Movie Success Prediction and Sentiment Study

Abstract
This project analyzes movie data to understand the factors influencing movie profitability. Exploratory Data Analysis (EDA) and a Random Forest regression model were used to predict movie success. The analysis shows that worldwide gross and audience ratings significantly impact profitability.

Introduction
The film industry produces many movies every year, but only a few become profitable. This project aims to predict movie success using machine learning and analyze trends in ratings and revenue.

Dataset Description
The dataset contains movie details such as film name, genre, lead studio, audience score, Rotten Tomatoes score, worldwide gross, profitability, and release year.

Tools Used
Python, Pandas, Matplotlib, Scikit-learn, Google Colab

Methodology
Data cleaning was performed by converting monetary values to numeric format and handling missing values. EDA was used to identify relationships between ratings and profitability. A Random Forest model was trained using audience score, critic score, and worldwide gross as features.

Results
The analysis showed that movies with higher audience and critic scores tend to be more profitable. Worldwide gross was the most influential feature in predicting movie success.

Conclusion
The project successfully demonstrates how data analysis and machine learning can predict movie profitability. Future improvements can include sentiment analysis on movie reviews and additional features.