IMDB Movies Data Analysis using SQL

1. Project Overview

The IMDB Movies Data Analysis project focuses on analyzing movie and director information to uncover insights related to movie popularity, revenue, ratings, and director performance.

Using SQL queries on a relational database, the project answers key business questions related to movie success, director productivity, and profitability.

2. Dataset Summary

The dataset is stored in a remote MySQL database and consists of two tables: Movies and Directors.

Tables:

- Movies
- Directors

Key Attributes:

- Movie details: title, budget, revenue, popularity, ratings
- Director details: name, gender, department

The tables are connected using a primary–foreign key relationship between directors.id and movies.director_id.

3. Project Objectives

The project is designed to answer the following business questions:

1. Retrieve all movie data
2. Retrieve all director data
3. Count total movies in IMDB
4. Identify specific directors
5. Analyze directors by name patterns
6. Count female directors
7. Identify ranking among female directors
8. Find popular and bankable movies
9. Analyze movie ratings after 2000
10. Identify movies by a specific director
11. Determine the most productive director
12. Identify the most bankable director

4. Data Analysis using SQL

Below are the primary SQL queries used to answer the business questions.

1. Can you get all data about movies?

```
1 • USE project_movie_database;
2
3   #1-Can you get all data about movies?
4 • SELECT * FROM movies;
5
```

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

id	original_title	budget	popularity	release_date	revenue	title	vote_average	vote_count
43597	Avatar	237000000	150	2009-12-10	2787965087	Avatar	7.2	11800
43598	Pirates of the Caribbean: At World"s End	300000000	139	2007-05-19	961000000	Pirates of the Caribbean: At World"s End	6.9	4500
43599	Spectre	245000000	107	2015-10-26	880674609	Spectre	6.3	4466
43600	The Dark Knight Rises	250000000	112	2012-07-16	1084939099	The Dark Knight Rises	7.6	9106
43601	John Carter	260000000	43	2012-03-07	284139100	John Carter	6.1	2124
43602	Spider-Man 3	258000000	115	2007-05-01	890871626	Spider-Man 3	5.9	3576
43603	Tangled	260000000	48	2010-11-24	591794936	Tangled	7.4	3330
43604	Avengers: Age of Ultron	280000000	134	2015-04-22	1405403694	Avengers: Age of Ultron	7.3	6767
43605	Harry Potter and the Half-Blood Prince	250000000	98	2009-07-07	933959197	Harry Potter and the Half-Blood Prince	7.4	5293
43607	Superman Returns	270000000	57	2006-06-28	391081192	Superman Returns	5.4	1400
43608	Quantum of Solace	200000000	107	2008-10-30	586090727	Quantum of Solace	6.1	2965
43609	Pirates of the Caribbean: Dead Man"s C...	200000000	145	2006-06-20	1065659812	Pirates of the Caribbean: Dead Man"s C...	7	5246
43610	The Lone Ranger	255000000	49	2013-07-03	89289910	The Lone Ranger	5.9	2311
43611	Man of Steel	225000000	99	2013-06-12	662845518	Man of Steel	6.5	6359
43612	The Chronicles of Narnia: Prince Caspian	225000000	53	2008-05-15	419651413	The Chronicles of Narnia: Prince Caspian	6.3	1630
43613	The Avengers	220000000	144	2012-04-25	1519557910	The Avengers	7.4	11776
43614	Pirates of the Caribbean: On Stranger T...	380000000	135	2011-05-14	1045713802	Pirates of the Caribbean: On Stranger T...	6.4	4948
43615	Men in Black 3	225000000	52	2012-05-23	624026776	Men in Black 3	6.2	4160
43616	The Hobbit: The Battle of the Five Armies	250000000	120	2014-12-10	956019788	The Hobbit: The Battle of the Five Armies	7.1	4760
43617	The Amazing Spider-Man	215000000	89	2012-06-27	752215857	The Amazing Spider-Man	6.5	6586
43618	Robin Hood	200000000	37	2010-05-12	310669540	Robin Hood	6.2	1398
43619	The Hobbit: The Desolation of Smaug	250000000	94	2013-12-11	958400000	The Hobbit: The Desolation of Smaug	7.6	4524
43620	The Golden Compass	180000000	42	2007-12-04	372234864	The Golden Compass	5.8	1303
43621	King Kong	207000000	61	2005-12-14	550000000	King Kong	6.6	2337

movies 12 | Apply | Revert

2. How do you get all data about directors?

```
7
8   #2-How do you get all data about directors?
9 • SELECT * FROM directors;
10
11
```

