DATAMITES - INTERNSHIP

PRSQ-02-IMDB MOVIES

Project ID: **PRSQL-01 IMDB MOVIES**

Project Team ID: **PTID-CDA-SEP-25-1036**

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

Language**: SQL**

Project By**: RAKSHITA SHANKRAPPA APPAJI**

Table of Content

Si. no	Content
1	Introduction
2	Project Overview
3	Quires need to be performed
4	Exploring the database
5	Quires and their Output
6	Conclusion

Introduction

- This report presents an overview and analysis of the IMDB Movies database, which contains structured information about movies and their directors. The objective of this analysis is to use SQL queries to explore the dataset and extract meaningful insights related to movie performance, director activity, popularity, revenue, and audience ratings.

Project Overview

- This report presents an overview and analysis of the IMDB Movies database, which contains structured information about movies and their directors. The objective of this analysis is to use SQL queries to explore the dataset and extract meaningful insights related to movie performance, director activity, popularity, revenue, and audience ratings.

The database consists of two main tables:

- **Directors Table**: Contains director details such as name, gender, department, and a unique director ID.
- **Movies Table**: Contains movie-related information including title, budget, revenue, popularity, release date, ratings, and a director reference ID.

Quires 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?

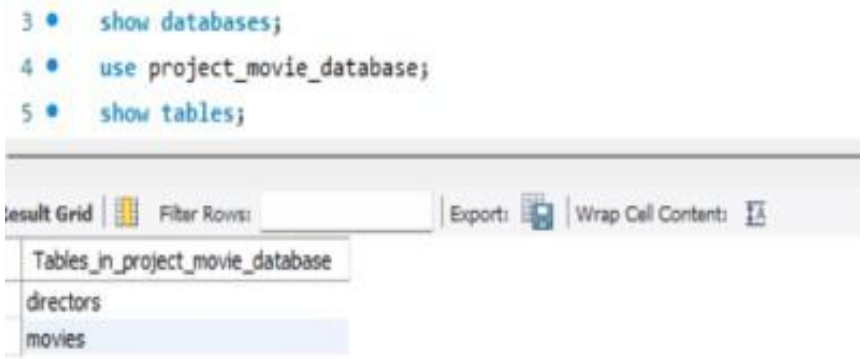
Exploring the database

- Exploring the database means first understanding **what data is available and how it is structured**. In this project, the database contains two main tables: **Movies** and **Directors**. The Movies table stores details about films such as movie names, release dates, popularity, budget, revenue, and audience ratings. The Directors table stores information about the directors, including their names, gender, and department.
- During this step, we view all the tables in the database, check the number of records present, and understand how the Movies and Directors tables are connected using the **director ID**. This helps in identifying the columns, data types, and relationships between tables

show database;

Use project_movie_database;

Show table;



a. Can you get all data about movies?

```
Select
    *
  from
      movies;
```

```
8  SELECT
9      *
10 FROM
11     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
43597	Avatar	237000000	150	2009-12-10	2787965087	Avatar	7.2	11800	In the 22nd century, a paraplegic Marine is disp...	Enter th
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	Captain Barbossa, long believed to be dead, ha...	At the e
43599	Spectre	245000000	107	2015-10-26	880674609	Spectre	6.3	4466	A cryptic message from Bonds past sends him o...	A Plan N
43600	The Dark Knight Rises	250000000	112	2012-07-16	1084939099	The Dark Knight Rises	7.6	9106	Following the death of District Attorney Harvey ...	The Lega
43601	John Carter	260000000	43	2012-03-07	284139100	John Carter	6.1	2124	John Carter is a war-weary, former military cap...	Lost in o
43602	Spider-Man 3	258000000	115	2007-05-01	890871626	Spider-Man 3	5.9	3576	The seemingly invincible Spider-Man goes up ag...	The batt
43603	Tangled	260000000	48	2010-11-24	591794936	Tangled	7.4	3330	When the kingdom's most wanted-and most ch...	They're
43604	Avengers: Age of Ultron	280000000	134	2015-04-22	1405403694	Avengers: Age of Ultron	7.3	6767	When Tony Stark tries to jumpstart a dormant p...	A New A
43605	Harry Potter and the Half-Blood Prince	250000000	98	2009-07-07	933959197	Harry Potter and the Half-Blood Prince	7.4	5293	As Harry begins his sixth year at Hogwarts, he ...	Dark Sec
43607	Superman Returns	270000000	57	2006-06-28	391081192	Superman Returns	5.4	1400	Superman returns to discover his 5-year absenc	NULL

b. How do you get all data about directors?

SELECT * FROM directors;

