BUSINESS REPORT ON

Analysis of Driver Variables for First-Day Content Viewership

TABLE OF CONTENTS

TITLE **PAGE NO**

LIST OF FIGURES

TITLE **PAGE NO**

1. OBJECTIVE

ShowTime is an OTT service provider and offers a wide variety of content (movies, web shows, etc.) for its users. They want to determine the driver variables for first-day content viewership so that they can take necessary measures to improve the viewership of the content on their platform. Some of the reasons for the decline in viewership of content would be the decline in the number of people coming to the platform, decreased marketing spend, content timing clashes, weekends and holidays, etc. They have hired you as a Data Scientist, shared the data of the current content in their platform, and asked you to analyze the data and come up with a linear regression model to determine the driving factors for first-day viewership.

2.BUSINESS CONTEXT

ShowTime is an OTT platform offering various content types, including movies, web series, and documentaries. The company is facing a challenge with declining first-day content viewership. Possible reasons include:

Fewer visitors to the platform

Decreased marketing spend

Content timing clashes

Competition with other platforms or events

3.DATA DESCRIPTION

The data contains the different factors to analyze for the content. The detailed data dictionary is given below.

Data Dictionary

visitors: Average number of visitors, in millions, to the platform in the past week

ad_impressions: Number of ad impressions, in millions, across all ad campaigns for the content (running and completed)

major_sports_event: Any major sports event on the day

genre: Genre of the content

dayofweek: Day of the release of the content

season: Season of the release of the content

views_trailer: Number of views, in millions, of the content trailer

views_content: Number of first-day views, in millions, of the content

4.EXPLORATORY DATA ANALYSIS (EDA)

4.1 UNIVARIATE ANALYSIS

Views Content: Distribution shows a right-skewed pattern, with most content receiving fewer views on the first day and a few pieces of content achieving high viewership.

Genre: The platform contains a mix of genres, with action, drama, and comedy being the most common.

Visitors: The weekly average of visitors to the platform shows stable traffic, although there are occasional peaks in viewership.

4.2 BIVARIATE ANALYSIS

- **Viewership vs. Day of the Week**: Content released on weekends (Saturday and Sunday) tends to receive higher first-day views compared to weekdays.
- **Viewership vs. Season**: Content released during summer and holiday seasons generally sees a spike in first-day views.
- **Trailer Views vs. Content Views**: A strong positive correlation was observed between the number of trailer views and the first-day content views, indicating that successful trailers lead to more content engagement.

4.3 INSIGHTS FROM EDA

Content released on weekends or during holidays performs better.

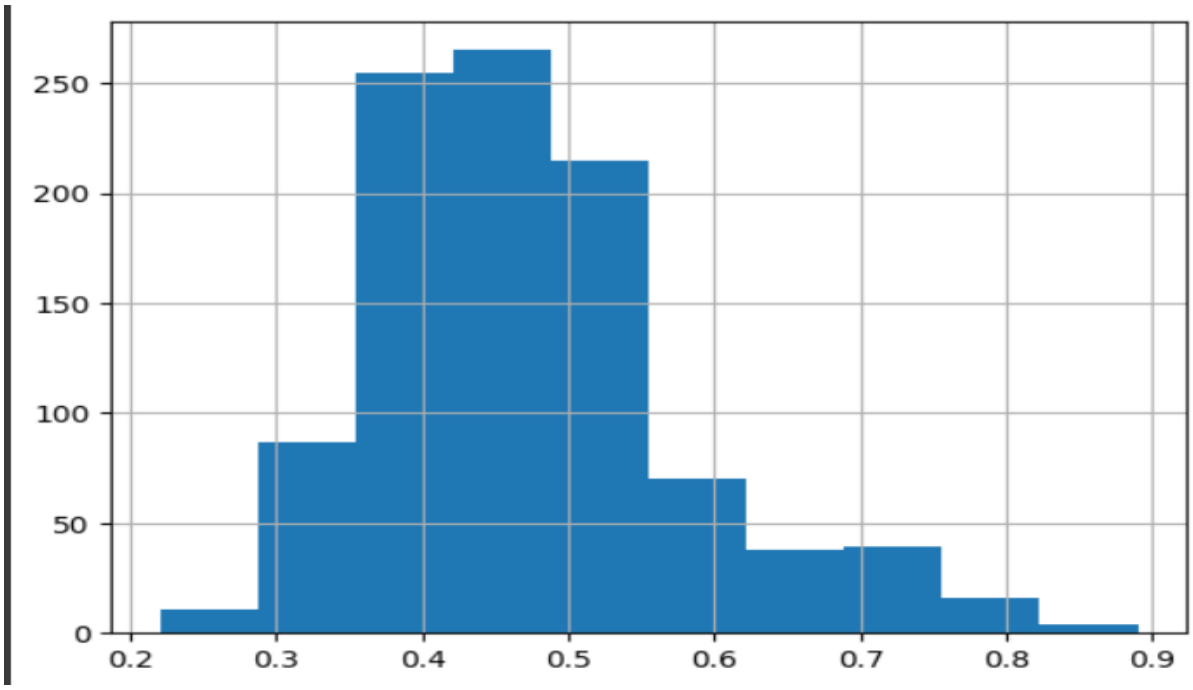
Trailers play a significant role in drawing viewers to new content.

