

Big Data Management and Analysis



Project Report On Data Visualisation using MongoDB Atlas Charts

Submitted to:
Prof (Dr.) Amarnath Mitra

Submitted By:
Group 29
Ashwin Khandelwal (055008)
Soumyadeep Das (055048)

PGDM Big Data Analytics-05 | Section K

1. Project Introduction

The M.Flix Database System is designed to efficiently manage movie-related data, including film details, ratings, genres, and award statistics. To ensure the reliability of the system, the database must maintain data integrity, proper indexing, and performance efficiency under high query loads. This report documents a series of validation and stress tests conducted to verify compliance with data structuring principles, enforce relational integrity where applicable, and evaluate system performance under large-scale data operations. These tests help confirm that the database can handle real-world query demands effectively.

2. Description of Data

2.1. Shape of dataset:

Number of Variables: 135

Number of Observations: 21,349

2.2. Data Structure & Variable Types

2.2.1. Text Variables (Categorical)

2.2.1.1. Movie Metadata:

- a) _id (Unique Movie ID)
- b) plot (Movie Synopsis)
- c) genres[0], genres[1], genres[2] (Movie Genres)
- d) cast[0], cast[1], cast[2], ... (Actors)
- e) writers[0], writers[1], ... (Writers)
- f) directors[0], directors[1], ... (Directors)
- g) language (Primary Language)
- h) country (Production Country)
- i) awards (Awards Won/Nominated)
- j) production (Production Company)

2.2.1.2. Rotten Tomatoes Metadata:

- a) tomatoes.consensus (Critics' Summary)
- b) tomatoes.website (Official Movie Website)
- c) tomatoes.production (Production Studio)

2.2.2. Numerical Variables

2.2.2.1. Movie Metrics:

- a) runtime (Duration in minutes)
- b) metacritic (Metacritic Score)
- c) tomatoes.boxOffice (Box Office Revenue)

2.2.3. Time Variables

- a) tomatoes.dvd (DVD Release Date)
- b) releaseDate (Theatrical Release Date)

3. Problem Objectives

A well-structured movie database must meet the following criteria:

- 3.1. Ensure data consistency through unique identifiers and reference constraints where necessary.
- 3.2. Maintain data accuracy to prevent duplicate or inconsistent records.
- 3.3. Optimize indexing strategies to enhance search efficiency.
- 3.4. Handle concurrent queries efficiently without performance bottlenecks.
- 3.5. Support complex queries, such as filtering by multiple attributes (genres, ratings, release years).
- 3.6. Maintain performance even under high data loads and bulk update operations.
- 3.7. To address these challenges, a series of validation and stress tests were conducted, evaluating the M.Flix database's ability to process large datasets while maintaining accuracy, efficiency, and scalability.

4. Tools Used

The following tools and technologies were used for database testing:

- 4.1. **MongoDB** – NoSQL database system used for managing and querying movie data.
- 4.2. **MongoDB Query Language (MQL)** – Used to perform validation checks, retrieve data, and test indexing strategies.
- 4.3. **MongoDB Compass** – A tool used for analyzing queries and monitoring performance metrics.
- 4.4. **MongoDB Atlas charts** – A graphical tool used for creating charts for data visualisation.
- 4.5. **Command Line Interface (CLI)** – Used to execute bulk data operations and stress testing.

5. Operations Performed

5.1. Import

The screenshot shows the MongoDB Compass interface for a cluster named 'cluster0.2alsq.mongodb.net'. The database is 'sample_mflix' and the collection is 'movies'. A modal window titled 'Import' is open, showing the path 'To collection sample_mflix.movies' and the file 'Import file: insert_movies_compass.txt'. The 'Stop on errors' option is checked. Below the modal, the main interface shows the document count as 21.4K. A message at the bottom left says 'Import completed. 20 documents imported.'

Import file: insert_movies_compass.txt

Options

Stop on errors

Cancel Import

Import completed.

20 documents imported.

more fields

5.2. Retrieve

5.2.1. Retrieval of data for year 2000

```
{  
  "year": {  
    "$gt": 2000  
  }  
}
```

MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies

Connections Edit View Collection Help

Compass

My Queries

cluster0.2alsq.mongodb.net > sample_mflix > movies

Documents 21.4K Aggregations Schema Indexes 2 Validation

Open MongoDB shell

Tell Compass what documents to find (e.g. which movies were released in 2000)

Find Explain Reset Options

ADD DATA EXPORT DATA UPDATE DELETE

100 1-100 of 11988

French
released : 2001-12-25T00:00:00+00:00
directors : Array (1)
0: "James Mangold"
writers : Array (3)
0: "Steven Rogers (story)"
1: "James Mangold (screenplay)"
2: "Steven Rogers (screenplay)"
awards : Object
win : 2
nominations : 4
text : "Nominated for 1 Oscar. Another 1 win & 4 nominations."
lastupdated : "2015-08-31 00:19:09.717000000"
year : 2001
imdb : Object
rating : 6.3
votes : 59951
id : 35423
countries : Array (1)
0: "USA"
type : "movie"
tomatoes : Object
website : "http://www.kateandleopold-themovie.com"

Search

ENG IN 01:48 PM 17/03/2025

The screenshot shows the MongoDB Compass interface. In the top-left, there's a code editor window containing a MongoDB query: `{"year": {"\$gt": 2000}}`. Below it, the main Compass window displays the 'movies' collection from the 'sample_mflix' database. The left sidebar shows the connection 'cluster0.2alsq.mongodb.net' and the 'sample_mflix' database, with the 'movies' collection selected. The main pane shows the results of the query, listing movie documents. One document is expanded to show its full structure, including fields like 'year', 'directors', 'writers', 'awards', and 'imdb'. The bottom status bar shows system information like 'ENG IN' and the date '17/03/2025'.

5.2.2. Retrieval of data for year 1990

```
{  
  "year": {  
    "$lt": 1990  
  }  
}
```

The screenshot shows the MongoDB Compass interface. The top navigation bar includes 'Connections', 'Edit', 'View', 'Collection', and 'Help'. Below the bar, the title is 'MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies'. The main area has tabs for 'Documents' (21.4K), 'Aggregations', 'Schema', 'Indexes' (2), and 'Validation'. A search bar says 'Tell Compass what documents to find (e.g. which movies were released in 2000)'. Below it is a query builder with the following code:

```
{  
  "year": {  
    "$lt": 1990  
  }  
}
```

Buttons for 'EXPLAIN', 'RESET', 'Find', and 'Options' are visible. On the left, the 'Connections' sidebar lists 'cluster0.2alsq.mongodb.net' with databases 'admin', 'config', 'library', 'local', and 'sample_mflix'. Under 'sample_mflix', 'movies' is selected. The main pane displays a document from the 'movies' collection with fields like '_id', 'countries', 'imdb', 'lastupdated', 'rated', 'titles', 'type', 'tomatoes', and 'year'. The bottom status bar shows '100 1-100 of 5187' and a timestamp '01:49 PM 17/03/2025'.

5.2.3. Retrieval of data for runtime 120 minutes

```
{  
  "runtime": {  
    "$gte": 120  
  }  
}
```

The screenshot shows the MongoDB Compass interface. The top navigation bar includes 'MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies', 'Connections', 'Edit', 'View', 'Collection', and 'Help'. Below the navigation is a 'Compass' sidebar with 'My Queries' and a 'CONNECTIONS (1)' section listing 'cluster0.2alsq.mongodb.net' with databases 'admin', 'config', 'library', 'local', 'sample_mflix' (which is expanded to show 'comments', 'embedded_movies', 'movies' (selected), 'sessions', 'theaters', and 'users'), and 'engines'. The main workspace is titled 'cluster0.2alsq.mongodb.net > sample_mflix > movies'. It displays a table with 'Documents' (21.4K), 'Aggregations', 'Schema', 'Indexes' (2), and 'Validation'. A search bar at the top says 'Tell Compass what documents to find (e.g. which movies were released in 2000)'. Below it is a query builder with the following code:

```
{  
  "runtime": {  
    "$gte": 120  
  }  
}
```

Buttons for 'Explain', 'Reset', 'Find' (highlighted in green), and 'Options' are present. At the bottom of the table are buttons for 'ADD DATA', 'EXPORT DATA', 'UPDATE', and 'DELETE'. The status bar at the bottom right shows '100 1–100 of 3651' and various system icons.

5.2.4. Retrieval of data for runtime 90 minutes

```
{  
  "runtime": {  
    "$lte": 90  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the connection tree is visible, showing a cluster named 'cluster0.2alsq.mongodb.net' with databases like 'admin', 'config', 'library', 'local', 'sample_mflix', and 'movies'. The 'sample_mflix' database is selected. In the center, the 'movies' collection is displayed with 21.4K documents. A search bar at the top says 'Tell Compass what documents to find (e.g. which movies were released in 2000)'. Below it, a query builder shows the following code: { "runtime": { "\$lte": 90 } }. The results pane shows a single document from the collection, which includes fields such as _id, plot, runtime, genres, cast, poster, title, fullplot, languages, released, directors, rated, awards, lastupdated, year, imdb, countries, type, tomatoes, and num_mflix_comments. The document's plot field contains a detailed description of the movie 'The Great Train Robbery'. The bottom status bar shows system information like battery level, signal strength, and the date/time (17/03/2025, 01:51 PM).

5.2.5. Retrieval of data for runtime 120 minutes

```
{  
  "year": {  
    "$ne": 2010  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the connection tree shows 'cluster0.2alsq.mongodb.net' with its databases: admin, config, library, local, sample_mflix, comments, embedded_movies, sessions, theaters, and users. The 'sample_mflix' database is selected, and its 'movies' collection is highlighted. In the center, a search bar contains the query: { "year": { "\$ne": 2010 } }. Below the search bar, the results pane displays a single document from the collection. The document details are as follows:

```
_id: ObjectId('573a1390f29313cabcd42e8')  
plot: "A group of bandits stage a brazen train hold-up, only to find a determined..."  
genres: Array (2)  
runtime: 11  
cast: Array (4)  
poster: "https://m.media-amazon.com/images/M/MVSBMTU3NjE5NzYtYTYYNS00MDVmLWIwYj..."  
title: "The Great Train Robbery"  
fullplot: "Among the earliest existing films in American cinema - notable as the ..."  
languages: Array (1)  
released: 1903-12-01T00:00:00+00:00  
directors: Array (1)  
rated: "TV-14"  
awards: Object  
lastupdated: "2015-08-13 00:27:59.177000000"  
year: 1903  
imdb: Object  
countries: Array (1)  
type: "movie"  
tomatoes: Object  
num_mflix_comments: 0
```

The bottom status bar shows system information: ENG IN, 01:51 PM, 17/03/2025.

5.2.6. Retrieval of data for genre action

```
{  
  "$or": [  
    {  
      "genres": "Action"  
    },  
    {  
      "genres": "Comedy"  
    }  
  ]  
}
```

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays connections and the current database and collection (sample_mflix.movies). The main area shows a search bar with the query: { \$or: [{ genres: "Action" }, { genres: "Comedy" }] }. Below the query, a detailed document view is shown for a specific movie entry. The document includes fields like _id, plot, genres (with values Animation, Short, Comedy), runtime, cast (with value Winsor McCay), num_mflix_comments, poster URL, title, and fullplot. The bottom status bar shows system information including battery level, signal strength, and the date/time (01:52 PM, 17/03/2025).

```
_id: ObjectId('573a1390f29313caabcd4803')  
plot: "Cartoon figures announce, via comic strip balloons, that they will mov..."  
genres: Array (3)  
  0: "Animation"  
  1: "Short"  
  2: "Comedy"  
runtime: 7  
cast: Array (1)  
  0: "Winsor McCay"  
num_mflix_comments: 0  
poster: "https://m.media-amazon.com/images/M/MVSBYzg2NjNHNTctHjUxM100ZWU4LWI3Zj..."  
title: "Winsor McCay, the Famous Cartoonist of the N.Y. Herald and His Moving..."  
fullplot: "Cartoonist Winsor McCay agrees to create a large set of drawings that..."  
languages: Array (1)  
  0: "English"  
released: 1911-04-08T00:00:00.000+00:00  
directors: Array (2)  
  0: "Winsor McCay"  
  1: "J. Stuart Blackton"
```

5.2.7. Retrieval of data for duration between year 1995 and 2010

```
{  
  "$and": [  
    {  
      "year": {  
        "$gt": 1995  
      }  
    },  
    {  
      "year": {  
        "$lt": 2010  
      }  
    }  
  ]  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays the database connections and collections, with the 'sample_mflix' database and 'movies' collection selected. The main pane shows the 'Find' interface where the previously defined query is entered:

```
{  
  "$and": [  
    {  
      "year": {  
        "$gt": 1995  
      }  
    },  
    {  
      "year": {  
        "$lt": 2010  
      }  
    }  
  ]  
}
```

Below the query, the results are displayed as a list of movie documents. One document is shown in full:

```
 1: "French"  
  released : 2001-12-25T00:00:00.000+00:00  
  directors : Array (1)  
   0: "James Mangold"  
  writers : Array (3)  
   0: "Steven Rogers (story)"  
   1: "James Mangold (screenplay)"  
   2: "Steven Rogers (screenplay)"  
  awards : Object  
   wins : 2  
   nominations : 4  
   text : "Nominated for 1 Oscar. Another 1 win & 4 nominations."  
  lastupdated : "2015-08-31 00:19:09.717000000"  
  year : 2001  
  imdb : Object  
   rating : 6.3  
   votes : 59951  
   id : 35423  
  countries : Array (1)  
   0: "USA"
```

The bottom status bar shows system information like battery level, signal strength, and the date/time.

5.2.8. Retrieval of data where Actors are Tom Hanks and Robert Downey Jr.

