–––USE imdb;

/\* Now that you have imported the data sets, let’s explore some of the tables.

To begin with, it is beneficial to know the shape of the tables and whether any column has null values.

Further in this segment, you will take a look at 'movies' and 'genre' tables.\*/

-- Segment 1:

-- Q1. Find the total number of rows in each table of the schema?

-- Type your code below:

1.practice=# select count(\*) from movie ;

count

-------

7997

(1 row)

2.practice=# select count(\*) from genre ;

count

-------

14662

(1 row)

3.practice=# select count(\*) from director\_mapping ;

count

-------

3867

(1 row)

-- Q2. Which columns in the movie table have null values?

-- Type your code below:

practice=# select \* from movie where title = null and year = null and date\_published = null and duration = null and country = null and 'movie.worldwide\_groos\_income' = null and languages = null and production\_company = null;

id | title | year | date\_published | duration | country | worlwide\_gross\_income | languages | production\_company

----+-------+------+----------------+----------+---------+-----------------------+-----------+--------------------

(0 rows)

-- Now as you can see four columns of the movie table has null values. Let's look at the at the movies released each year.

-- Q3. Find the total number of movies released each year? How does the trend look month wise? (Output expected)

/\* Output format for the first part:

+---------------+-------------------+

| Year | number\_of\_movies|

+-------------------+----------------

| 2017 | 2134 |

| 2018 | . |

| 2019 | . |

+---------------+-------------------+

Actual o/p:-

practice=# select year as release\_year, count(title) as number\_of\_movies from movie group by year;

release\_year | number\_of\_movies

--------------+------------------

2018 | 2944

2017 | 3052

2019 | 2001

(3 rows)

Output format for the second part of the question:

+---------------+-------------------+

| month\_num | number\_of\_movies|

+---------------+----------------

| 1 | 134 |

| 2 | 231 |

| . | . |

+---------------+-------------------+ \*/

-- Type your code below:

practice=# select date\_trunc('month', date\_published)month\_num,count(title) as number\_of\_movies from movie group by month\_num order by month\_num;

practice=#

month\_num | number\_of\_movies

---------------------------+------------------

2017-01-01 00:00:00+05:30 | 291

2017-02-01 00:00:00+05:30 | 228

2017-03-01 00:00:00+05:30 | 298

2017-04-01 00:00:00+05:30 | 249

2017-05-01 00:00:00+05:30 | 205

2017-06-01 00:00:00+05:30 | 226

2017-07-01 00:00:00+05:30 | 188

2017-08-01 00:00:00+05:30 | 246

2017-09-01 00:00:00+05:30 | 327

2017-10-01 00:00:00+05:30 | 303

2017-11-01 00:00:00+05:30 | 276

2017-12-01 00:00:00+05:30 | 215

2018-01-01 00:00:00+05:30 | 302

2018-02-01 00:00:00+05:30 | 215

2018-03-01 00:00:00+05:30 | 285

2018-04-01 00:00:00+05:30 | 247

2018-05-01 00:00:00+05:30 | 229

2018-06-01 00:00:00+05:30 | 193

2018-07-01 00:00:00+05:30 | 167

2018-08-01 00:00:00+05:30 | 247

2018-09-01 00:00:00+05:30 | 276

2018-10-01 00:00:00+05:30 | 324

2018-11-01 00:00:00+05:30 | 252

2018-12-01 00:00:00+05:30 | 207

:

/\*The highest number of movies is produced in the month of March.

So, now that you have understood the month-wise trend of movies, let’s take a look at the other details in the movies table.

We know USA and India produces huge number of movies each year. Lets find the number of movies produced by USA or India for the last year.\*/

-- Q4. How many movies were produced in the USA or India in the year 2019??

-- Type your code below:

practice=# select country,count(title) as total\_number\_of\_movies from movie where (country ilike 'india' or country ilike 'usa')and year=2019 group by country;

country | total\_number\_of\_movies

---------+------------------------

India | 295

USA | 592

(2 rows)

/\* USA and India produced more than a thousand movies(you know the exact number!) in the year 2019.

Exploring table Genre would be fun!!

Let’s find out the different genres in the dataset.\*/

-- Q5. Find the unique list of the genres present in the data set?

-- Type your code below:

practice=# select distinct genre from genre ;

genre

-----------

Crime

Romance

Horror

Drama

Action

Mystery

Fantasy

Thriller

Sci-Fi

Comedy

Adventure

Others

Family

(13 rows)

/\* So, RSVP Movies plans to make a movie of one of these genres.

Now, wouldn’t you want to know which genre had the highest number of movies produced in the last year?

