PROJECT REPORT

On

Data Analyzing Of Amazon Prime Movies And TV Shows

By

Neelofar Shaik - AP21110010047

Under the guidance of

K. Meenakshi, APSSDC

K. Narmada, APSSDC

Prepared in the partial fulfilment of the

Summer Internship Course At

Andhra Pradesh State Skill Development Corporation

(APSSDC)

3rd floor, Infosight Building, survey No. 78/2, Near Pathuru Road

Junction, NH-16 Service Road, Tadepalli - 522501, Guntur District, Andhra Pradesh

Table of Contents

ACKNOWLEDGEMENTS

I extend my heartfelt gratitude to my mentors, Meenakshi Mam and Narmada Mam, for their unwavering dedication and guidance throughout the internship. Your valuable insights and thoughtful suggestions significantly contributed to the successful completion of the project. I am truly appreciative of your support.

I also wish to convey my sincere appreciation to the APSSDC management for their continuous guidance, supervision, and provision of essential project-related information. Your unwavering support has been instrumental in the project's accomplishment.

ABSTRACT

This data analysis project is focused on exploring the vast content landscape of Amazon Prime Movies and TV Shows. The project utilises a comprehensive dataset sourced from kaggle.com, comprising two CSV files: titles.csv and credits.csv. The primary objectives of this project include an in-depth analysis of various content aspects, such as genres, ratings, user preferences, and temporal trends. The project will provide a comprehensive report depicting key insights and findings of digital entertainment.

As I conclude this data analysis project, this report offers a unique perspective on the content realm of Amazon Prime Movies and TV Shows. By Data Analysis using Python, we provide content creators and producers with actionable insights to optimize content curation, enhance user experience, and stay ahead in the fiercely competitive digital entertainment field.

SYSTEM REQUIREMENTS

- PC(Minimum 4 GB Ram and 256 SSD)

- OS: Windows 7 and above

- Python (Version 3.10)

- Jupyter Notebook

INTRODUCTION

This data analysis project is focused on exploring the vast content of Amazon Prime Movies and TV Shows. The project aims to enhance user preference understanding, better content decision-making, and increased customer satisfaction by analysing the dataset of amazon prime, which is obtained from Kaggle.com source. The dataset acquired from kaggle.com encompasses over 10,000 titles and more than 140,000 credits of actors and directors. This vast data repository forms the foundation for our analysis. The project will provide a comprehensive report depicting key insights and findings of digital entertainment.

Data Cleaning and Organization: Missing values and inconsistencies are addressed to create a solid foundation for subsequent analysis.

Data Exploration: thoroughly acquaint ourselves with the dataset's intricacies.

Data Visualization: Data visualisation techniques, such as histograms, bar plots, Scatter plots, and box plots, are used to represent the distribution visually, Relationships and patterns within the data.

Statistical Analysis: It was applied to delve deeper into genre distributions and user behaviour trends.

Interpretation and Conclusion: we distilled valuable insights regarding content popularity and emerging trends, enhancing our overall understanding.

Libraries Used for Data Analysis

PANDAS:

- Used for data manipulation and analysis.
- Offers powerful data structures like DataFrames for handling structured data.
- Provides functionalities for filtering, grouping, and transforming data efficiently.

NUMPY:

- Used for numerical computations and array operations.

MATPLOTLIB:

- Used for creating basic static visualisations like line plots, bar charts, and histograms.

SEABORN:

- Built on top of Matplotlib, it enhances visualisations with minimal code.
- Provides attractive statistical graphics like scatter plots, box plots, and heatmaps.

Warnings:

- Used to suppress unnecessary warnings during the analysis and visualisation process.
- Prevents warning messages from cluttering the output.

Plotly Express and Graph Objects:

- Used for creating interactive and visually appealing visualisations.

Missingno:

- Used for visualising and understanding missing data patterns.
- Helps to identify missing values in the dataset and potential data quality issues.

TextBlob:

- Used for text processing and sentiment analysis.
- Enables sentiment polarity detection (positive, negative, neutral) in textual data.

WordCloud:

- Used for creating word clouds, a popular and engaging visualisation for text data.
- Presents word frequency in a graphical format, with larger words indicating higher frequency.

Main Purpose Of The Project

The primary objective of this data analysis project is to derive insightful conclusions from an extensive dataset comprising Amazon Prime Movies and TV shows. By employing thorough data exploration, statistical methodologies, and advanced analytical techniques, the project aims to uncover underlying patterns and trends in user behaviour, content attributes, and temporal dynamics. The project seeks to illuminate latent relationships within diverse attributes, including genres, content ratings, and release years.

By employing statistical techniques and data visualization, the project seeks to map out the popularity and prevalence of various genres within the dataset. The analysis of heatmaps and graphical representations will facilitate a clearer understanding of the relationships between different attributes, enabling Amazon Prime to make informed content-related decisions.

We intend to discern viewers' emotional responses and perceptions associated with different content by conducting sentiment analysis on the descriptions of movies and shows. This precision-driven approach is poised to significantly enhance user engagement, satisfaction, and the overall streaming experience, ultimately boosting sales.

OBJECTIVE AND SCOPE OF THE PROJECT

The primary objective of this data analysis project is to extract valuable insights from a comprehensive dataset containing Amazon Prime Movies and TV shows. By analysing user preferences, our aim is to provide Amazon Prime with a deeper understanding of its audience, ultimately facilitating enhancements to the user experience and boosting sales.

