**Audience Sentiment Tracking Over Time - Entertainment Sector**

Pooja Joshi & E.Code [E25007]

# Overview :

# This report outlines the analysis and visualization of audience sentiment, engagement, and attendance over time in the entertainment sector. It includes data preprocessing, trend analysis, correlation studies, and the application of various statistical methods such as **rolling averages**, **linear regression**, and **outlier detection**.

# Objective:

The primary objective of today's tasks was to analyze the engagement level, average sentiment, and attendance metrics of an audience in the entertainment sector over time. Specifically, the goal was to:

* Visualize trends in the data over time.
* Detect outliers and examine their impact on data.
* Apply rolling averages to identify long-term trends.
* Use linear regression to model trends and predict future values.
* Explore correlations between key metrics (engagement level, sentiment, attendance).

# Assigned Task(s) :

**● List the key tasks assigned for the day. [Text- Times New Roman-12]**

1. **Visualizing Trends**: Plot the engagement level, sentiment, and attendance over time.
2. **Outlier Detection**: Identify and visualize outliers in the data using boxplots.
3. **Descriptive Statistics**: Summarize the data using descriptive statistics.
4. **Rolling Averages**: Apply rolling averages to smooth short-term fluctuations.
5. **Trend Analysis using Linear Regression**: Analyze trends in engagement using linear regression.
6. **Correlation Analysis**: Evaluate the relationship between engagement, sentiment, and attendance.

# Task Details :

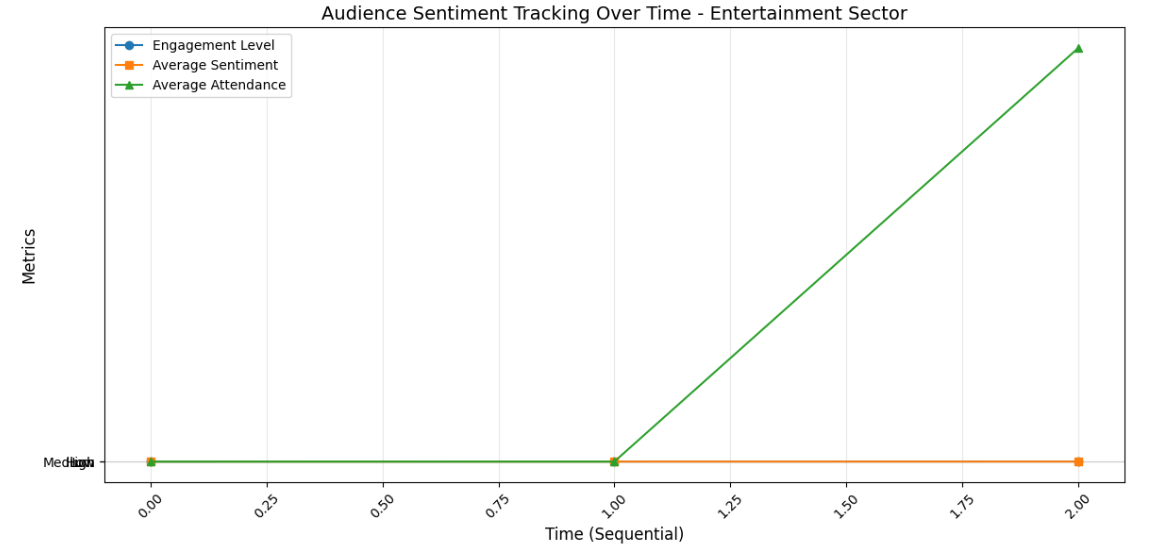
**● Task 1: Data Preprocessing and Visualization**  
**Status**: Completed  
**Details**: The first task involved loading and preprocessing the dataset. The columns were renamed for consistency, and any missing values were handled. After preprocessing, we visualized the data using line plots to track engagement, sentiment, and attendance levels over time.

