

Big Data Management and Analysis



Project Report On Data Visualisation using MongoDB Atlas Charts

Submitted to:
Prof (Dr.) Amarnath Mitra

Submitted By:
Preksha Verma (055032)
Suvra Datta Banik (055049)

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 the 'sample_mflix' database. A modal window titled 'Import' is open, indicating the file 'insert_movies_compass.txt' is being imported into the 'movies' collection. The 'Stop on errors' option is checked. The main interface shows the 'movies' collection with 21.4K documents.

Import
To collection sample_mflix.movies
Import file: insert_movies_compass.txt
Options
Stop on errors

Documents 21.4K **Aggregations** **Schema** **Indexes** 2 **Validation**

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

Connections (1) **Search connections**

cluster0.2alsq.mongodb.net
 └ movies

ADD DATA **EXPORT DATA** **UPDATE** **DELETE**

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)

Explain Reset Find Options

ADD DATA EXPORT DATA UPDATE DELETE

100 1 - 100 of 11988

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
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. On the left sidebar, the connection 'cluster0.2alsq.mongodb.net' is selected, showing its contents: admin, config, library, local, sample_mflix, comments, embedded_movies, and movies. The 'movies' folder is currently selected. The main pane shows the document structure for a movie released in 2001, including fields like 'French', 'released', 'directors', 'writers', 'awards', 'imdb', 'countries', and 'type'. The bottom status bar shows the system language as 'ENG' and the date and time as '01:48 PM 17/03/2025'.

5.2.2. Retrieval of data for year 1990

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

The screenshot shows the MongoDB Compass interface. On the left, the connection tree displays a cluster named 'cluster0.2alsq.mongodb.net' with databases 'admin', 'config', 'library', 'local', and 'sample_mflix'. The 'sample_mflix' database is expanded, showing collections 'comments', 'embedded_movies', 'movies', 'sessions', 'theaters', and 'users'. The 'movies' collection is selected and highlighted in grey. The main pane shows a search bar with the query: { "year": { "\$lt": 1990 } }. Below the search bar, a document is displayed with the following fields and values:

```
  "_id": "43E...",  
  "adult": false,  
  "country": "USA",  
  "imdb": {  
    "rating": 8.5,  
    "votes": 9847  
  },  
  "lastupdated": "2015-08-13 00:27:59.177000000",  
  "rated": "TV-14",  
  "titles": {  
    "original": "The Shawshank Redemption",  
    "primary": "The Shawshank Redemption",  
    "secondary": "The Shawshank Redemption",  
    "tertiary": null  
  },  
  "tomatoes": {  
    "meter": 75,  
    "rating": 3.7,  
    "numReviews": 2559  
  },  
  "year": 1994
```

At the bottom of the main pane, there are buttons for 'ADD DATA', 'EXPORT DATA', 'UPDATE', and 'DELETE'. The status bar at the bottom right shows the date and time: 17/03/2025, 01:49 PM.

5.2.3. Retrieval of data for runtime 120 minutes

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

The screenshot shows the MongoDB Compass interface. The left sidebar displays 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, a query builder shows the following expression:

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

The results pane shows a single document from the 'movies' collection, which includes fields like _id, plot, genres, runtime, cast, languages, num_mflix_comments, poster, title, fullplot, released, directors, writers, and sessions.

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 displays 'cluster0.2alsq.mongodb.net' with its databases: admin, config, library, local, sample_mflix, comments, embedded_movies, and movies. The 'movies' database is selected. In the center, the 'sample_mflix' collection is shown 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 } }. To the right of the code are buttons for Explain, Reset, Find, Options, and Generate. The results pane shows a single document from the collection, including fields like _id, plot, runtime, cast, poster, title, fullplot, genres, languages, awards, directors, year, imdb, countries, type, tomatoes, and num_mflix_comments. The bottom status bar shows system information: ENG IN, 01:51 PM, 17/03/2025.

5.2.5. Retrieval of data for runtime 120 minutes

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

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays connections and collections, with 'sample_mflix' selected. The main area shows a search bar with the query: { "year": { "\$ne": 2010 } }. Below the search bar, a document preview is shown, starting with the '_id' field. The bottom status bar shows system information like battery level, signal strength, and the date.

```
_id: ObjectId('573a1390f29313caabcd42e8')  
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/MVSBMTU3NjE5NzYtYTYYNS00MDVmWiwYj..."  
title: "The Great Train Robbery"  
fullplot: "Among the earliest existing films in American cinema - notable as the ..."  
languages: Array (1)  
released: 1903-12-01T00:00:00.000+00:00  
directors: Array (1)  
rated: "TV-G"  
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
```

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 pane shows a search query entered into the search bar: { \$or: [{ genres: "Action" }, { genres: "Comedy" }] }. Below the query, a detailed document view is shown for a movie entry. The document includes fields like _id, plot, genres (with values Animation, Short, Comedy), runtime, cast (with Winsor McCay), num_mflix_comments, poster URL, title, fullplot, languages (English), release date (1911-04-08T00:00:00+00:00), and directors (Winsor McCay, J. Stuart Blackton). The bottom status bar shows system information including battery level, signal strength, and the date/time (17/03/2025).

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. The main area shows the 'movies' collection with 21.4K documents. A query builder window is open, displaying the aggregation pipeline stage you provided:

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

The results pane shows the first few documents of the filtered collection.

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 with the following details:

- Connections:** cluster0.2aisq.mongodb.net / sample_mflix.movies
- Collection:** movies
- Documents:** 21.4K
- Aggregations:** Schema, Indexes, 2, Validation
- Search Bar:** Tell Compass what documents to find (e.g., which movies were released in 2000)
- Query Preview:** A JSON snippet of the search query is shown:

```
{  
  "cast": {  
    "$in": [  
      "Tom Hanks",  
      "Robert Downey Jr."  
    ]  
  }  
}
```
- Buttons:** Explain, Reset, Find, Options
- Document Preview:** The first few documents are listed, including:
 - 0: "Tom Hanks"
 - 1: "Daryl Hannah"
 - 2: "Eugene Levy"
 - 3: "John Candy"

poster: "https://m.media-amazon.com/images/M/MVSBZDRmZTQ9MDQtNvReY500NDhhLTg2Nz..."
title: "Splash"
fullplot: "Allen Bauer is rescued from drowning as a young boy off Cape Cod by a..."
languages: Array (2)
 - 0: "English"
 - 1: "Swedish"
released: 1984-03-09T00:00:00.000+00:00
directors: Array (1)
 - 0: "Ron Howard"
writers: Array (5)
 - 0: "Brian Grazer (story)"
 - 1: "Bruce Jay Friedman (screen story)"
 - 2: "Lowell Ganz (screenplay)"
 - 3: "Babلوو ماندل (screenplay)"
 - 4: "Bruce Jay Friedman (screenplay)"
awards: Object
wins: 4
- Toolbar:** ADD DATA, EXPORT DATA, UPDATE, DELETE
- Status Bar:** 100 1 - 72 of 72, navigation icons, and system status (ENG IN, 01:54 PM, 17/03/2025)

5.2.9. Retrieval of data where information about the director exists

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

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 the results of a query, starting with the aggregation stage:

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

Below this, the document structure is expanded to show fields like poster, title, fullplot, languages, released, directors, awards, and countries. The 'directors' field is shown as an array containing two elements: "George Barnes" and "Justus D. Barnes". The 'titles' field contains "The Great Train Robbery". The 'languages' field is an array with one element: "English". The 'released' field is a timestamp: 1903-12-01T00:00:00.000+00:00. The 'directors' array has two elements, both labeled '0'. The 'awards' object has 'wins' and 'nominations' fields, both set to 1. The 'countries' array has one element: "USA". The 'imdb' object has 'rating' (8.5), 'votes' (9847), and an 'id' (436). The 'lastupdated' field is a timestamp: 2015-08-13 00:27:59.177000000. The 'year' field is 1903.

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

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

The screenshot shows the MongoDB Compass interface. The left sidebar displays the database connections, 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, a query editor contains the aggregation pipeline:

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

The results pane shows a single document from the collection, which includes fields like _id, plot, genres, runtime, cast, num_mflix_comments, poster, title, fullplot, languages, released, awards, and year.

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 the database structure with a selected 'movies' collection. The main pane shows a search query entered into the search bar: { "genres": { "\$size": 3 } }. Below the query, the results are displayed as a list of movie documents. One document is expanded to show its full details, including fields like _id, plot, genres (which contains three genres: Short, Comedy, Animation), cast, runtime, languages, released date, directors, writers, and awards. The bottom status bar shows system information like network connection, battery level, and the date/time (17/03/2025).

```
_id: ObjectId('573a1390f29313caabcd4803')  
plot: "Cartoon figures announce, via comic strip balloons, that they will mov..."  
genres: Array (3)  
  0: "Short"  
  1: "Comedy"  
  2: "Animation"  
runtime: 7  
cast: Array (1)  
  0: "Winsor McCay"  
num_mflix_comments: 0  
poster: "https://m.media-amazon.com/images/M/MVSBYzg2NjNHNTctMjUxMi00ZWU4LWI3Zj_..."  
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"  
writers: Array (2)  
  0: "Winsor McCay (comic strip "Little Nemo in Slumberland")"  
  1: "Winsor McCay (screenplay)"  
awards: Object
```

5.2.12.

