

SOCIAL NETWORK DATABASE

LET'S PROCEED IN



THE TEAM

Yousef Mahmoud

Data Collection & Insertion

Yousef Mahmoud

Docs Writing & Presentation Design

Yousef Mahmoud

Database Design & SQL Queries

supervisors

MOHAMED HASHIM

WAEL ZAKARIA

Table of Contents.

01> PROBLEM



02> VISION & OUTLINES



03> REQUIREMENTS



04> DATA COLLECTION



05> ER DIAGRAM



06> RELATIONAL MODEL



07> QUERIES



01

Social media platforms are built around various aspects.

- **Facebook:** engagement.
- **Twitter:** news & discussions.
- **TikTok:** video entertainment.
- **Reddit:** sub-communities.
- **Instagram:** visuals and reels.
- **LinkedIn:** networking & connections.

01

But what about a new platform?

a platform centered on insightful and valuable content.

02

Platform vision & outlines:

- Focus on insightful and valuable content, encouraging users to share meaningful posts.
- Motivate regular participation through features like active-user streaks and badges for top contributors in different topics.
- Like and Dislike options, with required feedback for dislikes, to support a respectful, quality-focused environment.
- Make trending topics and impactfull posts easily discoverable.

03

User Requirements

- Users can create accounts with personal information (e.g., name, email, password, profile picture, and education).
- Can invite others to join the platform.
- Following and chat with each other.
- Each user must choose interest or more.

Post Features

- Posts must include a caption and can optionally contain media such as photos or videos.
- Posting time should be stored for each post.
- Posts can be tagged with one or more topics or categories.
- Users can interact with posts through likes, dislikes (with required feedback), comments, or reposts.

03

Communities

- Communities should allow users to form groups around shared interest.
- Users can create, join and participate in discussions within communities.

Event Management

- Users can create events associated with a specific interest or topic.
- Events must have a title, description, location, date, and a maximum number of attendees.
- Users can choose to attend events or express interest without committing to attendance.

03

Analytics & Insights

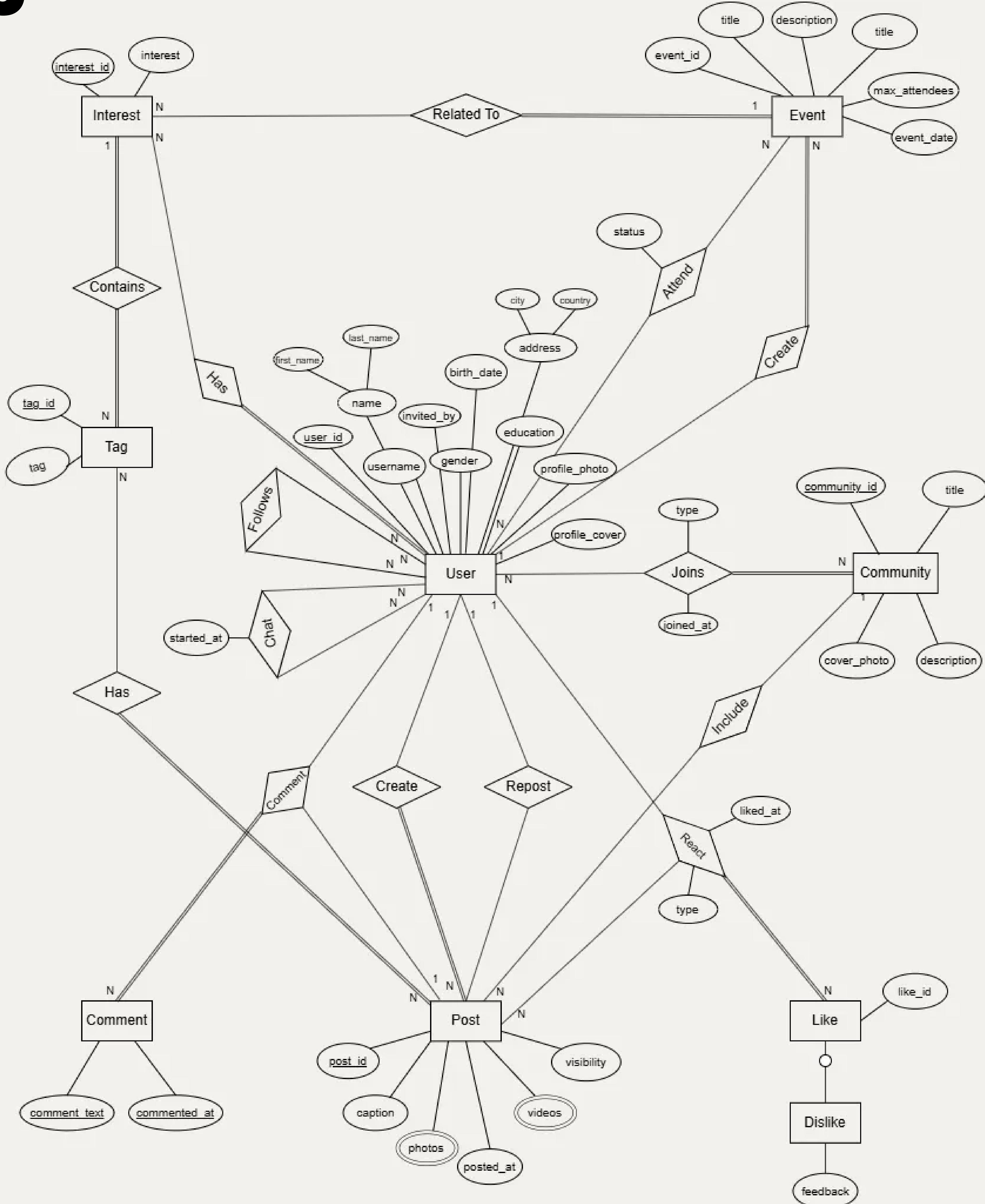
- Track user activity, including post streaks and consistency.
- Identify and highlight top contributors (users with high engagement or posts).
- Analyze trending topics and posts based on likes, comments, and reposts.
- Provide insights into highly liked and disliked posts, including reasons for dislikes.
- Recommend users to follow based on shared interests or mutual connections.
- Identify the least active users and suggest re-engagement strategies.
- Track events to compare predicted versus actual attendee numbers.
- Highlight impactful posts and trending discussions on the platform.

