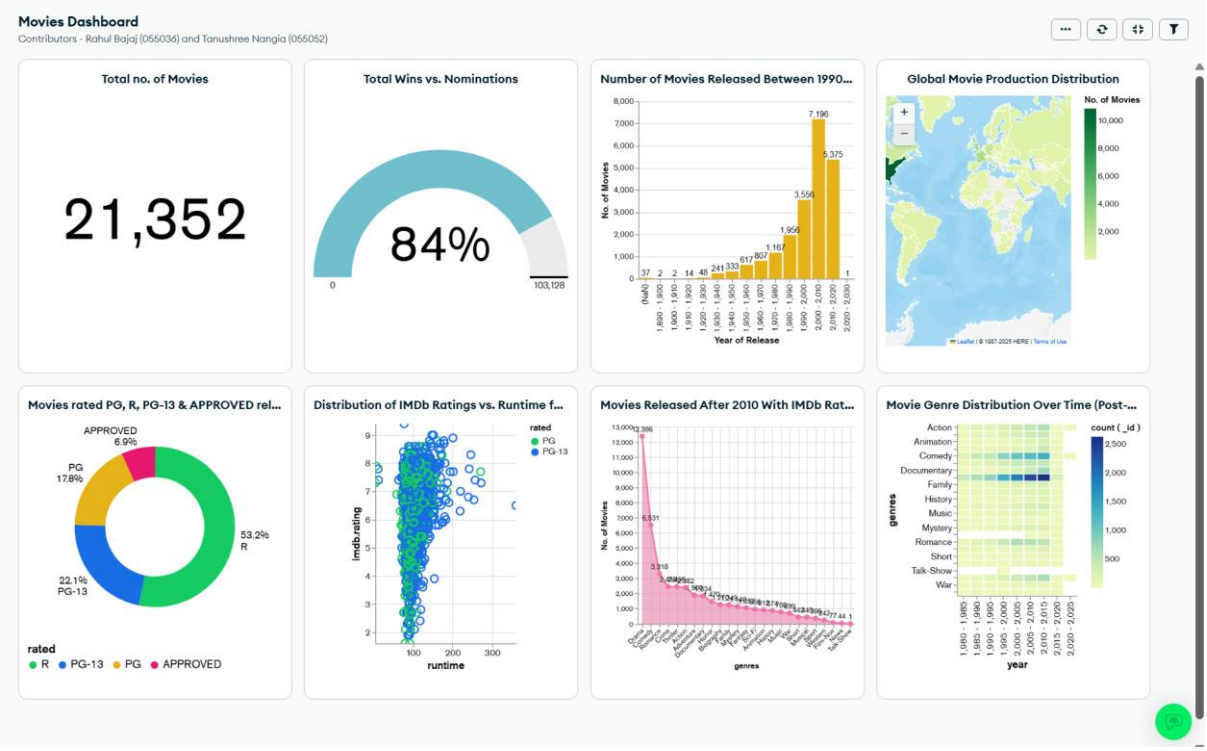
Data-Driven Movie Insights: Building a Dynamic Dashboard with MongoDB CRUD and Aggregation Framework



Submitted by:

Rahul Bajaj & Tanushree Nangia

Roll No: 055036, 055052

Group: 7

Section: K

1. Project Information

Project Title: Data Analysis and Dashboard Creation using MongoDB on Sample Movies Dataset

Technology Used: MongoDB, Aggregation Framework, CRUD Operations

Dataset: Sample Movies Dataset provided by MongoDB

Tools & Environment: MongoDB Atlas

2. Description of Data

The Sample Movies dataset comprises movie-related data with the following attributes:

- **_id:** Unique identifier for each movie.
- **Title:** Name of the movie.
- **Year:** Release year of the movie.
- **Genres:** Type/category of the movie (e.g., Drama, Sci-Fi, Action).
- **IMDB Ratings:**
    - **rating:** Viewer ratings of the movie.
    - **votes:** Number of votes received on IMDB.
- **Directors and Cast:**
    - **directors:** List of directors associated with the movie.
    - **cast:** List of actors starring in the movie.
- **Runtime:** Duration of the movie in minutes.
- **Countries:** Countries where the movie was produced.
- **Languages:** Languages spoken in the movie.
- **Awards:**
    - **wins:** Number of awards won.
    - **nominations:** Number of nominations received.
    - **text:** Summary of awards.
- **Ratings from Other Sources:**
    - **metacritic:** Metacritic rating.
    - **tomatoes:** Includes various Rotten Tomatoes metrics like box office revenue, critic ratings, and viewer ratings.

- **Additional Information:**
    - **plot:** Short description of the movie.
    - **fullplot:** Detailed movie summary.
    - **rated:** Movie rating classification (e.g., PG-13, R).
    - **released:** Official release date.
    - **writers:** List of writers involved in scriptwriting.
    - **website:** Official website link if available.

3. Project Objectives | Problem Statements

The main objectives of the project are:

- To explore and analyse the Sample Movies dataset using MongoDB.
- To implement CRUD operations (Create, Read, Update, Delete) for data manipulation.
- To apply advanced queries including comparison operators, logical operators, and aggregation.
- To visualize key insights using a dashboard.
- To derive meaningful managerial insights based on the dataset.

4. Analysis of Data

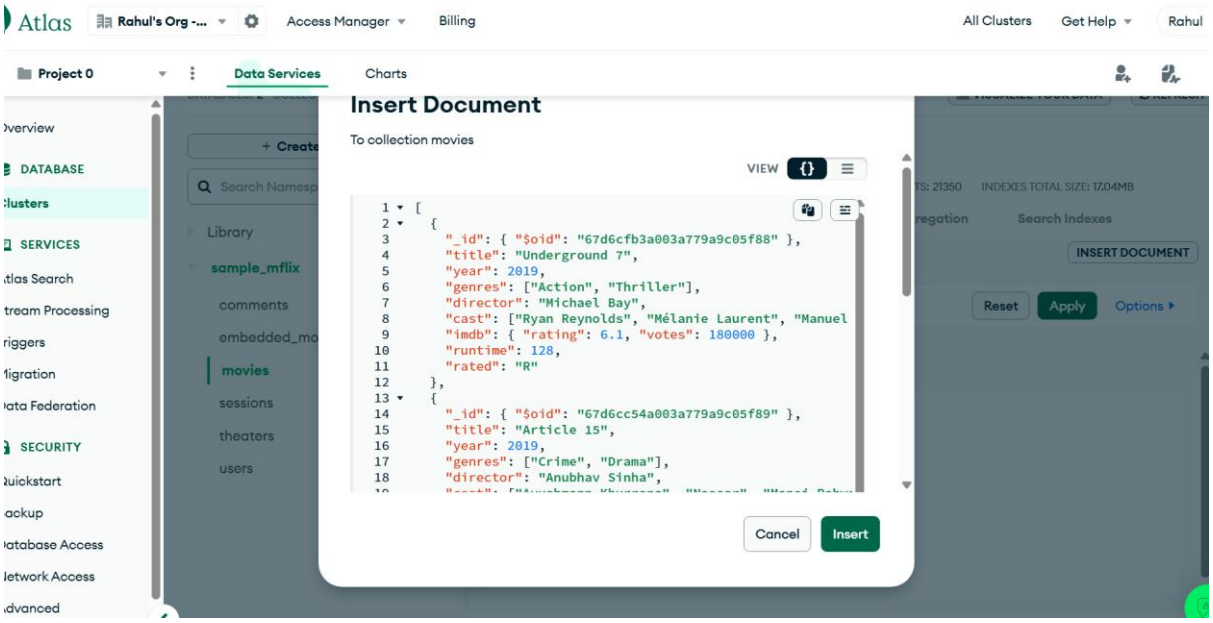
CRUD Operations in MongoDB

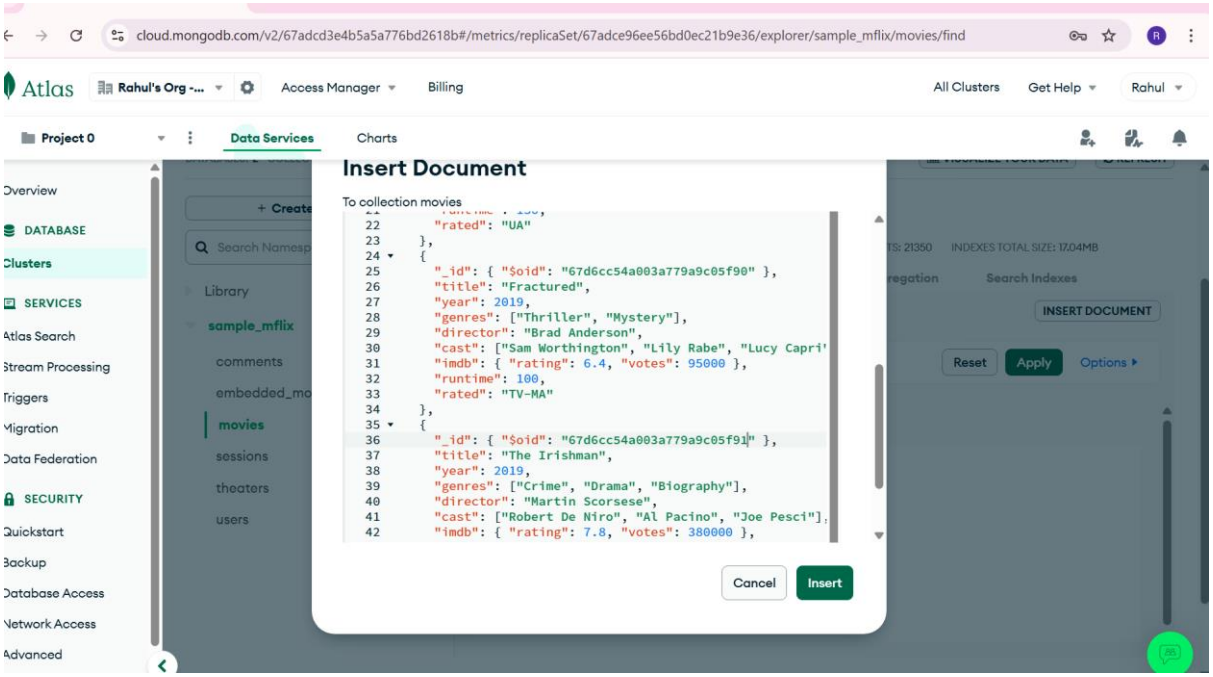
MongoDB is a NoSQL database that allows flexible and efficient handling of unstructured and semi-structured data. The core operations performed on a MongoDB database follow the CRUD paradigm:

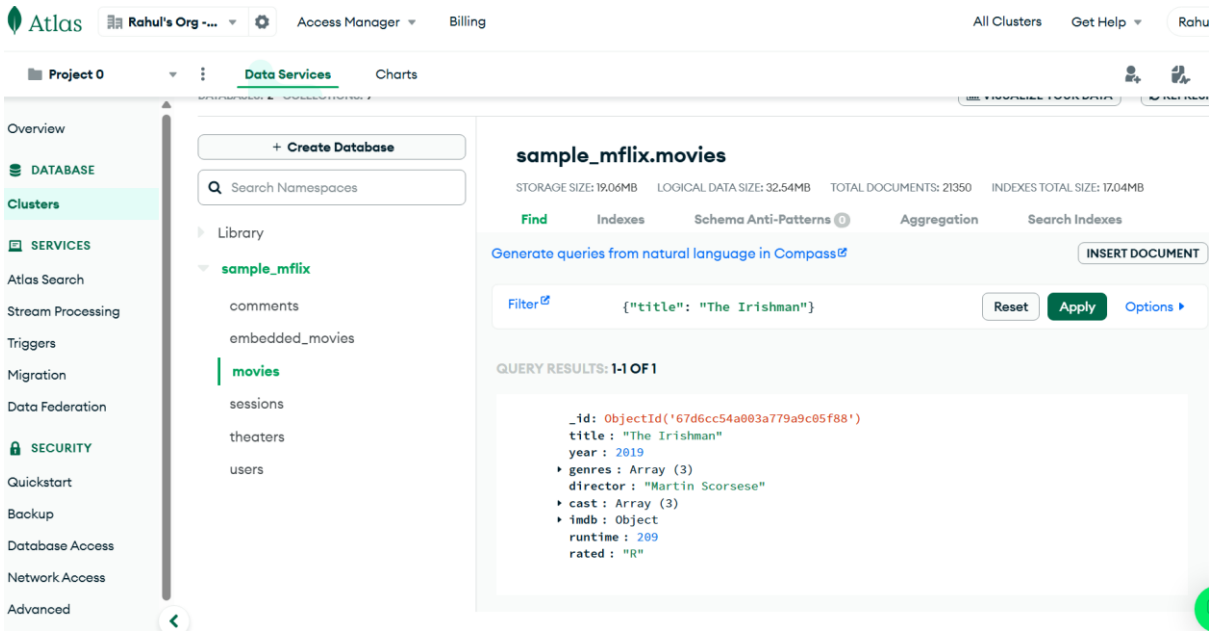
Create (Insert Operations)

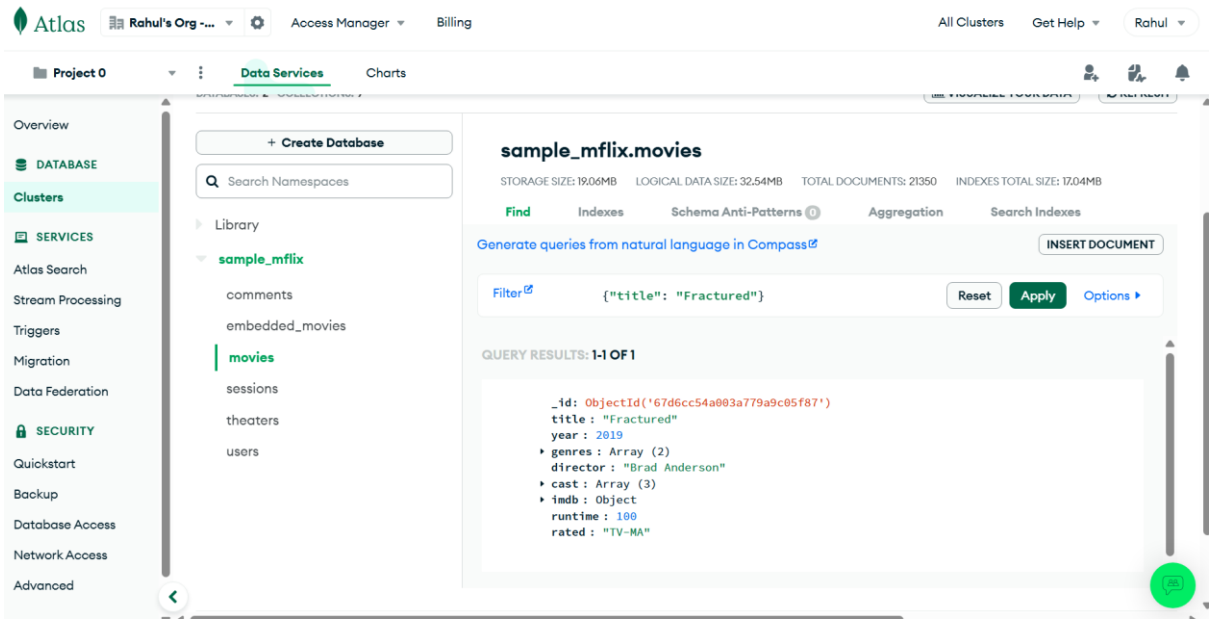
Inserting new documents into the collection.