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

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 a 'Find' button. Below the navigation is a search bar with placeholder text 'Tell Compass what documents to find (e.g. which movies were released in 2000)'. A query builder window displays the following aggregation stage:

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

Below the query builder, the results pane shows a single document from the 'movies' collection. The document details are as follows:

```
_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/MVSBZJiMTU2NGQtNWRxN100ZjExLWxM..."  
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+00:00  
directors : Array (1)  
  0: "Charles Chaplin"  
writers : Array (1)
```

The bottom of the screen shows a Windows taskbar with various icons and the system tray indicating the date and time as 17/03/2025 at 01:56 PM.

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 selected. 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:

```
{  
  "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', 'num_mflix_comments', 'poster', 'title', 'fullplot', 'languages', 'released', 'directors', and 'writers'. The 'Find' button is highlighted in green at the bottom of the query builder. The bottom right corner of the screen shows system status icons for battery, signal, and date/time (01:57 PM, 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 the following fields and values:

- `_id`: `ObjectId('573a1391f29313caabcd93aa')`
- `plot`: `"A cabaret singer and a Legionnaire fall in love, but their relationship..."`
- `genres`: `Array (2)
 0: "Romance"
 1: "Drama"`
- `runtime`: `92`
- `rated`: `"PASSED"`
- `cast`: `Array (4)
 0: "Gary Cooper"
 1: "Marlene Dietrich"
 2: "Adolphe Menjou"
 3: "Ullrich Haupt"`
- `num_mflix_comments`: `8`
- `title`: `"Morocco"`
- `fullplot`: `"The Foreign Legion marches in to Mogador with booze and women in mind..."`
- `languages`: `Array (5)
 0: "English"
 1: "French"
 2: "Spanish"
 3: "Arabic"
 4: "Italian"`
- `released`: `1936-12-06T00:00:00+00:00`

The results pane shows 100 documents, with the current view being 1-100 of 4110. The bottom status bar indicates the system is ENG IN, the time is 01:57 PM, and the date is 17/03/2025.

5.2.15.

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

The screenshot shows the MongoDB Compass interface. On the left, the connection tree displays '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 center, the 'Documents' tab shows a search query: { \$or: [{ directors: "Steven Spielberg" }, { directors: "Christopher Nolan" }] }. Below the query, a list of movie documents is shown, starting with one from 'The Usual Suspects'. At the bottom right, the system status bar shows 'ENG IN' and the date '17/03/2025'.

5.2.16.

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

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays connections and collections, with 'sample_mflix' selected under 'movies'. The main area shows a search bar with the query: { "runtime": { "\$ne": null } }. Below the search bar, 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 lawman who will stop at nothing to bring them to justice."  
genres: Array (2)  
  0: "Short"  
  1: "Western"  
runtime: 11  
cast: Array (4)  
  0: "A.C. Abadie"  
  1: "Gilbert M. 'Broncho Billy' Anderson"  
  2: "George Barnes"  
  3: "Justin D. Barnes"  
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)  
  0: "English"  
released: ISODate("1903-12-01T00:00:00.000+00:00")  
directors: Array (1)  
  0: "Edwin S. Porter"  
rated: "TV-G"  
awards: Object  
  wins: 1  
  nominations: 0
```

The bottom status bar shows system information: ENG IN, 01:58 PM, 17/03/2025.

5.2.17.

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

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar displays connections and collections, with 'sample_mflix.movies' selected. The main area shows a search bar with the query: { "tomatoes.viewer.rating": { "\$lt": 6.0 } }. Below the search bar, the results are displayed as a list of movie documents. One document is expanded to show its full structure, including fields like _id, plot, genres, cast, languages, released, directors, rated, awards, and nominations. The document ID is `_id: ObjectId('573a1390f29313caabcd42e8')`. The plot summary is: "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." The genres are "Short" and "Western". The runtime is 114 minutes. The cast includes "A.C. Abadie", "Gilbert M. 'Broncho Billy' Anderson", "George Barnes", and "Justin D. Barnes". The poster URL is https://m.media-amazon.com/images/M/MVSBMTU3NjE5NzYtYTYYNS00MDVmLWiwYj_. The title is "The Great Train Robbery". The full plot summary is: "Among the earliest existing films in American cinema - notable as the ..." The language is English. It was released on 1903-12-01T00:00:00.000+00:00. The director is Edwin S. Porter. The rating is TV-G. There are no awards listed. The document has 1 win and 8 nominations.

5.2.18.

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

The screenshot shows the MongoDB Compass interface connected to a cluster. The left sidebar lists connections and collections, with 'sample_mflix' selected. The main area shows a search query builder for the 'movies' collection, displaying the aggregation pipeline from the previous step. Below the query, a specific movie document is expanded, showing fields like _id, fullplot, imdb, rating, votes, id, year, plot, genres, title, lastupdated, languages, and writers. The bottom status bar shows system information and the date/time.

```
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)
```

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

```
_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 the 'sample_mflix' connection selected and its 'movies' collection highlighted. The main pane shows a query builder with the following query:

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

The results pane shows a list of documents from the 'movies' collection, with the first document partially visible:

```
_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 Morton (novel)"  
awards: Object
```

The bottom status bar shows system information like battery level, signal strength, and the date/time (02:00 PM, 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 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 shows a complex query with nested objects and arrays. The results pane displays a single document's details, including its ID, plot, genres (Romance, Drama), runtime (84), rating (PASSED), cast (Norma Shearer, Chester Morris, Conrad Nagel, Robert Montgomery), and poster URL. The document also includes full plot, languages (English, French), and release date (1930-04-19T00:00:00.000+00:00). The bottom status bar shows system information like battery level, signal strength, and the date/time (17/03/2025).

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 title bar reads "MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies". The main window displays a modal dialog titled "Update 20930 documents" for the "sample_mflix.movies" collection. The "Filter" field contains the query: { runtime: { \$ne: null } }. The "Update" field contains the update operation: { "\$addToSet": { "awards.wins": 1 } }. The "Preview" section shows a sample document structure with fields like tomatoes.viewer.meter, tomatoes.fresh, tomatoes.critic.rating, tomatoes.critic.numReviews, tomatoes.critic.meter, tomatoes.rotten, tomatoes.lastUpdated, num_mflix_comments, genres, cast, languages, directors, countries, awards (an object with wins: Array), and _id. The "Save" and "Update 20930 documents" buttons are at the bottom of the modal. The background shows the navigation pane with connections and collections, and the bottom status bar indicates the system is in English, battery level, and the date/time.

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 a 'sample_mflix' database selected. The 'movies' collection is currently being edited. A modal window titled 'Update 9442 documents' is open, containing the following update query:

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

The 'Preview' section shows an example of three documents. One document is expanded to show its fields and values, including the '_id' field, plot, genres (which includes 'Fantasy'), runtime, metacritic, cast, poster, title, fullplot, languages, released, directors, writers, awards, lastupdated, year, imdb, and countries. The 'Update 9442 documents' button at the bottom right of the modal is highlighted in green.

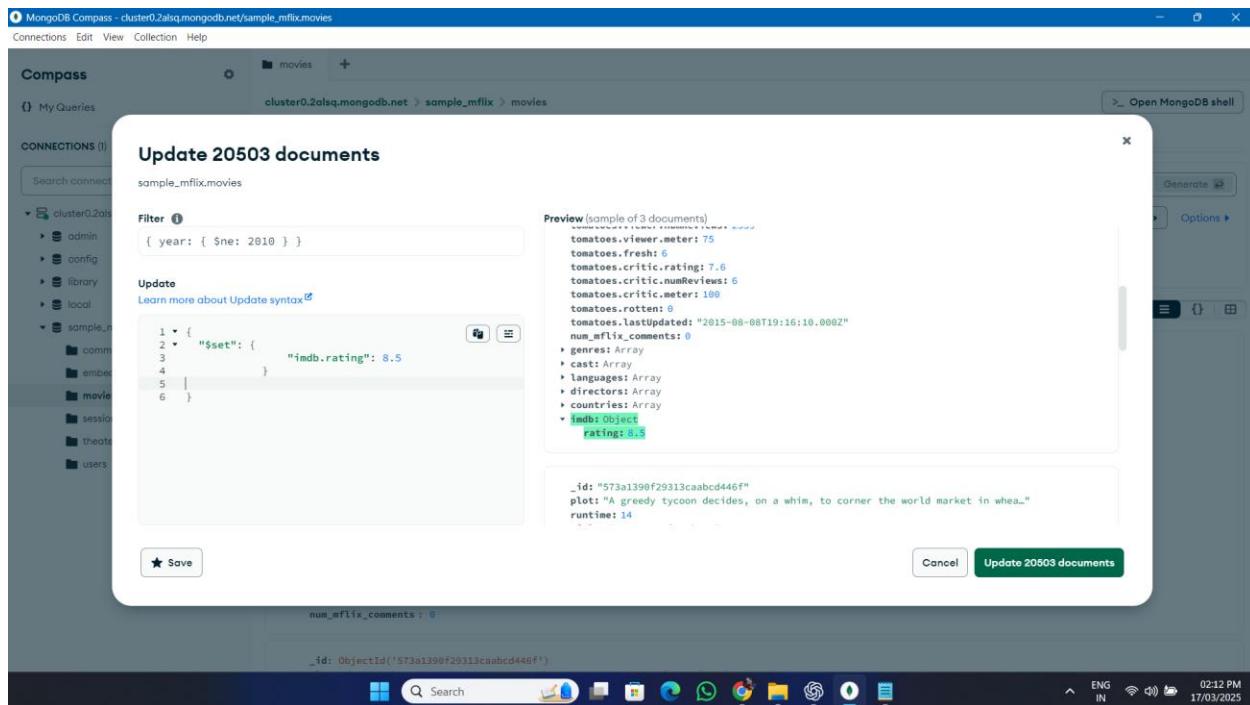
5.3.3.

