INTRAIN
TECH

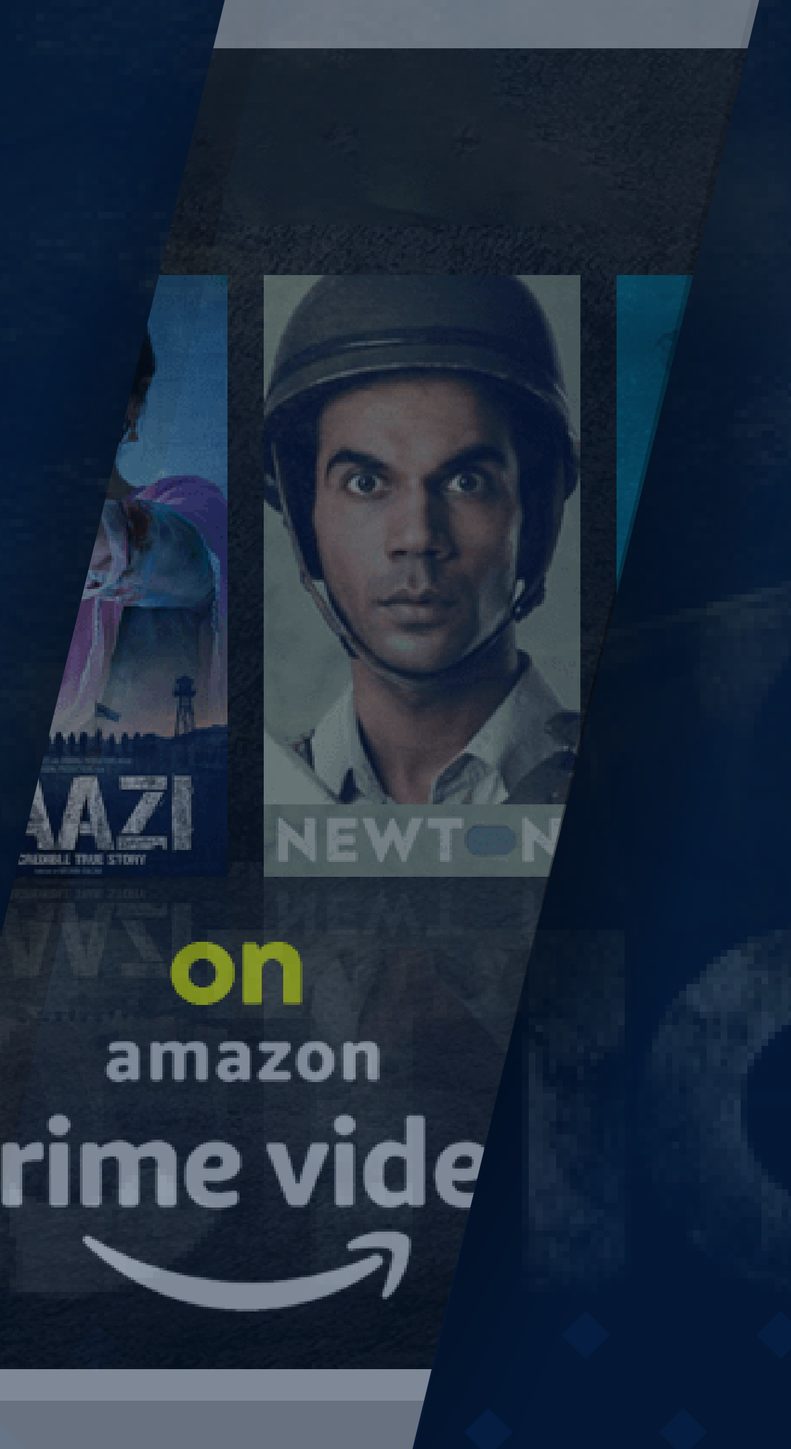
AMAZON PRIME MOVIES AND SHOWS

ANALYSIS

OVERVIEW

INTRODUCTION

- Amazon Prime Video has become a powerhouse in the entertainment industry, offering a vast library of TV shows, movies, and original content.
- Analyzing viewer data can provide valuable insights for content,creators,producers,and platform managers to enhance their offerings and cater to the evolving demands of the audience.



