DATAMITES-INTERNSHIP
MOVIE DATABASE – SQL PROJECT

This professionally enhanced report now contains detailed explanations below every SQL query screenshot for better understanding and presentation quality.

Project ID : **PRSQ-02-IMDB Movies**

Project Team ID : **PTID-CDA-DEC-25-1033**

Database Management System(DBMS) **: SQL SERVER**

Language **: SQL**

Project By **: PRIYANKA C METI**

TABLE OF CONTENTS

CONTENTS

INTRODUCTION:

This project focuses on analyzing the **Movie Database** using SQL queries to extract meaningful insights from movie and director data. The database contains important information such as movie titles, popularity, revenue, ratings, release dates, and director details. By performing different SQL operations like filtering, sorting, grouping, and joining tables, this project helps in understanding movie performance, director contribution, and overall industry trends in a simple and effective way.

PROJECT OVERVIEW:

• In this project with all the data present in the **movies** and **director's** tables, various SQL queries were performed to interpret the data and find useful insights related to movie popularity, revenue generation, ratings, and director performance.

• The movies table consists of data like:

👉 id
👉 original title
👉 title
👉 budget
👉 revenue
👉 popularity
👉 release date
👉 vote average
👉 vote count
👉 overview
👉 tagline
👉 uid
👉 director_id

• The directors table consists of data like:

👉 id
👉 name
👉 gender (0/2 = Male, 1 = Female)
👉 department
👉 uid

QUERIES NEED TO BE PERFORMED:

a) Can you get all data about movies?

b) How do you get all data about directors?

c) Check how many movies are present in IMDB.

d) Find these 3 directors: James Cameron ; Luc Besson ; John Woo

e) Find all directors with name starting with S.

f) Count female directors.

g) Find the name of the 10th first women directors?

h) What are the 3 most popular movies?

i) What are the 3 most bankable movies?

j) What is the most awarded average vote since the January 1st, 2000?

k) Which movie(s) were directed by Brenda Chapman?

l) Which director made the most movies?

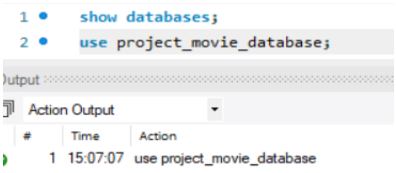
m) Which director is the most bankable?

EXPLORING THE DATABASE

First, we have to load the database to clearly understand the data present in it. We will explore the database step-by-step using SQL queries.

To load and access the database, we first connect to the **SQL Server** using the provided login credentials. After successfully connecting, we access our database project_movie_database and start exploring the data stored in it.

• SHOW DATABASES;



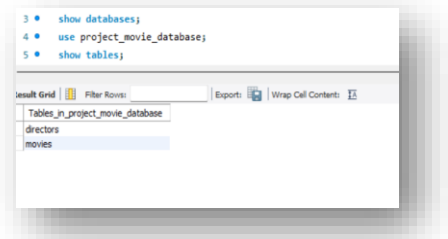
This query is used to display all the databases available in the SQL Server.
From this, we can identify our working database named:

👉 project_movie_database

• USE project_movie_database;

This query is used to select the specific database project_movie_database to work with in the Database Management System (DBMS).
It allows us to perform all further SQL operations inside this database.

• SHOW TABLES;



This query displays all the tables present in the selected database.
The tables available in this database are:

👉 movies
👉 directors

QUERIES AND THIRE OUTPUTS

Query A - Get All Movies

SELECT * FROM movies

```
1 • show databases;
2 • use project_movie_database;
3 • show tables;
4   -- a) Get all movies
5 • select * from movies;
```

