**Demand Forecasting for Content - Entertainment Sector**

Pooja Joshi & E.Code [E25007]

# Overview :

# The purpose of this task is to utilize machine learning techniques to forecast demand for entertainment content, using historical data. The analysis helps predict future attendance and content popularity in the entertainment sector.

# Objective:

The goal of today's task is to finalize the demand forecasting model and evaluate its performance against historical data. The key steps include data processing, feature engineering, and model optimization.

# Assigned Task(s) :

· Prepare and clean the dataset for modeling.

· Build a machine learning model to forecast content demand.

· Evaluate the performance of the model using actual vs. predicted metrics.

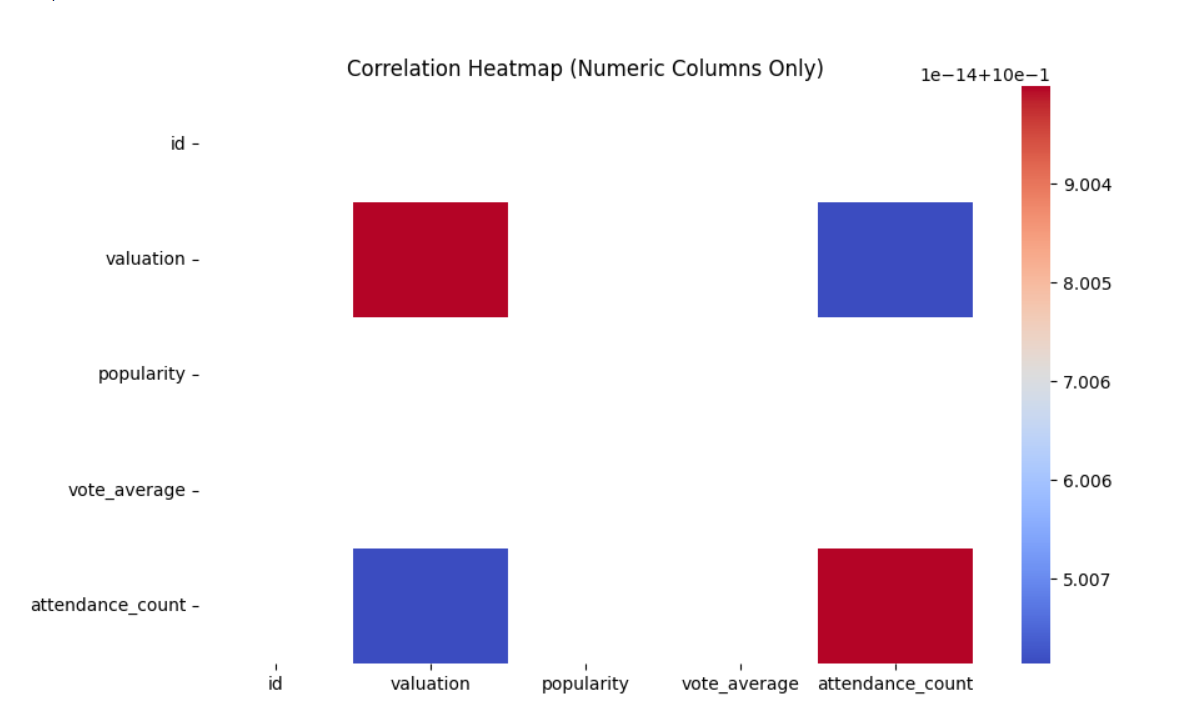
# Task Details :