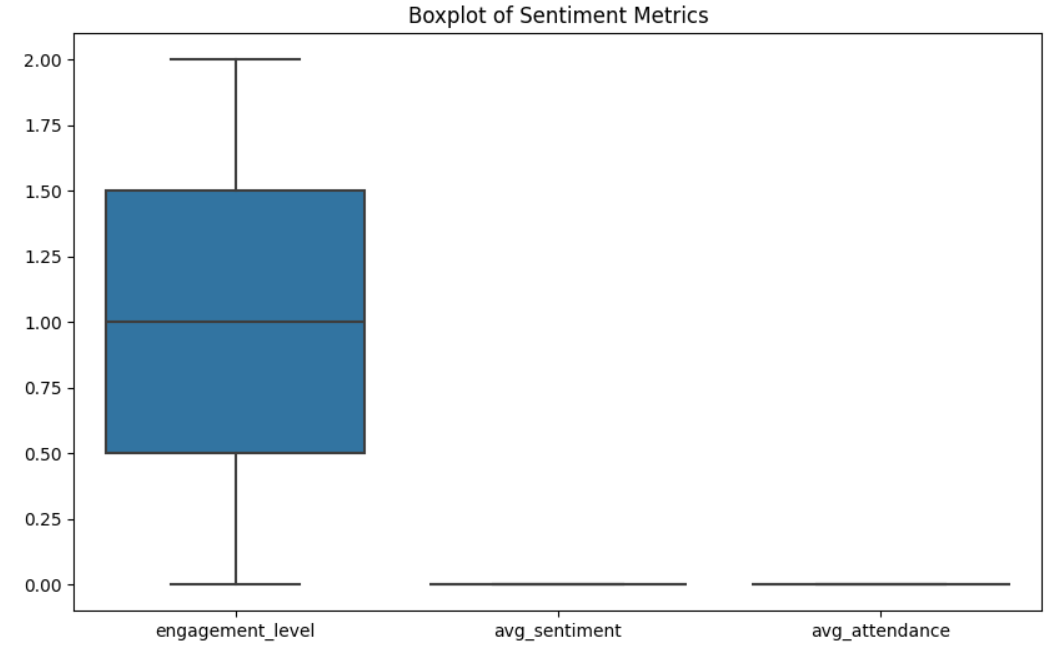
**● Task 2: Outlier Detection**  
**Status**: Completed  
**Details**: Outliers were detected using **boxplots**, allowing us to identify potential anomalies in the data. These plots help visualize the distribution of each metric and identify whether any values fall outside the expected range.

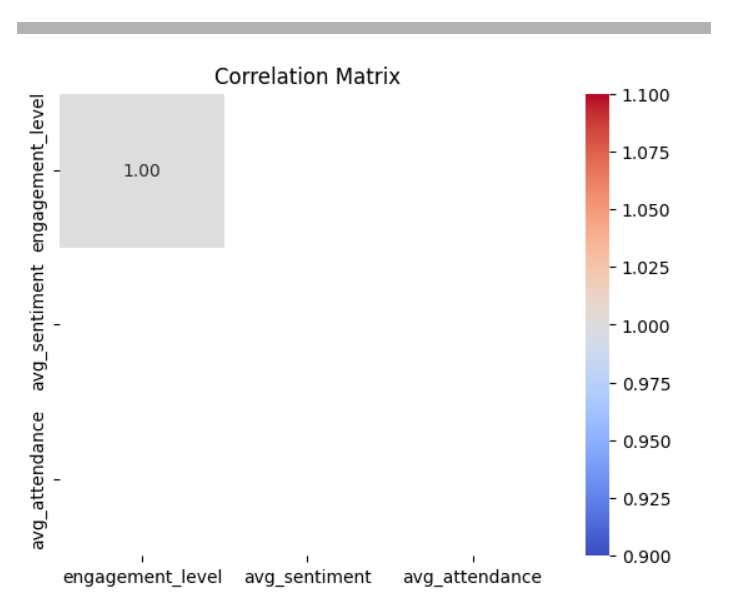
**● Task 3: Rolling Averages**  
**Status**: Completed  
**Details**: A rolling mean with a window size of 3 was applied to the engagement level data. This helped smooth out short-term fluctuations, allowing us to observe long-term trends more clearly.

