IMDB DATA REVIEW

MYSQL Project

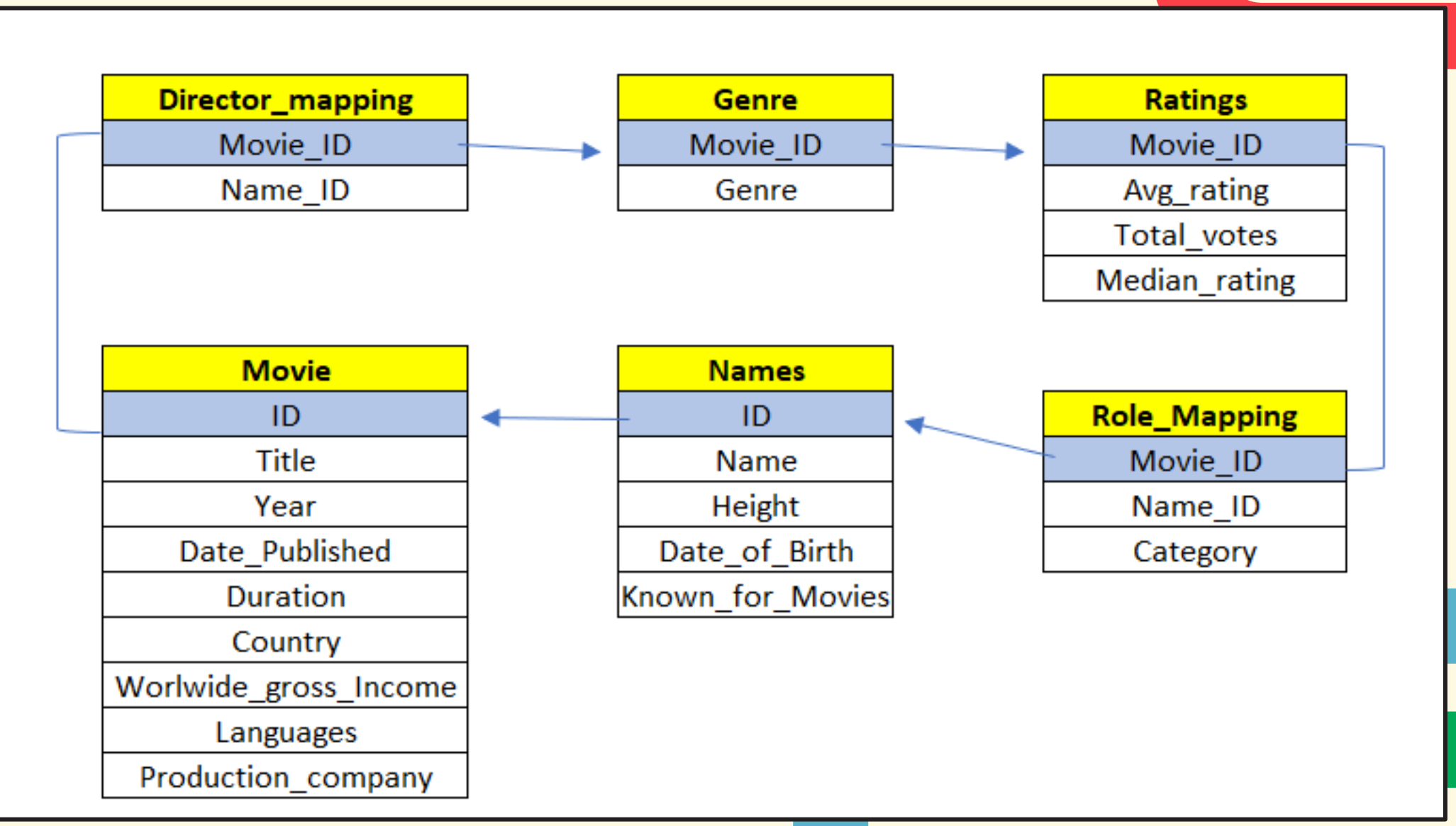
➢ Objective

- Analyze and derive insights from the IMDB movie database using SQL queries.
- Answer specific questions related to movie production trends, genres, and regional data.
- Explore correlations between various movie attributes, such as duration and genre distribution.

➢ Question Modes

- Easy – Queries Include : SELECT, GROUP BY, ORDER BY, LIMIT,DESC.
- Moderate – Queries Include : JOINS, GROUP BY, ORDER BY, LIMIT.
- Advance – Queries Include : CTE (COMMON TABLE EXPRESSION.