```
{  
    "cast": {  
        "$in": [  
            "Tom Hanks",  
            "Robert Downey Jr."  
        ]  
    }  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays connections and databases, with 'sample_mflix' selected. The main area shows the 'movies' collection with 21.4K documents. A search bar at the top asks 'Tell Compass what documents to find (e.g. which movies were released in 2000)'. Below it, a query builder window contains the following code:

```
{  
    "cast": {  
        "$in": [  
            "Tom Hanks",  
            "Robert Downey Jr."  
        ]  
    }  
}
```

Below the query, the results list the first few documents, all of which star Tom Hanks and Robert Downey Jr. The results table includes columns for '_id', 'title', 'poster', 'fullplot', 'languages', 'released', 'directors', 'writers', and 'awards'. The bottom right corner of the screen shows the Windows taskbar with various icons and the date/time: 17/03/2025, 01:54 PM.

5.2.9. Retrieval of data where information about the director exists

```
{  
    "directors": {  
        "$exists": true  
    }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the 'Connections' sidebar lists 'cluster0.2alsq.mongodb.net' with databases 'admin', 'config', 'library', 'local', and 'sample_mflix'. The 'sample_mflix' database is selected, and its 'movies' collection is highlighted. The main pane displays the results of a query. The query stage is:

```
{  
    "directors": {  
        "$exists": true  
    }  
}
```

The results show a single document from the 'movies' collection:

```
2: "George Barnes"  
3: "Justus D. Barnes"  
poster: "https://m.media-amazon.com/images/M/MVSBMTU3NjE5NzYtYTtyNS00MDVmLwIwYj..."  
title: "The Great Train Robbery"  
fullplot: "Among the earliest existing films in American cinema - notable as the ..."  
languages: Array (1)  
0: "English"  
released: 1903-12-01T00:00:00.000Z  
directors: Array (1)  
0: "Edwin S. Porter"  
rated: "TV-G"  
awards: Object  
wins: 1  
nominations: 0  
text: "1 win."  
lastupdated: "2015-08-13 00:27:59.177000000"  
year: 1903  
imdb: Object  
rating: 8.5  
votes: 9847  
id: 436  
countries: Array (1)  
0: "USA"  
+more...  
+less...
```

Below the results, there are buttons for 'ADD DATA', 'EXPORT DATA', 'UPDATE', and 'DELETE'. The status bar at the bottom right shows 'ENG IN' and the date '17/03/2025'.

5.2.10. Retrieval of data where information about the director does not exist

```
{  
  "directors": {  
    "$exists": false  
  }  
}
```

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar lists connections and databases, with 'sample_mflix' selected. The main area displays the results of an aggregation query on the 'movies' collection. The query is:

```
{  
  "directors": {  
    "$exists": false  
  }  
}
```

The results show a single document from the 'movies' collection with the following fields and values:

- `_id`: ObjectId('573a1395f29313caabce2f03')
- `plot`: "The extended Forsyte family live a more than pleasant upper middle class life."
- `genres`: ["Drama"]
- `runtime`: 50
- `cast`: ["Eric Porter", "Margaret Tyzack", "Nyree Dawn Porter", "June Barry"]
- `num_mflix_comments`: 0
- `poster`: "https://m.media-amazon.com/images/M/MVSBMTQzNjgzODI2MF5BM15BanBnXkFtZT..."
- `title`: "The Forsyte Saga"
- `fullplot`: "The extended Forsyte family live a more than pleasant upper middle class life."
- `languages`: ["English"]
- `released`: 1969-10-05T00:00:00+00:00
- `awards`: Object with fields: wins: 4, nominations: 2, text: "Won 1 Primetime Emmy. Another 3 wins & 2 nominations."
- `lastupdated`: "2015-09-05 00:46:25.133000000"
- `year`: 1967

5.2.11. Retrieval of data where genre size is 3.

```
{  
  "genres": {  
    "$size": 3  
  }  
}
```

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays database connections and collections, with 'sample_mflix' selected. The main area shows a search bar with the query: { "genres": { "\$size": 3 } }. Below the query, a document preview is shown with fields like _id, plot, genres (containing Animation, Short, Comedy), runtime, cast, num_mflix_comments, poster, title, fullplot, languages, released, directors, writers, and awards. The bottom status bar shows system icons and the date/time: 17/03/2025, 01:56 PM, ENG IN.

5.2.12.

```
{  
  "plot": {  
    "$regex": "love",  
    "$options": "i"  
  }  
}
```

MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies

Connections Edit View Collection Help

Compass

My Queries

Connections (1)

- cluster0.2alsq.mongodb.net
 - admin
 - config
 - library
 - local
 - sample_mflix
 - comments
 - embedded_movies
 - movies**
 - sessions
 - theaters
 - users

cluster0.2alsq.mongodb.net > sample_mflix > movies

Documents 21.4K Aggregations Schema Indexes 2 Validation

Tell Compass what documents to find (e.g. which movies were released in 2000)

Find Explain Reset Options ▾

```
{  
  "plot": {  
    "$regex": "love",  
    "$options": "i"  
  }  
}
```

ADD DATA EXPORT DATA UPDATE DELETE

100 1 - 100 of 1703

_id: ObjectId('573a1391f29313caabcd7a34')
plot: "A kept woman runs into her one-time fianc  and finds herself torn betw..."
genres: Array (2)
0: "Drama"
1: "Romance"
runtime: 78
rated: "TV-PG"
cast: Array (4)
0: "Edna Purviance"
1: "Clarence Geldart"
2: "Carl Miller"
3: "Lydia Knott"
num_mflix_comments: 1
poster: "https://m.media-amazon.com/images/M/MVSBZjJiMTU2NGQtNWRkNi00ZjExLWEzMT..."
title: "A Woman of Paris: A Drama of Fate"
fullplot: "Marie St. Clair believes she has been jilted by her artist fiance Jean..."
countries: Array (1)
0: "USA"
released: 1923-11-04T00:00:00.000+00:00
directors: Array (1)
0: "Charles Chaplin"
writers: Array (1)

Search

01:56 PM 17/03/2025

5.2.13.

```
{  
  "year": {  
    "$gte": 1980,  
    "$lte": 1990  
  }  
}
```

The screenshot shows the MongoDB Compass interface. The top navigation bar includes 'Connections', 'Edit', 'View', 'Collection', and 'Help'. Below the bar, the title 'MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies' is displayed. The main area has tabs for 'Documents' (21.4K), 'Aggregations', 'Schema', 'Indexes', and 'Validation'. The 'Documents' tab is active. A search bar at the top says 'Tell Compass what documents to find (e.g. which movies were released in 2000)'. Below it, a query builder shows the following code:

```
i {  
  "year": {  
    "$gte": 1980,  
    "$lte": 1990  
  }  
}
```

Below the query, the results pane displays a single document from the 'movies' collection. The document contains fields like '_id', 'plot', 'genres', 'runtime', 'cast', 'fullplot', 'title', 'num_mflix_comments', 'poster', 'released', 'directors', 'writers', and 'languages'. The 'Languages' field is expanded to show 'Mandarin' and 'Cantonese'. The bottom right corner of the screen shows system status: ENG IN, 01:57 PM, and 17/03/2025.

5.2.14.

```
{  
  "awards.wins": {  
    "$gt": 5  
  }  
}
```

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays the database structure with the 'movies' collection selected. The main pane shows a query builder with the following query:

```
{  
  "awards.wins": {  
    "$gt": 5  
  }  
}
```

Below the query, the results of the search are displayed, showing a single document from the 'movies' collection. The document contains fields such as _id, plot, genres, runtime, rated, cast, num_mflix_comments, title, fullplot, languages, and released. The document ID is ObjectId('573a1391f29313caabcd93aa').

5.2.15.

```
{  
  "$or": [  
    {  
      "directors": "Steven Spielberg"  
    },  
    {  
      "directors": "Christopher Nolan"  
    }  
  ]  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays connections and collections, with 'sample_mflix' selected. The main area shows a search bar with the query: { \$or: [{ directors: "Steven Spielberg" }, { directors: "Christopher Nolan" }] }. Below the search bar, the results list shows one document from the 'movies' collection. The document details are as follows:

```
_id: ObjectId('573a1396f29313caabce449a')  
plot: "A business commuter is pursued and terrorized by a malevolent driver o..."  
genres: Array (1)  
  0: "Thriller"  
runtime: 90  
rated: "PG"  
cast: Array (4)  
  0: "Dennis Weaver"  
  1: "Jacqueline Scott"  
  2: "Eddie Firestone"  
  3: "Lou Frizzell"  
num_mflix_comments: 0  
poster: "https://m.media-amazon.com/images/M/MVSBNzRkNjE4NjQtZTQ2NC0BYTg1LTh1ZT..."  
title: "Due"  
fullplot: "While traveling through the desert for an appointment with a client, t..."  
languages: Array (1)  
  0: "English"  
released: 1971-11-13T00:00:00.000+00:00  
directors: Array (1)
```

The status bar at the bottom right indicates the system is in English (ENG), connected to the internet (IN), and the date is 17/03/2025.

5.2.16.

```
{  
  "runtime": {  
    "$ne": null  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the 'Connections' sidebar lists 'cluster0.2alsq.mongodb.net' with its databases: admin, config, library, local, sample_mflix, comments, embedded_movies, movies, sessions, theaters, and users. The 'movies' database is selected. In the main pane, the 'Documents' tab shows 21.4K documents. A search bar at the top says 'Tell Compass what documents to find (e.g. which movies were released in 2000)'. Below it, a query builder shows the following search query:

```
{  
  "runtime": {  
    "$ne": null  
  }  
}
```

The results pane displays a single document from the 'movies' collection. The document details are as follows:

```
_id: ObjectId("573a1390f29313caabcd42e8")  
plot: "A group of bandits stage a brazen train hold-up, only to find a determined  
 0: "short"  
 1: "Western"  
runtime: 111  
cast: Array (4)  
 0: "A.C. Abbie"  
 1: "Gilbert M. 'Broncho Billy' Anderson"  
 2: "George Barnes"  
 3: "Justin D. Barnes"  
poster: "https://m.media-amazon.com/images/M/MVSBHTU3NjE5NzYtYTYYNS00MDVmLWIwYj_"  
title: "The Great Train Robbery"  
fullplot: "Among the earliest existing films in American cinema - notable as the ..."  
languages: Array (1)  
 0: "English"  
released: 1903-12-01T00:00:00.000+00:00  
directors: Array (1)  
 0: "Edwin S. Porter"  
rated: "TV-G"  
awards: Object  
  wins: 1  
  nominations: 0
```

At the bottom of the screen, the taskbar shows various application icons, and the system tray indicates the date and time as 17/03/2025, 01:58 PM.

5.2.17.

```
{  
  "tomatoes.viewer.rating": {  
    "$lt": 6.0  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the connection tree displays 'cluster0.2alsq.mongodb.net' with databases 'admin', 'config', 'library', 'local', and 'sample_mflix'. The 'sample_mflix' database is selected, and its collections 'comments', 'embedded_movies', 'movies', 'sessions', 'theaters', and 'users' are listed. The 'movies' collection is currently selected. The main pane shows a search query in the 'Find' field:

```
{  
  "tomatoes.viewer.rating": {  
    "$lt": 6.0  
  }  
}
```

Below the query, the results table shows a single document with the following fields and values:

- `_id`: `ObjectId('573a1390f29313caabcd42e8')`
- `plot`: "A group of bandits stage a brazen train hold-up, only to find a determined lawman who won't let them get away with it."
- `genres`: Array [2]
 - 0: "Short"
 - 1: "Western"
- `runtime`: 11
- `cast`: Array [4]
 - 0: "A.C. Abbie"
 - 1: "Gilbert M. 'Broncho Billy' Anderson"
 - 2: "George Barnes"
 - 3: "Justin D. Barnes"
- `poster`: "https://m.media-amazon.com/images/M/MVSBHTU3NjE5NzYtYTYYNS00MDVmLWIwYj_"
- `title`: "The Great Train Robbery"
- `fullplot`: "Among the earliest existing films in American cinema - notable as the ..."
- `languages`: Array [1]
 - 0: "English"
- `released`: `1903-12-01T00:00:00.000+00:00`
- `directors`: Array [1]
 - 0: "Edwin S. Porter"
- `rated`: "TV-G"
- `awards`: Object
 - `wins`: 1
 - `nominations`: 0

5.2.18.

```
{  
  "$and": [  
    {  
      "year": {  
        "$gte": 2000  
      }  
    },  
    {  
      "genres": "Sci-Fi"  
    }  
  ]  
}
```

The screenshot shows the MongoDB Compass interface. The top bar displays the title "MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies". The left sidebar shows connections and collections, with "sample_mflix" selected. The main area shows the "movies" collection with 21.4K documents. A search bar at the top says "Tell Compass what documents to find (e.g. which movies were released in 2000)". Below it is the query builder with the "\$and" query. The results pane shows a single document extracted from the collection, including fields like _id, fullplot, imdb, year, genres, plot, and languages. The bottom status bar shows system information like battery level, signal strength, and date/time.

```
cluster0.2alsq.mongodb.net > sample_mflix > movies
```