Combining both the movie and genres table can give more interesting insights. \*/

-- Q6.Which genre had the highest number of movies produced overall?

-- Type your code below:

practice=# select count(genre),genre from movie full outer join genre on id=movie\_id group by genre order by count desc;

count | genre

-------+-----------

4285 | Drama

2412 | Comedy

1484 | Thriller

1289 | Action

1208 | Horror

906 | Romance

813 | Crime

591 | Adventure

555 | Mystery

375 | Sci-Fi

342 | Fantasy

302 | Family

100 | Others

(13 rows)

/\* So, based on the insight that you just drew, RSVP Movies should focus on the ‘Drama’ genre.

But wait, it is too early to decide. A movie can belong to two or more genres.

So, let’s find out the count of movies that belong to only one genre.\*/

-- Q7. How many movies belong to only one genre?

-- Type your code below:

/\* There are more than three thousand movies which has only one genre associated with them.

So, this figure appears significant.

Now, let's find out the possible duration of RSVP Movies’ next project.\*/

-- Q8.What is the average duration of movies in each genre?

-- (Note: The same movie can belong to multiple genres.)

/\* Output format:

+---------------+-------------------+

| genre | avg\_duration |

+-------------------+----------------

| thriller | 105 |

| . | . |

| . | . |

+---------------+-------------------+ \*/

-- Type your code below:

practice=# select genre,avg(duration)as avg from movie full outer join genre on id=movie\_id group by genre order by avg desc;

genre | avg

-----------+----------------------

Action | 112.8828549262994569

Romance | 109.5342163355408389

Crime | 107.0516605166051661

Drama | 106.7745624270711785

Fantasy | 105.1403508771929825

Comedy | 102.6227197346600332

Adventure | 101.8714043993231810

Mystery | 101.8000000000000000

Thriller | 101.5761455525606469

Family | 100.9668874172185430

Others | 100.1600000000000000

Sci-Fi | 97.9413333333333333

Horror | 92.7243377483443709

(13 rows)

/\* Now you know, movies of genre 'Drama' (produced highest in number in 2019) has the average duration of 106.77 mins.

Lets find where the movies of genre 'thriller' on the basis of number of movies.\*/

-- Q9.What is the rank of the ‘thriller’ genre of movies among all the genres in terms of number of movies produced?

-- (Hint: Use the Rank function)

/\* Output format:

+---------------+-------------------+---------------------+

| genre | movie\_count | genre\_rank |

+---------------+-------------------+---------------------+

|drama | 2312 | 2 |

+---------------+-------------------+---------------------+\*/

-- Type your code below:

practice=# select genre,count(genre) as count,rank() over(order by count(genre)) rank\_number from movie full outer join genre on id=movie\_id group by genre ;

genre | count | rank\_number

-----------+-------+-------------

Others | 100 | 1

Family | 302 | 2

Fantasy | 342 | 3

Sci-Fi | 375 | 4

Mystery | 555 | 5

Adventure | 591 | 6

Crime | 813 | 7

Romance | 906 | 8

Horror | 1208 | 9

Action | 1289 | 10

Thriller | 1484 | 11

Comedy | 2412 | 12

Drama | 4285 | 13

(13 rows)

/\*Thriller movies is in top 3 among all genres in terms of number of movies

In the previous segment, you analysed the movies and genres tables.

In this segment, you will analyse the ratings table as well.

To start with lets get the min and max values of different columns in the table\*/

-- Segment 2:

-- Q10. Find the minimum and maximum values in each column of the ratings table except the movie\_id column?

/\* Output format:

+---------------+-------------------+---------------------+----------------------+-----------------+-----------------+

| min\_avg\_rating| max\_avg\_rating | min\_total\_votes | max\_total\_votes |min\_median\_rating|min\_median\_rating|

+---------------+-------------------+---------------------+----------------------+-----------------+-----------------+

| 0 | 5 | 177 | 2000 | 0 | 8 |

+---------------+-------------------+---------------------+----------------------+-----------------+-----------------+\*/

-- Type your code below:

practice=# select min(avg\_rating) as min\_avg\_rating,max(avg\_rating) as max\_avg\_rating,min(total\_votes) as min\_total\_votes,max(total\_votes) as max\_total\_votes,min(median\_rating) as min\_median\_rating,max(median\_rating) as max\_median\_rating from ratings ;

min\_avg\_rating | max\_avg\_rating | min\_total\_votes | max\_total\_votes | min\_median\_rating | max\_median\_rating

----------------+----------------+-----------------+-----------------+-------------------+-------------------

1.0 | 10.0 | 100 | 725138 | 1 | 10

(1 row)

/\* So, the minimum and maximum values in each column of the ratings table are in the expected range.

