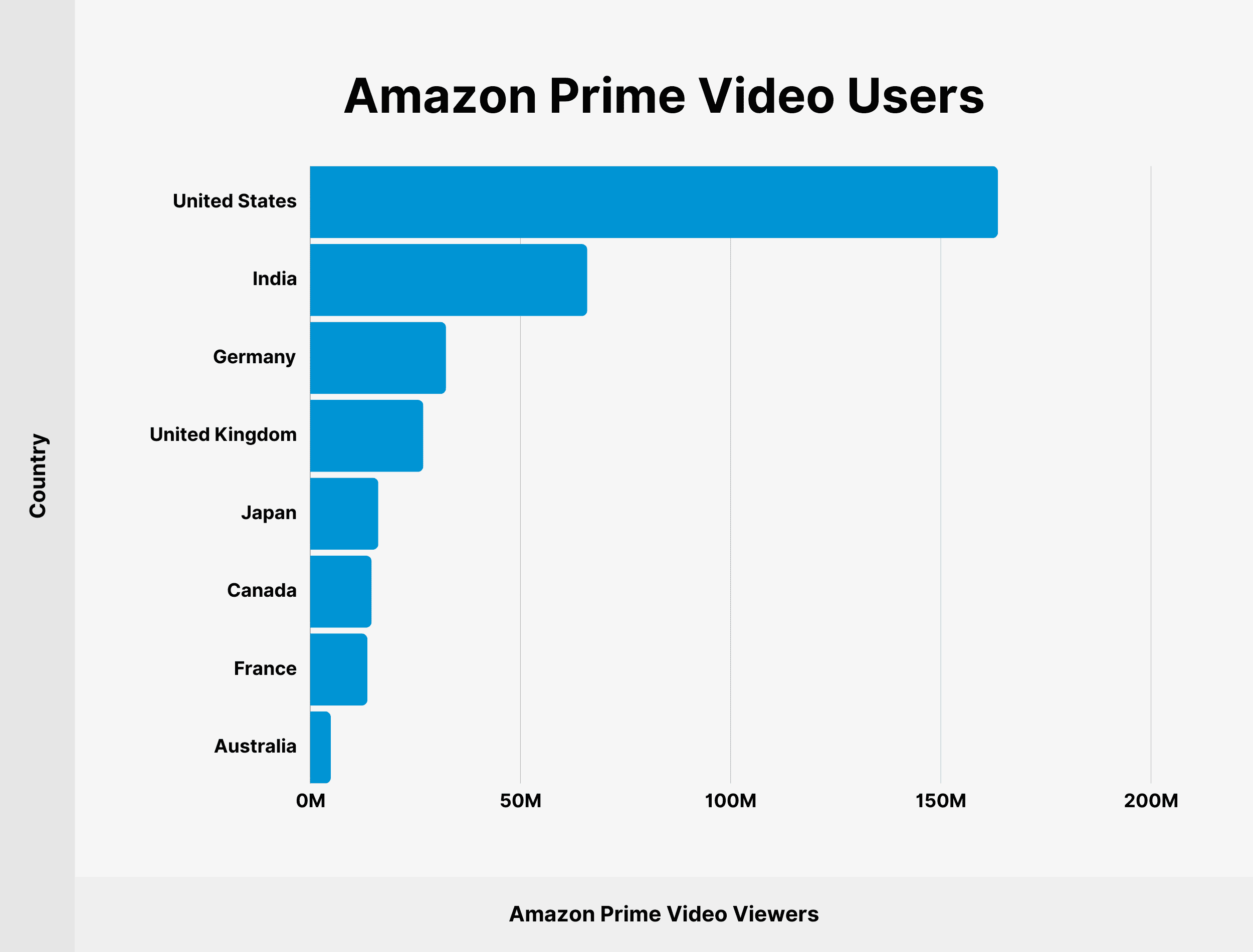




Rating of the content

## **What are the top Categories?**





**Challenges**

Analysing data from Amazon Prime Video can present several challenges:

