ADVANCED MY SQL REINFORCEMENT PROJECT

1.COUNT THE TOTAL NUMBER OF RECORDS IN EACH TABLE OF THE DATABASE.

- The query gets the total number of records from multiple tables: `MOVIE`, `GENRE`, `DIRECTOR_MAPPING`, `ROLE_MAPPING`, `NAMES`, and `RATINGS`.
- For each table, it returns the table name and the count of its rows.
- The results from all tables are combined using `UNION ALL` to show a list of tables and their record counts.

```
select 'MOVIE 'AS TABLE_NAME ,count(*) AS TOTAL_RECORDS from MOVIE

UNION ALL

select 'GENRE',count(*) from genre

UNION ALL

select 'DIRECTOR_MAPPING', count(*) from director_mapping

UNION ALL

select' ROLE_MAPPING',count(*)  from role_mapping

UNION ALL

SELECT 'NAMES', COUNT(*)  FROM NAMES

UNION ALL

select 'NAMES',count(*)  FROM ratings;
```

Result Grid | Filter Rows:

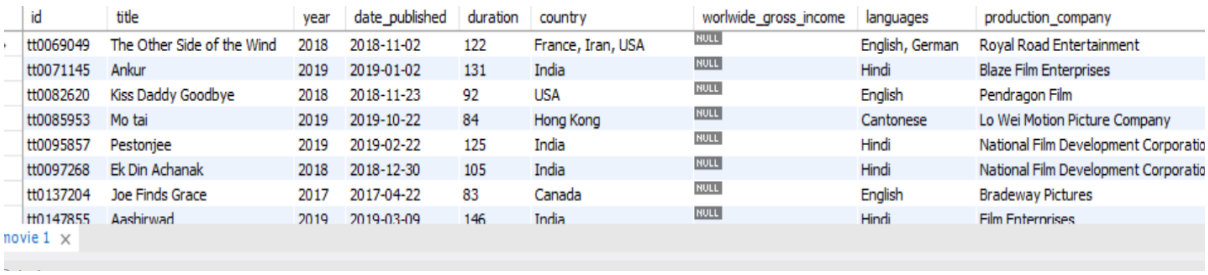
TABLE_NAME	TOTAL_RECORDS
MOVIE	7997
GENRE	14662
DIRECTOR_MAPPING	3867
ROLE_MAPPING	15615
NAMES	25735
NAMES	7997

2.IDENTIFY WHICH COLUMN IN THE MOVIE TABLE CONTAIN NULL VALUES.

- • The query selects all records from the `movie` table where **any** of these columns have **NULL** values:
  `id`, `title`, `year`, `date_published`, `duration`, `worldwide_gross_income`, `languages`, or `production_company`.
- • It helps identify rows with missing important information.

select * from movie

where id is null or title is null or year is null or date_published is null or duration is null or worlwide_gross_income is null or languages is null or production_company is null ;



id	title	year	date_published	duration	country	worlwide_gross_income	languages	production_company
tt0069049	The Other Side of the Wind	2018	2018-11-02	122	France, Iran, USA	NULL	English, German	Royal Road Entertainment
tt0071145	Ankur	2019	2019-01-02	131	India	NULL	Hindi	Blaze Film Enterprises
tt0082620	Kiss Daddy Goodbye	2018	2018-11-23	92	USA	NULL	English	Pendragon Film
tt0085953	Mo tai	2019	2019-10-22	84	Hong Kong	NULL	Cantonese	Lo Wei Motion Picture Company
tt0095857	Pestonjee	2019	2019-02-22	125	India	NULL	Hindi	National Film Development Corporatio
tt0097268	Ek Din Achanak	2018	2018-12-30	105	India	NULL	Hindi	National Film Development Corporatio
tt0137204	Joe Finds Grace	2017	2017-04-22	83	Canada	NULL	English	Bradeway Pictures
tt0147855	Aashirwad	2019	2019-03-09	146	India	NULL	Hindi	Film Enterprises

