

IMDB Movies SQL Data Analysis Project

INTRODUCTION

The goal of this project is to analyze the IMDB Movies dataset using SQL and derive meaningful insights about movies, directors, popularity, profitability, gender distribution, and film industry patterns.

The dataset consists of two tables:

1. Directors
2. Movies

Using SQL queries, we explored trends such as the most popular movies, directors with the highest contributions, female director representation, bankability (profit), and top-rated movies since the year 2000.

A final interactive dashboard was created to visualize key patterns and support better storytelling and decision-making.

DATASET OVERVIEW

1.Directors Table

- Column Description
- ID Unique ID of the director
- Name Director's name
- Gender 0/2 = Male, 1 = Female
- Department Film Department

2.Movies Table

- Column Description
- ID Unique Movie ID
- Title Movie title
- Original Title Original movie name
- Budget Movie budget
- Revenue Box office revenue
- Popularity IMDB popularity score
- Vote Average IMDB rating
- Vote Count Number of votes

- Release Date Release date of the movie
- Director ID Foreign key referencing director

To merge the tables, Movies.Director_ID was linked to Directors.ID.

OBJECTIVES AND ANALYSIS

The project aims to answer questions such as:

1. Can you get all data about movies?
2. How do you get all data about directors?
3. Check how many movies are present in IMDB.
4. Find these 3 directors: James Cameron ; Luc Besson ; John Woo
5. Find all directors with name starting with S.
6. Count female directors.
7. Find the name of the 10th first women directors?
8. What are the 3 most popular movies?
9. What are the 3 most bankable movies?
10. What is the most awarded average vote since the January 1st, 2000?
11. Which movie(s) were directed by Brenda Chapman?
12. Which director made the most movies?
13. Which director is the most bankable?

SQL QUERIES

Task1

```
select*from movies;
```

Interpretation:

This query retrieves the entire Movies table. It gives a full overview of each movie, including budget, revenue, ratings, popularity, and release dates. This helps understand the dataset's structure before performing deeper analysis.

Task2

```
select*from directors;
```

Interpretation:

Shows the complete Directors table. This helps identify all directors, their gender distribution, and their unique IDs used to join with the Movies table.

Task3

select count(original_title) as movies from movies;

Movies	47
--------	----

Interpretation:

The dataset contains a total of 47 movies. All further insights—popularity, revenue rankings, director performance—are based on these 47 movies.

Task4

select * from directors where name in ('james cameron','luc besson','john woo');

Name	ID	Gender	UID	Department
James Cameron	4762	2	2710	Directing
John Woo	4893	2	11401	Directing
Luc Besson	4949	2	59	Directing

Interpretation:

The database contains complete details about these three well-known directors. All three are male directors in the Directing department. Their IDs can be used to check their movie contributions.

Task5

select name from directors where name like 'S%';

Sam Mendes	Simon West
Sam Raimi	Stefen Fangmeier
Shane Black	Spike Jonze
Steven Spielberg	Steve Martino

Stephen Sommers	Sergei Bodrov
Shawn Levy	Sydney Pollack
Steve Hickner	Sylvester Stallone
Simon Wells	Seth Gordon
Steven Soderbergh	Scott Derrickson

Interpretation:

The result shows a long list of directors beginning with “S”. This highlights that names starting with S are very common in the film industry, including popular ones like Steven Spielberg and Sam Raimi.

Task6

```
select count( gender) from directors where gender='1';
```

Female(count)	150
---------------	-----

Interpretation:

There are 150 female directors recorded, showing significant female representation in the dataset. This can be used to analyze gender diversity in the directing profession.

Task7

```
select name from directors where gender=1 order by name asc limit 1 offset 9;
```

Name	Amy Holden Jones
------	------------------

Interpretation:

Based on alphabetical sorting of female directors by name, Amy Holden Jones is the 10th listed. This helps verify sorting, indexing, and correct application of ordering in SQL.

Task8

```
select original_title as movie from movies order by popularity desc limit 3;
```

Jurassic World
Captain America: Civil War
Avatar

Interpretation:

These movies have the highest IMDB popularity scores in the dataset. All three are major blockbusters, showing that high-budget, franchise, or sci-fi films tend to attract more viewers and online interest.

Task9

```
select original_title as movie from movies order by revenue desc limit 3;
```

Movies
Avatar
Titanic
The Avengers

Interpretation:

These are the highest-grossing films in the dataset. Notably, two of them (Avatar and Titanic) were directed by James Cameron, showing his strong contribution to box-office success.

Task10

```
select vote_average,release_date from movies where year(release_date)>=2000 order by vote_average desc limit 1;
```

vote_avg	Release date
7.6	16/07/2012

Interpretation:

The highest-rated movie released after 2000 has a rating of 7.6. This result highlights strong audience reception for the film released on 16 July 2012 (likely *The Dark Knight Rises*).