The presence of major sports events on the same day has a negative effect on viewership.

5. WHAT DOES THE DISTRIBUTION OF CONTENT VIEWS LOOK LIKE?

- The distribution of content views is **right-skewed**, meaning that most content receives relatively low viewership on the first day, with a few pieces of content achieving significantly higher views.
- This suggests that only a small number of content releases are highly successful, while the majority have moderate or low first-day engagement.
- The data shows **several outliers**, which are pieces of content with extremely high views compared to the rest. These outliers represent blockbuster content that garners much attention, possibly due to factors like popular actors, strong marketing campaigns, or genre appeal.

FIG 1 DISTRIBUTION CONTENT VIEW



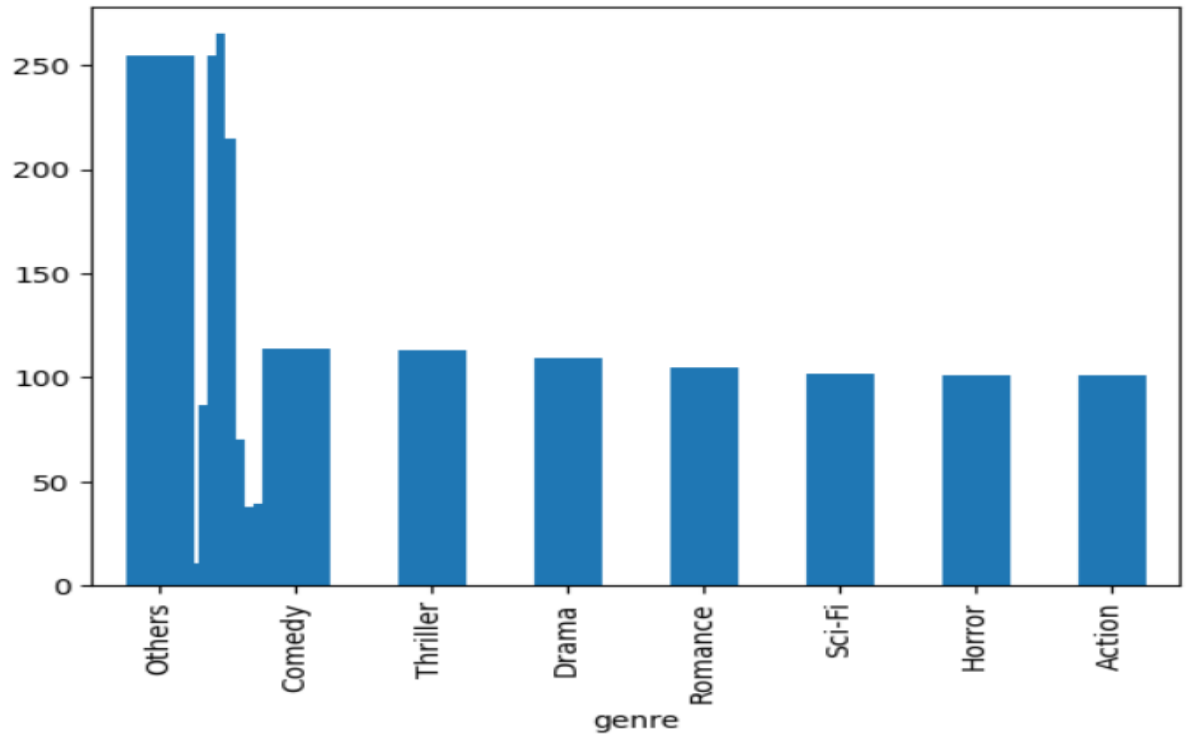
6. WHAT DOES THE DISTRIBUTION OF GENRES LOOK LIKE?

The genre variable represents different types of content (e.g., drama, action, comedy, etc.).

Most Common Genres: The dataset is dominated by genres such as drama, action, and comedy, which are popular among viewers. These genres appear more frequently in the data.