movie 1 ×

COUNT OF NULL COLUMNS

SELECT COUNT(*) AS TOTAL_ROWS,

COUNT(ID) AS ID_NOTNULL,

COUNT(WORLWIDE_GROSS_INCOME) AS WG_NOT_NULL,

COUNT(TITLE) AS TTILE_NOTNULL,

COUNT(DATE_PUBLISHED)AS DP_NOTNULL,

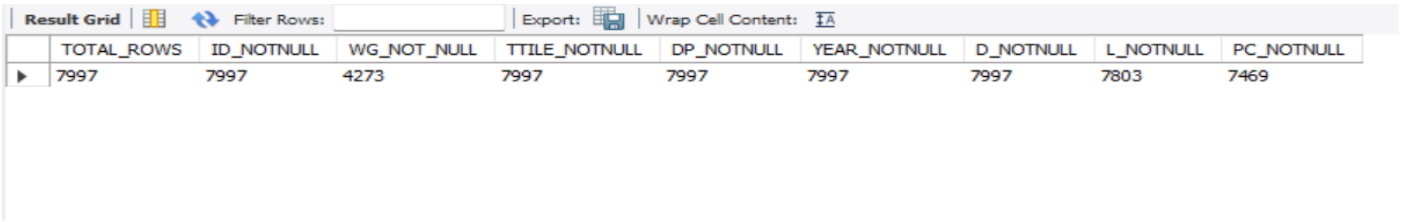
COUNT(YEAR) AS YEAR_NOTNULL,

COUNT(DURATION) AS D_NOTNULL,

COUNT(LANGUAGES) AS L_NOTNULL,

COUNT(PRODUCTION_COMPANY) AS PC_NOTNULL

FROM MOVIE;



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

TOTAL_ROWS	ID_NOTNULL	WG_NOT_NULL	TTILE_NOTNULL	DP_NOTNULL	YEAR_NOTNULL	D_NOTNULL	L_NOTNULL	PC_NOTNULL
7997	7997	4273	7997	7997	7997	7997	7803	7469

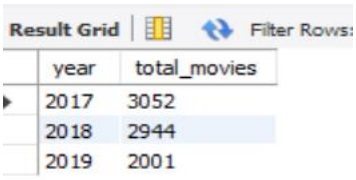
3. **1)TOTAL NUMBER OF MOVIES EACH YEAR**

- The query counts the total number of movies released in each **year**.
- It groups the results by the year column.
- The output shows each year along with the number of movies released in that year.

select year,count(year) as total_movies

from movie

group by year;

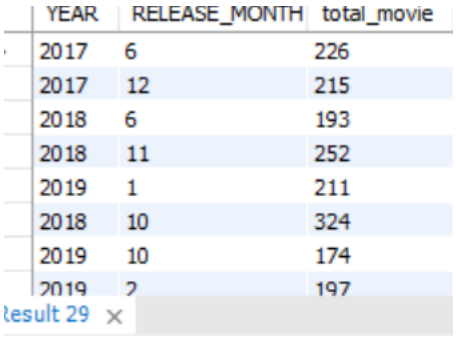


year	total_movies
2017	3052
2018	2944
2019	2001

2) MONTH WISE RELEASE EACH YEAR.

- The query counts how many movies were released in each **year** and **month**.
- It extracts the **month** from the DATE_PUBLISHED column.
- Then, it groups the data by **year** and **month**.
- The result shows the year, release month, and total number of movies released in that month.

```
SELECT YEAR,
 EXTRACT(MONTH FROM DATE_PUBLISHED) AS RELEASE_MONTH,
 COUNT(title) as total_movie
 FROM MOVIE
 GROUP BY
 YEAR,RELEASE_MONTH;
```



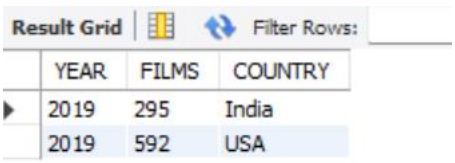
YEAR	RELEASE_MONTH	total_movie
2017	6	226
2017	12	215
2018	6	193
2018	11	252
2019	1	211
2018	10	324
2019	10	174
2019	2	197

Result 29

4.HOW MANY MOVIES WERE PRODUCED IN EITHER THE USA OR INDIAIN THE YEAR 2019.

- The query counts how many movies were released in **2019** from the countries **USA** and **India**.
- It groups the results by **year** and **country**.
- It uses the **HAVING** clause to filter only the year **2019**.
- The output shows the year, country, and the number of movies released.

```
SELECT YEAR,COUNT(TITLE) AS FILMS,COUNTRY
 FROM MOVIE
 WHERE COUNTRY in( 'USA', 'INDIA')
 GROUP BY YEAR,COUNTRY
having  year ='2019' ;
```

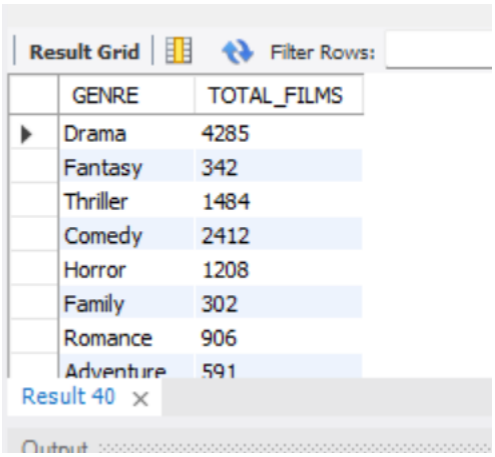


YEAR	FILMS	COUNTRY
2019	295	India
2019	592	USA

5.LIST THE UNIQUE GENRE ,AND COUNT HOW MANY MOVIES BELONG EXCLUSIVELY TO ONE GENRE.