This implies there are no outliers in the table.

Now, let’s find out the top 10 movies based on average rating.\*/

-- Q11. Which are the top 10 movies based on average rating?

/\* Output format:

+---------------+-------------------+---------------------+

| title | avg\_rating | movie\_rank |

+---------------+-------------------+---------------------+

| Fan | 9.6 | 5 |

| . | . | . |

| . | . | . |

| . | . | . |

+---------------+-------------------+---------------------+\*/

-- Type your code below:

-- It's ok if RANK() or DENSE\_RANK() is used too

practice=# select title,avg\_rating, rank() over( order by avg\_rating desc)ranking from movie full outer join ratings on id=movie\_id limit 10;

title | avg\_rating | ranking

--------------------------------+------------+---------

Love in Kilnerry | 10.0 | 1

Kirket | 10.0 | 1

Gini Helida Kathe | 9.8 | 3

Runam | 9.7 | 4

Fan | 9.6 | 5

Android Kunjappan Version 5.25 | 9.6 | 5

Safe | 9.5 | 7

Yeh Suhaagraat Impossible | 9.5 | 7

The Brighton Miracle | 9.5 | 7

Ananthu V/S Nusrath | 9.4 | 10

(10 rows)

/\* Do you find you favourite movie FAN in the top 10 movies with an average rating of 9.6? If not, please check your code again!!

So, now that you know the top 10 movies, do you think character actors and filler actors can be from these movies?

Summarising the ratings table based on the movie counts by median rating can give an excellent insight.\*/

-- Q12. Summarise the ratings table based on the movie counts by median ratings.

/\* Output format:

+---------------+-------------------+

| median\_rating | movie\_count |

+-------------------+----------------

| 1 | 105 |

| . | . |

| . | . |

+---------------+-------------------+ \*/

-- Type your code below:

-- Order by is good to have

practice=# select median\_rating,count(median\_rating) as movie\_count from ratings group by median\_rating order by median\_rating asc;

median\_rating | movie\_count

---------------+-------------

1 | 94

2 | 119

3 | 283

4 | 479

5 | 985

6 | 1975

7 | 2257

8 | 1030

9 | 429

10 | 346

(10 rows)

/\* Movies with a median rating of 7 is highest in number.

Now, let's find out the production house with which RSVP Movies can partner for its next project.\*/

-- Q13. Which production house has produced the most number of hit movies (average rating > 8)??

/\* Output format:

+------------------+-------------------+---------------------+

|production\_company|movie\_count | prod\_company\_rank|

+------------------+-------------------+---------------------+

| The Archers | 1 | 1 |

+------------------+-------------------+---------------------+\*/

-- Type your code below:

practice=# select production\_company, rank() over( order by count(title) asc)prod\_company\_rank from movie group by production\_company limit 10;

production\_company | prod\_company\_rank

----------------------------------+-------------------

NewBeTV | 1

Ida Rose | 1

Little Rock Productions | 1

14 Reels Entertainment | 1

Blackhall Entertainment Ventures | 1

Beijing J.Q. Spring Pictures | 1

Erdbeermund Filmproduktion | 1

Pacific Electric Picture Company | 1

Do It Creative | 1

Ghoul Film | 1

(10 rows)

-- It's ok if RANK() or DENSE\_RANK() is used too

-- Answer can be Dream Warrior Pictures or National Theatre Live or both

-- Q14. How many movies released in each genre during March 2017 in the USA had more than 1,000 votes?

/\* Output format:

+---------------+-------------------+

| genre | movie\_count |

+-------------------+----------------

| thriller | 105 |

| . | . |

| . | . |

+---------------+-------------------+ \*/

-- Type your code below:

practice=# select m.title, count(title) from movie m full outer join genre g on m.id=g.movie\_id join ratings r on g.movie\_id=r.movie\_id where extract(month from date\_published)= 3 and r.total\_votes>1000 group by title limit 10;

select id , extract(hour from hours) as hr from user\_log,rank() over(order by hr desc);

select extract(hour from timestamp) as hour, count(id) as log from Weblog group by hour order by log desc;

title | count

----------------------+-------

1:54 | 1

1971: Beyond Borders | 1

4 latas | 2

Abduction | 3

Abnormal Attraction | 3

Acrimony | 3

A Few Less Men | 2

Ailecek Saskiniz | 1

A.I. Rising | 3

All Nighter | 2

(10 rows)

-- Lets try to analyse with a unique problem statement.

