Introduction To Data Science Assignment – 1

Kirti Katiyar

Question -1

Write a query to list the distinct theaters name, movie name, movie length playing the shortest in length pg-13 movies.

```
kk23m@bl2:~                  ×    +   ∨

mysql> SELECT DISTINCT
    ->     theaters.Name,
    ->     movies.Title,
    ->     movies.Length
    -> FROM movies
    -> JOIN movies2theaters ON movies.MovieCode = movies2theaters.MovieCode
    -> JOIN theaters ON movies2theaters.TheaterCode = theaters.TheaterCode
    -> WHERE movies.Rating = 'PG-13'
    -> ORDER BY movies.Length ASC
    -> LIMIT 1;
+--------------------------+--------------+--------+
| Name                     | Title        | Length |
+--------------------------+--------------+--------+
| Landmark Neptune Theatre | Best in Show |     90 |
+--------------------------+--------------+--------+
1 row in set (0.00 sec)
```

Question – 2

Write a sql query of all the movies played in Downtown district showing distinct district name, theater name, movie name, movie rating and movie length where the movie rating is NR and the query is sorted by length.

```
mysql> SELECT DISTINCT
    ->    districts.Name,
    ->    theaters.Name,
    ->    movies.Title,
    ->    movies.Rating,
    ->    movies.Length
    -> FROM movies
    -> JOIN movies2theaters ON movies.MovieCode = movies2theaters.MovieCode
de = theaters.TheaterCode
JOIN theaters2di    -> JOIN theaters ON movies2theaters.TheaterCode = theaters.TheaterCode
    -> JOIN theaters2districts ON theaters.TheaterCode = theaters2districts.TheaterCode
    -> JOIN districts ON theaters2districts.DistrictCode = districts.DistrictCode
Name = 'Downtown'
ORDER BY movies.Length ASC;     -> WHERE movies.Rating = 'NR'
    -> AND districts.Name = 'Downtown'
    -> ORDER BY movies.Length ASC;
+----------+--------------------------+---------------------+--------+--------+
| Name     | Name                     | Title               | Rating | Length |
+----------+--------------------------+---------------------+--------+--------+
| DOWNTOWN | Landmark Varsity Theatre | Dark Days           | NR     |     84 |
| DOWNTOWN | Cineplex Odeon Uptown    | Barenaked in America | NR     |     89 |
| DOWNTOWN | Cineplex Odeon Uptown    | Human Resources     | NR     |    100 |
| DOWNTOWN | Cineplex Odeon Uptown    | One                 | NR     |    102 |
+----------+--------------------------+---------------------+--------+--------+
4 rows in set (0.00 sec)
```

Question – 3

Write a query that shows the district name, movie rating and movie count where we want a distinct count of all movies played for each rating in Downtown district.

```
mysql> SELECT
    ->    districts.Name,
    ->    movies.Rating,
    ->    COUNT(DISTINCT movies.Title) AS MovieCount
    -> FROM movies
    -> JOIN movies2theaters ON movies.MovieCode = movies2theaters.MovieCode
    -> JOIN theaters ON movies2theaters.TheaterCode = theaters.TheaterCode
    -> JOIN theaters2districts ON theaters.TheaterCode = theaters2districts.TheaterCode
    -> JOIN districts ON theaters2districts.DistrictCode = districts.DistrictCode
    -> WHERE districts.Name = 'Downtown'
    -> GROUP BY districts.Name, movies.Rating
    -> ORDER BY movies.Rating ASC;
+----------+--------+------------+
| Name     | Rating | MovieCount |
+----------+--------+------------+
| DOWNTOWN | G      |          1 |
| DOWNTOWN | NR     |          4 |
| DOWNTOWN | PG     |          2 |
| DOWNTOWN | PG-13  |          4 |
| DOWNTOWN | R      |         18 |
+----------+--------+------------+
5 rows in set (0.01 sec)
```

