**Machine Learning for Viewer Predictions - Entertainment Sector**

E.Code [E25007]

# Overview :

Today’s task focused on building a predictive model for viewer count in the entertainment sector using a structured machine learning approach. This involved data cleaning, exploratory data analysis, and initial model preparation steps, including visualizations and feature selection.

# Objective:

Analyze and prepare data to predict viewer count in the entertainment sector accurately, using machine learning techniques.

# Assigned Task(s) :

* Conduct data preparation and visualization to identify patterns and relationships within the data.
* Resolve encountered errors related to data types and missing columns.
* Build initial visualizations to understand feature relationships and guide model preparation.

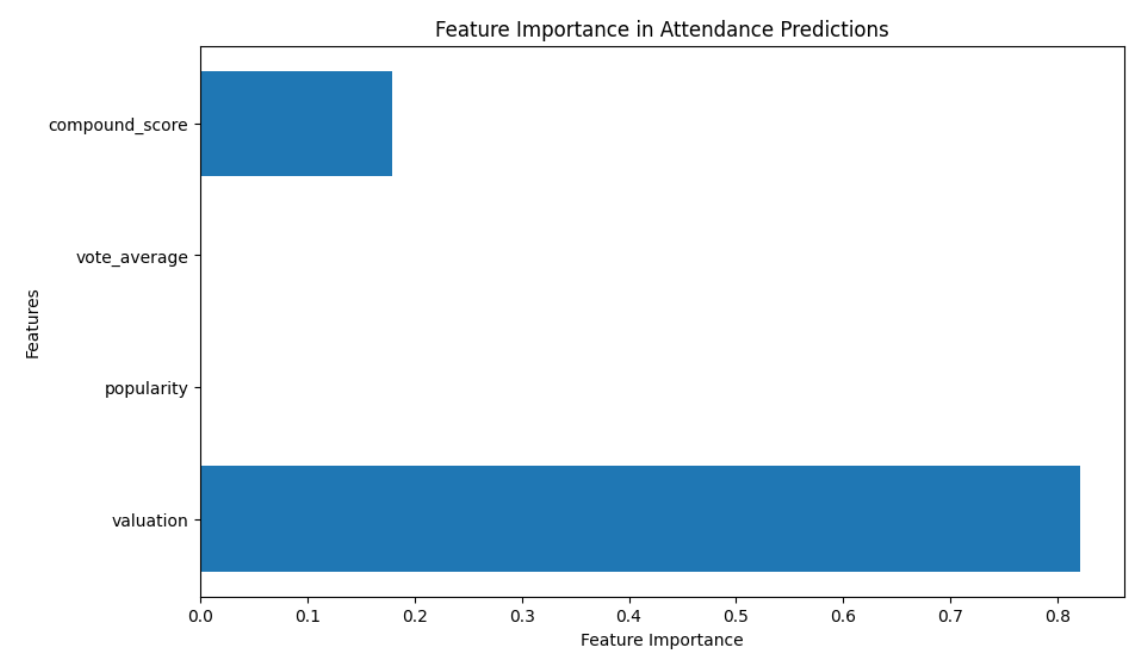
# Task Details :

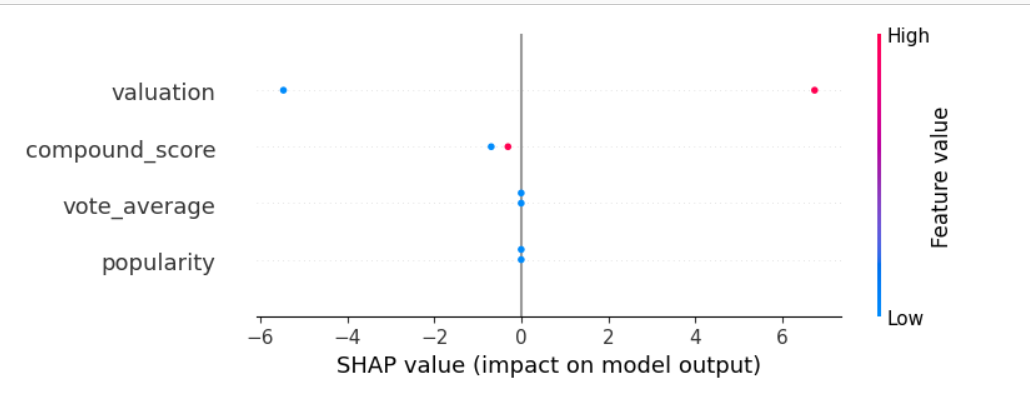
#### Task 1: Data Cleaning and Exploration

