Movie Success Prediction Project – Summary

Project Title: Movie Success Prediction and Analysis

Objective:
The objective of this project is to analyze movie data and predict movie profitability using data analysis and machine learning techniques. The project focuses on understanding how audience scores, critic ratings, and worldwide gross contribute to movie success.

Dataset Used:
The project uses a movie dataset containing details such as film name, genre, lead studio, audience score, Rotten Tomatoes score, worldwide gross, profitability, and year of release.

Tools and Technologies:
Python, Pandas, Matplotlib, Scikit-learn, IDLE, Tableau, GitHub

Project Workflow:
1. The raw dataset (movies.csv) was stored locally and loaded using Python in IDLE.
2. A data cleaning script was written to convert monetary and percentage columns into numeric format and remove missing values. The cleaned data was saved as movies_cleaned.csv.
3. Exploratory Data Analysis (EDA) was performed using a separate Python script. Graphs such as genre distribution, profitability distribution, and rating vs profitability plots were generated and saved as image files.
4. A Random Forest regression model was trained to predict movie profitability using audience score, critic score, and worldwide gross as features. Model performance was evaluated using MAE, RMSE, and R^2 score, and results were saved to a text file.
5. A Tableau dashboard was created using the cleaned dataset to visually present key insights and trends.
6. The final project report and all scripts were organized into a structured folder and pushed to a GitHub repository for version control and submission.

Key Results:
The analysis showed that worldwide gross is the most influential factor in predicting movie profitability, followed by audience and critic ratings. Movies with higher audience reception generally achieved better profitability.

Conclusion:
This project demonstrates the complete data analytics workflow, including data cleaning, analysis, visualization, machine learning, and project deployment. It highlights how data-driven approaches can help predict movie success and support decision-making in the entertainment industry.