# Common Twitter Queries

**Q1.** What user posted this tweet?

**Query**

```
SELECT
   u.name, t.twitter_text
FROM
   twitter_user AS u
      INNER JOIN
   twitter_tweets AS t ON u.twitter_handle = t.twitter_handle
WHERE
   t.tweet_id = 1590790945641230337
```

**Relational Algebra**

```
π u.name, t.twitter_text
 σ t.tweet_id = 1590790945641230337
  (ρ u twitter_user ⋈ u.twitter_handle = t.twitter_handle
   ρ t twitter_tweets)
```

**Q2.** When did the user post this tweet?

**Query**

```
SELECT
   u.name, t.twitter_text, t.created_at
FROM
   twitter_user as u
      INNER JOIN
   twitter_tweets as t ON u.twitter_handle = t.twitter_handle
WHERE
   t.tweet_id = 1590790945641230337
```

**Relational Algebra**

```
π u.name, t.twitter_text, t.created_at
 σ t.tweet_id = 1590790945641230337
  (ρ u twitter_user ⋈ u.twitter_handle = t.twitter_handle
   ρ t twitter_tweets)
```

**Q3.** What tweets have this user posted in the past 24 hours?

**Query**

```
SELECT
   u.name, t.twitter_text, t.created_at
FROM
   twitter_user AS u
      INNER JOIN
   twitter_tweets AS t ON u.twitter_handle = t.twitter_handle
WHERE
   u.twitter_handle = 'gaugetherange' and t.created_at > "2022-11-11 19:00:00.0000" and
t.created_at < "2022-11-12 18:59:59.0000"
```

**Relational Algebra**

$$\pi_{u.name,\ t.twitter\_text,\ t.created\_at}$$
$$\sigma_{u.twitter\_handle = \text{"gaugetherange"}\ AND\ t.created\_at > \text{"2022-11-11 19:00:00.0000"}\ AND\ t.created\_at < \text{"2022-11-12 18:59:59.0000"}}$$
$$(\rho_{u}\, twitter\_user \bowtie_{u.twitter\_handle = t.twitter\_handle}$$
$$\rho_{t}\, twitter\_tweets)$$

**Q4.** How many tweets have this user posted in the past 24 hours?

**Query**

```
SELECT
   u.name, count(t.twitter_text), t.created_at
FROM
   twitter_user AS u
      INNER JOIN
   twitter_tweets AS t ON u.twitter_handle = t.twitter_handle
WHERE
   u.twitter_handle = 'gaugetherange' and t.created_at > "2022-11-11 19:00:00.0000" and
t.created_at < "2022-11-12 18:59:59.0000"
```

**Relational Algebra**

$$\pi_{u.name,\ COUNT\ (twitter\_text),\ t.created\_at}$$
$$\gamma_{COUNT\ (twitter\_text)}$$
$$\sigma_{u.twitter\_handle = \text{"gaugetherange"}\ AND\ t.created\_at > \text{"2022-11-11 19:00:00.0000"}\ AND\ t.created\_at < \text{"2022-11-12 18:59:59.0000"}}$$
$$(\rho_{u}\, twitter\_user \bowtie_{u.twitter\_handle = t.twitter\_handle}$$
$$\rho_{t}\, twitter\_tweets)$$

**Q5.** When did this user join Twitter?

**Query**

```
SELECT
   u.twitter_handle, u.created_at
FROM
   twitter_user AS u
WHERE
   u.twitter_handle = 'gaugetherange'
```

**Relational Algebra**

$$\pi_{u.\,twitter\_handle,\ u.\,created\_at}$$
$$\sigma_{u.\,twitter\_handle = "gaugetherange"}$$
$$\rho_{u}\,twitter\_user$$

**Q6.** What keywords/ hashtags are popular?

**Query**

```
SELECT
   t.tag, COUNT(t.tag)
FROM
   tweet_tags AS t
GROUP BY t.tag
ORDER BY COUNT(t.tag) DESC
LIMIT 10
```

**Relational Algebra**

$$\tau_{COUNT\ (tag)\ \downarrow}$$
$$\gamma_{tag,\ COUNT\ (tag)}$$
$$\rho_{t}\,tweet\_tags$$

**Q7.** What tweets are popular?

**Query**

```
SELECT
   t.twitter_text, t.retweet
FROM
   twitter_tweets AS t
ORDER BY t.retweet DESC
LIMIT 10;
```

**Relational Algebra**

$$\tau_{t.\,retweet\downarrow}$$
$$\pi_{t.\,twitter\_text,\,t.\,retweet}$$
$$\rho_{t}\,twitter\_tweets$$