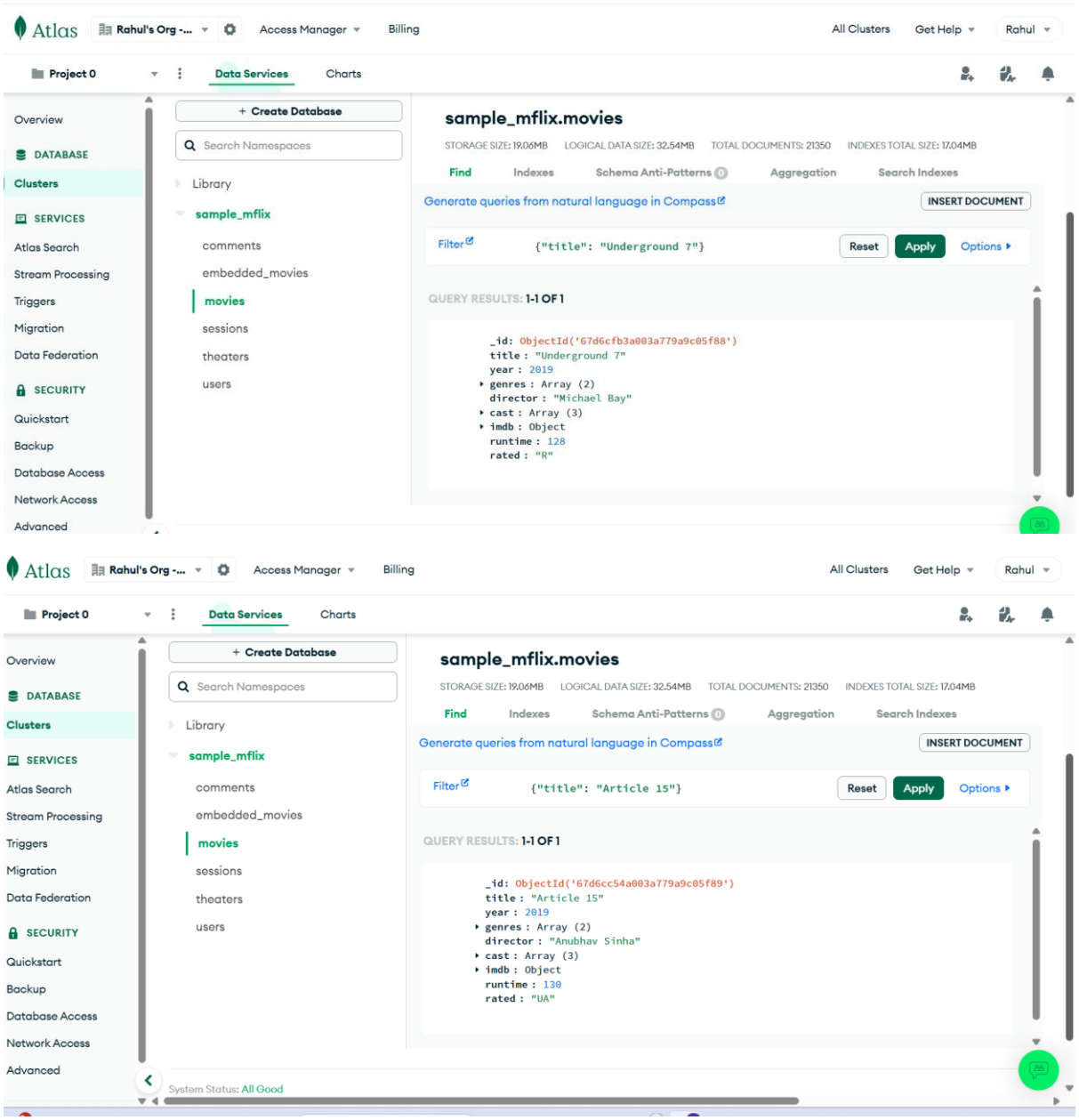
Movies: Underground 7, Fractured, The Irishman and Article 15.


Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Insert Document
To collection movies
VIEW
{}
1 [
2 {
3 "_id": { "$oid": "67d6cfb3a003a779a9c05f88" },
4 "title": "Underground 7",
5 "year": 2019,
6 "genres": ["Action", "Thriller"],
7 "director": "Michael Bay",
8 "cast": ["Ryan Reynolds", "Mélanie Laurent", "Manuel
9 "imdb": { "rating": 6.1, "votes": 180000 },
10 "runtime": 128,
11 "rated": "R"
12 },
13 {
14 "_id": { "$oid": "67d6cc54a003a779a9c05f89" },
15 "title": "Article 15",
16 "year": 2019,
17 "genres": ["Crime", "Drama"],
18 "director": "Anubhav Sinha",
Cancel
Insert
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Advanced
Library
sample_mflix
comments
embedded_mo
movies
sessions
theaters
users
INSERT DOCUMENT
Reset
Apply
Options
Search Indexes


cloud.mongodb.com/v2/67adcd3e4b5a5a776bd2618b#/metrics/replicaSet/67adce96ee56bd0ec21b9e36/explorer/sample_mflix/movies/find
Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Insert Document
To collection movies
22 "rated": "UA"
23 },
24 {
25 "_id": { "$oid": "67d6cc54a003a779a9c05f90" },
26 "title": "Fractured",
27 "year": 2019,
28 "genres": ["Thriller", "Mystery"],
29 "director": "Brad Anderson",
30 "cast": ["Sam Worthington", "Lily Rabe", "Lucy Capri
31 "imdb": { "rating": 6.4, "votes": 95000 },
32 "runtime": 100,
33 "rated": "TV-MA"
34 },
35 {
36 "_id": { "$oid": "67d6cc54a003a779a9c05f91" },
37 "title": "The Irishman",
38 "year": 2019,
39 "genres": ["Crime", "Drama", "Biography"],
40 "director": "Martin Scorsese",
41 "cast": ["Robert De Niro", "Al Pacino", "Joe Pesci"],
42 "imdb": { "rating": 7.8, "votes": 380000 },
Cancel
Insert
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Advanced
Library
sample_mflix
comments
embedded_mo
movies
sessions
theaters
users
INSERT DOCUMENT
Reset
Apply
Options
Search Indexes


Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahu
Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Advanced
+ Create Database
Search Namespaces
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
sample_mflix.movies
STORAGE SIZE: 19.06MB
LOGICAL DATA SIZE: 32.54MB
TOTAL DOCUMENTS: 21350
INDEXES TOTAL SIZE: 17.04MB
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Generate queries from natural language in Compass
INSERT DOCUMENT
Filter
{"title": "The Irishman"}
Reset
Apply
Options
QUERY RESULTS: 1-1 OF 1
_id: ObjectId('67d6cc54a003a779a9c05f88')
title : "The Irishman"
year : 2019
genres : Array (3)
director : "Martin Scorsese"
cast : Array (3)
imdb : Object
runtime : 209
rated : "R"

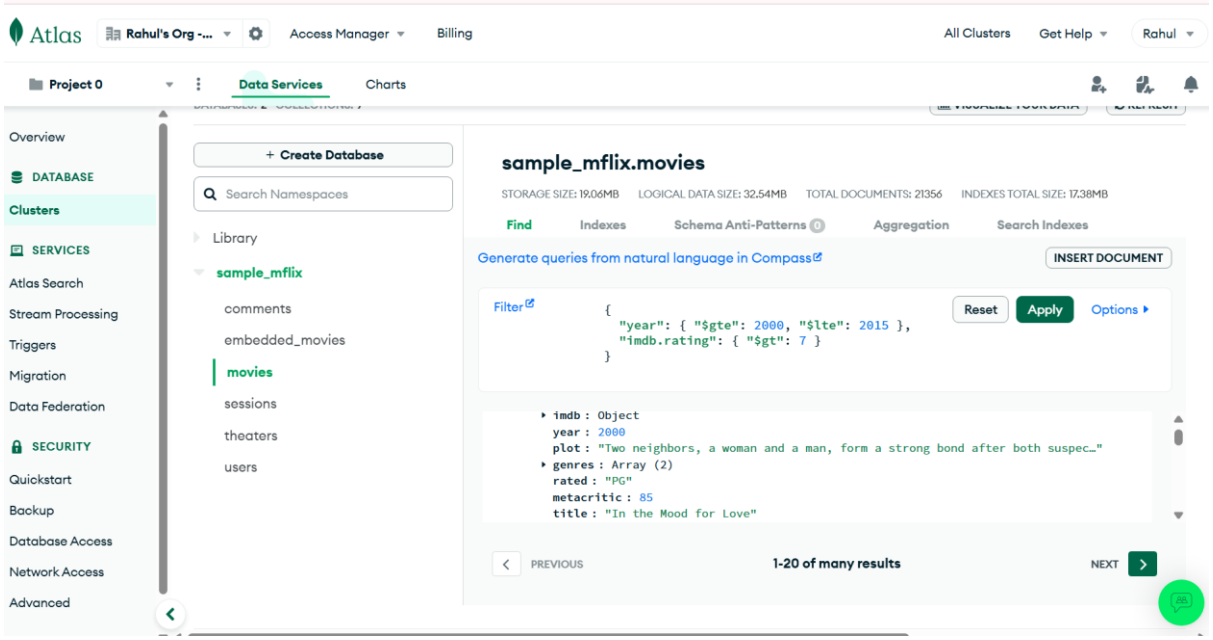

Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Advanced
+ Create Database
Search Namespaces
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
sample_mflix.movies
STORAGE SIZE: 19.06MB
LOGICAL DATA SIZE: 32.54MB
TOTAL DOCUMENTS: 21350
INDEXES TOTAL SIZE: 17.04MB
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Generate queries from natural language in Compass
INSERT DOCUMENT
Filter
{"title": "Fractured"}
Reset
Apply
Options
QUERY RESULTS: 1-1 OF 1
_id: ObjectId('67d6cc54a003a779a9c05f87')
title : "Fractured"
year : 2019
genres : Array (2)
director : "Brad Anderson"
cast : Array (3)
imdb : Object
runtime : 100
rated : "TV-MA"

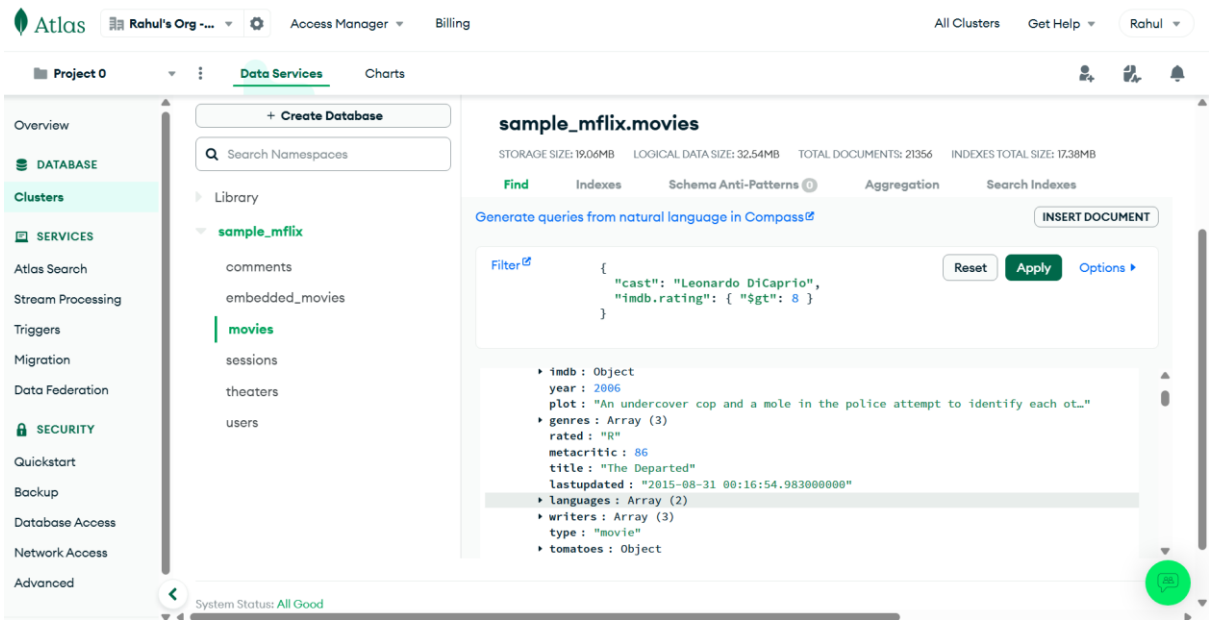


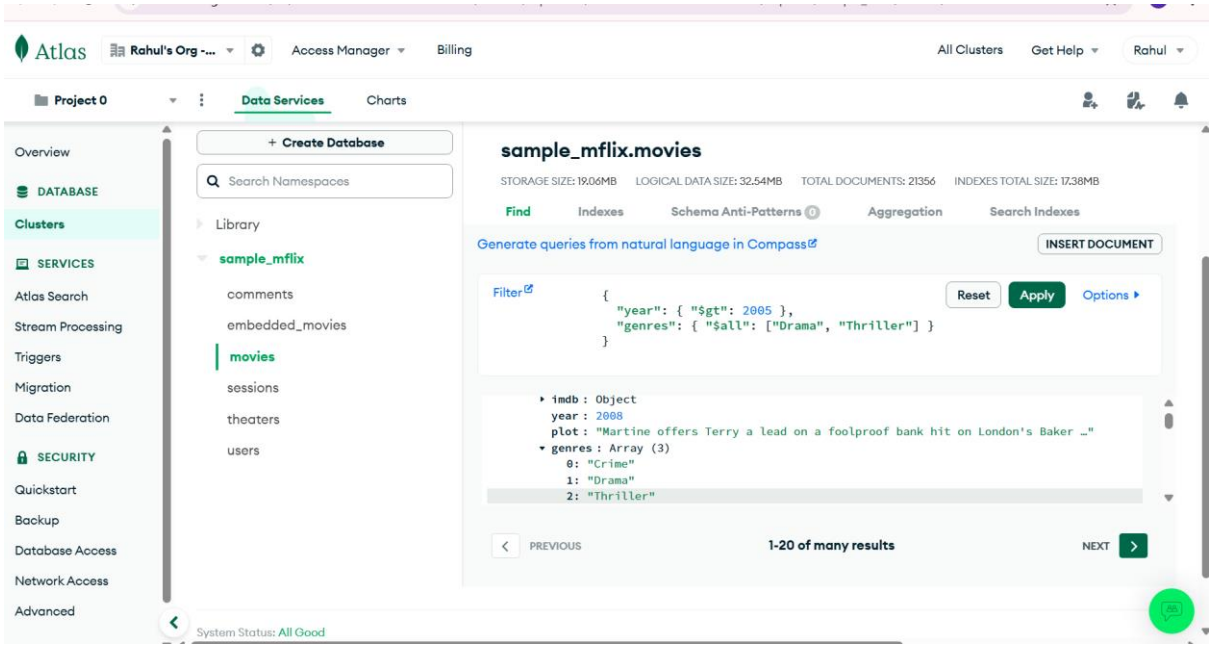
Successfully inserted 4 movie records into the movies collection.