-- Q15. Find movies of each genre that start with the word ‘The’ and which have an average rating > 8?

/\* Output format:

+---------------+-------------------+---------------------+

| title | avg\_rating | genre |

+---------------+-------------------+---------------------+

| Theeran | 8.3 | Thriller |

| . | . | . |

| . | . | . |

| . | . | . |

+---------------+-------------------+---------------------+\*/

-- Type your code below:

practice=# select m.title, r.avg\_rating,g.genre from movie m full outer 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 r.avg\_rating > 8 limit 10;

title | avg\_rating | genre

--------------------------+------------+----------

The Blue Elephant 2 | 8.8 | Drama

The Blue Elephant 2 | 8.8 | Horror

The Blue Elephant 2 | 8.8 | Mystery

The Brighton Miracle | 9.5 | Drama

The Irishman | 8.7 | Crime

The Irishman | 8.7 | Drama

The Colour of Darkness | 9.1 | Drama

Theeran Adhigaaram Ondru | 8.3 | Action

Theeran Adhigaaram Ondru | 8.3 | Crime

Theeran Adhigaaram Ondru | 8.3 | Thriller

(10 rows)

-- You should also try your hand at median rating and check whether the ‘median rating’ column gives any significant insights.

-- Q16. Of the movies released between 1 April 2018 and 1 April 2019, how many were given a median rating of 8?

-- Type your code below:

practice=# select ratings.median\_rating,count(ratings.median\_rating) from movie full outer join ratings on movie.id=ratings.movie\_id where (date\_published between '2018-04-01' and '2019-04-01') and ratings.median\_rating =8 group by ratings.median\_rating limit 10;

median\_rating | count

---------------+-------

8 | 361

(1 row)

practice=# select title, date\_published,ratings.median\_rating from movie full outer join ratings on movie.id=ratings.movie\_id where (date\_published between '2018-04-01' and '2019-04-01') and ratings.median\_rating =8 limit 10;

title | date\_published | median\_rating

---------------------+----------------+---------------

Le roi de coeur | 2018-06-08 | 8

Ankur | 2019-01-02 | 8

The Evil Dead | 2018-10-31 | 8

Pestonjee | 2019-02-22 | 8

Aashirwad | 2019-03-09 | 8

Kaminnyy khrest | 2018-10-30 | 8

Ek Hasina Thi | 2019-01-02 | 8

Alita: Battle Angel | 2019-02-06 | 8

Dukun | 2018-04-05 | 8

Back Roads | 2018-12-07 | 8

(10 rows)

-- Once again, try to solve the problem given below.

-- Q17. Do German movies get more votes than Italian movies?

-- Hint: Here you have to find the total number of votes for both German and Italian movies.

-- Type your code below:

practice=# select movie.country, sum(ratings.total\_votes) from movie full outer join ratings on movie.id=ratings.movie\_id where movie.country ilike 'germany' or movie.country ilike 'italy' group by country;

country | sum

---------+--------

Germany | 106710

Italy | 77965

(2 rows)

-- Answer is Yes

/\* Now that you have analysed the movies, genres and ratings tables, let us now analyse another table, the names table.

Let’s begin by searching for null values in the tables.\*/

-- Segment 3:

-- Q18. Which columns in the names table have null values??

/\*Hint: You can find null values for individual columns or follow below output format

+---------------+-------------------+---------------------+----------------------+

| name\_nulls | height\_nulls |date\_of\_birth\_nulls |known\_for\_movies\_nulls|

+---------------+-------------------+---------------------+----------------------+

| 0 | 123 | 1234 | 12345 |

+---------------+-------------------+---------------------+----------------------+\*/

-- Type your code below:

practice=# select count(case when names is null then 'ok' end) as count\_names,

practice-# count(case when height is null then 'ok' end) as count\_height,

practice-# count(case when date\_of\_birth is null then 'ok' end) as date\_of\_birth\_nulls,

practice-# count(case when known\_for\_movies is null then 'ok' end) as count\_known\_for\_movies from names;

count\_names | count\_height | date\_of\_birth\_nulls | count\_known\_for\_movies

-------------+--------------+---------------------+------------------------

0 | 17335 | 13431 | 15226

(1 row)

/\* There are no Null value in the column 'name'.

The director is the most important person in a movie crew.

Let’s find out the top three directors in the top three genres who can be hired by RSVP Movies.\*/

-- Q19. Who are the top three directors in the top three genres whose movies have an average rating > 8?

-- (Hint: The top three genres would have the most number of movies with an average rating > 8.)