```
14 •  SELECT
15        *
16    FROM
17        directors;
```

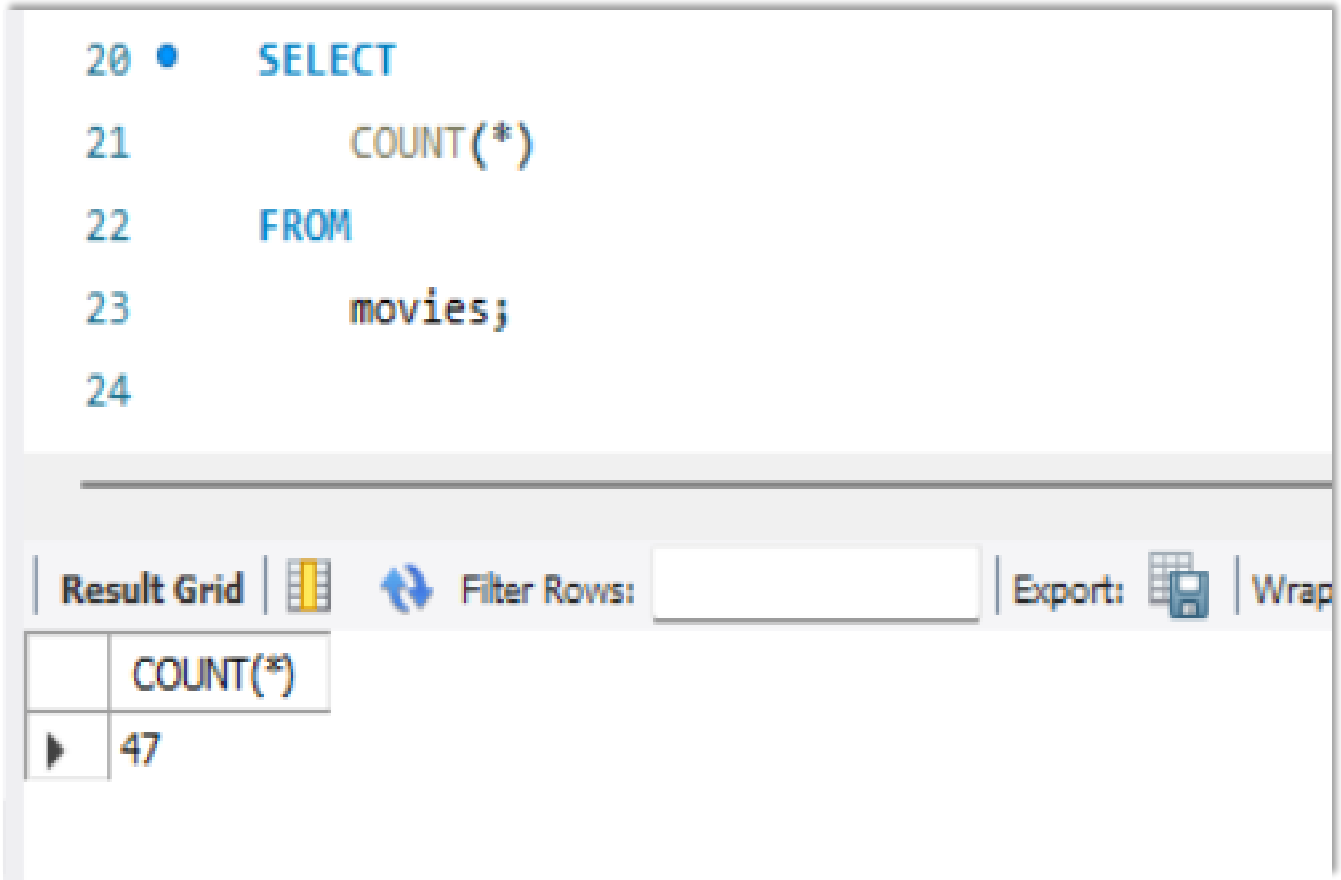
Result Grid | Filter Rows: | Edit:

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

directors 4

c. Check how many movies are present in IMDB.

SELECT COUNT (*) AS total_movies FROM movies;



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