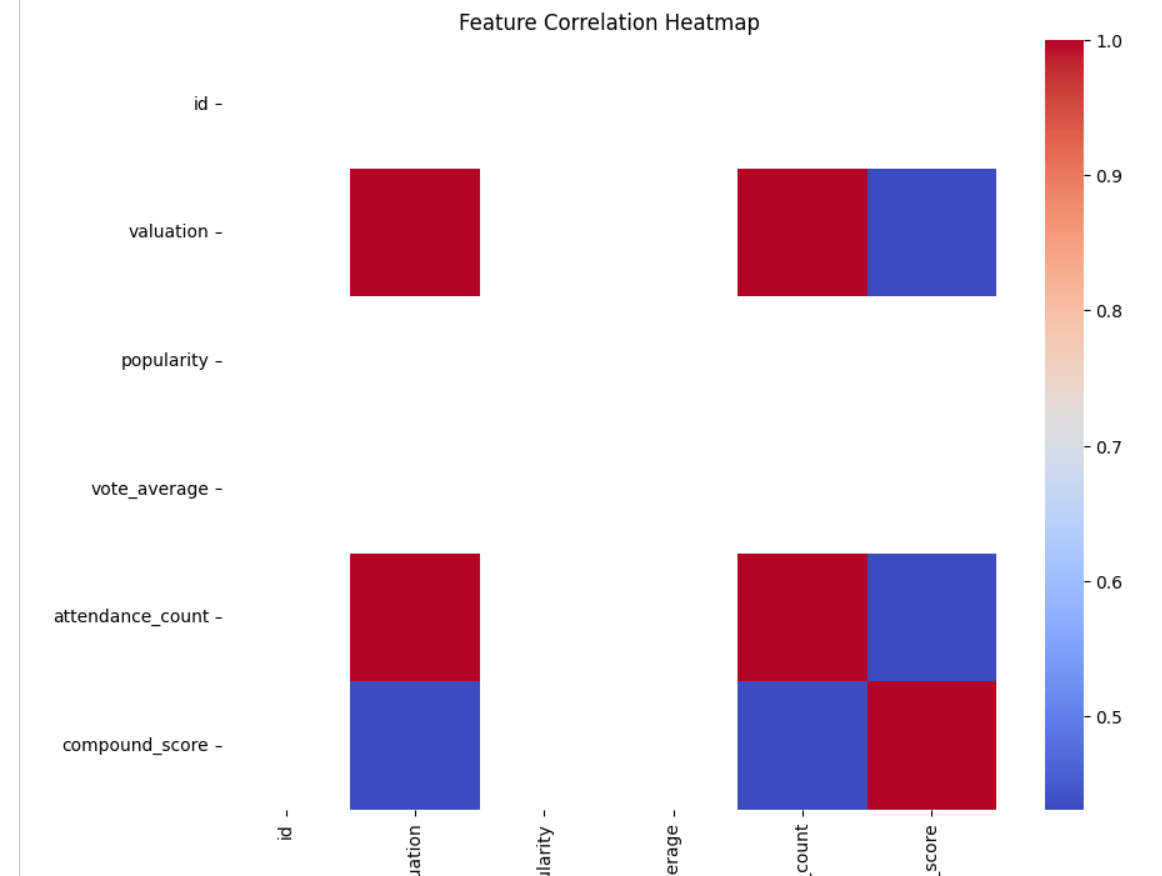
* **Status:** Completed
* **Details:** Addressed issues with missing columns and incompatible data types, ensuring all necessary columns for viewer prediction, such as attendance\_count, popularity, and sentiment, were properly formatted and accessible for further processing.

