IMOVIE VIEWER ANALYTICS & DASHBOARD REPORT

Big Data Analysis Project – December 2025

Contents

1. Introduction

The IMovie dataset contains monthly film-viewing records, including viewer ratings, categories, number of views, languages, and film names.

The objective of this project is to analyse user engagement on the platform and develop a Streamlit dashboard to provide insights for marketing strategy and decision-making.

The dashboard enables IMovie stakeholders to explore

- Category-level performance

- Language-based viewing trends

- Top-performing films

- Viewer preferences

- Simple predictive indicators (based on popularity score)

2. Dataset Description

The dataset used in this project consists of the following fields

Column Name	Description
Film_Name	Title of the movie
Category	Genre / category of the movie
Language	Primary language of the movie
Viewing_Month	Date the movie was viewed
Number_of_Views	Total number of views in the given month
Viewer_Rate	Overall viewer rating (1–5)

Additional fields were created during preprocessing

New Column	Description
Year	Extracted from Viewing_Month
Popularity_Score	Computed as Viewer_Rate × Number_of_Views

This score was used as an aggregated measure of engagement and satisfaction.

3. Data Preprocessing

3.1 Cleaning

- Converted Viewing_Month to datetime format.

- Removed missing or inconsistent values.

- Ensured numeric columns (Viewer_Rate, Number_of_Views) were properly converted.
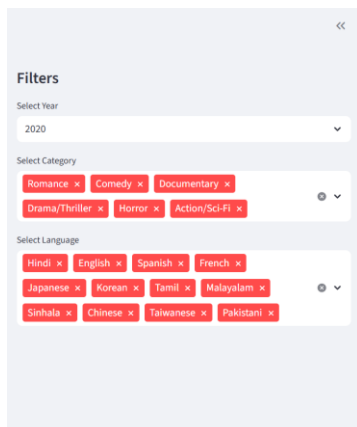
3.2 Feature Engineering

- Extracted Year for filtering.

- Created Popularity_Score to quantify content performance..

- Created a unique_ID to identify movies by film name, category, language, release date

3.3 Filtering Logic

Users can filter data by

- Year

- Category

- Language

These filters make the dashboard interactive and responsive.

4. Dashboard Development (Streamlit)

The dashboard was built using **Streamlit 1.10.0**, with the following structure

4.1 KPIs (Top Row)

The main performance indicators include

- Total Views

- Average Viewer Rating

- Most Popular Category

- Most Watched Language

These metrics provide a quick overview of platform health.

📈 Total Views	⭐ Average Viewer Rate	🏆 Most Popular Category	🌐 Top Language
42,289,031	4.22	Action/Sci-Fi	Hindi

KPI cards simplify complex metrics into one actionable number, which is critical for quick managerial decision-making.

4.2 Visualizations & Insights (Tabs)

📊 Category Insights   🌐 Language Insights   🔥 Top Films   📈 Prediction for 2026

Category Insights (Tab 1)

Visuals included

- Bar chart of views by category – best visualization for comparing discrete categories and displaying rankings.

- Line chart of average viewer rating by category – category wise rate fluctuations can be identify clearly

Key insights

Category popularity varies significantly.

Some categories have fewer views but high ratings, indicating strong niche potential.

Language Insights (Tab 2)

Visuals included

- Bar chart of views by language – simple representation of language distribution

Insights

Language strongly impacts movie engagement.

Popular languages tend to maintain high viewer volumes across months.

Top Films (Tab 3)

- Top 10 films ranked by **Popularity Score**

- Popularity Score helps identify films with both high views and strong ratings

- Table view of the top films gives and in detail idea about the top movies and bar graph gives an visual graphical idea to grasps for the strategic decisions.

Insights

Certain films consistently outperform others due to genre appeal or star power.

Viewer engagement is strongly correlated with film category and language choice.
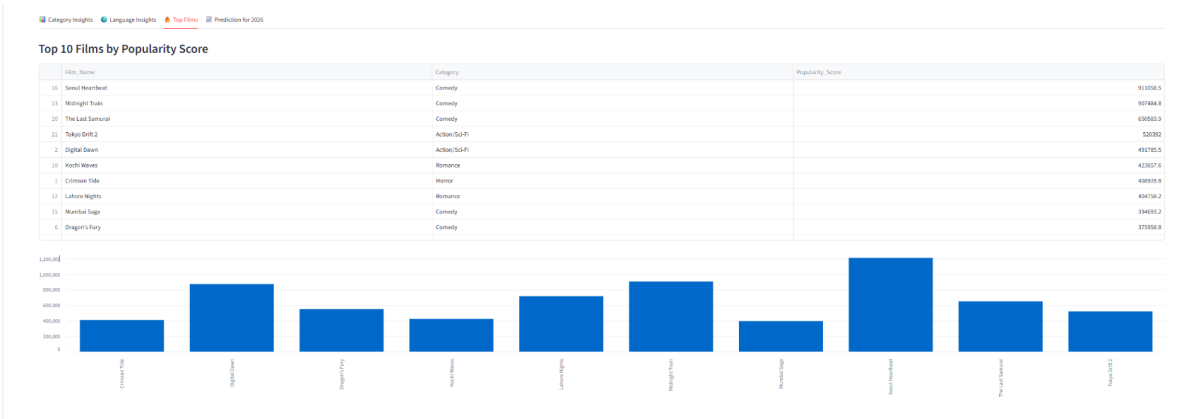
Prediction for 2026 (Tab4)