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

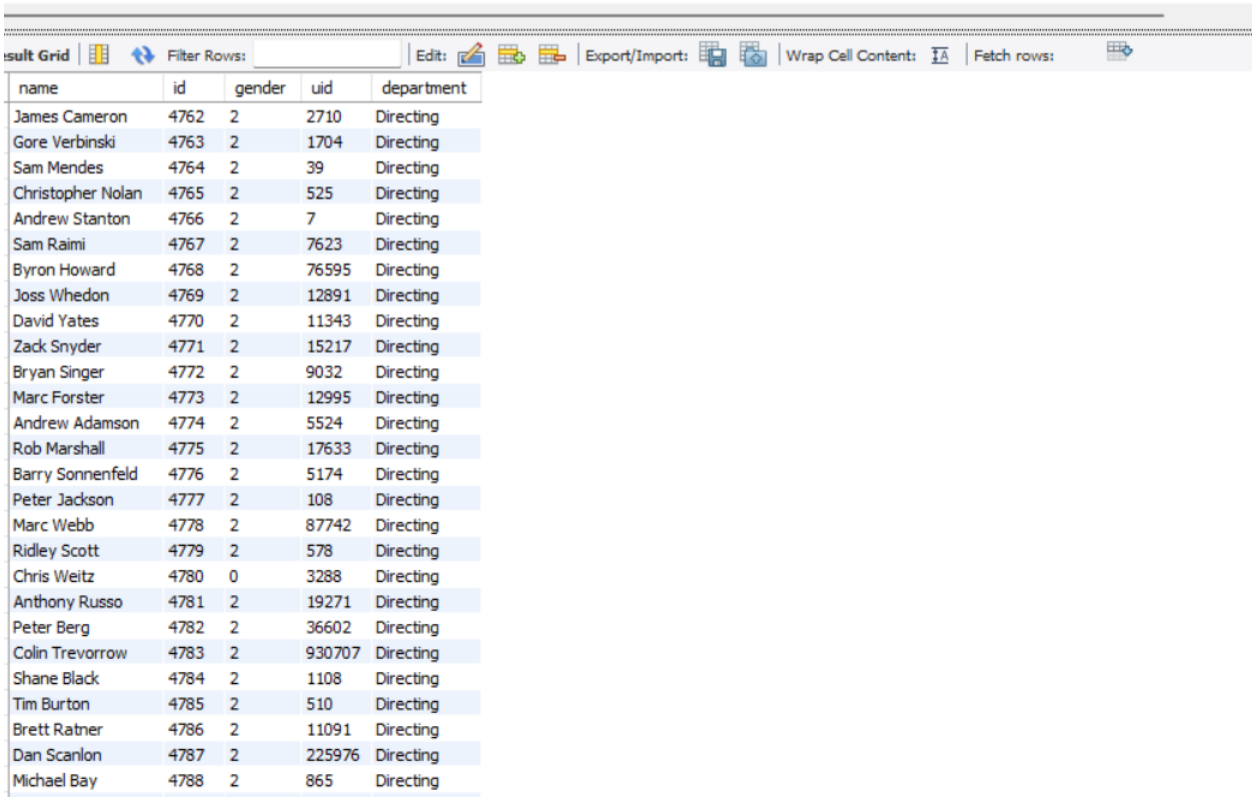
id	original_title	budget	popularity	release_date	revenue	title	vote_average	vote_count	overview	tagline
43620	The Golden Compass	180000000	42	2007-12-04	372234864	The Golden Compass	5.8	1303	After overhearing a shocking secret, precocious...	There are worlds be
43621	King Kong	207000000	61	2005-12-14	550000000	King Kong	6.6	2337	In 1933 New York, an overly ambitious movie pr...	The eighth wonder
43622	Titanic	200000000	100	1997-11-18	1845034188	Titanic	7.5	7562	84 years later, a 101-year-old woman named R...	Nothing on Earth co
43623	Captain America: Civil War	250000000	198	2016-04-27	1153304495	Captain America: Civil War	7.1	7241	Following the events of Age of Ultron, the colle...	Divided We Fall
43624	Battleship	209000000	64	2012-04-11	303025485	Battleship	5.5	2114	When mankind beams a radio signal into space, ...	The Battle for Earth
43625	Jurassic World	150000000	418	2015-06-09	1513528810	Jurassic World	6.5	8662	Twenty-two years after the events of Jurassic ...	The park is open.
43626	Skyfall	200000000	93	2012-10-25	1108561013	Skyfall	6.9	7604	When Bond"s latest assignment goes gravely w...	Think on your sins.
43627	Spider-Man 2	200000000	35	2004-06-25	783766341	Spider-Man 2	6.7	4321	Peter Parker is going through a major identity cr...	There"s a hero in all
43628	Iron Man 3	200000000	77	2013-04-18	1215439994	Iron Man 3	6.8	8806	When Tony Stark"s world is torn apart by a for...	Unleash the power
43629	Alice in Wonderland	200000000	78	2010-03-03	1025491110	Alice in Wonderland	6.4	4645	Alice, an unpretentious and individual 19-year-o...	You're invited to a
43630	X-Men: The Last Stand	210000000	3	2006-05-24	459359555	X-Men: The Last Stand	6.3	3525	When a cure is found to treat mutations, lines a...	Take a Stand
43631	Monsters University	200000000	89	2013-06-20	743559607	Monsters University	7	3528	A look at the relationship between Mike and Sull...	School never looked
43632	Transformers: Revenge of the Fallen	150000000	21	2009-06-19	836297228	Transformers: Revenge of the Fallen	6	3138	Sam Witwicky leaves the Autobots behind for a ...	Revenge is coming.