Task11

```
select m.original_title as movie from directors d join movies m on d.id=m.director_id where d.name= 'Brenda Chapman';
```

Interpretation:

There is no movies directed by Brenda Chapman.

Task12

```
select d.name,count(m.id) as total_movie from directors as d join movies as m on  
d.id=m.director_id group by d.id order by total_movie desc limit 1;
```

Name	movie count
Gore Verbinski	3
Sam Raimi	3
Peter Jackson	3

Interpretation:

These three directors share the highest count with 3 movies each. They are highly active and influential directors in big-film franchises (Pirates of the Caribbean, Spider-Man, Lord of the Rings).

Task13

```
select d.name,sum(revenue) as total_revenue from directors as d join movies as m on  
d.id=m.director_id group by d.id order by total_revenue desc;
```

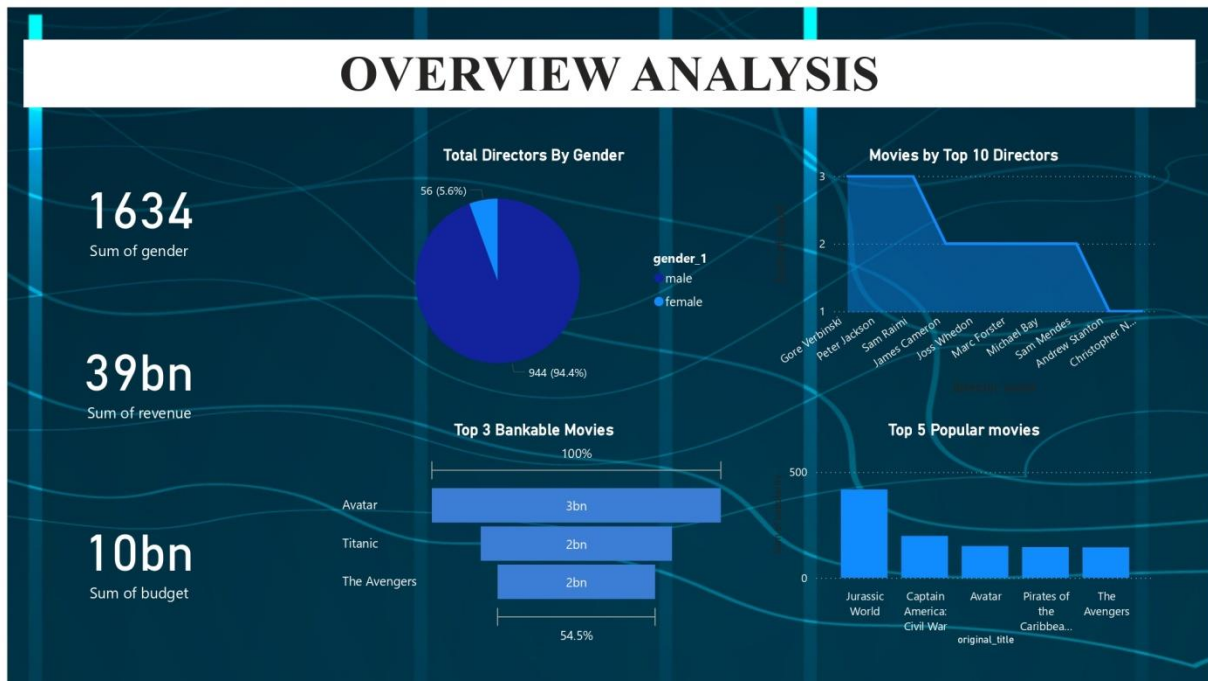
Name	Revenue
James Cameron	4632999275
Joss Whedon	2924961604
Peter Jackson	2464419788
Sam Raimi	2166506515
Gore Verbinski	2115949722
Sam Mendes	1989235622

Interpretation:

James Cameron leads by a huge margin, making him the most financially successful director in this dataset. His movies *Avatar* and *Titanic* contribute significantly to his revenue ranking.

DASHBOARD ANALYSIS.

1.OVERVIEW ANALYSYS DASHBOARD



1. Gender Distribution of Directors

- **Male Directors:** 944 (94.4%)
- **Female Directors:** 56 (5.6%)

Insight:

There is a clear gender imbalance. The industry is dominated by male directors, with female representation being extremely low.

2. Movies by Top 10 Directors

Directors such as Gore Verbinski, Peter Jackson, and Sam Raimi top the list with the highest movie counts.

Insight:

These directors consistently deliver high-budget, commercially successful films, making them major contributors to cinema output.