```
{  
  "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 area displays a modal dialog titled 'Update 18566 documents' for the 'sample_mflix.movies' collection. The 'Filter' field contains the query: '{ "tomatoes.viewer.rating": { \$lt: 6 } }'. The 'Update' field contains the update operation: '1. { "\$set": { "imdb.rating": 8.5 } }'. A preview section shows the structure of three sample documents, including fields like '_id', 'plot', 'title', 'year', 'imdb', 'languages', 'countries', and 'tomatoes'. At the bottom of the modal are 'Save' and 'Update 18566 documents' buttons. The background of the Compass interface shows the database structure with collections like 'movies', 'sample_mflix', and 'sample_movies'.

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. On the left, the 'Filter' section shows a query: '{ countries: { \$nin: ['USA'] } }'. Below it, the 'Update' section shows the command: '1 { 2 \$pull: { 3 genres: "Drama" 4 } 5 } 6'. To the right, a 'Preview' window shows a sample of three documents, and at the bottom, there are 'Save' and 'Update 10428 documents' buttons. The status bar at the bottom right shows system information like battery level, signal strength, and the date/time (17/03/2025).

5.3.6.

```
{  
  "filter": {  
    "countries": {  
      "$nin": [  
        "USA"  
      ]  
    },  
    "update": {  
      "$set": {  
        "updated_at": "2025-03-16"  
      }  
    },  
    "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', and 'Help'. Below the navigation is a toolbar with icons for 'New Connection', 'New Collection', 'Open MongoDB shell', 'Generate', and 'Options'.

The main area displays a modal titled 'Update 10428 documents' for the collection 'sample_mflix.movies'. The modal has two sections: 'Filter' and 'Update'.

Filter: A dropdown menu shows the query: '{ countries: { \$nin: ['USA'] } }'.

Update: The update command is: { '\$set': { 'updated_at': "2025-03-16" } }.

Preview: A preview section shows a sample of three documents. One document is expanded to show its fields: _id, num_mflix_comments, _id, plot, genres, runtime, cast, num_mflix_comments, poster, and updated_at.

Buttons: At the bottom of the modal are 'Save', 'Cancel', and a large green 'Update 10428 documents' button.

The bottom of the screen shows the Windows taskbar with icons for Search, Start, Task View, File Explorer, Edge, WhatsApp, Google Chrome, and others. The system tray shows battery level, signal strength, and the date and time (17/03/2025, 02:16 PM).

5.3.7.

```
{  
  "filter": {  
    "directors": {  
      "$exists": true  
    }  
  },  
  "update": {  
    "$set": {  
      "imdb.rating": 8.5  
    }  
  },  
  "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', and 'Help'. Below the navigation is a toolbar with icons for 'movies' (selected), 'collection', 'query', and 'shell'. The left sidebar shows 'Compass' and 'sample_mflix' collections, with 'movies' selected. The main area displays an 'Update 21127 documents' dialog. The 'Filter' field contains the query: { directors: { \$exists: true } }. The 'Update' field contains the update script:
1 ▼ {
2 ▼ \$set: {
3 "imdb.rating": 8.5
4 },
5 },
6 }
The 'Preview' section shows a sample of three documents from the 'movies' collection, including fields like _id, tomatoes, num_mflix_comments, genres, cast, languages, directors, countries, and an expanded imdb object with rating: 8.5. At the bottom of the dialog are 'Save' and 'Cancel' buttons, and a large green 'Update 21127 documents' button. The bottom right corner of the screen shows a taskbar with various application icons and the date/time: 02:17 PM, 17/03/2025.

5.3.8.

```
{  
  "filter": {  
    "runtime": {  
      "$gte": 120  
    }  
  },  
  "update": {  
    "$push": {  
      "genres": "New Genre"  
    }  
  },  
  "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 3651 documents' for the 'sample_mflix.movies' collection. The 'Filter' field contains the query: '{ runtime: { \$gte: 120 } }'. The 'Update' field contains the update operation: '1. { 2. \$push: { 3. "genres": "New Genre" 4. } 5. }'. A preview window on the right shows sample document data, including fields like 'awards.text', 'lastupdated', 'year', 'imdb.rating', 'imdb.votes', 'imdb.id', 'type', 'tomatoes.viewer.rating', 'tomatoes.viewer.numReviews', 'tomatoes.lastUpdated', 'num_mflix_comments', 'tomatoes.production', 'genres' (containing 'Action' and 'New Genre'), and arrays for 'cast', 'languages', 'directors', 'countries', and 'writers'. At the bottom of the modal are 'Save', 'Cancel', and 'Update 3651 documents' buttons. The status bar at the bottom indicates the document ID as '_id: ObjectId('573a1391f29313caab7064')' and the system time as '02:18 PM 17/03/2025'.

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 the 'movies' collection. A modal dialog titled 'Update 10428 documents' is open, containing the following update query:

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

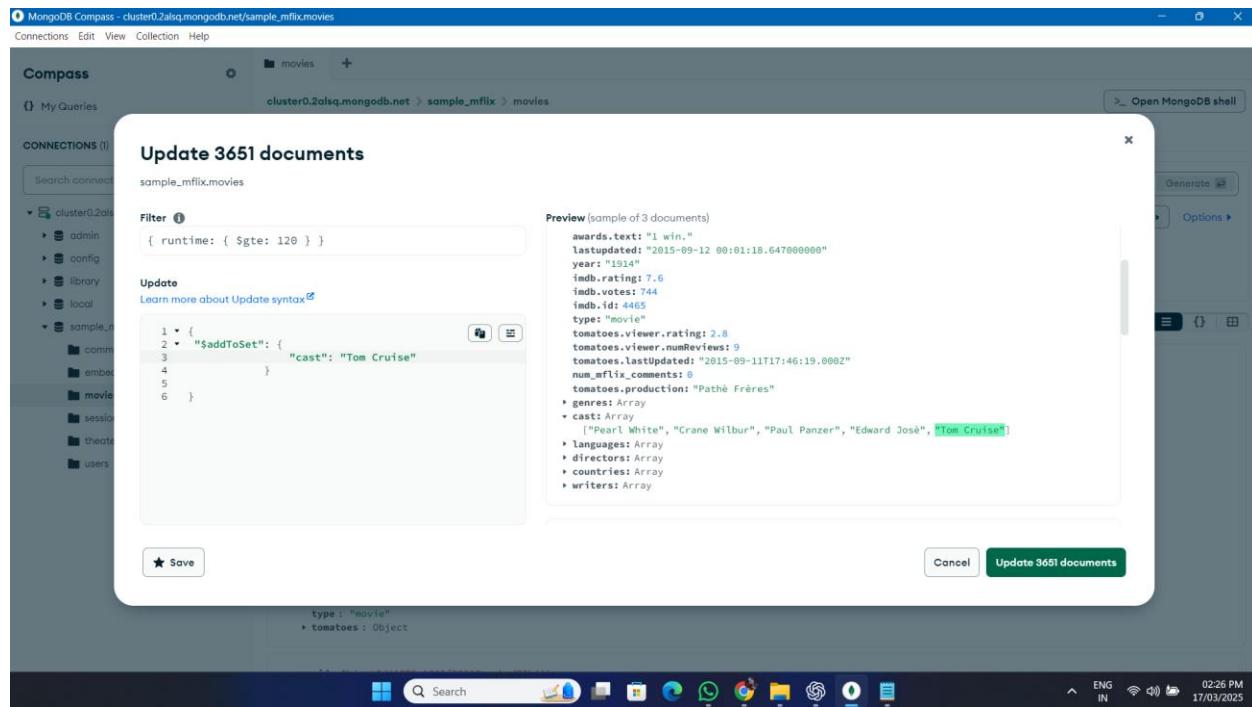
The 'Preview' section shows a sample of three documents. One document is expanded to show its fields and values:

