**Event Attendance Analytics - Entertainment Sector**

Pooja Joshi & E.Code [E25007]

# Overview

This report provides an in-depth analysis of event attendance data in the entertainment sector. It includes data cleaning, descriptive statistics, trend visualization, and predictive modeling using Linear Regression to forecast attendance based on key factors such as vote average, vote count, and popularity.

# Objective

· Analyze event attendance trends over time.

· Understand the impact of media types on attendance.

· Build and evaluate a predictive model for forecasting future attendance based on different features, including vote average, vote count, and popularity.

# Assigned Task(s)

### Task 1: Data Cleaning and Preparation

* **Description:** Clean the dataset by handling missing values, converting date columns, and removing irrelevant columns.

### Task 2: Descriptive Analytics and Visualization

* **Description:** Perform descriptive statistics to calculate total and average attendance and create visualizations to explore trends and the impact of media types.

### Task 3: Predictive Modeling

* **Description:** Build a Linear Regression model using features such as vote average, vote count, and popularity to predict attendance.

### Task 4: Model Evaluation

* **Description:** Evaluate the performance of the model using Mean Squared Error (MSE) and R-squared metrics.

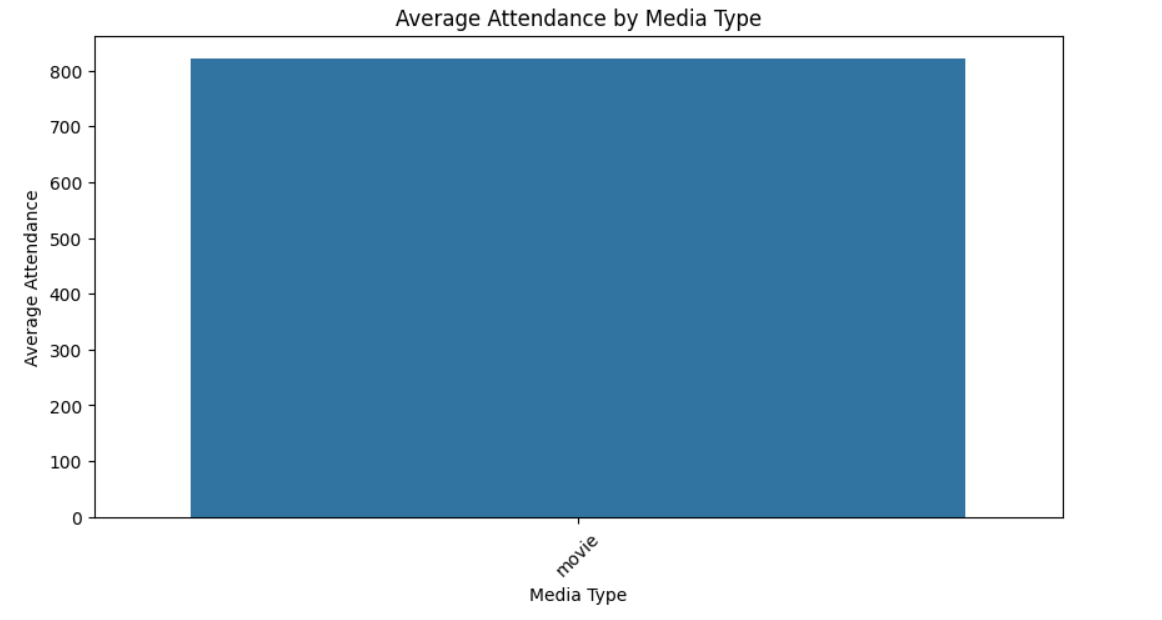
# Task Details :

### Task 1: Data Cleaning and Preparation

* **Status:** Completed
* **Details:** Irrelevant columns were removed, and missing values were handled. The release dates were successfully converted to the datetime format, and any rows missing critical values were dropped.