- The query counts the total number of films for each **distinct** genre.
- It groups the records by genre and calculates the count for each genre.
- The result shows each genre along with the total number of films in that genre.

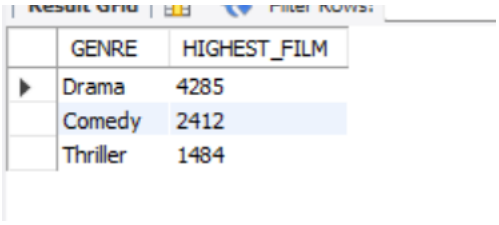
```
SELECT distinct GENRE ,COUNT(GENRE) AS TOTAL_FILMS

FROM GENRE

GROUP BY GENRE;
```



6. . Which genre has the highest total number of movies produced?

- The query counts how many movies belong to each genre.
- It groups the data by genre.
- Then, it orders the genres by the number of movies in descending order.
- Finally, it shows the top 3 genres with the highest number of movies.

```
SELECT GENRE ,COUNT(GENRE) AS HIGHEST_FILM

FROM GENRE

GROUP BY GENRE

ORDER BY HIGHEST_FILM DESC

LIMIT 3;
```



7. Calculate the average movie duration for each genre.

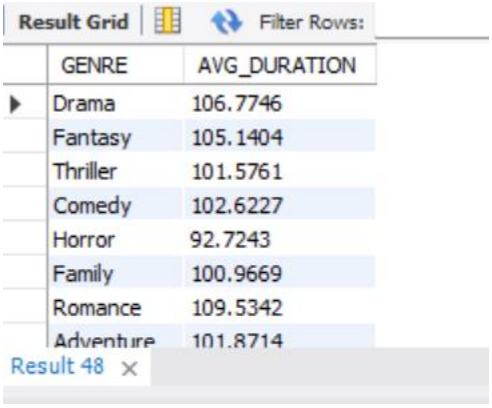
- The query calculates the **average duration** of movies for each genre.
- It joins the MOVIE and GENRE tables using the movie ID.
- Then, it groups the results by genre and computes the average duration for each group.

- The output shows each genre and its average movie duration
- SELECT G.GENRE,AVG(M.DURATION)AS AVG_DURATION

FROM MOVIE M

JOIN GENRE G ON M.ID=G.MOVIE_ID

GROUP BY G.GENRE;



8. Identify actors or actresses who have appeared in more than three movies with an average

rating below 5.

- It Finds the COUNT() function along with the GROUP BY clause to count the number of movies by their IDs**.**
- It a JOIN to combine **two tables**, linking each category to its average rating value**.**
- Then, I used the HAVING clause to filter results where the average rating is greater than 3**.**

SELECT RM.CATEGORY,R.MOVIE_ID, COUNT(R.MOVIE_ID)AS MOVIES,R.AVG_RATING

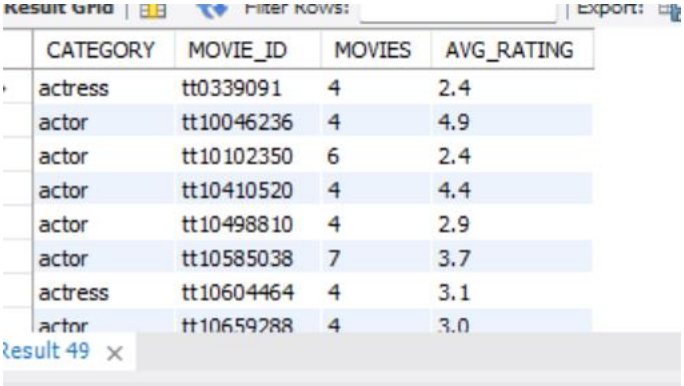
FROM ROLE_MAPPING RM

JOIN RATINGS R ON RM.MOVIE_ID=R.MOVIE_ID

WHERE AVG_RATING<5

GROUP BY RM.CATEGORY,R.MOVIE_ID,R.AVG_RATING

having COUNT(R.MOVIE_ID)>3;