➢ IMDB Database



Level - Basic

Q.1. Find the total number of rows in Movie table of the schema ?

- **Input**

```
select count(*) as movie_row_count from movie;
```

- **Output**

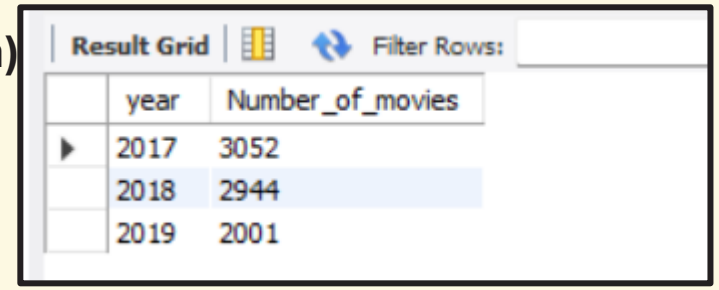
Level - Basic

Q.2. (a) Find the total number of movies released each year? (b)How does the trend look month wise?

- **Input(a)**

```
select year, count(*) as Number_of_movies from movie group by year;
```

- **Output(a)**



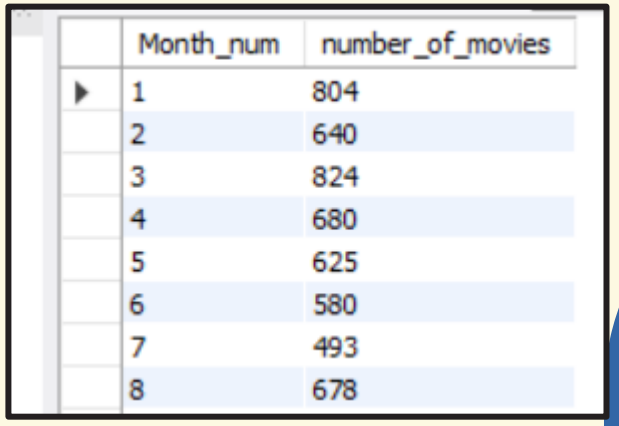
Result Grid | Filter Rows:

year	Number_of_movies
2017	3052
2018	2944
2019	2001

- **Input(b)**

```
Select month(date_published) as Month_num, count(*) as number_of_movies from movie group by month_num order by Month_num asc;
```

- **Output(b)**



Month_num	number_of_movies
1	804
2	640
3	824
4	680
5	625
6	580
7	493
8	678

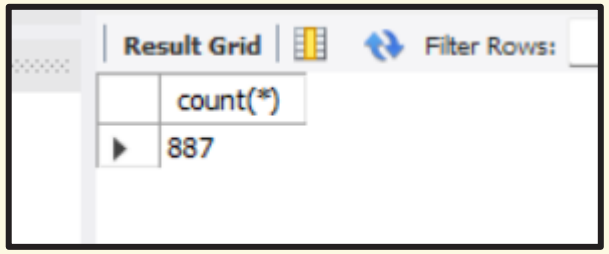
Level - Basic

Q.3. How many movies were produced in the USA or India in the year 2019 ?

- **Input**

```
select count(*) from movie where year=2019 and (country like 'India' or
country like 'usa');
```

- **Output**



Level - Basic

Q.4. Find the unique list of the genres present in the data set ?

- Input

```
select distinct genre from genre;
```

- Output

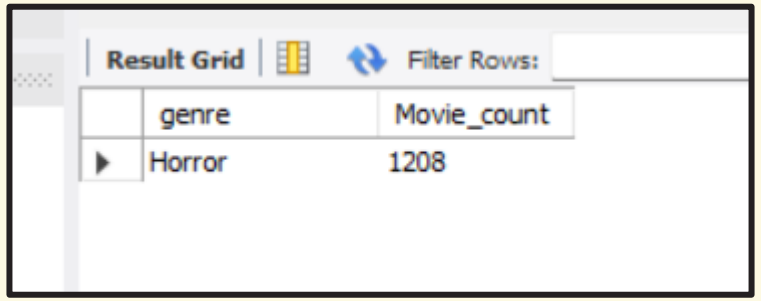
Level - Basic

Q.5. Which genre had the highest number of movies produced overall ?