43633	Transformers: Age of Extinction	210000000	116	2014-06-25	1091405097	Transformers: Age of Extinction	5.8	3095	As humanity picks up the pieces, following the c...	This is not war. It"s
43634	Oz: The Great and Powerful	200000000	46	2013-03-07	491868548	Oz: The Great and Powerful	5.7	3530	Oscar Diggs, a small-time circus illusionist and co...	In Oz, nothing is wh
43636	TRON: Legacy	170000000	73	2010-12-10	400062763	TRON: Legacy	6.3	2841	Sam Flynn, the tech-savvy and daring son of K...	The Game Has Char
43637	Cars 2	200000000	49	2011-06-11	559852396	Cars 2	5.8	2033	Star race car Lightning McQueen and his pal Ma...	Ka-ciao!
43638	Green Lantern	200000000	51	2011-06-16	219851172	Green Lantern	5.1	2487	For centuries, a small but powerful force of war...	In our darkest hour
43639	Toy Story 3	200000000	59	2010-06-16	1066969703	Toy Story 3	7.6	4597	Woody, Buzz, and the rest of Andy"s toys have...	No toy gets left beh
43640	Terminator Salvation	200000000	71	2009-05-20	371353001	Terminator Salvation	5.9	2463	All grown up in post-apocalyptic 2018, John Co...	The End Begins.
43641	Furious 7	190000000	102	2015-04-01	1506249360	Furious 7	7.3	4176	Deckard Shaw seeks revenge against Dominic T...	Vengeance Hits Hon
43642	World War Z	200000000	81	2013-06-20	531865000	World War Z	6.7	5560	Life for former United Nations investigator Gerr...	Remember Philly!
43644	Star Trek Into Darkness	190000000	78	2013-05-05	467365246	Star Trek Into Darkness	7.4	4418	When the crew of the Enterprise is called back ...	Earth Will Fall
43646	The Great Gatsby	105000000	61	2013-05-10	351040419	The Great Gatsby	7.3	3769	An adaptation of F. Scott Fitzgerald"s Long Isla...	Reserving judgment
43647	Prince of Persia: The Sands of Time	150000000	62	2010-05-19	335154643	Prince of Persia: The Sands of Time	6.2	2317	A rogue prince reluctantly joins forces with a m...	Defy the Future
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Description: This query displays complete information of all movies stored in the database using SELECT * FROM movies. It retrieves columns such as movie ID, title, budget, revenue, popularity, release date, vote average and vote count. This helps to understand the full dataset before performing analysis.