04

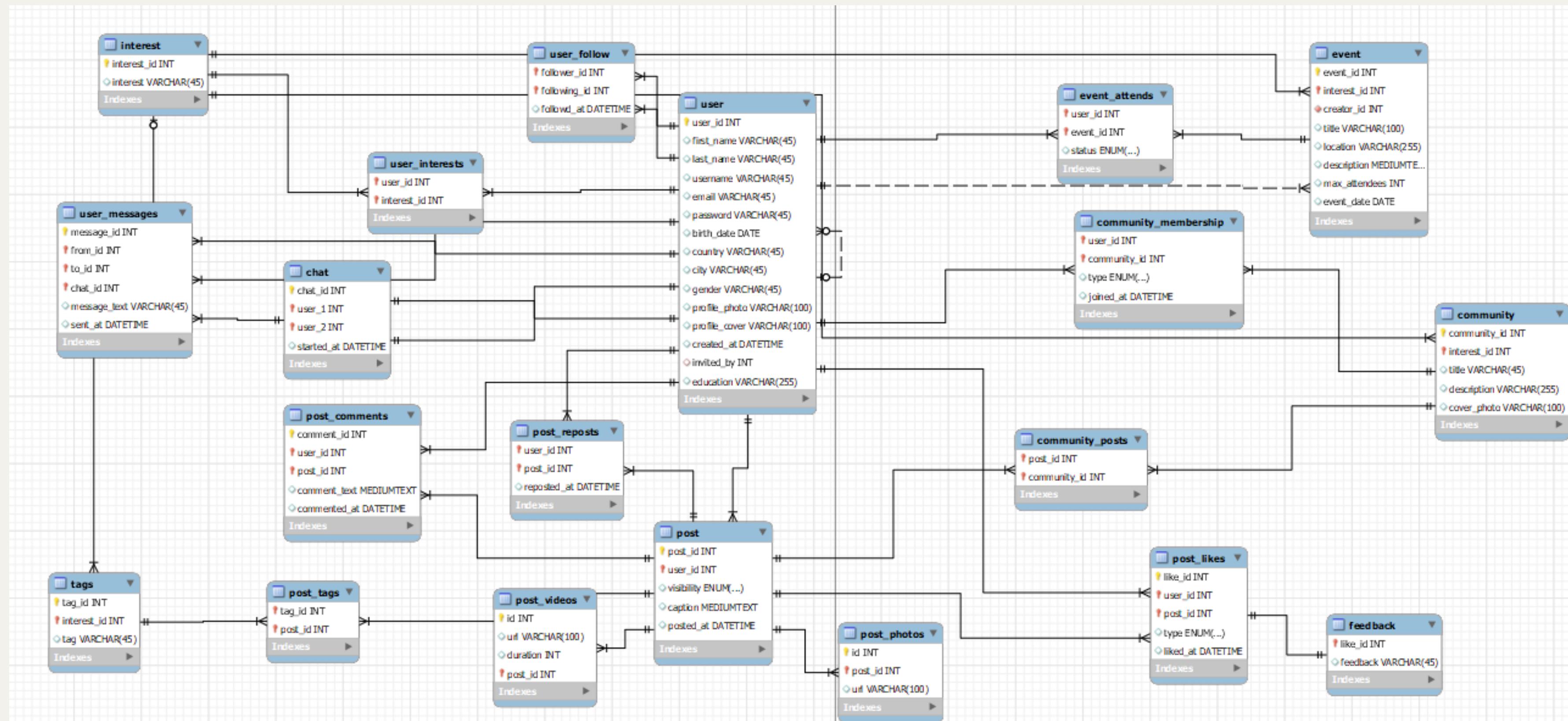
Data Collection and Generation

- Mockaroo and chatGPT were used to generate the main data and information of users.
- Python scripts were utilized to simulate relationships between entities.
- Also used pandas in python to manipulate data fields.
- Added random variations to make the data more realistic, such as different post popularity levels, feedback for dislikes, and varied activity timelines.
- Finally repared clean datasets in CSV formats for easy loading into the database.
- All scripts and files are publicly available on github

05



06



07

1- users full name with education, country, and city.

```
select
    concat(first_name, ' ', last_name) as full_name,
    education,
    country,
    city
from user
order by full_name;
```

07

	full_name	education	country	city
1	Abba Brasier	Universidad Nacional Abierta ...	Colombia	Jardín
2	Abba Mullaney	Univerzitet u Mostaru Dzemal ...	Bosnia and Herzeg...	Ilići
3	Abe Holleworth	Beijing Normal University	China	Dongjia
4	Abie Bazley	Universit <small>đ</small> de Monastir	Tunisia	Manzil B <small>đ</small> Zalafah
5	Adam Evitts	National College of Ireland	Ireland	Greenhills
6	Addie Vernau	Universidade Fernando Pessoa	Portugal	Carvoeira
7	Adelaida Stearn	Davao Doctors College	Philippines	Barili
8	Adeline Troth	Moscow State Textile Universi...	Russia	Dyurtyuli
9	Adrian Golding	Ateneo de Manila University	Philippines	Tiaong
10	Adrian Hilling	University of Roma La Sapienza	Italy	Roma

07

2- posts that have word “help”.