Documents 21.4K Aggregations Schema Indexes 2 Validation

Search connections

cluster0.2alsq.mongodb.net

- admin
- config
- library
- local
- sample_mflix
 - comments
 - embedded_movies
 - movies**
 - sessions
 - theaters
 - users

ADD DATA EXPORT DATA UPDATE DELETE

`_id: ObjectId('573a139af29313caabcf0e9d')
fullplot: "Reed Richards, a brilliant but timid and bankrupt scientist, is convin..."
imdb: Object
 rating: 5.7
 votes: 259168
 id: 128667
 year: 2005
 plot: "A group of astronauts gain superpowers after a cosmic radiation exposu..."
 genres: Array (3)
 0: "Action"
 1: "Adventure"
 2: "Sci-Fi"
 rated: "PG-13"
 metacritic: 40
 title: "Fantastic Four"
 lastupdated: "2015-09-16 13:02:16.957000000"
 languages: Array (1)
 0: "English"
 writers: Array (4)`

5.2.19.

```
{  
  "countries": {  
    "$nin": [  
      "USA"  
    ]  
  }  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays connections and collections, with 'sample_mflix' selected. The main area shows a query in the search bar:

```
{
  "countries": {
    "$nin": [
      "USA"
    ]
}
```

The results pane shows a list of movie documents. One document is expanded to show its full structure:

```
_id: ObjectId('573a1391f29313caabcd7db6')
plot: "In Paris a wild girl becomes possessed by the soul of her twin who die..."
genres: Array (1)
  0: "Drama"
runtime: 82
cast: Array (4)
  0: "Betty Compson"
  1: "Clive Brook"
  2: "Henry Victor"
  3: "A.B. Imeson"
title: "White Shadows"
fullplot: "In Paris a wild girl becomes possessed by the soul of her twin who die..."
languages: Array (1)
  0: "English"
released: 1924-05-05T00:00:00.000+00:00
directors: Array (1)
  0: "Graham Cutts"
writers: Array (2)
  0: "Alfred Hitchcock"
  1: "Michael Norton (novel)"
awards: Object
```

The status bar at the bottom indicates the system is ENG IN, the time is 02:00 PM, and the date is 17/03/2025.

5.2.20.

```
{  
  "languages": {  
    "$all": [  
      "English",  
      "French"  
    ]  
  }  
}
```

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays connections and collections, with 'sample_mflix' selected. The main pane shows the 'movies' collection, which contains 21.4K documents. A specific document is expanded to show its fields:

```
_id: ObjectId('573a1391f29313caabed9264')  
plot: "When a woman discovers that her husband has been unfaithful to her, sh..."  
genres: Array (2)  
  0: "Romance"  
  1: "Drama"  
runtime: 84  
rated: "PASSED"  
cast: Array (4)  
  0: "Norma Shearer"  
  1: "Chester Morris"  
  2: "Conrad Nagel"  
  3: "Robert Montgomery"  
num_mflix_comments: 0  
poster: "https://m.media-amazon.com/images/M/MVSBMtk5NzQzNjEwNV58M15BnXkFtZT..."  
title: "The Divorcee"  
fullplot: "Jerry and Ted are young, in love, and part of the New York 'in-crowd'..."  
languages: Array (2)  
  0: "English"  
  1: "French"  
released: 1930-04-19T00:00:00.000+00:00
```

5.3. Update

```
{  
  "filter": {  
    "runtime": {  
      "$ne": null  
    }  
  },  
  "update": {  
    "$addToSet": {  
      "awards.wins": 1  
    }  
  },  
  "multi": true  
}
```

5.3.1.

The screenshot shows the MongoDB Compass interface. The top bar indicates the connection is to 'MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies'. The left sidebar shows the database structure with 'sample_mflix' selected, containing 'movies', 'sessions', 'theaters', and 'users' collections. The main area is titled 'Update 20930 documents' for the 'movies' collection. It displays a 'Filter' field with the query '{ runtime: { \$ne: null } }'. Below it is a 'Update' field containing the provided code: '1 { 2 "\$addToSet": { 3 "awards.wins": 2 4 } 5 }'. To the right, a 'Preview' section shows an example document with fields like _id, plot, awards, and wins. At the bottom are 'Save', 'Cancel', and a large green 'Update 20930 documents' button. The status bar at the bottom right shows the date and time: '17/03/2025 02:05 PM'.

5.3.2.

```
{  
  "filter": {  
    "year": {  
      "$gte": 1995,  
      "$lt": 2010  
    }  
  },  
  "update": {  
    "$pull": {  
      "genres": "Fantasy"  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays the database structure with 'sample_mflix' selected. The main area shows a modal dialog titled 'Update 9442 documents'. Inside the dialog, there is a 'Filter' section with the query: { year: { \$gte: 1995, \$lt: 2010 } }. Below it is a 'Update' section containing the update command: 1. { 2. '\$pull': { 3. 'genres': 'Fantasy' 4. } 5. }. To the right, a 'Preview' section shows a sample of three documents. One document is fully visible:

```
_id: ObjectId('573a1393f29313caabcdcb42')  
plot: "Kate and her actor brother live in N.Y. in the 21st Century. Her ex-bo..."  
genres: Array  
  [ "Comedy", "Fantasy", "Romance" ]  
runtime: 118  
metacritic: 44  
rated: "PG-13"  
cast: Array  
poster: "https://m.media-amazon.com/images/M/MV5BNmNlN2ViOTctYTdHS00NzUxLTg0ZG..."  
title: "Kate & Leopold"  
fullplot: "Kate and her actor brother live in N.Y. in the 21st Century. Her ex-bo..."  
languages: Array  
released: 2001-12-25T00:00:00.000+0000  
directors: Array  
writers: Array  
awards: Object  
lastupdated: "2015-08-31 00:19:09.717000000"  
year: 2001  
imdb: Object  
countries: Array
```

At the bottom of the dialog, there are 'Save' and 'Cancel' buttons, and a large green 'Update 9442 documents' button. The status bar at the bottom indicates the system is in English (ENG IN), the time is 02:07 PM, and the date is 17/03/2025.

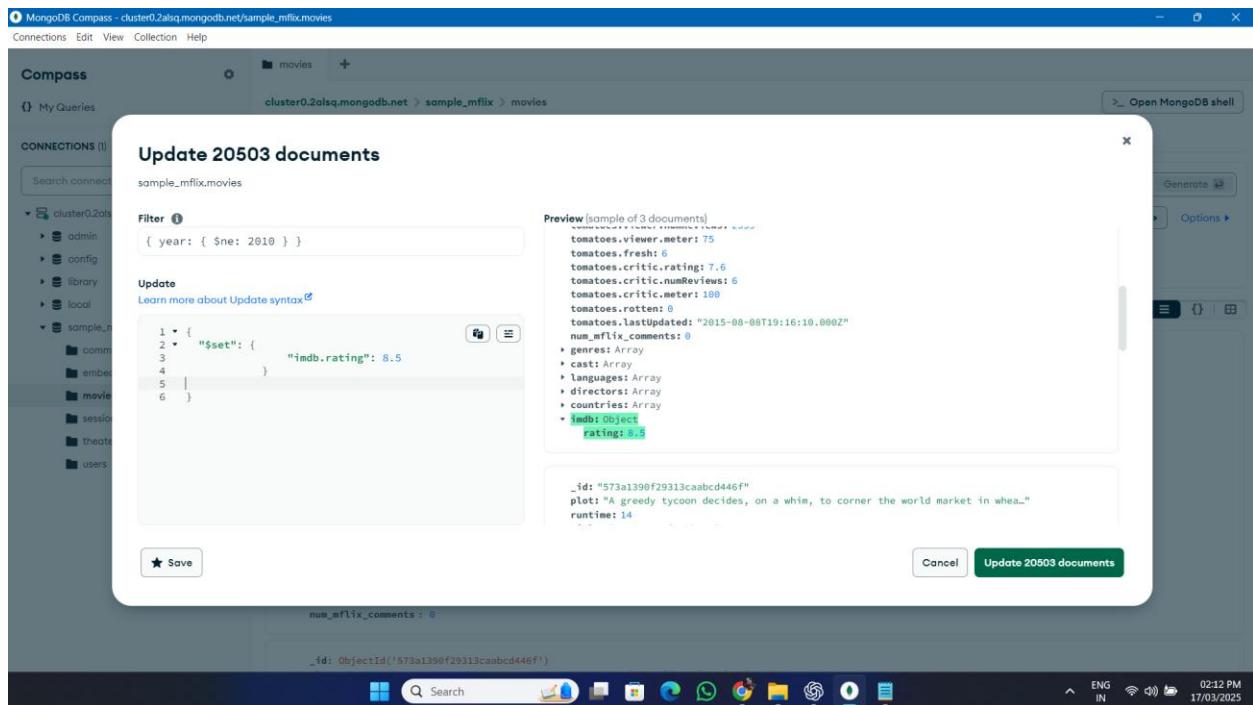
5.3.3.

```
{  
  "filter": {  
    "tomatoes.viewer.rating": {  
      "$lt": 6.0  
    }  
  },  
  "update": {  
    "$set": {  
      "imdb.rating": 8.5  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The top navigation bar includes 'Connections', 'Edit', 'View', 'Collection', and 'Help'. The main window displays a 'Compass' sidebar with 'My Queries' and a 'sample_mflix' connection. Under 'sample_mflix', there are collections: 'admin', 'config', 'library', 'local', and 'sample_mflix' which contains 'comments', 'embeds', 'movies', 'sessions', 'theatre', and 'users'. A modal dialog titled 'Update 1856 documents' is open, showing a 'Filter' field with the query: '{ "tomatoes.viewer.rating": { \$lt: 6 } }'. Below it is an 'Update' section with the command: '1. {
 2. "\$set": {
 3. "imdb.rating": 8.5
 4. }
 5. }'. To the right, a 'Preview' section shows a sample of three documents, including one for 'Winsor McCay'. At the bottom of the dialog are 'Save', 'Cancel', and 'Update 1856 documents' buttons. The status bar at the bottom shows system icons and the date/time: '17/03/2025 02:10 PM'.

5.3.4.

```
{  
  "filter": {  
    "year": {  
      "$ne": 2010  
    }  
  },  
  "update": {  
    "$set": {  
      "imdb.rating": 8.5  
    }  
  },  
  "multi": true  
}
```



5.3.5.

```
{  
  "filter": {  
    "year": {  
      "$ne": 2010  
    }  
  },  
  "update": {  
    "$set": {  
      "imdb.rating": 8.5  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The top bar indicates the connection is to 'cluster0.2alsq.mongodb.net' and the collection is 'sample_mflix.movies'. The main area displays an 'Update 10428 documents' dialog. The 'Filter' section contains the query: '{ countries: { \$nin: ['USA'] } }'. The 'Update' section shows the command: '1 { 2 \$pull: { 3 genres: "Drama" 4 } 5 } 6'. A preview window shows a sample document with fields like _id, plot, genres, runtime, cast, poster, title, fullplot, countries, released, directors, writers, awards, lastupdated, and year. The bottom right of the dialog has 'Cancel' and 'Update 10428 documents' buttons. The bottom of the screen shows a Windows taskbar with various icons and the date/time '17/03/2025 02:15 PM'.

5.3.6.

```
{  
  "filter": {  
    "countries": {  
      "$nin": [  
        "USA"  
      ]  
    },  
    "update": {  
      "$set": {  
        "updated_at": "2025-03-16"  
      }  
    },  
    "multi": true  
  }  
}
```

The screenshot shows the MongoDB Compass interface with the following details:

- Header:** MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies
- Navigation:** Connections, Edit, View, Collection, Help
- Collection:** movies
- Preview:** Shows a sample of 3 documents with fields like directors, writers, awards, lastupdated, year, imdb, countries, type, tomatoes, num_mflix_comments, and updated_at.
- Update Dialog:**
 - Filter:** { countries: { \$nin: ['USA'] } }
 - Update:** {
 "\$set": {
 "updated_at": "2025-03-16"
 }
}
This indicates an update operation where all documents without 'USA' in their countries array will have their updated_at field set to '2025-03-16'.
 - Buttons:** Save, Cancel, Update 10428 documents
- Bottom Status:** num_mflix_comments : 0
- System Bar:** _id: ObjectId('573a1391f29313caabcd883d'), plot: "A film about the French Field Marshal's youth and early military career.", genres: Array, runtime: 240, cast: Array, num_mflix_comments: 0, poster: "https://m.media-amazon.com/images/M/MV5BOGY4M2ZlMGQtMzM4ZC08YjhLTkxOG...".
- System Icons:** ENG IN, 02:16 PM, 17/03/2025

5.3.7.

```
{  
  "filter": {  
    "directors": {  
      "$exists": true  
    }  
  },  
  "update": {  
    "$set": {  
      "imdb.rating": 8.5  
    }  
  },  
  "multi": true  
}
```