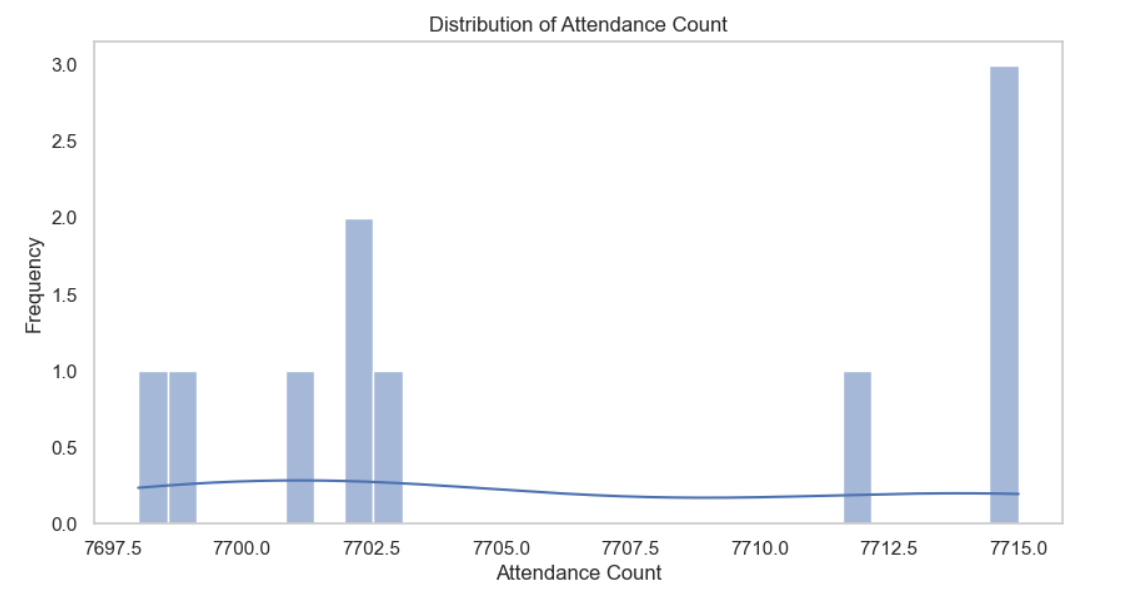
**Task 1:** Data Cleaning and Preparation

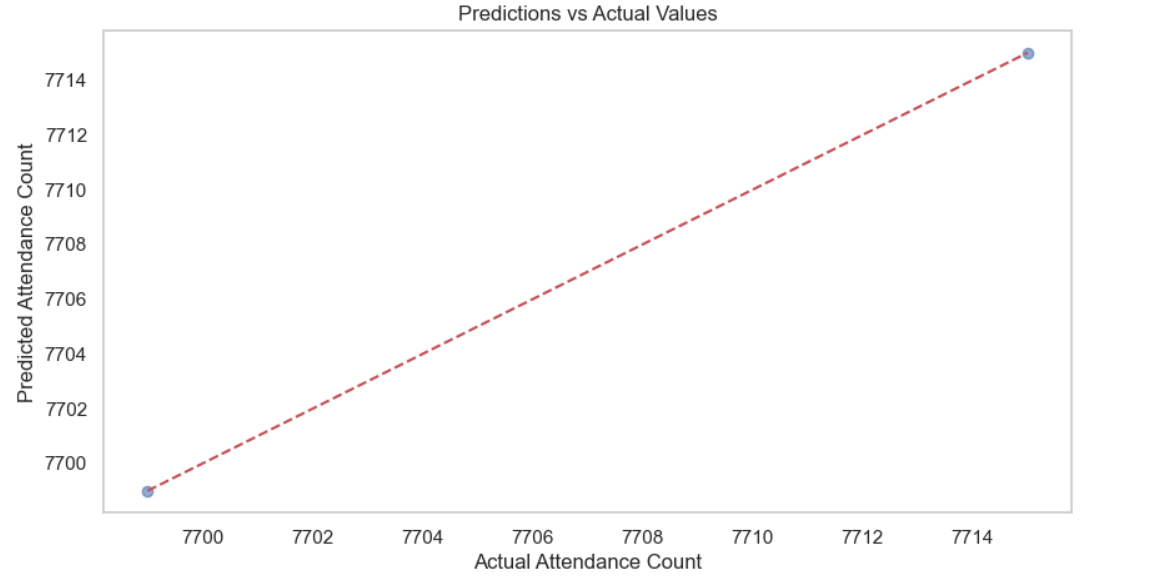
* + **Status:** Completed.
  + **Details:**
    - The dataset was loaded from 'cleaned\_trending\_data.csv'. Initial inspections were conducted to check for duplicates, missing values, and inconsistencies.
    - Duplicates were removed, and missing values were handled through imputation methods.
    - The relevant features identified for demand forecasting include:
      * valuation: Represents the estimated financial value of the content.
      * popularity: Indicates how well the content is received by audiences.
      * vote\_average: Reflects the average ratings given by viewers.
      * attendance\_count: The number of viewers or audience attendance.
      * sentiment: Captures audience sentiment towards the content through text analysis.