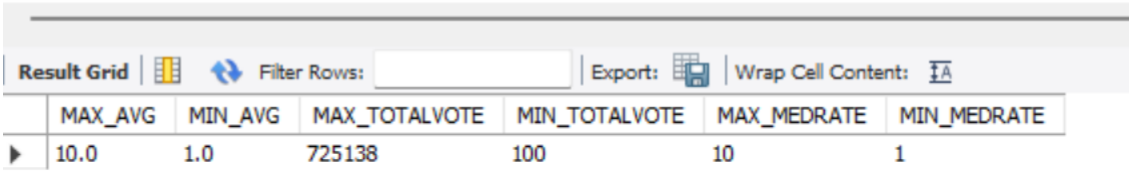


9. Find the minimum and maximum values for each column in the ratings table, excluding the

movie_id column.

- It selects data from the RATINGS table.
- It finds the maximum and minimum values for:
  - Average Rating (AVG_RATING)
  - Total Votes (TOTAL_VOTES)
  - Median Rating (MEDIAN_RATING)
- The result will be a single row showing the highest and lowest values for each of the three columns.

```
SELECT MAX(AVG_RATING) AS MAX_AVG ,MIN(AVG_RATING)AS MIN_AVG,

MAX(TOTAL_VOTES) AS MAX_TOTALVOTE,MIN(TOTAL_VOTES) AS MIN_TOTALVOTE,

MAX(MEDIAN_RATING) AS MAX_MEDRATE,MIN(MEDIAN_RATING)AS MIN_MEDRATE

FROM RATINGS;
```



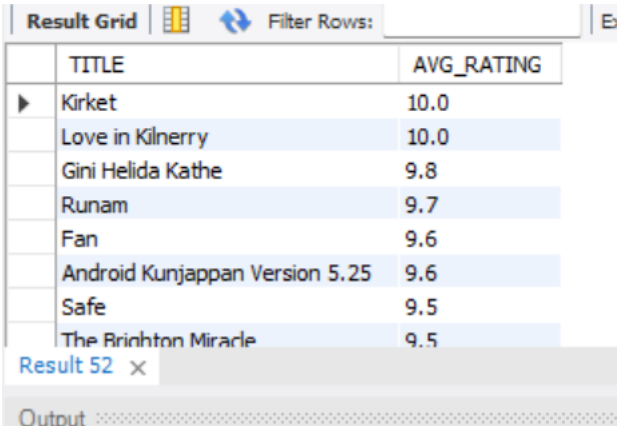
MAX_AVG	MIN_AVG	MAX_TOTALVOTE	MIN_TOTALVOTE	MAX_MEDRATE	MIN_MEDRATE
10.0	1.0	725138	100	10	1

10. Which are the top 10 movies based on their average rating?

- The query selects the **movie title** (M.TITLE) and its **average rating** (R.AVG_RATING).
- It **joins** the MOVIE table (M) with the RATINGS table (R) using the movie ID (M.ID = R.MOVIE_ID).
- Then, it **orders** the movies by their average rating in **descending order** (highest rated first).
- Finally, it **limits** the result to show only the **top 10** movies with the highest average ratings.

```
SELECT M.TITLE ,R.AVG_RATING

FROM MOVIE M

JOIN RATINGS R ON M.ID=R.MOVIE_ID

ORDER BY R.AVG_RATING DESC
```

LIMIT 10;



TITLE	AVG_RATING
Kirket	10.0
Love in Kilnerry	10.0
Gini Helida Kathe	9.8
Runam	9.7
Fan	9.6
Android Kunjappan Version 5.25	9.6
Safe	9.5
The Brighton Miracle	9.5

11. Summarize the ratings table by grouping movies based on their median ratings

- This query selects each movie's title and its median rating by joining the MOVIE and RATINGS tables on movie ID.
- It groups the results by movie title and median rating.

```
SELECT  m.title,R.MEDIAN_RATING AS MID_RATING

FROM MOVIE M

JOIN RATINGS R ON M.ID=R.MOVIE_ID

group by m.title,r.median_rating ;
```

title	MID_RATING
Der müde Tod	8
A Matter of Life and Death	8
Le roi de coeur	8
The Other Side of the Wind	7
Ankur	8
Kiss Daddy Goodbye	3
The Evil Dead	8
Mo tai	6

12. How many movies, released in March 2017 in the USA within a specific genre, had more

than 1,000 votes?