Result Grid | Filter Rows: | Edit: | Export

name	id	gender	uid	department
James Cameron	4762	2	2710	Directing
Gore Verbinski	4763	2	1704	Directing
Sam Mendes	4764	2	39	Directing
Christopher Nolan	4765	2	525	Directing
Andrew Stanton	4766	2	7	Directing
Sam Raimi	4767	2	7623	Directing
Byron Howard	4768	2	76595	Directing
Joss Whedon	4769	2	12891	Directing
David Yates	4770	2	11343	Directing
Zack Snyder	4771	2	15217	Directing
Bryan Singer	4772	2	9032	Directing
Marc Forster	4773	2	12995	Directing
Andrew Adamson	4774	2	5524	Directing
Rob Marshall	4775	2	17633	Directing
Barry Sonnenfeld	4776	2	5174	Directing
Peter Jackson	4777	2	108	Directing
Marc Webb	4778	2	87742	Directing
Ridley Scott	4779	2	578	Directing
Chris Weitz	4780	0	3288	Directing
Anthony Russo	4781	2	19271	Directing
Peter Berg	4782	2	36602	Directing
Colin Trevorrow	4783	2	930707	Directing
Shane Black	4784	2	1108	Directing
Tim Burton	4785	2	510	Directing
Brett Ratner	4786	2	11091	Directing

directors 23 | Result 24 | directors 25 | directors 26 | Result 27

3. Check how many movies are present in IMDB.

```
12      #3-Check how many movies are present in IMDB.
13 •    SELECT COUNT(*) AS total_movies FROM movies;
14
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content

total_movies
47

4. Find these 3 directors: James Cameron; Luc Besson; John Woo

Limit to 1000 rows

```
16      #4-Find these 3 directors: James Cameron ; Luc Besson ; John Woo
17 •    SELECT *
18      FROM directors
19      WHERE name IN ('James Cameron', 'Luc Besson', 'John Woo');
20
```

Result Grid | Filter Rows: | Edit: | Export/Import:

name	id	gender	uid	department
James Cameron	4762	2	2710	Directing
John Woo	4893	2	11401	Directing
Luc Besson	4949	2	59	Directing
NULL	NULL	NULL	NULL	NULL

5. Find all directors with name starting with S.

```
22      #5-Find all directors with name starting with S.
23 •    SELECT *
24      FROM directors
25      WHERE name LIKE 'S%';
26
```

Result Grid | Filter Rows: | Edit: | Export/Impo

name	id	gender	uid	department
Sam Mendes	4764	2	39	Directing
Sam Raimi	4767	2	7623	Directing
Shane Black	4784	2	1108	Directing
Steven Spielberg	4799	2	488	Directing
Stephen Sommers	4815	2	7775	Directing
Shawn Levy	4842	2	17825	Directing
Steve Hickner	4852	2	44113	Directing
Simon Wells	4855	2	21879	Directing
Steven Soderbergh	4909	2	1884	Directing
Simon West	4930	2	12786	Directing
Stefen Fangmeier	4931	0	25453	Directing
Spike Jonze	4932	2	5953	Directing
Steve Martino	4943	2	71729	Directing
Sergei Bodrov	4952	0	130938	Directing
Sydney Pollack	4965	2	2226	Directing
Sylvester Stallone	4992	2	16483	Directing
Seth Gordon	4997	2	71600	Directing
Scott Derrickson	5004	2	55499	Directing
Stephen Hopkins	5008	2	2042	Directing
Steven Brill	5013	2	32593	Directing
Stephen Norrington	5028	2	10808	Directing
Steve Carr	5048	2	52112	Directing
Seth MacFarlane	5075	2	52139	Directing
Scott Waugh	5081	2	293911	Directing
Stanley Kubrick	5089	2	240	Directing

directors 23 | Result 24 | directors 25 | directors 26 | Result 27 | director

6. Count female directors.

```
28      #6-Count female directors.
29 •    SELECT COUNT(*) AS female_directors
30      FROM directors
31      WHERE gender = 1;
32
```

Result Grid | Filter Rows: | Export: | Wrap Cell Co

female_directors
150

7. Find the name of the 10th first women directors?