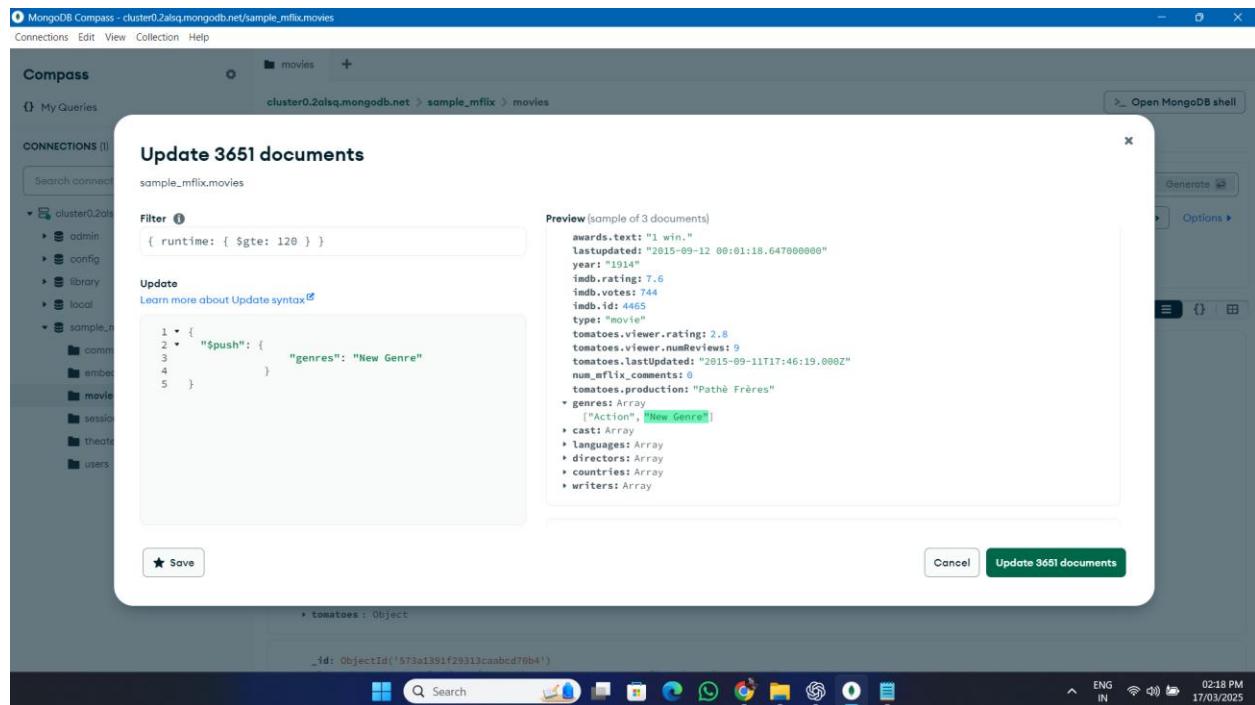
The screenshot shows the MongoDB Compass interface. The top bar indicates the connection is to 'cluster0.2alsq.mongodb.net' and the database is 'sample_mflix'. A modal window titled 'Update 21127 documents' is open, showing the following update command:

```
{  
  "filter": {  
    "directors": {  
      "$exists": true  
    }  
  },  
  "update": {  
    "$set": {  
      "imdb.rating": 8.5  
    }  
  },  
  "multi": true  
}
```

The 'Preview' section shows a sample of 3 documents from the 'movies' collection, which includes fields like _id, plot, runtime, and various rating metrics. At the bottom of the modal, there are 'Save' and 'Update 21127 documents' buttons. The bottom right corner of the screen shows the Windows taskbar with the date and time (02:17 PM, 17/03/2025).

5.3.8.

```
{  
  "filter": {  
    "runtime": {  
      "$gte": 120  
    }  
  },  
  "update": {  
    "$push": {  
      "genres": "New Genre"  
    }  
  },  
  "multi": true  
}
```



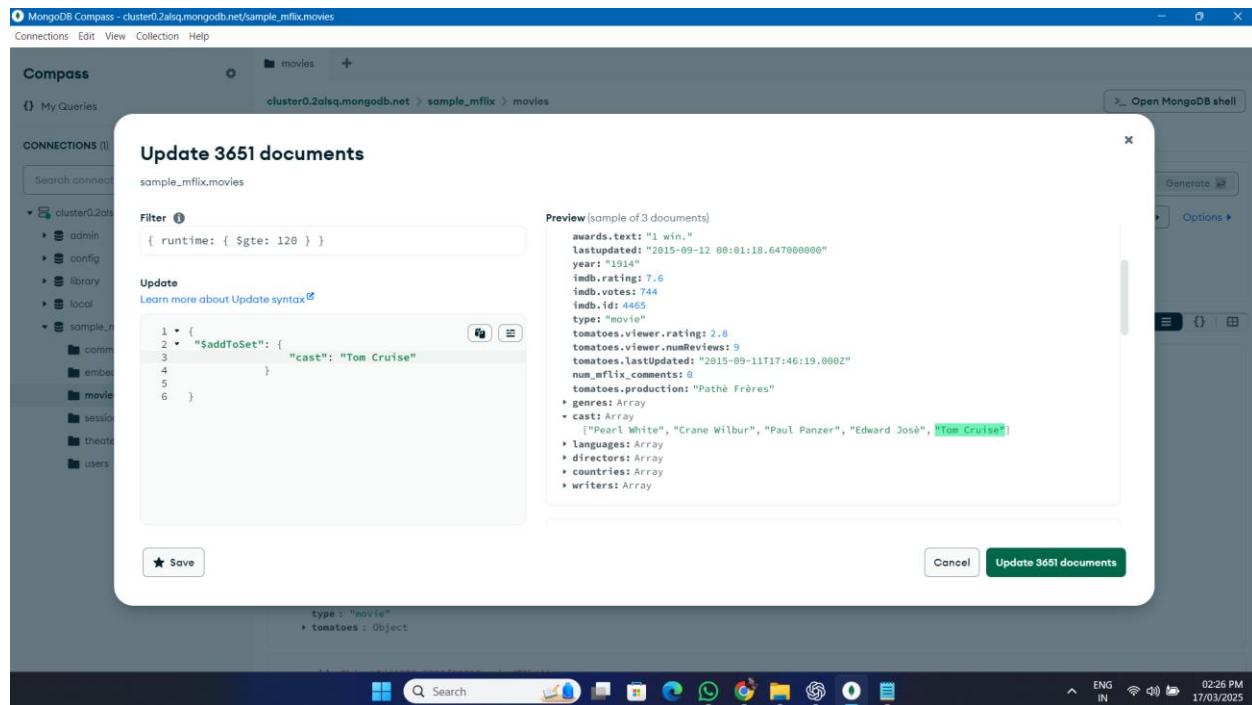
5.3.9.

```
{  
  "filter": {  
    "countries": {  
      "$nin": [  
        "USA"  
      ]  
    }  
  },  
  "update": {  
    "$push": {  
      "genres": "New Genre"  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays connections and databases, with 'sample_mflix' selected. The main area shows a modal dialog titled 'Update 10428 documents' for the 'sample_mflix.movies' collection. The 'Filter' field contains the query: { countries: { \$nin: ['USA'] } }. The 'Update' field contains the update command: { "\$push": { "genres": "New Genre" } }. A preview window on the right shows sample document data, including fields like _id, plot, genres, runtime, cast, title, languages, released, awards, lastupdated, year, imdb, countries, type, tomatoes, and num_mflix_comments. At the bottom of the modal are 'Save' and 'Update 10428 documents' buttons. The status bar at the bottom of the screen shows system information: ENG IN, 02:19 PM, 17/03/2025.

5.3.10.

```
{  
  "filter": {  
    "runtime": {  
      "$gte": 120  
    }  
  },  
  "update": {  
    "$addToSet": {  
      "cast": "Tom Cruise"  
    }  
  },  
  "multi": true  
}
```



5.3.11.

```
{  
  "filter": {  
    "$and": [  
      {  
        "year": {  
          "$gte": 2000  
        }  
      },  
      {  
        "genres": "Sci-Fi"  
      }  
    ]  
  },  
  "update": {  
    "$set": {  
      "updated_at": "2025-03-16"  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays a connection named 'cluster0.2alsq.mongodb.net' with a single database 'sample_mflix'. The 'movies' collection is selected. A modal window titled 'Update 527 documents' is open, containing the following update query:

```
1 - {  
2   "$set": {  
3     "updated_at": "2025-03-16"  
4   }  
5 }
```

The 'Preview' section shows a sample of three documents. One document is expanded to show its full structure:

_id: ObjectId('573a139af29313caabc0ff51')
fullplot: "In a world where both Mutants and Humans fear each other, Marie, bette..."
awards: Object
countries: Array
cast: Array
directors: Array
runtimes: 108
updated_at: "2025-03-16"

At the bottom of the modal, there are 'Save' and 'Cancel' buttons, and a large green 'Update 527 documents' button.

5.3.12.

```
{  
  "filter": {  
    "year": {  
      "$gte": 1980,  
      "$lte": 1990  
    }  
  },  
  "update": {  
    "$push": {  
      "genres": "New Genre"  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The top navigation bar includes 'MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies', 'Connections', 'Edit', 'View', 'Collection', 'Help', and 'Open MongoDB shell'. The left sidebar shows 'Compass' and 'sample_mflix.movies' collections. The main area displays an 'Update 2181 documents' dialog. The 'Filter' field contains the query: { year: { \$gte: 1980, \$lte: 1990 } }. The 'Update' field contains the update operation: { '\$push': { 'genres': 'New Genre' } }. A preview window shows a sample of three documents, one of which is expanded to show fields like _id, plot, genres, runtime, cast, num_mflix_comments, poster, title, fullplot, languages, released, directors, writers, awards, lastupdated, year, imdb, countries, type, and tomatoes. At the bottom of the dialog are 'Save' and 'Update 2181 documents' buttons. The bottom right corner of the screen shows a taskbar with various icons and the date/time: 02:32 PM, 17/03/2025.

5.3.13.

```
{  
  "filter": {  
    "year": {  
      "$ne": 2010  
    }  
  },  
  "update": {  
    "$set": {  
      "status": "updated"  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays the connection 'cluster0.2alsq.mongodb.net' and the database 'sample_mflix'. The 'movies' collection is selected. A modal window titled 'Update 20503 documents' is open, containing the following update command:

```
{  
  "filter": {  
    "year": {  
      "$ne": 2010  
    }  
  },  
  "update": {  
    "$set": {  
      "status": "updated"  
    }  
  },  
  "multi": true  
}
```

The 'Preview' section shows a sample of three documents from the collection. One document is highlighted, showing its fields and the updated 'status' value:

```
_id: "573a1390f29313caabcd446f"  
plot: "A greedy tycoon decides, on a whim, to corner the world market in whea..."  
runtime: 14  
title: "A Corner in Wheat"  
...  
status: "updated"
```

At the bottom right of the modal, there are 'Cancel' and 'Update 20503 documents' buttons. The status bar at the bottom of the screen indicates the date and time as 17/03/2025, 02:33 PM.

5.3.14.

```
{  
  "filter": {  
    "cast": {  
      "$in": [  
        "Tom Hanks",  
        "Robert Downey Jr."  
      ]  
    }  
  },  
  "update": {  
    "$addToSet": {  
      "imdb.top": 1  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The left sidebar displays the database structure with 'sample_mflix' selected. The main area shows a modal dialog titled 'Update 72 documents'. Inside the dialog, the 'Filter' field contains the query: { cast: { \$in: ['Tom Hanks', 'Robert Downey Jr.'] } }. Below it, the 'Update' section contains the update command: { "\$addToSet": { "imdb.top": 1 } }. To the right, a 'Preview' section shows a sample of three documents. One document is highlighted, showing fields like _id, year, languages, released, directors, writers, awards, lastupdated, imdb (with rating, votes, id), top (an array containing one element), countries, type, and num_mflix_comments. At the bottom of the dialog are 'Save', 'Cancel', and 'Update 72 documents' buttons. The status bar at the bottom of the screen shows system information including battery level, signal strength, and the date/time (17/03/2025).

5.3.15.

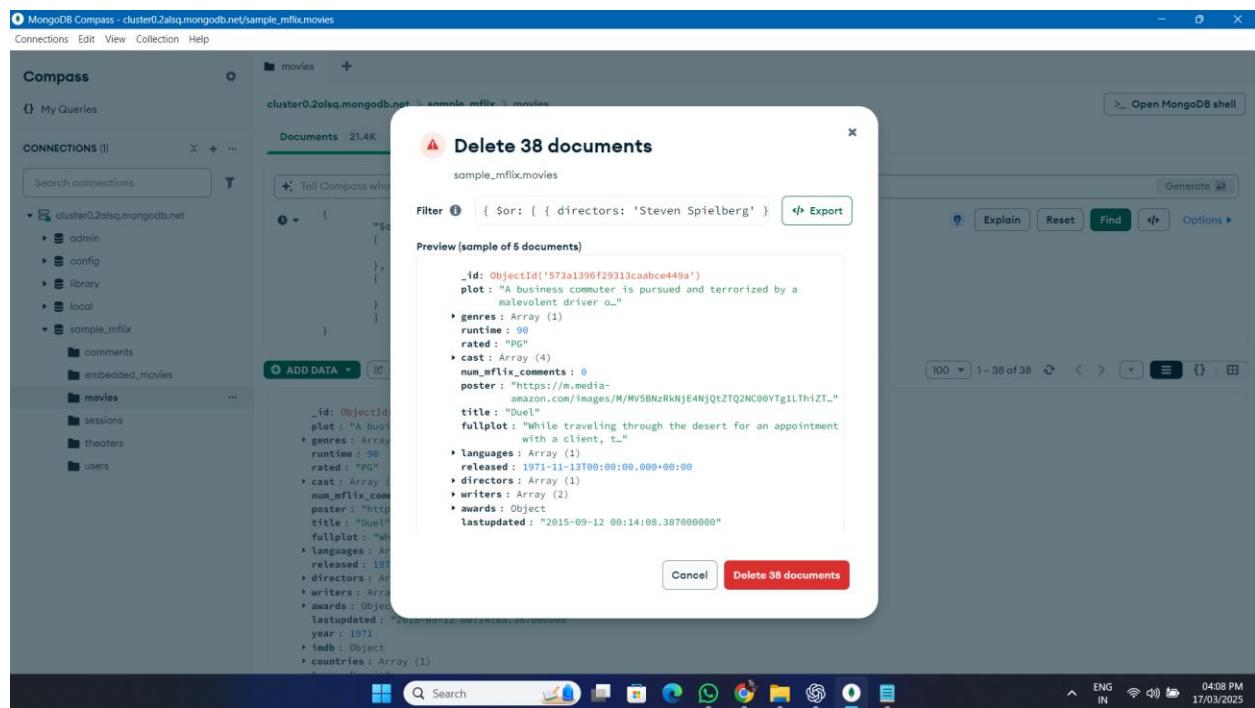
```
{  
  "filter": {  
    "languages": {  
      "$all": [  
        "English",  
        "French"  
      ]  
    }  
  },  
  "update": {  
    "$pull": {  
      "genres": "Romance"  
    }  
  },  
  "multi": true  
}
```

The screenshot shows the MongoDB Compass interface. The top navigation bar includes 'MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies', 'Connections', 'Edit', 'View', 'Collection', 'Help', and 'Open MongoDB shell'. The left sidebar shows 'Compass' and 'sample_mflix.movies' collections. The main area displays an 'Update 1247 documents' dialog. The 'Filter' field contains the query: { languages: { \$all: ['English', 'French'] } }. The 'Update' section shows the '\$pull' operation: { "genres": "Romance" }. A preview window shows the first three documents being updated, including fields like _id, plot, genres, cast, num_mflix_comments, poster, title, fullplot, awards, languages, released, directors, writers, year, imdb, and countries. At the bottom right of the dialog are 'Cancel' and 'Update 1247 documents' buttons. The bottom of the screen shows a Windows taskbar with icons for File Explorer, Edge, and other applications, along with system status indicators.

5.4. Delete

5.4.1.

