MOVIES DATABASE ANALYSIS – PROJECT REPORT

Group N0 : 03
Contributors:
Anjala 055005

Dimple 055009

Introduction

In this project, we are analyzing trends and insights in the film and television industry using MongoDB Atlas Charts and MongoDB Shell. By leveraging the dataset, we aim to identify patterns that influence movie success, audience engagement, and industry trends.

Objective

The primary objective of this project is to analyze patterns in the movie industry using the MFlix Movies dataset within MongoDB Atlas. The dataset allows us to explore insights such as movie release trends, IMDb ratings, directors' contributions, global distribution, and audience engagement.

Key areas of analysis include:

- Movie distribution across genres
- IMDb and Metacritic rating trends
- Viewer vs. critic perception analysis
- Awards and recognition-based performance evaluation
- People analytics in the movie industry
- Global and TV content contribution

A dashboard has been created to visualize these insights, and CRUD operations have been performed for deeper analysis.

Dataset Description

The MFlix Movies dataset includes a diverse range of movie-related data spanning multiple decades. Key attributes in the dataset include:

Variable	Description
Title	The name of the movie

Year	The release year of the movie
Genres	The categories the movie falls into (e.g., Action, Drama, Comedy)
IMDb Rating	The average IMDb user rating of the movie
IMDb Votes	The number of audience votes received on IMDb
Directors	The individuals responsible for directing the movie
Runtime	The duration of the movie in minutes
Languages	The primary language(s) spoken in the film
Countries	The country where the movie was produced
Other Metadata	Additional fields such as production companies, box office performance, and cast information

---

Major Problem Statements

The movie industry has evolved over time, making it essential to analyze trends in content, quality, and audience reception. This project addresses the following challenges:

1. **Genre and Content Distribution Analysis**
    - Identifying dominant genres and their frequency.
    - Examining trends in runtime and ratings for different genres.
2. **Quality and Performance Evaluation**
    - Analyzing IMDb and Metacritic scores across genres and time periods.
    - Identifying high-performing movies.
3. **Audience vs. Critic Perception Analysis**
    - Comparing IMDb ratings with Metacritic and Rotten Tomatoes scores.
    - Identifying gaps in critic vs. audience reception.
4. **Awards and Recognition-Based Evaluation**
    - Examining the influence of awards on movie success.
    - Identifying top-rated directors and award-winning movies.
5. **People Analytics in Movies**

- Analyzing the influence of directors and actors on movie success.
- Identifying top actors and directors based on average IMDb ratings.

6. **Global Film Industry Contributions**
    - Understanding global trends in movie production.
    - Identifying top-producing countries and production houses.

Analysis & Observations

1. **Movies per Genre**
    - Drama and Comedy are the most frequent genres in the dataset, indicating strong audience preference and profitability.
2. **Number of Movies with 70+ Metacritic Score Historically**
    - The number of critically acclaimed movies fluctuates over time, with peaks in certain years.
3. **Genre Distribution for 95+ Metacritic Score**
    - Drama, Biography, and Crime dominate the high-rated movie category.
4. **Mean Metacritic Score Ratings**
    - The average Metacritic score fluctuates between 55-75, with notable peaks and dips over different years.
5. **IMDb vs. Rotten Tomatoes Ratings**
    - Critics tend to rate movies higher than audiences, highlighting a preference gap.
6. **Most Critically Panned Movies**
    - Movies like "Disaster Movie" and "Superbabies: Baby Geniuses 2" received poor ratings from both critics and audiences.
7. **Top 10 Movies with Most Award Wins**
    - "12 Years a Slave" and "Gravity" are among the most awarded films.
8. **Correlation Between Awards and IMDb Ratings**
    - A moderate correlation exists between the number of awards won and IMDb ratings.
9. **Top 10 TV Shows by IMDb Ratings**

    - "Band of Brothers," "Planet Earth," and "The Civil War" are among the highest-rated TV shows.

10. **Top 20 Actors with Most Movies**

- Actors like Robert De Niro, Samuel L. Jackson, and Tom Hanks have extensive filmographies.

11. **Runtime Trends Over the Years**

- The average runtime of movies has increased over the years, particularly for Drama and Biography genres.

12. **Directors with Highest Average IMDb Ratings**

- Christopher Nolan, Martin Scorsese, and Steven Spielberg consistently produce high-rated movies.

13. **Most Frequently Remade Movies**

- "The Journey," "The Hole," and "Jane Eyre" have been remade multiple times.

14. **Total Number of Directors**

- Over 9,000 unique directors are present in the dataset, indicating a diverse film industry.

15. **Number of Movies by Country**

- The USA leads in movie production, followed by the UK and France.

16. **Production Houses by Average Freshness Level**

- Disney and Warner Bros. consistently produce highly-rated films.

MongoDB Dashboard (055005 & 055009)

Number of Movies

23,539

TOTAL NUMBER OF DIRECTORS

11,537

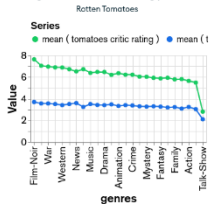
20 movies with most Remakes/Duplicate T...

title	count (title)
The Journey	6
The Terrorist	6
Jane Eyre	6
The Hole	6
Hamlet	6
First Love	5
Total	103

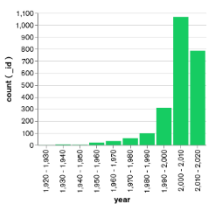
Top 10 TV Shows by IMDB Ratings

title
Band of Brothers
Planet Earth
A Brave Heart: The Lizzie Velasquez Story
The Civil War
Cosmos
The Real Miyagi
Total

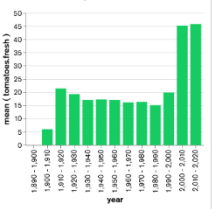
Average Viewer Vs Critic Rating per Genre



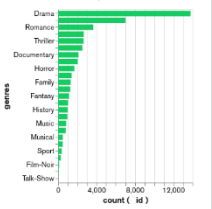
Number of Movies with 70+ Metacritic Sco...