# Manashree's Queries

**Q1.** What is the average rating of comedy genre movies?

**Query**

```
SELECT AVG(m.rating) AS Average_Rating FROM Movie m
JOIN Movie_Genre mg ON m.movie_id=mg.movie_id
JOIN Genre g ON g.genre_id = mg.genre_id and g.genre_id=5;
```

**Relational Algebra**

$\pi_{\text{AVG (rating)} \rightarrow \text{average\_rating}}$

$\gamma_{\text{AVG (rating)}}$

$(\rho_{m}\,\text{movie} \bowtie_{\text{m . movie\_id = mg . movie\_id}}$

$\rho_{mg}\,\text{movie\_genre} \bowtie_{\text{g . genre\_id = mg . genre\_id AND g . genre\_id = 5}}$

$\rho_{g}\,\text{genre})$

**Q2.** What are the top 10 hit movies by Ajay Devgn?

**Query**

```
SELECT m.name, m.rating, m.movie_id from movie m
INNER JOIN movie_stars ma on ma.movie_id = m.movie_id and ma.star_id=(SELECT s.star_id
from stars s where s.name='Ajay Devgn')
ORDER BY m.rating DESC
Limit 10;
```

**Relational Algebra**

$\tau_{\text{m . rating} \downarrow}$

$\pi_{\text{m . name, m . rating, m . movie\_id}}$

$(\rho_{m}\,\text{movie} \bowtie_{\text{ma . movie\_id = m . movie\_id}}$

$\rho_{ma}\,\text{movie\_stars} \bowtie_{\text{ma . star\_id = 112}}$

$\rho_{s}\,\text{stars})$

**Q3.** What is the total number of theaters in Ahmedabad?

**Query**

```
SELECT COUNT(theater_id) AS NoOfTheaters FROM Theaters t where t.city_id=(Select
c.city_id from city c where c.name='Ahmedabad') ;
```

**Relational Algebra**

$\pi_{\text{COUNT (theater\_id)} \rightarrow \text{nooftheaters}}$

$\gamma_{\text{COUNT (theater\_id)}}$

$\sigma_{\text{t . city\_id} = \pi_{\text{c . city\_id}}}$

$\sigma_{\text{c . name = "Ahmedabad"}}$

$\rho_{c}\,\text{city}$

$\rho_{t}\,\text{theaters}$

**Q4.** Name the movies with minimum runtime.

**Query**

```
SELECT N.NAME FROM MOVIE N WHERE N.RUNTIME=(SELECT MIN(M.RUNTIME)
FROM MOVIE M)
```

**Relational Algebra**

$\pi_{\text{n . name}}$

$\sigma_{\text{n . runtime} = \pi_{\text{MIN (runtime)}}}$

$\gamma_{\text{MIN (runtime)}}$

$\rho_{m}\,\text{movie}$

$\rho_{n}\,\text{movie}$

**Q5.** List the movie with the least number of tweets between 2022-11-1 between 2022-11-12.

**Query**

```
Select m.name, count(t.movie_id) as count from movie m inner join twitter_tweets t on
t.movie_id=m.movie_id
group by m.name
order by count
limit 1;
```

**Relational Algebra**

$\tau_{\text{COUNT (movie\_id)}}$

$\gamma_{\text{name, COUNT (movie\_id)}}$

$(\rho_{m}\,\text{movie} \bowtie_{\text{t . movie\_id} = \text{m . movie\_id}}$

$\rho_{t}\,\text{twitter\_tweets})$

**Q6.** List actors of movies with highest positive reviews fetched from tweets between 2022-11-1 between 2022-11-12.

**Query**

```
Select s.name from stars s
inner join movie_stars ms on s.star_id=ms.star_id and
ms.movie_id=(Select m.movie_id  from movie m inner join twitter_tweets t on
t.movie_id=m.movie_id
group by m.name
order by sum(t.sentiment)
limit 1);
```

**Relational Algebra**

$$\pi_{s.name}$$
$$(\rho_{s}\ stars \bowtie_{s.star\_id = ms.star\_id\ AND\ ms.movie\_id = \tau\ SUM\ (sentiment)}$$
$$\pi\ m.movie\_id$$
$$\gamma\ name,$$
$$(\rho\ m\ movie \bowtie t.movie\_id = m.movie\_id$$
$$\rho\ t\ twitter\_tweets)$$
$$\rho_{ms}\ movie\_stars)$$

# Anshul's Queries

**Q1.** Top 5 actors with maximum number of movie releases in the year 2016?

**Query**

```
SELECT count(stars.star_id) as top_actors ,stars.name, movie.released_year FROM stars
INNER JOIN movie_stars ON movie_stars.star_id = stars.star_id
INNER JOIN movie ON movie.movie_id = movie_stars.movie_id and movie.released_year =
2016
GROUP BY stars.star_id
ORDER BY top_actors DESC
LIMIT 5;
```