Question – 4

Write a query that list theater name, movie name and movie length that shows movies played with runtime less than 90 minutes in Cineplex Odeon Uptown

```
mysql> SELECT
    ->    theaters.Name,
    ->    movies.Title,
    ->    movies.Length
    -> FROM movies
    -> JOIN movies2theaters ON movies.MovieCode = movies2theaters.MovieCode
    -> JOIN theaters ON movies2theaters.TheaterCode = theaters.TheaterCode
    -> WHERE theaters.Name = 'Cineplex Odeon Uptown'
    -> AND movies.Length < 90
    -> ORDER BY movies.Length ASC;
+-----------------------+----------------------+--------+
| Name                  | Title                | Length |
+-----------------------+----------------------+--------+
| Cineplex Odeon Uptown | Chicken Run          |     85 |
| Cineplex Odeon Uptown | Barenaked in America |     89 |
| Cineplex Odeon Uptown | Barenaked in America |     89 |
| Cineplex Odeon Uptown | Barenaked in America |     89 |
+-----------------------+----------------------+--------+
4 rows in set (0.01 sec)
```

Question – 5

Write a query that lists district name, theater name, movie name, movie rating and how many times the movie is played in each theater in every district. The rating of the movie is NR and it should be the shortest NR movie.

```
kk23m@bl2:~                    ×    +   ∨

mysql> SELECT
    ->    districts.Name,
    ->    theaters.Name,
    ->    movies.Title,
    ->    movies.Rating,
    ->    COUNT(*) AS Count
    -> FROM movies
    -> JOIN movies2theaters ON movies.MovieCode = movies2theaters.MovieCode
    -> JOIN theaters ON movies2theaters.TheaterCode = theaters.TheaterCode
    -> JOIN theaters2districts ON theaters.TheaterCode = theaters2districts.TheaterCode
    -> JOIN districts ON theaters2districts.DistrictCode = districts.DistrictCode
    -> WHERE movies.Rating = 'NR'
    -> GROUP BY districts.Name, theaters.Name, movies.Title, movies.Rating
    -> ORDER BY movies.Length ASC
    -> LIMIT 1;
+-----------------+--------------------------+------------+--------+--------+
| Name            | Name                     | Title      | Rating | Count  |
+-----------------+--------------------------+------------+--------+--------+
| THE U DISTRICT  | Landmark Varsity Theatre | Dark Days  | NR     |     1  |
+-----------------+--------------------------+------------+--------+--------+
1 row in set (0.01 sec)
```

Question – 6

Write a sql query that lists theater name, movie name, movie length and how many times the movie is played. The theater should play the longest movie in length and the shortest movie in length, (if there are multiple movies having shortest or longest time it should pick one, Also the theater should play the shortest as well as the longest movie in length).

```
kk23m@bl2:~                    ×    +    ∨

mysql> SELECT
    ->    theaters.Name,
    ->    movies.Title,
    ->    movies.Length,
    ->    COUNT(*) AS MovieCount
    -> FROM
    ->    theaters
    -> INNER JOIN
    ->    movies2theaters ON theaters.TheaterCode = movies2theaters.TheaterCode
    -> INNER JOIN
    ->    movies ON movies2theaters.MovieCode = movies.MovieCode
    -> WHERE
    ->    movies.Length = (SELECT MAX(Length) FROM movies)
    ->    OR movies.Length = (SELECT MIN(Length) FROM movies)
    -> GROUP BY
    ->    theaters.Name,
    ->    movies.Title,
    ->    movies.Length
    -> HAVING
    ->    movies.Length IN (
    ->      SELECT MAX(Length) FROM movies UNION ALL SELECT MIN(Length) FROM movies
    ->    )
    -> LIMIT 2;
+--------------------------+----------------+--------+------------+
| Name                     | Title          | Length | MovieCount |
+--------------------------+----------------+--------+------------+
| Cineplex Odeon Meridian  | Get Carter     |   147  |         10 |
| Cineplex Odeon Meridian  | The Ladies Man |    84  |         10 |
+--------------------------+----------------+--------+------------+
2 rows in set (0.00 sec)
```