/\* Output format:

+---------------+-------------------+

| director\_name | movie\_count |

+---------------+-------------------|

|James Mangold | 4 |

| . | . |

| . | . |

+---------------+-------------------+ \*/

-- Type your code below:

practice=# select n.name as director\_name, count(m.title) as movie\_count from names n full outer join director\_mapping d on n.id=d.name\_id full outer join movie m on d.movie\_id=m.id full outer join ratings r on m.id=r.movie\_id full outer join genre g on m.id=g.movie\_id where r.avg\_rating > 8 group by director\_name order by movie\_count desc limit 15;

director\_name | movie\_count

-----------------------+-------------

| 244

Anthony Russo | 6

Joe Russo | 6

James Mangold | 5

Madhu C. Narayanan | 3

Milos Avramovic | 3

Khalid Rahman | 3

Soubin Shahir | 3

Adesh Prasad | 3

Kleber Mendonça Filho | 3

Anand Gandhi | 3

Juliano Dornelles | 3

Aaron K. Carter | 3

Venkat Ramji | 3

Jeral Clyde Jr. | 3

(15 rows)

practice=# select n.name as director\_name, count(m.title) as movie\_count from names n full outer join director\_mapping d on n.id=d.name\_id full outer join movie m on d.movie\_id=m.id full outer join ratings r on m.id=r.movie\_id full outer join genre g on m.id=g.movie\_id where r.avg\_rating > 8 group by director\_name order by movie\_count desc limit 15;

director\_name | movie\_count

-----------------------+-------------

| 244

Anthony Russo | 6

Joe Russo | 6

James Mangold | 5

Madhu C. Narayanan | 3

Milos Avramovic | 3

Khalid Rahman | 3

Soubin Shahir | 3

Adesh Prasad | 3

Kleber Mendonça Filho | 3

Anand Gandhi | 3

Juliano Dornelles | 3

Aaron K. Carter | 3

Venkat Ramji | 3

Jeral Clyde Jr. | 3

(15 rows)

/\* James Mangold can be hired as the director for RSVP's next project. Do you remeber his movies, 'Logan' and 'The Wolverine'.

Now, let’s find out the top two actors.\*/

-- Q20. Who are the top two actors whose movies have a median rating >= 8?

/\* Output format:

+---------------+-------------------+

| actor\_name | movie\_count |

+-------------------+----------------

|Christain Bale | 10 |

| . | . |

+---------------+-------------------+ \*/

-- Type your code below:

practice=# select n.name as actor, count(m.title)as movie\_count, rank() over(order by count(m.title) desc) ranking from names n full outer join role\_mapping rr on n.id=rr.name\_id full outer join movie m on rr.movie\_id=m.id full outer join ratings r on m.id=r.movie\_id where median\_rating >=8 group by n.name limit 15;

actor | movie\_count | ranking

-------------------+-------------+---------

| 954 | 1

Mammootty | 8 | 2

Mohanlal | 5 | 3

Adinia Wirasti | 4 | 4

Amit Sadh | 4 | 4

Dulquer Salmaan | 4 | 4

Tovino Thomas | 4 | 4

Amrinder Gill | 4 | 4

Dileesh Pothan | 4 | 4

Fahadh Faasil | 4 | 4

Pankaj Tripathi | 4 | 4

Johnny Yong Bosch | 4 | 4

Rajkummar Rao | 4 | 4

Manju Warrier | 4 | 4

Siddique | 4 | 4

(15 rows)

practice=# select n.name as actor, count(m.title)as movie\_count from names n full outer join role\_mapping rr on n.id=rr.name\_id full outer join movie m on rr.movie\_id=m.id full outer join ratings r on m.id=r.movie\_id where median\_rating >=8 group by n.name order by movie\_count desc limit 15;

actor | movie\_count

--------------------+-------------

| 954

Mammootty | 8

Mohanlal | 5

Rajkummar Rao | 4

Ayushmann Khurrana | 4

Siddique | 4

Fahadh Faasil | 4

Rashmika Mandanna | 4

Simi Chahal | 4

Amrinder Gill | 4

Joju George | 4

Pankaj Tripathi | 4

Adinia Wirasti | 4

Amit Sadh | 4

Manju Warrier | 4

(15 rows)

/\* Have you find your favourite actor 'Mohanlal' in the list. If no, please check your code again.

RSVP Movies plans to partner with other global production houses.

Let’s find out the top three production houses in the world.\*/

-- Q21. Which are the top three production houses based on the number of votes received by their movies?