```
_id: ObjectId('573a1391f29313caabcd7db6')  
plot: "In Paris a wild girl becomes possessed by the soul of her twin who die..."  
+ genres: Array  
  ["Drama", "New Genre"]  
+ runtime: 82  
+ cast: Array  
  title: "White Shadows"  
  fullPlot: "In Paris a wild girl becomes possessed by the soul of her twin who die..."  
+ languages: Array  
  released: 1924-05-05T00:00:00.000Z  
+ directors: Array  
+ writers: Array  
+ awards: Object  
  lastupdated: "2015-05-10 00:19:38.277000000"  
  year: 1924  
+ imdb: Object  
+ countries: Array  
  type: "movie"  
+ tomatoes: Object  
  num_mflix_comments: 0
```

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 like battery level, signal strength, and date/time.

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 with the following details:

- Connections:** cluster0.2alsq.mongodb.net / sample_mflix.movies
- Collection:** movies
- Update Dialog:** "Update 527 documents" for sample_mflix.movies
- Filter:** { \$and: [{ year: { \$gte: 2000 } }, { genres: 'Sci-Fi' }] }
- Update Script:** { "\$set": { updated_at: "2025-03-16" } }
- Preview:** Shows a sample of 3 documents with fields like lastupdated, languages, writers, type, tomatoes, poster, num_mflix_comments, released, awards, countries, cast, directors, runtime, and updated_at.
- Buttons:** Save, Cancel, and a large green "Update 527 documents" button.
- Bottom Status Bar:** num_mflix_comments: 134, released: 2005-07-08T00:00:00+00:00, awards: Object, countries: Array (2).
- System Icons:** Search bar, taskbar icons (File Explorer, Edge, File, Mail, Photos, Task View, Taskbar Icons, Taskbar Buttons), and system status indicators (ENG IN, 02:27 PM, 17/03/2025).

5.3.12.

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

The screenshot shows the MongoDB Compass interface. The title bar reads "MongoDB Compass - cluster0.2alsq.mongodb.net/sample_mflix.movies". The left sidebar shows connections and collections, with "sample_mflix.movies" selected. A modal dialog titled "Update 2181 documents" is open. It contains a "Filter" field with the query: { year: { \$gte: 1980, \$lte: 1990 } }. Below it is an "Update" field containing the update operation: 1. { 2. "\$push": { 3. "genres": "New Genre" 4. } 5. }. To the right, a "Preview" section shows a sample of three documents. One document 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 modal are "Save" and "Cancel" buttons, and a large green "Update 2181 documents" button.

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 database structure with 'sample_mflix' selected. The main area shows the 'movies' collection. A modal window titled 'Update 20503 documents' is open, containing the update query from the previous code block. The 'Filter' field contains the query '{ year: { \$ne: 2010 } }'. The 'Update' field contains the update object: { "\$set": { "status": "updated" } }. To the right, a 'Preview' section shows a sample of three documents, all of which have their 'status' field updated to 'updated'. At the bottom of the modal are 'Save' and 'Update 20503 documents' buttons, along with 'Cancel' and 'Generate' buttons. The status bar at the bottom indicates the operation will affect 20503 documents.

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' for the 'movies' collection. The 'Filter' field contains the query: { cast: { \$in: ['Tom Hanks', 'Robert Downey Jr.'] } }. The 'Update' field contains the update operation: { "\$addToSet": { "imdb.top": 1 } }. A preview window on the right shows three sample documents. One document's 'imdb' field is expanded, showing it has a 'top' array containing the value 1. The bottom of the dialog has a 'Save' button and an 'Update 72 documents' button. The status bar at the bottom indicates the system is ENG IN, the date is 17/03/2025, and the time is 02:37 PM.

5.3.15.

```
{  
  "filter": {  
    "languages": {  
      "$all": [  
        "English",  
        "French"  
      ]  
    }  
  },  
  "update": {  
    "$pull": {  
      "genres": "Romance"  
    }  
  },  
  "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'. A modal window titled 'Update 1247 documents' is open. On the left, the 'Filter' section contains the query: { languages: { \$all: ['English', 'French'] } }. On the right, the 'Preview' section shows a sample of three documents. One document is expanded to show its full structure:

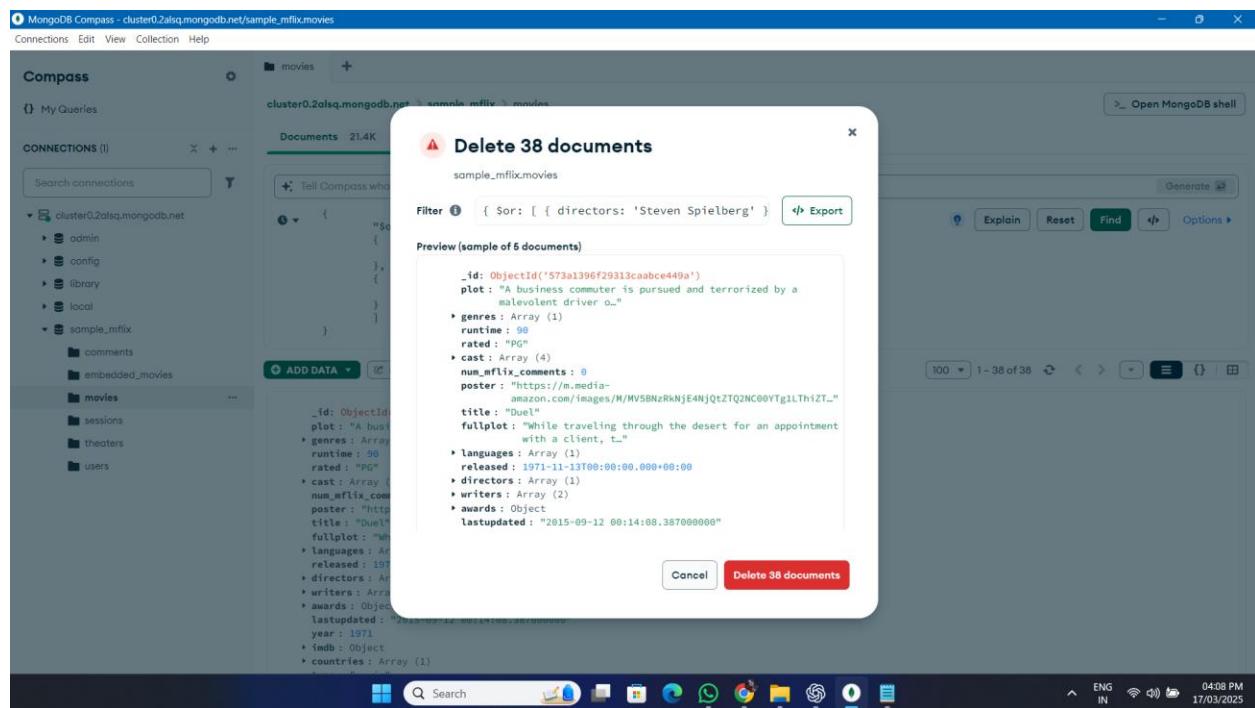
```
_id: ObjectId('573a1391f29313caabcd9264')  
plot: "When a woman discovers that her husband has been unfaithful to her, sh..."  
genres: Array  
  ["Romance", "Drama"]  
runtime: 84  
rated: "PASSED"  
cast: Array  
num_mflix_comments: 0  
poster: "https://m.media-amazon.com/images/M/MV5BMTk5NzQzNjEwNV5Bml5BanBnXkFtZT_..."  
title: "The Divorcee"  
fullPlot: "Jerry and Ted are young, in love, and part of the New York 'in-crowd'..."  
languages: Array  
released: 1930-04-19T00:00:00.000+00:00  
directors: Array  
writers: Array  
awards: Object  
lastupdated: "2015-08-25 00:14:05.103000000"  
year: 1930  
imdb: Object  
countries: Array
```

At the bottom of the modal, there are 'Save' and 'Update 1247 documents' buttons. The status bar at the bottom of the screen shows the date and time as 17/03/2025 and 02:40 PM.

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. The preview includes 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' 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 area, the 'Documents' tab shows 21.4K documents. A modal dialog titled 'Delete 9070 documents' is open over the interface. The dialog contains a warning icon and the text 'Delete 9070 documents'. It shows a preview of 5 documents with their IDs and titles. One document is expanded to show its full structure, including fields like _id, plot, genres, runtime, metacritic, rated, cast, poster, title, fullplot, languages, released, directors, writers, awards, lastupdated, year, imdb, and countries. At the bottom of the dialog are 'Cancel' and 'Delete 9070 documents' buttons. The background of the interface shows the mongo shell command: { \$and: [{ year: { \$gt: 1995 } }, { year: { \$lt: 2010 } }] }.

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 databases, with 'sample_mflix' selected. In the center, the 'movies' collection is shown with 21.4K documents. A modal dialog titled 'Delete 72 documents' is open, showing a preview of one document. The document details are as follows:

```
_id: ObjectId('573a1398f29313caabce943a')  
plot: "A young man is reunited with a mermaid who saves him from drowning as ..."   
genres: Array (3)  
runtime: 111  
rated: "PG"  
cast: Array (4)  
poster: "https://m.media-amazon.com/images/M/MV5BZDRmZTQ0MDQtNWRYSG0NDhhLTgzNz.."  
title: "Splash"  
fullplot: "Allie Bauer is rescued from drowning as a young boy off Cape Cod by a ..."   
languages: Array (2)  
released: 1984-03-09T00:00:00+00:00  
directors: Array (1)  
writers: Array (5)  
awards: Object  
lastupdated: "2015-08-15 00:27:58.600000000"  
year: 1984
```

At the bottom of the dialog are 'Cancel' and 'Delete 72 documents' buttons. The background shows the main Compass interface with various toolbars and a status bar at the bottom indicating network and system information.

5.4.5.

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

The screenshot shows the MongoDB Compass interface. On the left, the connection tree displays 'sample_mflix' and its collections: 'comments', 'embedded_movies', 'movies', 'sessions', 'theaters', and 'users'. The 'movies' collection is selected, showing 21.4K documents. A search bar at the top is empty. In the center, a modal dialog titled 'Delete 8398 documents' is open, showing a preview of five documents. The preview includes fields like '_id', 'plot', 'genres', 'cast', 'runtime', 'languages', 'released', 'directors', 'writers', 'awards', and 'lastupdated'. At the bottom of the dialog are 'Cancel' and 'Delete 8398 documents' buttons. The background shows the main Compass interface with various toolbars and status indicators at the bottom.