Query B - Get All Directors

SELECT * FROM directors;

```
7     -- b) Get all directors
8 •   select * from directors;
```

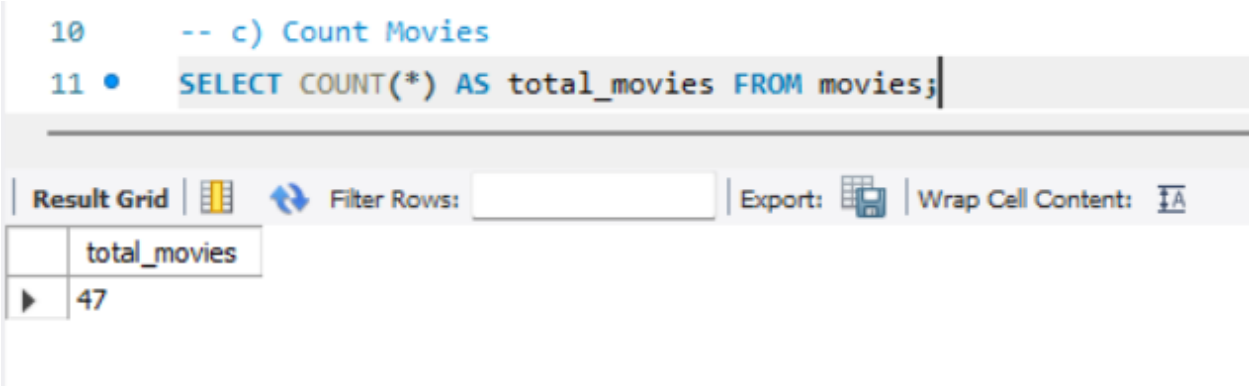


name	id	gender	uid	department
James Cameron	4762	2	2710	Directing
Gore Verbinski	4763	2	1704	Directing
Sam Mendes	4764	2	39	Directing
Christopher Nolan	4765	2	525	Directing
Andrew Stanton	4766	2	7	Directing
Sam Raimi	4767	2	7623	Directing
Byron Howard	4768	2	76595	Directing
Joss Whedon	4769	2	12891	Directing
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
Dan Scanlon	4787	2	225976	Directing
Michael Bay	4788	2	865	Directing

Description: This query retrieves all records from the directors table. It lists director names, gender, department and their unique identifiers. This helps to understand the number of directors and their attributes in the dataset.

Query C - Count Movies

SELECT COUNT(*) AS total_movies FROM movies;



Description: This query counts how many movies are stored in the database using COUNT(*) function. It returns a single numeric value representing the total number of movies available for analysis.

Query D - Find Specific Directors

SELECT * FROM directors

WHERE name IN ('James Cameron', 'Luc Besson', 'John Woo');

```
13      -- d) Find James Cameron, Luc Besson, John Woo
14 •    SELECT * FROM directors
15      WHERE name IN ('James Cameron','Luc Besson','John Woo');
```

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

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

Description: This query filters and retrieves information about three famous directors: James Cameron, Luc Besson and John Woo. The WHERE clause with IN condition is used to match multiple names in a single query.

Query E - Directors Starting With S

SELECT * FROM directors

WHERE name LIKE 'S%';

```
17      -- e) Directors starting with S
18 •    SELECT * FROM directors
19      WHERE name LIKE 'S%';
```

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

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
Stuart Beattie	5095	0	1707	Directing
Steven Zaillian	5117	2	2260	Directing

Description: This query displays all directors whose names start with the letter 'S' using LIKE 'S%'. It is useful for pattern matching and filtering based on text values.

Query F - Count Female Directors

SELECT COUNT(*) AS female_directors

FROM directors

WHERE gender = 1;

```
21      -- f) Count Female Directors
22 •    SELECT COUNT(*) AS female_directors
23      FROM directors
24      WHERE gender = 1;
```

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

female_directors
150

Description: This query counts total female directors in the database by checking gender = 1. It helps in gender-based analysis and representation study.

