**Second-screen Experience Analytics - Entertainment Sector**

Pooja Joshi & E.Code [E25007]

# Overview :

# This task focuses on analyzing second-screen engagement by examining key content metrics such as popularity, vote average, attendance count, and sentiment. Visualizations are used to highlight insights into attendance by sentiment, distribution of user reviews, and relationships between content metrics and sentiment. Predictive models and statistical analysis are applied to identify significant drivers of second-screen engagement.

# Objective:

The main objective is to analyze relationships between content metrics and user sentiment, followed by predictive modeling to identify key factors contributing to second-screen engagement. The goal is also to visualize insights around attendance and sentiment scores.

# Assigned Task(s) :

· Analyze content metrics such as popularity, vote average, attendance count, and sentiment.

· Create visualizations to highlight key insights.

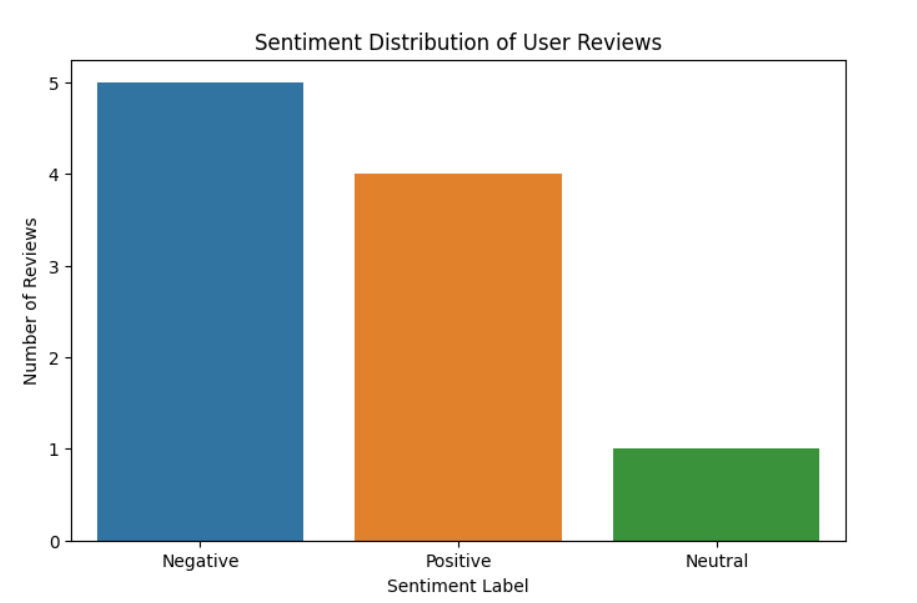
· Perform linear regression to predict attendance and sentiment from key metrics.

# Task Details :

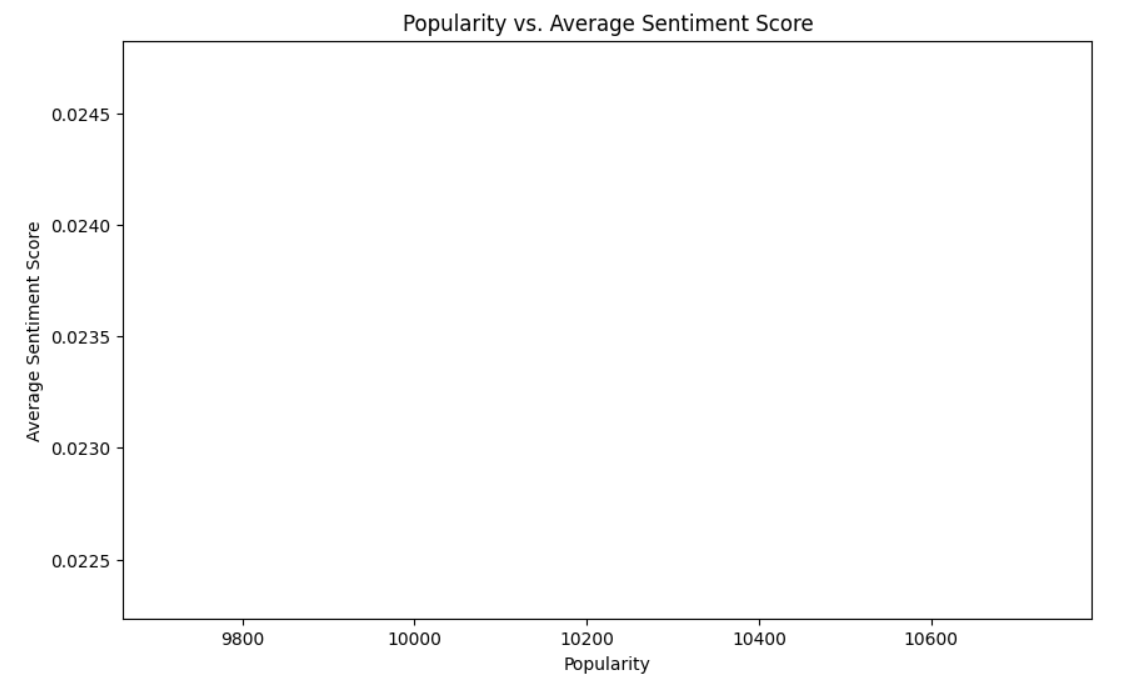
**Task 1: Analyze correlation between content metrics and second-screen engagement.**  
● Status: Completed  
● Details: A correlation matrix was created to analyze the relationships between popularity, vote average, attendance count, and sentiment scores. This helped identify the strength of relationships across these variables.