5.4.6.

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

The screenshot shows the MongoDB Compass interface. On the left, the connection tree displays 'sample_mflix' under 'sample_mflix'. In the center, the 'movies' collection is selected. A modal dialog titled 'Delete 21127 documents' is open, showing a preview of 5 documents. The filter used is '{ directors: { \$exists: true } }'. The preview shows a single document's fields: _id, plot, genres, runtime, cast, languages, released, director, awards, lastupdated, year, and imdb. The document's plot field is described as 'A group of bandits stage a brazen train hold-up, only to find a determ...'. The 'Delete 21127 documents' button at the bottom right of the dialog is highlighted in red. The status bar at the bottom right shows 'ENG IN' and the date '17/03/2025'.

5.4.7.

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

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 1703 documents' is open, showing a preview of five documents. The preview includes a JSON snippet of one document and a warning message: 'A kept woman runs into her one-time fiance and finds herself torn between them.' The dialog has 'Cancel' and 'Delete 1703 documents' buttons. The status bar at the bottom right shows '04:13 PM 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 modal dialog titled 'Delete 2181 documents' is open. The dialog contains a preview of five documents matching the filter: { year: { \$gte: 1980, \$lte: 1990 } }. One document is expanded to show its full JSON structure, including fields like _id, plot, genres, runtime, cast, num_mflix_comments, poster, title, fullplot, languages, released, directors, writers, awards, lastupdated, year, and type. At the bottom of the dialog are 'Cancel' and 'Delete 2181 documents' buttons. The background shows the Compass interface with a list of movies and a toolbar with various options like Explain, Reset, Find, and Options.

5.4.9.

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

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. A search bar at the top contains the query: { _id: '573a1390f29313caabcd5ea4' }. Below the search bar, the results table shows one document with the following details:

Document	Value
_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	8
tomatoes.production	"MCA/Universal Pictures"
genres	Array (1)
cast	Array (4)
languages	Array (1)

5.4.10.

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

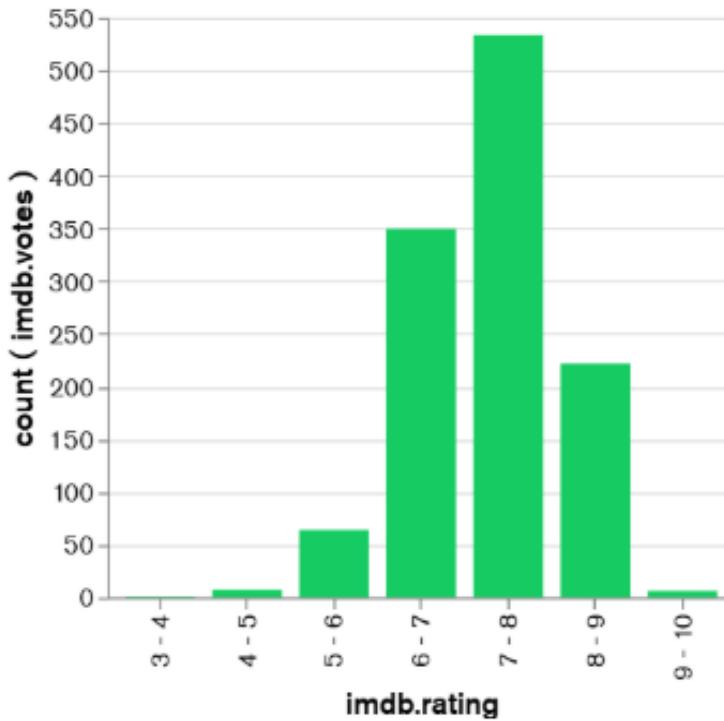
The screenshot shows the MongoDB Compass interface. On the left, the sidebar displays connections and databases, with 'sample_mflix' selected. Under 'sample_mflix', the 'movies' collection is highlighted. In the main area, a search bar and filter dropdown are visible. A modal dialog box titled 'Delete 1247 documents' is open, showing a preview of five documents that match the filter. The preview includes fields like '_id', 'plot', 'genres', 'rated', 'cast', 'num_mflix_comments', 'poster', 'title', 'fullplot', 'languages', 'released', 'directors', 'writers', 'awards', and 'lastupdated'. At the bottom of the dialog are 'Cancel' and 'Delete 1247 documents' buttons. The status bar at the bottom right shows system information: ENG IN, 04:17 PM, 17/03/2025.

5.5. Key Observations and Insights

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

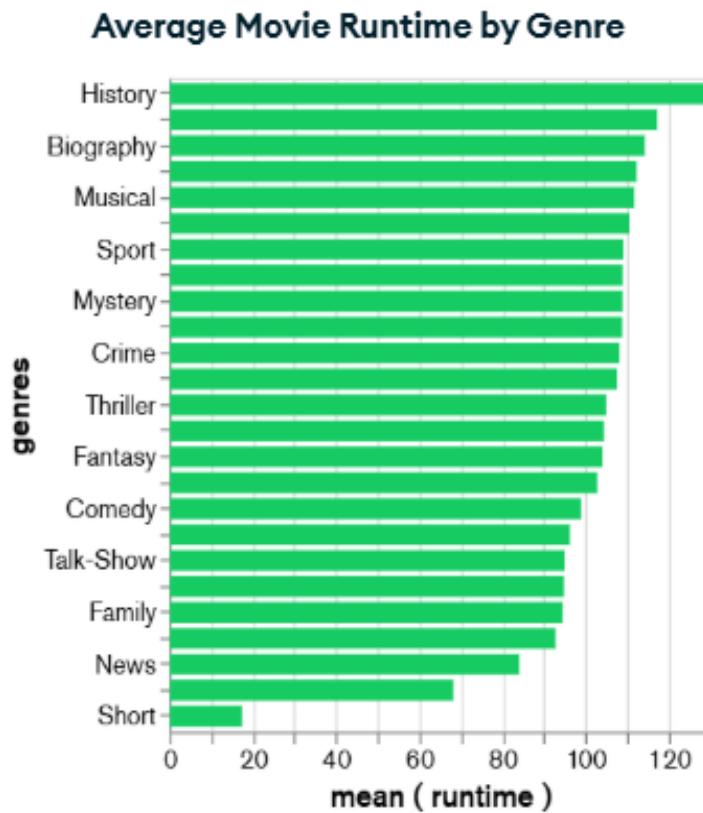
- Most movies fall within the **7.0–8.0 rating range**, indicating that audiences generally rate popular movies positively.
- A significant drop is observed for movies rated **below 6.0**, showing that fewer widely popular movies receive lower ratings.
- Very few movies have ratings below **4.0 or above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

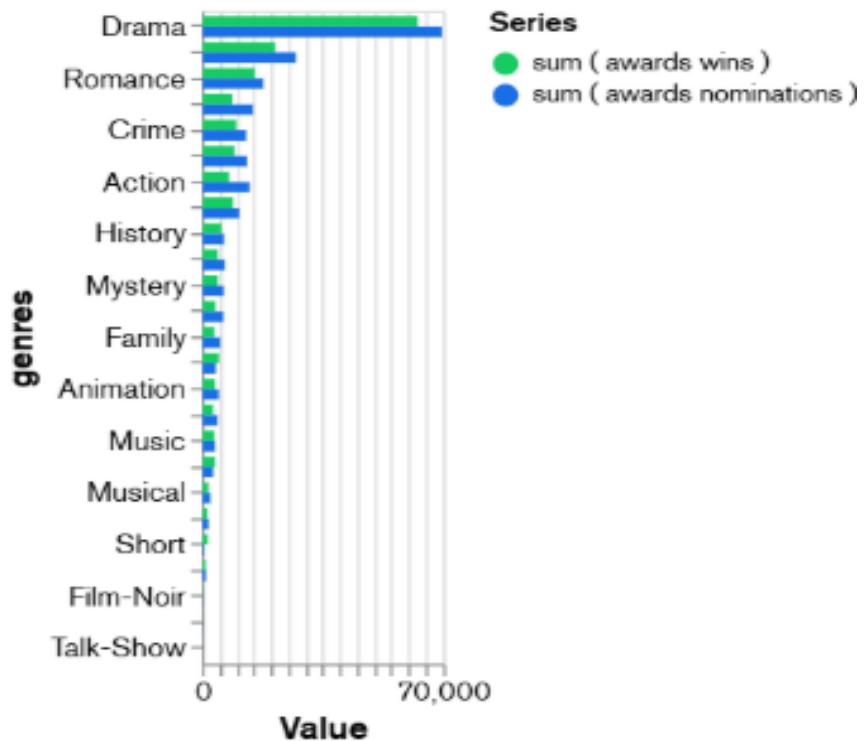
- **Historical movies have the longest runtime**, averaging over 120 minutes, which aligns with the genre's need for extensive storytelling.
- **Short films have the lowest runtime**, as expected, reinforcing the distinction in duration between different formats.
- Biographical, musical, and sports genres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