3. Top 3 Bankable Movies

- Avatar – 3B
- Titanic – 2B
- The Avengers – 2B

Insight:

James Cameron leads as the most bankable director with two of the top-grossing movies ever made. Franchise movies have strong earning potential.

4. Top 5 Popular Movies

- Jurassic World
- Captain America: Civil War
- Avatar
- Pirates of the Caribbean
- The Avengers

Insight:

Popularity is heavily driven by brand loyalty, sequel demand, action/adventure themes, and global fan-bases.

5. Financial Summary

- **Total Revenue:** 39 Billion
- **Total Budget:** 10 Billion

Insight:

The industry shows **4× return on investment**, demonstrating strong profitability and effective commercial play.

2.DIRECTOR ANALYSIS DASHBOARD**1. Director Gender Summary**

- **Total gender entries:** 1634
This ensures broad director coverage within the dataset.

2. Directors With Most Movies

The directors who contributed the most movies include:

- Gore Verbinski
- Peter Jackson
- Sam Raimi
- James Cameron
- Joss Whedon

Insight:

These top directors are strongly associated with blockbuster franchises like Pirates of the Caribbean, Lord of the Rings, Spider-Man, Avatar, and The Avengers. Their high output reflects their strong influence and industry demand.

3. Top 10 Rated Movies

Highest-rated movies include:

- Titanic
- The Avengers
- Furious 7
- The Dark Knight
- Toy Story 3
- Harry Potter
- The Hobbit
- Star Trek Into Darkness

Insight:

These movies excel due to strong storytelling, emotional depth, advanced cinematography, and massive franchise followership. High ratings indicate widespread audience satisfaction.

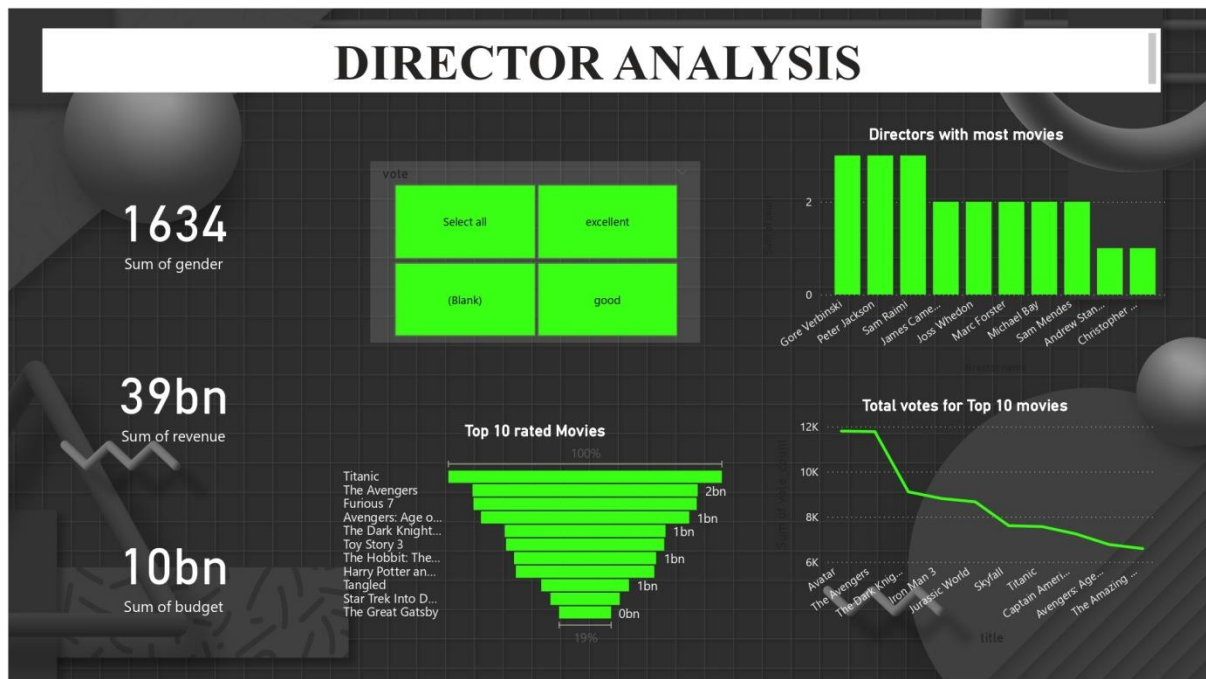
4. Total Votes for Top 10 Movies

Movies such as Avatar, The Dark Knight, and The Avengers have the highest vote counts.

