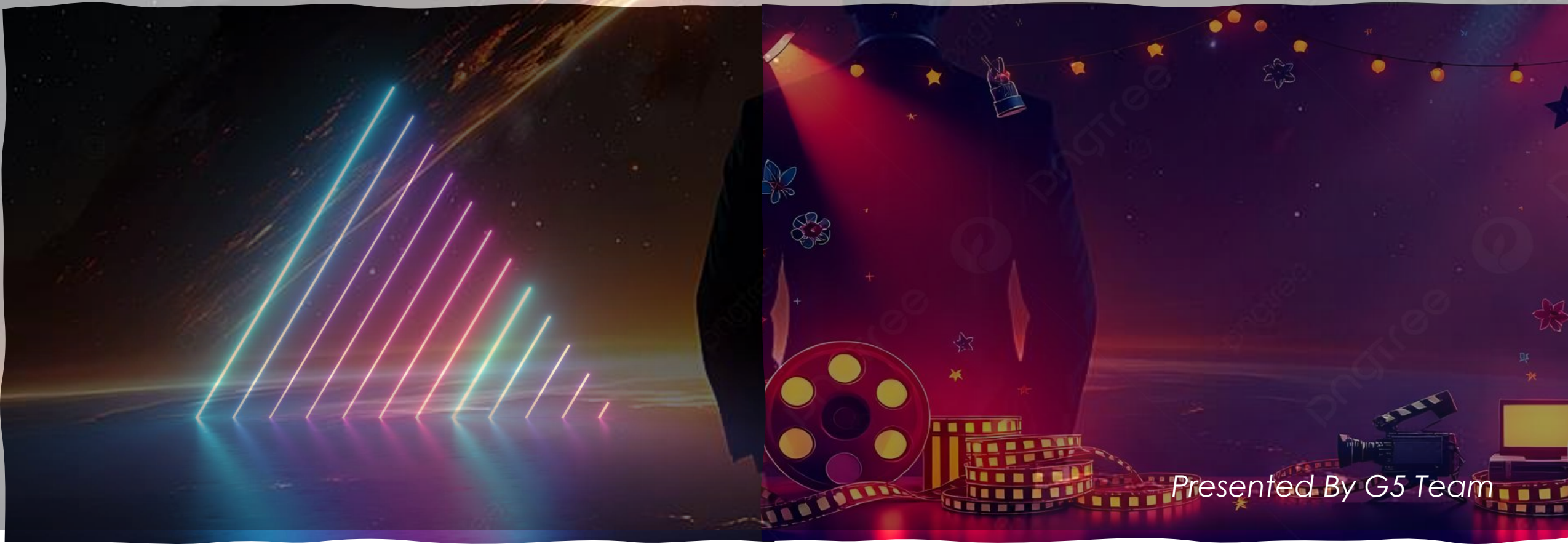


# G5- CINEMAXX

## The Big Screen Experience!!



# Introduction

**G5 Cinemax Company Limited** has observed the trend of major corporations investing in original video content and aims to establish its own competitive movie studio.

As the directors/heads lack existing industry expertise, they mandated **G5-Moringa Data Enthusiasts** to analyze current box office performance data to identify the most successful film genres and trends.

The G5-Moringa Data Enthusiasts will then synthesize this analysis into a strategic recommendations report to guide the studio head's decision-making on what types of films to produce which will be most profitable and have high ROI.



# Overview

---

**Strategy:** The G5 - Data Enthusiasts group stipulated the **G5 Cinemax Studio Strategy** which its main objective is to analyze the box office performance data and provide actionable insights to guide the studio's content creation strategy.



# Business Understanding

The core business problem is one of **de-risking market entry**. G5 Cinemax is a new player with no experience in film production.

The goal is not just to identify what is "popular," but to identify the **most financially sound and strategically viable** paths to success. Key business questions include:

- **Profitability:** Which genres yield the highest Return on Investment (ROI), not just the highest revenue?
- **Audience Appeal:** Which genres have the strongest domestic vs. international (worldwide) appeal?
- **Strategic Fit:** What is the optimal balance between high-budget "tentpole" films and lower-budget, high-ROI projects?
- **Critical Mass:** Is there a correlation between a film's popularity (votes, rating) and its financial success?