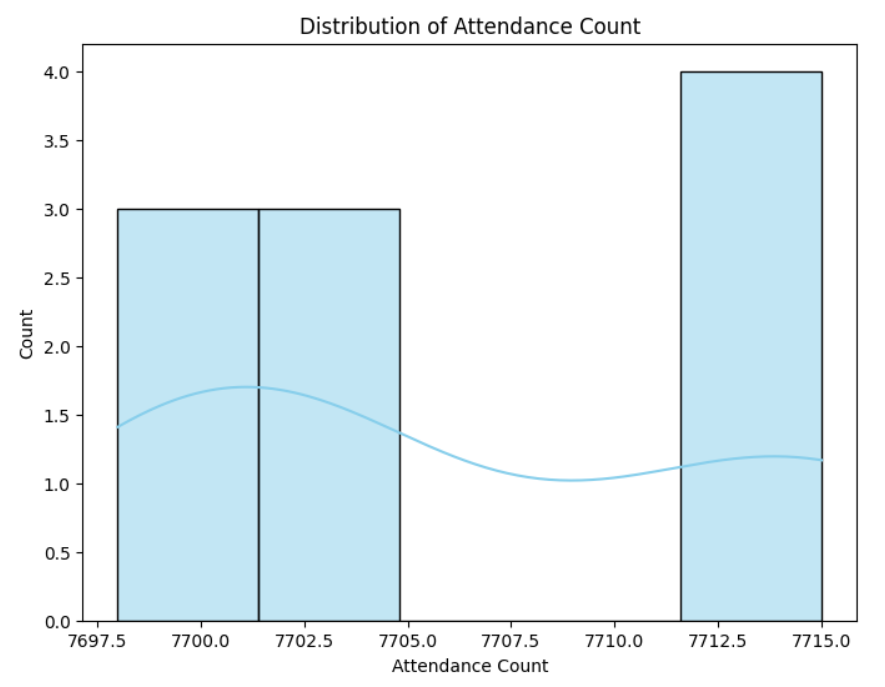
#### Task 2: Data Visualization

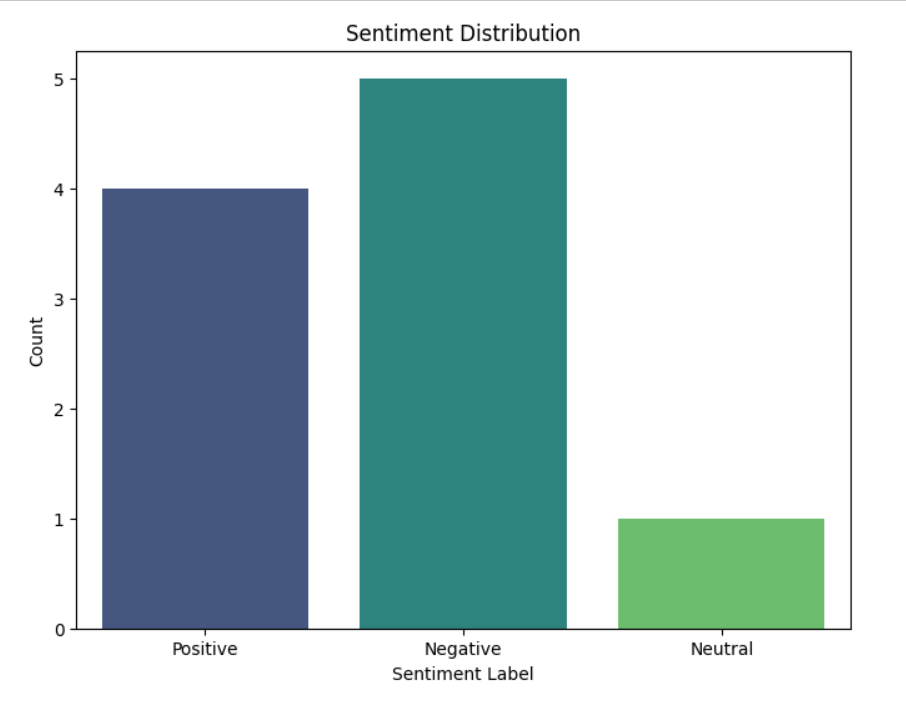
* **Status:** In Progress
* **Details:** Generated initial visualizations to explore correlations among features, specifically using correlation matrices, histograms, and scatter plots. This provided insight into viewer count distribution and its relationships with variables like popularity and sentiment.