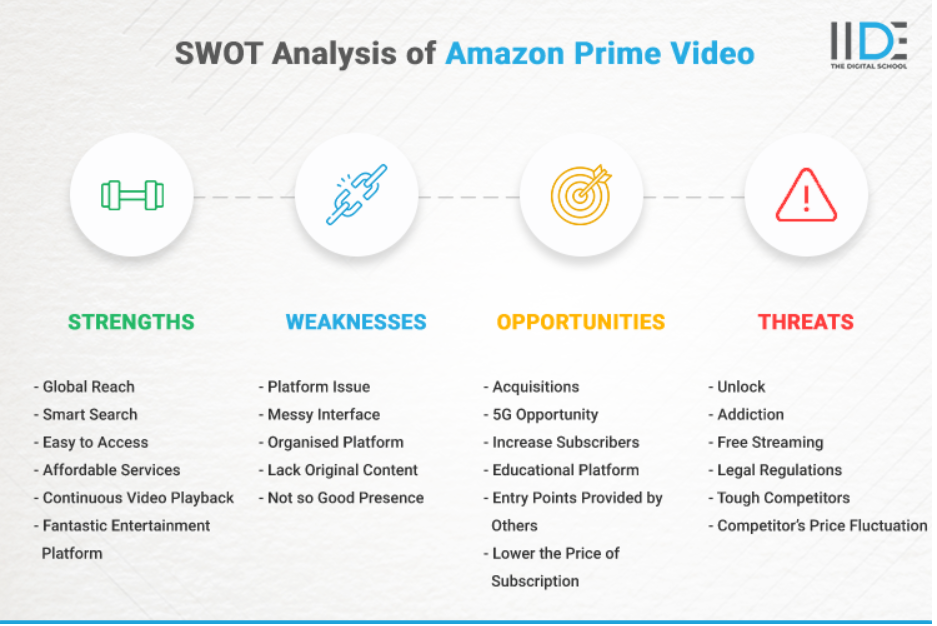
1. **Data Accessibility**:
   * **Limited Public Data**: Many metrics, like viewer counts or demographic data, are not publicly available.
   * **API Limitations**: If using an API, there might be restrictions on the volume or type of data you can retrieve.
2. **Data Quality**:
   * **Inconsistent Ratings**: Ratings may vary across different platforms (e.g., IMDb, Rotten Tomatoes), leading to inconsistencies.
   * **User Bias**: Reviews can be influenced by factors like marketing or social media trends, affecting the reliability of ratings.
3. **Sample Size**:
   * **Small Sample Sizes**: For niche shows, there may not be enough ratings or reviews to draw meaningful conclusions.
4. **Dynamic Content**:
   * **Changing Library**: Amazon Prime’s content library is frequently updated, which can make longitudinal studies difficult.
   * **Seasonal Variability**: Viewer ratings may fluctuate based on external factors, such as trends or global events.
5. **Genre Classification**:
   * **Overlapping Genres**: Shows may fit into multiple genres, complicating analysis and comparisons.
   * **Subjectivity in Genre**: Different viewers may categorize shows differently, leading to inconsistencies.
6. **Interpretation of Data**:
   * **Contextual Factors**: Understanding the context behind ratings (e.g., production quality, marketing, competition) is crucial for accurate analysis.
   * **Correlation vs. Causation**: It can be challenging to determine if high ratings directly correlate with viewer satisfaction or are influenced by external factors.
7. **User Engagement Metrics**:
   * **Lack of Comprehensive Metrics**: Metrics like watch time, viewer retention, and engagement are often not disclosed, limiting in-depth analysis.
8. **Privacy Concerns**:
   * **Data Anonymity**: Ensuring user privacy while collecting and analysing data can restrict the types of insights that can be gleaned.

**Outcomes**

****

Analysing data from Amazon Prime Video can yield several valuable outcomes that inform strategic decisions and enhance viewer experience. Here are some key outcomes:

1. **Viewer Preferences**:
   * **Content Trends**: Understanding which genres, themes, or formats are most popular helps guide future content acquisition and production.
   * **Rating Insights**: Identifying shows with high ratings can inform marketing strategies and highlight potential successes.
2. **Audience Segmentation**:
   * **Demographic Insights**: Analysing viewer demographics can help tailor content and marketing efforts to specific audience segments.
   * **Behaviour Patterns**: Understanding viewing habits (e.g., binge-watching vs. episodic viewing) can optimize release strategies.
3. **Content Strategy**:
   * **Acquisition Decisions**: Data can guide the acquisition of new titles or franchises that align with viewer interests.
   * **Production Focus**: Insights from ratings and reviews can inform decisions on which types of original content to produce.
4. **Performance Metrics**:
   * **Engagement Levels**: Metrics like average watch time and completion rates help assess how well content retains viewers.
   * **Churn Analysis**: Identifying trends in subscriber retention can highlight areas for improvement in content offerings.
5. **Marketing Effectiveness**:
   * **Campaign Evaluation**: Analysing viewer response to marketing campaigns can inform future promotional strategies.
   * **Social Media Impact**: Understanding how social media buzz correlates with viewership can optimize engagement efforts.
6. **Competitive Analysis**:
   * **Benchmarking**: Comparing Amazon Prime’s offerings and performance against competitors helps identify strengths and weaknesses.
   * **Market Trends**: Keeping abreast of broader industry trends can inform strategic positioning.
7. **User Experience Improvement**:
   * **Interface Optimization**: Analysing how users navigate the platform can lead to improvements in user interface design.
   * **Personalization**: Data-driven recommendations can enhance user satisfaction by providing tailored content suggestions.
8. **Feedback Loop**:
   * **Real-time Adjustments**: Continuous data analysis allows for agile responses to viewer feedback and preferences, ensuring the platform remains relevant.