Query G - 10th Alphabetical Female Director

SELECT name

FROM directors

WHERE gender = 1

ORDER BY name

LIMIT 1 OFFSET 9;

```
26      -- g) 10th Female Director Alphabetically
27 •    SELECT name
28      FROM directors
29      WHERE gender = 1
30      ORDER BY name
31      LIMIT 1 OFFSET 9;
```

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

name
Amy Holden Jones

Description: This query sorts all female directors alphabetically and retrieves the 10th director using LIMIT and OFFSET. It demonstrates ordering and positional record extraction techniques.

Query H - Top 3 Most Popular Movies

SELECT original_title, popularity

FROM movies

ORDER BY popularity DESC

LIMIT 3;

```
33      -- h) 3 Most Popular Movies
34 •    SELECT original_title, popularity
35      FROM movies
36      ORDER BY popularity DESC
37      LIMIT 3;
```

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

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

Description: This query retrieves the three most popular movies by sorting popularity in descending order. It helps identify movies with highest audience interest levels.

Query I - Top 3 Most Bankable Movies

SELECT original_title, revenue

FROM movies

ORDER BY revenue DESC

LIMIT 3;

```
38
39      -- i) 3 Most Bankable Movies (Revenue)
40 •    SELECT original_title, revenue
41      FROM movies
42      ORDER BY revenue DESC
43      LIMIT 3;
```

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

original_title	revenue
Avatar	2787965087
Titanic	1845034188
The Avengers	1519557910

Description: This query lists the top three highest revenue generating movies. It is useful for financial performance analysis and profitability evaluation.

Query J - Highest Rated Movie Since 2000

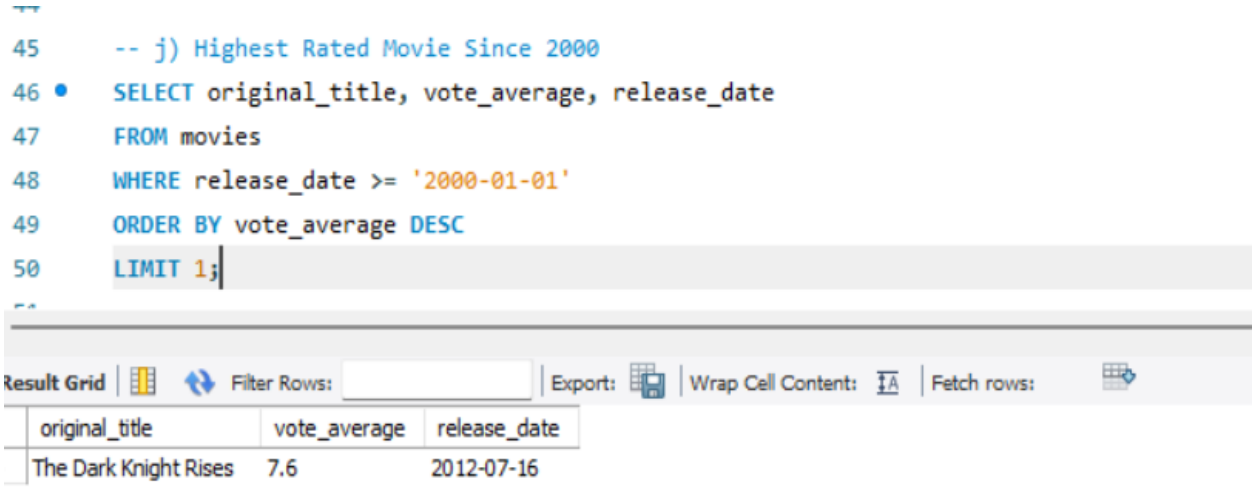
```
SELECT original_title, vote_average, release_date

FROM movies

WHERE release_date >= '2000-01-01'

ORDER BY vote_average DESC

LIMIT 1;
```



Description: This query finds the highest voted/rated movie released after January 1st 2000. It filters movies by date, sorts by ratings and picks the best performing one.