/\* Output format:

+------------------+--------------------+---------------------+

|production\_company|vote\_count | prod\_comp\_rank|

+------------------+--------------------+---------------------+

| The Archers | 830 | 1 |

| . | . | . |

| . | . | . |

+-------------------+-------------------+---------------------+\*/

-- Type your code below:

practice=# select m.production\_company, sum(r.total\_votes)as total\_votes, rank() over(order by sum(r.total\_votes) desc) from movie m full outer join ratings r on m.id=r.movie\_id group by m.production\_company limit 3;

production\_company | total\_votes | rank

-----------------------+-------------+------

Marvel Studios | 2656967 | 1

Twentieth Century Fox | 2411163 | 2

Warner Bros. | 2396057 | 3

(3 rows)

/\*Yes Marvel Studios rules the movie world.

So, these are the top three production houses based on the number of votes received by the movies they have produced.

Since RSVP Movies is based out of Mumbai, India also wants to woo its local audience.

RSVP Movies also wants to hire a few Indian actors for its upcoming project to give a regional feel.

Let’s find who these actors could be.\*/

-- Q22. Rank actors with movies released in India based on their average ratings. Which actor is at the top of the list?

-- Note: The actor should have acted in at least five Indian movies.

-- (Hint: You should use the weighted average based on votes. If the ratings clash, then the total number of votes should act as the tie breaker.)

/\* Output format:

+---------------+-------------------+---------------------+----------------------+-----------------+

| actor\_name | total\_votes | movie\_count | actor\_avg\_rating |actor\_rank |

+---------------+-------------------+---------------------+----------------------+-----------------+

| Yogi Babu | 3455 | 11 | 8.42 | 1 |

| . | . | . | . | . |

| . | . | . | . | . |

| . | . | . | . | . |

+---------------+-------------------+---------------------+----------------------+-----------------+\*/

-- Type your code below:

practice=# select n.name as actor\_name,sum(r.total\_votes)as total\_votes, count(m.title)as movie\_count,avg(r.avg\_rating) as actor\_avg\_rating, rank() over(order by sum(r.total\_votes)desc)ranking from names n full outer join role\_mapping rr on n.id=rr.name\_id full outer join movie m on rr.movie\_id=m.id full outer join ratings r on m.id=r.movie\_id where m.country='India' group by actor\_name having count(m.title)>= 5 limit 10;

actor\_name | total\_votes | movie\_count | actor\_avg\_rating | ranking

------------------+-------------+-------------+--------------------+---------

| 1290678 | 549 | 6.5180327868852459 | 1

Rajkummar Rao | 42560 | 6 | 6.8666666666666667 | 2

Pankaj Tripathi | 40728 | 5 | 7.4000000000000000 | 3

Vijay Sethupathi | 23114 | 5 | 6.1000000000000000 | 4

Taapsee Pannu | 18895 | 5 | 6.9800000000000000 | 5

Dulquer Salmaan | 17666 | 5 | 6.7000000000000000 | 6

Mohanlal | 17244 | 6 | 6.3500000000000000 | 7

Fahadh Faasil | 13557 | 5 | 7.7400000000000000 | 8

Amit Sadh | 13355 | 5 | 7.2200000000000000 | 9

Mammootty | 12613 | 8 | 6.7125000000000000 | 10

(10 rows)

-- Top actor is Vijay Sethupathi

-- Q23.Find out the top five actresses in Hindi movies released in India based on their average ratings?

-- Note: The actresses should have acted in at least three Indian movies.

-- (Hint: You should use the weighted average based on votes. If the ratings clash, then the total number of votes should act as the tie breaker.)

/\* Output format:

+---------------+-------------------+---------------------+----------------------+-----------------+

| actress\_name | total\_votes | movie\_count | actress\_avg\_rating |actress\_rank |

+---------------+-------------------+---------------------+----------------------+-----------------+

| Tabu | 3455 | 11 | 8.42 | 1 |

| . | . | . | . | . |

| . | . | . | . | . |

| . | . | . | . | . |

+---------------+-------------------+---------------------+----------------------+-----------------+\*/