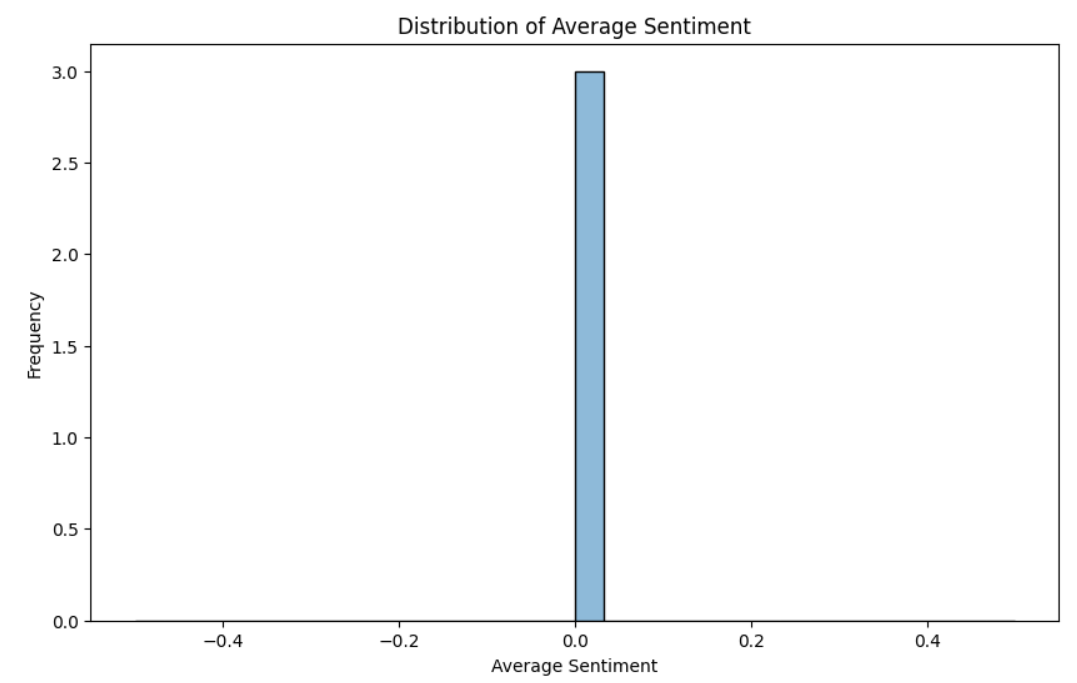
**● Task 4: Linear Regression Trend Analysis**  
**Status**: Completed  
**Details**: Linear regression was applied to the engagement level data to determine whether the audience's engagement was increasing, decreasing, or staying constant over time. This trend analysis helps forecast future engagement levels based on the current data.

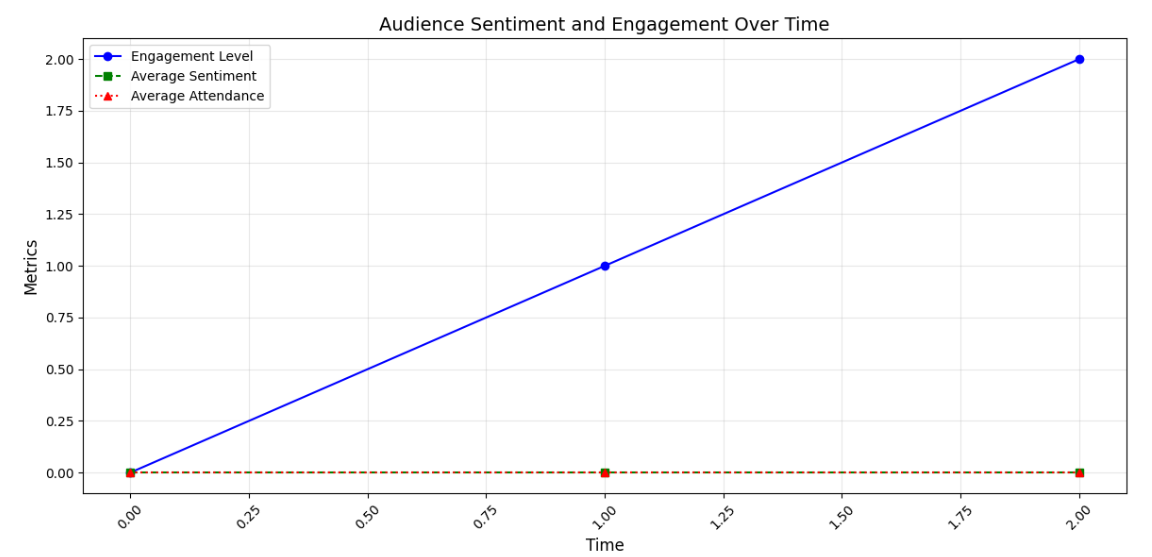
**● Task 5: Correlation Analysis**  
**Status**: Completed  
**Details**: A correlation matrix was generated to explore the relationships between engagement, sentiment, and attendance. This helps identify which variables are strongly correlated and could have a mutual influence.