```
34      #7-Find the name of the 10th first women directors?
35 •    SELECT name
36      FROM directors
37      WHERE gender = 1
38      ORDER BY id
39      LIMIT 1 OFFSET 9;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

name
Angelina Jolie

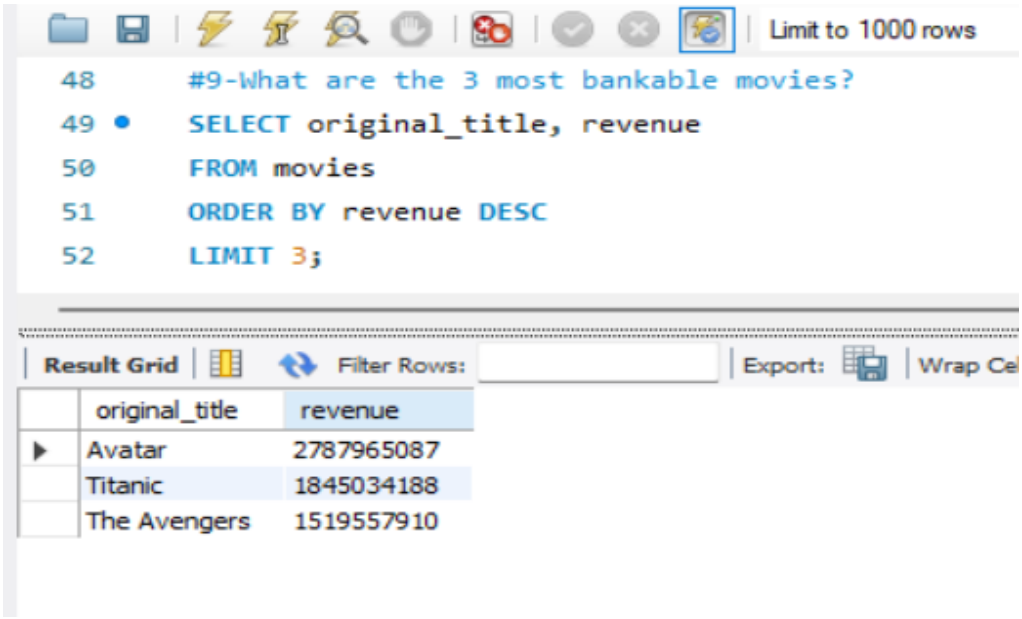
8. What are the 3 most popular movies?

```
41      #8-What are the 3 most popular movies?
42 •    SELECT original_title, popularity
43      FROM movies
44      ORDER BY popularity DESC
45      LIMIT 3;
46
```

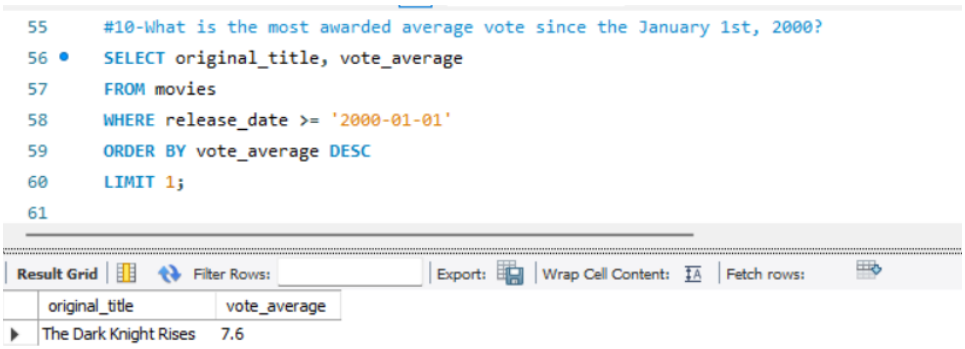
Result Grid | Filter Rows: | Export: | Wrap

original_title	popularity
Jurassic World	418
Captain America: Civil War	198
Avatar	150

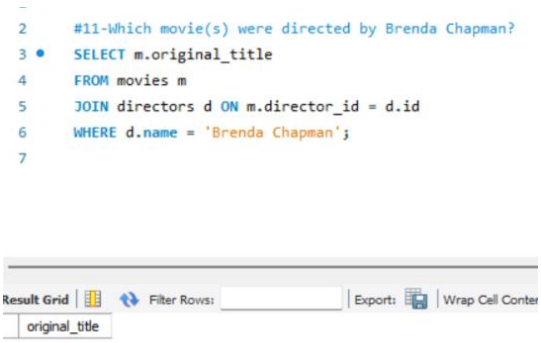
9. What are the 3 most bankable movies?