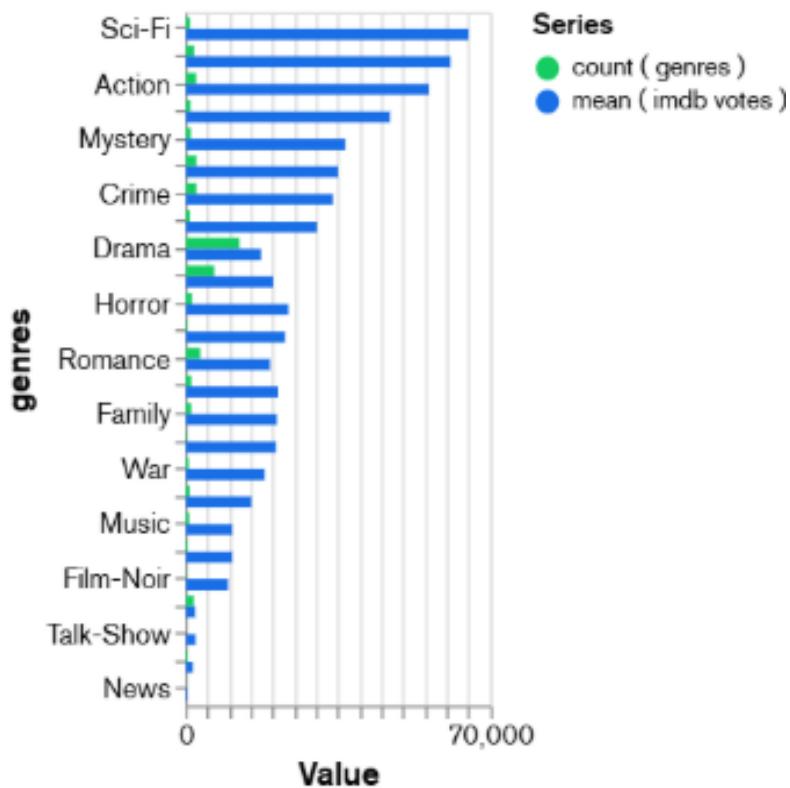
Award Wins and Nominations by Genre: A ...



5.5.4. Comparing Average IMDb Votes Across Different Film Genres

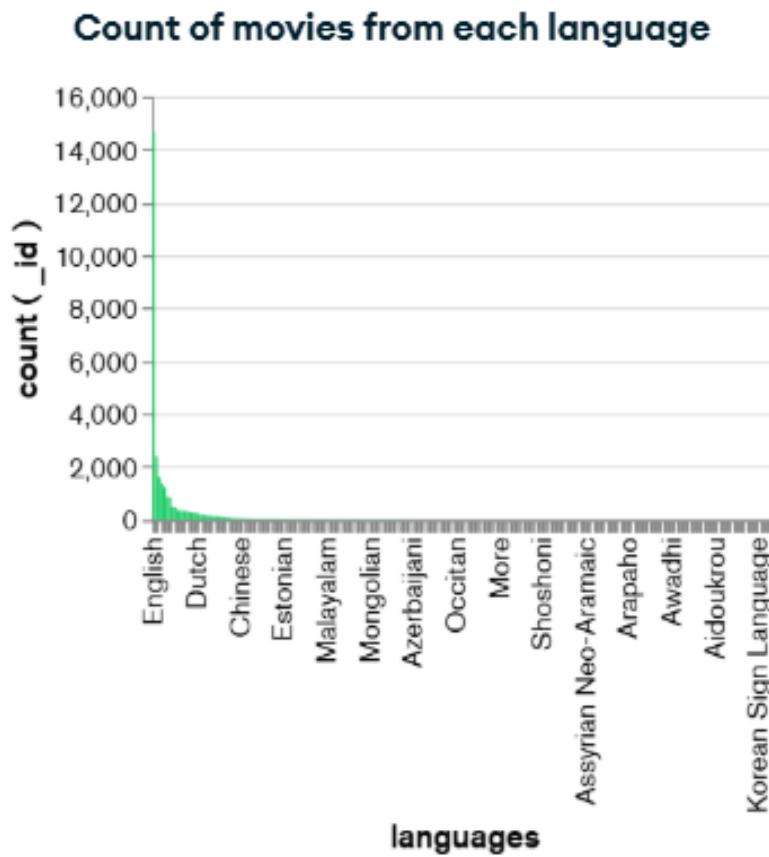
- **Sci-Fi and Action** genres receive the **highest average IMDb votes**, suggesting these genres attract a larger audience and fan engagement.
- **Drama, Horror, and Mystery** also perform well, showing that audience interest spans across both intense and emotional storytelling.
- **Genres like Talk-Show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

Comparing Average IMDb Votes Across Di...



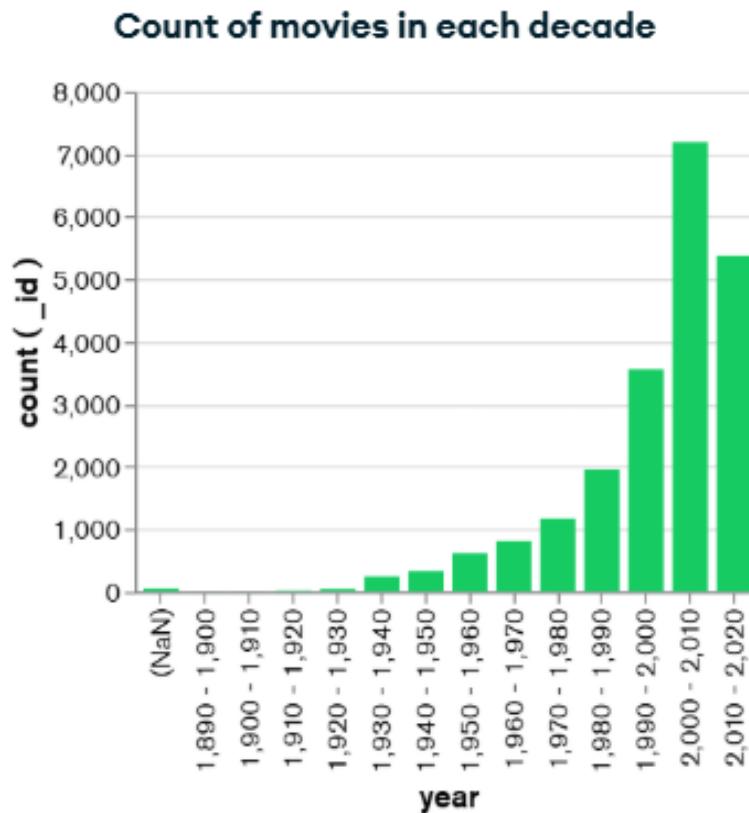
5.5.5. Count of Movies from Each Language

- **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

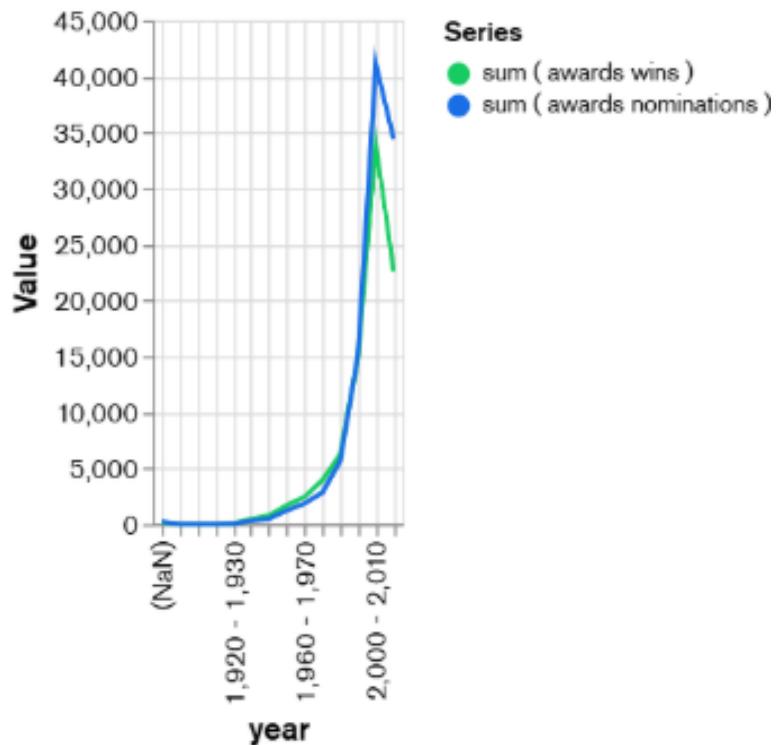
- The number of movies has **increased drastically over time**, with the highest count in the **2000s and 2010s**.
- Early decades (1890–1950) have **very few movies**, indicating limited production capacity and technology.
- The rise in movie production **post-1980** aligns with technological advancements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

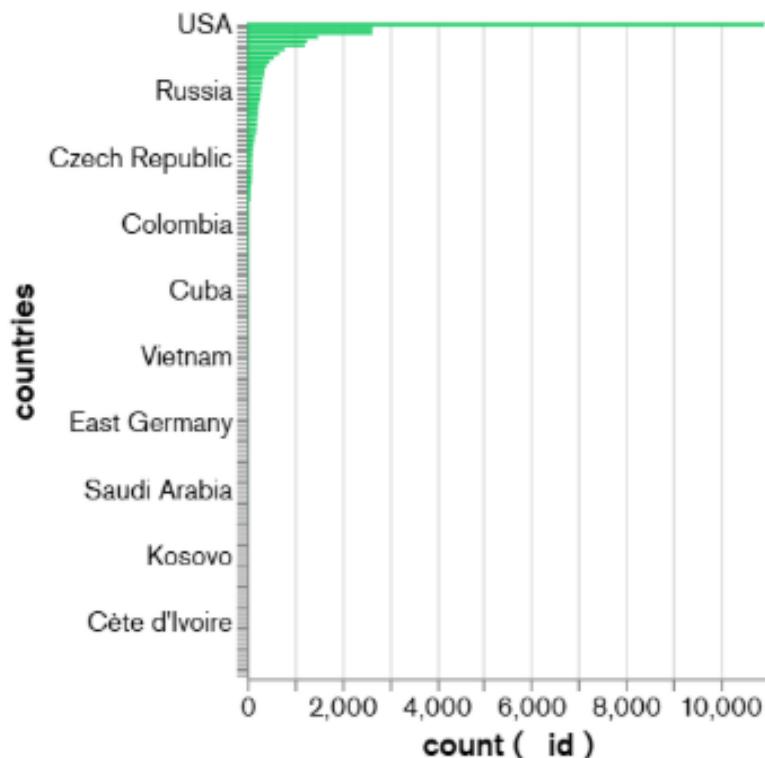
Decadal Trends in Movie Wins and Nomina...