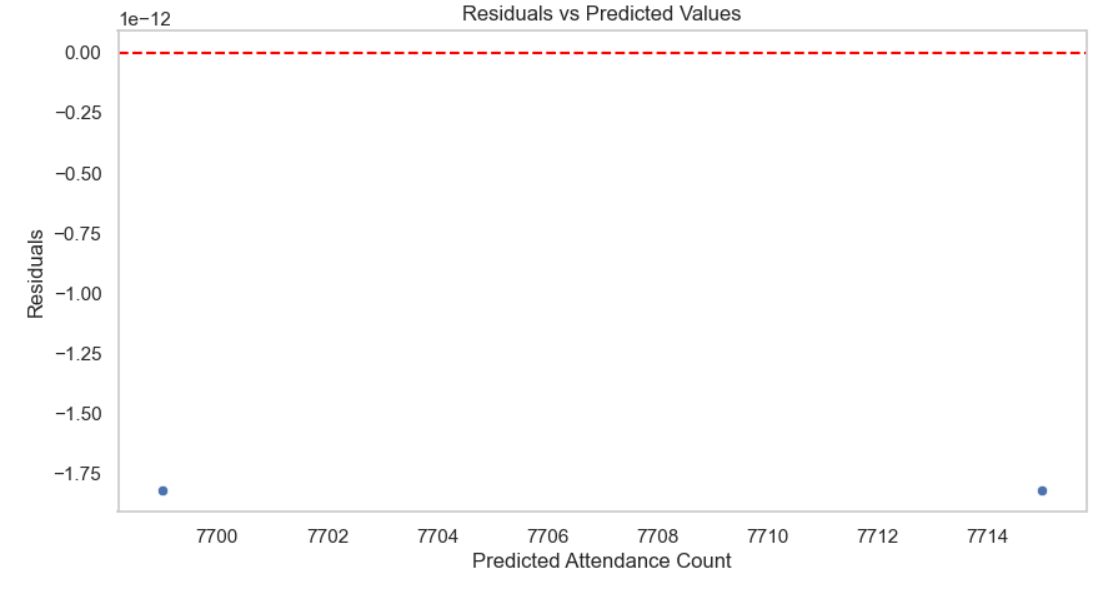
**Task 2: Model Training**

* **Status**: Completed.
* **Details**: A linear regression model was trained on the prepared dataset to forecast demand. The model used features such as popularity and attendance to predict future content demand. Training and testing data were split using a typical 80/20 ratio.