Since the project did not require machine-learning forecasting, prediction was performed using

Average Popularity Score per Category

Predicted_Popularity = Mean(Popularity_Score) for each category

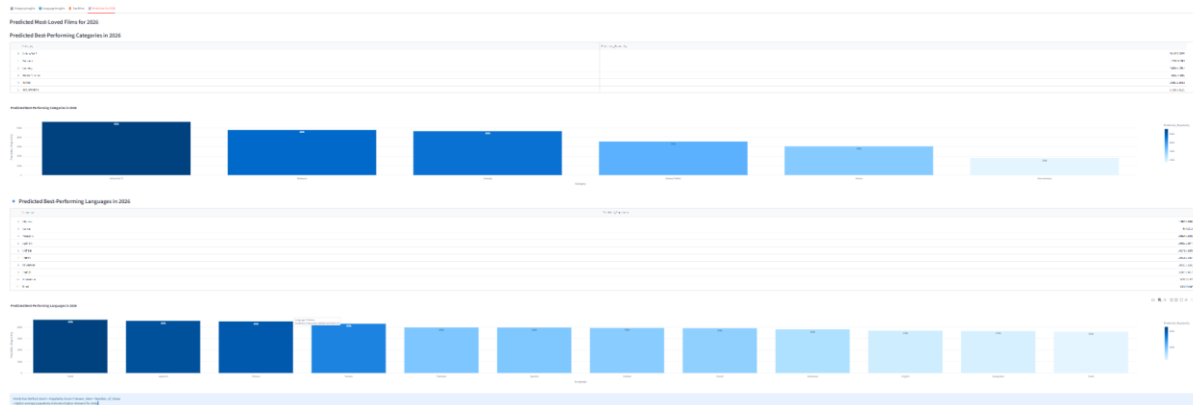**Average Popularity Score per Language**

Predicted_Popularity = Mean(Popularity_Score) for each language

Interpretation

Higher average popularity score indicates expected demand for 2026.

Limitations

- Not a time-series model

- Does not account for monthly or seasonal trends

- Purely descriptive (not predictive)

5. Key Findings

Category-Level Insights

- Certain genres (e.g., Action, Drama) dominate overall views.

- Some categories have high ratings but low views - marketing opportunity for those categories.

Language Trends

- Languages like Sinhala, Tamil, and English show consistently high engagement.

- Viewer preferences are often tied to language familiarity and cultural resonance.

Viewership Patterns

- Films with high viewer ratings typically correlate with high popularity scores.

- Categories with consistent performance across months contribute more to yearly total views.

6. Contribution to the project

We started the project as a Power BI project but as we continue we found it quiet difficult with its functionalities.

Streamlit provide flexible adjustable dashboard preferences.

W M V V Alagoda   Fc212029   fc110602 – dash board design

Sanjaya Weerakoon   fC110550   Fc211023 - dashboard design

N N H Kodithuwakku   fc111011   fc212047 – data preprocessing/ report design

Ravindu Vidurusinghe fC110551   fc211024 – dash board architecture design

Kavindu Nimesh   fc110556   fc211029 – feature engineering/ report design

7. Limitations of the Project

- Predictions are simple averages, not actual ML forecasts.

- Dataset size may affect generalizability.

- Missing film metadata (duration, actors, region) limits deeper analysis.

- Streamlit version constraints (1.10.0) limit advanced components like tabs & themes.

8. Conclusion

This project demonstrates the power of interactive dashboards for big data analysis. The IMovie dashboard provides a clear visualization of

- Film category performance

- Language preferences

- Top-performing films

- Basic predictive indicators

Marketing teams can use the dashboard to optimize

Content promotion
Language targeting
Category-based content investment
Audience segmentation strategies

With future enhancements like machine-learning forecasting and richer film metadata, the dashboard can evolve into a fully automated recommendation and prediction system.

9. Future Enhancements

Integrate Machine Learning

- RandomForest for rating prediction

- Clustering for viewer segmentation