Read (Retrieve Operations)

Retrieving data from the collection.





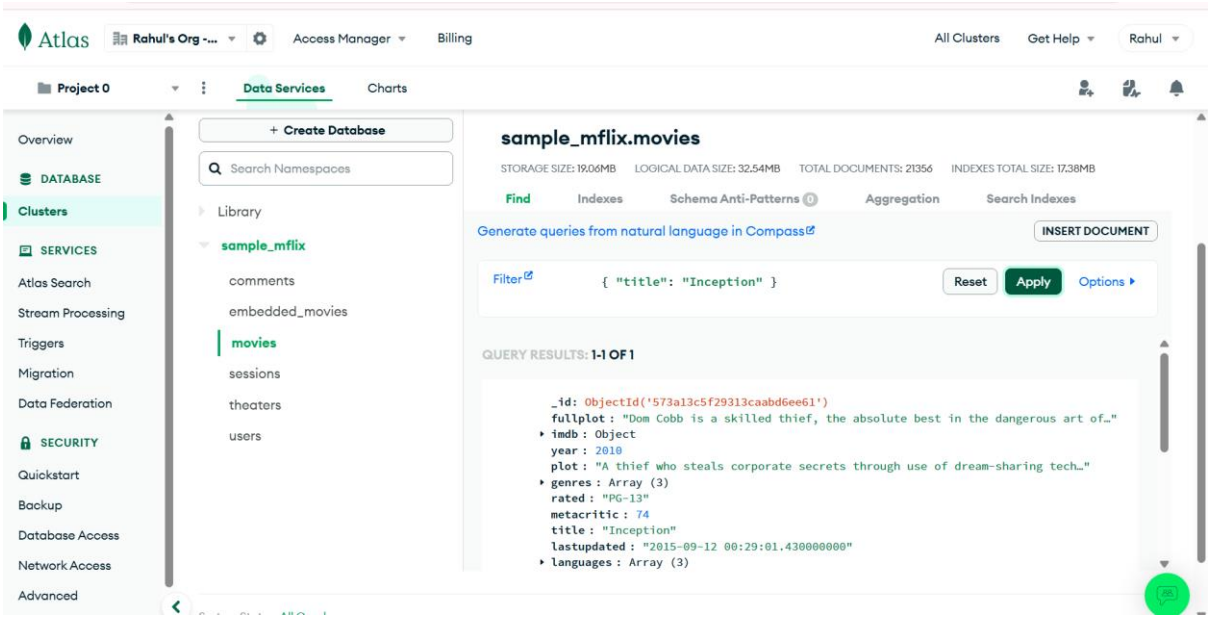


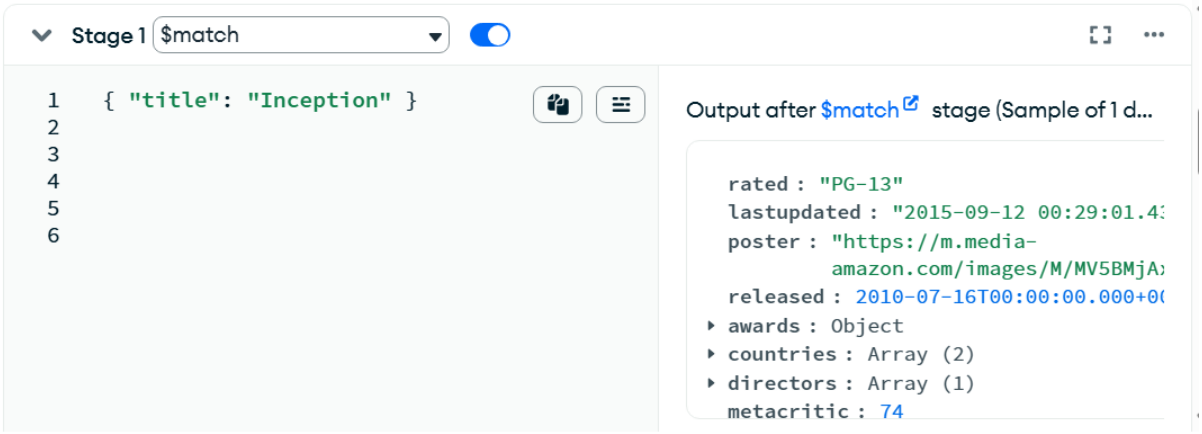
Successfully retrieved details for movies with different ratings.

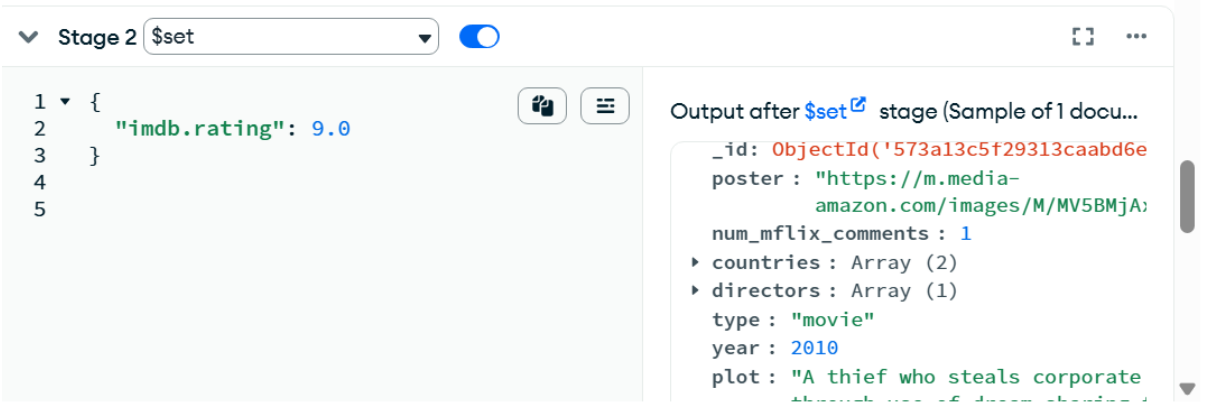
Update (Modify Operations)

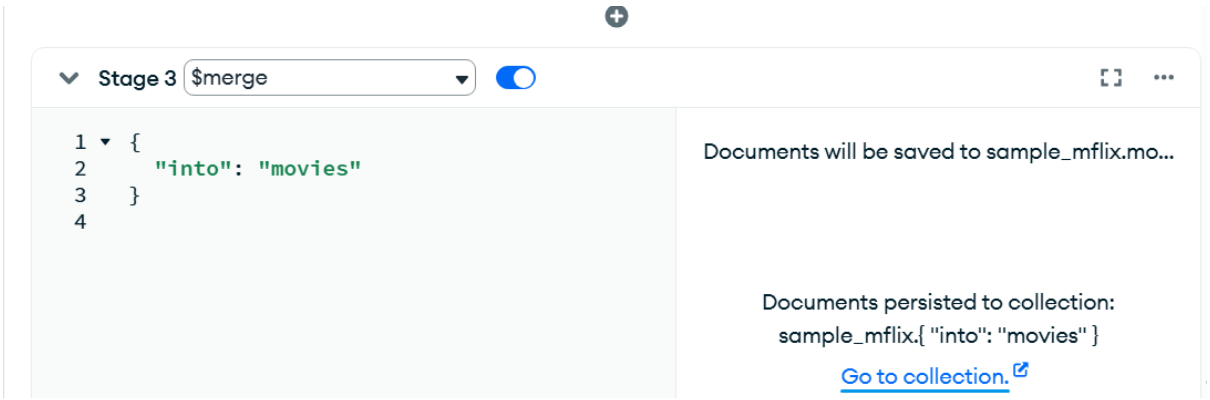
Modifying existing documents in the collection.