```
{  
  "$or": [  
    {  
      "directors": "Steven Spielberg"  
    },  
    {  
      "directors": "Christopher Nolan"  
    }  
  ]  
}
```



5.4.2.

```
{  
  "$and": [  
    {  
      "year": {  
        "$gte": 2000  
      }  
    },  
    {  
      "genres": "Sci-Fi"  
    }  
  ]  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays connections and collections, with 'sample_mflix' selected. In the main area, the 'movies' collection is shown with 21.4K documents. A modal dialog titled 'Delete 527 documents' is open, showing a preview of five documents. One document is expanded to show its full JSON structure, including fields like _id, fullplot, year, plot, genres, rated, metacritic, title, lastupdated, languages, writers, type, tomatoes, poster, and num_mflix_comments. At the bottom of the dialog are 'Cancel' and 'Delete 527 documents' buttons. The status bar at the bottom right shows '04:09 PM 17/03/2025'.

5.4.3.

```
{  
  "$and": [  
    {  
      "year": {  
        "$gt": 1995  
      }  
    },  
    {  
      "year": {  
        "$lt": 2010  
      }  
    }  
  ]  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the 'Connections' sidebar lists 'cluster0.2alsq.mongodb.net'. The main area shows the 'movies' collection with 21,4K documents. A modal dialog titled 'Delete 9070 documents' is open, displaying a query filter: { \$and: [{ year: { \$gt: 1995 } }, { year: { \$lt: 2010 } }]}. Below the filter, a preview shows the first five documents of the selection. The preview includes fields like _id, plot, genres, runtime, metacritic, rated, cast, poster, title, fullplot, languages, released, directors, writers, awards, and lastupdated. At the bottom of the dialog are 'Cancel' and 'Delete 9070 documents' buttons. The status bar at the bottom right shows '04:10 PM 17/03/2025'.

5.4.4.

```
{  
  "cast": {  
    "$in": [  
      "Tom Hanks",  
      "Robert Downey Jr."  
    ]  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays connections and collections, with 'sample_mflix' selected. The main area shows a list of documents under the 'movies' collection. A modal dialog titled 'Delete 72 documents' is open, showing a preview of the query and the first document's details. The document preview includes fields like _id, plot, genres, rated, cast, poster, title, fullplot, languages, released, directors, writers, awards, lastupdated, year, and various arrays for sessions, theaters, and comments. At the bottom of the dialog are 'Cancel' and 'Delete 72 documents' buttons. The status bar at the bottom right shows system information: ENG IN, 04:00 PM, 17/03/2025.

5.4.5.

```
{  
    "$or": [  
        {  
            "genres": "Action"  
        },  
        {  
            "genres": "Comedy"  
        }  
    ]  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays connections and collections, with 'sample_mflix' selected. The main area shows a list of documents under the 'movies' collection. A modal dialog titled 'Delete 8398 documents' is open, showing a query filter: '{ \$or: [{ genres: 'Action' }, { genres: 'Comedy' }]}'. Below the filter, a preview section shows the first few fields of one document, including _id, plot, genres, cast, runtime, num_mflix_comments, poster, title, fullplot, languages, released, directors, writers, awards, lastupdated, year, imdb, countries, and type. At the bottom of the dialog are 'Cancel' and 'Delete 8398 documents' buttons. The status bar at the bottom right indicates the date and time as 17/03/2025, 04:11 PM.

5.4.6.

```
{  
  "directors": {  
    "$exists": true  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the connection tree displays a cluster named 'cluster0.2aisq.mongodb.net' with a database 'sample_mflix' and a collection 'movies'. A modal dialog box is centered over the interface, titled 'Delete 21127 documents'. The dialog contains a warning icon and the text 'sample_mflix.movies'. Below this, a 'Filter' field shows the query: '{ directors: { \$exists: true } }'. A preview section shows the first document of 5 samples, which is a movie document with fields like _id, plot, genres, cast, languages, title, fullplot, released, directors, awards, lastupdated, year, imdb, and tomatoes. At the bottom of the dialog are 'Cancel' and 'Delete 21127 documents' buttons. The background of the interface shows the list of movies in the 'movies' collection.

5.4.7.

```
{  
  "plot": {  
    "$regex": "love",  
    "$options": "i"  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the 'Connections' sidebar lists 'cluster0.2alsq.mongodb.net' with databases 'admin', 'config', 'local', and 'sample_mflix'. The 'movies' database is selected. In the main pane, the 'sample_mflix.movies' collection has 21.4K documents. A search bar at the top right contains the query: { plot: { \$regex: 'love', \$options: 'i' } }. A modal dialog titled 'Delete 1703 documents' is open, showing a preview of five documents. One document's details are visible:

```
_id: ObjectId('573a1391f29313caabcd7a34')  
plot: "A kept woman runs into her one-time fiance and finds herself torn betw...  
genres: Array (2)  
  rated: "TV-PG"  
  runtime: 78  
cast: Array (4)  
  num_mflix_comments: 1  
  poster: "https://m.media-amazon.com/images/M/MVSBZjJ1HTU2NGQtNWRKN100ZjExLWExMT_..."  
  title: "A Woman of Paris: A Drama of Fate"  
  fullplot: "Marie St. Clair believes she has been jilted by her artist fiance Jean."  
countries: Array (1)  
  released: 1923-11-04T00:00:00+00:00  
directors: Array (1)  
  writers: Array (1)  
awards: Object  
lastupdated: "2015-09-02 00:22:09.303000000"  
imdb: Object  
type: "movie"  
tomatoes: Object
```

The dialog includes 'Cancel' and 'Delete 1703 documents' buttons. The bottom status bar shows 'ENG IN' and the date '17/03/2025'.

5.4.8.

```
{  
  "year": {  
    "$gte": 1980,  
    "$lte": 1990  
  }  
}
```

The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays connections and collections, with 'sample_mflix' selected. In the main area, a search bar and filter dropdown are visible. A modal dialog titled 'Delete 2181 documents' is open, containing a warning icon and a preview of five documents. The preview shows a movie document with fields like _id, plot, genres, cast, num_mflix_comments, poster, title, fullplot, languages, released, directors, writers, awards, lastupdated, year, and various URLs for images and objects. At the bottom of the dialog are 'Cancel' and 'Delete 2181 documents' buttons. The background shows the 'movies' collection with 21.4K documents. The status bar at the bottom right indicates the date and time as 17/03/2025 and 04:13 PM.

5.4.9.

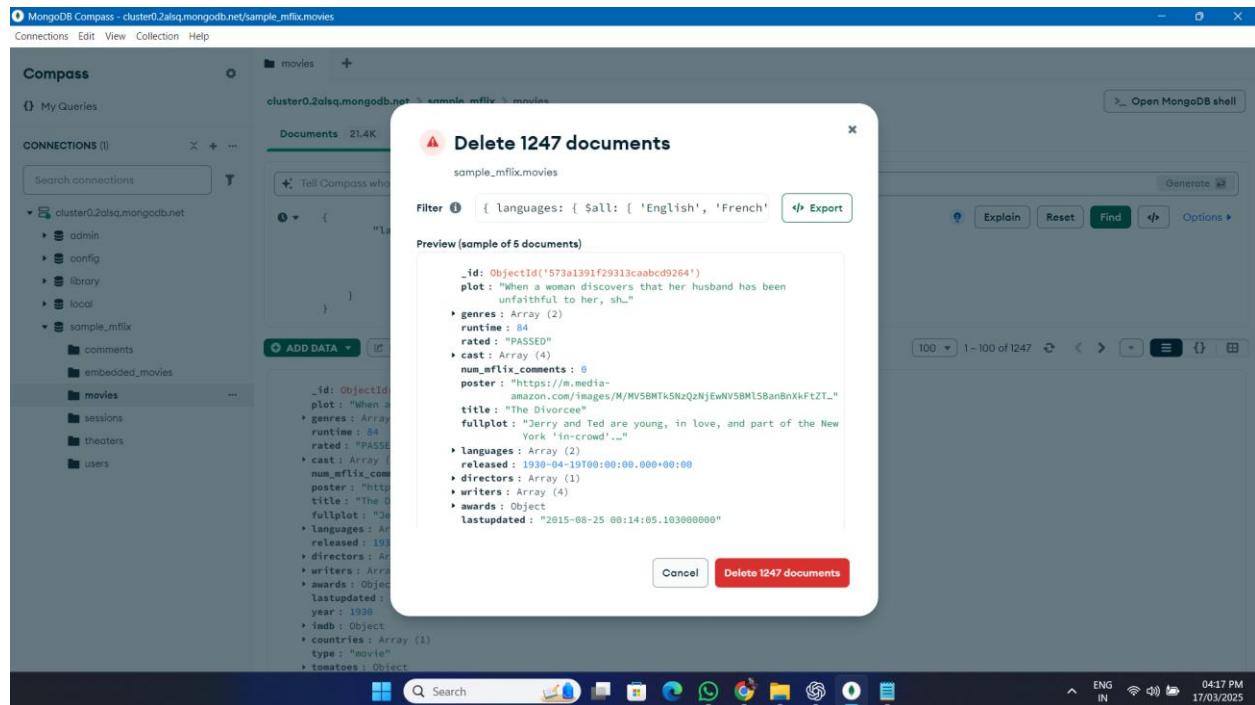
```
{  
    _id:'573a1390f29313caabcd5ea4'  
}
```

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays connections and databases, with 'sample_mflix' selected. The main area shows the 'movies' collection with 21.4K documents. A search bar at the top contains the query: { _id: '573a1390f29313caabcd5ea4' }. Below the search bar, the results are displayed as a single document:

```
_id: "573a1390f29313caabcd5ea4"  
plot: "A District Attorney's outspoken stand on abortion gets him in trouble..."  
runtime: 62  
title: "Where Are My Children?"  
fullplot: "While prosecuting a physician for the death of a client after an abort..."  
released: "1916-05-01T00:00:00Z"  
rated: "Approved"  
awards.wins: 1  
awards.nominations: 0  
awards.text: "I win."  
lastupdated: "2015-09-07 00:51:32.560000000"  
year: "1916"  
imdb.rating: 5.9  
imdb.votes: 247  
imdb.id: 7558  
type: "movie"  
tomatoes.viewer.rating: 3.1  
tomatoes.viewer.numReviews: 34  
tomatoes.viewer.meter: 50  
tomatoes.lastupdated: "2015-08-06T19:49:17.000Z"  
num_mflix.comments: 0  
tomatoes.production: "MCA/Universal Pictures"  
genres: Array (1)  
cast: Array (4)  
languages: Array (1)
```

5.4.10.