-- Type your code below:

practice=# select n.name as atress\_name, sum(r.total\_votes)as total\_votes, count(m.title) as movie\_count,avg(r.avg\_rating)as actoress\_avg\_rating, rank() over(order by sum(r.total\_votes) desc) from names n full outer join role\_mapping rr on n.id=rr.name\_id full outer join movie m on rr.movie\_id=m.id full outer join ratings r on m.id=r.movie\_id where m.country='India' and rr.category='actress' group by atress\_name having count(m.title)>=3 limit 10;

atress\_name | total\_votes | movie\_count | actoress\_avg\_rating | rank

----------------------+-------------+-------------+---------------------+------

Kajal Aggarwal | 35358 | 3 | 7.1000000000000000 | 1

Shraddha Kapoor | 26779 | 3 | 6.3000000000000000 | 2

Shraddha Srinath | 26498 | 4 | 7.5000000000000000 | 3

Samantha Ruth Prabhu | 25143 | 3 | 6.7333333333333333 | 4

Kriti Sanon | 21967 | 3 | 5.9666666666666667 | 5

Taapsee Pannu | 18895 | 5 | 6.9800000000000000 | 6

Rakul Preet Singh | 12327 | 3 | 5.4000000000000000 | 7

Manju Warrier | 11276 | 5 | 6.6800000000000000 | 8

Rashmika Mandanna | 10975 | 4 | 7.0500000000000000 | 9

Vidya Balan | 10794 | 3 | 6.9666666666666667 | 10

(10 rows)

/\* Taapsee Pannu tops with average rating 7.74.

Now let us divide all the thriller movies in the following categories and find out their numbers.\*/

/\* Q24. Select thriller movies as per avg rating and classify them in the following category:

Rating > 8: Superhit movies

Rating between 7 and 8: Hit movies

Rating between 5 and 7: One-time-watch movies

Rating < 5: Flop movies

--------------------------------------------------------------------------------------------\*/

-- Type your code below:

practice=# select m.title, r.avg\_rating, case when r.avg\_rating > 8 then 'SUPERDUPER' when r.avg\_rating between 7 and 8 then 'HIT' when r.avg\_rating between 5 and 7 then '1 t w' else 'Tapasee Pannuu (flop)' end as descriptions from movie m full outer join ratings r on m.id=r.movie\_id full outer join genre g on r.movie\_id=g.movie\_id where g.genre ilike 'thriller' limit 10;

title | avg\_rating | descriptions

----------------+------------+-----------------------

Der müde Tod | 7.7 | HIT

Fahrenheit 451 | 4.9 | Tapasee Pannuu (flop)

Pet Sematary | 5.8 | 1 t w

Dukun | 6.9 | 1 t w

Back Roads | 7.0 | HIT

Countdown | 5.4 | 1 t w

Staged Killer | 3.3 | Tapasee Pannuu (flop)

Vellaipookal | 7.3 | HIT

Uriyadi 2 | 7.3 | HIT

Incitement | 7.5 | HIT

(10 rows)

/\* Until now, you have analysed various tables of the data set.

Now, you will perform some tasks that will give you a broader understanding of the data in this segment.\*/

-- Segment 4:

-- Q25. What is the genre-wise running total and moving average of the average movie duration?

-- (Note: You need to show the output table in the question.)

/\* Output format:

+---------------+-------------------+---------------------+----------------------+

| genre | avg\_duration |running\_total\_duration|moving\_avg\_duration |

+---------------+-------------------+---------------------+----------------------+

| comdy | 145 | 106.2 | 128.42 |

| . | . | . | . |

| . | . | . | . |

| . | . | . | . |

+---------------+-------------------+---------------------+----------------------+\*/practice=# select g.genre, round(avg(m.duration)) as avg\_duration, sum(m.duration) as running\_total\_duration from genre g full outer join movie m on m.id= g.movie\_id group by g.genre;

genre | avg\_duration | running\_total\_duration

-----------+--------------+------------------------

Crime | 107 | 87033

Romance | 110 | 99238

Horror | 93 | 112011

Drama | 107 | 457529

Action | 113 | 145506

Mystery | 102 | 56499

Fantasy | 105 | 35958

Thriller | 102 | 150739

Sci-Fi | 98 | 36728

Comedy | 103 | 247526

Adventure | 102 | 60206

Others | 100 | 10016

Family | 101 | 30492

(13 rows)

-- Type your code below:

-- Round is good to have and not a must have; Same thing applies to sorting

-- Let us find top 5 movies of each year with top 3 genres.

-- Q26. Which are the five highest-grossing movies of each year that belong to the top three genres?

-- (Note: The top 3 genres would have the most number of movies.)

/\* Output format:

+---------------+-------------------+---------------------+----------------------+-----------------+

| genre | year | movie\_name |worldwide\_gross\_income|movie\_rank |

+---------------+-------------------+---------------------+----------------------+-----------------+

| comedy | 2017 | indian | $103244842 | 1 |

| . | . | . | . | . |

| . | . | . | . | . |

| . | . | . | . | . |

+---------------+-------------------+---------------------+----------------------+-----------------+\*/

-- Type your code below:

-- Top 3 Genres based on most number of movies

-- Finally, let’s find out the names of the top two production houses that have produced the highest number of hits among multilingual movies.