5.5.8. Distribution of Movie Counts Across Different Countries

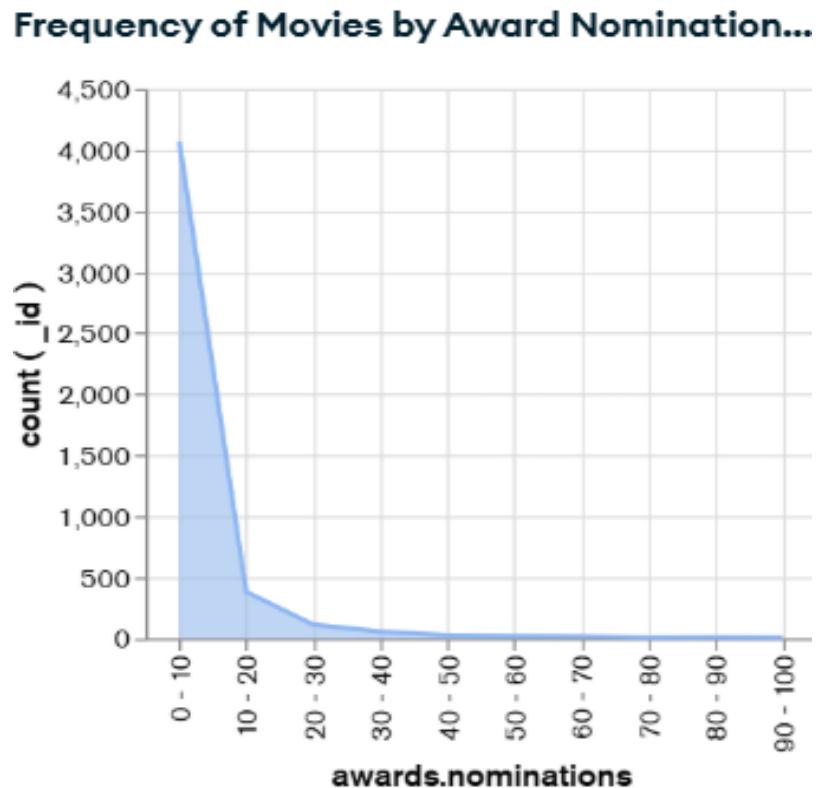
- The USA has an overwhelmingly high number of movies compared to other countries.
- Russia follows, but no other country comes close, highlighting Hollywood's **dominant influence** in film production.

Distribution of Movie Counts Across Differ...



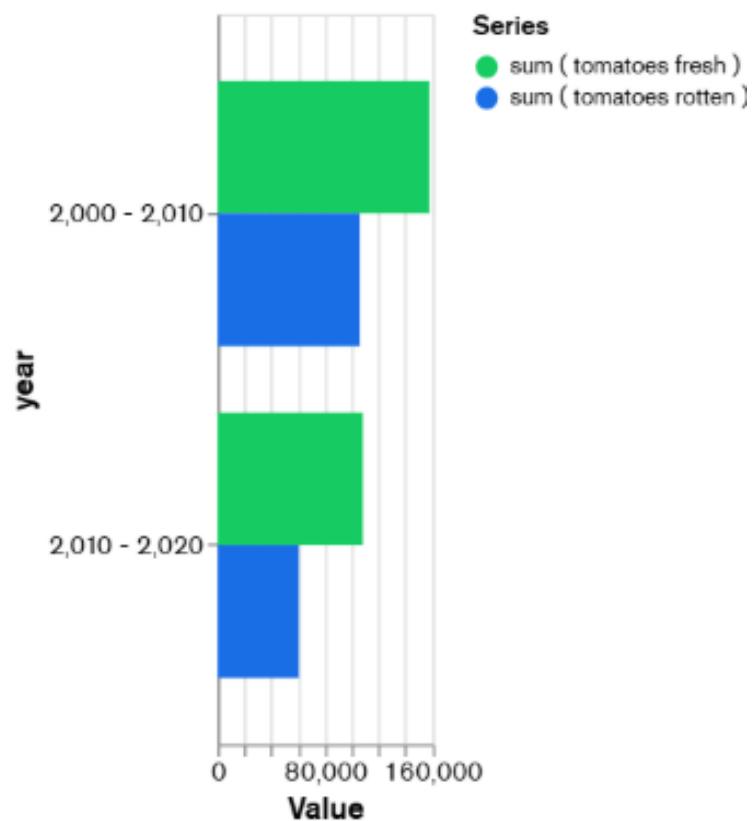
5.5.9. Frequency of Movies by Award Nominations Count

- 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)

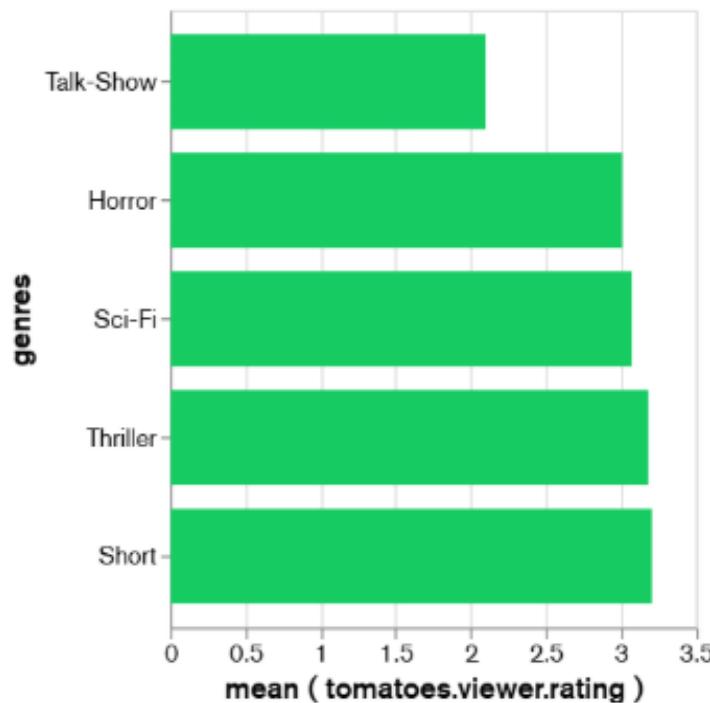
- More movies received "Fresh" ratings than "Rotten," suggesting a general tendency for movies to be rated positively.
- However, there is still a **substantial number** of negatively rated movies.



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

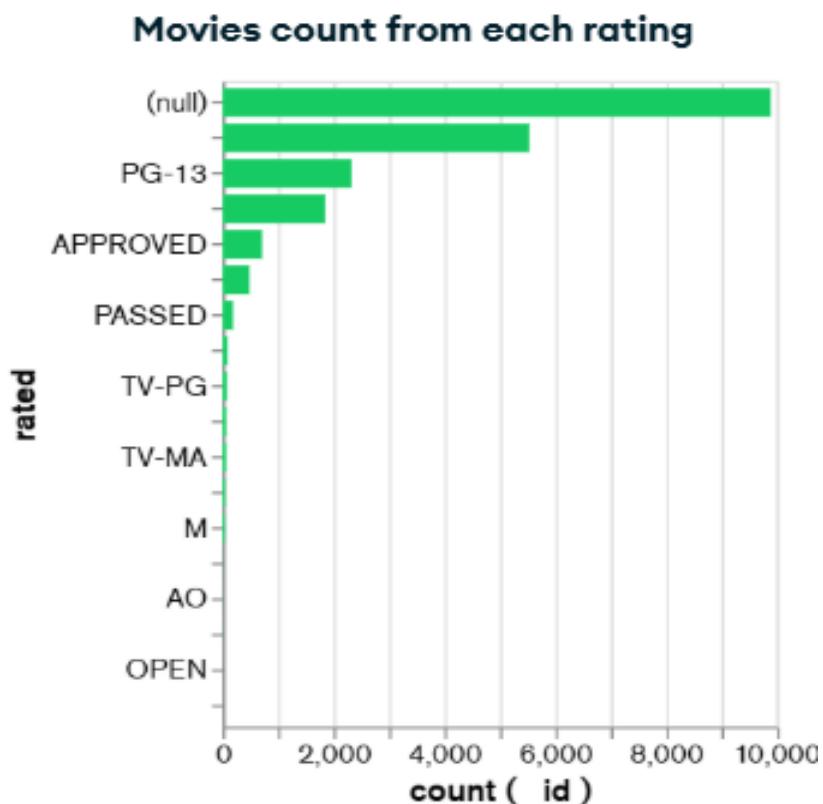
- The majority of movies (over **19,000**) have won between **0-10 awards**.
- Only **a few hundred** movies have won more than **20 awards**, making highly awarded films extremely rare.