- The query gets movies from the USA released in 2017 with more than 1000 votes, shows their title, release month, year, country, genre, and total votes, and counts how many movies are in each genre.
- It filters the results to only include movies released in March.

```
select M.TITLE, (extract(month from m.date_published)) as RELEASE_MONTH,m.year,m.country,g.genre,r.total_votes,

COUNT(M.TITLE)OVER (PARTITION BY G.GENRE)AS MOVIE_COUNT

from  movie m

join genre g on m.id=g.movie_id
```

join ratings r on g.movie_id=r.movie_id

where m.country ='usa' and r.total_votes>1000 and m.year=2017

having RELEASE_MONTH= 3;

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

TITLE	RELEASE_MONTH	year	country	genre	total_votes	MOVIE_COUNT
CHIPS	3	2017	USA	Action	38693	4
Security	3	2017	USA	Action	14114	4
Bloodrunners	3	2017	USA	Action	1095	4
Logan	3	2017	USA	Action	586106	4
Wilson	3	2017	USA	Comedy	9219	8
Deidra & Laney Rob a Train	3	2017	USA	Comedy	1402	8
Fist Fight	3	2017	USA	Comedy	36526	8
All Nighter	3	2017	USA	Comedy	5476	8

Result 65

Output

Action Output

13. Find movies from each genre that begin with the word “The” and have an average rating greater than 8.

- The query selects the movie title, genre, and average rating for movies whose titles start with "the " and have an average rating greater than 8.
- It joins the movie, genre, and ratings tables by movie ID.

select m.title,g.genre,r.avg_rating

from movie m

join genre g on m.id=g.movie_id

join ratings r on g.movie_id=r.movie_id

where m.title like 'the %%%%' and avg_rating > 8;

Result Grid | Filter Rows: | Export: | Wr

title	genre	avg_rating
The Blue Elephant 2	Drama	8.8
The Blue Elephant 2	Horror	8.8
The Blue Elephant 2	Mystery	8.8
The Brighton Miracle	Drama	9.5
The Irishman	Crime	8.7
The Irishman	Drama	8.7
The Colour of Darkness	Drama	9.1
The Mystery of Godliness: The Sequel	Drama	8.5

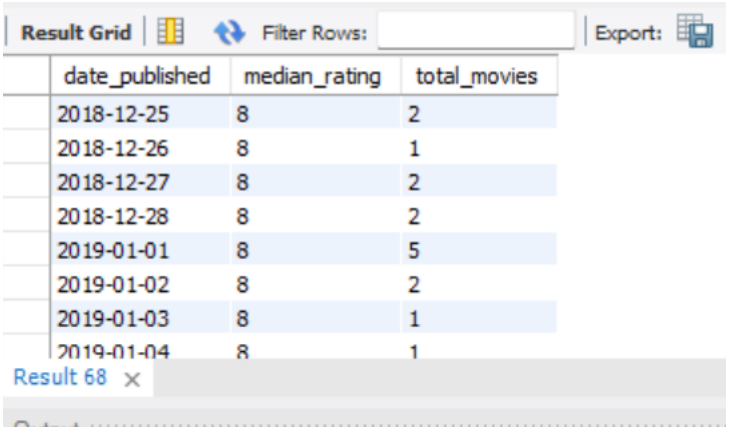
Result 66

Output

14. Of the movies released between April 1, 2018, and April 1, 2019, how many received a median rating of 8?

- The query counts the number of movies published between April 1, 2018, and April 1, 2019, that have a median rating of 8.
- It shows the date published, median rating, and total movies for each date, ordered by the publishing date from oldest to newest.

```
select m.date_published,r.median_rating,count(m.title)as total_movies
from movie m
join ratings r on m.id=r.movie_id
where date_published between ' 2018-4-1' and '2019-4-1' and r.median_rating=8
group by m.date_published,r.median_rating
order by date_published asc;
```



Result Grid | Filter Rows: | Export:

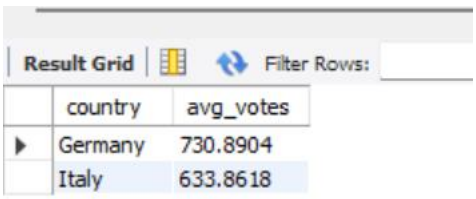
date_published	median_rating	total_movies
2018-12-25	8	2
2018-12-26	8	1
2018-12-27	8	2
2018-12-28	8	2
2019-01-01	8	5
2019-01-02	8	2
2019-01-03	8	1
2019-01-04	8	1

Result 68

15. Do German movies receive more votes on average than Italian movies?

- The query calculates the average total votes for movies from Germany and Italy.
- It groups the results by country to show the average votes for each.

```
select m.country,avg(r.total_votes) as avg_votes
from movie m
join ratings r on m.id=r.movie_id
where country in ('germany','italy')
group by m.country;
```