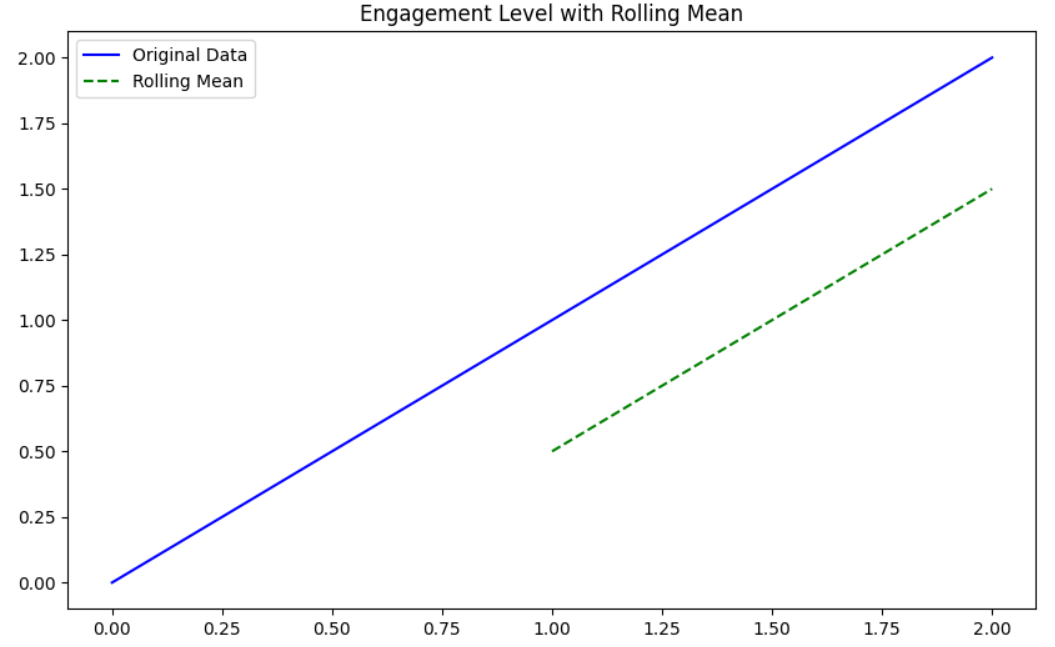


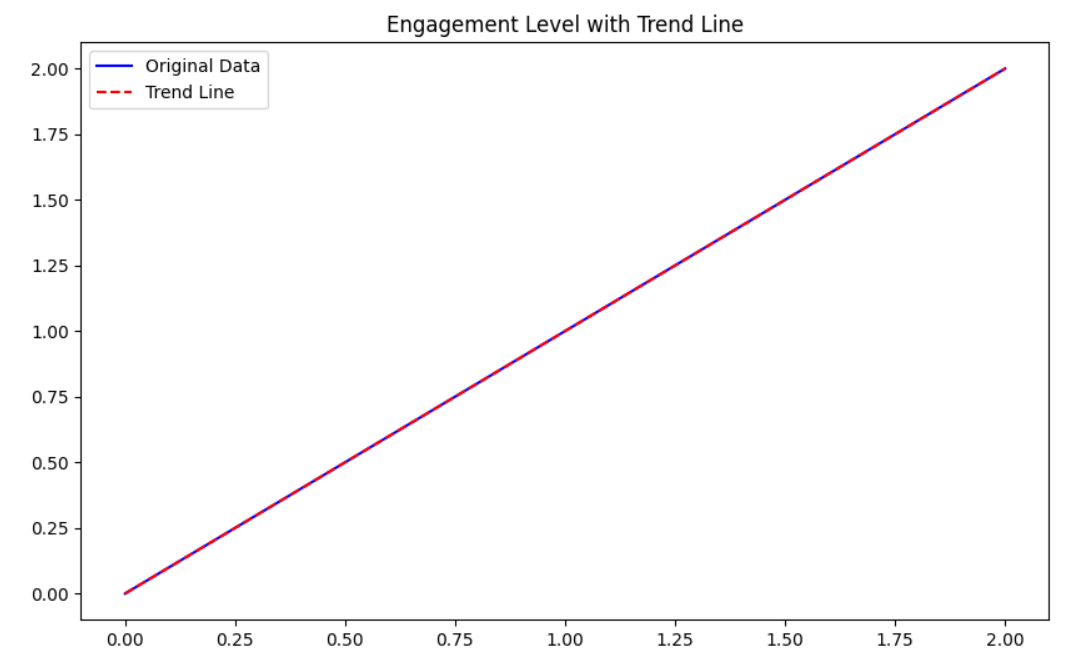


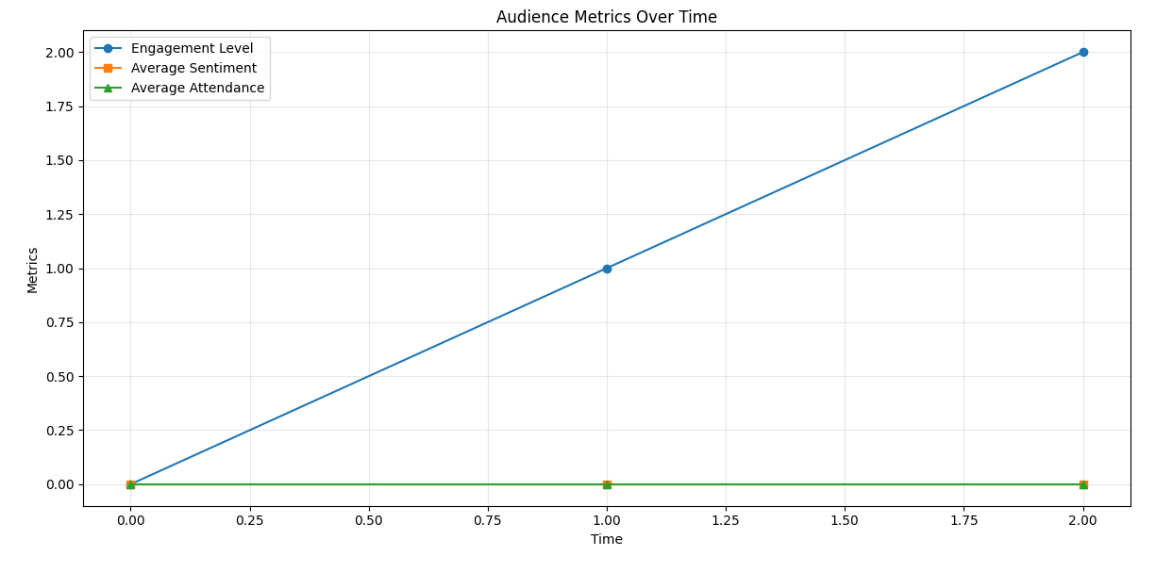






-----





**Progress :**

# Challenges and Solutions :

**● Accomplishments**:  
Several key tasks were completed successfully:

* Trends in engagement, sentiment, and attendance were visualized.
* Rolling averages were applied to highlight long-term trends.
* A linear regression model was developed to track engagement trends.
* The correlation between key metrics was analyzed, revealing important relationships.

**● Metrics**:

* **Engagement Level (Rolling Mean)**: A smoothing effect was applied, reducing fluctuations and highlighting key trends.
* **Linear Regression Trend**: The linear regression model showed a steady increase in engagement level over time.
* **Correlation**: The correlation analysis showed strong positive relationships between **engagement level** and **attendance**.

### Challenges Faced

**● Challenges Faced**:

* **Limited Data Points**: The dataset had only 3 observations, which limited the ability to perform deeper time series analysis (e.g., seasonal decomposition).
* **Outliers**: A few outliers were detected in the data, which could distort analysis if not handled properly.

**● Solutions Implemented**:

* **Data Size Limitation**: Seasonal decomposition was skipped due to the limited number of data points. Instead, rolling averages and linear regression were used to analyze trends.
* **Outlier Handling**: Outliers were identified and visualized using boxplots, and a decision was made to investigate the impact of these outliers on the results rather than removing them.

**Next Steps :**

**● Upcoming Tasks**:

* Gather more data points to perform more robust seasonal decomposition and time series forecasting.
* Consider applying more sophisticated forecasting models (e.g., ARIMA) for longer time series analysis.
* Investigate how external factors (such as seasonality, marketing campaigns, etc.) influence engagement.

**● Goals**:

* Increase the dataset size to improve the accuracy and reliability of time series analysis.
* Explore deeper analytical techniques, including ARIMA or exponential smoothing, for forecasting future trends.
* Conduct a deeper analysis of correlations and their potential causal relationships.

# Conclusion :

### Summary: The task successfully explored audience sentiment tracking over time in the entertainment sector. Despite challenges related- to the small dataset size, trends were visualized, correlations were explored, and predictive models were developed. Further analysis can be done as more data is collected.

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