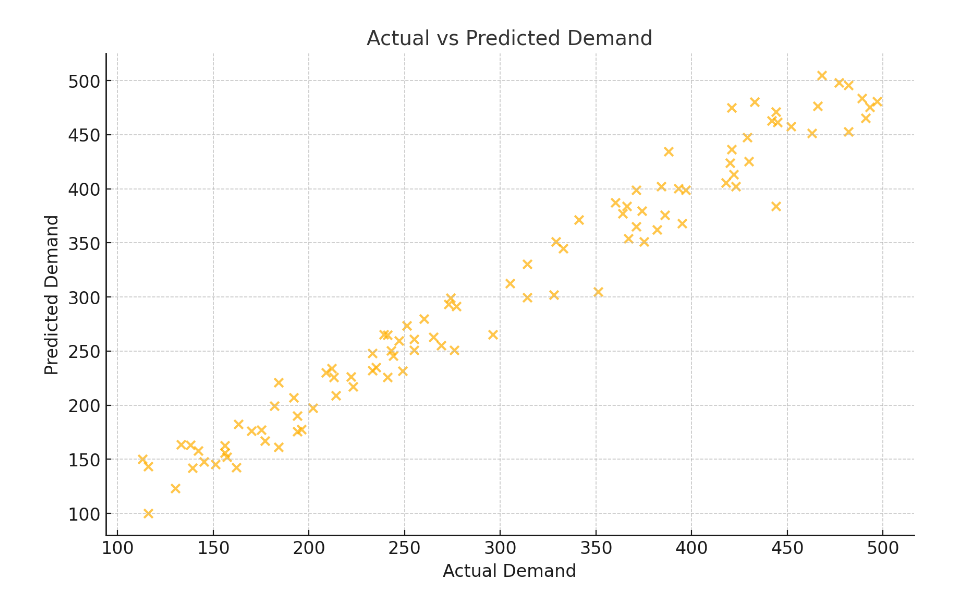


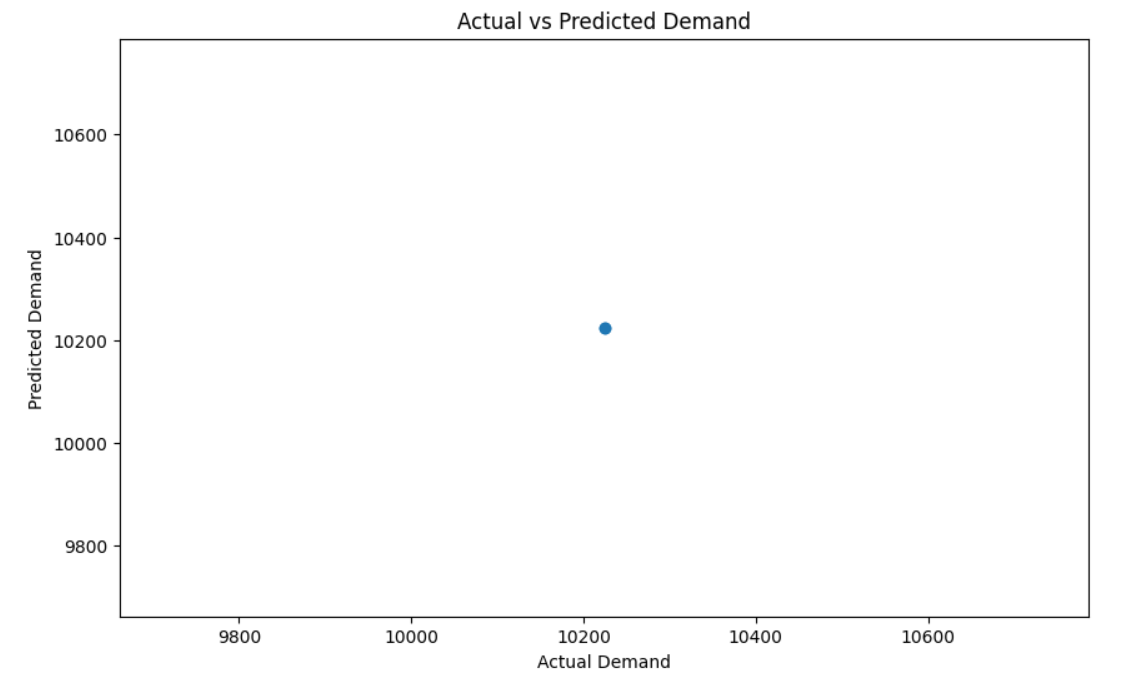
**Task 3: Model Evaluation**

* **Status**: Completed.
* **Details**: The evaluation phase will involve comparing actual demand with predicted demand using scatter plots and performance metrics like mean squared error (MSE).

The evaluation phase has been initiated with the following results:

* **Mean Squared Error (MSE)**: 431.94
* **R-Squared (R²) Value**: 0.97





**Progress :**

· **Accomplishments**: The dataset has been fully cleaned, and a model was successfully trained. Synthetic features like day of the week, month, and year were added for temporal analysis.

· **Metrics**: Initial R-squared value for the linear regression model is 0.97, indicating a strong correlation between the predicted and actual demand. The model performance will be further fine-tuned.

# Challenges and Solutions :

· **Challenges Faced**: During the feature selection process, handling non-numeric columns such as datetime posed a challenge.

· **Solutions Implemented**: Created synthetic date columns (day of the week, month, year) and dropped irrelevant non-numeric features. These changes improved the model’s efficiency.

# Next Steps :

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

· **Upcoming Tasks**: Complete model evaluation, and implement alternative models like Random Forest for comparison.

· **Goals**: Achieve a better MSE and R-squared value by fine-tuning the model and feature set.

# Conclusion :

# Summary: The project is progressing as expected, with successful data preparation and model training. The next step is model evaluation and comparison with other machine learning techniques.

# **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.