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

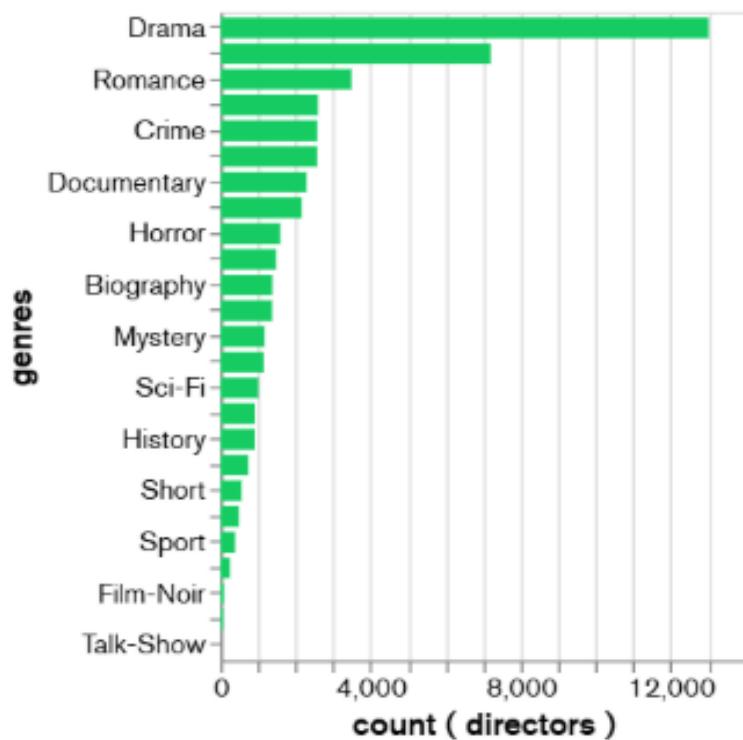
- A significant number of movies **lack a specified rating** (`null`), indicating missing or unclassified rating data.
- The most common explicit rating is **PG-13**, suggesting that a large portion of movies cater to a wide audience.
- Older rating systems like **APPROVED** and **PASSED** have significantly 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.

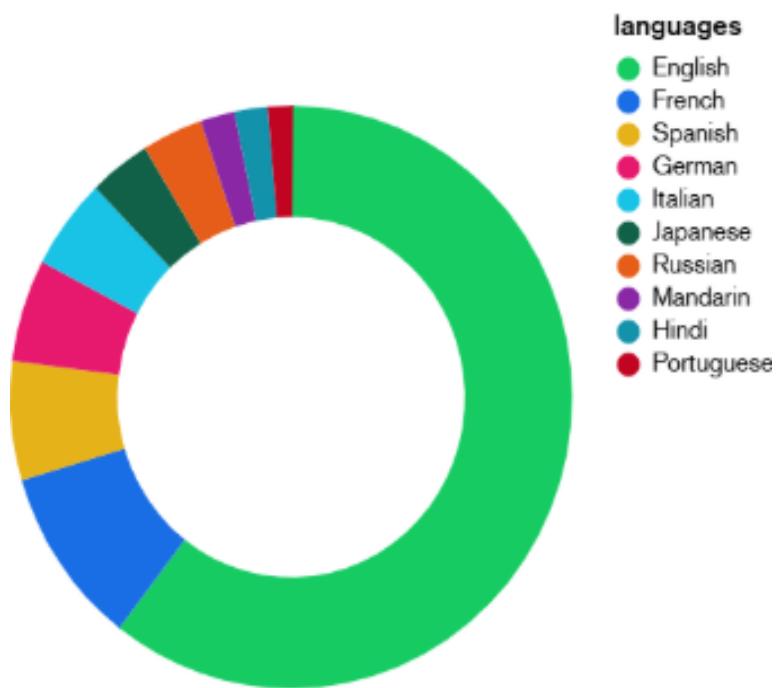
Number of directors in each movie category



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.

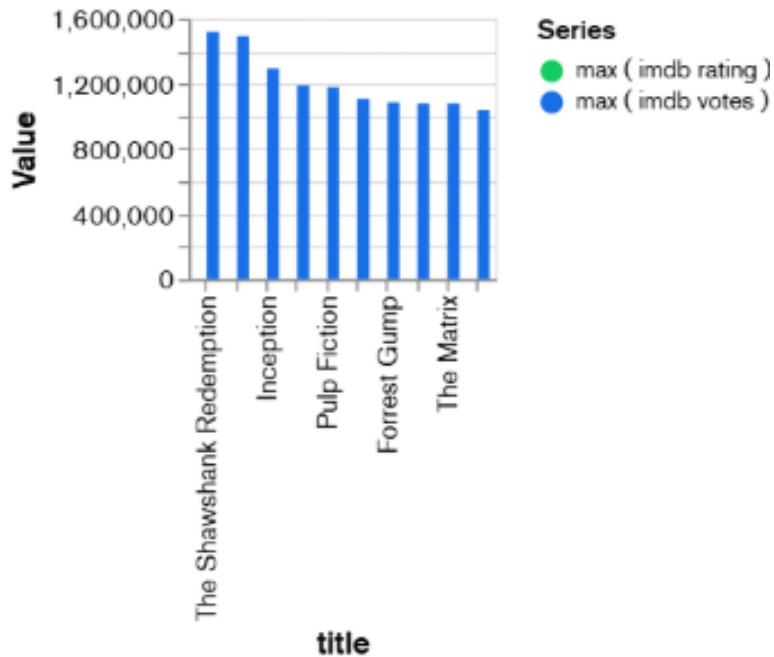
Share of top 10 most popular languages in ...



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.

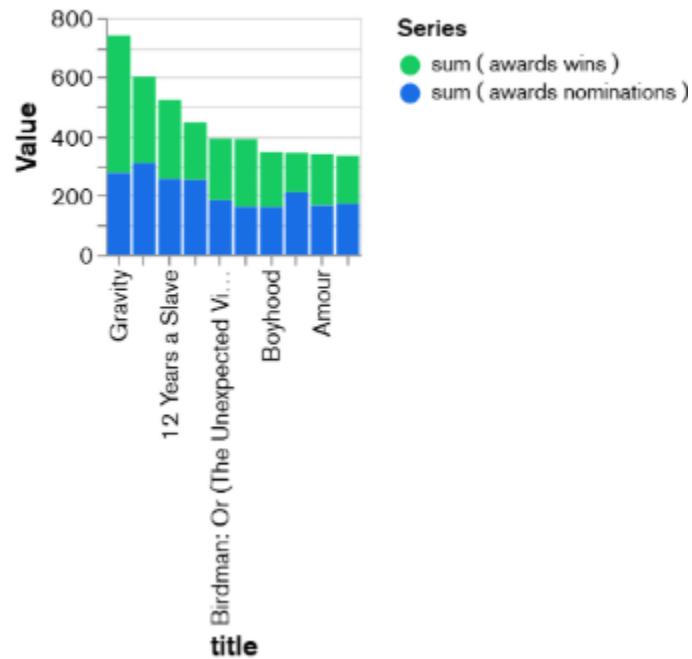
Top 10 Highest Rated Movies with Most Vot...



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

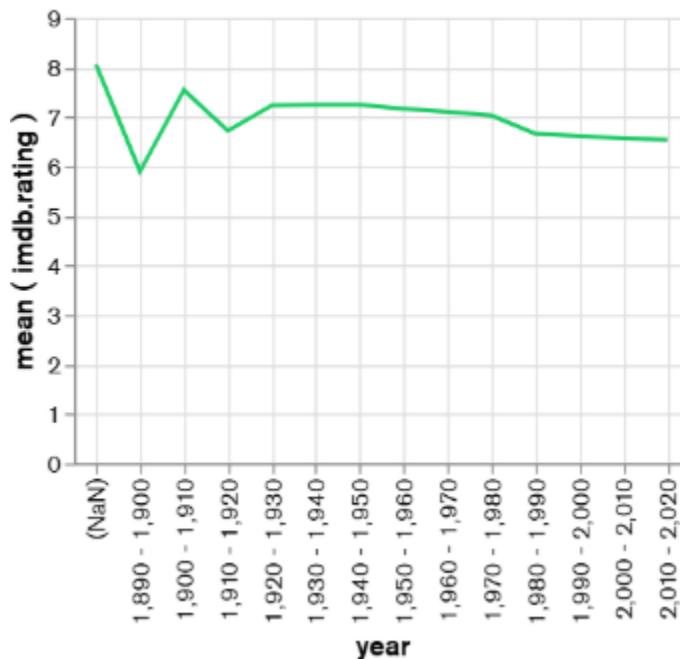
title ↑↓	max (imdb.rat...)
éxtasis	6.9
évocateur: The Morton Downey Jr. Movie	6.8
éríðo Movie	6.7
éon Flux	5.5
élisa	6.7
ékte vare	6.2
éke och hans värld	6.7
édipussi	7.6
éber uns das All	6.7
éVivan las Antípodas!	7.1

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...