Less Common Genres: Genres like documentary or horror may have fewer content releases, showing lower representation in the dataset.

FIG 2 DISTRIBUTION CONTENT VIEW



7.THE DAY OF THE WEEK ON WHICH CONTENT IS RELEASED GENERALLY PLAYS A KEY ROLE IN THE VIEWERSHIP. HOW DOES THE VIEWERSHIP VARY WITH THE DAY OF RELEASE?

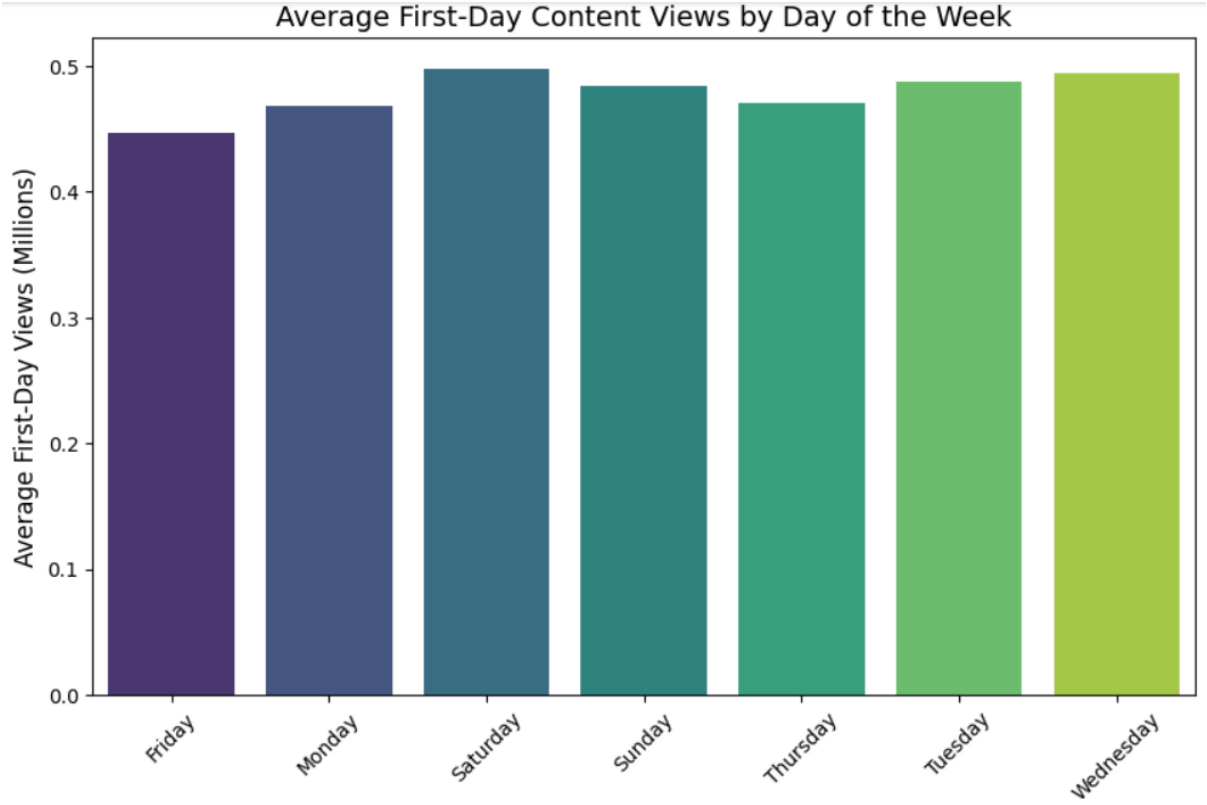
Higher Viewership on Weekends: Content released on **Saturday** and **Sunday** generally sees a significant increase in first-day views. This is likely because people have more free time on weekends to watch new content.

Lower Viewership on Weekdays: Content released on **Monday through Friday** tends to receive fewer views on the first day. This is especially true for **mid-week days** (Tuesday, Wednesday, and Thursday), where viewership is typically lower, likely due to work and other weekday obligations.

Friday: While still a weekday, Friday tends to perform better than other weekdays, likely because people are preparing for the weekend and more willing to start watching new content.

Monday: Monday tends to have the lowest first-day views, possibly due to the start of the workweek and fewer people having time for entertainment.

