## Title : Movie Success Prediction and Sentiment Analysis using IMDb Data

### Introduction

This project predicts the commercial success of movies using IMDb data and analyzes viewer sentiment through natural language processing. It mimics how entertainment companies forecast content performance before release.

### Abstract

The dataset of top 1000 IMDb movies was analyzed. Using Python and VADER sentiment analysis, the overview text of each movie was scored. A regression model was trained to predict box office earnings based on features like genre, rating, votes and sentiment. The final model provided insights into what drives movie revenue and how sentiment plays a role.

### Tools used:

- Python (Pandas, Seaborn, Scikit-learn, NLTK/VADER)
- Jupyter Notebook

### Steps Involved:

1. Loaded and cleaned the IMDb dataset.
2. Engineered features like primary genre, gross earnings (in millions), and sentiment score.
3. Performed sentiment analysis using VADER on movie overviews.
4. Visualized sentiment trends across genres.
5. Built and evaluated a linear regression model to predict Gross Million.
6. Interpreted feature influence using coefficients.

### Conclusion

Sentiment scores from overviews contributed moderately to predicting gross revenue. Features like votes, runtime and genre also played key roles. This approach demonstrates how analytics can support decision-making in media production and marketing.