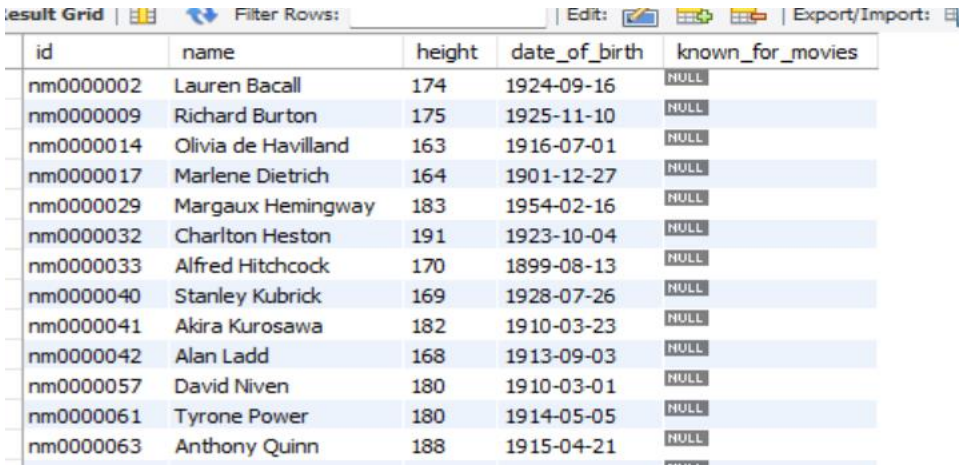
Result Grid | Filter Rows:

country	avg_votes
Germany	730.8904
Italy	633.8618

16. Identify the columns in the names table that contain null values

- The query selects all rows from the names table where any of the following columns have NULL values: id, name, height, date_of_birth, or known_for_movies.

```
select * from names

where id is null or name is null or height is null or date_of_birth is null or known_for_movies is null;
```

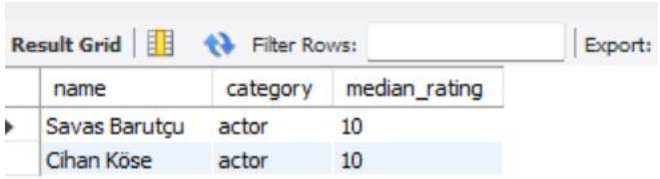


id	name	height	date_of_birth	known_for_movies
nm0000002	Lauren Bacall	174	1924-09-16	NULL
nm0000009	Richard Burton	175	1925-11-10	NULL
nm0000014	Olivia de Havilland	163	1916-07-01	NULL
nm0000017	Marlene Dietrich	164	1901-12-27	NULL
nm0000029	Margaux Hemingway	183	1954-02-16	NULL
nm0000032	Charlton Heston	191	1923-10-04	NULL
nm0000033	Alfred Hitchcock	170	1899-08-13	NULL
nm0000040	Stanley Kubrick	169	1928-07-26	NULL
nm0000041	Akira Kurosawa	182	1910-03-23	NULL
nm0000042	Alan Ladd	168	1913-09-03	NULL
nm0000057	David Niven	180	1910-03-01	NULL
nm0000061	Tyrone Power	180	1914-05-05	NULL
nm0000063	Anthony Quinn	188	1915-04-21	NULL

17. Who are the top two actors whose movies have a median rating of 8 or higher?

- The query lists the top 2 actors (rm.category = 'actor') with movies that have a median rating of 8 or higher.
- It shows the actor's name, their category, and the movie's median rating, sorted from highest to lowest rating.

```
select  n.name,rm.category,r.median_rating

from names n

join role_mapping rm on n.id=rm.name_id

join  ratings r on rm.movie_id= r.movie_id

where rm.category ='actor' and r.median_rating >=8

order by r.median_rating desc

limit 2;
```



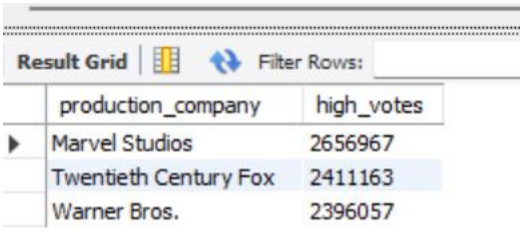
name	category	median_rating
Savas Barutçu	actor	10
Cihan Köse	actor	10

18. Which are the top three production companies based on the total number of votes their

movies received?