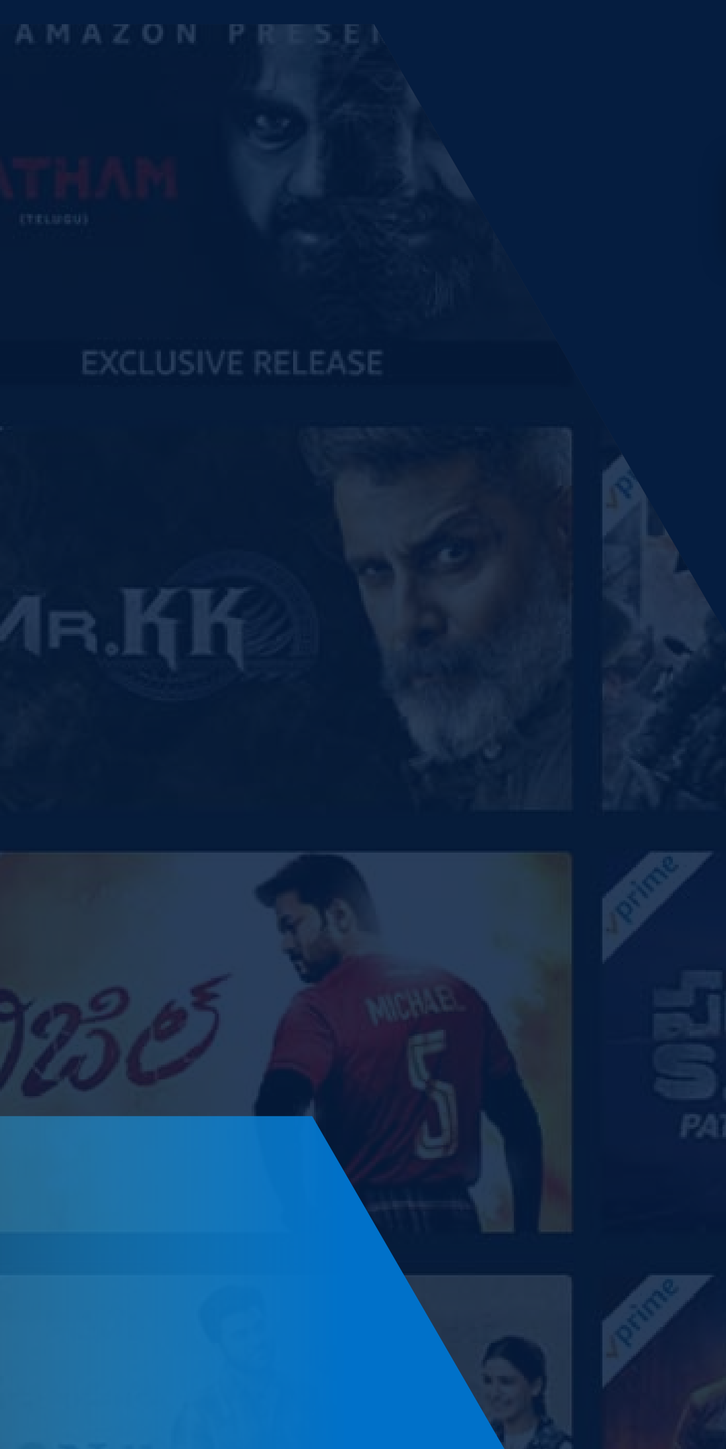
WHY THIS?

- The data we have collected is historical data, as data scientists we can bring out insights from the dataset.
- The insights generated from this analysis can provide valuable strategic direction for Amazon Prime, such as optimizing content acquisition, enhancing user experience, and improving customer retention.
- Since this project includes domains of computer science and statistics, we have decided to explore more and derive useful insights.
- Also, the format of the dataset is compatible with our analysis tools and methods, allowing for seamless integration into our analysis process.

OBJECTIVES AND GOALS

- Exploratory Data Analysis (EDA): Objective: Perform exploratory data analysis to understand the distribution and characteristics of the available movies and shows

- Analyze the distribution of genres, release years, ratings
- Identify Trends
- Identify trends and patterns in user ratings and reviews.
- Visualize key insights using plots and charts.