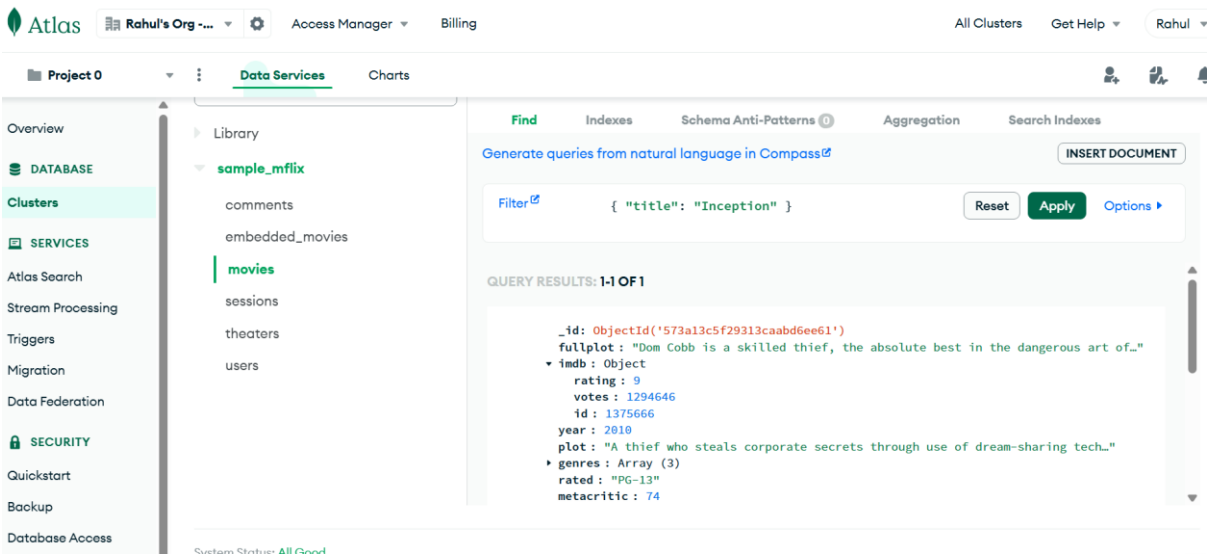
“Inception” IMDb rating updated to 9.0



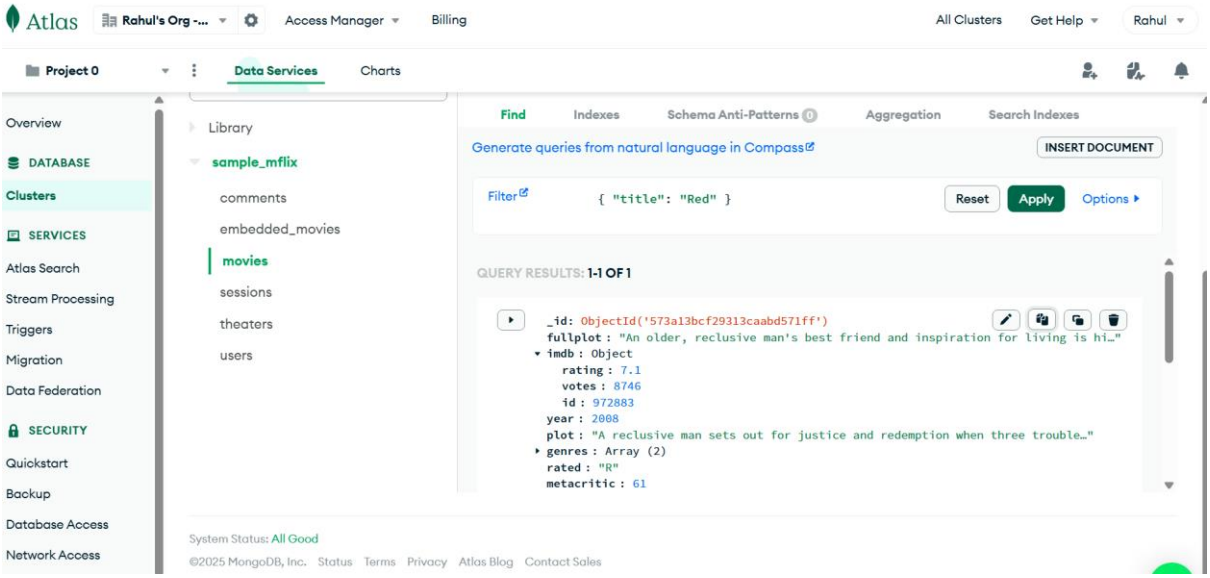


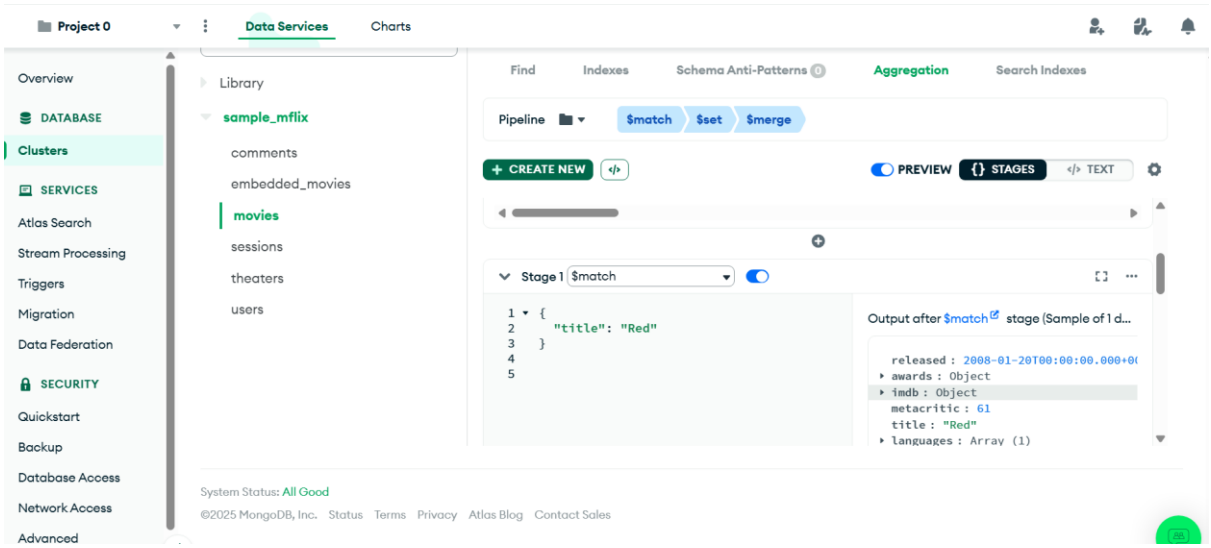


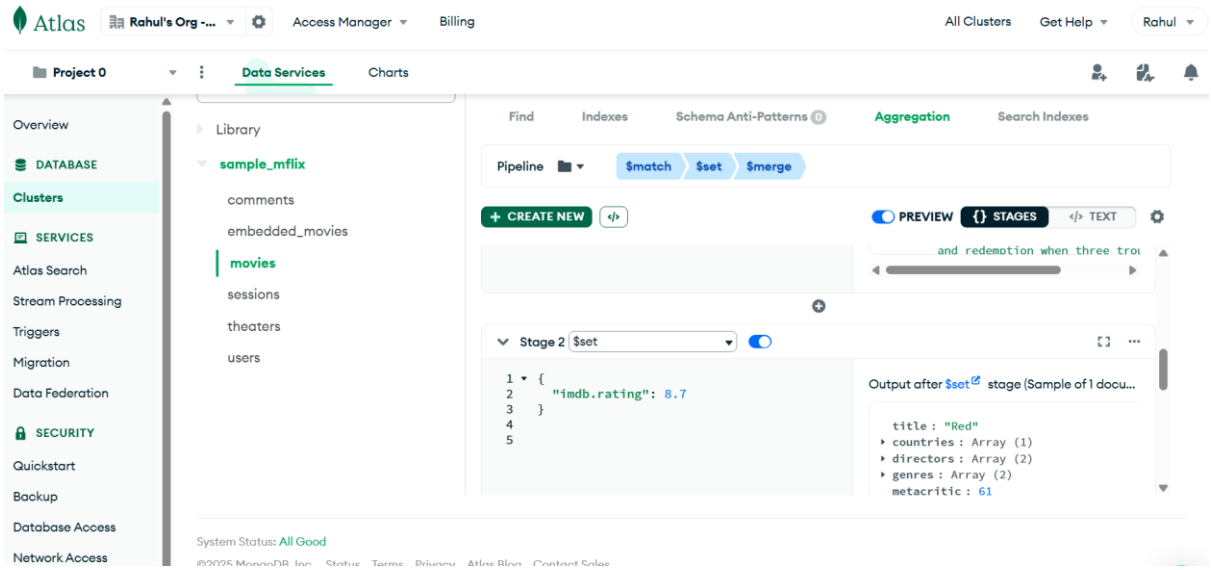


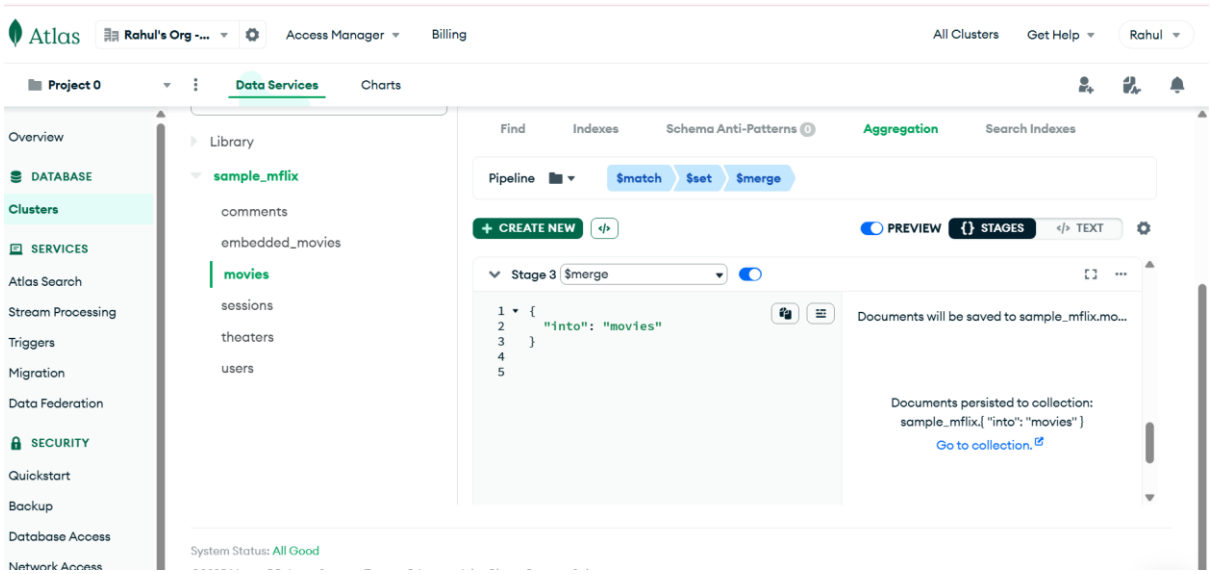


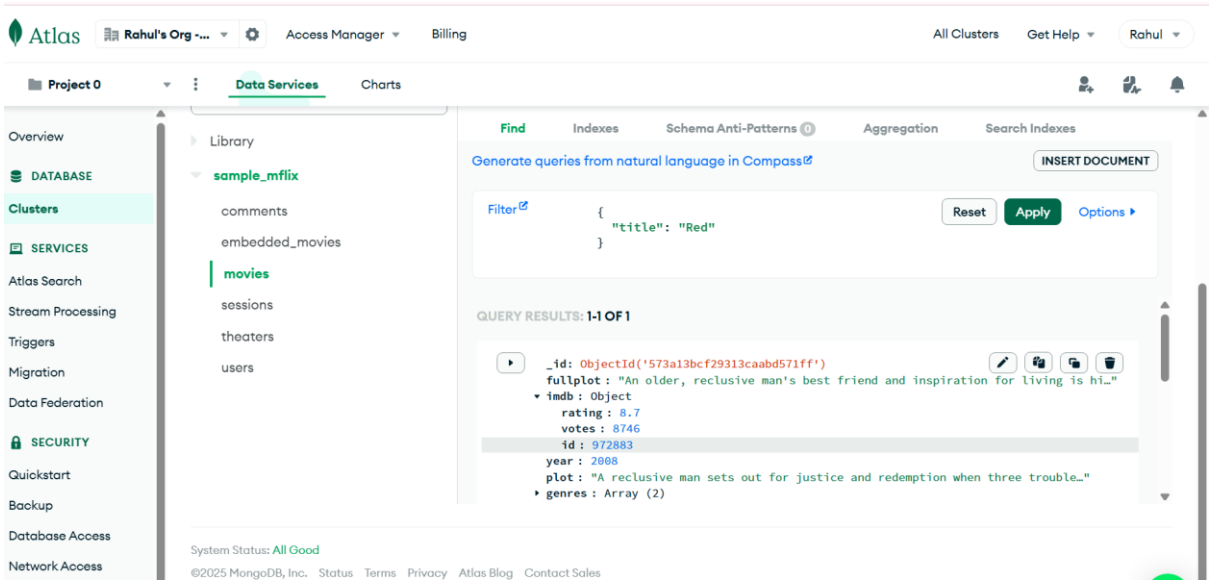
“Red” IMDb rating updated from 7.1 to 8.7



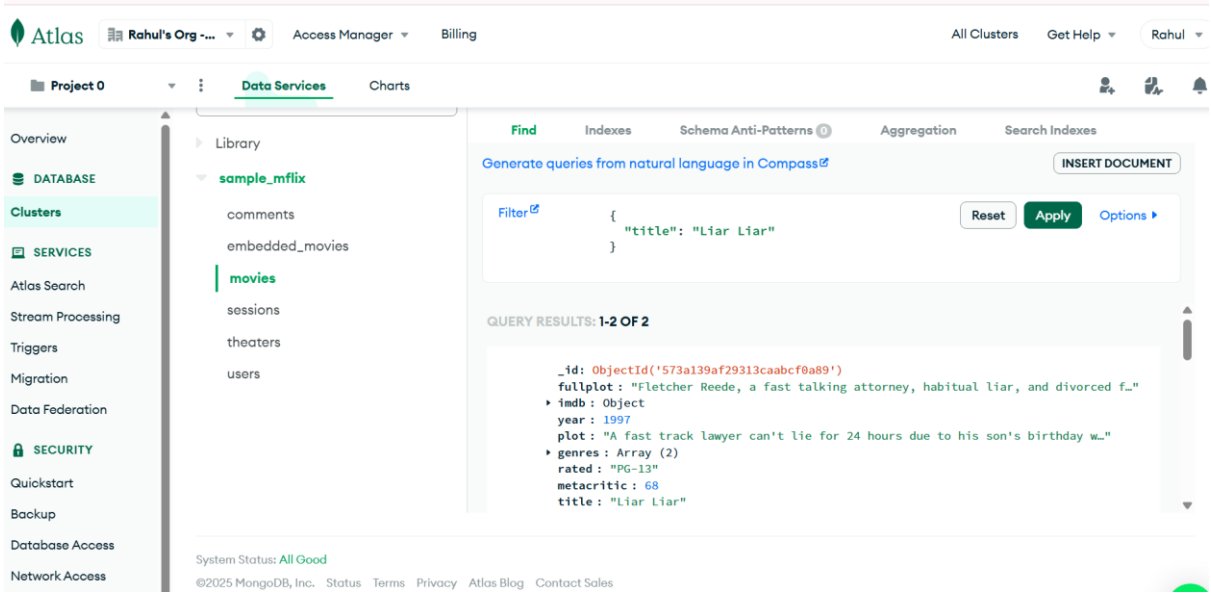

Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Advanced
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Pipeline
$match
$set
$merge
CREATE NEW
PREVIEW
STAGES
TEXT
Stage 1 $match
{
"title": "Red"
}
Output after $match stage (Sample of 1 d...
released : 2008-01-20T00:00:00.000+0(
awards : Object
imdb : Object
metacritic : 61
title : "Red"
languages : Array (1)
System Status: All Good
©2025 MongoDB, Inc. Status Terms Privacy Atlas Blog Contact Sales

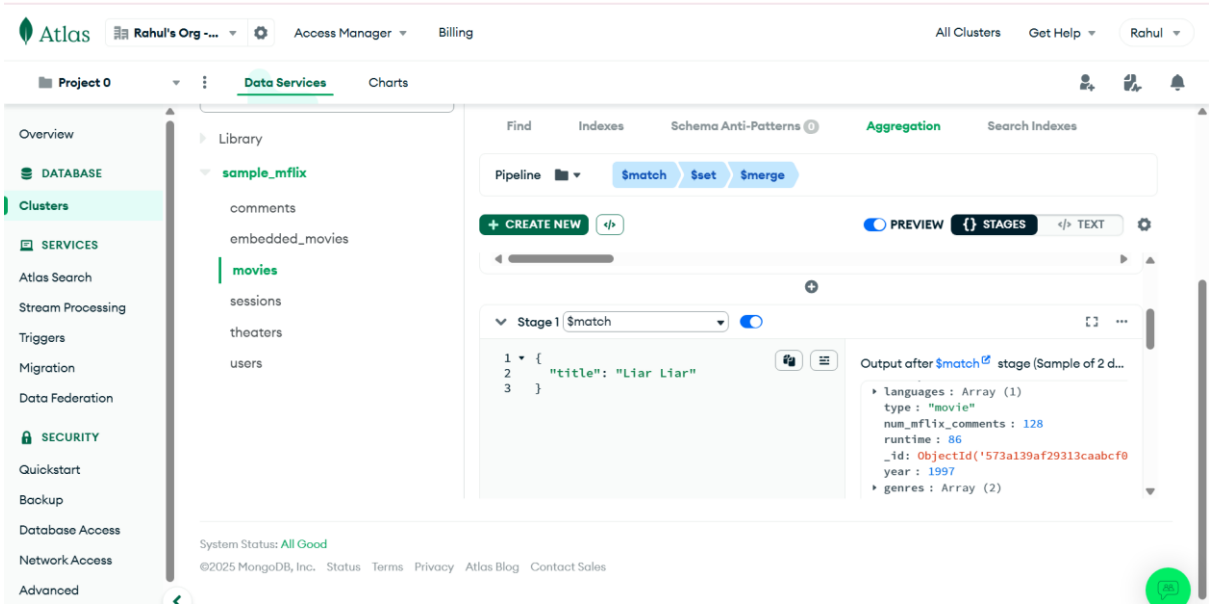

Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Pipeline
$match
$set
$merge
CREATE NEW
PREVIEW
STAGES
TEXT
and redemption when three trou
Stage 2 $set
{
"imdb.rating": 8.7
}
Output after $set stage (Sample of 1 docu...
title : "Red"
countries : Array (1)
directors : Array (2)
genres : Array (2)
metacritic : 61
System Status: All Good

