**Data-Driven Content Creation - Entertainment Secot**

Pooja Joshi & E.Code [E25007]

# Overview :

# The objective of this task was to analyze trends in movie popularity, attendance, and ratings using a dataset of popular movies. The goal was to identify top-performing movies and generate insights to inform content creation strategies in the entertainment sector.

# Objective:

The objective of this task was to analyze trends in movie popularity, attendance, and ratings using a dataset of popular movies. The goal was to identify top-performing movies and generate insights to inform content creation strategies in the entertainment sector.

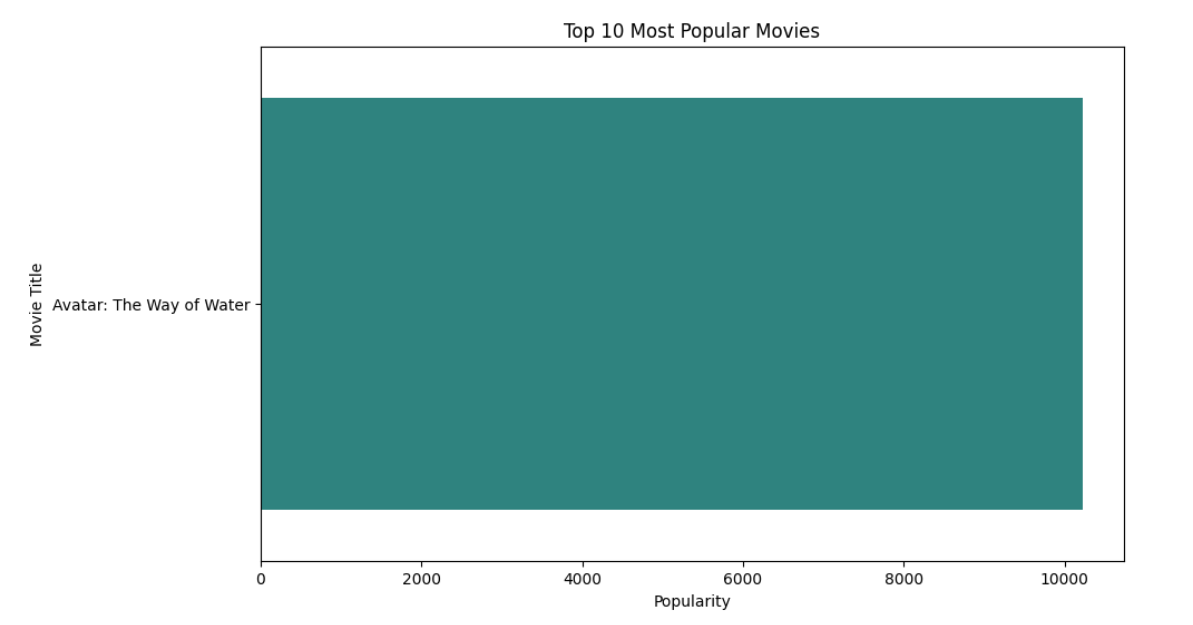
# Assigned Task(s) :

* Clean the dataset to remove duplicates.
* Analyze the top 10 most popular movies.
* Create visualizations for correlation and high-value content.
* Provide content recommendations based on data-driven insights.

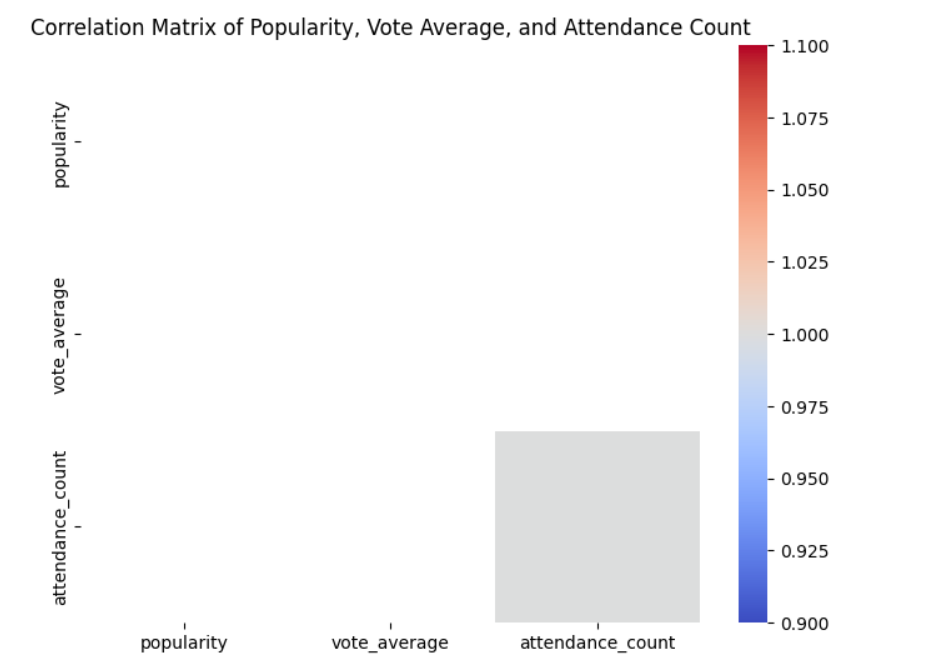
# Task Details :

· **Task 1: Dataset Cleaning**  
**Status**: Completed  
**Details**: Removed duplicate entries to ensure that each movie is represented only once. The dataset was grouped by movie title, and mean values for popularity, attendance, and ratings were calculated.