Question – 7

Write a SQL query that lists only the phone number and district name of General Cinema Cinerama.

```
kk23m@bl2:~                    ×    +    ∨

mysql> SELECT
    ->    theaters.Phone,
    ->    districts.Name
    -> FROM theaters
    -> JOIN theaters2districts ON theaters.TheaterCode = theaters2districts.TheaterCode
    -> JOIN districts ON theaters2districts.DistrictCode = districts.DistrictCode
    -> WHERE theaters.Name = 'General Cinema Cinerama';
+----------------+---------------+
| Phone          | Name          |
+----------------+---------------+
| (206) 441-3080 | CAPITOL HILL  |
| (206) 441-3080 | DOWNTOWN      |
+----------------+---------------+
2 rows in set (0.00 sec)
```
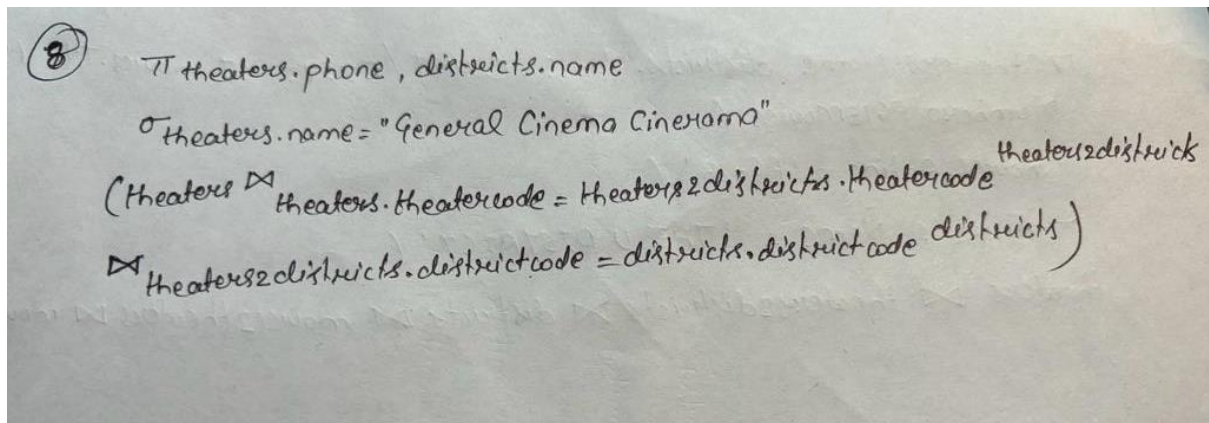
Question – 8

Write a Relational Algebra Expression for Question 7

π theaters.phone, districts.name

σ theaters.name = "General Cinema Cinerama"

(theaters $\bowtie$ theaters.theatercode = theaters2districts.theatercode theaters2districts

$\bowtie$ theaters2districts.districtcode = districts.districtcode districts)

Question – 9

Write a SQL query that lists theater name and movie name where each movie is played in both theaters (Landmark Metro Cinemas and General Cinema Pacific Place 11). The movies should be common in both theaters, it cannot be played only in one of them.

```
mysql> SELECT DISTINCT
    ->    t1.Theater_Name, t2.Theater_Name, t1.Movie_Name
    -> FROM
    ->    (SELECT t.Name AS Theater_Name, m.Title AS Movie_Name
    ->    FROM movies2theaters mt
    ->    JOIN theaters t ON mt.TheaterCode = t.TheaterCode
    ->    JOIN movies m ON mt.MovieCode = m.MovieCode
    ->    WHERE t.Name = 'Landmark Metro Cinemas') AS t1
    -> JOIN
    ->    (SELECT t.Name AS Theater_Name, m.Title AS Movie_Name
    ->    FROM movies2theaters mt
    ->    JOIN theaters t ON mt.TheaterCode = t.TheaterCode
    ->    JOIN movies m ON mt.MovieCode = m.MovieCode
    ->    WHERE t.Name = 'General Cinema Pacific Place 11') AS t2
    -> ON
    ->    t1.Movie_Name = t2.Movie_Name;
+------------------------+----------------------------------+---------------------+
| Theater_Name           | Theater_Name                     | Movie_Name          |
+------------------------+----------------------------------+---------------------+
| Landmark Metro Cinemas | General Cinema Pacific Place 11   | Girlfight           |
| Landmark Metro Cinemas | General Cinema Pacific Place 11   | Remember the Titans |
| Landmark Metro Cinemas | General Cinema Pacific Place 11   | The Contender       |
+------------------------+----------------------------------+---------------------+
3 rows in set (0.01 sec)
```

Answer – 10

Write a Relational Algebra Expression for Question 9.