```
SELECT * FROM
directors
    WHERE name IN ('James
Cameron','Luc Besson','John
Woo');
```

```
26 •  SELECT
27        *
28    FROM
29        directors
30    WHERE
31        name IN ('James Cameron' , 'Luc Besson', 'John Woo');
32
```

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

e. Find all directors with name starting with S.

SELECT * FROM directors
WHERE name LIKE 'S%';

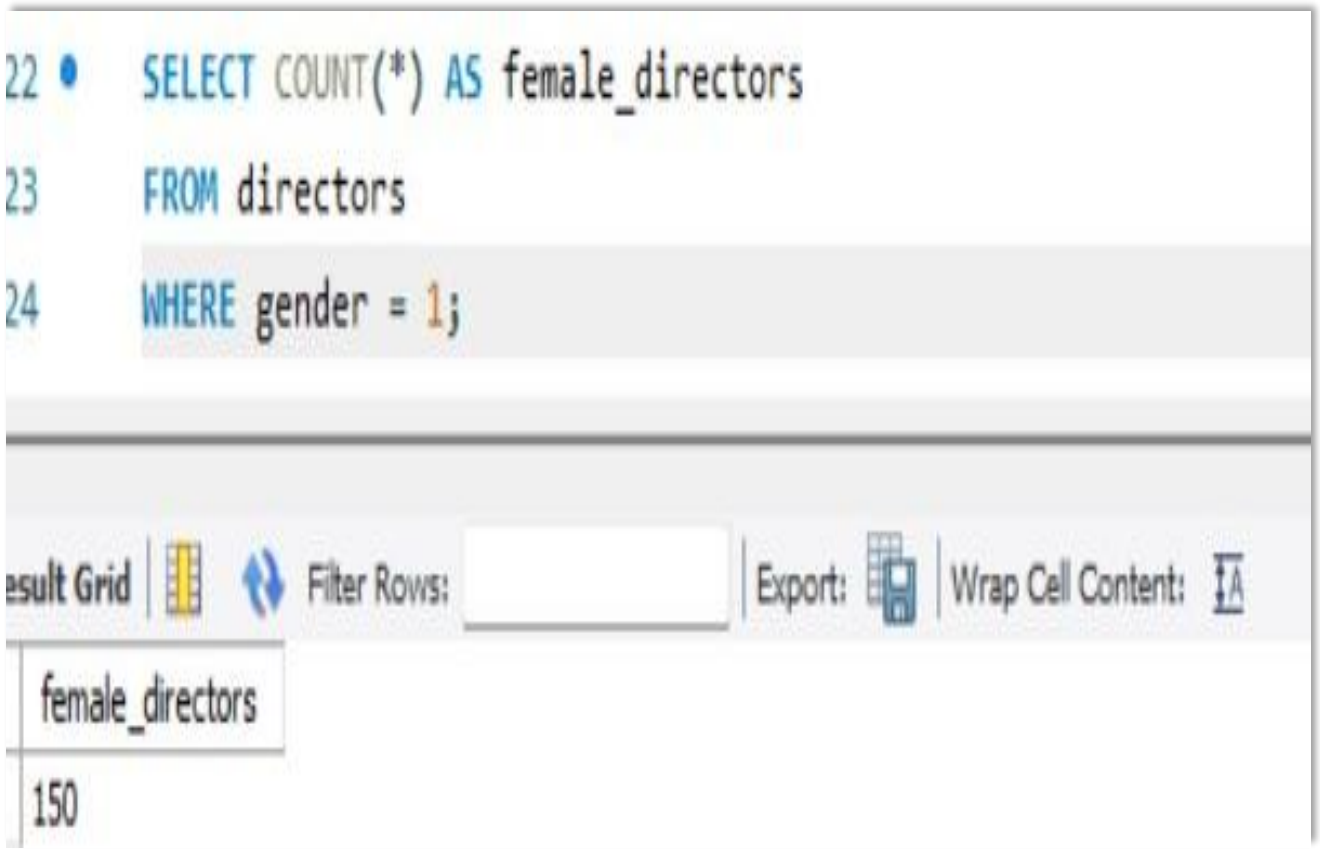
```
34 •  SELECT
35        *
36    FROM
37        directors
38    WHERE
39        name LIKE 'a%';
```

Result Grid | Filter Rows: | Edit:

name	id	gender	uid	department
Andrew Stanton	4766	2	7	Directing
Andrew Adamson	4774	2	5524	Directing
Anthony Russo	4781	2	19271	Directing
Alan Taylor	4837	2	47005	Directing