**Success Metric:** *Clear recommendations on genre, budget, and release strategy.*



# Data Understanding

---

The team analyzed all the data sets to understand how the data seated. The data sets reviewed were:

- [Box Office Mojo](#),
- [IMDB](#),
- [Rotten Tomatoes](#),
- [TheMovieDB](#)
- [And The Numbers](#).

The objective of data understanding was to keenly select the data set/s which would bring out the best data driven analytical approach by picking the most important fields to enable the analysis statistically.

These fields included;

**Identifiers, Film Metadata** e.g genres, **Audience & Critical Reception** e.g computed metrics and **Derived/Engineered Features** e.g log\_numvotes, combined\_score etc.

# Data Preparation

---

This phase involves cleaning and transforming the raw data into a reliable analysis-ready dataset.

## Key Preparation Steps:

- **Handling Missing Values:** Dropping rows with critical missing financial data
- **Data Type Conversion:** Ensuring production\_budget\_num, domestic\_gross\_num, and worldwide\_gross\_num are stored as numeric integers, not strings.
- **Genre Standardization:** The genres field is likely a pipe-separated string (e.g., "Action | Adventure | Sci-Fi"). This will be split to create a list of genres for each movie, enabling analysis by individual genre.
- **Outlier Treatment:** Using techniques like winsorization (as seen in runtime\_winsorized) to cap extreme values in financial data and runtimes that could skew averages.
- **Calculating Key Metrics:** Creating the essential roi\_num and profit\_num columns if not already present.

**Filtering:** Focusing on a modern era (e.g., films from 2010 onwards) to ensure recommendations reflect current market trends.

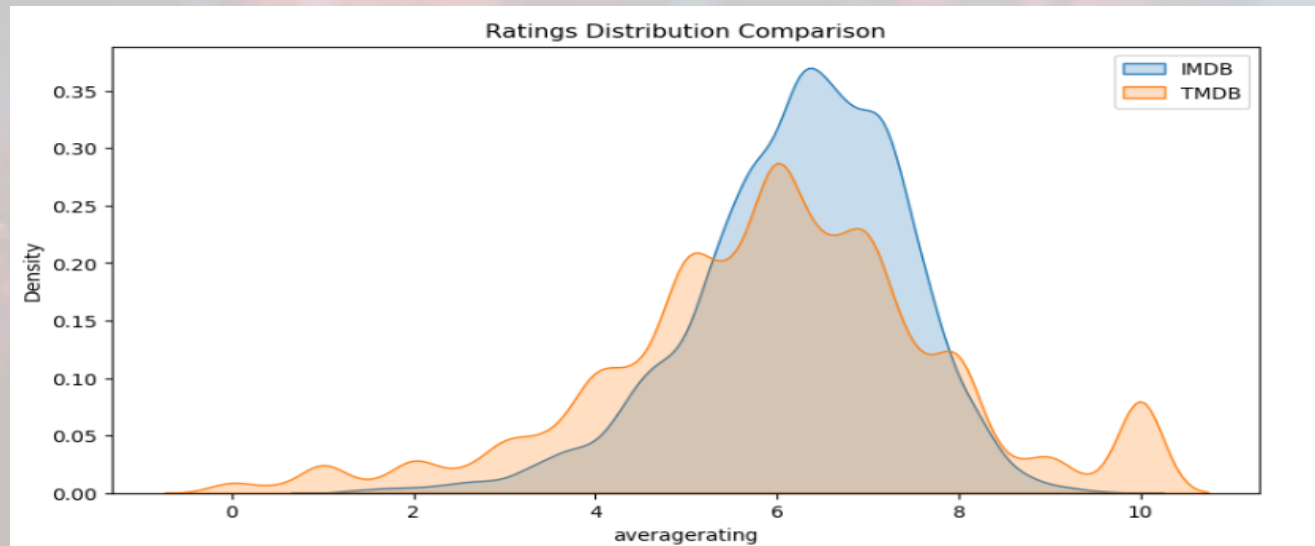
After the data preparation process, the team identified three data sets which were combined, and a clean data set was generated named: **finalcombined.csv which contained all the important fields to aid the statistical analysis.**

# Data Analysis

---

This phase involves analysis steps done on the selected dataset and test hypotheses to drive strategic decisions.