DATA COLLECTION

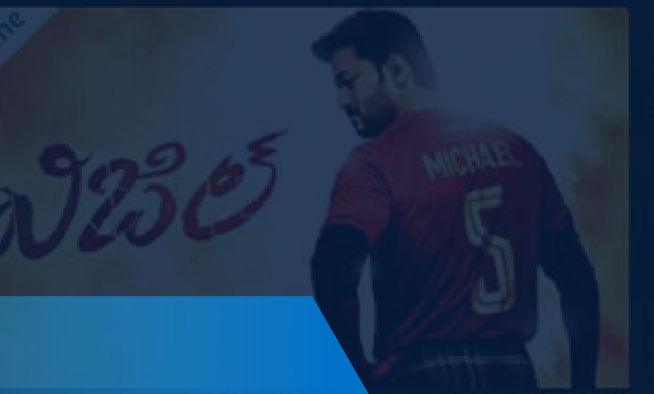
Dataset link: https://www.kaggle.com/datasets/shivamb/amazon-prime-movies-and-tv-shows

The dataset has following attributes:

- show_id: Unique identifier for each entry in the dataset.
- type: Indicates the type of content, such as movie or TV show.
- title: The title of the movie or TV show.
- director: The director(s) of the movie or TV show.
- cast: The cast or actors appearing in the movie or TV show.
- country: The country or countries where the movie or TV show was produced.
- date_added: The date when the movie or TV show was added to the streaming platform.
- release_year: The year when the movie or TV show was originally released.
- rating: The rating assigned to the movie or TV show, such as G, PG, PG-13, etc.
- duration: The duration or length of the movie or TV show.
- listed_in: The category or genre under which the movie or TV show is listed.
- description: A brief description or summary of the movie or TV show.

CLEANING METHOD

- **Handling Missing Values:** Check for missing values in important columns (e.g., show ratings, genres) and decide on a strategy (e.g., imputation, removal)
- **Standardizing Data Formats:** Ensure consistency in data formats (e.g date formats, text casing) for easy analysis.
- **Removing Duplicates:** Check for and remove duplicate entries to avoid skewing analysis results.
- **Dealing with Outliers**: Identify and handle outliers in numerical columns appropriately to prevent them from affecting the analysiis.
- **Addressing Inconsistencies**: Check for inconsistencies in categorical data (e.g., different spellings of the same genre) and standardize them.
- **Normalization:** Normalize numerical data if necessary to ensure fair comparison between different features.
- **Feature Engineering:** Create new features if needed to enhance analysis (e.g., combining genres into broader categories).
- **Data Validation**: Validate data against a known source or dataset to ensure accuracy.
- **Data Transformation:** Transform data into a format suitable for analysis (e.g., converting categorical variables into numerical ones).
- **Quality Check**: Perform a final quality check to ensure the dataset is clean and ready for analysis.



ANALYSIS

Content Trends

Trends: Content addition over time, popular genres, movie/TV show distribution.

Top Contributors

Insights: Prolific directors, actors analyzed for their content impact.