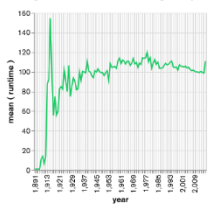
Freshness Rating Distribution Over Years



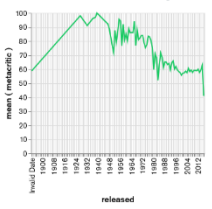
Movies per Genre



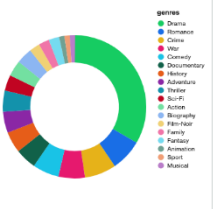
Average Runtime of Movies Through the y...



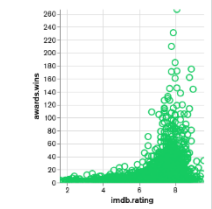
Mean Metacritic Score Ratings



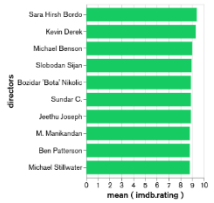
Genre Distribution for 95+ Metacritic Score



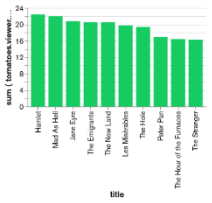
Correlation between Award Wins and IMD...



Top 10 Directors with the Highest Average ...



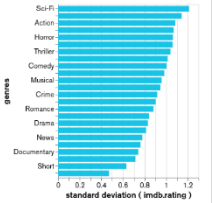
Top 10 Most Rotten Movies with Viewer Rat...



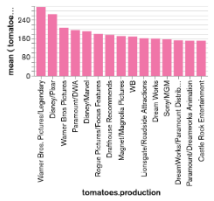
Top 10 Movies with Most Award Wins

title
12 Years a Slave
Gravity
Birdman: Or (The Unexpected Virtue of Ignorance)
Boyhood
The Lord of the Rings: The Return of the King
No Country for Old Men
Total

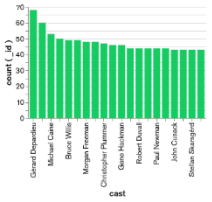
Standard Deviation of IMDB Ratings acros...



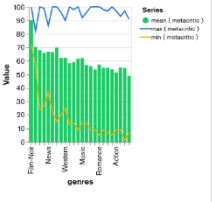
Top 15 Production Houses by Average Fres...



Top 20 Actors with Most Movies

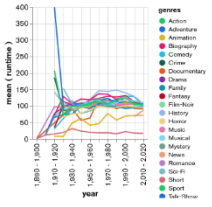


Metacritic Rating per Genre

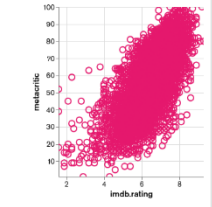


Top 15 Number of Movies by Country

Runtime Trend Per Genre Over Years

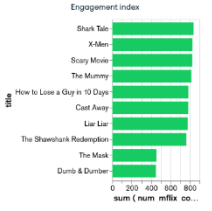


Correlation - IMDB Rating vs Metacritic Sc...



Top 10 movies with the number of Comme...

Engagement index



Querying

CRUD Operations – We did some querying using compass and Py Mongo to understand the workings of MongoDB syntax and commands.

1.CREATE Queries

```
> use sample_mflix
< switched to db sample_mflix
> db.movies.insertOne({
      title: "Inception",
      year: 2010,
      genres: ["Sci-Fi", "Thriller"],
      imdb: { rating: 8.8, votes: 2200000 },
      directors: ["Christopher Nolan"],
      runtime: 148,
      languages: ["English"],
      countries: ["USA", "UK"]
  })
< {
    acknowledged: true,
    insertedId: ObjectId('67d826390162f9c216898df5')
  }
Atlas atlas-rrgaki-shard-0 [primary] sample_mflix>
```

2.RETRIEVE Queries

Retrieve Top 5 Highest Rated Movies

3.UPDATE Queries

Update IMDb Rating of a Specific Movie

4.DELETE Queries

delete one document where the director field is null

Insights & Recommendations

Insights:

- **Drama and Comedy dominate**: These genres consistently perform well, reflecting audience preference.

- **Viewer-Critic Rating Gap**: Critics often rate movies higher than audiences, emphasizing a perception gap.
- **Longer Movies Are Becoming More Common**: Storytelling trends are shifting towards longer formats.
- **Strong Director Influence**: Renowned directors consistently produce high-rated films.
- **Production Houses Matter**: Studios like Disney and Warner Bros. maintain high production quality.
- **Global Film Contributions Are Expanding**: Countries like the UK, India, and France are producing more content.

Recommendations:

- **Encourage Genre Diversification**: Invest in underrepresented genres like Documentaries and Animation.
- **Balance Critical and Audience Appeal**: Strive to create content that appeals to both critics and mass audiences.
- **Support Emerging Directors**: Provide opportunities for new filmmakers to bring fresh perspectives.
- **Optimize Movie Length**: Consider audience attention spans while determining runtime.
- **Expand International Collaborations**: Increase co-productions to diversify content portfolios.
- **Leverage Award-Winning Films**: Market critically acclaimed movies to maximize returns.

Project Stats

- **Dataset Size**: 21,000 movies
- **Visualization Tools**: MongoDB Atlas Charts
- **Database Queries**: CRUD operations performed in MongoDB Shell

This analysis provides valuable insights into the evolving trends of the movie industry and helps stakeholders make data-driven decisions.