Assignment - 3

Project Cont'd

Consider an online video-sharing platform like YouTube which hosts tens of thousands of channels and crores of users.

You have to analyse the data and provide meaningful insights on the type of content that drives engagement, users growth, and many more to all the stakeholders. Let's roll our sleeves up for an insightful analysis!

Database

The sample database consists of tables that store the information of users, channels, videos, genres and likes/dislikes.

Note:

channel_user table

channel_id	user_id	subscribed_datetime
100	1	2020-12-10 10:30:45
100	7	2020-10-10 11:30:45
...	...	...

`channel_user` table stores the data of the channel_ids and their subscribers' user_ids.

First row in the table represents that the user with user_id = 1 is subscribed to the channel with channel_id = 100 at `2020-12-10 10:30:45`

user_likes table

user_id	video_id	reaction_type	reacted_at
1	10	LIKE	2020-12-10 10:30:45
7	10	DISLIKE	2020-10-10 11:30:45
...	...	...	...

Similarly, `user_likes` table stores the data of video_id and the user_ids who reacted to the video.

video_genre table

video_id	genre_id
10	201
10	202
...	...

Similarly, `video_genre` table stores the data of video_id and the ids of the genres that the corresponding video belongs to.

Let's dive in to analyze the in and outs of each part of the data. Here we go!

1. Fetch the top 10 videos having more number of views.

   Note:

   Sort the output in the descending order of no_of_views, and then in the ascending order of name.

   Expected Output Format:

name	no_of_views
...	...

SELECT name, no_of_views FROM video

ORDER BY no_of_views DESC,name ASC LIMIT 10

2. Fetch the top 10 videos having more number of views, along with the channel details.

   Note:

   Sort the output in the descending order of no_of_views, and then in the ascending order of channel_name.

   Expected Output Format:

video_name	no_of_views	channel_name
...	...	...

```
SELECT video.name AS video_name, video.no_of_views AS no_of_views, channel.name AS channel_name
FROM video INNER JOIN channel ON channel.channel_id = video.channel_id
ORDER BY no_of_views DESC, channel_name ASC LIMIT 10
```

3. Get all the music videos released before the year 2016.

Note:

- You can consider the videos which contain "music" in name as music videos.
- Get the year in the integer format
- Sort the output in the descending order of year, and then in the ascending order of name.

Expected Output Format:

name	no_of_views	year
...	...	...

```
SELECT name,no_of_views,cast(strftime("%Y", published_datetime) AS integer) AS year FROM video
WHERE name LIKE '%music%' AND cast(strftime("%Y", published_datetime) AS integer) < 2016
ORDER BY year DESC,name ASC
```

4. Get all distinct channels which published music videos before 2016.

Note:

- You can consider the videos which contain "music" in name as music videos.
- Sort the output in the ascending order of chanel_name.

Expected Output Format:

channel_id	channel_name
...	...

```
SELECT DISTINCT channel.channel_id, channel.name AS channel_name FROM channel
 INNER JOIN video ON channel.channel_id = video.channel_id
WHERE video.name LIKE '%music%'
AND cast(strftime('%Y', published_datetime) AS integer) < 2016
ORDER BY channel_name ASC
```

5. Get all the review videos, i.e., videos which contain "review" in their name.

Note:

- Sort the output in the descending order of no_of_views.

Expected Output Format:

name	no_of_views
...	...

SELECT name,no_of_views FROM video WHERE name LIKE '%review%' ORDER BY no_of_views DESC

6. Get all the unique channels that published review videos.

Note:

- You can consider the videos which contain "review" in name as review videos.
- Sort the output in the ascending order of channel_name.

Expected Output Format:

channel_id	channel_name
...	...

SELECT DISTINCT channel.channel_id, channel.name AS channel_name

FROM video INNER JOIN channel ON channel.channel_id = video.channel_id

WHERE video.name LIKE '%review%' ORDER BY channel_name ASC

7. Get all the videos that belong to "Action" genre (genre_id = 201) and have more than 1 lakh views.

Note:

- Sort the output in the ascending order of video_id.

Expected Output Format:

video_id	name	genre_id
...	...	...

```
SELECT video.video_id, video.name, video_genre.genre_id
FROM video INNER JOIN video_genre ON video_genre.video_id = video.video_id
WHERE genre_id = 201 AND video.no_of_views > 100000
ORDER BY video.video_id ASC
```

8. Get all the Indian users details whose age is below 30 years and liked the video (video_id = 1011) in the year 2020.

Note:

- Consider the name of the country as "INDIA"
- Consider reaction_type LIKE as liked.
- Sort the output in the ascending order of name.

Expected Output Format:

name	gender	age	country	premium_membership
...	...	...	...	...

```
SELECT user.name, user.gender, user.age, user.country, user.premium_membership
FROM user
 INNER JOIN user_likes ON user.user_id = user_likes.user_id
 INNER JOIN video ON video.video_id = user_likes.video_id
WHERE user.country = 'INDIA' AND user_likes.reaction_type = 'LIKE' AND video.video_id = 1011 AND user.age < 30
ORDER BY user.name ASC
```

9. Find the number of videos published between the years 2010 & 2016.

Note:

- Sort the output in the ascending order of the year
- Keep the year in the integer format

Expected Output Format:

year	no_of_videos
...	...

```
SELECT
  cast(strftime('%Y', published_datetime) AS integer) AS year,
  count(*) AS no_of_videos
FROM video
WHERE cast(strftime('%Y', published_datetime) AS integer) BETWEEN 2010 AND 2016
GROUP BY year
```

10. Between 2010 & 2020, find the number of videos released in each of the below genres.

Note:

- genre_ids : 201, 202, 204, 205, 206, 207
- Sort the output in the descending order of no_of_videos, and then in the ascending order of genre_id.