- **Input**

```
Select genre, Count(*) as Movie_count from genre group by genre having genre='horror';
```

- **Output**



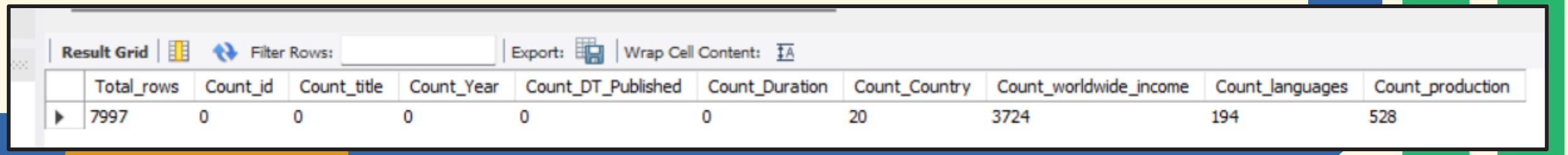
Level - Moderate

Q.6. Which columns in the movie table have null values ?

- Input

```
Select count(*) as Total_rows,sum(case when id is null then 1 else 0 end) Count_id,sum(case when title is null then 1 else 0 end) Count_title,sum(case when year is null then 1 else 0 end) Count_Year,sum(case when date_published is null then 1 else 0 end) Count_DT_Published,sum(case when duration is null then 1 else 0 end) Count_Duration,sum(case when country is null then 1 else 0 end) Count_Country,sum(case when worlwide_gross_income is null then 1 else 0 end) Count_worldwide_income,sum(case when languages is null then 1 else 0 end) Count_languages,sum(case when production_company is null then 1 else 0 end) Count_production from movie;
```

- Output



Total_rows	Count_id	Count_title	Count_Year	Count_DT_Published	Count_Duration	Count_Country	Count_worldwide_income	Count_languages	Count_production
7997	0	0	0	0	0	20	3724	194	528

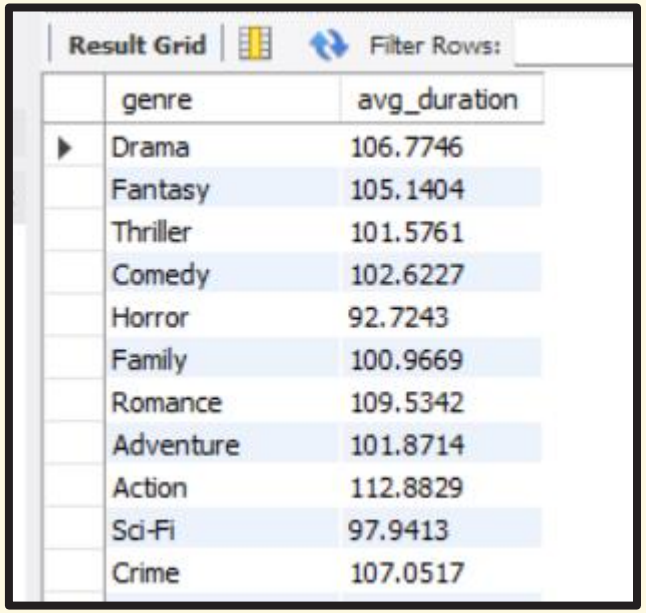
Level - Moderate

Q.7. What is the average duration of movies in each genre ?

- Input

```
Select genre.genre, avg(movie.duration) as avg_duration from genre inner join movie
on genre.movie_id = movie.id group by genre.genre;
```

- Output



Result Grid | Filter Rows:

genre	avg_duration
Drama	106.7746
Fantasy	105.1404
Thriller	101.5761
Comedy	102.6227
Horror	92.7243
Family	100.9669
Romance	109.5342
Adventure	101.8714
Action	112.8829
Sci-Fi	97.9413
Crime	107.0517

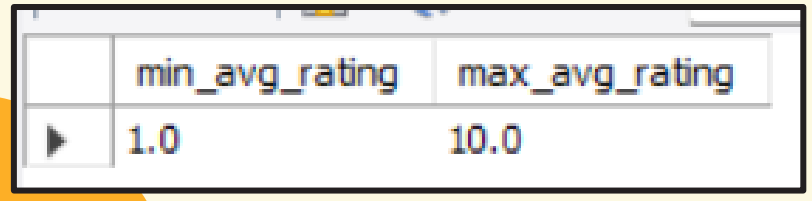
Level - Moderate

Q.8. Find the minimum and maximum values in each column of the ratings table except the movie_id column ?

