2025

Data Visualization and Interpretation

Autumn 2025

Final Project

Name:Sanya Agarwal

Roll no.:R2142230839

Sap_ID:500120145

Program: B.Tech CSE

Semester: 05

Batch: 01 (Data Science)

Date of experiment: 03-12–25

School of Computer Science,

University of Petroleum and Energy Studies,

Dehradun

Final Project Report

Tool Used: Tableau Desktop Public

Dataset: IMDB_Movies Dataset

Introduction

The Movie Sales Dashboard is designed to provide a visual and data-driven understanding of how movies perform across different markets. By analyzing key metrics such as domestic and international gross, opening weekend revenue, per-theater averages, and return on investment, the dashboard helps studios, distributors, and theaters make informed business decisions. It offers a centralized platform to track movie performance, identify trends, and optimize strategies related to production, marketing, distribution, and scheduling. This ensures better planning, improved profitability, and enhanced audience engagement.

Objectives

1.Strategic Investment Guidance (Studio Focus): Identify high-ROI movies, genres, and release windows to inform future production greenlights and promotional expenditure.

2.Distribution Network Optimization (Distributor Focus): Benchmark performance across territories and individual theater chains to negotiate favorable terms and allocate prints/digital assets efficiently.

3.Exhibitor Scheduling Efficiency (Theater Focus): Provide granular per-theater average data to optimize local film scheduling, showing the right movie at the right time.

4.Financial Transparency: Deliver real-time, consolidated worldwide performance metrics (Total Gross) and profitability analysis (ROI).

Dataset Description

The dataset contains historical (19240-2020) movie records with attributes related to production, cast, budget, ratings, revenue, and audience engagement.

Column Name	Description
movie_imdb_link	Direct URL to the movie's IMDb page.
color	Indicates whether the movie is in color or black & white.
movie_title	Official title of the movie.
language	Primary language in which the movie was released.
title_year	Release year of the movie.
duration	Total runtime of the movie in minutes.
director_name	Name of the movie's director.
actor_1_name	Lead/Primary actor’s name.
actor_2_name	Secondary supporting actor’s name.
actor_3_name	Third listed actor’s name.
genres	Movie genres (e.g., Action, Drama), often multiple.
country	Country where the movie was produced.
budget	Estimated production budget in USD.
gross	Total worldwide revenue collected by the movie.
plot_keywords	Key themes or objects related to the movie plot.
content_rating	Movie censorship rating (e.g., PG-13, R).
aspect_ratio	Screen format ratio (e.g., 16:9, 2.35:1).
imdb_score	Movie rating on IMDb (out of 10).
num_critic_for_reviews	Number of critic reviews counted.
num_user_for_reviews	Number of user reviews on IMDb.
num_voted_users	Number of IMDb users who voted/rated the movie.
director_facebook_likes	Number of Facebook likes on the director’s page.
actor_1_facebook_likes	Facebook likes for Actor 1.
actor_2_facebook_likes	Facebook likes for Actor 2.
actor_3_facebook_likes	Facebook likes for Actor 3.
cast_total_facebook_likes	Total Facebook likes for the entire cast.
movie_facebook_likes	Facebook likes on the movie’s official page.
facenumber_in_poster	Number of faces present on the movie poster.

Data Processing Steps

1. Data Ingestion

This step involves collecting and importing the raw movie data from its source (CSV file). The dataset is loaded into the analysis environment where all further processing and exploration can take place.

2. Data Cleaning & Standardization

In this step, missing values, duplicate entries, and inconsistent formats are handled. Movie titles, genres, and language fields are standardized, numeric columns like budget and gross are cleaned, and year/duration values are validated to ensure accuracy and uniformity.

3. Metric Calculation Engine

New analytical metrics are derived from existing columns to support meaningful insights. For the movie dataset, this includes calculating total revenue, ROI, profit margin, rating performance, and creating additional attributes such as decade group, popularity score, or genre categories.