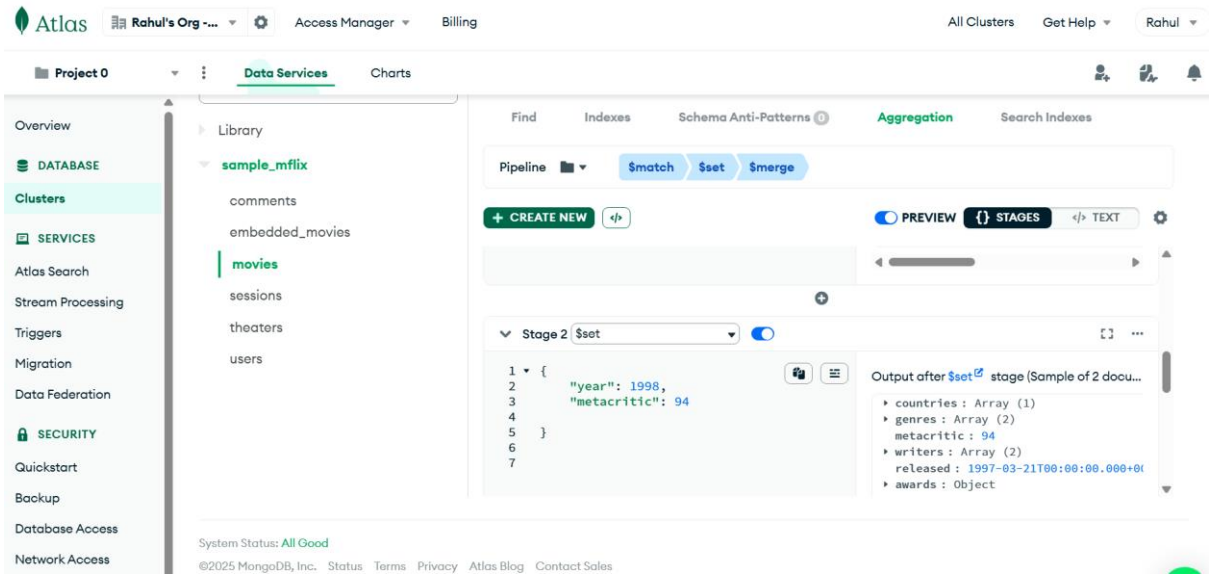

Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Pipeline
$match
$set
$merge
CREATE NEW
PREVIEW
STAGES
TEXT
Stage 3 $merge
{
"into": "movies"
}
Documents will be saved to sample_mflix.mo...
Documents persisted to collection:
sample_mflix.["into": "movies"]
Go to collection.
System Status: All Good

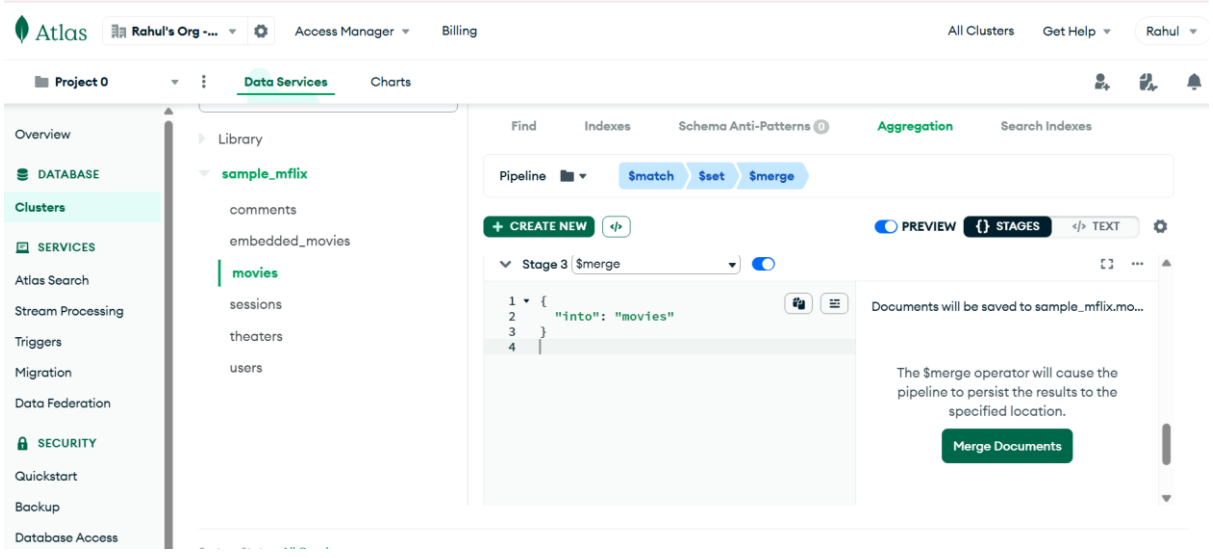


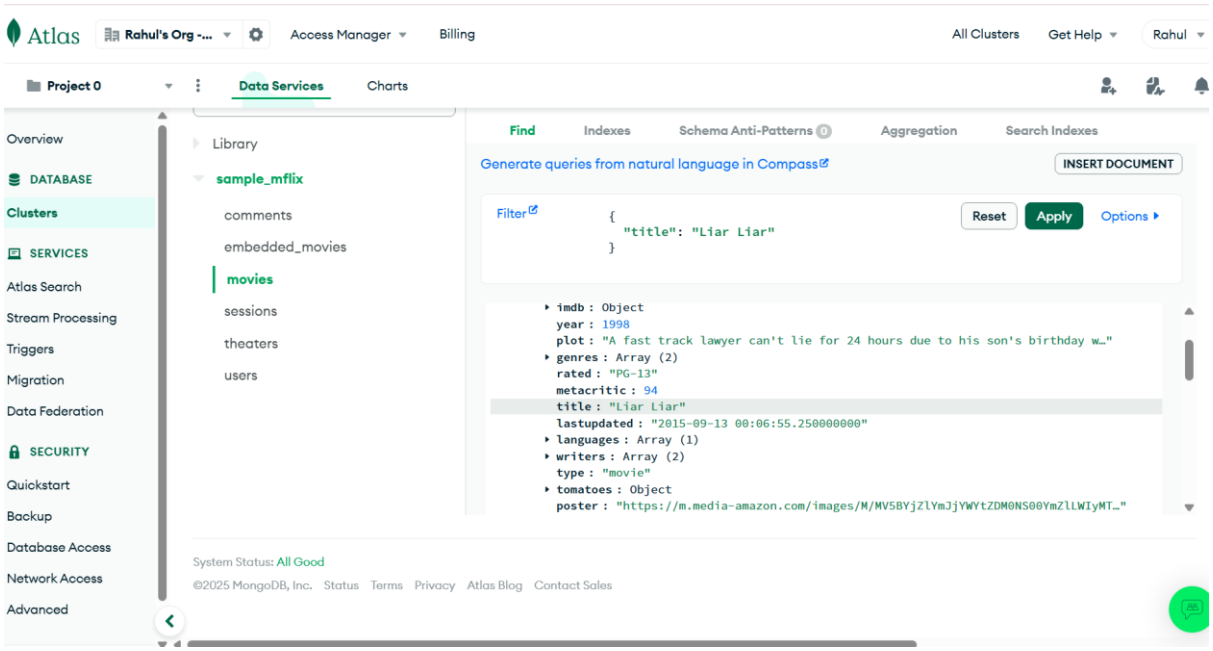
“Liar Liar” – year updated from 1997 to 1998 and Metacritic score from 68 to 94




Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Advanced
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Pipeline
$match
$set
$merge
CREATE NEW
PREVIEW
STAGES
TEXT
Stage 1 $match
{
"title": "Liar Liar"
}
Output after $match stage (Sample of 2 d...
languages : Array (1)
type : "movie"
num_mflix_comments : 128
runtime : 86
_id: ObjectId('573a139af29313caabcf0
year : 1997
genres : Array (2)
System Status: All Good
©2025 MongoDB, Inc. Status Terms Privacy Atlas Blog Contact Sales


Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Network Access
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Pipeline
$match
$set
$merge
CREATE NEW
PREVIEW
STAGES
TEXT
Stage 2 $set
{
"year": 1998,
"metacritic": 94
}
Output after $set stage (Sample of 2 docu...
countries : Array (1)
genres : Array (2)
metacritic : 94
writers : Array (2)
released : 1997-03-21T00:00:00.000+0(
awards : Object
System Status: All Good
©2025 MongoDB, Inc. Status Terms Privacy Atlas Blog Contact Sales


Atlas
Rahul's Org -...
Access Manager
Billing
All Clusters
Get Help
Rahul
Project 0
Data Services
Charts
Overview
DATABASE
Clusters
SERVICES
Atlas Search
Stream Processing
Triggers
Migration
Data Federation
SECURITY
Quickstart
Backup
Database Access
Library
sample_mflix
comments
embedded_movies
movies
sessions
theaters
users
Find
Indexes
Schema Anti-Patterns
Aggregation
Search Indexes
Pipeline
$match
$set
$merge
CREATE NEW
PREVIEW
STAGES
TEXT
Stage 3 $merge
{
"into": "movies"
}
Documents will be saved to sample_mflix.mo...
The $merge operator will cause the pipeline to persist the results to the specified location.
Merge Documents

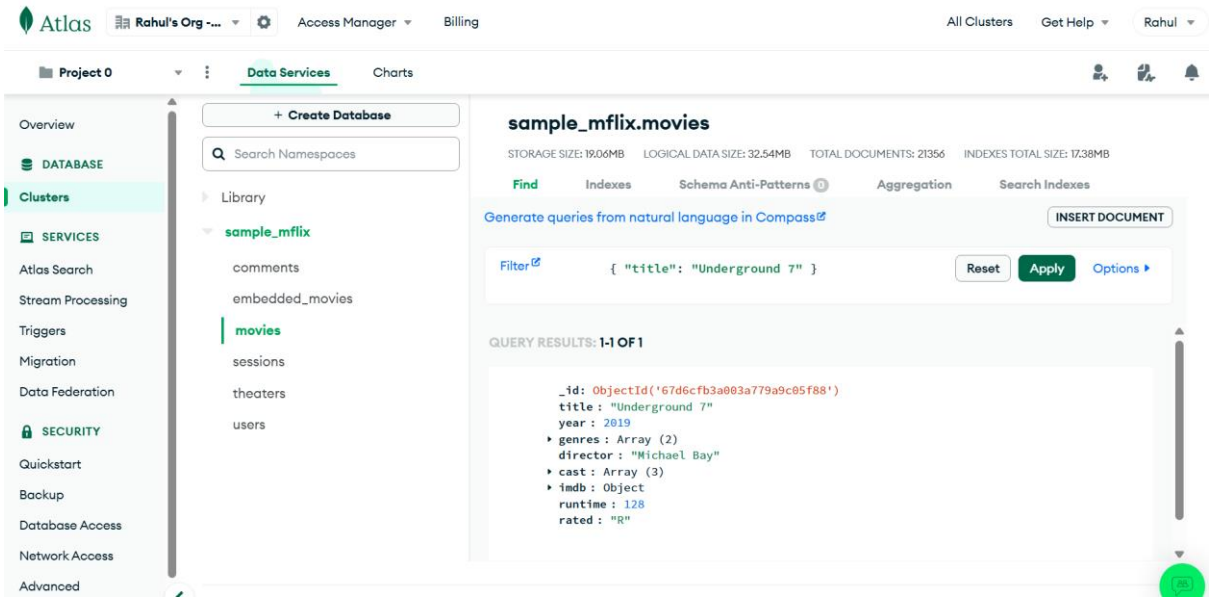


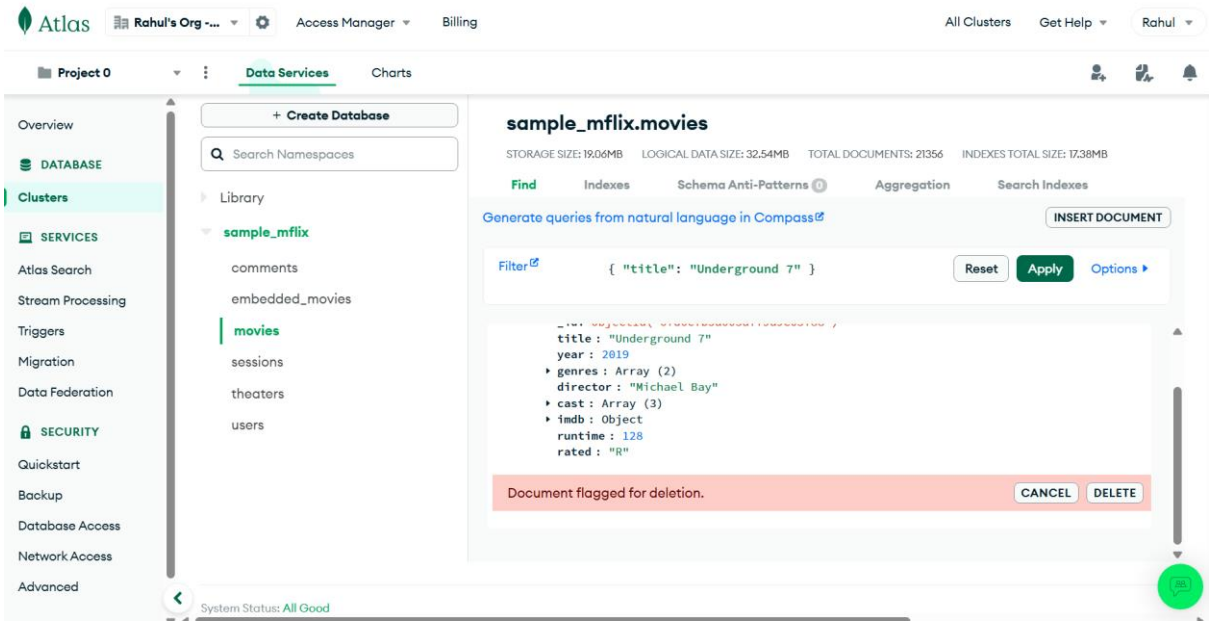
Successfully updated records for "Inception", "RED", and "Liar Liar".

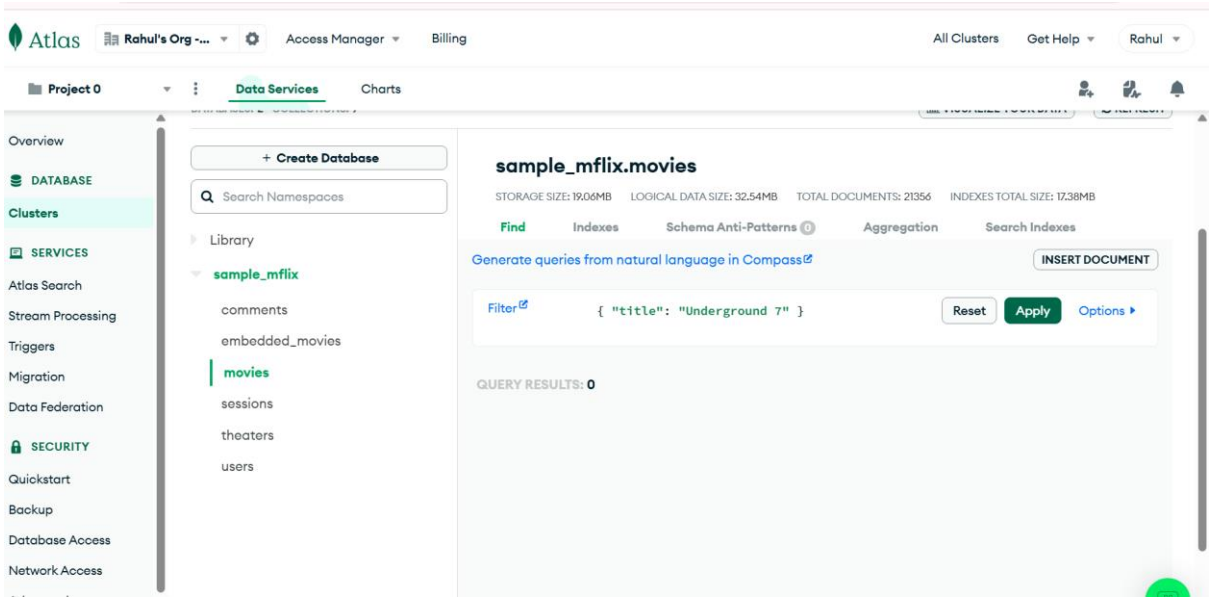
Delete (Remove Operations)

Removing documents from the collection.