****These outcomes not only enhance Amazon Prime’s content strategy but also improve overall viewer satisfaction and engagement.

**Project Deliverables**

This project aims to conduct a comprehensive analysis of Amazon Prime’s content using Excel. The primary outputs are designed to deliver clear insights into the platform’s offerings and trends, ultimately supporting strategic decision-making. Below are the key deliverables:

1. **Excel Workbook**  
   The Excel workbook used for the analysis will be shared, containing:
   * **Organized and Cleaned Dataset**: A well-structured dataset, prepped for analysis with clear documentation.
   * **Formulas and Functions**: Comprehensive use of Excel functions for sorting, filtering, and aggregating data, ensuring transparency in analytical processes.
   * **Integrated Charts and Graphs**: All visualizations embedded within the workbook for immediate reference.
   * **Data Dictionary**: An explanation of key metrics and terminology used throughout the analysis to enhance understanding.  
     This deliverable allows others to review and replicate the analysis process, ensuring transparency and usability.
2. **Data Analysis Report**  
   A detailed report summarizing findings from the dataset, highlighting critical insights such as:
   * **Top-Rated Content**: Identification and analysis of the highest-rated shows and movies, including viewer demographics and engagement metrics.
   * **Content Production Trends**: Examination of how content production has evolved over the years, focusing on notable spikes and trends tied to external factors (e.g., global events, holiday seasons).
   * **Genre Popularity**: In-depth exploration of the most popular genres on the platform, including sub-genre analysis and viewer preferences.
   * **Country-wise Content Contributions**: Insights into the contributions of various countries to the content library, analysing cultural influences and regional trends.
   * **Viewer Behaviour Insights**: Analysis of viewing habits, such as binge-watching trends, peak viewing times, and average watch duration.  
     The report will feature written explanations alongside visual aids to support the conclusions drawn from the analysis, ensuring a comprehensive understanding of the data.
3. **Visual Representations**  
   A collection of visually engaging charts and graphs created in Excel to effectively convey findings, including:
   * **Bar Charts**: Highlighting the top 10 rated shows, popular genres, and distribution of content by country.
   * **Line Graphs**: Illustrating trends in content production over time, with annotations for significant events or changes.
   * **Pie Charts**: Depicting genre popularity and market share among different content types.
   * **Geographical Heat map**: Visualizing the number of shows produced by each country, showcasing regional contributions and cultural diversity.
   * **Interactive Dashboards**: If feasible, a user-friendly dashboard allowing stakeholders to explore the data dynamically.  
     Each visual will be clearly labelled and formatted for easy comprehension and included in both the report and the presentation.
4. **Presentation Slides**  
   A professional set of presentation slides summarizing the analysis, key findings, and visualizations. These slides will include:
   * **Executive Summary**: A brief overview of the project objectives and key insights.
   * **Visual Highlights**: Selected charts and graphs that effectively communicate the main findings.
   * **Key Recommendations**: Actionable insights based on the analysis, tailored for stakeholders to support strategic decisions.
   * **Q&A Section**: A slide dedicated to addressing potential questions from the audience, facilitating engagement during the presentation.  
     These slides will provide a clear, concise overview of the project and can be adapted for presentations to various audiences, including stakeholders and academic settings.
5. **Feedback Mechanism**  
   A structured approach for gathering feedback on the analysis and presentation:
   * **Surveys or Polls**: Short surveys to collect audience feedback on insights and areas for further exploration.
   * **Follow-Up Discussion**: Opportunities for stakeholders to discuss findings and implications for future content strategies.  
     This deliverable will help refine future analyses and ensure continuous improvement in the approach.