(10) 8

$\pi_{t1.theater\_name, t2.theater\_name, t1.movie\_name}$

$(\rho_{t1}$

$\pi_{t.name \rightarrow theater\_name, m.title \rightarrow movie\_name}$

$\sigma_{t.name = "Landmark \ metro \ Cinemas"}$

$(\rho_{mt} \ movies2theaters \bowtie_{mt.theatercode = t.theatercode}$

$\rho_t \ theaters \bowtie_{mt.moviecode = m.moviecode}$

$\rho_{t2} \rho_m \ movies) \bowtie_{t1.movie\_name = t2.movie\_name}$

$\pi_{t.name \rightarrow theater\_name, m.title \rightarrow movie\_name}$

$\sigma_{t.name = "General \ Cinema \ Pacific \ Place \ 11"}$

$(\rho_{mt} \ movies2theaters \bowtie_{mt.theatercode = t.theatercode}$

$\rho_t \ theaters \bowtie_{mt.moviecode = m.moviecode}$
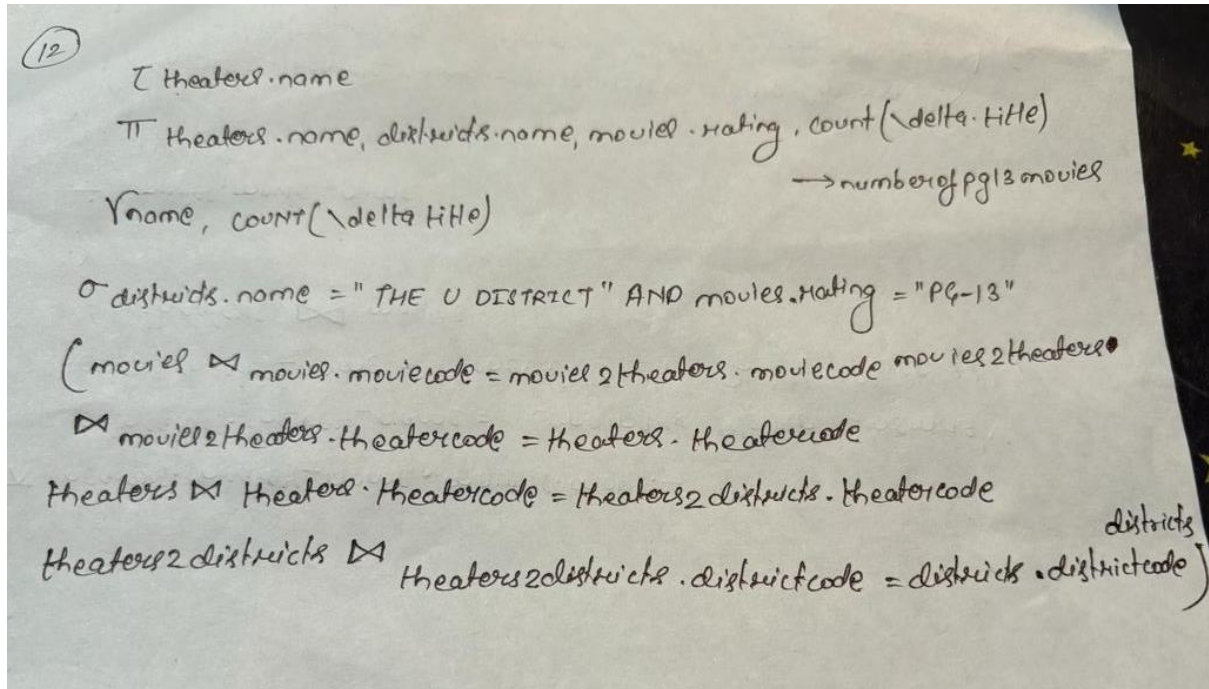
$\rho_m \ movies)))$