10. What is the most awarded average vote since the January 1st, 2000?



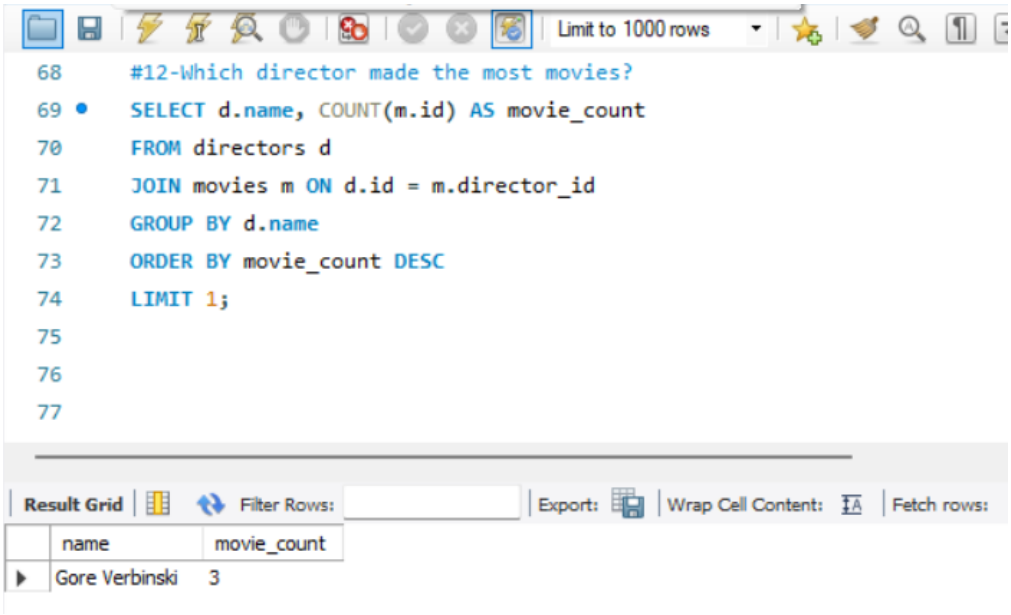
11. Which movie(s) were directed by Brenda Chapman?



Result:

No movies were mapped to this director in the dataset.

12. Which director made the most movies?



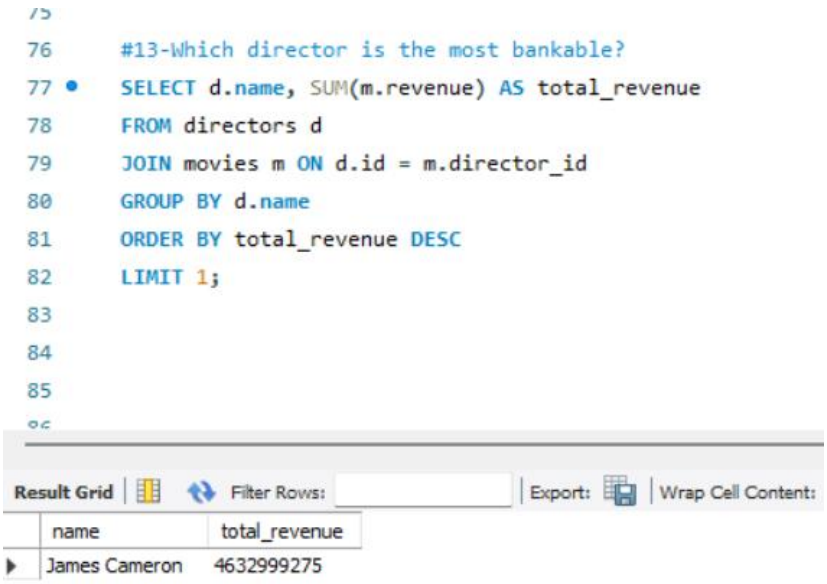
```
#12-Which director made the most movies?
SELECT d.name, COUNT(m.id) AS movie_count
FROM directors d
JOIN movies m ON d.id = m.director_id
GROUP BY d.name
ORDER BY movie_count DESC
LIMIT 1;
```

Limit to 1000 rows

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

name	movie_count
Gore Verbinski	3

13. Which director is the most bankable?



```
#13-Which director is the most bankable?
SELECT d.name, SUM(m.revenue) AS total_revenue
FROM directors d
JOIN movies m ON d.id = m.director_id
GROUP BY d.name
ORDER BY total_revenue DESC
LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

name	total_revenue
James Cameron	4632999275

5. Business Summary

1. A small number of directors account for a large share of total movie output.
2. Revenue distribution shows that only a few directors are highly bankable.
3. Popularity and revenue do not always correlate with higher ratings.
4. Data inconsistencies highlight the importance of accurate relational mapping.

7. Conclusion

This project demonstrates strong SQL skills including joins, aggregations, filtering, sorting, and real-world data analysis. The analysis provides actionable insights into movie performance and director success.