**Checking The correlation between ratings for the combined data sets.**



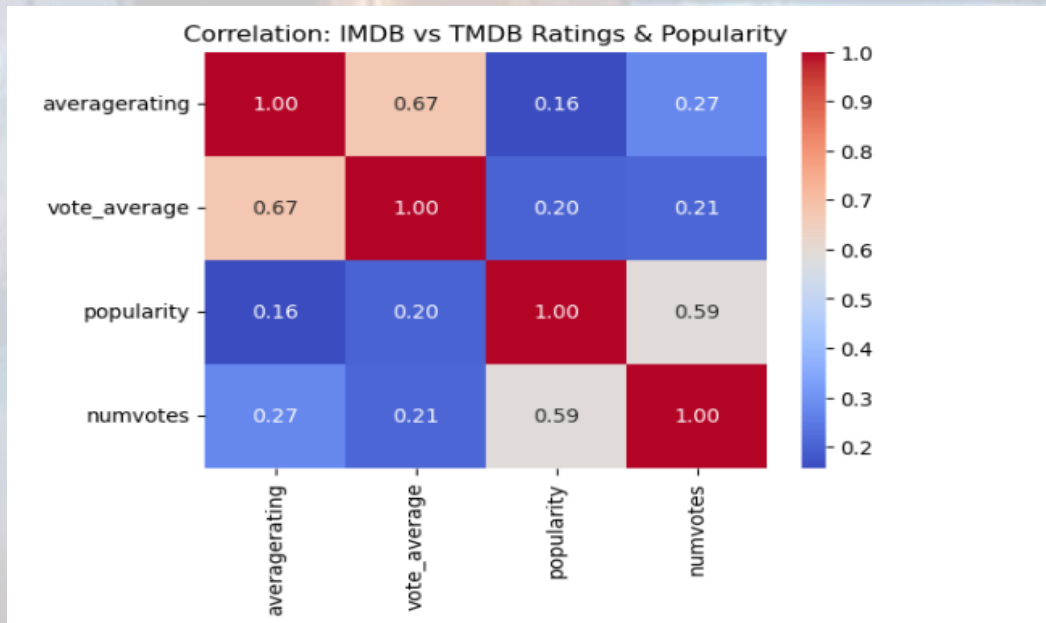
**The insights arrived on this :**

There is high correlation between the ratings i.e when IMDB ratings are high, the corresponding TMDB ratings are high as well



# Checking the Correlations

## Checking the Correlations on Ratings and Popularity - Visualized



### Key Insights from the Heatmap

1. There is a moderate strong correlation (0.67) between IMDB Average rating and TMDB Average rating.  
Correlation = 0.67 → Movies rated high on IMDb tend to also be rated high on TMDB
2. Correlation on Ratings against Popularity - Weak correlation (0.16 – 0.20) → A highly rated movie is not necessarily popular.
3. Popularity drives audience engagement - More popular movies tend to attract more votes. Popularity and audience size move together.