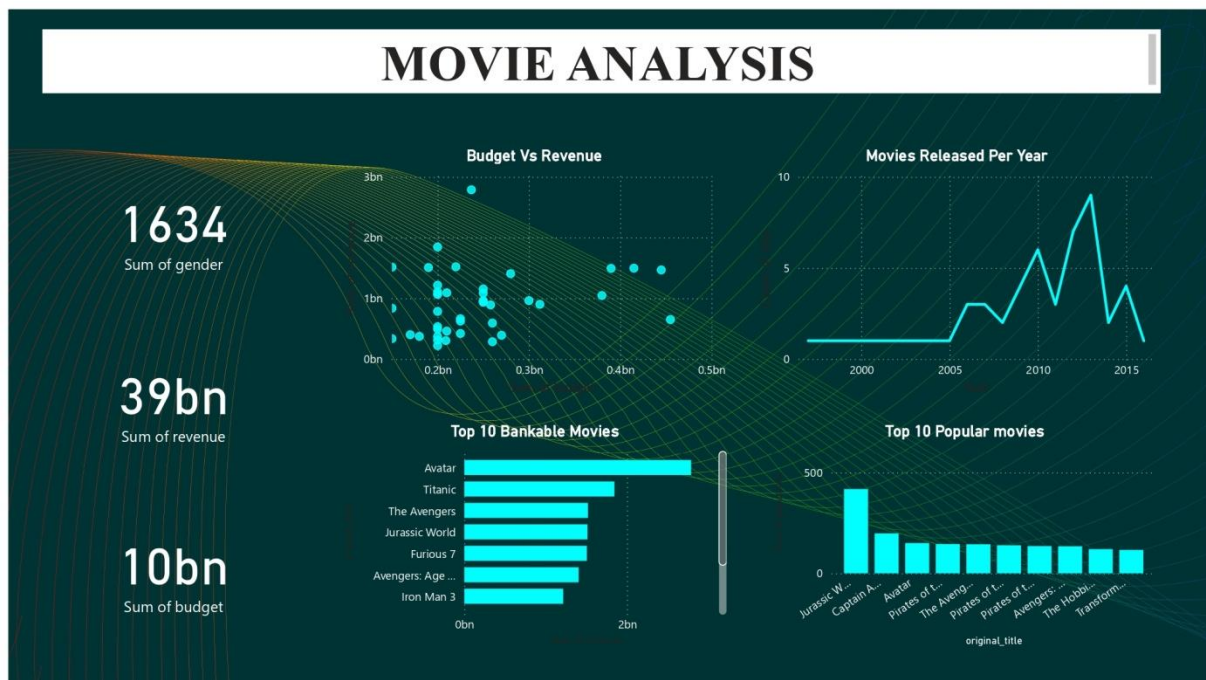
Insight:

Vote count represents audience engagement. Movies with high votes tend to be widely watched,

frequently reviewed, and globally recognized.



3.MOVIE ANALYSIS DASHBOARD



1. Budget vs Revenue

- Most movies earned significantly higher revenue than their production budget.
- Several movies earned over **2–3 billion dollars**.

Insight:

There is a positive budget-revenue relationship. High investment generally results in higher global returns.

2. Movies Released Per Year

A noticeable increase in releases occurred after 2010, with a peak around 2015.

Insight:

The industry expanded significantly during the 2010s due to growth in global cinema demand and streaming platforms.

3. Top 10 Bankable Movies

Including:

- Avatar
- Titanic
- The Avengers
- Jurassic World
- Furious 7
- Iron Man 3

Insight:

High-revenue movies share common traits:

- Franchise connection
- Global distribution
- Strong marketing
- High production value

4. Top 10 Popular Movies

- Jurassic World
- Captain America: Civil War
- Avatar
- Pirates of the Caribbean
- The Avengers

Insight:

Popularity is strongly dependent on franchise recognition and action-centric storytelling.

Key Findings & Insights**✓Strong Revenue Performance**

Movies generate almost **4 times their total budget**, reflecting high profitability.

✓Franchise Dominance

Top movies by revenue, popularity, votes, and ratings mostly belong to:

- Marvel Cinematic Universe
- Pirates of the Caribbean
- Jurassic World
- Avatar
- Fast & Furious series

✓Director Influence

A small group of directors contributes the most movies and generates the highest revenue.

✓Gender Disparity

Female directors make up only **5.6%**, showing a clear imbalance in opportunities.

✓Increasing Trend in Movie Releases

Film production has steadily increased, especially between 2010–2015.

Conclusion

The IMDB Movie Analysis project reveals that the movie industry is highly profitable and strongly influenced by a few top directors and global franchises. Movies with higher budgets tend to generate higher revenues, proving the advantage of investment in large-scale productions. However, the industry shows a significant gender gap among directors. The dashboards offer a comprehensive view of movie performance, director contributions, audience engagement, and global cinematic trends.