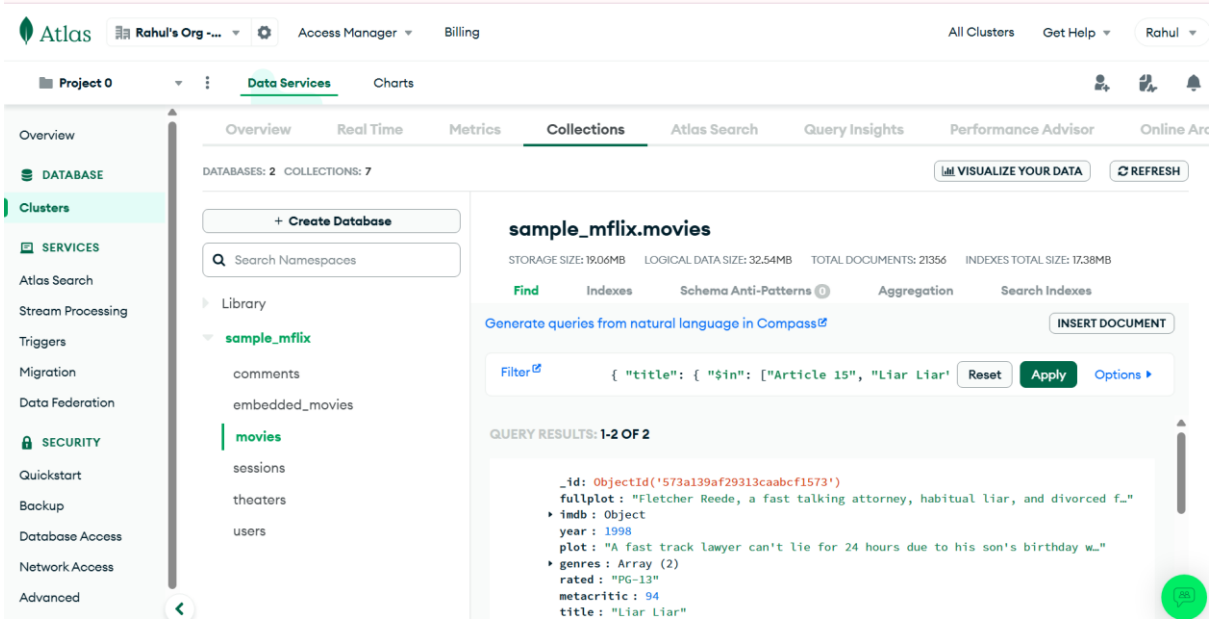
Delete record underground 7

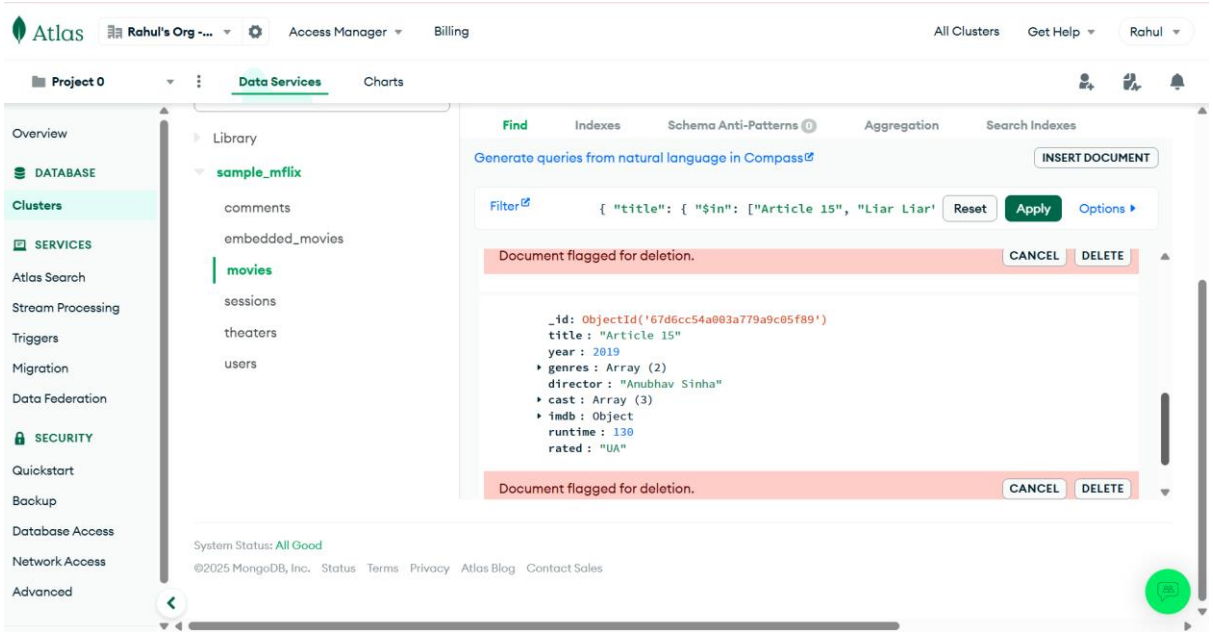


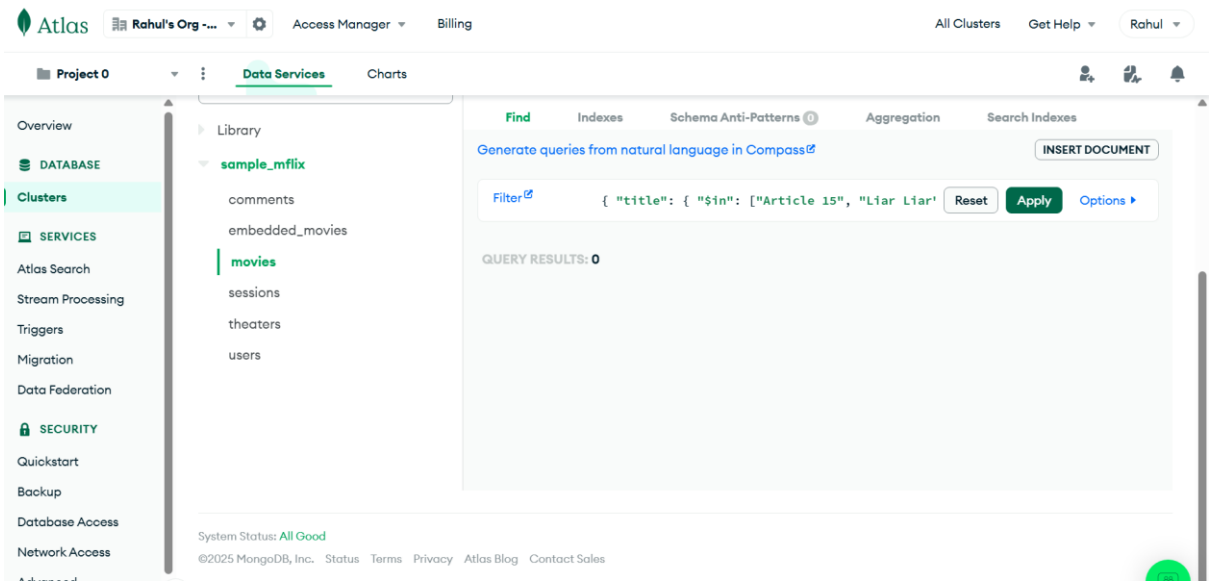




Deleted records – “Liar Liar” and “Article 15”







Successfully deleted "Underground 7", "Liar Liar", and "Article 15".

Summary of CRUD Operations:

Operation	Movies Affected	Details
Create	4	Added Underground 7, The Irishman, Fractured, Article 15
Read	3	Retrieved Inception, Underground 7, Article 15
Update	3	Updated Inception, RED, Liar Liar
Delete	3	Removed Underground 7, Liar Liar, Article 15

Querying with Comparison Operators

- Find movies with an IMDB rating above 8 ($gt operator).
- List movies released after 2010 ($gte operator).
- Identify movies with a runtime between 90 and 150 minutes ($gte and $lte operators).

Filtering with Logical Operators

- Find movies that are either Action or Adventure ($or operator).
- Retrieve movies that are not in the Horror genre ($not operator).
- Find movies directed by a specific director AND having a high rating ($and operator).

Aggregation Queries

- Compute the average IMDB rating for each genre ($group by genre).
- Find the top 5 highest-rated movies ($sort and $limit).
- Count the number of movies produced per country ($group by country).

Movies Dashboard

1. Total no. of Movies

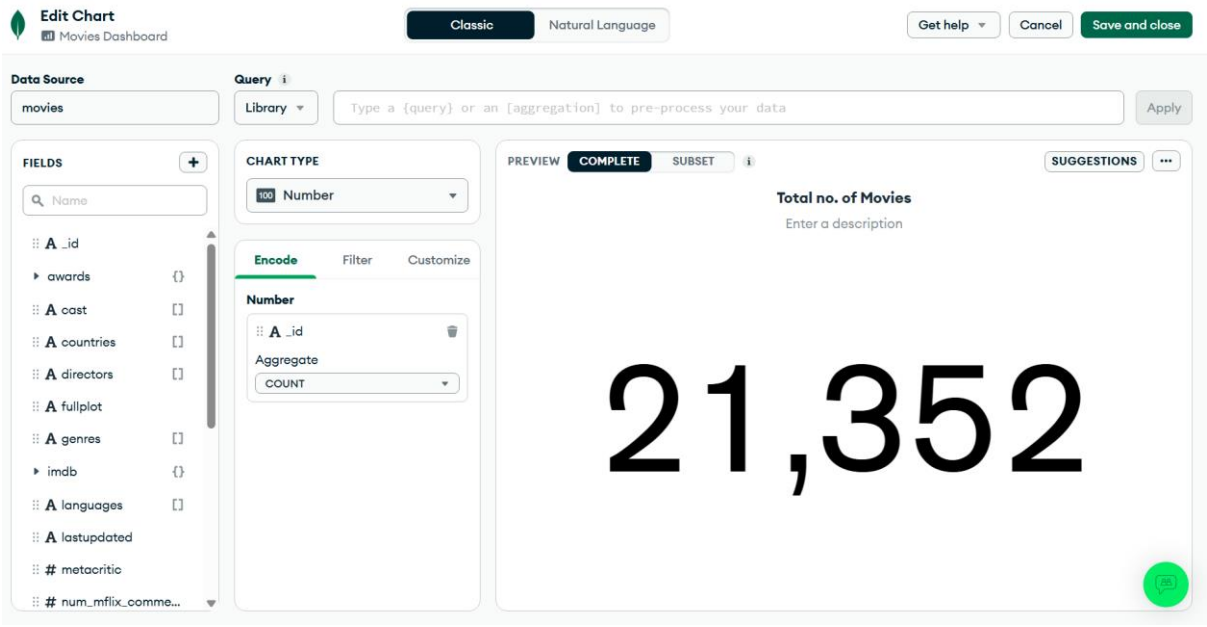


Chart Type Used: Number Chart

- Displays a single numeric value, making it ideal for key performance indicators (KPIs).
- Aggregates the count of unique movie entries using _id.

Purpose & Interpretation:

- The dataset contains 21,352 movies, representing the total number of records available.
- This metric serves as a foundational statistic, helping users understand the dataset's scale.
- Knowing the total count of movies is essential for further analysis, such as tracking industry trends, categorization by genre, or comparing production volumes over time.

How It Helps in Analysis:

- Data Completeness Check: Ensures that the dataset contains a significant number of records for meaningful analysis.
- Comparison Across Filters: Can be used as a reference point when breaking down data by genres, countries, or languages.

- Decision-Making Basis: If used in a business scenario, understanding the number of movies helps in content acquisition, marketing strategies, or trend analysis in the film industry.

2. Total Wins vs Nominations

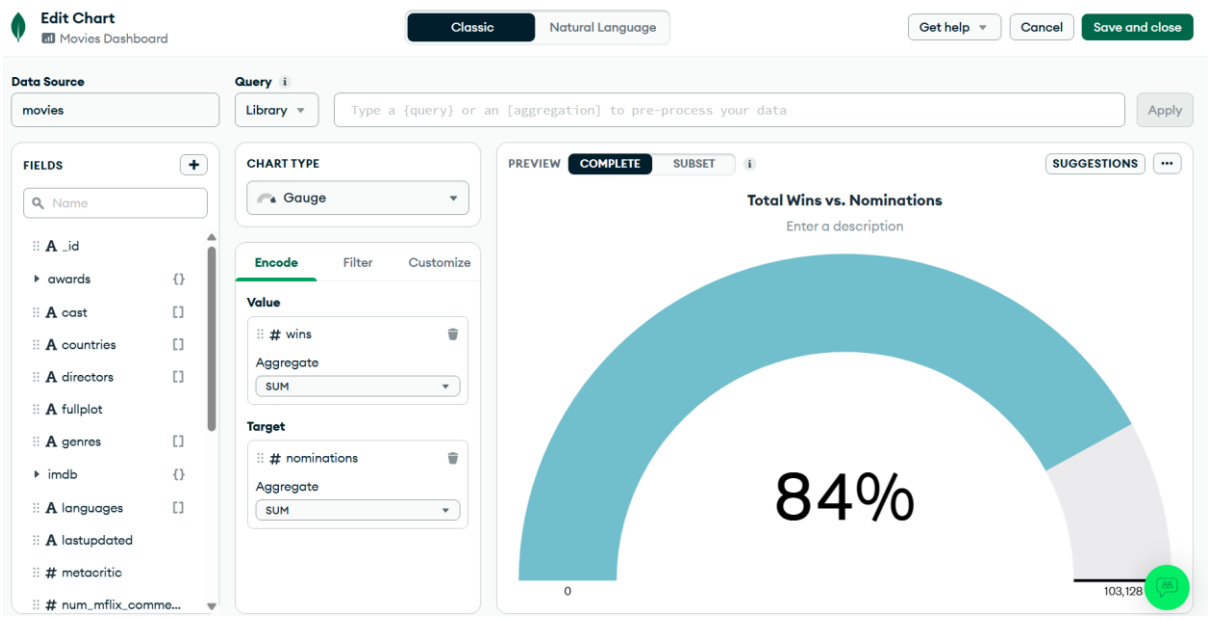


Chart Type Used: Gauge Chart

- Represents a proportion (percentage of wins out of total nominations).
- Uses SUM(wins) / SUM(nominations) * 100 to calculate the success rate of movies in awards.