Query K - Movies Directed by Brenda Chapman

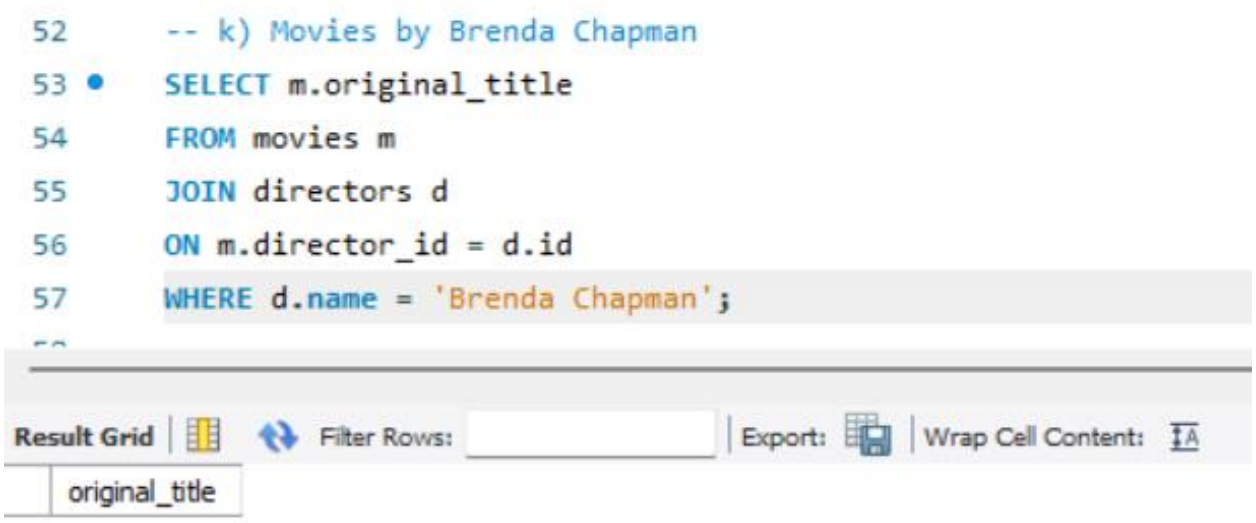
```
SELECT m.original_title

FROM movies m

JOIN directors d

ON m.director_id = d.id

WHERE d.name = 'Brenda Chapman';
```



Description: This query identifies movies directed by Brenda Chapman using JOIN between movies and directors tables. It demonstrates relational database linking and lookup.

Query L - Director With Most Movies

SELECT d.name, COUNT(m.id) AS movie_count

FROM directors d

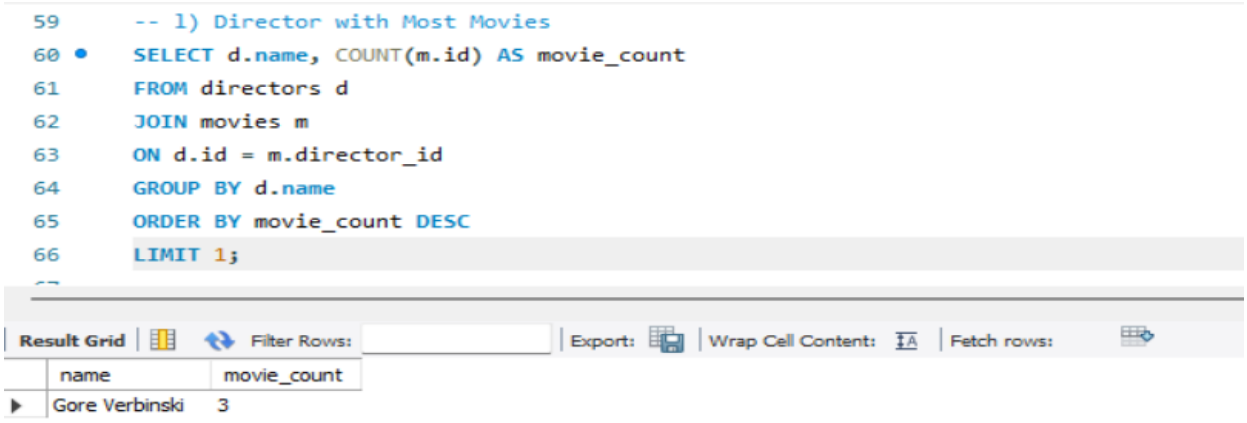
JOIN movies m

ON d.id = m.director_id

GROUP BY d.name

ORDER BY movie_count DESC

LIMIT 1;