- The query finds the top 3 production companies with the highest total votes across all their movies.
- It sums the total votes for each company, groups by production company, and sorts the results from highest to lowest total votes.

```
select m.production_company ,sum(r.total_votes) as high_votes

from movie m

join ratings r on m.id=r.movie_id

group by m.production_company

order by high_votes desc

limit 3;
```



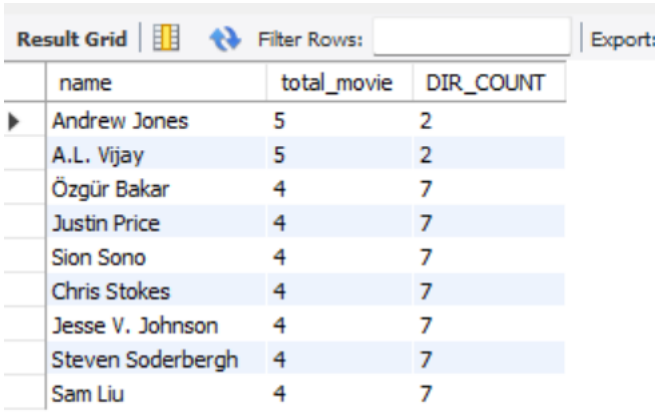
Result Grid | Filter Rows:

production_company	high_votes
Marvel Studios	2656967
Twentieth Century Fox	2411163
Warner Bros.	2396057

19. How many directors have worked on more than three movies?

- The query counts how many movies each director has directed (more than 3), lists their names and total movies, and tries to count how many directors share the same movie count .
- Results are ordered by total movies directed, highest first.

```
select n.name ,count(dm.movie_id) as total_movie,

COUNT(N.NAME)OVER (PARTITION BY  count(dm.movie_id))AS DIR_COUNT

from names n

join director_mapping dm on n.id=dm.name_id

group by n.name ,dm.name_id

having total_movie >3

order by total_movie desc;
```



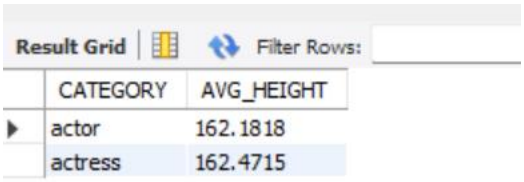
Result Grid | Filter Rows: | Export:

name	total_movie	DIR_COUNT
Andrew Jones	5	2
A.L. Vijay	5	2
Özgür Bakar	4	7
Justin Price	4	7
Sion Sono	4	7
Chris Stokes	4	7
Jesse V. Johnson	4	7
Steven Soderbergh	4	7
Sam Liu	4	7

20. Calculate the average height of actors and actresses separately.

- The query calculates the average height of people in the categories 'ACTOR' and 'ACTRESS'.
- It groups the results by category and shows the average height for each group.

```
SELECT RM.CATEGORY,AVG(N.HEIGHT) AS AVG_HEIGHT

FROM NAMES N

JOIN ROLE_MAPPING RM ON N.ID=RM.NAME_ID

WHERE CATEGORY ='ACTOR' OR CATEGORY ='ACTRESS'

GROUP BY RM.CATEGORY;
```



Result Grid | Filter Rows:

CATEGORY	AVG_HEIGHT
actor	162.1818
actress	162.4715

21. List the 10 oldest movies in the dataset along with their title, country, and director.

- The query retrieves the title, country, publication date of movies, and the director's name.
- It lists the 10 oldest movies by sorting the movies from the earliest to latest date published.

```
SELECT M.TITLE,M.COUNTRY AS OLDEST_MOVIE ,M.DATE_PUBLISHED,N.NAME

FROM MOVIE M

JOIN DIRECTOR_MAPPING DM ON M.ID=DM.MOVIE_ID

JOIN NAMES N ON DM.NAME_ID=N.ID
```

ORDER BY M.DATE_PUBLISHED asc

LIMIT 10;

TITLE	OLDEST_MOVIE	DATE_PUBLISHED	NAME
Sleeping Beauties	USA	2017-01-01	Dean McKendrick
Yol: The Full Version	Switzerland, Turkey	2017-01-01	Serif Gören
The Darkest	France	2017-01-01	Robin Entreinger
Sobre Nós	Brazil	2017-01-01	Mauro Carvalho
Nagarkirtan	India	2017-01-01	Kaushik Ganguly
Destined	USA	2017-01-01	Qasim Basir
The Beautiful Ones	USA	2017-01-01	Jesse V. Johnson
Furthest Witness	UK, USA	2017-01-01	Adam Del Giudice
Haunted	Italy	2017-01-01	Eros D Antona
Fetish Factory	USA	2017-01-01	Staci Layne Wilson

22. List the top 5 movies with the highest total votes, along with their genres.

- The query gets data from three tables: MOVIE, RATINGS, and GENRE.
- It joins these tables using the movie ID to combine information about each movie's title, total votes, and genre.
- It sorts all movies by the total number of votes they received, from the highest to the lowest.
- Finally, it shows the top 5 movies with the most votes, displaying their title, total votes, and genre.