Alex Proyas	4865	2	21085	Directing
Ang Lee	4867	2	1614	Directing
Alejandro González Iñárritu	4874	2	223	Directing
Alfonso Cuarón	4882	2	11218	Directing
Adam McKay	4925	2	55710	Directing
Ash Brannon	4946	2	12905	Directing
Alan J. Pakula	4969	2	6349	Directing

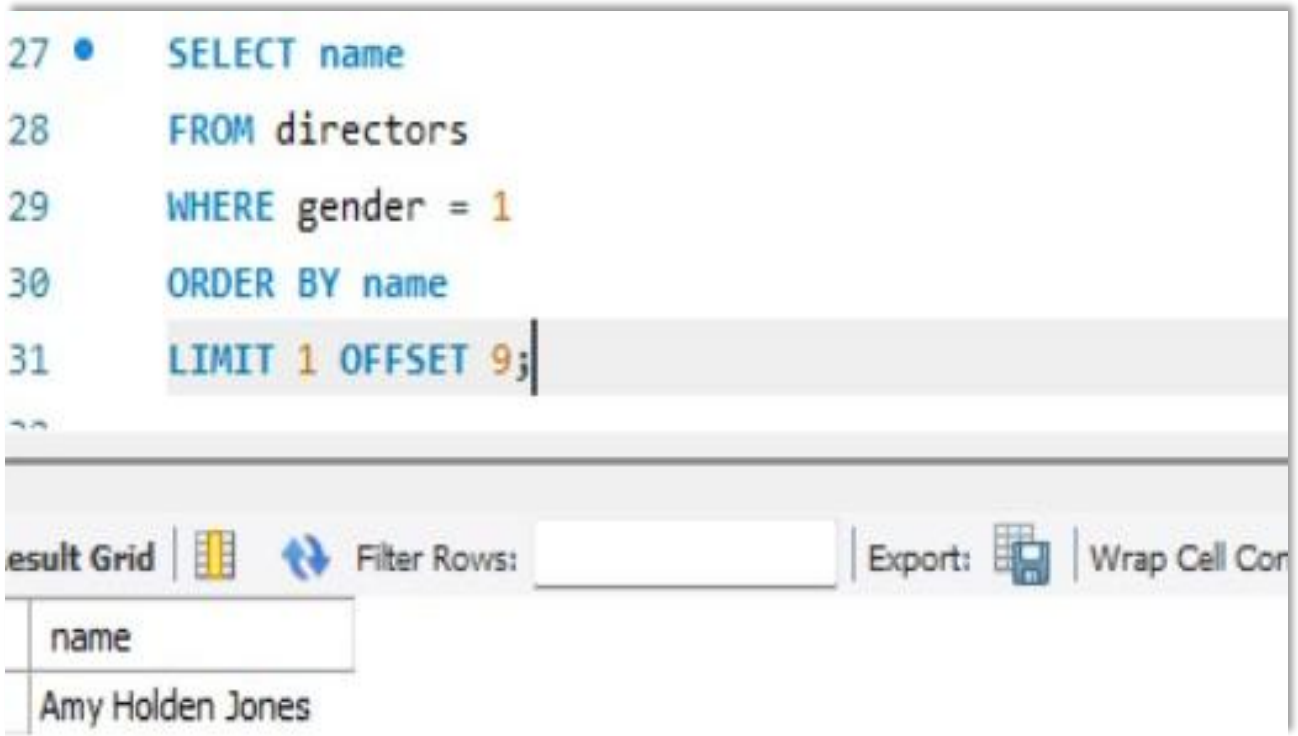
f. Count female directors.

SELECT COUNT (*) AS female_directors FROM director WHERE gender = 1;



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

SELECT name FROM directors WHERE gender = 1 ORDER BY name LIMIT 1 OFFSET 9;



h. What are the 3 most popular movies?

SELECT original_title, popularity FROM movies ORDER BY popularity DESC LIMIT 3;

```
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

i. What are the 3 most bankable movies?