The project scope encompasses:

Data Exploration and Preprocessing:

Cleaning, structuring, and preprocessing the dataset to ensure accuracy and consistency for analysis.

Genre Analysis:

Identifying the distribution of genres and uncovering trends in viewer preferences.

Content Rating Investigation:

Exploring content ratings and their correlations with genres, aiming to understand viewer age group preferences.

Runtime Trends:

Analyzing runtime patterns and their influence on content popularity.

Release Year Analysis:

Investigating content releases across years to discern temporal viewing trends.

User Sentiment Analysis:

Employing sentiment analysis on descriptions to gauge viewers' emotional responses to content.

Strategic Recommendations:

Formulating data-driven recommendations for content diversification, release planning, and user-centric enhancements.

The project scope encompasses a thorough exploration of user behaviour, content attributes, and viewer sentiment to provide actionable insights for Amazon Prime's content strategy enhancement.

Architecture Of The Project

Data Collection & Preprocessing

↓

Exploratory Data Analysis

↓

Genre Analysis

↓

Content Ratings Investigation

↓

Runtime Trends & Release Years

↓

User Sentiment Analysis

↓

Strategic Recommendations

↓

Visualization & Reporting

↓

Presentation & Delivery

Amazon Prime Movie And TV Shows

Data Collection and Monitoring

Data Collection and Monitoring:

This section delves into the pivotal phases of data collection and ongoing monitoring, forming the bedrock of our comprehensive analysis of the Amazon Prime Movies and TV shows dataset. The success of our analysis hinges upon the meticulous curation of a robust dataset and the continuous vigilance exercised to maintain data quality and accuracy.

Data Collection:

The data collection process commenced with acquiring the Amazon Prime Movies and TV shows dataset from the primary data source — Kaggle.com. The dataset, comprising a diverse range of content attributes, including genres, content ratings, descriptions, runtimes, and release years, was extracted with the utmost attention to fidelity.

Data Preprocessing:

Before embarking on the analysis, a significant emphasis was placed on data preprocessing. This phase involved cleansing the dataset to rectify any inconsistencies, errors, or missing values. The data was transformed into a structured format conducive to effective analysis, ensuring the accuracy of subsequent insights drawn.

Data Monitoring:

The aspect of data monitoring proved to be equally critical in maintaining data integrity throughout the project lifecycle. This ongoing process allowed us to promptly identify and rectify any discrepancies that could potentially impact the validity of our conclusions.

Recommendations To Enhance User Experience

And Boost Sales

Content Diversification:

Recognizing the popularity of the top 5 genres – drama, comedy, documentation, drama-romance, and horror – it's an opportune time to diversify your content offerings. Introducing additional genres such as action, thriller, science fiction, and fantasy can help capture a broader audience, ensuring a well-rounded entertainment experience that resonates with various viewer preferences.

Content Type Distribution:

Within your dataset, movies make up the majority, constituting 85.7% of the content, while series comprise the remaining 14.3%. This distribution underscores the prominence of movies, possibly indicating a preference among users for standalone, shorter-form content. Understanding this balance can help Amazon Prime tailor its content strategy to cater to the differing preferences of its audience.

Balancing Parental Guidance Ratings:

Given the prominence of "PG-13: Parents Strongly Cautioned" and "PG: Parental Guidance" ratings, focus on producing content that caters to these guidelines. Develop family-friendly series and movies that strike a balance between entertainment and age-appropriateness. This approach ensures that Amazon Prime offers content suitable for both parents and children, expanding the platform's appeal to a broader audience while maintaining viewer trust.

Diversify Content Lengths:

Given the variety of popular runtimes, it's evident that viewers appreciate the content of different durations. To engage a wider audience, consider producing a mix of content with runtimes around the most frequently occurring durations. This could involve creating shorter-form series or movies for time-constrained viewers and longer narratives for those seeking immersive experiences. By diversifying content lengths, Amazon Prime can effectively cater to varying preferences and enhance viewer satisfaction.

Strategic Release Planning Based on Historical Trends:

Given the surge in releases during 2021, followed by 2020 and 2019, adopt a strategic approach. Analyze successful content from 2021 to identify patterns and themes that resonated. Promote standout releases from that year, leverage nostalgia from 2020 and 2019 content, and use data to expand genres intelligently. This data-driven strategy ensures you capture momentum, maintain viewer interest, and guide future content decisions effectively.

Conclusion

In wrapping up Amazon Prime Movies and TV shows data analysis project, we've gleaned meaningful insights that are valuable for content curation and user engagement. The project's multi-faceted approach enabled us to unlock key trends and patterns, ultimately enhancing the streaming experience for viewers.

Advantages and Contributions:

Informed Decision-Making: Armed with data-backed insights, Amazon Prime can now make informed decisions about content selection, releasing strategies, and more.

Tailored User Experience: By understanding genre preferences, content ratings, and sentiments, Amazon Prime can tailor its offerings to match user expectations better.

Strategic Planning: Our Data Analysis aid in strategic content planning, optimizing audience engagement and satisfaction.

Actionable Recommendations: Our project's strategic recommendations provide a clear roadmap for enhancing content diversity and overall user experience.

Future Scope:

Personalization: Exploring more personalized recommendation systems could deepen viewer engagement.

Real-time Analytics: Integrating real-time data for up-to-date insights could refine content strategies.

Experimentation: Conducting content, A/B tests could provide empirical insights for continuous improvement.

This project stands as a stepping stone towards Amazon Prime's continued growth, ensuring it remains a preferred choice for viewers seeking tailored, engaging content.