SELECT M.TITLE, R.TOTAL_VOTES,G.GENRE

FROM MOVIE M

JOIN RATINGS R ON M.ID=R.MOVIE_ID

JOIN GENRE G ON R.MOVIE_ID=G.MOVIE_ID

ORDER BY R.TOTAL_VOTES desc

LIMIT 15;

TITLE	TOTAL_VOTES	GENRE
Avengers: Infinity War	725138	Action
Avengers: Infinity War	725138	Adventure
Avengers: Infinity War	725138	Sci-Fi
Avengers: Endgame	602792	Action
Avengers: Endgame	602792	Adventure
Avengers: Endgame	602792	Drama
Logan	586106	Action
Logan	586106	Drama
Logan	586106	Sci-Fi
Black Panther	551245	Action
Black Panther	551245	Adventure
Black Panther	551245	Sci-Fi
Thor: Ragnarok	518571	Action

Result 90

23. Identify the movie with the longest duration, along with its genre and production company.

- The query selects the **movie title**, **production company**, and **genre**.
- It calculates the **longest duration** for each movie title using a window function (MAX(DURATION) OVER (PARTITION BY M.TITLE)), though since it's grouped by title, this will just be the movie's own duration if titles are unique.
- It joins the MOVIE and GENRE tables by movie ID.
- Finally, it orders all movies by their duration in descending order and shows only the **movie with the longest duration**.

```
SELECT  M.TITLE,M.PRODUCTION_COMPANY,G.GENRE,

MAX(DURATION)OVER (PARTITION BY M.TITLE) AS LONGEST_DURATION

FROM MOVIE M

JOIN GENRE G ON M.ID=G.MOVIE_ID

ORDER BY M.DURATION DESC

LIMIT 1;
```

TITLE	PRODUCTION_COMPANY	GENRE	LONGEST_DURATION
La flor	El Pampero Cine	Drama	808

24. Determine the total number of votes for each movie released in 2018.

- The query selects the **movie title**, **year**, and **total votes** for movies released in the year **2018**.
- It joins the `MOVIE` and `RATINGS` tables using the movie ID.
- It filters the results to include only movies from **2018**.
- Then, it orders the movies by the number of votes in **descending order**, showing the most voted movies first.

```
SELECT M.TITLE,M.YEAR,R.TOTAL_VOTES AS VOTES

FROM  MOVIE M

JOIN RATINGS R ON M.ID=R.MOVIE_ID

WHERE YEAR ='2018'

ORDER BY VOTES DESC ;
```

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

TITLE	YEAR	VOTES
Avengers: Infinity War	2018	725138
Black Panther	2018	551245
Deadpool 2	2018	424573
Bohemian Rhapsody	2018	392599
Three Billboards Outside Ebbing, Missouri	2018	367832
A Quiet Place	2018	333041
The Shape of Water	2018	326832
Ready Player One	2018	320181
Venom	2018	312437
Aquaman	2018	311374
Coco	2018	305662
A Star Is Born	2018	284195
Spider-Man: Into the Spider-Verse	2018	275653

25 . What is the most common language in which movies were produced?

- The query selects the **movie title**, **languages**, and **production company** from the `movie` table.
- It uses a window function `COUNT(LANGUAGES) OVER (PARTITION BY LANGUAGES)` to count how many movies there are for each language.
- The result includes a new column `movie_count` showing the total number of movies for that language.
- Finally, it orders the results so that movies in the languages with the most movies appear first.

```
SELECT TITLE,languages,PRODUCTION_COMPANY,

count(LANGUAGES) OVER (PARTITION BY LANGUAGES) as movie_count

from movie

order by movie_count desc;
```

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

TITLE	languages	PRODUCTION_COMPANY	movie_count
Book of Monsters	English	Dark Rift Films	3095
The Huntress: Rune of the Dead	English	ITN Films	3095
Deadcon	English	Gunpowder & Sky	3095
I Love You, Daddy	English	Circus King Productions	3095
A.M.I.	English	1160594 B.C.	3095
Instakiller	English	Cartel Pictures	3095
#Followme	English	Samurai Films	3095
A Christmas Prince: The Royal Wedding	English	Motion Picture Corporation of America (MPCA)	3095
The Manson Family Massacre	English	North Bank Entertainment	3095
Demon	English	WOWNow Entertainment	3095
I Trapped the Devil	English	NULL	3095