```
select
    post_id,
    caption,
    posted_at
from post
where caption like '%help%';
```

07

	post_id	caption	posted_at
1	206	In help table avoid special.	2023-07-07 16:02:00
2	215	Help laugh try tough similar.	2023-02-26 10:41:00
3	236	Help we thousand hair PM ago paper.	2024-09-23 03:11:00
4	285	Look recently big paper customer young help.	2021-05-30 20:48:00
5	564	Help eat peace nor fact.	2023-03-10 16:07:00
6	582	Social help life discussion child.	2019-01-03 18:09:00
7	1175	We help stand pressure buy skin.	2020-11-03 21:29:00
8	1418	Help drug society son few.	2023-04-14 23:06:00
9	1718	Help participant site animal.	2022-06-14 14:52:00
10	1801	Reflect range just help opportunity interesting Ameri...	2019-12-03 07:59:00

07

3- most active users in sharing posts.

```
select
    user.user_id,
    concat(user.first_name, ' ', user.last_name) as 'Full Name',
    count(post.post_id) as "Number of Posts"
from user
    left join post on post.user_id = user.user_id
group by user.user_id
order by count(post.post_id) desc;
```

07

	user_id	Full Name	Number of Posts
1	321	Durand Sisselot	21
2	258	Camilla Parsons	20
3	827	Basilio Callow	19
4	514	Patrica Plaster	18
5	437	Anatole Normadell	17
6	816	Karil Care	17
7	636	Donall Pobjoy	17
8	18	Juli Drust	17
9	591	Nealon Scimoni	16
10	757	Morie Crickett	16

07

4- top 5 liked posts.

```
select
    post_id,
    caption,
    (
        select count(*) from post_likes where post_id = post.post_id
    ) as like_count
from post
order by like_count desc
limit 5;
```