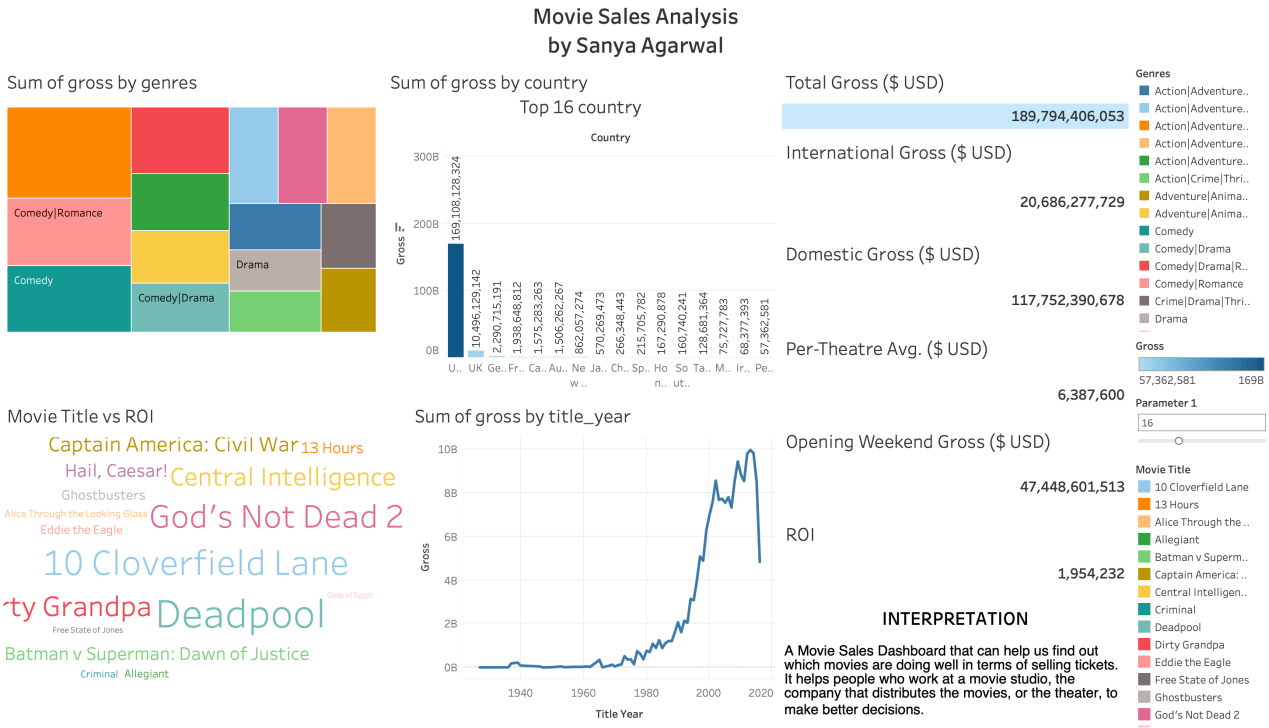
4. Real-Time Visualization

The processed dataset is then connected to a dashboard or visualization tool where charts, KPIs, and filters allow users to explore patterns such as top-grossing movies, highest-rated films, budget vs. revenue trends, and performance by year, genre, or language.

Dashboard Overview

- **KPI Cards:** Total Gross, Domestic Gross, International Gross, Opening Weekend Gross, Per-Theatre Average, ROI
- **Tree Map:** Sum of Gross by Genres
- **Bar Chart:** Sum of Gross by Top 16 Countries (Parameter Applied)
- **Line Chart:** Sum of Gross by Title Year (1927–2016)
- **Word Cloud:** Movie Title vs ROI

- **Slicers/Filters:** Genres, Gross Range, Movie Title, Parameter Selector
- **Interpretation Box:** Summary explanation describing business insights from the dashboard



Insights

Revenue Trends

- Total gross shows significant growth over the decades, with a sharp rise after the 1980s.
- Peak earnings occur between **2010–2016**, indicating strong box-office performance in recent years.

Genre Performance

- Genres such as **Action, Adventure, and Comedy** dominate gross revenue.
- Mixed-genre combinations (e.g., *Comedy|Romance*, *Action|Adventure*) perform exceptionally well in global earnings.

Country-wise Earnings

- The **United States** contributes the highest revenue globally.
- Other strong-performing regions include **UK, France, Germany, Japan, and Canada**, highlighting international market strength.

ROI Insights

- Movies like **Deadpool, Central Intelligence, and 10 Cloverfield Lane** deliver high ROI, reflecting strong profitability relative to budget.
- Some high-budget films show lower ROI, indicating that cost does not always correlate with returns.

Per-Theatre & Opening Weekend Performance

- Per-theatre average is strong for blockbuster films, showing high audience turnout during initial weeks.
- Opening weekend gross significantly impacts total earnings, emphasizing the importance of early marketing and promotions.

Year-wise Gross Trends

- Gross revenue starts low in early decades but rises consistently, demonstrating the industry's long-term expansion.
- A spike in the 2000s–2010s reflects global market growth and improved distribution strategies.

Overall Interpretation

- The dashboard highlights which genres, countries, and movies are financially successful.
- Insights help studios, distributors, and theaters optimize movie production, marketing spend, and release strategies.

Socio-Economic Impact

- High-grossing movies boost employment and stimulate economic activity across production, distribution, and theaters.

- Strong international and domestic revenues encourage investment in new films, talent, and technology.
- Successful movies promote cultural exchange and enhance global audience engagement.

Conclusion

1.Profitability Focus: The main goal is transforming raw revenue into crucial business metrics like Net Profit and ROI, moving the focus from simple sales figures to true financial gain.

2.Audience Insight: Features like Performance Decay and Per-Theater Average (PTA) are engineered to model audience behavior, telling stakeholders how fast interest drops and where demand is most concentrated.

3.Context and Comparison: By creating Budget Tiers and Contextual Flags (e.g., seasonality), the dashboard allows for fair peer-to-peer comparisons and helps isolate external factors impacting a movie's performance.