```
{  
  "languages": {  
    "$all": [  
      "English",  
      "French"  
    ]  
  }  
}
```

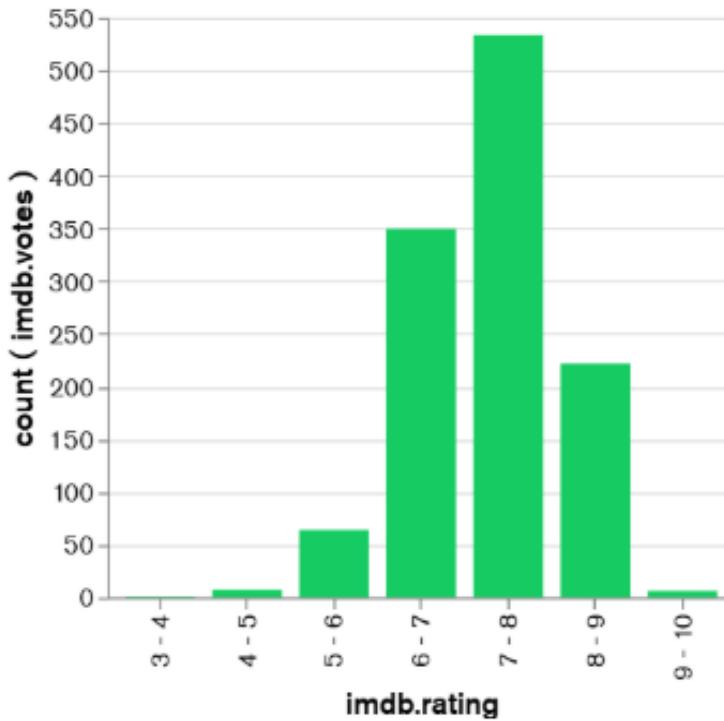


5.5. Key Observations and Insights

5.5.1. Distribution of IMDb Ratings for Movies with More Than 100,000 Votes

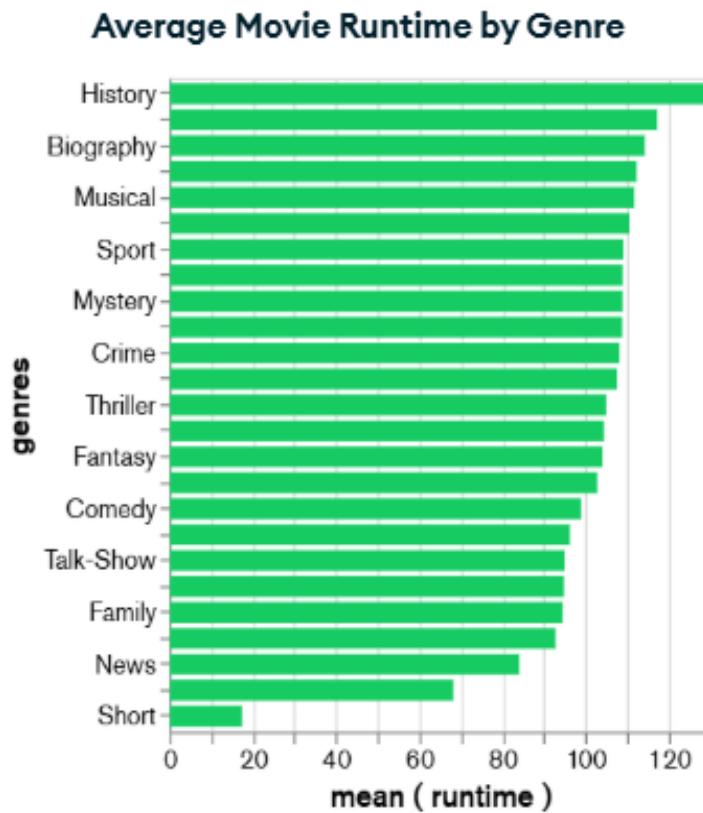
- Most movies fall within the **7.0–8.0 rating range**, indicating that audiences generally rate popular movies positively.
- A significant drop is observed for movies rated **below 6.0**, showing that fewer widely popular movies receive lower ratings.
- Very few movies have ratings below **4.0 or above 9.0**, suggesting that extreme ratings (very low or very high) are rare.

Distribution of IMDb ratings for movies wi...



5.5.2. Average Movie Runtime by Genre

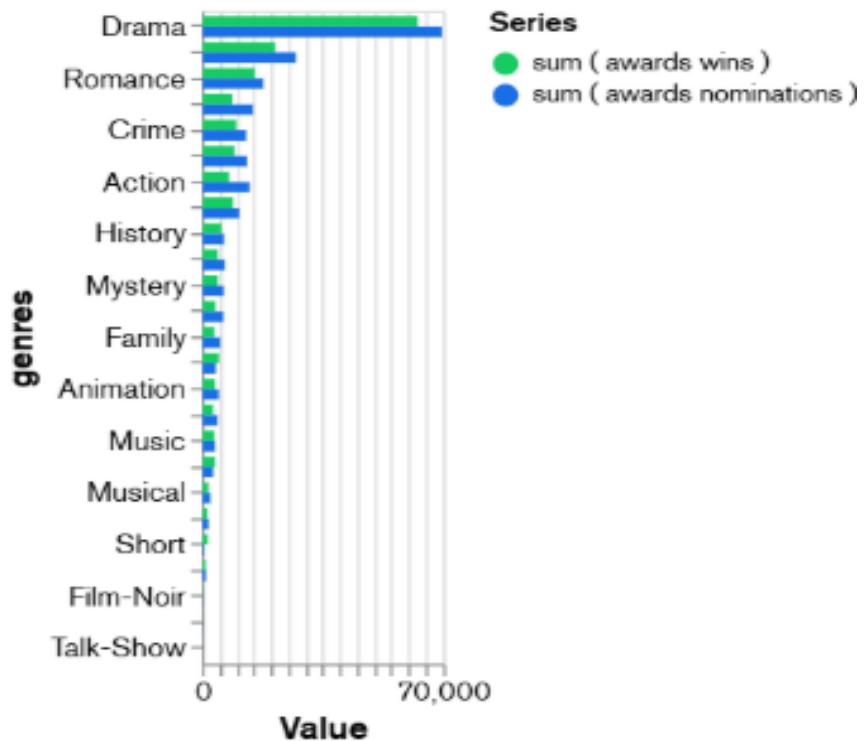
- **Historical movies have the longest runtime**, averaging over 120 minutes, which aligns with the genre's need for extensive storytelling.
- **Short films have the lowest runtime**, as expected, reinforcing the distinction in duration between different formats.
- Biographical, musical, and sports genres also tend to have **longer runtimes**, possibly due to the need for detailed character development.



5.5.3. Award Wins and Nominations by Genre: A Comparative Analysis

- **Drama dominates** in award wins and nominations, suggesting that this genre is highly favored by award committees.
- **Romance, Crime, and Action** follow in recognition, although their numbers are significantly lower than Drama.
- **Genres like Talk-Show and Film-Noir** receive very few awards, indicating they may not be mainstream choices for major award bodies.

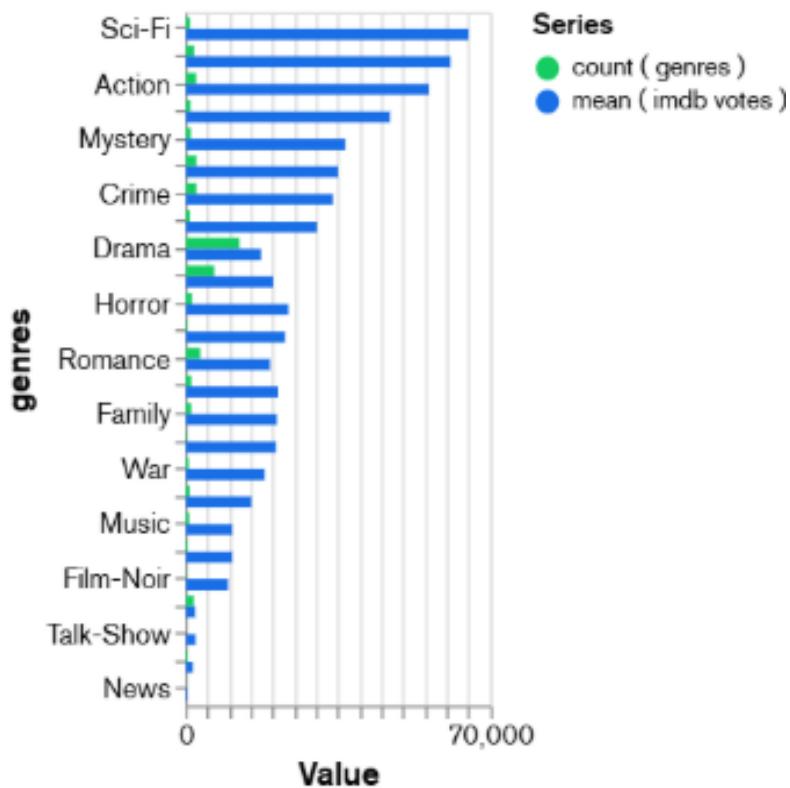
Award Wins and Nominations by Genre: A ...



5.5.4. Comparing Average IMDb Votes Across Different Film Genres

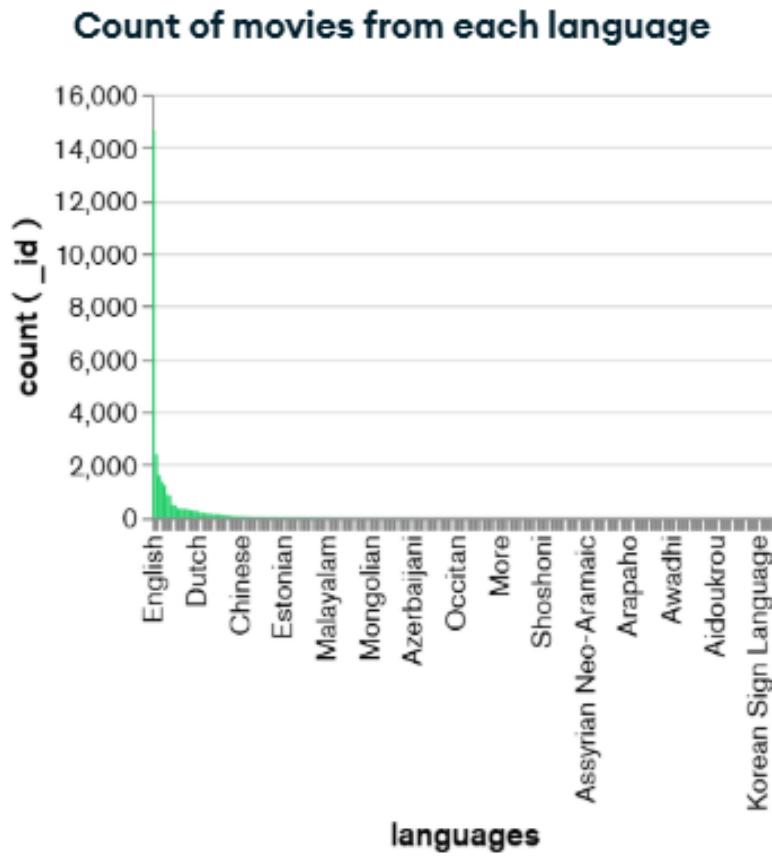
- **Sci-Fi and Action** genres receive the **highest average IMDb votes**, suggesting these genres attract a larger audience and fan engagement.
- **Drama, Horror, and Mystery** also perform well, showing that audience interest spans across both intense and emotional storytelling.
- **Genres like Talk-Show and News** receive significantly fewer votes, indicating a niche audience rather than widespread appeal.

Comparing Average IMDb Votes Across Di...



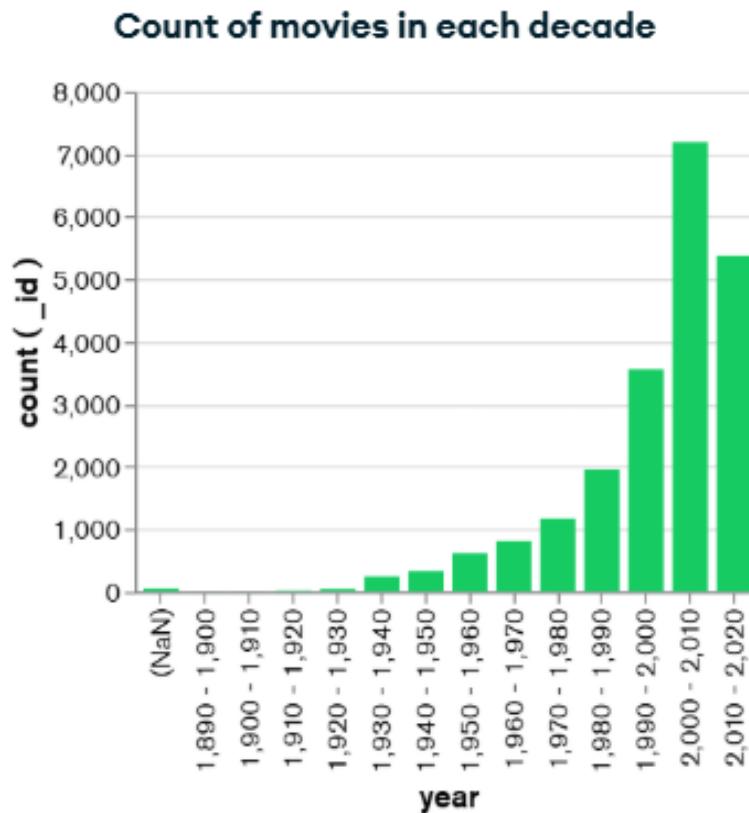
5.5.5. Count of Movies from Each Language

- **English dominates** the dataset, with a significantly higher number of movies compared to any other language.
- The distribution follows a **long-tail pattern**, where a few languages have a high number of films, but most languages have very few.
- **Languages like Dutch, Chinese, and Estonian** have a noticeable presence, but they are far behind English.



5.5.6. Count of Movies in Each Decade

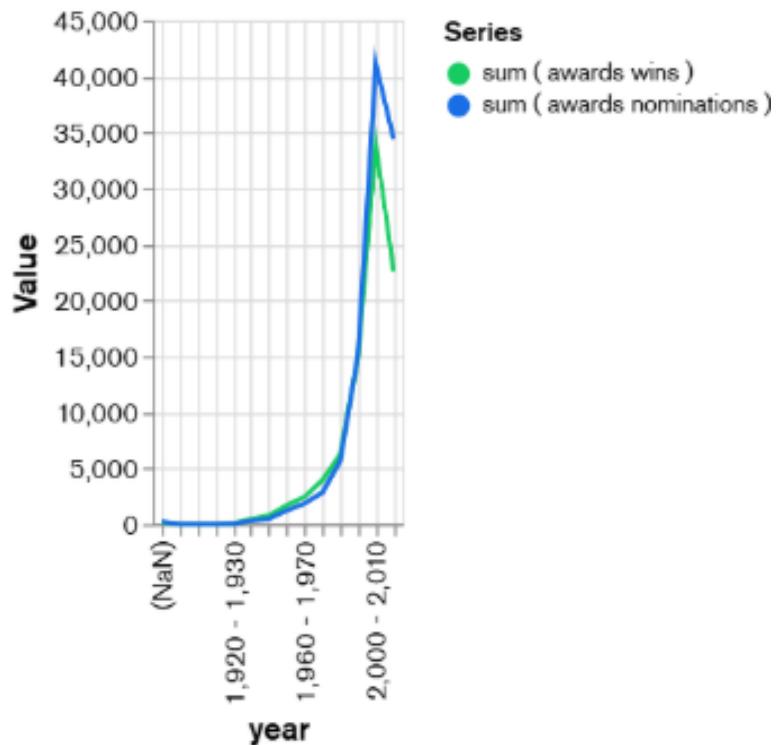
- The number of movies has **increased drastically over time**, with the highest count in the **2000s and 2010s**.
- Early decades (1890–1950) have **very few movies**, indicating limited production capacity and technology.
- The rise in movie production **post-1980** aligns with technological advancements and the growth of the entertainment industry.



5.5.7. Decadal Trends in Movie Wins and Nominations

- Award wins and nominations have **increased significantly since the 1970s**, peaking in the 2000s.
- This may indicate a **rise in film festivals, streaming content, and award events**.

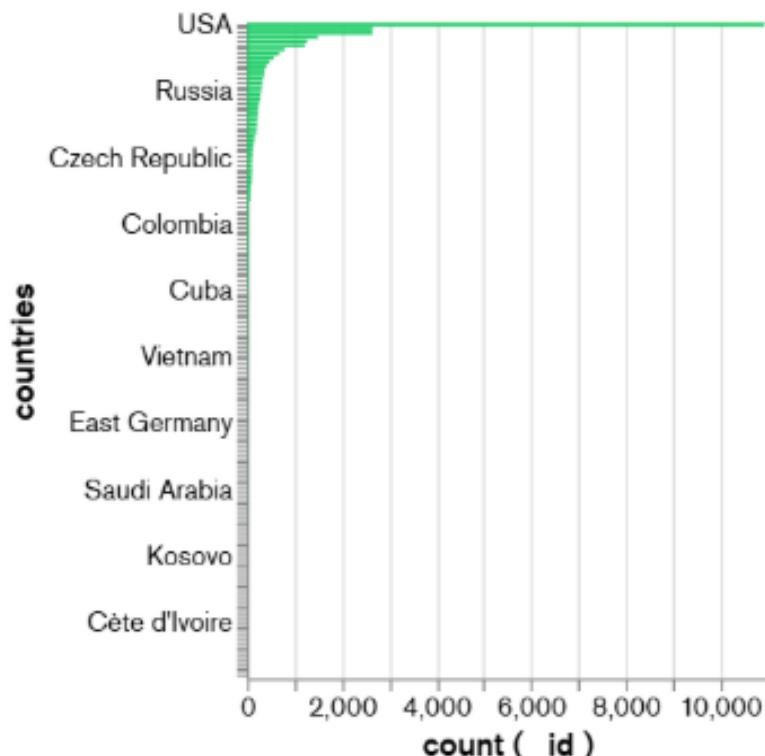
Decadal Trends in Movie Wins and Nomina...



5.5.8. Distribution of Movie Counts Across Different Countries