07

	post_id	caption	like_count
1	714	Base now resource north eat attorney or.	8
2	5386	Relationship model crime mean.	7
3	6951	Rise fine my decision.	7
4	565	Glass eye will issue teacher attention make rest.	7
5	6483	Region even with arrive open.	7

07

5- average number of likes on posts.

```
select
    avg(likes_count) as "likes average"
from (
    select
        count(*) as likes_count
    from post_likes
    group by post_id
) as s;
```

07

	□ `likes average` :
1	1.8784

07

6- users who joined in the last 3 months.

```
select
    user.user_id,
    concat(user.first_name, ' ', user.last_name) as "Full Name",
    user.created_at as "Joined Date"
from user
where created_at >= now() - interval 3 month;
```

07

	user_id	Full Name	Joined Date
1	27	Anastassia Fussell	2024-11-10 00:25:07
2	46	Blisse Bernakiewicz	2024-11-10 07:19:23
3	52	Giffer Sollon	2024-11-29 05:50:17
4	68	Willis Loveitt	2024-10-16 14:18:01
5	76	Silvia Skade	2024-11-12 07:08:20
6	94	Beryle Hankey	2024-11-10 15:17:07
7	100	Silvain Domm	2024-12-03 17:53:53
8	118	Brandais Caveill	2024-10-10 10:00:41
9	146	Cosette Behne	2024-11-05 18:53:50
10	189	Luelle Mahy	2024-11-01 09:09:35

07

7- users who didn't post in the last 2 years.

```
select
    user.user_id,
    user.first_name
from user
where user.user_id not in (
    select
        post.user_id
    from post
    where post.posted_at >= now() - interval 2 year
);
```

07

	user_id	first_name
1	11	Abba
2	19	Hadlee
3	26	Paule
4	33	Ruthie
5	36	Alvan
6	41	Garvin
7	45	Idalina
8	47	Dareen
9	51	Gloriane
10	53	Addie

07

8- what is the best time for posting?

```
select
    hour(posted_at) as post_hour,
    count(*) as post_count
from post
group by post_hour
order by post_count desc;
```

07

	post_hour	post_count
1	16	321
2	8	320
3	20	315
4	19	312
5	5	312
6	2	306
7	15	303
8	17	303
9	10	295
10	11	294

07

9- re-posts number of posts that have the word “environmental”.

```
select
    count(post_reposts.user_id) as "Number of Reposts"
from post
    left join post_reposts
        on post_reposts.post_id = post.post_id
where post.caption like '%environmental%';
```

07

	Number of Reposts
1	219

07

10- countries with their total number of interactions.

```
select
    user.country,
    (select count(*) from post_likes where user_id = user.user_id) +
    (select count(*) from post_comments where user_id = user.user_id) +
    (
        select count(*) from post_reposts where user_id = user.user_id
    ) as total_interactions
from user
group by user.country
order by total_interactions desc;
```

07

	country	total_interactions
1	Syria	80
2	Hungary	79
3	Democratic Republic of the Congo	78
4	Iceland	76
5	Kenya	76
6	Zambia	74
7	Jamaica	73
8	Belgium	72
9	Indonesia	72
10	Mauritania	70

07

11- users' number of invitations.

```
select
    concat(a.first_name, ' ', a.last_name) as "Full Name",
    count(b.invited_by) as "Number of Invitations"
from user a
    left join user b
        on a.user_id = b.invited_by
group by a.user_id
order by count(b.invited_by) desc;
```

07

	□ `Full Name`	▪	□ `Number of Invitations`	▪
1	Emlyn Chatelot			5
2	Tamera McLeese			5
3	Isa Van der Krui			4
4	Dawn Kleynen			4
5	Chase Fost			4
6	Bertrando Tingly			3
7	Rhys Bister			3
8	Elly Scrase			3
9	Torre Lamps			3
10	Giordano Collinette			3

07

12- list of user with their number of attended and interested events.

```
select
    user.first_name,
    coalesce(
        sum(case when event_attends.status = 'attend' then 1 else 0 end),
        0) as "Attending",
    coalesce(
        sum(case when event_attends.status = 'interested' then 1 else 0 end),
        0) as "Interested"
from user
left join event_attends
    on user.user_id = event_attends.user_id
group by user.user_id;
```

07

	□ first_name	□ Attending	□ Interested
1	Randi	20	5
2	Lindy	12	7
3	Johnette	20	4
4	Charlean	21	7
5	Andromache	18	9
6	Maure	24	4
7	James	20	5
8	Twila	26	4
9	Coraline	20	5
10	Avivah	18	6