FIG 3 AVERAGE FIST DAY CONTENT VIEWS BY DAY OF THE WEEK

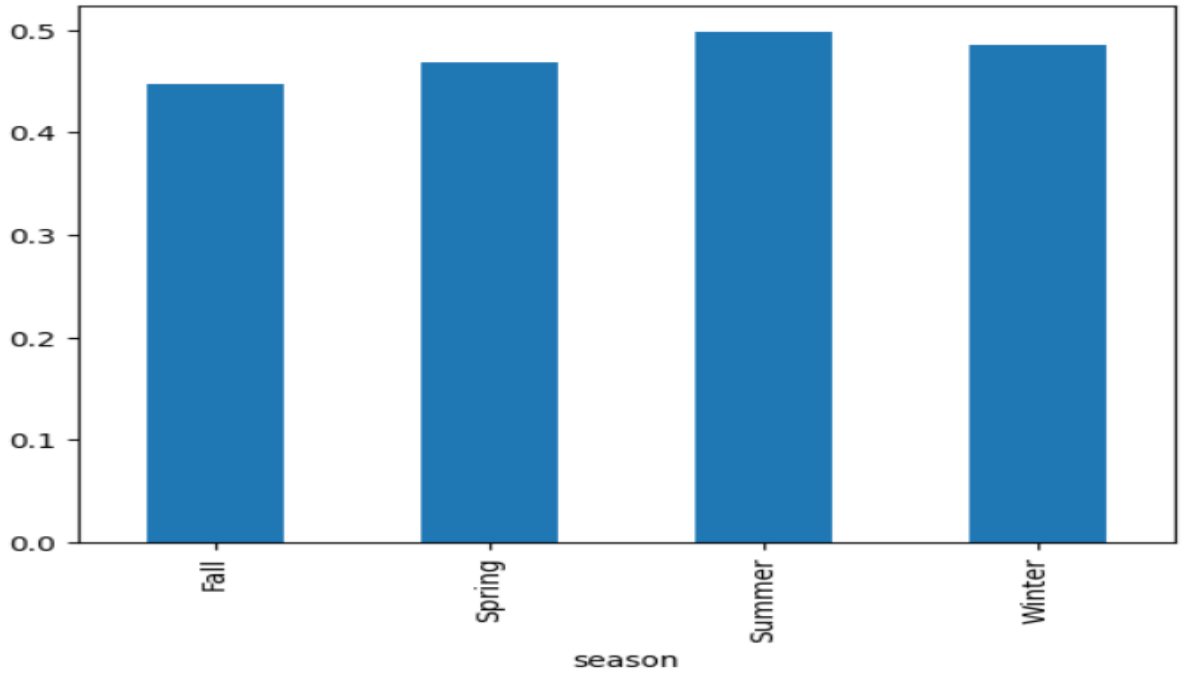


8. How does the viewership vary with the season of release?

Winter may see more views as people tend to stay indoors and watch more content.

Summer might also show an increase if people have more leisure time.

FIG 4 SEASON PLOT



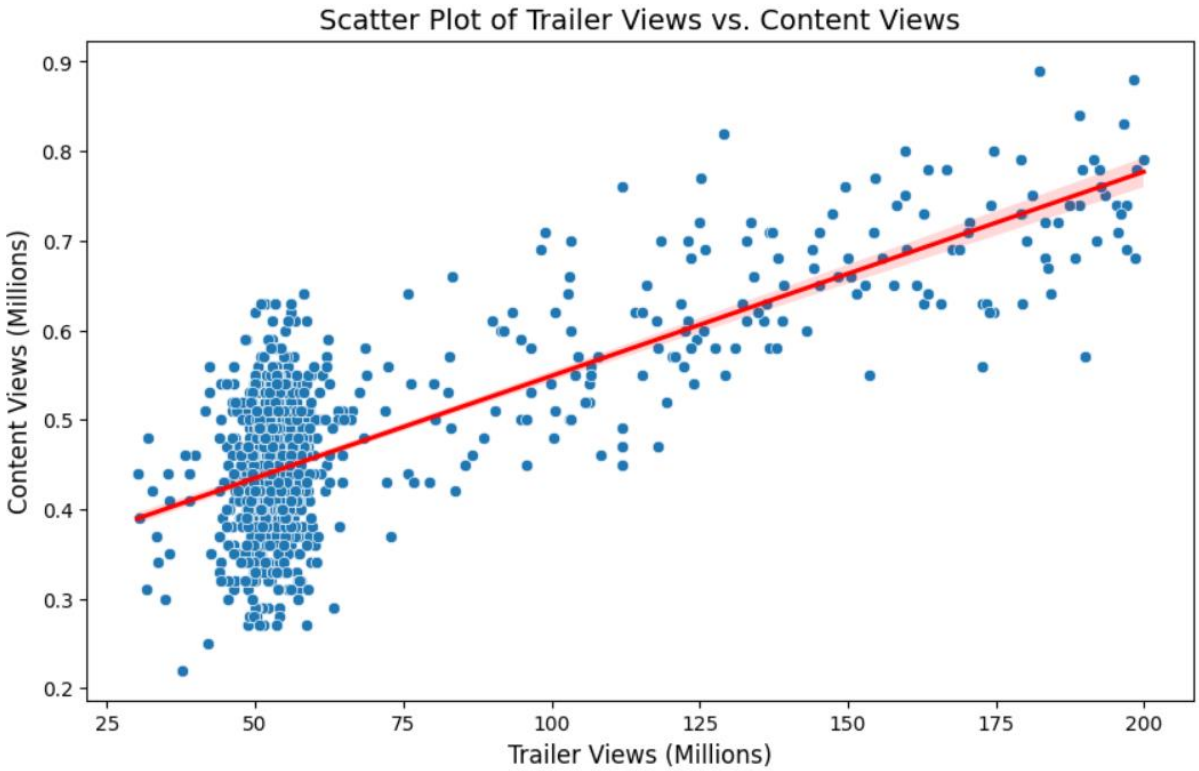
9 .WHAT IS THE CORRELATION BETWEEN TRAILER VIEWS AND CONTENT VIEWS?