Purpose & Interpretation:

- The dataset shows that 84% of award nominations resulted in wins, indicating a high success rate.
- This metric provides a quick snapshot of how well movies in the dataset perform in awards.
- A high percentage suggests that movies in the dataset generally have strong critical or industry recognition.

How It Helps in Analysis:

- Performance Benchmarking: Helps in evaluating whether the movies in this dataset are generally more successful in awards compared to industry standards.
- Comparative Insights: If segmented by genre, studio, or director, this metric can reveal which categories have the highest success rates.

- Strategic Decision Making: Useful for stakeholders in media and entertainment to analyse past success rates and make informed decisions on movie investments or promotions.

3. Number of Movies Released Between 1990 and 2015

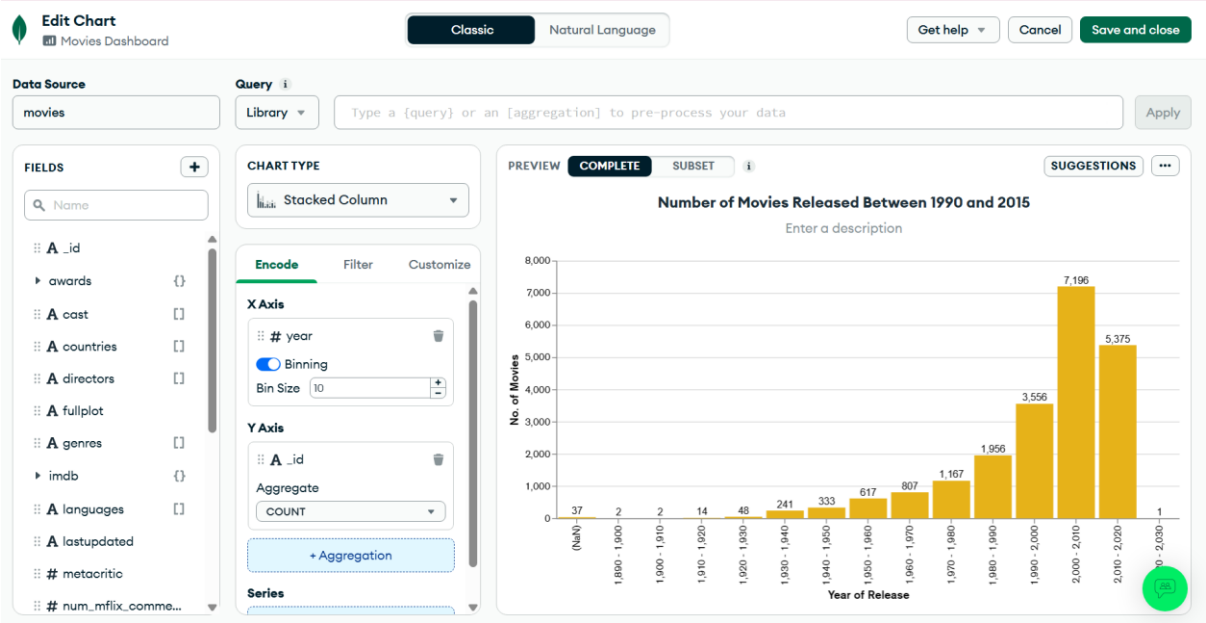


Chart Type Used: Stacked Column Chart

- Represents the count of movies released across different years.
- Uses binning on the X-axis to group data into time intervals (e.g., every 5 years).
- The Y-axis represents the total number of movies released in each bin.

Purpose & Interpretation:

- The trend shows a sharp increase in movie production over time, especially after 2000.
- The highest number of movies was released between 2005 and 2015, indicating an industry boom.
- The dataset suggests that movie production has grown significantly over the years, likely due to advancements in technology, streaming platforms, and global demand.

How It Helps in Analysis:

- Trend Analysis: Shows how movie production evolved over time, which can help in forecasting future trends.
- Industry Growth Insight: The increase in movies post-2000 may indicate a shift in production strategies, audience preferences, or technological improvements.

- Filtering for Deeper Insights: Further breakdown by genre or country could reveal which segments contributed most to this growth.

4. Global Movie Production Distribution

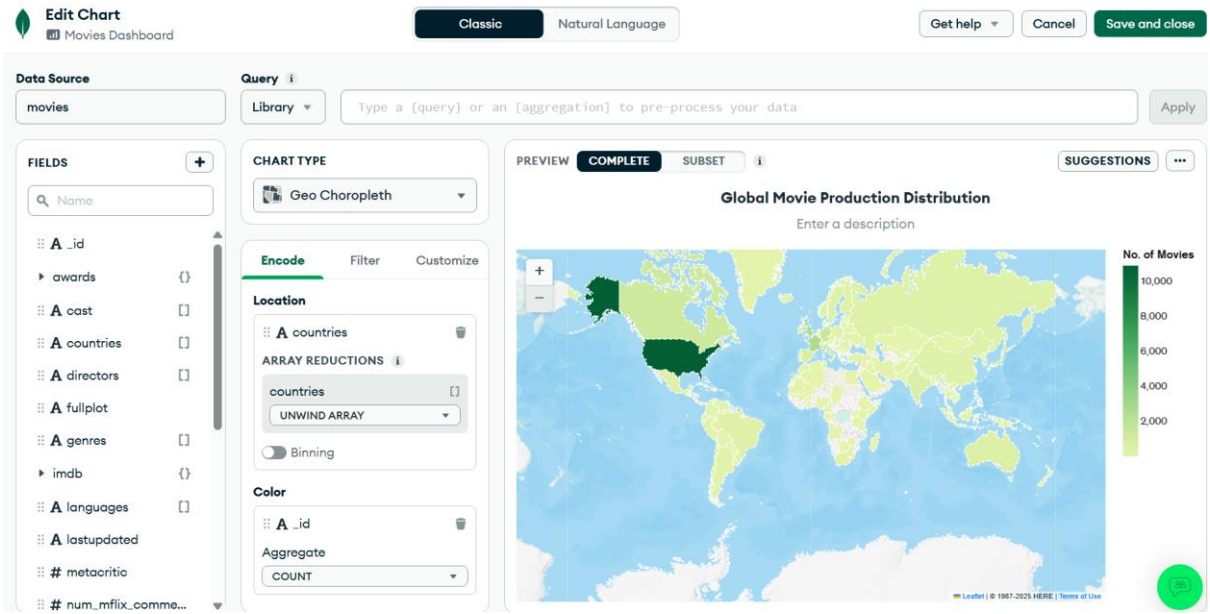


Chart Type Used: Geo Choropleth Map

- Represents the geographic distribution of movies produced across different countries.
- Darker shades indicate higher production volumes, while lighter shades represent fewer movies.

Purpose & Interpretation:

- The map reveals that certain regions dominate movie production, with countries like the USA, India, and China producing the most films.
- Some regions (such as Africa and parts of South America) have significantly fewer movies, likely due to smaller film industries, lower budgets, or fewer production studios.

How It Helps in Analysis:

- Market Insights: Helps identify which countries contribute most to global cinema and where the industry is growing.
- Strategic Decision Making: Useful for businesses looking to expand into new markets or for streaming platforms aiming to source content from high-production regions.

- Comparing Regional Growth: Could be enhanced by overlaying factors like box office revenue, streaming trends, or government film subsidies to understand why certain countries dominate.

5. Movies rated PG, R, PG-13 & APPROVED released before 2015

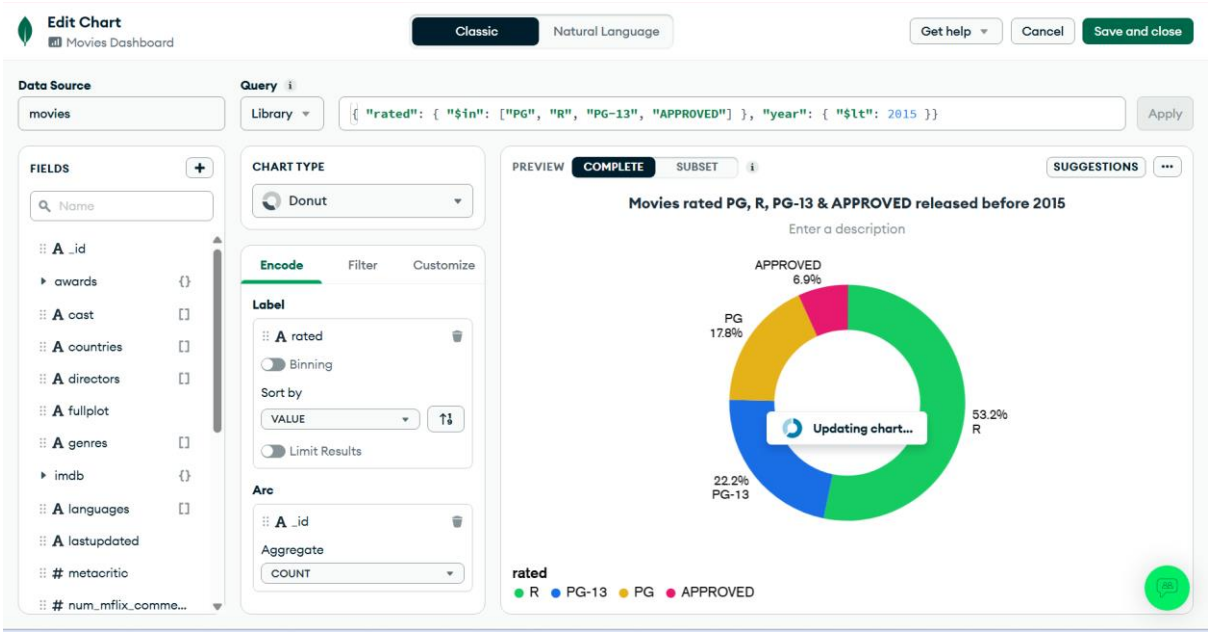


Chart Type Used: Donut Chart

- Represents the proportion of movies with ratings PG, R, PG-13, and APPROVED that were released before 2015.
- The different segments represent the share of each rating in the dataset.

Purpose & Interpretation:

- The majority of movies in this dataset were rated R (Restricted), followed by PG-13.
- PG-rated and APPROVED movies make up a smaller portion of the dataset, suggesting that fewer family-friendly movies were produced compared to R-rated films.
- The dominance of R-rated movies could indicate a preference for mature content in the industry before 2015.

How It Helps in Analysis:

- Content Preferences: Helps understand the rating distribution over time and whether the industry favoured mature audiences.
- Market Strategy: Useful for companies looking to analyse audience preferences and decide which type of content to produce or acquire.

- Further Breakdown: Could be improved by breaking it down by genre (e.g., are R-rated movies mostly action or thriller?).

Query Used:

```
{
  "rated": { "$in": ["PG", "R", "PG-13", "APPROVED"] },
  "year": { "$lt": 2015 }
}
```

Breakdown of the Query Components

1. **"rated": { "$in": ["PG", "R", "PG-13", "APPROVED"] }**
    - Filters movies where the "rated" field matches any of the specified values.
    - The "$in" operator ensures only movies with these ratings are included.
    - Useful for selecting specific classifications while excluding others.
2. **"year": { "$lt": 2015 }**
    - Selects movies released before 2015.
    - The "$lt" (less than) operator ensures only movies from earlier years are included.

This query retrieves movies that were released before 2015 and have a rating of PG, R, PG-13, or APPROVED. It helps analyse older movies, track rating trends over time, or filter content based on historical data.

6. Distribution of IMDb Ratings vs. Runtime for PG and PG-13 Movies (1990-2015)

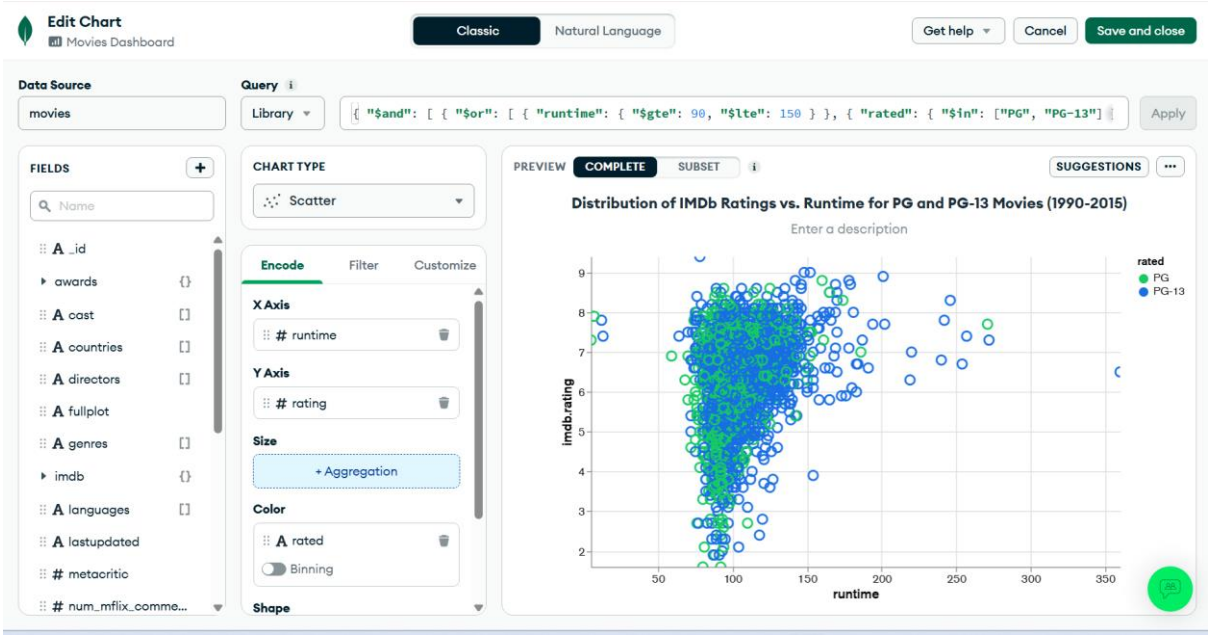


Chart Type Used: Scatter Plot

- Shows the relationship between movie runtime and IMDb ratings for PG and PG-13 movies released between 1990 and 2015.
- Colour coding is used to distinguish rating categories.

