Movie Success Prediction and Sentiment Report

Introduction

The success of a movie is influenced by numerous factors such as budget, genre, IMDB rating, sentiment of user reviews, and marketing efforts. This project aims to identify the relationship between these variables and movie success, using data-driven analysis to predict profitability and popularity.

Abstract

This project combines exploratory data analysis, sentiment analysis, and visual analytics to derive insights from a curated dataset of movies. By utilizing Python, Excel, and Power BI, we assess key drivers of revenue and profit. The sentiment polarity of audience reviews is also evaluated using VADER from NLTK to understand its impact on movie performance.

Tools Used

- **Python** (Pandas,Scikit-learn, Matplotlib)
- **Excel** (PivotTables, Bar Charts)
- **Power BI** (Interactive Dashboards, Slicers)
- **Jupyter Notebook** (for data preprocessing and modelling.

Steps Involved in Building the Project

a. Data Collection & Cleaning:

- Merged datasets including budget, revenue, reviews, and IMDB ratings.
- Handled missing values, removed duplicates, and normalized values.

b. Sentiment Analysis:

- Cleaned review text using re, tokenization, and stop-word removal.
- Used SentimentIntensityAnalyzer from NLTK's VADER to generate sentiment scores.
- Created a new feature: *compound_sentiment*.

c. Exploratory Data Analysis (EDA):

- Used Excel to create PivotTables and visualize the relationship between IMDB ratings, profit, and revenue by sentiment polarity (TRUE/FALSE).
- Identified trends such as higher profit and revenue in cases where sentiment polarity was negative (FALSE), which may indicate blockbuster movies receive mixed or polarized reviews.

d. Dashboard Creation:

- Built an interactive Power BI dashboard filtering by genre and rating.
- Key Metrics Visualized: Total Profit, Revenue, and Sentiment Score by Genre.

Conclusion

This project demonstrates how combining structured and unstructured data enhances predictive insights. Despite positive sentiment being an indicator of audience liking, movies with diverse or polarized opinions (FALSE polarity) have shown higher cumulative revenue, potentially due to broader attention. Interactive dashboards empower stakeholders to dive deep into genres, sentiment trends, and profitability drivers.