**Task 2: Create Visualizations for Sentiment and Engagement Analysis.**  
● Status: Completed  
● Details: Several key visualizations were created:

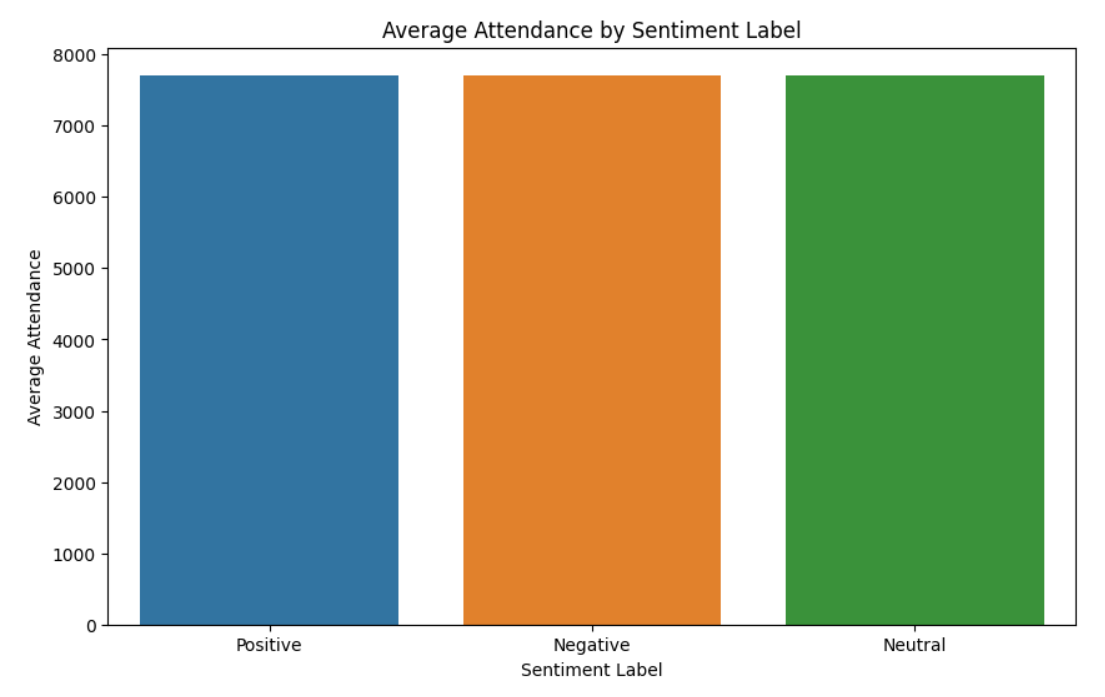
* **Average Attendance by Sentiment Label**: Shows how user sentiment (positive, negative) correlates with attendance figures.