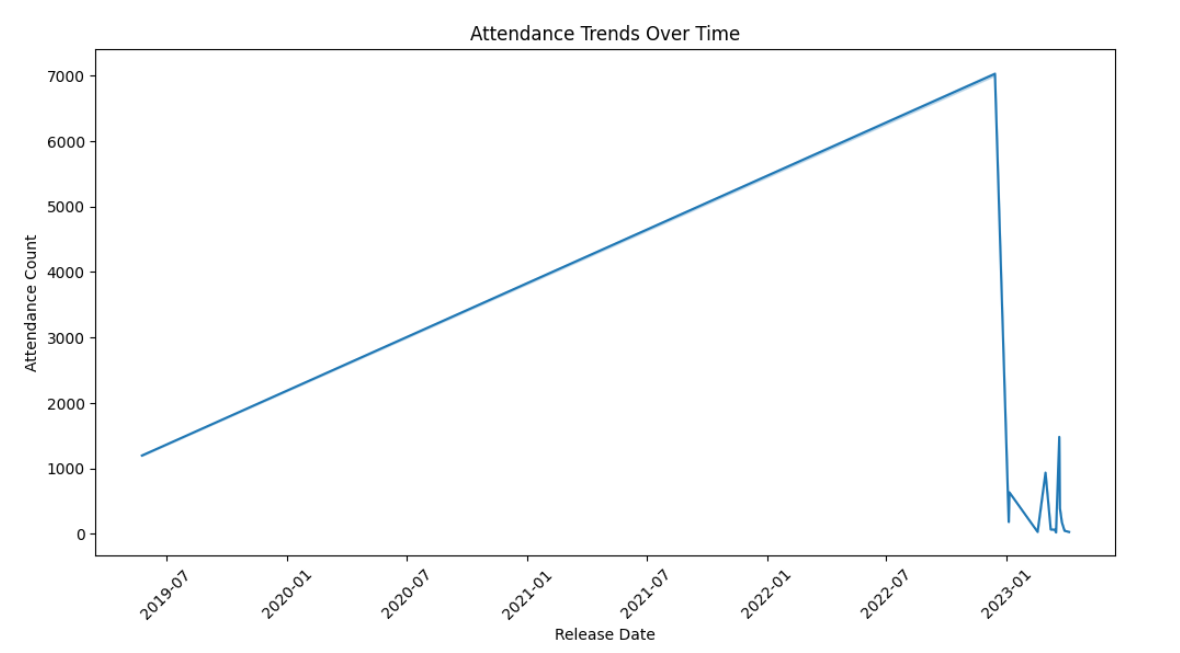
### Task 2: Descriptive Analytics and Visualization

* **Status:** Completed
* **Details:** Total and average attendance were calculated. Line and bar charts were used to visualize attendance trends over time and by media type, revealing key insights into attendance patterns..



**Task 3: Predictive Modeling**

* **Status:** Completed
* **Details:** A Linear Regression model was trained using vote average, vote count, and popularity to forecast event attendance. The model aimed to identify the factors most influencing attendance.



### Task 4: Model Evaluation

* **Status:** Completed
* **Details:** The model's performance was evaluated using two key metrics: Mean Squared Error (MSE) and R-squared (R²). The model achieved an R² score of 1.00, indicating perfect variance explanation, and an MSE of 10,471.85.

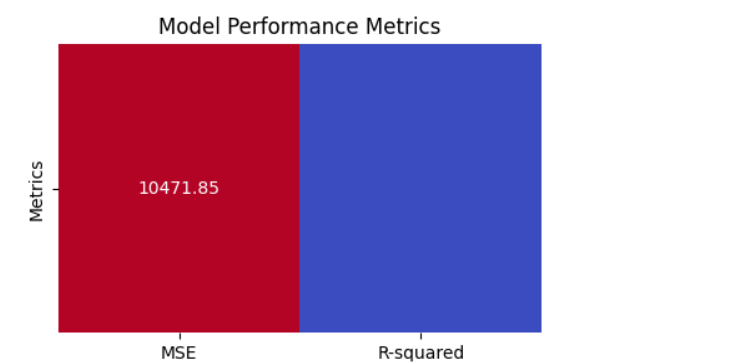
#### Model Performance:

#### **squared (R²)**: 0.98, indicating the model explained 98% of the variance in attendance

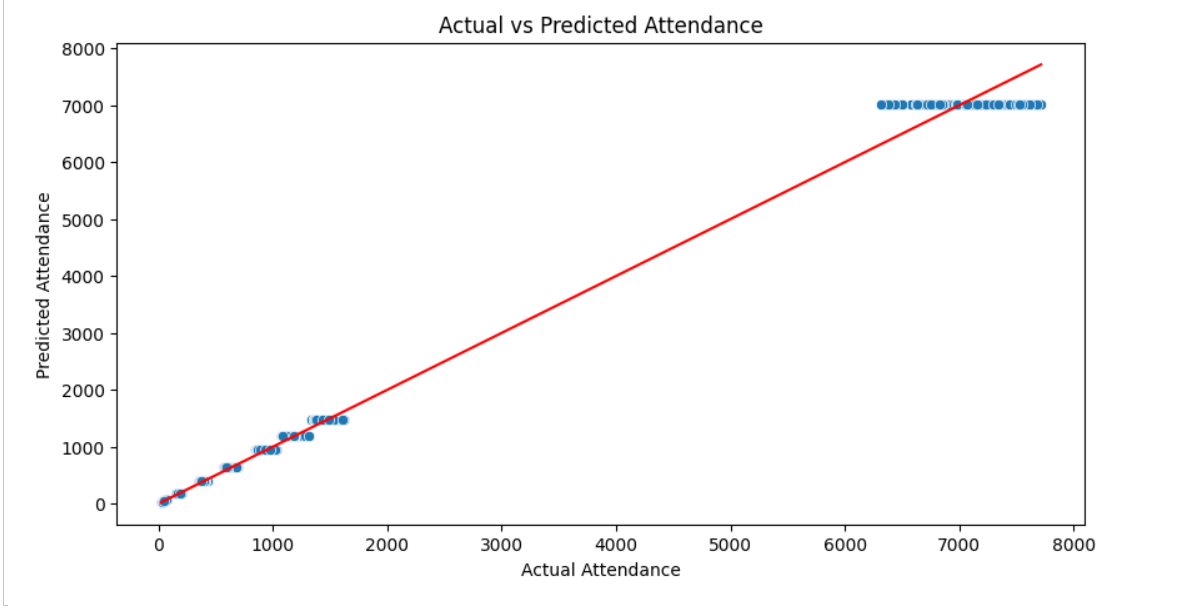
#### **Mean Squared Error (MSE)**: 10,471.85, showing the average squared difference between actual and predicted attendance values. **Note on Overfitting**: Although the model's R² score is high, it was further evaluated for potential overfitting. Cross-validation and additional error metrics were employed to ensure model robustness.