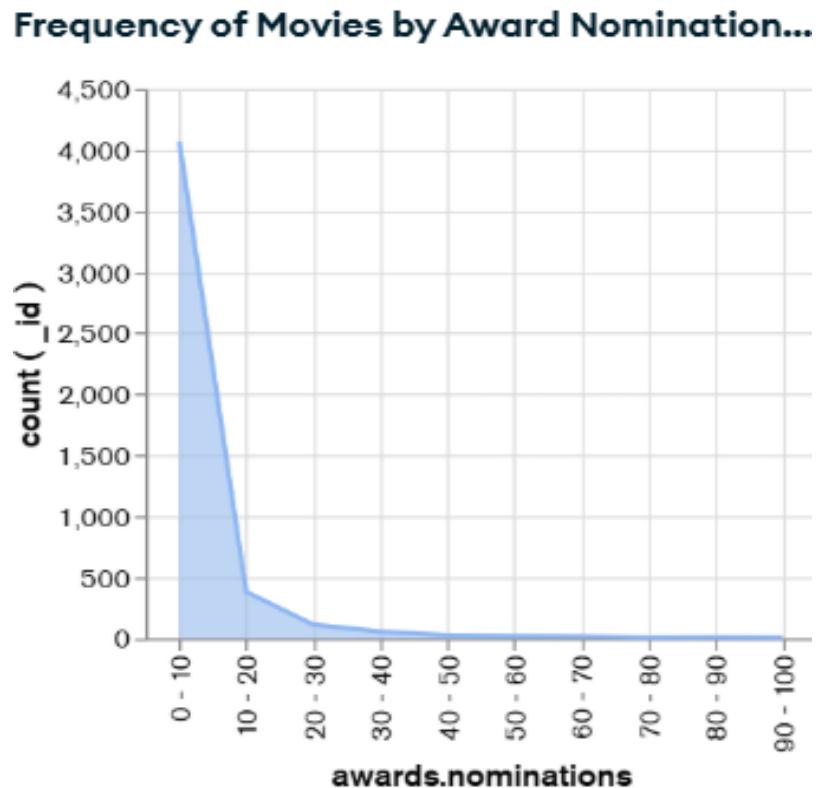
- The USA has an overwhelmingly high number of movies compared to other countries.
- Russia follows, but no other country comes close, highlighting Hollywood's **dominant influence** in film production.

Distribution of Movie Counts Across Differ...



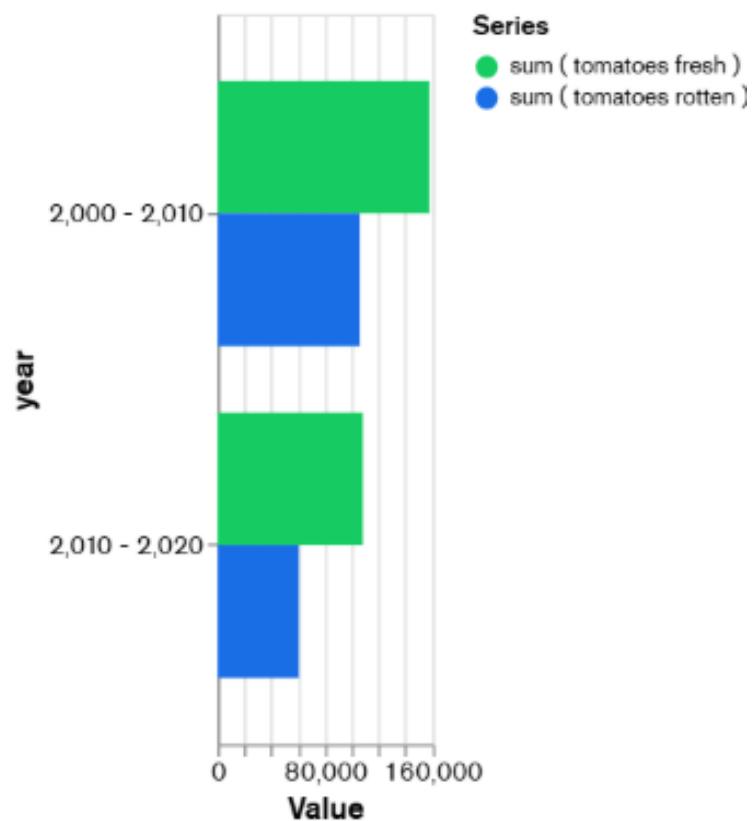
5.5.9. Frequency of Movies by Award Nominations Count

- Most movies receive **fewer than 10 nominations**, while very few movies get high nominations (above 50).
- This suggests that **highly nominated films are rare** and prestigious.



5.5.10. Grouped Bar Chart of Rotten Tomatoes Ratings (Fresh vs. Rotten)

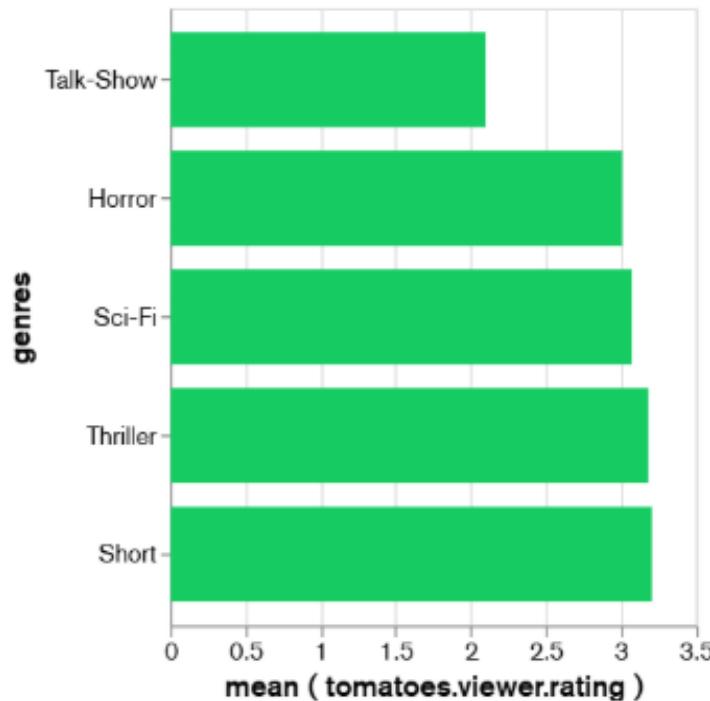
- More movies received "Fresh" ratings than "Rotten," suggesting a general tendency for movies to be rated positively.
- However, there is still a **substantial number** of negatively rated movies.



5.5.11. Lowest Rated Genres Based on Average Viewer Ratings

- **Talk-Show, Horror, and Sci-Fi** have the lowest average viewer ratings.
- **Thriller and Short** films also rank low, indicating that audiences might find them less appealing.

Lowest Rated Genres Based on Average Vi...



5.5.12. Movie Count Categorized by Award Wins

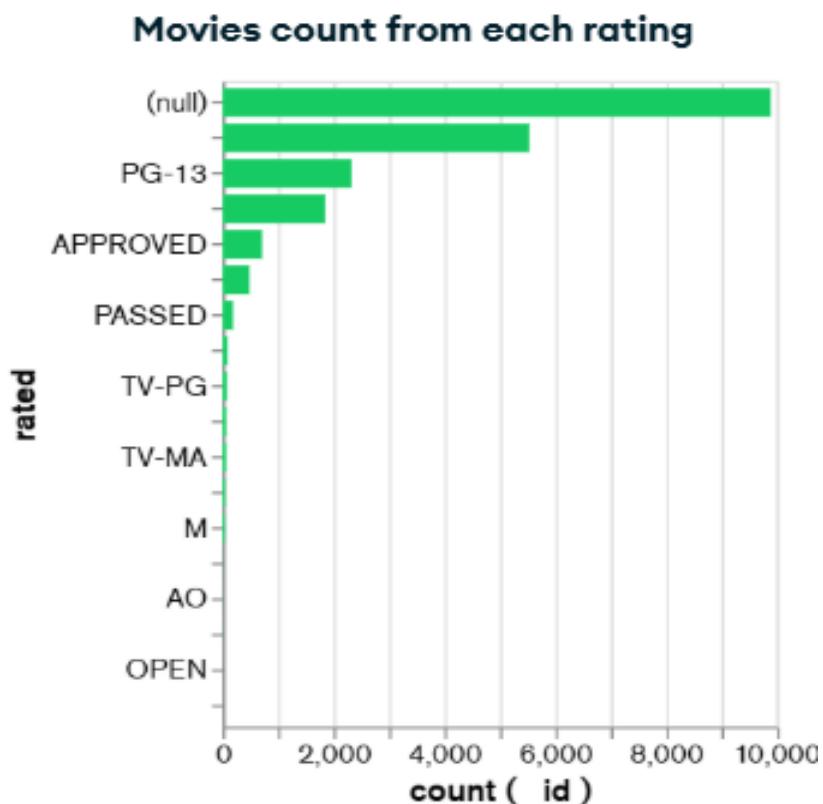
- The majority of movies (over **19,000**) have won between **0-10 awards**.
- Only **a few hundred** movies have won more than **20 awards**, making highly awarded films extremely rare.

Movie Count Categorized by Award Wins

awards.wins	count (_id)
(NaN)	20
0 - 10	19,223
10 - 20	1,416
20 - 30	360
30 - 40	129
40 - 50	74
Total	21,369

5.5.13. Movies Count from Each Rating

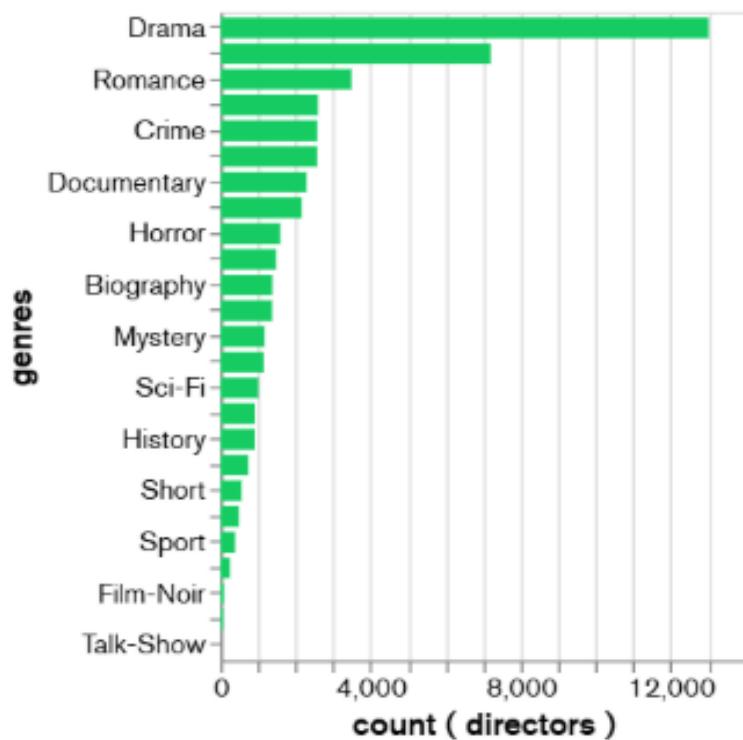
- A significant number of movies **lack a specified rating** (`null`), indicating missing or unclassified rating data.
- The most common explicit rating is **PG-13**, suggesting that a large portion of movies cater to a wide audience.
- Older rating systems like **APPROVED** and **PASSED** have significantly fewer movies, indicating that most films in the dataset are from the modern rating system.
- Ratings such as **TV-MA** and **AO** have very low counts, suggesting that explicit or adult content films are less frequent in this dataset.



5.5.14. Number of Directors in Each Movie Category

- **Drama** has the highest number of directors, indicating its popularity and the variety of storytelling in this genre.
- **Romance and Crime** also have significant director counts, showing their strong presence in the film industry.
- **Documentary** has a high director count, suggesting a strong representation of real-world narratives in filmmaking.
- Niche genres like **Film-Noir and Talk-Show** have the lowest number of directors, which aligns with their relatively smaller market share.

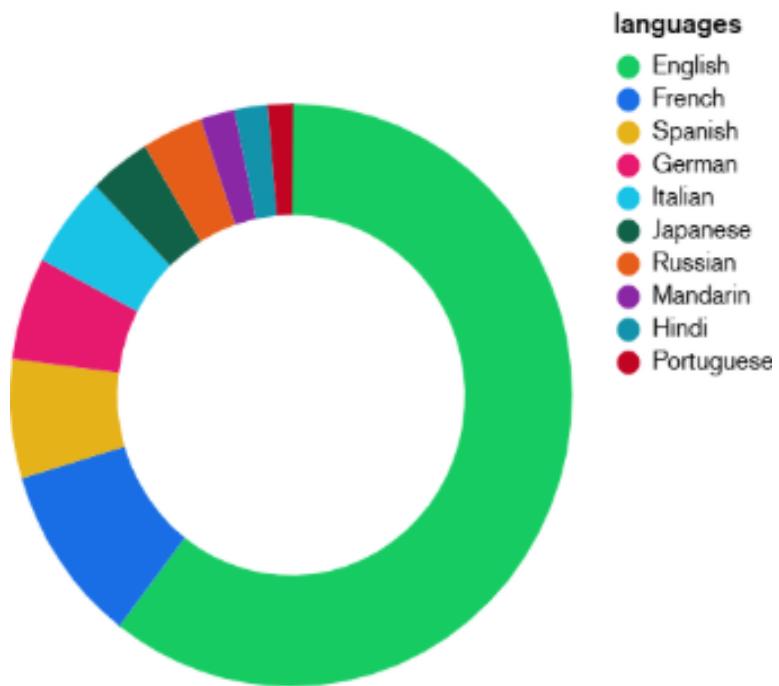
Number of directors in each movie category



5.5.15. Share of Top 10 Most Popular Languages in Movies

- **English** dominates the dataset, implying that most of the movies in this collection are from English-speaking countries.
- **French, Spanish, and German** have a noticeable share, showing the global presence of European cinema.
- **Hindi and Mandarin** represent the influence of Bollywood and Chinese cinema, but their share is smaller compared to English-language films.
- Other languages such as **Japanese and Russian** also have a presence, highlighting international diversity in filmmaking.

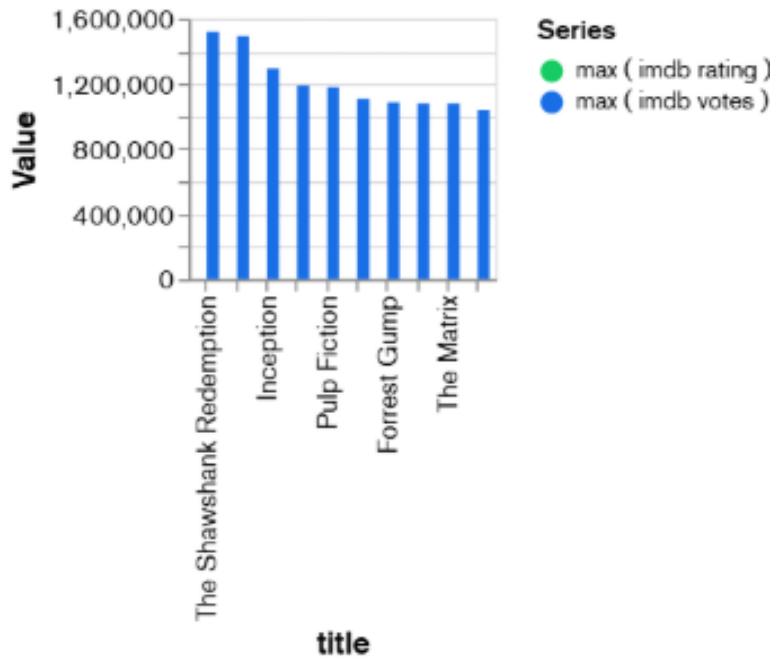
Share of top 10 most popular languages in ...



5.5.16. Top 10 Highest Rated Movies with Most Votes on IMDb