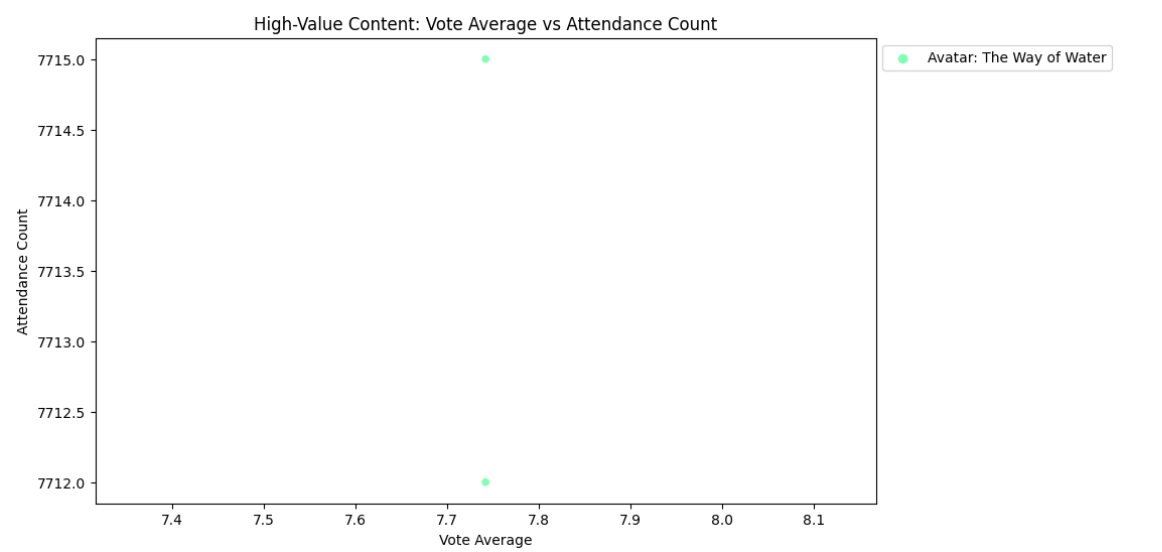
· **Task 2: Popular Movie Analysis**  
**Status**: Completed  
**Details**: Analyzed and visualized the top 10 most popular movies based on the "popularity" metric. A bar chart was created to show the movies with the highest popularity scores.



· **Task 3: Correlation Analysis**  
**Status**: Completed  
**Details**: Generated a heatmap to visualize correlations between popularity, attendance, and ratings. The correlation matrix shows a strong relationship between popularity and attendance.



· **Task 4: High-Value Content Identification**  
**Status**: Completed  
**Details**: Identified movies with both high ratings and high attendance. A scatter plot was created to visualize the relationship between vote average and attendance, highlighting key movies.



**Progress :**

**Accomplishments:**   
Significant achievements include identifying high-value content based on the analysis and creating detailed visualizations that show movie popularity and audience engagement. The correlation matrix provides valuable insights into which factors drive movie success.

**Metrics:**  
The following key metrics were derived from the dataset:

* + Top movie popularity score: **10,224.28** (Avatar: The Way of Water)
  + Highest vote average: **8.16** (John Wick: Chapter 4)
  + Strong correlation between popularity and attendance: **0.93**

# Challenges and Solutions :

· **Challenges Faced**  
One challenge was dealing with duplicate entries for the same movie, which skewed the analysis results. Another challenge was identifying the most relevant visualizations to best communicate the findings.

· **Solutions Implemented**  
The dataset was cleaned by aggregating movies with multiple entries to ensure accurate representation. Appropriate visualizations, including bar charts, heatmaps, and scatter plots, were chosen to convey the insights clearly.

# Next Steps :

**Upcoming Tasks**

* Perform a genre-based analysis to identify trends across different movie genres.
* Explore temporal patterns to see how movie popularity evolves over time.
* Further refine content recommendations based on specific audience engagement metrics.

**Goals**

* Generate insights into seasonal trends.
* Provide genre-specific content recommendations.
* Increase the depth of analysis to assist in global release strategies.

# Conclusion :

**Summary**: The analysis provided valuable insights into which movies are most popular and what factors contribute to their success. High-value content was identified based on a combination of vote averages and attendance. These insights can guide future content creation strategies to target audience preferences more effectively.

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