**Relational Algebra**

$\tau_{\text{top\_actors} \downarrow}$

$\pi_{\text{COUNT (star\_id)} \rightarrow \text{top\_actors, stars . name, movie . released\_year}}$

$\gamma_{\text{star\_id, COUNT (star\_id)}} (\text{stars} \bowtie_{\text{movie\_stars . star\_id = stars . star\_id}} \text{movie\_stars} \bowtie_{\text{movie . movie\_id = movie\_stars . movie\_id AND movie . released\_year = 2016}} \text{movie})$

**Q2.** Total number of screens in theaters in the city of Mumbai?

**Query**

```
SELECT c.name, SUM(t.screens) FROM theaters AS t INNER JOIN city AS c ON t.city_id =
c.city_id AND c.name = 'Mumbai';
```

**Relational Algebra**

$\pi_{\text{c . name, SUM (screens)}}$

$\gamma_{\text{SUM (screens)}}$

$(\rho_{\text{t}} \text{theaters} \bowtie_{\text{t . city\_id = c . city\_id AND c . name = "Mumbai"}}$

$\rho_{\text{c}} \text{city})$

**Q3.** List the actors whose movies were released between 2018 and 2021.

**Query**

```
SELECT s.name, m.name, m.released_year
FROM movie AS m INNER JOIN movie_stars AS ms ON m.movie_id = ms.movie_id AND
m.released_year BETWEEN 2018 AND 2021 INNER JOIN stars AS s ON s.star_id =
ms.star_id;
```

**Relational Algebra**

$$\pi_{s.name,\ m.name,\ m.released\_year}$$
$$(\rho_{m}\ movie \bowtie_{m.movie\_id = ms.movie\_id\ AND\ (2018 <= m.released\_year\ AND\ m.released\_year <= 2021)}$$
$$\rho_{ms}\ movie\_stars \bowtie_{s.star\_id = ms.star\_id}$$
$$\rho_{s}\ stars)$$

**Q4.** List the actors of movies with highest number of retweets between "2022-11-01" and "2022-11-12"

**Query**

```
SELECT s.name
FROM stars AS s INNER JOIN movie_stars AS ms ON s.star_id = ms.star_id AND
ms.movie_id = (SELECT m.movie_id
FROM movie AS m INNER JOIN twitter_tweets AS t ON m.movie_id = t.movie_id
WHERE t.created_at BETWEEN '2022-11-01' AND '2022-11-12'
ORDER BY t.retweet DESC
LIMIT 1);
```

**Relational Algebra**

$$\pi_{s.name}$$
$$(\rho_{s}\ stars \bowtie_{s.star\_id = ms.star\_id\ AND\ ms.movie\_id = (\tau\ t.retweet \downarrow}$$
$$\pi\ m.movie\_id$$
$$\sigma\ "2022\text{-}11\text{-}01" <= t.created\_at\ AND\ t.created\_at <= "2022\text{-}11\text{-}12"$$
$$(\rho\ m\ movie \bowtie m.movie\_id = t.movie\_id$$
$$\rho\ t\ twitter\_tweets))$$
$$\rho_{ms}\ movie\_stars)$$

**Q5.** List the movie with most negative reviews fetched from tweets between 2022-11-1 between 2022-11-12.

**Query**

```
SELECT m.name, SUM(t.sentiment) AS sentiment
FROM movie AS m INNER JOIN twitter_tweets AS t ON t.movie_id = m.movie_id
GROUP BY m.name
ORDER BY SUM(sentiment) ASC
LIMIT 1;
```

**Relational Algebra**

$\tau_{\text{SUM (sentiment)}}$
$\pi_{\text{m . name, SUM (sentiment)} \rightarrow \text{sentiment}}$
$\gamma_{\text{name, SUM (sentiment)}}$
$(\rho_{\text{m}}\,\text{movie} \bowtie_{\text{t . movie\_id = m . movie\_id}}$
$\rho_{\text{t}}\,\text{twitter\_tweets})$

**Q6.** List movies with most and distinct tweet mentions fetched from tweets between 2022-11-1 between 2022-11-12.

**Query**

```
SELECT m.name, COUNT(DISTINCT tt.target_user) as all_mentions FROM movie AS m
INNER JOIN twitter_tweets AS t ON m.movie_id = t.movie_id INNER JOIN tweet_mentions
AS tt ON t.tweet_id = tt.tweet_id
WHERE t.created_at BETWEEN '2022-11-1' AND '2022-11-12'
GROUP BY m.name
ORDER BY all_mentions DESC
LIMIT 1;
```

