**Movie Success Prediction & Sentiment Analysis**

### ****1. Project Overview****

This project, developed by vaibhaviBachu, focuses on predicting the commercial success of movies using structured data and assessing public sentiment through user reviews. The goal is to identify key drivers of success in the film industry using data science and natural language processing.

**GitHub Repository:** [ElevateLabs\_Project](https://github.com/vaibhaviBachu/ElevateLabs_Project" \t "_new)

**Objectives:**

* Analyze metadata from TMDB to discover what features (budget, popularity, genres, etc.) influence revenue.
* Apply sentiment analysis on movie reviews to capture audience emotions and preferences.
* Build a predictive model to estimate movie revenue.

### ****2. Tools & Technologies Used****

* **Python** – Primary language used for analysis.
* **Pandas, NumPy** – Data cleaning, manipulation, and exploration.
* **Matplotlib, Seaborn** – For visualizations and plots.
* **Scikit-learn** – Machine learning model (Linear Regression).
* **NLTK + VADER** – Sentiment analysis using rule-based scoring.
* **Jupyter Notebook** – Analysis workflow documentation.
* **Excel** – Preliminary data checks.

### ****3. Dataset Description****

* **Source:** TMDB 5000 Movies Dataset from Kaggle.  
  **Files Used:** tmdb\_5000\_movies.csv

**Key Features:**

* budget, revenue, popularity, vote\_average, runtime, genres
* JSON-like nested fields such as genres were converted into readable formats.

**Data Preprocessing Steps:**

* Dropped irrelevant or null-heavy columns.
* Cleaned and formatted genre fields.
* Handled missing values in revenue/budget.
* Created new metrics like ROI (Return on Investment).
* Performed correlation analysis to find influential features.

### ****4. Sentiment Analysis Approach****

**Tool Used:** VADER Sentiment Analyzer from NLTK  
**Process:**

* Simulated or sampled user reviews (textual data).
* Scored each review using compound sentiment score:
* 0.05 → Positive
* <-0.05 → Negative
* In between → Neutral
* Analyzed sentiment trends across genres (Action, Drama, Horror, etc.)

### ****5. Key Insights and Takeaways****

* Budget is a powerful indicator of success when paired with good popularity metrics.
* Genre plays a critical role not just in financial performance but also in viewer perception.
* ROI-based metrics help evaluate smaller-budget films that perform exceptionally.
* Viewer sentiment varies by genre and should be considered for strategic decisions.

### ****6. Suggestions for Future Improvement****

* Include a proper README with setup steps, data schema, and result interpretation.
* Add classification models to predict hit/flop labels instead of just revenue.
* Use actual review datasets (e.g., IMDB reviews or Twitter data).
* Include evaluation metrics like MAE and adjusted R² for a fuller picture.
* Build an interactive dashboard using Plotly, Dash, or Streamlit for better presentation.

### ****8. Conclusion****

This project demonstrates a practical and holistic approach to movie success prediction using data science. It integrates exploratory analysis, feature engineering, regression modeling, and sentiment analysis to understand both business metrics and audience emotions. Such a dual-lens framework is essential in industries where success is driven by both numbers and narratives.