Movie Success Prediction & Sentiment Study

Introduction:

This project explores the relationship between viewer sentiment and a movie's commercial success. By analyzing movie reviews using sentiment analysis and combining that with basic movie metadata, we build a model to predict box office performance.

Abstract:

The aim of this project is to combine natural language processing and machine learning to understand how public sentiment influences or correlates with a movie's financial success. We use a dataset of 10,000 synthetic movie reviews with attributes such as rating, genre, budget, and box office earnings.

Tools Used:

- Python (Pandas, Scikit-learn, Matplotlib, Seaborn)
- VADER Sentiment Analyzer (NLTK)
- Jupyter Notebook
- Excel (for data storage and export)

Steps Involved in Building the Project:

1. Import and preprocess a dataset with movie reviews and metadata.
2. Apply VADER sentiment analysis to calculate sentiment scores and classify each review.
3. Visualize sentiment trends across different movie genres.
4. Build a linear regression model using rating, budget, and sentiment score to predict gross earnings.
5. Evaluate the model's performance using RMSE and R^2.

6. Save the analyzed dataset and visualizations.

Conclusion:

The project demonstrates that viewer sentiment, alongside traditional features like rating and budget, can help predict a movie's box office performance. It highlights the power of sentiment analysis in understanding public opinion and offers a simple yet insightful approach to revenue prediction in the film industry.