**Relational Algebra**

$\tau_{\text{all\_mentions} \downarrow}$
$\pi_{\text{m . name, COUNT (\delta target\_user)} \rightarrow \text{all\_mentions}}$
$\gamma_{\text{name, COUNT (\delta target\_user)}}$
$\sigma_{\text{"2022-11-1" <= t . created\_at AND t . created\_at <= "2022-11-12"}}$
$(\rho_{\text{m}}\,\text{movie} \bowtie_{\text{m . movie\_id = t . movie\_id}}$
$\rho_{\text{t}}\,\text{twitter\_tweets} \bowtie_{\text{t . tweet\_id = tt . tweet\_id}}$
$\rho_{\text{tt}}\,\text{tweet\_mentions})$

# Soham's Queries

**Q1.** Best rated movie between 2015 to 2016.

**Query -**

```
SELECT name, max(rating)
FROM movie
WHERE released_year BETWEEN 2015 AND 2016
ORDER BY name ASC;
```

**Relational Algebra**

$$\tau_{\text{name}}$$
$$\pi_{\text{name, MAX (rating)}}$$
$$\gamma_{\text{MAX (rating)}}$$
$$\sigma_{\text{2015 <= released\_year AND released\_year <= 2016}}\ \text{movie}$$

**Q2.** Which theater had the highest sales?

**Query -**

```
SELECT t.name, MAX(s.ticket_price*s.ticket_sold)
FROM theaters t
INNER JOIN screens sc ON t.theater_id = sc.theater_id
INNER JOIN screen_shows s ON sc.screen_id = s.screen_id;
```

**Relational Algebra**

$$\pi_{\text{t . name, MAX (?column?)}}$$
$$\gamma_{\text{MAX (?column?)}}$$
$$(\rho_{\text{t}}\ \text{theaters} \bowtie_{\text{t . theater\_id = sc . theater\_id}}$$
$$\rho_{\text{sc}}\ \text{screens} \bowtie_{\text{sc . screen\_id = s . screen\_id}}$$
$$\rho_{\text{s}}\ \text{screen\_shows})$$

**Q3.** Year with least number of movies

**Query -**

```
SELECT released_year, COUNT(movie_ID) as movie_count
FROM movie
GROUP BY released_year
ORDER BY movie_count ASC
LIMIT 1;
```

**Relational Algebra**

$$\tau_{\text{movie\_count}}$$
$$\pi_{\text{released\_year, COUNT (movie\_id)} \rightarrow \text{movie\_count}}$$
$$\gamma_{\text{released\_year, COUNT (movie\_id)}}\ \text{movie}$$

**Q4.** List movies with most tweet tags fetched from tweets between 2022-11-1 between 2022-11-12.

**Query**

```
SELECT m.name, COUNT (tt.tag) FROM movie AS m INNER JOIN twitter_tweets AS t ON
m.movie_id = t.movie_id INNER JOIN tweet_tags AS tt ON t.tweet_id = tt.tweet_id
WHERE t.created_at BETWEEN '2022-11-1' AND '2022-11-12'
GROUP BY m.name
ORDER BY COUNT (tt.tag) DESC
LIMIT 1;
```

**Relational Algebra**

$$\tau_{\text{COUNT (tag)}\downarrow}$$
$$\gamma_{\text{name, COUNT (tag)}}$$
$$\sigma_{\text{"2022-11-1" <= t . created\_at AND t . created\_at <= "2022-11-12"}}$$
$$(\rho_{m}\,\text{movie} \bowtie_{\text{m . movie\_id = t . movie\_id}}$$
$$\rho_{t}\,\text{twitter\_tweets} \bowtie_{\text{t . tweet\_id = tt . tweet\_id}}$$
$$\rho_{tt}\,\text{tweet\_tags})$$

**Q5.** List movies with most tweets fetched from tweets between 2022-11-1 between 2022-11-12.

**Query**

```
SELECT m . movie_id, COUNT (t . tweet_id)AS tweet_count
FROM movie AS m INNER JOIN twitter_tweets AS t ON m . movie_id = t . movie_id
GROUP BY t . movie_id
ORDER BY tweet_count DESC
LIMIT 1;
```

**Relational Algebra**

$$\tau_{\text{tweet\_count}\downarrow}$$
$$\pi_{\text{m . movie\_id, COUNT (tweet\_id)} \rightarrow \text{tweet\_count}}$$
$$\gamma_{\text{movie\_id, COUNT (tweet\_id)}}$$
$$(\rho_{m}\,\text{movie} \bowtie_{\text{m . movie\_id = t . movie\_id}}$$
$$\rho_{t}\,\text{twitter\_tweets})$$