SELECT original_title, revenue
FROM movies
ORDER BY revenue DESC
LIMIT 3;

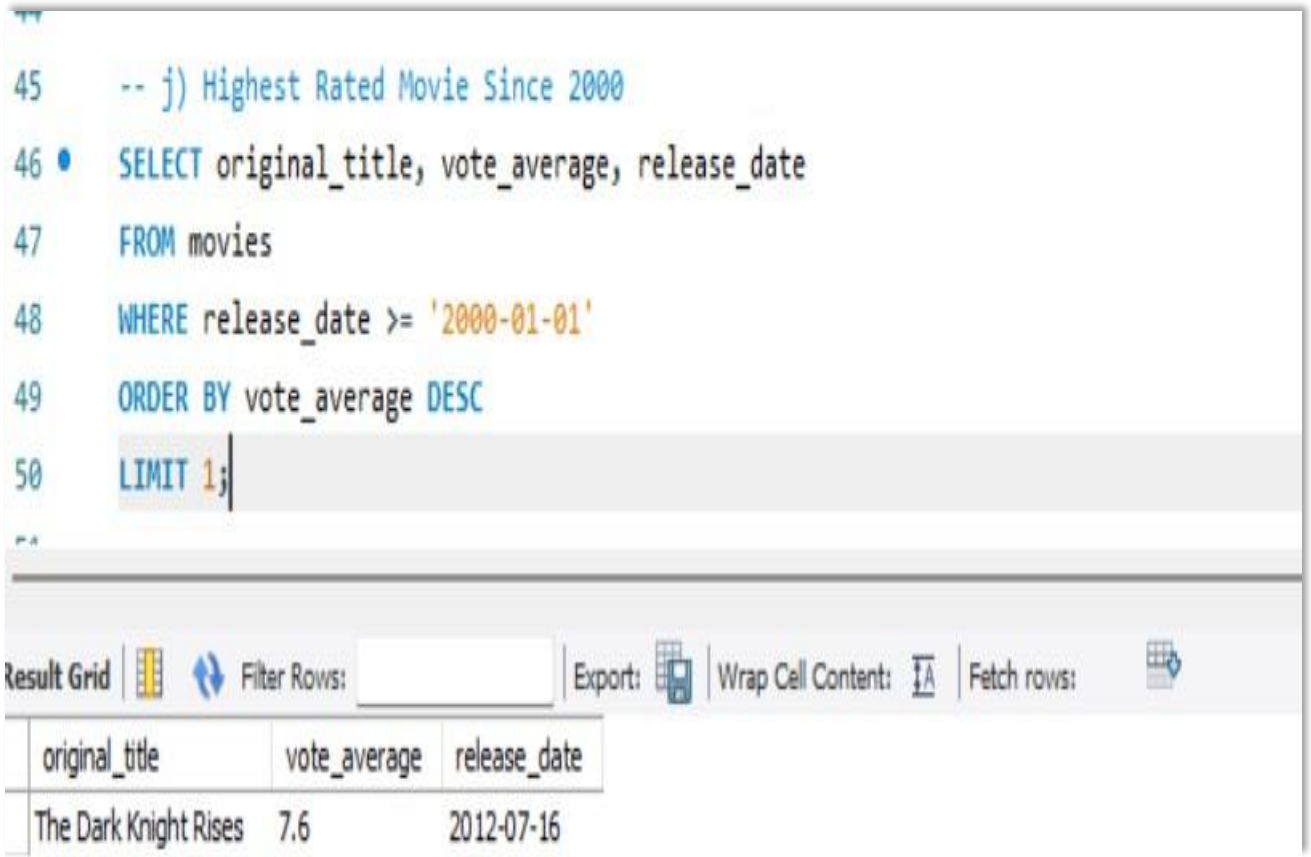
```
40 •  SELECT original_title, revenue
41    FROM movies
42    ORDER BY revenue DESC
43    LIMIT 3;
```