Purpose & Interpretation:

- Most movies fall within the runtime range of 90 to 150 minutes, suggesting that PG and PG-13 movies rarely exceed 2.5 hours.
- The cluster of higher IMDb ratings appears in the 90–120-minute range, suggesting that audiences favour movies with moderate runtime.
- There are some outliers with longer runtimes (200+ minutes), but their IMDb ratings do not appear significantly higher.

How It Helps in Analysis:

- Optimal Runtime for Audience Reception: If movies with 90–120-minute runtimes tend to receive higher IMDb ratings, this could indicate an industry standard for successful family-friendly films.
- Content Planning for Filmmakers: Helps studios decide the ideal length for movies targeting PG and PG-13 audiences.
- Further Enhancement: Additional filters could analyse which genres within PG/PG-13 perform best at different runtimes.

Query Used:

```
{
 "$and": [
  { "runtime": { "$gte": 90, "$lte": 150 } },
  { "year": { "$gte": 1990, "$lte": 2015 } },
  { "rated": { "$in": ["PG", "PG-13"] } }
 ]
}
```

Breakdown of the Query Components

1. **"$and": [...]**
    - Ensures that all conditions inside must be met.
2. **"$or": [...]**
    - A movie qualifies if either:
        - Its runtime is between 90 and 150 minutes, or
        - It is rated PG or PG-13.
3. **"year": { "$gte": 1990, "$lte": 2015 }**
    - Filters movies released between 1990 and 2015.
4. **"rated": { "$in": ["PG", "PG-13"] }**
    - Ensures movies are rated PG or PG-13.

Key Observations

- The second rating filter is redundant since movies must already be PG or PG-13.
- The $or condition is unnecessary as the rating filter covers it.

Use Cases

- Film analysis of PG and PG-13 movies over time.
- Filtering mid-length, family-friendly movies for streaming services.
- Market research on runtime trends in movies.

7. Movies released after 2010 with IMDb rating above 8.0 by Genre

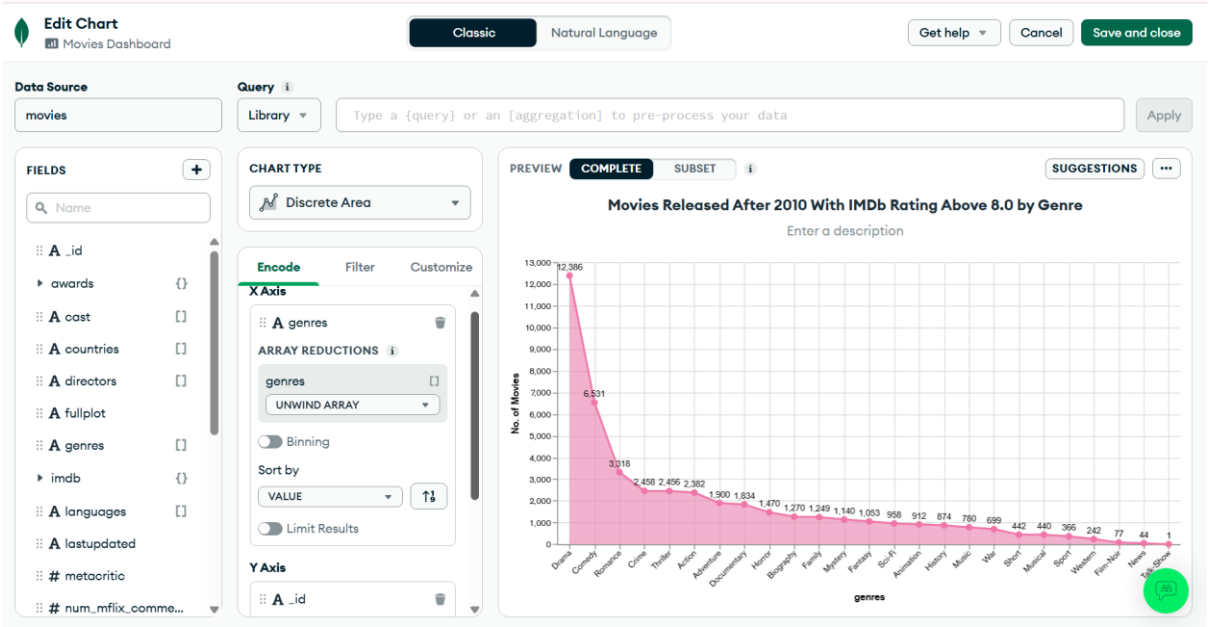


Chart Type Used: Discrete Area Chart

- Shows the count of highly rated movies (IMDb > 8.0) released after 2010, categorized by genre.
- The height of each bar indicates the number of movies per genre that meet the criteria.

Purpose & Interpretation:

- Drastic drop in the number of highly rated movies after the most frequent genre.
- Genres with the highest counts:
  - The leading genre (likely Drama or Action) dominates in producing high-rated films.
  - Other genres, such as Documentary and Biography, might also appear due to their critical appeal.
  - The long tail of less frequent genres suggests that only a few movies from those categories surpass the IMDb 8.0 threshold.

How It Helps in Analysis:

- Industry Trend Analysis: Determines which genres consistently receive high IMDb ratings post-2010.
- Filmmaking & Production Focus: Helps studios identify genres with high audience and critic approval.

- Possible Enhancements:
    - Adding budget or box office data to correlate high ratings with financial success.
    - Analysing directors or production houses responsible for these high-rated films.

8. Movie Genre Distribution Over Time (Post-1980)

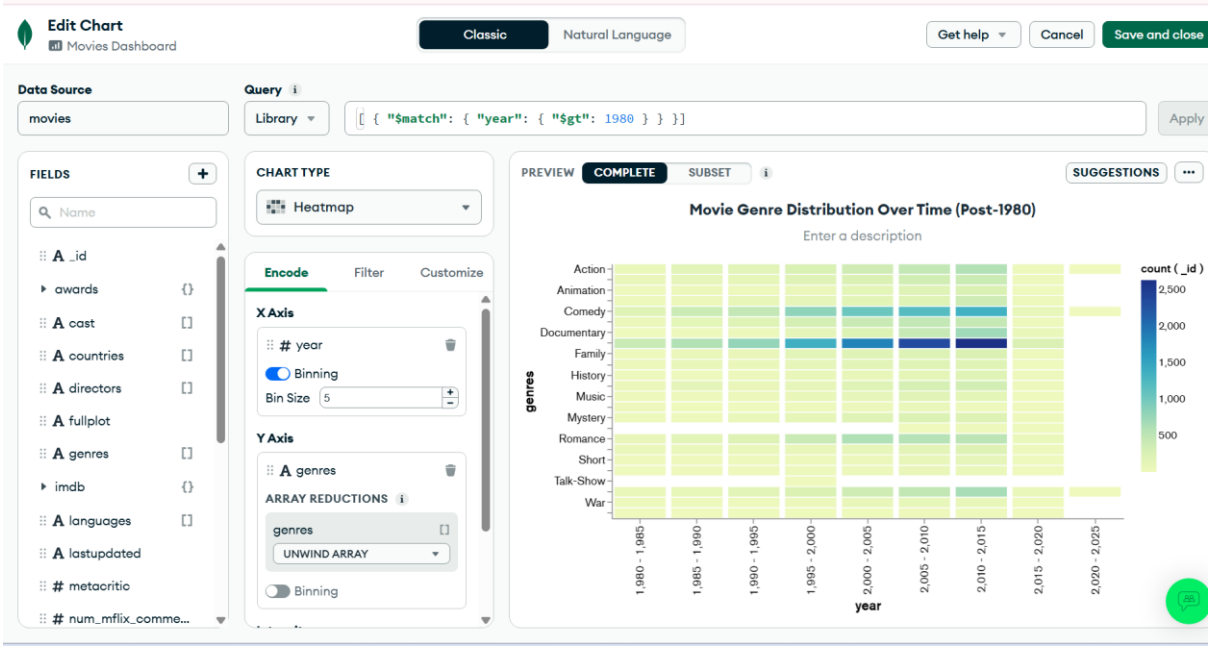


Chart Type Used: Heatmap

- Displays how different movie genres have evolved over time from 1980 onward.
- Darker shades indicate higher concentration of movies in that genre for a given year.

Purpose & Interpretation:

- Some genres (likely Action, Drama, Comedy) appear consistently over the decades, indicating their popularity.
- Others (such as Science Fiction or Fantasy) may show periodic spikes, aligning with trends in audience demand.
- The most recent decades seem more diverse, possibly due to the expansion of streaming services and indie productions.

How It Helps in Analysis:

- Genre Evolution Insight: Identifies trends in which genres have gained or lost traction over time.
- Content Strategy for Streaming Platforms: Helps services like Netflix or Prime curate content based on genre trends.
- Possible Enhancements:
  - A deeper dive into box office revenue per genre over time to correlate with financial success.
  - Comparing the rise of new genres (e.g., superhero movies) and their impact on the industry.

Query Used:

```
[
 {
  "$match": { "year": { "$gt": 1980 } }
 }
]
```

Breakdown of the Query

1. **"$match": { "year": { "$gt": 1980 } }**
   - Filters movies where the year is greater than 1980 (i.e., movies released after 1980).

Use Cases

- Analysing movie trends post-1980.
- Studying the evolution of genres, ratings, or other characteristics in modern cinema.
- Filtering datasets to focus only on relatively recent films.

5. Observations | Findings

Comprehensive Dataset Analysis:

- The dataset's size (21,352 movies) allows for granular analysis of various attributes. This includes examining correlations between runtime and rating, genre distribution over time, and geographical production patterns.
- The inclusion of diverse data points like IMDB ratings, award wins/nominations, and production countries provides a holistic view of movie performance and industry trends.

Critical Acclaim and Award Success:

- The high 84% win-to-nomination ratio suggests a strong correlation between nomination and actual award attainment within this dataset. This could indicate a selection bias towards critically acclaimed films or a high standard of quality within the dataset itself.
- This metric highlights the importance of critical recognition as a potential indicator of a film's overall success or perceived quality.

Temporal Production Trends:

- The significant increase in movie production post-2000, particularly between 2005 and 2015, points to a period of rapid industry expansion. This growth is likely fueled by factors such as:
    - Digital filmmaking technologies reducing production costs.
    - The rise of streaming platforms increasing demand for content.
    - Globalization of film markets.
- The heatmap displaying genre distribution over time also shows that more recent years contain a larger variety of genres, showing that the movie industry is becoming more diverse.

Geographic Production Disparities:

- The Geo Choropleth Map reveals a stark contrast in movie production volumes across different regions. This highlights the dominance of established film industries in countries like the USA, India, and China.
- Conversely, regions with lower production volumes may face challenges such as limited funding, infrastructure, or skilled labour.
- This distribution is also affected by cultural and language barriers.

Audience Rating Preferences and Content Maturity:

- The dominance of R-rated movies, especially before 2015, suggests a historical preference for mature content. This could reflect evolving audience tastes or industry trends towards more adult-oriented storytelling.
- The scatter plot analysis of PG and PG-13 movies indicates that audience preferences for runtime are relatively consistent, with a preference for movies between 90 and 120 minutes.
- The discrete area chart shows that even after 2010, when looking at movies with an IMDB rating of over 8, that there is still a large preference for specific genres.

Genre-Specific Performance and Evolution:

- Certain genres, such as Drama and Action, consistently demonstrate high performance in terms of both production volume and audience ratings.
- The heatmap analysis reveals the ebb and flow of genre popularity over time, with genres like Science Fiction and Fantasy experiencing periodic surges in popularity.
- The heatmap also shows a large increase in the amount of documentary films being created.

6. Managerial Insights | Recommendations

Strategic Content Investment and Development:

- **Genre-Focused Production:** Studios should prioritize production in high-performing genres, leveraging data insights to identify audience preferences and market trends.
- **Runtime Optimization:** Filmmakers should adhere to optimal runtime guidelines for specific rating categories (e.g., 90-120 minutes for PG/PG-13) to maximize audience engagement and satisfaction.

- **Content Diversification:** While mature content has historically been dominant, studios should consider diversifying their portfolios to cater to broader audiences, including family-friendly and international markets.
- **Documentary investment:** With the rise of documentary popularity, streaming services should invest in high quality documentary films.

Global Market Expansion and Distribution:

- **Targeted Market Entry:** Businesses should conduct thorough market research to identify high-potential regions for expansion, considering factors such as audience demographics, cultural preferences, and regulatory environments.
- **Localized Content Strategies:** Streaming platforms and distributors should develop localized content strategies to cater to specific regional audiences, including subtitling, dubbing, and culturally relevant programming.
- **International Co-productions:** Explore international co-production opportunities to leverage local expertise, resources, and distribution networks.

Data-Driven Decision-Making and Analytics:

- **Advanced Analytics Integration:** Implement advanced analytics tools and techniques, such as machine learning and predictive modelling, to gain deeper insights into audience behaviour, market trends, and content performance.
- **Real-Time Dashboard Monitoring:** Establish real-time dashboard monitoring systems to track key performance indicators (KPIs) and identify emerging trends and anomalies.
- **Audience Segmentation and Personalization:** Utilize audience segmentation techniques to personalize content recommendations and marketing campaigns, enhancing user engagement and satisfaction.
- **A/B Testing:** Use A/B testing on movie posters, trailers, and other media to find the most effective advertising.

Critical Acclaim and Award Strategy:

- **Quality-Focused Production:** Prioritize investments in high-quality productions that demonstrate artistic merit and storytelling excellence, increasing the likelihood of critical acclaim and award recognition.
- **Strategic Award Campaigns:** Develop comprehensive award campaign strategies, including targeted marketing, publicity, and outreach to industry professionals and critics.
- **Festival Submissions:** Submit films to prestigious film festivals to gain exposure, critical recognition, and distribution opportunities.

Future Trend Forecasting and Innovation:

    - **Technology Adoption:** Invest in and adopt emerging technologies, such as virtual reality (VR), augmented reality (AR), and artificial intelligence (AI), to enhance storytelling experiences and production workflows.
    - **Content Format Experimentation:** Experiment with new content formats, such as interactive storytelling, short-form video, and podcasting, to cater to evolving audience preferences.
    - **Data Security:** With the increase of data, ensure that all data is kept secure.
- **Content Strategy:** Streaming platforms should focus on acquiring more Drama and Sci-Fi movies, as they tend to have higher ratings.
- **Market Expansion:** Movie production companies should consider expanding operations in countries with emerging film industries.
- **User Engagement:** Platforms can highlight top-rated movies to enhance viewer engagement.
- **Data Enrichment:** More details on box office performance and audience demographics can further improve decision-making.