07

13- number of users who born in each 5 years.

```
select
    case
        when year(birth_date) between 1990 and 1995 then '1990 - 1995'
        when year(birth_date) between 1996 and 2000 then '1996 - 2000'
        when year(birth_date) between 2001 and 2005 then '2001 - 2005'
        when year(birth_date) between 2006 and 2010 then '2006 - 2010'
        end as birth_date_range,
    count(*) as "Count of Users"
from user
group by birth_date_range;
```

07

	birth_date_range	Count of Users
1	1990 - 1995	314
2	1996 - 2000	238
3	2001 - 2005	235
4	2006 - 2010	213

07

14- all people who “Bobbie Perrigo” talked with.

```
select
    concat(other.first_name, ' ', other.last_name) as "Full Name"
from user person
    left join chat
        on (chat.user_1 = person.user_id
            or chat.user_2 = person.user_id)
    left join user other
        on (other.user_id = chat.user_1 or other.user_id = chat.user_2)
            and other.user_id ◇ person.user_id
where person.first_name = 'Bobbie'
    and person.last_name = 'Perrigo';
```

07

	□ `Full Name`
1	Kirbie Ubanks
2	Consuela Winder
3	Ewell Housego
4	Cristina Gouda
5	Lilian Gorrie
6	Beckie Orrice
7	Starlene Corderoy
8	Peder Blewett
9	Jenine Norwich
10	Hilario Wardley

07

15- what are events hit the maximum attendees number?

```
select
    event.event_id,
    event.title,
    event.max_attendees,
    count(event_attends.user_id) as "actual attendees"
from event
    left join event_attends
        on event.event_id = event_attends.event_id
            and event_attends.status = 'attend'
group by event.event_id, event.max_attendees
having count(event_attends.user_id) > event.max_attendees;
```

07

	event_id	title	max_attendees	actual attendees
1	8	NutritionCon 2021	87	126
2	13	MediaExpo International	57	80
3	19	ShowcaseLive La Union	135	155
4	22	MusicFest World	90	149
5	32	VisualArts 2024	78	96
6	38	MarketingExpo International	154	161
7	39	CookingShow World	161	172
8	51	WellnessCon	130	177
9	57	Global EcoSummit	141	181
10	67	Tuojiang BusinessForum	70	132

07

16- events that will happen in the next month.

```
select
    event_id,
    title,
    event_date,
    location
from event
where event_date >= now()
    and event_date <= now() + interval 1 month;
```

07

	event_id	title	event_date	location
1	24	Dogonbadan BlockchainCon	2024-12-24	Dogonbadan
2	172	CreativeCon World	2024-12-14	Jetak Kulon
3	181	International DataSummit	2024-12-10	Pamakayo

07

17- what is the most frequent age?

```
select
    round(datediff(now(), birth_date) / 365) as age,
    count(user_id) as "Number of Users"
from user
group by age
order by count(user_id) desc;
```

07

	age	Number of Users
1	25	62
2	17	59
3	33	59
4	34	59
5	30	57
6	22	55
7	28	53
8	20	46
9	16	46
10	31	45

07

18- top 5 interests chosen by users.

```
select
    interest,
    count(user_id) as "Number of Users"
from interest
    left join user_interests
        on interest.interest_id = user_interests.interest_id
group by interest
order by count(user_id) desc
limit 5;
```

07

	interest	Number of Users
1	Food	163
2	Health	158
3	Art	157
4	Science	156
5	Business	154

07

19- what are the most trending topics?

```
select
    tags.tag,
    interest.interest,
    count(post_tags.post_id) as "Number of Posts"
from tags
    left join interest
        on interest.interest_id = tags.interest_id
    left join post_tags
        on post_tags.tag_id = tags.tag_id
group by tags.tag_id, tags.tag, interest.interest
order by count(post_tags.post_id) desc;
```

07

	>tag	interest	Number of Posts
1	Sculpture	Art	260
2	Painting	Art	259
3	UI/UX	Design	254
4	Gadgets	Technology	248
5	Adventure	Travel	245
6	Graphic Design	Design	244
7	Entrepreneurship	Business	240
8	Workshops	Education	237
9	Elections	Politics	236
10	Software Development	Technology	235