Description: This query calculates which director has directed the highest number of movies using GROUP BY and COUNT. It helps determine the most productive director.

Query M - Most Bankable Director

SELECT d.name, SUM(m.revenue) AS total_revenue

FROM directors d

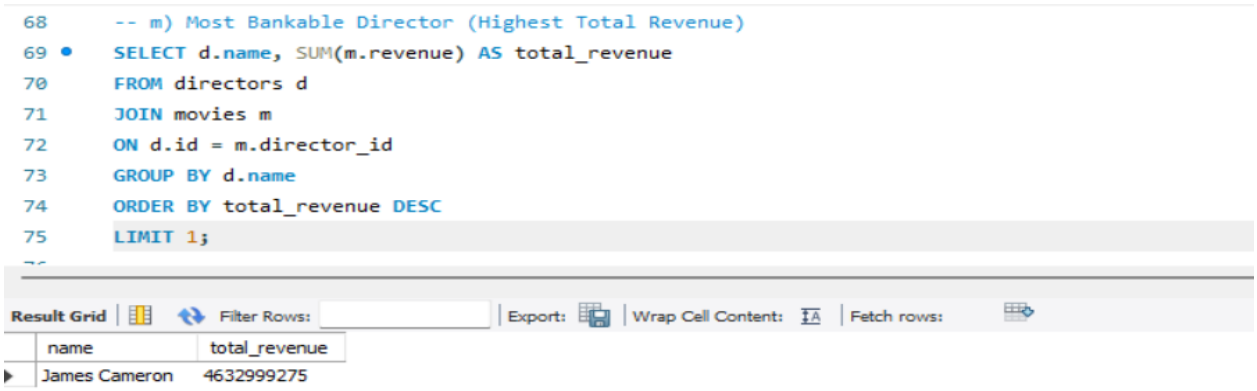
JOIN movies m

ON d.id = m.director_id

GROUP BY d.name

ORDER BY total_revenue DESC

LIMIT 1;



Description: This query identifies the most bankable director based on highest total movie revenue generated. SUM() aggregation with GROUP BY is used to compute revenue contribution of each director.

CONCLUSION:

In this project, an in-depth SQL analysis was performed on the Movie Database to extract meaningful insights from structured data. Through a series of systematically designed SQL queries, we explored essential analytical dimensions such as movie popularity, financial performance, director productivity, gender representation, and revenue contribution.

Business Understanding & Analytical Impact:
The executed queries enabled identification of:
• Total availability of movies and directors in the dataset
• Most popular and highest grossing movies
• Directors with the greatest industry impact
• Female representation within the directing category
• Highest rated movies after the year 2000

Technical Expertise Demonstrated:
Throughout this project, strong database analytical capabilities were demonstrated through:
• Use of filtering (WHERE), sorting (ORDER BY), limiting (LIMIT)
• Aggregation and summarization using COUNT() and SUM()
• Pattern matching using LIKE
• Relational understanding using JOIN operations
• Group-based analytics using GROUP BY and HAVING

Outcome & Value:
This project successfully highlights how SQL can transform raw data into insightful decision-support information. It not only enhances technical understanding but also strengthens analytical thinking and database interpretation skills. The results obtained can support data-driven decision making in movie analytics, revenue planning, business evaluation, and market understanding.

Overall, this project provided a complete learning experience covering database understanding, query execution, professional documentation, and interpretation of analytical outcomes, demonstrating strong readiness for Data Analytics and SQL-based roles.