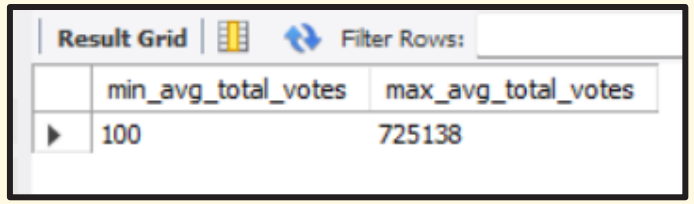
- Input

```
Select min(avg_rating) as min_avg_rating, max(avg_rating) as max_avg_rating from ratings;

select min(Median_rating) as min_avg_Median_Rating, max(median_rating) as
max_avg_Median_rating from ratings;

select min(Total_votes) as min_avg_total_votes, max(Total_votes) as max_avg_total_votes
from ratings;
```

- Output



min_avg_rating	max_avg_rating
1.0	10.0



Result Grid | Filter Rows:

min_avg_total_votes	max_avg_total_votes
100	725138

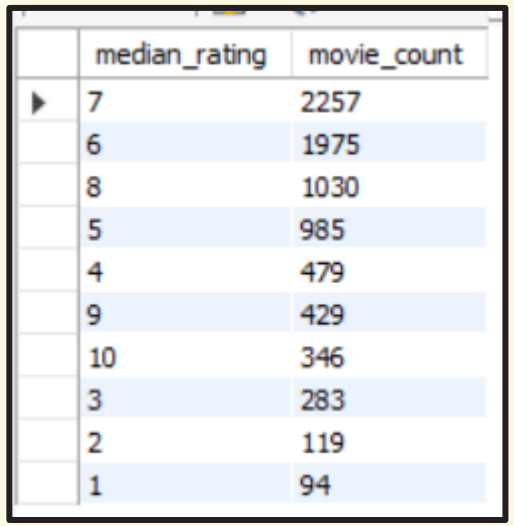
Level - Moderate

Q.9. Summaries the ratings table based on the movie counts by median ratings ?

- **Input**

```
select median_rating, count(movie_id) as movie_count from ratings group by median_rating
order by movie_count desc;
```

- **Output**



median_rating	movie_count
7	2257
6	1975
8	1030
5	985
4	479
9	429
10	346
3	283
2	119
1	94

Level - Moderate

Q.10. How many movies released in each genre during March 2017 in the USA had more than 1,000 votes ?

- **Input**

```
select g.genre, count(g.movie_id) as Movie_count from genre as g JOIN movie as m on m.id =
g.movie_idJOIN ratings as r on g.movie_id = r.Movie_idwhere year=2017 and
month(m.date_published) = 3 and country = 'usa' and r.total_votes > 1000group by g.genre order
by movie_count desc;
```

- **Output**

Result Grid | Filter Rows:

genre	Movie_count
Drama	16
Comedy	8
Crime	5
Horror	5
Action	4
Sci-Fi	4
Thriller	4
Romance	3
Fantasy	2
Mystery	2
Family	1

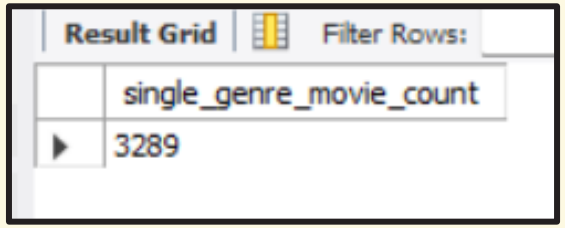
Level - Advance

Q.11. How many movies belong to only one genre ?

- Input

```
With movie_summary as
(select movie_id, count(*) as count_genre from genre group by movie_id)
select count(distinct Movie_id) as single_genre_movie_count from movie_summary where
count_genre=1;
```

- Output



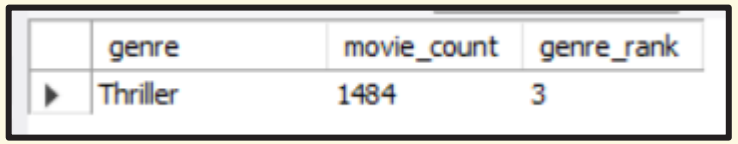
Level - Advance

Q.12. What is the rank of the ‘thriller’ genre of movies among all the genres in terms of number of movies produced ?