Expected Output Format:

genre_id	no_of_videos
...	...

```
SELECT
  video_genre.genre_id,
  count(video.video_id) AS no_of_videos
FROM video_genre
  INNER JOIN video ON video_genre.video_id = video.video_id
WHERE video_genre.genre_id IN (201, 202, 204, 205, 206, 207)
  AND cast(strftime('%Y', published_datetime) AS integer) BETWEEN 2010 AND 2020
GROUP BY genre_id
ORDER BY no_of_videos DESC, genre_id ASC
```

11. Get the total number of channels in the database.

Expected Output Format

no_of_channels
...

SELECT count(channel_id) AS no_of_channels FROM channel

12. Get the total number of users subscribed for the channel "Tyler Oakley" (channel_id = 376) in the year 2018.

Expected Output Format

no_of_subscribers
...

SELECT count(channel_user.user_id) AS no_of_subscribers

FROM channel_user

INNER JOIN channel ON channel.channel_id = channel_user.channel_id

WHERE channel.channel_id = 376

AND cast(strftime('%Y', subscribed_datetime) AS integer) = 2018

13. Get the total number of countries where the users of the platform are present.

Expected Output Format

country_count
...

SELECT count(DISTINCT country) AS country_count FROM user

14. Get the total number of countries where the subscribers of the Taylor Swift channel (channel_id = 399) are present.

Expected Output Format

country_count
...

SELECT count(DISTINCT country) AS country_count

FROM user

INNER JOIN channel_user ON user.user_id = channel_user.user_id

WHERE channel_user.channel_id = 399

15. Insights about users:
Get the geographic distribution of Taylor Swift channel (channel_id = 399) subscribers.

Note:

- Geographic distribution: Number of Taylor Swift subscribers present in each country. Ignore the countries where `no_of_subscribers` is 0.
- Order the result in the ascending order of the country name.

Expected Output Format

country	no_of_subscribers
...	...

```
SELECT
  user.country,
  count(channel_user.user_id) AS no_of_subscribers
FROM user
  INNER JOIN channel_user ON user.user_id = channel_user.user_id
WHERE channel_user.channel_id = 399
GROUP BY user.country
```

User Table

user_id	name	gender	age	country	premium_membership
2000	John White	M	63	AUSTRALIA	1
2001	John Andrews	M	67	AUSTRALIA	1
2002	April Robinson	F	60	SRILANKA	1
2003	Kathy Ryan	F	81	SRILANKA	0
2004	Megan Bradshaw	F	53	AUSTRALIA	1
2005	Melissa Sullivan	F	45	BANGLADESH	1
2006	Amber Smith	F	97	BANGLADESH	0
2007	Sara Higgins	F	85	ENGLAND	1

Channel Table

channel_id	name	owner_id	created_datetime
350	Motivation grid	1011	2014-10-05 17:32
351	Marvel	1011	2014-10-05 17:32
352	Disney	1012	2015-09-10 15:32
353	Tedx	1013	2016-10-05 17:32
354	ETV	1011	2012-08-17 19:32
355	Maa	1012	2007-03-07 11:32
356	PewDiePie	1012	2004-10-05 11:32
357	VanossGaming	1013	2005-09-10 15:32

Video Table

video_id	name	duration_in_secs	published_datetime	no_of_views	channel_id
1000	Getting My Driver's License | Lele Pons	3652	2011-12-05 19:00	10619	367
1001	Apple iPhone X Review: The Best Yet!	4556	2021-01-19 20:12	140012	362
1002	Victoria Beckham Gives Strangers Fashion Advice for $2 in Central Park | Vanity Fair	836	2021-01-19 20:12	75609	353
1003	4 Reasons I Don't Like Thanksgiving || Mayim Bialik	1751	2017-05-05 17:32	279351	350
1004	Maroon 5 - What Lovers Do (Live On The Ellen DeGeneres Show/2017)	3186	2018-02-17 19:32	94945	354
1005	Alicia Keys - When You Were Gone	3737	2018-04-10 12:32	35526	361
1006	U.S. Navy Three Carrier Formation in Western Pacific Ocean	3538	2021-01-19 20:12	180973	362
1007	DIY - Simon's Cat | NEW BLACK & WHITE!	1365	2012-09-07 11:32	106271	355

User_likes

user_id	video_id	reaction_type	reacted_at
2141	1529	LIKE	2012-04-12 04:46
2234	1529	LIKE	2012-04-30 17:47
2245	1529	DISLIKE	2012-05-03 21:24
2570	1529	LIKE	2012-04-14 23:08
2544	1529	LIKE	2012-04-17 22:58
2560	1529	LIKE	2012-05-01 01:20
2002	1529	LIKE	2012-04-10 16:53
2118	1529	LIKE	2012-04-24 08:32

Channel_User

channel_id	user_id	subscribed_datetime
376	2521	2018-02-06 20:43
376	2798	2018-07-10 06:12
376	2644	2017-11-11 19:20
376	2566	2018-06-27 17:32
376	2578	2017-12-26 17:11
376	2289	2017-11-10 08:40
376	2997	2017-09-26 21:52

Genre

genre_id	genre_type
201	ACTION
202	COMEDY
203	THRILLER
204	ROMANTIC
205	TECHNICAL
206	HEALTH CARE
207	GAMING
208	MUSIC VIDEOS
209	TEASER
210	INSPIRATIONAL
211	SCI-FI

Video_Genre

video_id	genre_id
1573	205
1574	205
1575	205
1576	205
1577	205
1578	205
1579	205
1580	205