Geographic Distribution

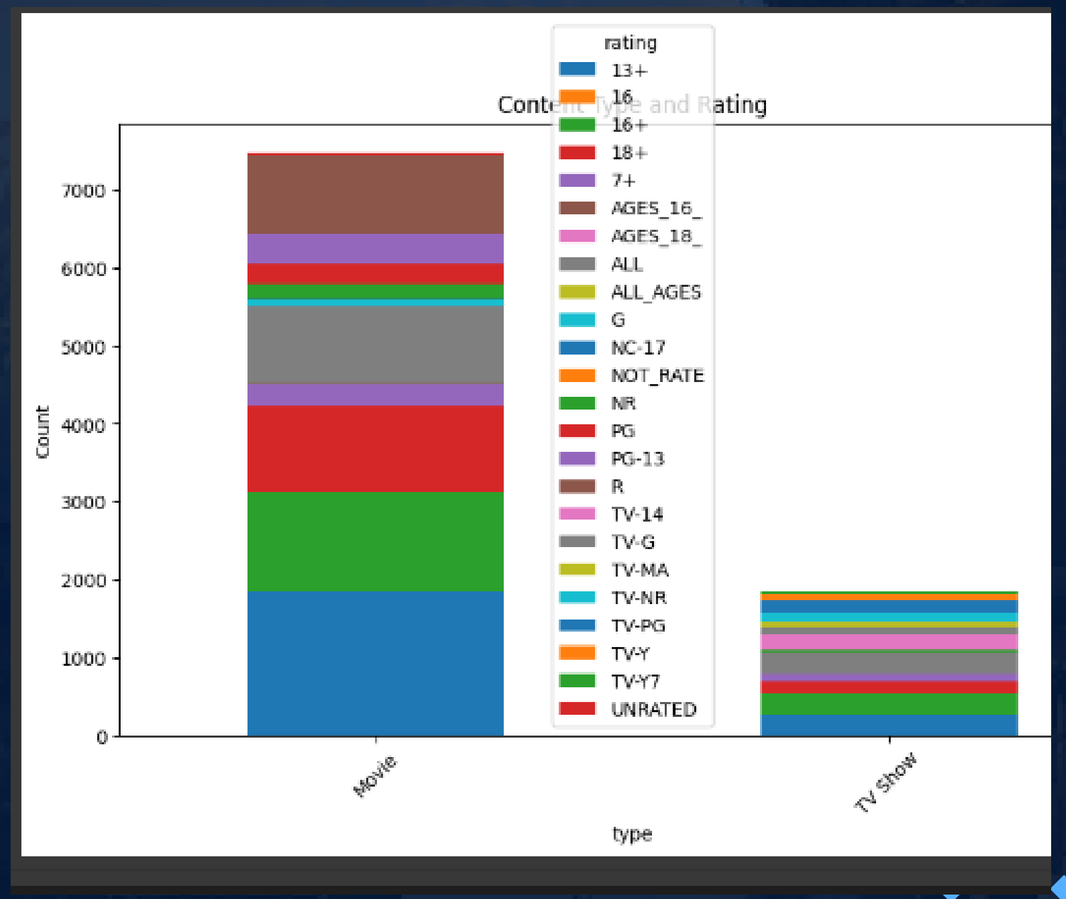
Regional content contributions on Amazon Prime explored, highlighting country preferences.

User Ratings

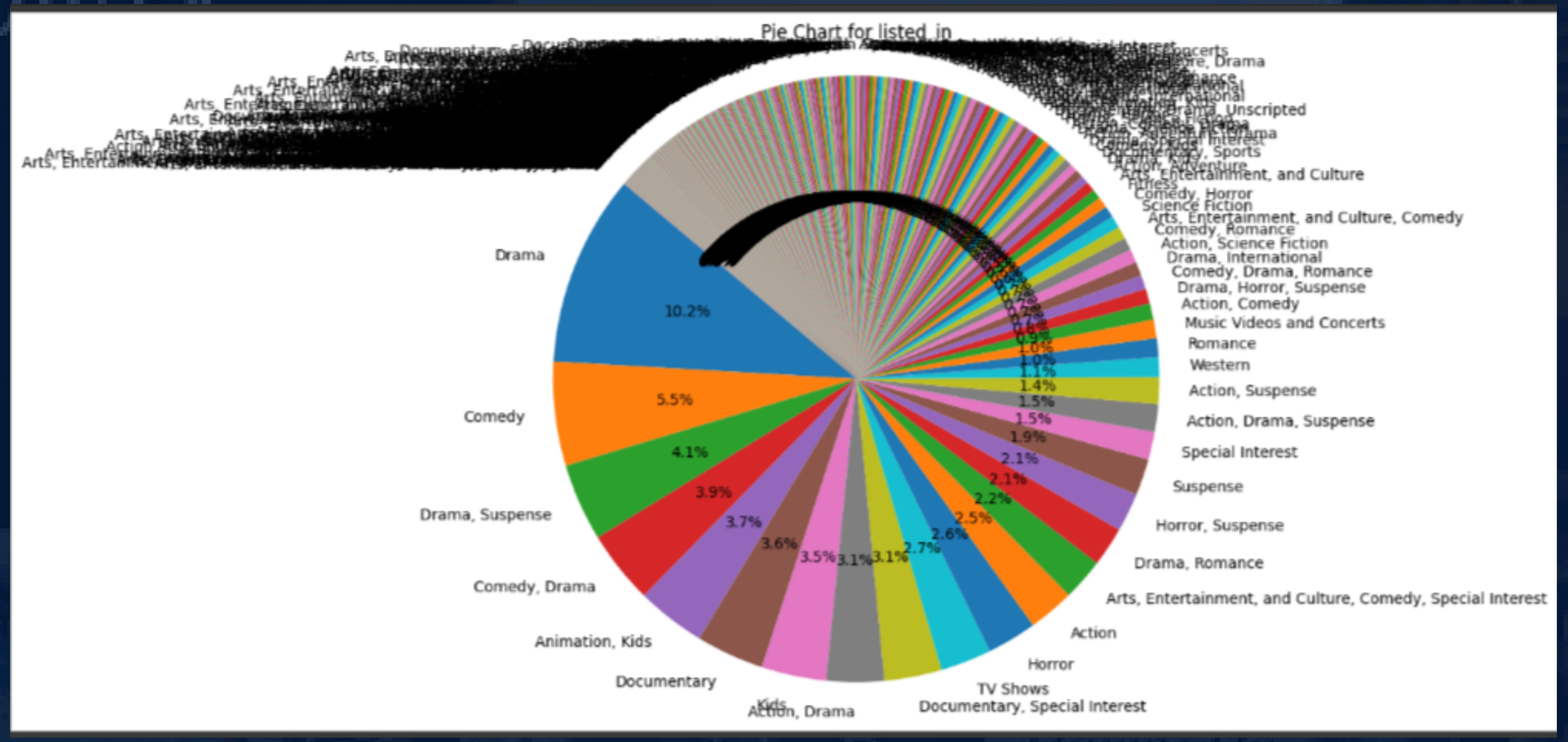
Explored show ratings distribution, gauging user satisfaction and preferences effectively.