# Interpretations based on Correlation:

---

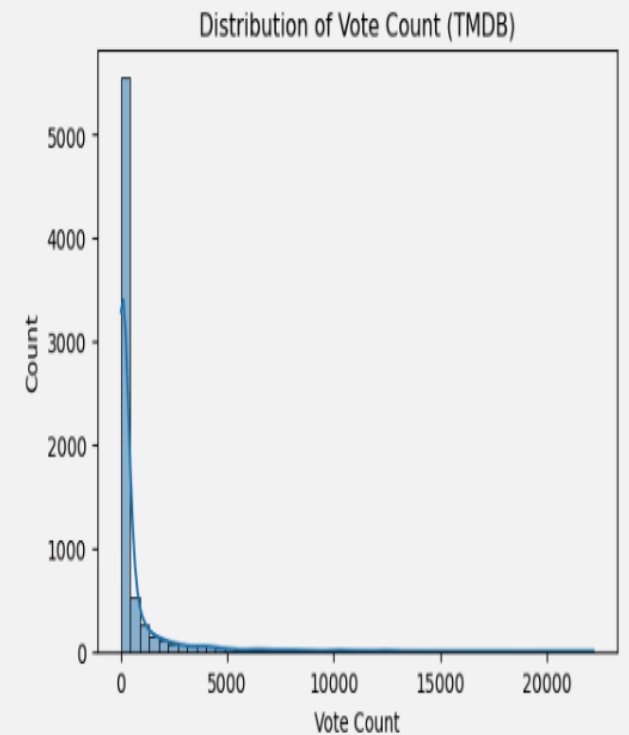
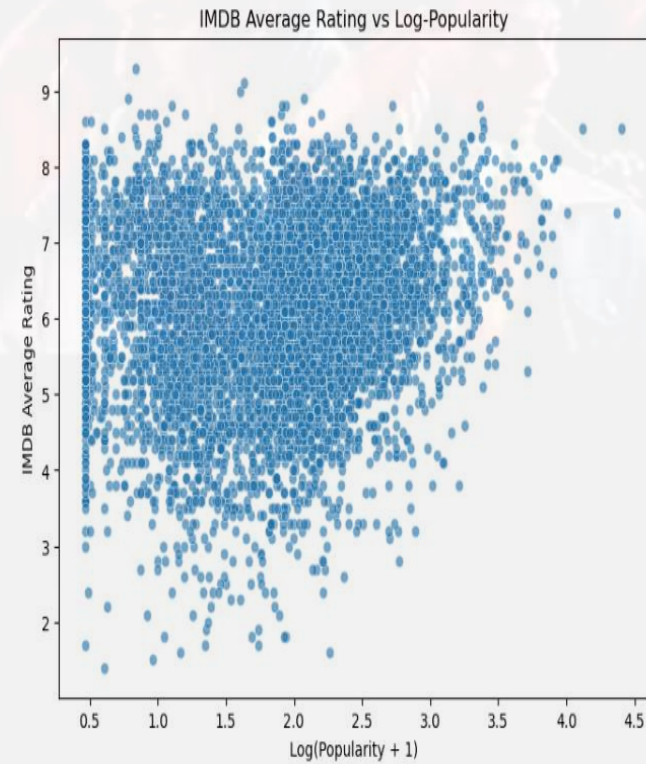
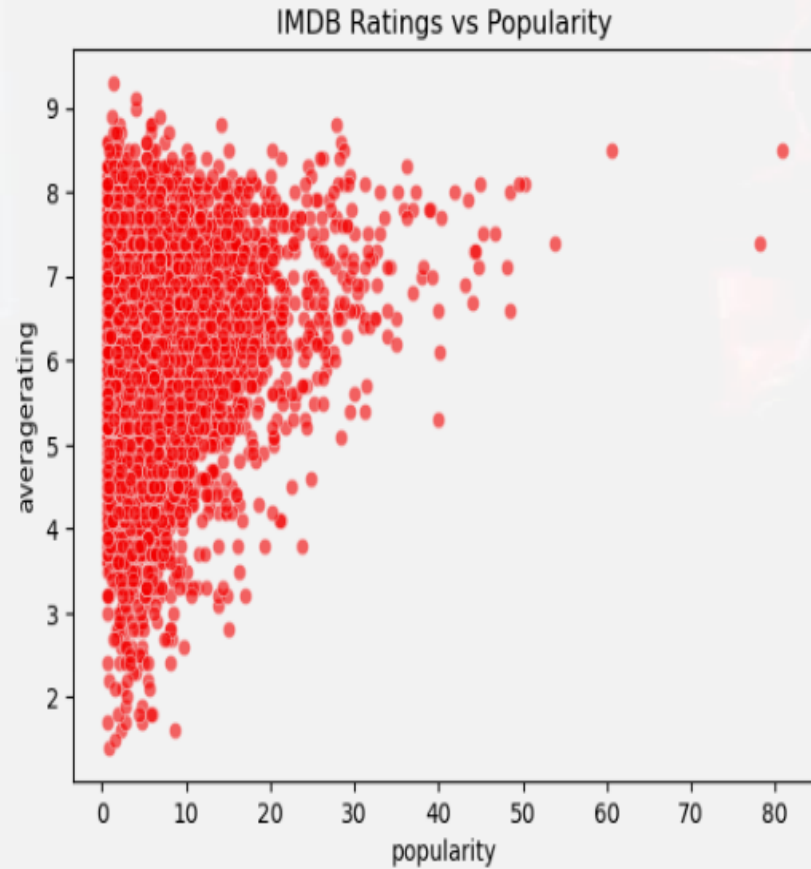
## Overall Conclusion on correlations:

- **Ratings (IMDb/TMDB) and popularity are not strongly linked** — quality (ratings) does not guarantee popularity.
- **Popularity and number of votes are connected**, indicating mass reach drives audience participation.
- **IMDb and TMDB ratings align well but** still leave room for differences across platforms.