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

original_title	revenue
Avatar	2787965087
Titanic	1845034188
The Avengers	1519557910

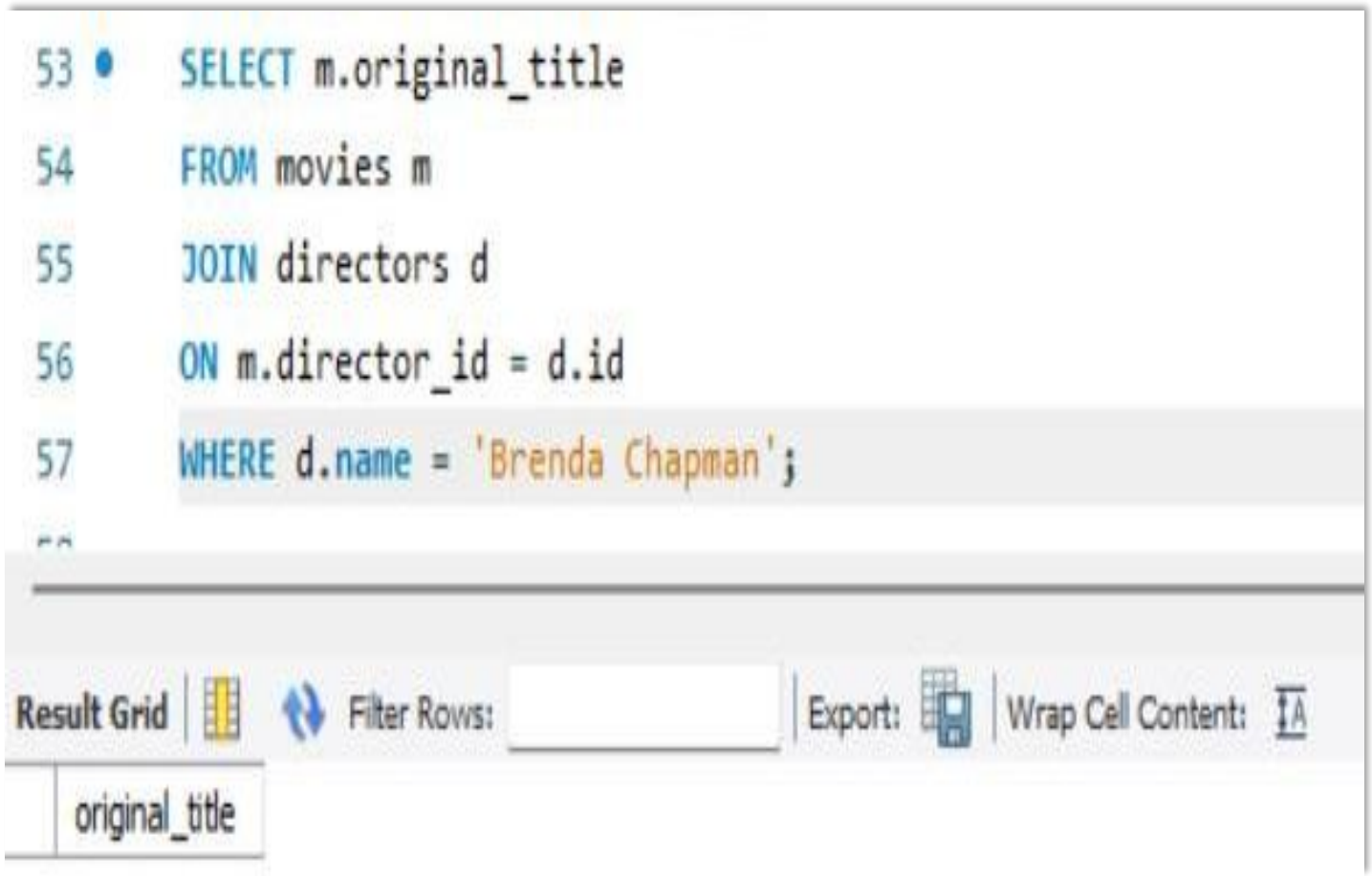
j. What is the most awarded average vote since the January 1st, 2000?