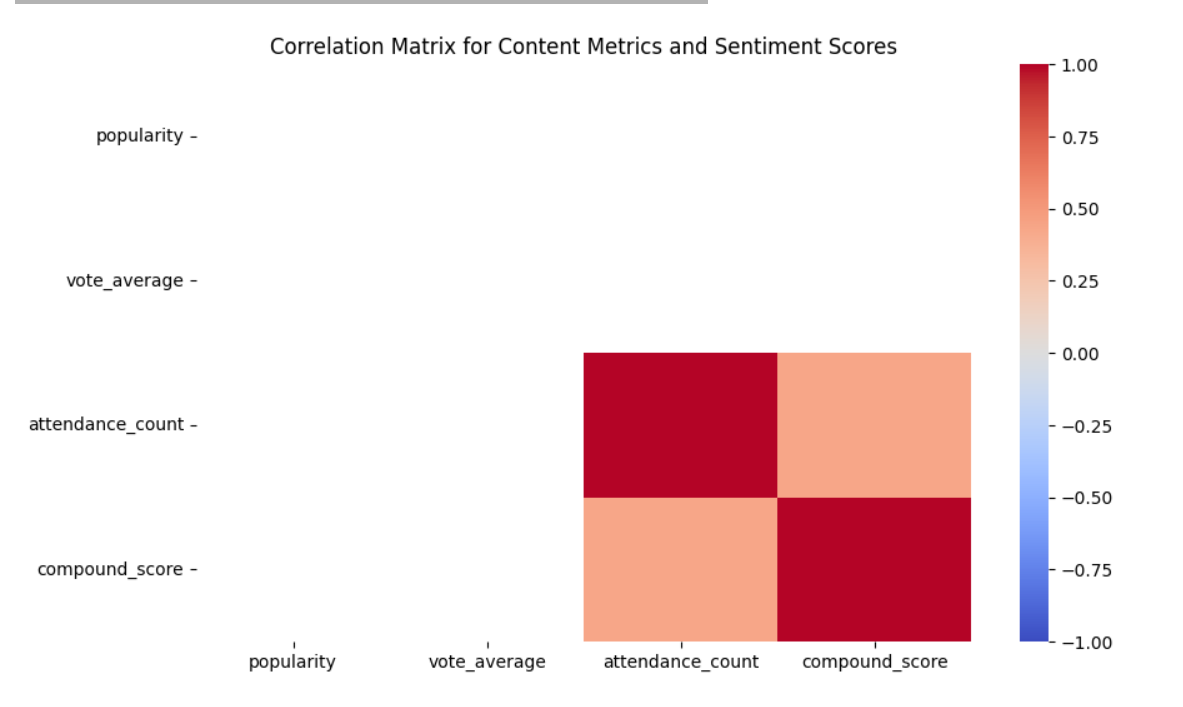
* **Sentiment Distribution of User Reviews**: Visualizes the breakdown of user sentiment, highlighting the proportion of positive vs. negative feedback.



* **Popularity vs. Average Sentiment Score**: Analyzes how a movie's popularity correlates with its sentiment score.



* **Correlation Matrix for Content Metrics and Sentiment Scores**: Visualizes correlations between all key metrics (popularity, vote average, attendance count, sentiment score).



**Task 3: Linear Regression to predict attendance count based on content metrics.**  
● Status: Completed  
● Details: Linear regression was applied to predict attendance count using the popularity and vote average. Model performance was evaluated based on R² and MSE, showing high predictive strength.

**Progress :**

**·Accomplishments:**

● Successfully cleaned the data and prepared it for time series modeling.  
● Fitted an ARIMA model and generated forecasts for future popularity trends.  
● Created a detailed visualization that compares historical and forecasted data, aiding in understanding potential future content demand.

**Metrics:**

● Data was resampled on a monthly basis.  
● ARIMA Model configuration: (5,1,0)  
● Forecasting time horizon: 12 months  
● Popularity trend visualization generated.

# Challenges and Solutions :

#### ****Challenges Faced:****

● Handling missing data within the popularity series, which could skew model accuracy.

#### ****Solutions Implemented:****

● Missing data was addressed by dropping null values from the time series. A sensitivity analysis will be performed later to assess the impact of different data imputation methods.

# Next Steps :

#### ****Upcoming Tasks:****

● Review the forecasting results with stakeholders to validate accuracy and determine its impact on licensing strategies.  
● Explore alternative ARIMA model configurations to ensure optimal forecast accuracy.

#### ****Goals:****

● Improve the accuracy of the forecast by testing additional parameters or alternative forecasting models such as SARIMA.  
● Incorporate external factors, such as marketing spend or major movie releases, to refine the popularity forecast.

# Conclusion :

# Summary: This report outlines the application of the ARIMA model to forecast movie popularity, providing valuable insights into potential future content demand. The next steps will focus on refining the model and validating its accuracy against real-world factors.

# **Acknowledgments**: Thank the audience for their time and attention.