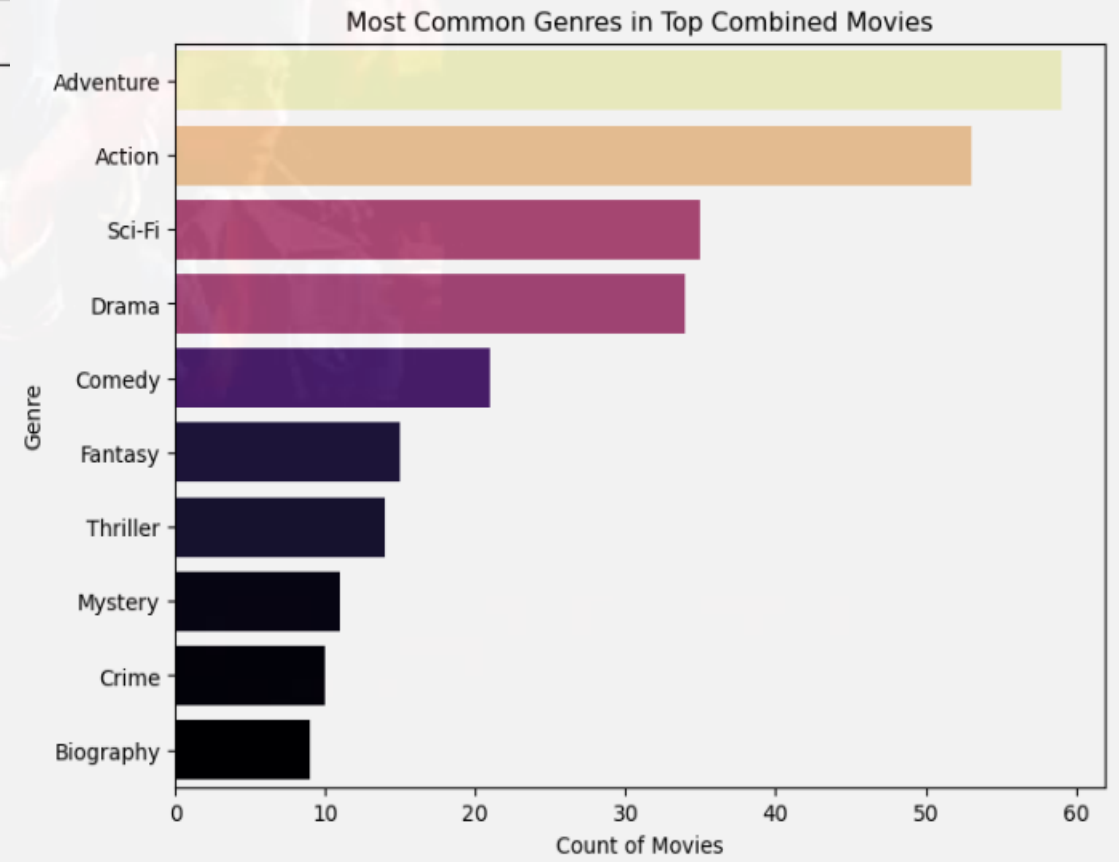
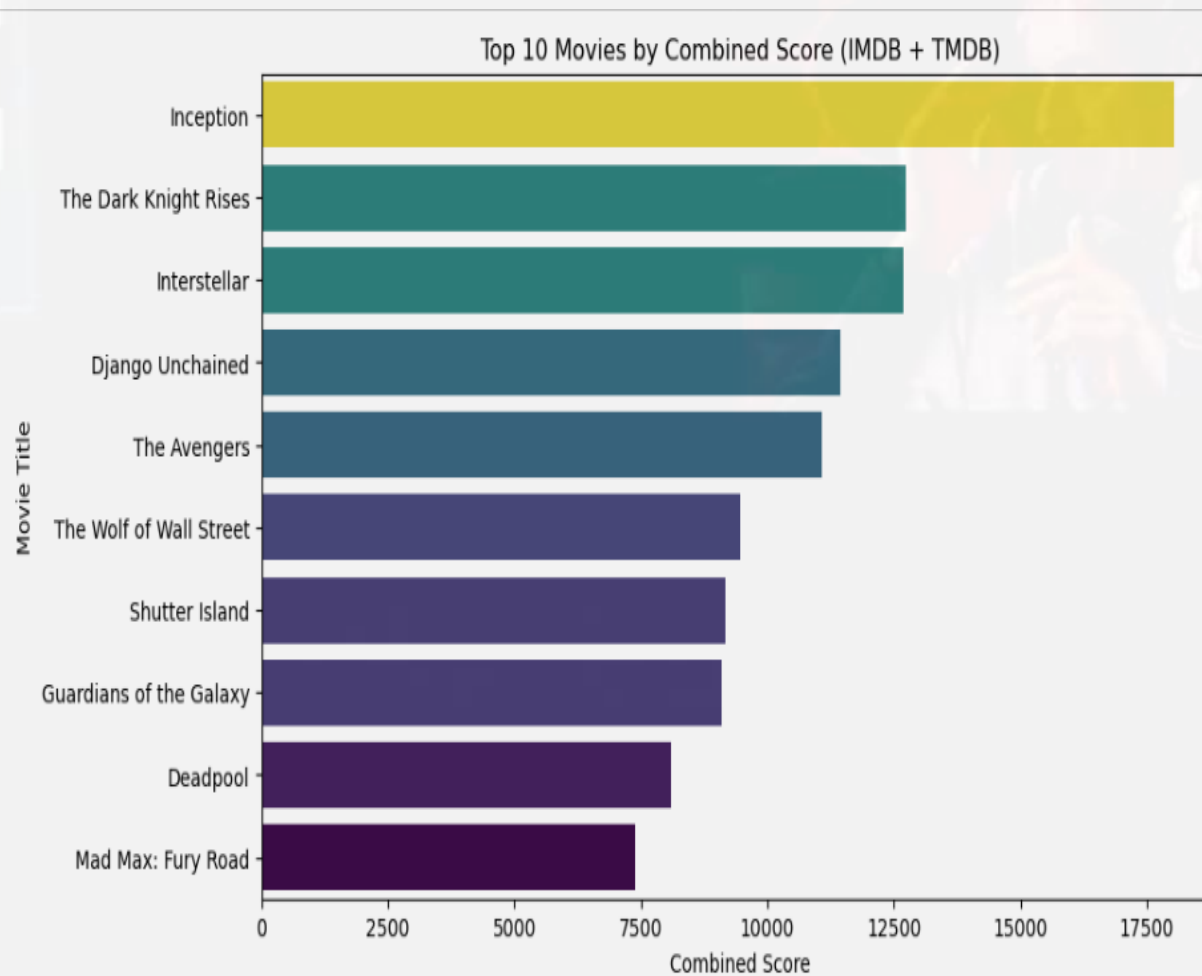
## Business Implication:

Focus on **both quality and mass appeal** — good ratings build credibility, but **popularity (marketing, distribution, trends)** drives audience size and revenue

# Additional Visualizations



# Visualizations





# Interpretation of the visualizations

---

## Interpretation: Top 10 Movies by Combined Score (IMDB + TMDB)

- **Inception** dominates the ranking with the **highest combined score ( $\approx 18,000$ )**, clearly outperforming the rest.
- **The Dark Knight Rises** and **Interstellar** follow closely, both above **12,000 combined score**, showing strong popularity and ratings across both platforms.
- **Django Unchained** and **The Avengers** occupy the middle tier with scores between **11,000 – 12,000**, reflecting broad audience appeal.
- Movies like **The Wolf of Wall Street**, **Shutter Island**, and **Guardians of the Galaxy** are in the **9,000 – 10,000 range**, indicating solid but slightly lower combined influence.
- **Deadpool** and **Mad Max: Fury Road** close the top 10 list with **7,000 – 8,000 combined score**, still strong but trailing the leaders.
- The **biggest gap** is between *Inception* and the second-highest (*The Dark Knight Rises*), suggesting *Inception* had both exceptional ratings and wide popularity.
- **Christopher Nolan's films (*Inception*, *Interstellar*, *The Dark Knight Rises*)** dominate the top, showing how his works consistently achieve high critical acclaim and popularity across platforms.

# Business Conclusions

---

From the critical data set analysis and visualizations, the business insights are concluded as below.

For G5-Cinemaxx movie studio to have the best profitable ventures in the new movie business, they need to consider;

- 1) **Adventure & Action blockbusters** (broad, global appeal) – Have better ratings and global appeal which will translate to higher returns.
- 2) **Sci-Fi and Drama** (satisfy demand for imaginative as well as deep storytelling).
- 3) Consider blending **Adventure + Sci-Fi** or **Action + Drama** for maximum impact.

# AOBS

---







**THANK YOU!!**