- **The Shawshank Redemption** has the highest votes and rating, reinforcing its status as one of the most beloved films.
- **Inception, Pulp Fiction, Forrest Gump, and The Matrix** also have high votes, indicating their widespread popularity and critical acclaim.
- The high number of votes suggests that these movies have a **strong fan base** and remain culturally significant.
- Most of these films belong to the **Drama, Thriller, or Sci-Fi** genres, suggesting these genres resonate well with IMDb users.

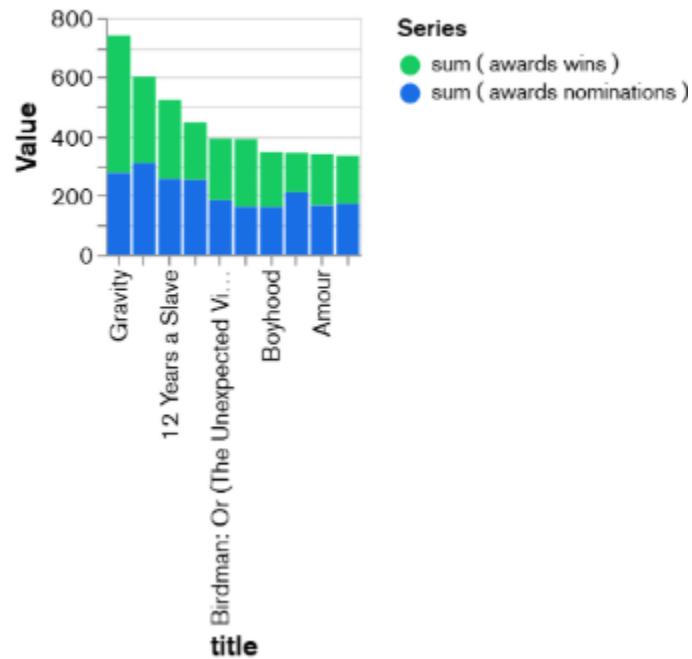
Top 10 Highest Rated Movies with Most Vot...



5.5.17. Top 10 IMDb Movies by Total Wins and Nominations

- **Gravity** has the highest combined wins and nominations, indicating strong critical and award recognition.
- **12 Years a Slave** and **Birdman** also have high award counts, suggesting the dominance of drama-based and socially relevant films in prestigious awards.
- The movies in this list generally have a **strong critical reception**, reinforcing the correlation between award recognition and quality filmmaking.
- A trend can be observed where **movies with strong narratives and technical excellence** receive the most awards.

Top 10 IMDb Movies by Total Wins and Nom...



5.5.18. Top 10 Movies with the Highest IMDb Rating

- The movies listed have ratings between **5.5** and **7.6**, which suggests that they are not the highest-rated films overall but still hold moderate appreciation.
- **Édipussi (7.6)** and **¡Vivan las Antípodas! (7.1)** are among the highest-rated, showing audience preference for unique storytelling or cultural significance.
- Some of the movies on this list have relatively **lower IMDb ratings** compared to what might be expected in a “Top 10” list, possibly indicating a **niche appeal** rather than mainstream popularity.

Top 10 movies with highest IMDB rating

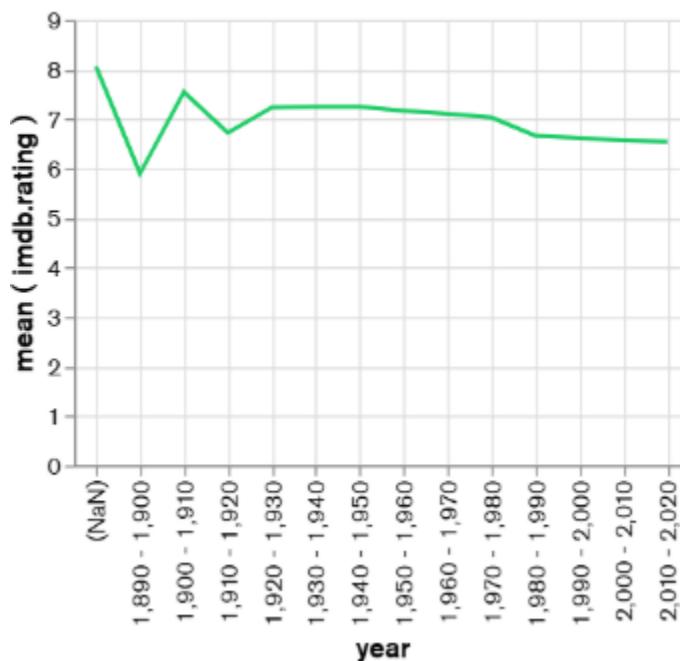
title ↑↓	max (imdb.rat...)
éxtasis	6.9
évocateur: The Morton Downey Jr. Movie	6.8
éríðo Movie	6.7
éon Flux	5.5
élisa	6.7
ékte vare	6.2
éke och hans värld	6.7
édipussi	7.6
éber uns das All	6.7
éVivan las Antipodas!	7.1

Total	66.9
-------	------

5.5.19. Trends in Average Movie Ratings by Decade

- The average IMDb rating of movies has been **gradually declining** over time.
- Movies from the early 20th century (**1900-1950**) generally had **higher average ratings**, possibly due to fewer releases and a focus on high-quality productions.
- A noticeable decline starts from the **1980s onwards**, which may be attributed to an increase in mass production of films, changes in audience preferences, and the diversification of movie genres.
- Modern films (**2000s–2020s**) tend to have **lower ratings** compared to earlier decades, which could also reflect stricter audience critique and more widespread access to ratings.

Trends in Average Movie Ratings by Decade



5.5.20. Trends in Long Movies – Yearly Count of Films with Over 120 Minutes Runtime

- There has been a **steady increase** in the number of long movies (over 120 minutes) over the decades.
- From the **1980s onwards**, there is a **sharp rise**, showing a trend where filmmakers prefer longer runtimes, possibly to accommodate more complex storytelling.
- The peak around the **2000s–2010s** suggests a golden era for long-form cinema, with high-budget productions often exceeding 2 hours.
- A **slight dip** towards the most recent years might indicate a shift in audience preference towards shorter, more digestible content, influenced by streaming services and digital consumption habits.

Trends in Long Movies: Yearly Count of Fil...