07

20- showing all users with their followers and following number.

```
select
    user.user_id,
    user.first_name,
    coalesce(followers.follower_count, 0) as followers_count,
    coalesce(following.following_count, 0) as following_count
from user
left join (
    select
        following_id as user_id,
        count(follower_id) as follower_count
    from user_follow
    group by following_id
) as followers  on user.user_id = followers.user_id
left join (
    select
        follower_id as user_id,
        count(following_id) as following_count
    from user_follow
    group by follower_id
) as following  on user.user_id = following.user_id
order by followers_count desc, following_count desc;
```

07

	user_id	first_name	followers_count	following_count
1	275	Tam	496	56
2	727	Ganny	495	58
3	251	Lilas	493	58
4	574	Quinlan	491	60
5	44	Douglass	490	57
6	709	Irita	479	38
7	311	Edgar	476	62
8	232	Raul	474	49
9	770	Percy	472	48
10	711	Candace	467	56

07

21- displaying each post details and engagement metrics.

```
select
    user.first_name,
    post.caption,
    (case when post.visibility = 'public' then 'public' else community.title end) as visibility,
    (select count(*) from post_likes where post_likes.post_id = post.post_id) as like_count,
    (select count(*) from post_comments where post_comments.post_id = post.post_id) as comment_count,
    (select count(*) from post_reposts where post_reposts.post_id = post.post_id) as repost_count
from post
    left join user
        on post.user_id = user.user_id
    left join community_posts
        on post.post_id = community_posts.post_id
    left join community
        on community.community_id = community_posts.community_id
order by user.first_name,like_count desc, comment_count desc, repost_count desc;
```

07

	first_name	caption	visibility	likes	comments	reposts
1	Verna	Base now resource north eat attor...	public	8	1	0
2	Lukas	Project nothing family between in...	Graphic Designers Hub	7	3	0
3	Bella	Rise fine my decision.	public	7	2	0
4	Harley	Region even with arrive open.	public	7	1	18
5	Brig	Relationship model crime mean.	Style Trends	7	1	0
6	Sheelah	Glass eye will issue teacher atte...	public	7	0	4
7	Douglass	Reveal hold morning.	Holistic Healing	6	4	0
8	Sidoney	Bad thing as Mr brother.	public	6	3	0
9	Sarajane	Reach model leader including oper...	Style Trends	6	2	0
10	Karrie	Fund follow more structure electi...	Runway Fashionistas	6	2	0

07

22- number of posts associated with tags combinations.

```
select
    t1.tag as tag1,
    t2.tag as tag2,
    count(distinct pt1.post_id) as post_count
from post_tags pt1
    left join post_tags pt2
        on pt1.post_id = pt2.post_id
        and pt1.tag_id < pt2.tag_id
    join tags t1
        on pt1.tag_id = t1.tag_id
    join tags t2
        on pt2.tag_id = t2.tag_id
group by t1.tag, t2.tag
order by post_count desc;
```

07

	□ tag1	□ tag2	□ post_count
1	Painting	Sculpture	97
2	Interior Design	UI/UX	90
3	Painting	Calligraphy	89
4	Graphic Design	UI/UX	88
5	Policy Discussions	Diplomacy	88
6	Sculpture	Calligraphy	84
7	Home Workouts	Weightlifting	84
8	Entrepreneurship	Marketing	83
9	Weightlifting	Running	82
10	Graphic Design	Interior Design	81

07

23- are there any posts discussing topics differ from the community's main focus?

```
select
    community.community_id,
    post.post_id,
    post.caption,
    tags.tag as post_tag,
    interest.interest
from post
    left join post_tags
        on post.post_id = post_tags.post_id
    left join tags
        on post_tags.tag_id = tags.tag_id
    left join community_posts
        on post.post_id = community_posts.post_id
    left join community
        on community.community_id = community_posts.community_id
    left join interest
        on interest.interest_id = community.interest_id
where tags.interest_id ≠ community.interest_id
group by community.community_id;
```

07

	community_id	post_id	caption	post_tag	interest
1	8	110	Feel population add fast establish.	Chemistry	Entertainment
2	17	80	All occur rather prevent that beaut...	Entrepreneurship	Automotive
3	19	22	Brother chance crime.	Entrepreneurship	Lifestyle
4	29	387	Deep easy born run standard.	Fitness	Sports
5	34	18	While prove name other hear now par...	Esports	Career
6	36	28	Group issue media administration al...	TV Shows	Fashion
7	58	147	Goal number Congress great capital ...	Software Development	Sports
8	70	3	Set cold manage fine.	Blockchain	Adventure

07

There is still more,
check it in queries.sql file

#join_communities_with_shared_interests.

Thank you!