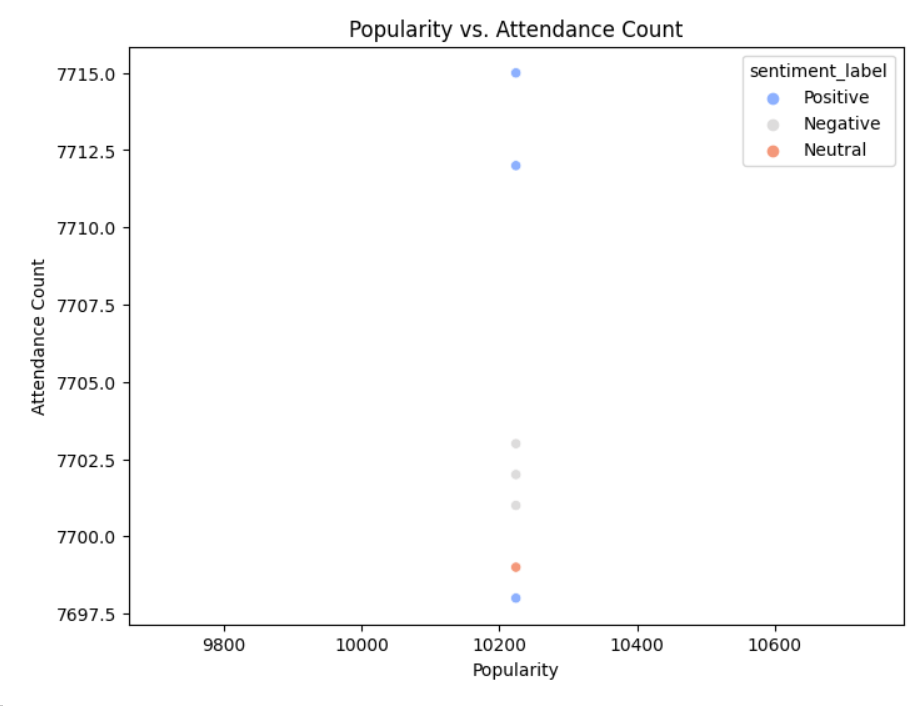














**Progress :**

* **Accomplishments:**
  + Successfully cleaned the dataset, resolving data type issues and ensuring all required columns were available.
  + Created visualizations, including correlation matrices and distribution plots, which provided valuable insights into feature relationships and viewer count trends.
* **Metrics:**
  + Completed data preprocessing and produced preliminary visualizations.
  + Correlation matrix insights indicated significant relationships, particularly between popularity, attendance\_count, and viewer count.

# Challenges and Solutions :

* **Challenges Faced:**
  + Encountered errors such as missing columns and data type mismatches, which initially hindered the data exploration process.
* **Solutions Implemented:**
  + Conducted targeted data cleaning by addressing column access issues and standardizing data types to allow for seamless analysis. Verified that all columns were usable for visualization and model input.

**Next Steps :**

* **Upcoming Tasks:**
  + Finalize feature selection and begin feature engineering to improve prediction accuracy.
  + Implement and evaluate a predictive model for viewer count using linear regression or decision trees.
* **Goals:**
  + Achieve a working model with acceptable accuracy metrics (e.g., R-squared, MAE) and document results and insights for report inclusion.

# Conclusion :

### Summary:

# Today's tasks achieved critical data cleaning and initial exploration milestones, setting a foundation for predictive modeling. Preliminary insights from visualizations have informed the selection of potential predictor variables.

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