Question – 11

Write an SQL Query that Lists the theaters which are playing movies with a PG-13 rating in the district named 'THE U DISTRICT'. Along with the theater names, fetch the number of distinct PG-13 movies each of those theaters is playing.

```
kk23m@bl2:~                    ×    +    ⌄

mysql> SELECT
    ->    theaters.Name,
    ->    districts.Name,
    ->    movies.Rating,
    ->    COUNT(DISTINCT movies.Title) AS NumberOfPG13Movies
    -> FROM
    ->    movies
    -> JOIN
    ->    movies2theaters ON movies.MovieCode = movies2theaters.MovieCode
    -> JOIN
    ->    theaters ON movies2theaters.TheaterCode = theaters.TheaterCode
    -> JOIN
    ->    theaters2districts ON theaters.TheaterCode = theaters2districts.TheaterCode
    -> JOIN
    ->    districts ON theaters2districts.DistrictCode = districts.DistrictCode
    -> WHERE
    ->    districts.Name = 'THE U DISTRICT'
    -> AND
    ->    movies.Rating = 'PG-13'
    -> GROUP BY
    ->    theaters.Name
    -> ORDER BY
    ->    theaters.Name ASC;
+--------------------------------+----------------+--------+--------------------+
| Name                           | Name           | Rating | NumberOfPG13Movies |
+--------------------------------+----------------+--------+--------------------+
| Cineplex Odeon Meridian        | THE U DISTRICT | PG-13  |                  2 |
| General Cinema Pacific Place 11 | THE U DISTRICT | PG-13  |                  1 |
| Landmark Broadway Market Cinemas | THE U DISTRICT | PG-13  |                1 |
| Landmark Guild 45th            | THE U DISTRICT | PG-13  |                  1 |
| Landmark Metro Cinemas         | THE U DISTRICT | PG-13  |                  1 |
| Landmark Neptune Theatre       | THE U DISTRICT | PG-13  |                  1 |
```
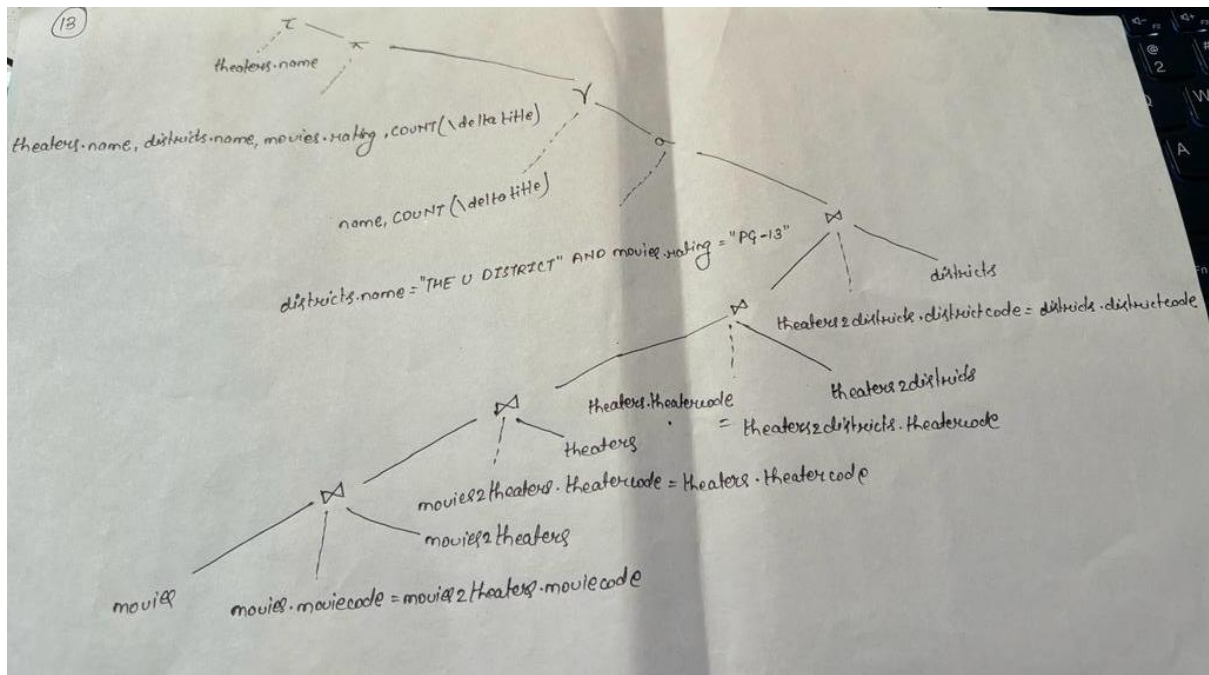
Question – 12