SELECT original_title,
vote_average, release_date
FROM movies
WHERE release_date >= '2000-01-01'
ORDER BY vote_average
DESC LIMIT 1;



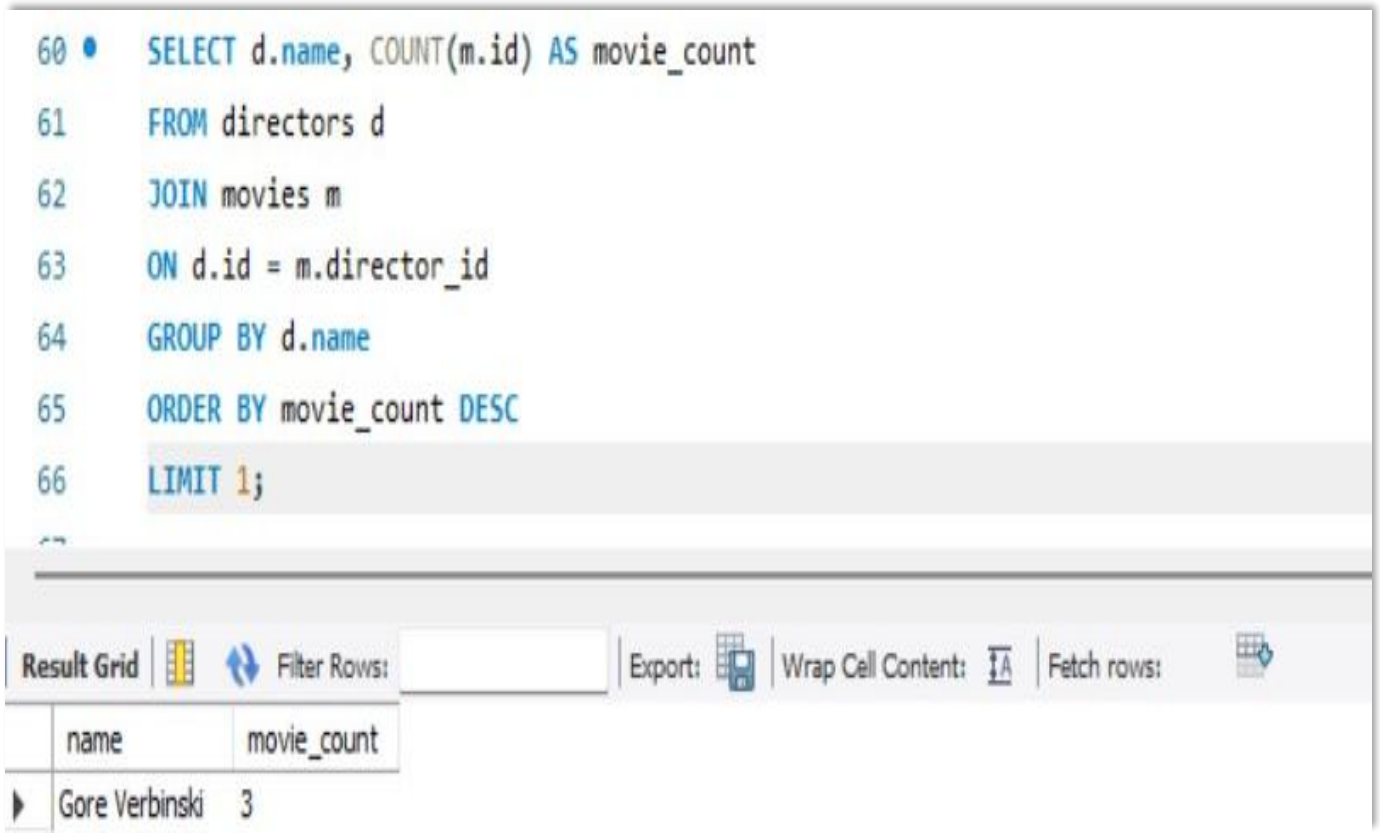
k. Which movie(s) were directed by Brenda Chapman?

```
SELECT m.original title
FROM movies m
JOIN directors d
ON m.director id = d.id
WHERE d.name = 'Brenda
Chapman';
```