- **Input**

```
With Sunil as
(select genre, count(*) as movie_count, rank() over (order by count(*) desc) as genre_rank from
genregroup by genre)
select * from sunil where genre='thriller’;
```

- **Output**



	genre	movie_count	genre_rank
▶	Thriller	1484	3

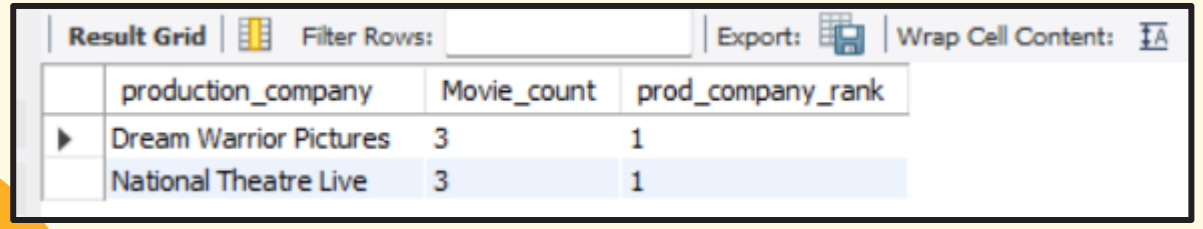
Level - Advance

Q.13. Which production house has produced the most number of hit movies (average rating > 8) ?

- **Input**

```
with Top_production as
(select m.production_company, count(m.id) as Movie_count, dense_rank() over (order by count(m.id) desc) asprod_company_rank from movie as m left join ratings as ron m.id = r.movie_id where r.avg_rating >8 AND m.production_company IS NOT NULL group by m.production_company)
select * from Top_production where prod_company_rank =1;
```

- **Output**



production_company	Movie_count	prod_company_rank
Dream Warrior Pictures	3	1
National Theatre Live	3	1

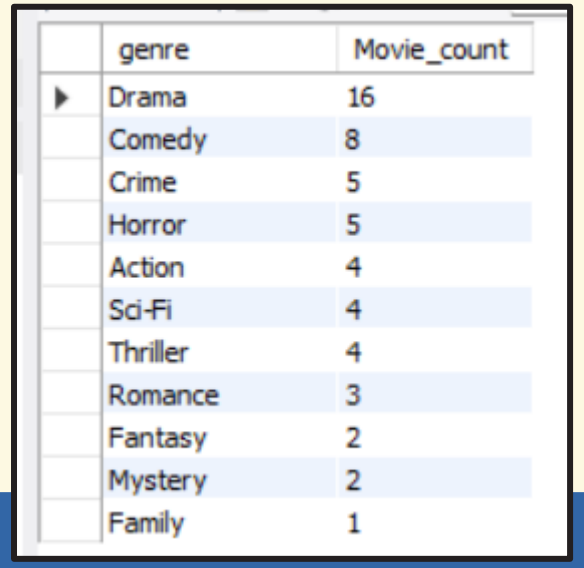
Level - Advance

Q.14. How many movies released in each genre during March 2017 in the USA had more than 1,000 votes?

- **Input**

```
select g.genre, count(g.movie_id) as Movie_count from genre as g JOIN movie as m on
m.id = g.movie_id JOIN ratings as r on g.movie_id = r.Movie_idwhere year=2017 and
month(m.date_published) = 3 and country = 'usa' and r.total_votes > 1000
group by g.genre order by movie_count desc;
```

- **Output**



genre	Movie_count
Drama	16
Comedy	8
Crime	5
Horror	5
Action	4
Sci-Fi	4
Thriller	4
Romance	3
Fantasy	2
Mystery	2
Family	1

Level - Advance

Q.15. Which columns in the names table have null values ?

- Input

```
select count(case when name is NULL then id END) as Name_null,
count(case when height is NULL then id END) AS height_nulls,
count(case when date_of_birth is NULL theN id END) as Date_of_Birth_nulls,
count(case when known_for_movies is NULL then id END) as known_for_movies_nulls
from names;
```

- Output

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

	Name_null	height_nulls	Date_of_Birth_nulls	known_for_movies_nulls
▶	0	17335	13431	15226

Project Resource –

CSV File Database Link : http://surl.li/nwulru

GitHub Project Link :

Profile Links :

LinkedIn Profile : www.linkedin.com/in/sunilk2498

Mail ID : sunilvishwanathkharat@gmail.com

Resume : http://surl.li/nmgavw