-- Q27. Which are the top two production houses that have produced the highest number of hits (median rating >= 8) among multilingual movies?

/\* Output format:

+-------------------+-------------------+---------------------+

|production\_company |movie\_count | prod\_comp\_rank|

+-------------------+-------------------+---------------------+

| The Archers | 830 | 1 |

| . | . | . |

| . | . | . |

+-------------------+-------------------+---------------------+\*/

-- Type your code below:

practice=# select m.production\_company, count(m.title)as movie\_count, rank() over(order by count(m.title) desc)as prod\_comp\_rank from movie m group by production\_company limit 10;

production\_company | movie\_count | prod\_comp\_rank

-----------------------+-------------+----------------

| 528 | 1

Netflix | 20 | 2

BRON Studios | 19 | 3

Warner Bros. | 19 | 3

Columbia Pictures | 17 | 5

Cartel Pictures | 16 | 6

Constantin Film | 16 | 6

Universal Pictures | 15 | 8

Blumhouse Productions | 15 | 8

BKM Film | 14 | 10

(10 rows)

-- Multilingual is the important piece in the above question. It was created using POSITION(',' IN languages)>0 logic

-- If there is a comma, that means the movie is of more than one language

-- Q28. Who are the top 3 actresses based on number of Super Hit movies (average rating >8) in drama genre?

/\* Output format:

+---------------+-------------------+---------------------+----------------------+-----------------+

| actress\_name | total\_votes | movie\_count |actress\_avg\_rating |actress\_rank |

+---------------+-------------------+---------------------+----------------------+-----------------+

| Laura Dern | 1016 | 1 | 9.60 | 1 |

| . | . | . | . | . |

| . | . | . | . | . |

+---------------+-------------------+---------------------+----------------------+-----------------+\*/

-- Type your code below:

practice=# select n.name, sum(r.total\_votes), count(m.title), avg(r.avg\_rating), rank() over( order by sum(r.total\_votes) desc) as ranking from names n full outer join role\_mapping rr on n.id=rr.name\_id full outer join movie m on rr.movie\_id=m.id full outer join ratings r on m.id=r.movie\_id where r.avg\_rating > 8 group by n.name limit 4;

name | sum | count | avg | ranking

-------------------+---------+-------+--------------------+---------

| 1958039 | 141 | 8.5780141843971631 | 1

Mark Ruffalo | 1327930 | 2 | 8.5000000000000000 | 2

Chris Evans | 1327930 | 2 | 8.5000000000000000 | 2

Robert Downey Jr. | 1327930 | 2 | 8.5000000000000000 | 2

(4 rows)

/\* Q29. Get the following details for top 9 directors (based on number of movies)

Director id

Name

Number of movies

Average inter movie duration in days

Average movie ratings

Total votes

Min rating

Max rating

total movie durations

Format:

+---------------+-------------------+---------------------+----------------------+--------------+--------------+------------+------------+----------------+

| director\_id | director\_name | number\_of\_movies | avg\_inter\_movie\_days | avg\_rating | total\_votes | min\_rating | max\_rating | total\_duration |

+---------------+-------------------+---------------------+----------------------+--------------+--------------+------------+------------+----------------+

|nm1777967 | A.L. Vijay | 5 | 177 | 5.65 | 1754 | 3.7 | 6.9 | 613 |

| . | . | . | . | . | . | . | . | . |

| . | . | . | . | . | . | . | . | . |

| . | . | . | . | . | . | . | . | . |

| . | . | . | . | . | . | . | . | . |

| . | . | . | . | . | . | . | . | . |

| . | . | . | . | . | . | . | . | . |

| . | . | . | . | . | . | . | . | . |

| . | . | . | . | . | . | . | . | . |

+---------------+-------------------+---------------------+----------------------+--------------+--------------+------------+------------+----------------+

--------------------------------------------------------------------------------------------\*/

-- Type you code below:

practice=# select d.name\_id as director\_id, n.name as director\_name, count(m.title), avg(m.duration) as avg\_inter\_movie\_days, avg(r.avg\_rating) as avg\_rating, sum(r.total\_votes)as total\_votes, min(r.avg\_rating) as min\_rating, max(r.avg\_rating) as max\_rating, sum(m.duration)as total\_duration from director\_mapping d full outer join names n on d.name\_id=n.id full outer join movie m on m.id=n.id full outer join ratings r on m.id=r.movie\_id group by director\_id, director\_name order by director\_name desc limit 3;