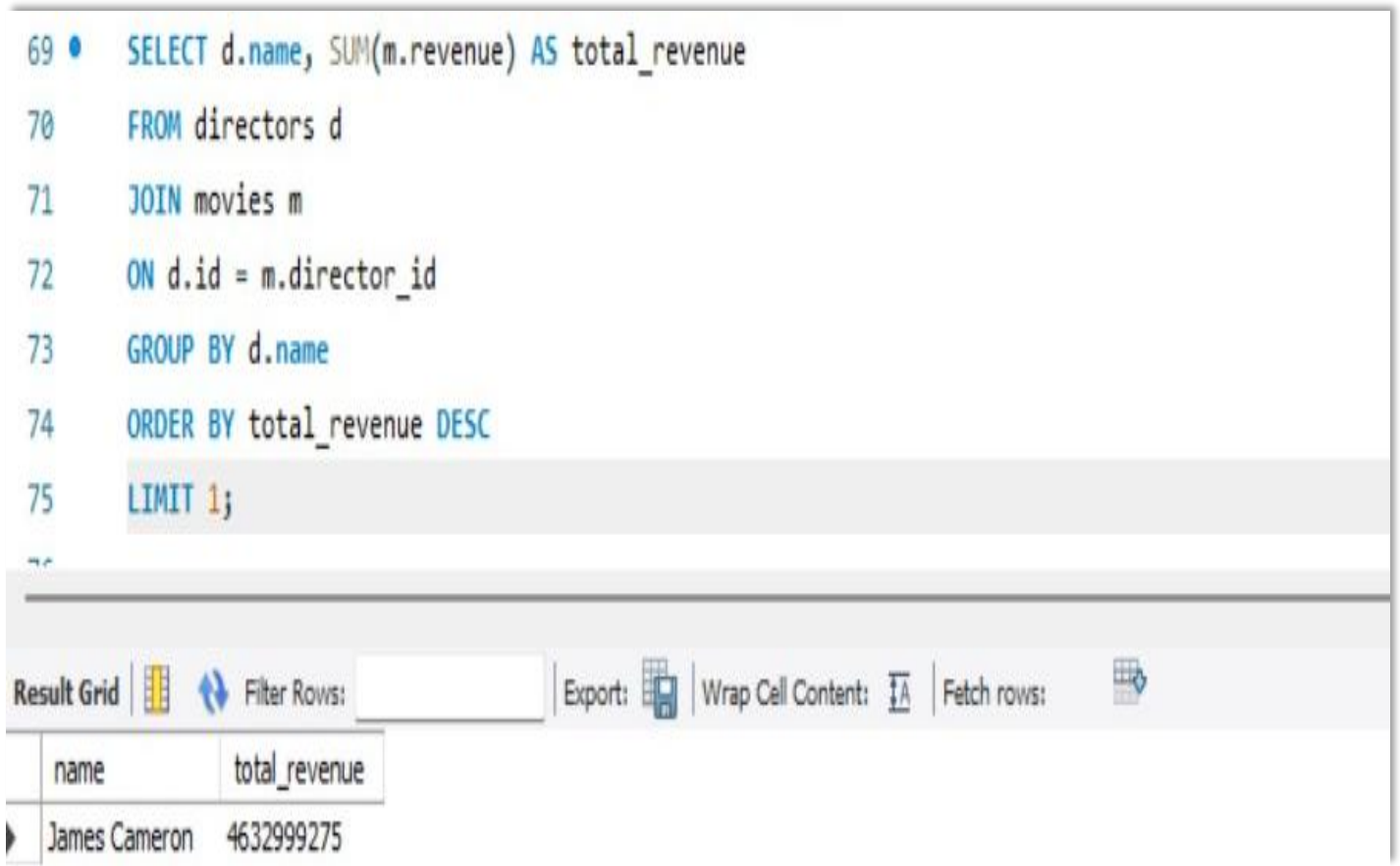
l. Which director made the most movies?

```
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;
```



m. 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
```



CONCLUSION

This report demonstrates how SQL can be effectively used to analyze a real-world IMDB-style dataset. By combining movie and director data, the analysis provides valuable insights into movie performance and director contributions. The assignment strengthens practical SQL skills such as querying, joining tables, aggregation, and analytical thinking, which are essential for data analysis roles.