EDA - RESULTS INTERPRETATION

- Stacked Bar Plot groups the dataset by both content type (movies or TV shows) and rating categories. It then creates a stacked bar plot where each bar represents a content type (movie or TV show), divided into segments representing different rating categories.
- This visualization allows for a clear comparison of how ratings are distributed across different types of content, aiding in understanding user preferences and satisfaction levels.



This pie -chart counts how often each category appears in the 'listed_in' column, and generates a pie chart to visually represent the proportion of each category. The chart displays labels for categories and their respective percentages, offering insight into content distribution on Amazon Prime.



RECOMMEDATIONS

- **Diversity Content:** Offer a variety of shows across different genres to cater to a wide audience.
- **Promote Top-Rated Shows:** Highlight and promote top-rated shows to attract more viewers.
- **Release New Content Strategically:** Release new shows during peak viewing periods to maximize viewership.
- **Engage with Viewers:** Encourage viewer engagement through interactive features to enhance the viewing experience.
- **Personalize Recommendations:** Use viewer data to personalize content recommendations for a more tailored viewing experience.
- **Monitor Competitor Trends:** Stay informed about trends on other streaming platforms to remain competitive.

CONCLUSION

- In conclusion, this project demonstrates the value of data analysis in understanding and optimizing content offerings on Amazon Prime.
- By exploring trends, preferences, and contributors within the dataset, actionable insights are derived to enhance user experience and strategic decision-making.
- The project highlights the role of data science in driving innovation and competitiveness in the streaming market, ultimately benefiting both the platform and its subscribers.

REFERENCES

https://www.kaggle.com/code/jyotmakadiya/popular-movies-and-tv-shows-data-analysis
https://www.kaggle.com/code/shivamb/amazon-prime-videos-exploratory-analysis
https://analyticswithamruha.hashnode.dev/analysis-of-amazon-prime-movies-and-tv-shows
https://hanidzikra.medium.com/exploratory-data-analysis-using-amazon-prime-video-repository-data-226ad36816d0
https://github.com/iambitttu/Analysis-of-Amazon-Prime-Videos

OUR TEAM

IREDDI RAKSHITHA
21BCE9144

THOKALA VASISRI
21BCE9145

AVANIGADDA SADHANA
21BCE9924

PINDI PRASANTH KUMAR
21BCE9141

DURGA ADITYA
21BCE8620

MONINGI RAMACHANDRA RAO
21BCE9183

BOTCHA TARAKA RUTESHA SAI
21BCE7158

SHIVA KARTHEEK
22BCE8751

VIJAYANAGARA SHARANYA
21BCE8002

INTRAIN
TECH

THANK YOU!!!