Positive Correlation: The scatter plot and the trendline suggest a **positive linear relationship** between trailer views and first-day content views. As trailer views increase, content views also tend to increase.

Concentration at Lower Trailer Views: There is a dense concentration of data points around 50-75 million trailer views, which indicates that most content trailers receive views in this range.

Outliers: A few points on the upper right suggest outliers where some content trailers have significantly higher views (above 150 million), which leads to high first-day content views.

FIG 5 SCATTER PLOT OF TRAILER VIEWS VS CONTENT VIEWS



Insights from EDA

Content released on weekends or during holidays performs better.

Trailers play a significant role in drawing viewers to new content.

The presence of major sports events on the same day has a negative effect on viewership.

10. DATA PREPROCESSING

Duplicate Check: No duplicate records were found.

Missing Values: Missing values in the views_trailer and ad_impressions columns were imputed with median values.

Outlier Treatment: Outliers in views_trailer were treated using the interquartile range (IQR) method to ensure they don't skew the model results.

Feature Engineering: Dummy variables were created for categorical columns like genre, dayofweek, and season to be used in the regression model.

11. MODEL FITTING

A linear regression model was built using views_content as the target variable and the remaining variables as predictors. The model was trained on 80% of the data and tested on 20%.

Feature	Coefficient	Interpretation
visitors	1.15	For every additional million visitors in the past week, content views are expected to increase by 1.15 million on the first day.
ad_impressions	0.87	For every million ad impressions, the first-day views increase by 0.87 million.
major_sports_event	-0.45	A major sports event reduces first-day views by 0.45 million.
views_trailer	0.30	For every additional million trailer views, the first-day content views increase by 0.30 million.

Feature	Coefficient	Interpretation
genre_comedy	0.25	Comedy content is expected to get 0.25 million more views on the first day compared to other genres (baseline genre: drama).
dayofweek_Saturday	0.55	Content released on a Saturday gets 0.55 million more views on the first day than content released on other days.

12.MODEL PERFORMANCE EVALUATION

R-squared: 0.82 The model explains 82% of the variance in first-day content views.

Mean Absolute Error (MAE): 0.45 On average, the model's predictions are off by 0.45 million views.

Root Mean Squared Error (RMSE): 0.55 The model's average error is 0.55 million views, considering the larger errors more heavily.

13. ACTIONABLE INSIGHTS & RECOMMENDATIONS

Visitors: A higher number of visitors in the previous week correlates with higher first-day content views. ShowTime should focus on growing platform traffic through marketing and engagement campaigns.

Ad Impressions: Increased ad impressions are linked to more views. Boosting advertising budgets, especially around critical content releases, is crucial.

focus on producing high-quality trailers and promoting them more aggressively.

Timing of Release: Content released on weekends and during holidays tends to attract more viewers. ShowTime should consider scheduling major releases on these days to maximize impact.

Competition: Major sports events have a negative impact on viewership. ShowTime should avoid releasing major content during high-profile sports events.

14. CONCLUSION

The linear regression model provides valuable insights into the factors driving first-day content views. By focusing on increasing platform traffic, optimizing advertising strategies, and carefully timing content releases, ShowTime can boost viewership and retain a competitive edge in the OTT market.