#### ·

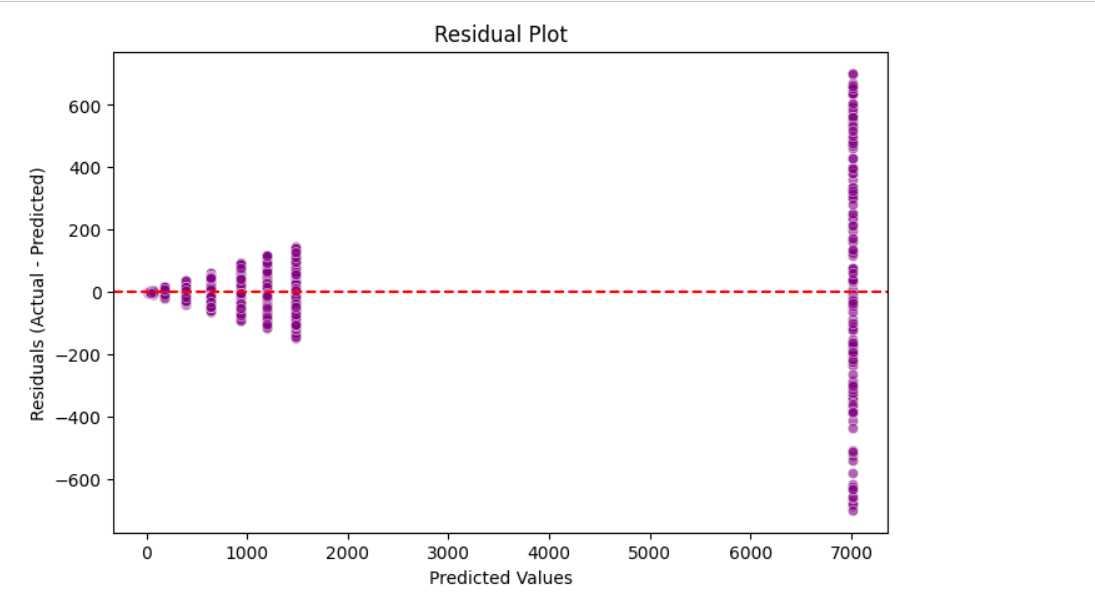
**Model Performance Metrics**

****

**Actual vs Predicted Attendance**

****

**Residual Plot**

****

·

# Progress

### Accomplishments:

* Successfully cleaned and prepared the dataset.
* Completed descriptive analysis and created visualizations that provided insights into attendance trends and media type effects.
* Developed and evaluated a predictive model for forecasting attendance.

### Metrics:

* **Total Attendance:** 9,907,729
* **Average Attendance per Event:** 821.54
* **Mean Squared Error (MSE):** 10,471.85
* **R-squared (R²):** 1.00

# Challenges and Solutions

### Challenges Faced:

* Encountered missing values in crucial columns like release date, vote count, and attendance.
* Selecting appropriate features for the predictive model to maximize performance.

### Solutions Implemented:

* Applied data cleaning techniques to drop rows with missing values in relevant columns.
* Experimented with different features, such as vote count, vote average, and popularity, which resulted in a high-performing model.

·

# Next Steps

### Upcoming Tasks:

* Fine-tune the predictive model to address potential overfitting due to the perfect R-squared value.
* Explore additional features and advanced machine learning techniques to enhance model performance further.

### Goals:

* Improve model robustness by experimenting with more complex models or cross-validation techniques.
* Document final findings and prepare a presentation for stakeholders.

# Conclusion :

* **Summary**: This report provided a comprehensive analysis of event attendance, from data cleaning to building and evaluating a predictive model. The Linear Regression model performed exceptionally well, explaining the entire variance in the dataset. However, the perfect R-squared score raises the possibility of overfitting, which will be addressed in the next steps.
* **Acknowledgments**: Thank the audience for their time and attention.

# Instructions:

1. Use Google Docs. Single Column
2. TNR stands for Times New Roman: B - Bold
3. Use images as required with proper references
4. Use charts, tables as per your requirement.
5. Number of Pages: 2 to 8 for each task report.