Write a Relational Algebra Expression for Question 11.

(12)

τ theaters.name

π theaters.name, districts.name, movies.rating, count(\delta.title)

$\longrightarrow$ number of pg13 movies

γ name, count(\delta title)

σ districts.name = "THE U DISTRICT" AND movies.rating = "PG-13"

(movies $\bowtie$ movies.moviecode = movies2theaters.moviecode movies2theaters

$\bowtie$ movies2theaters.theatercode = theaters.theatercode

theaters $\bowtie$ theaters.theatercode = theaters2districts.theatercode

theaters2districts $\bowtie$ theaters2districts.districtcode = districts.districtcode)

districts

Question – 13

Design a Query plan (tree) for Question 11.

(18)

τ theaters.name

π theaters.name, districts.name, movies.rating, COUNT(\delta title)

γ name, COUNT(\delta title)

σ districts.name = "THE U DISTRICT" AND movies.rating = "PG-13"

$\bowtie$ theaters2district.districtcode = districts.districtcode

districts

$\bowtie$ theaters2districts.theatercode = theaters.theatercode

theaters2districts

theaters.theatercode

theaters

$\bowtie$ movies2theaters.theatercode = theaters.theatercode

movies2theaters

$\bowtie$ movies.moviecode = movies2theaters.moviecode

movie

Question – 14

Write a SQL Query which tells how many movies each theater has played and also list the theaters that have played more movies than the average number of movies played across all theaters.

```
kk23m@bl2:~                    ×     +    ∨

mysql> SELECT
    ->    COUNT(mt.MovieCode) AS Count_of_movies,
    ->    t.Name AS Theater_Name
    -> FROM
    ->    theaters t
    -> JOIN
    ->    movies2theaters mt ON mt.TheaterCode = t.TheaterCode
    -> JOIN
    ->    movies m ON m.MovieCode = mt.MovieCode
    -> GROUP BY
    ->    t.Name
    -> HAVING
    ->    COUNT(mt.MovieCode) > (
    ->       SELECT AVG(MovieCount)
    ->       FROM (
    ->          SELECT COUNT(mt.MovieCode) AS MovieCount
    ->          FROM movies2theaters mt
    ->          GROUP BY TheaterCode
    ->       ) AS new_name
    ->    );
+-----------------+-----------------------------------+
| Count_of_movies | Theater_Name                      |
+-----------------+-----------------------------------+
|              73 | Cineplex Odeon Meridian           |
|              41 | General Cinema Pacific Place 11    |
|              34 | Landmark Metro Cinemas            |
+-----------------+-----------------------------------+
3 rows in set (0.01 sec)
```

Question – 15

Write a SQL Query which List the districts where the average movie length of the movies played is greater than the overall average movie length.

```
mysql> SELECT d. Name AS District_Name
    -> FROM districts d
    -> JOIN theaters2districts td ON d.DistrictCode=td. DistrictCode
    -> JOIN theaters t ON t. TheaterCode= td. TheaterCode
    -> JOIN movies2theaters mt ON t. TheaterCode= mt.TheaterCode
    -> JOIN movies m ON m.MovieCode= mt.MovieCode
    -> GROUP BY d.Name
    -> HAVING AVG(m.Length)> (SELECT AVG(Length) FROM movies);
+-----------------+
| District_Name   |
+-----------------+
| CAPITOL HILL    |
| DOWNTOWN        |
| THE U DISTRICT  |
+-----------------+
3 rows in set (0.00 sec)
```