INSTAGRAM USER ANALYTICS

Kushagra Singh
Feb 2023

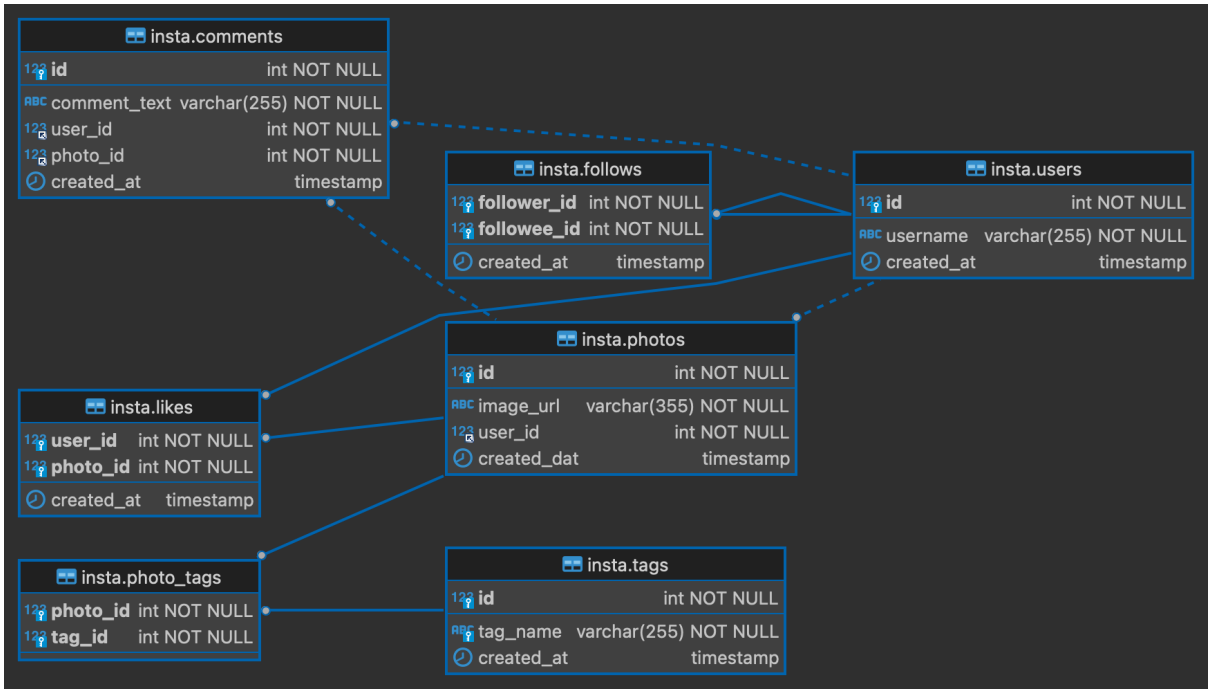
Description:

This project “Instagram User Analytics” tracks how users engage and interact with our digital product i.e. Instagram (in this case).
The project attempts to derive precise and intelligent business insights for marketing, product & development teams.
The insights gained are then can used by teams across the business to launch a new marketing campaign, decide on features to build, maintain and improvise Instagram to track the success of the platform by measuring user engagement and improve the experience altogether while helping the business grow.
In this I have imagined working with the product team of Instagram and the product manager who has asked me to provide insights on the questions asked by the management team.

Basic Approach:

I have prepared a database name ‘insta’ which contains important tables like users, photos, comments, likes, follows, tags and photo tags. And using SQL have answered the day-to-day challenging questions that must be faced by Instagram teams.
Here via Dbeaver (22.3) and SQL, I have used ER-diagram, Operators, Aggregated functions, Sorting functions, Joins and Sets to create a database, storing data in the form of tables, modifying and extracting it.
With all the above tools and approach, I have collected, transformed and organized data to discover, interpret and communicate significant patterns in data to draw conclusions, make predictions, and drive informed decision making.
Quite simply, this focuses on using insights derived from data to make more informed decisions that will help Instagram to increase sales, reach and profits.

ER Diagram:



Marketing: The marketing team wants to launch some campaigns, and they need your help with the following questions

Q1. Marketing team is rewarding most loyal users, the people who have been using the platform for the longest time.

SELECT * FROM insta.users ORDER BY created_at LIMIT 5		
id	**username**	**created_at**
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26

The 5 oldest users of the Instagram from the database used

Q2. Marketing team want to remind inactive users to start posting by sending them promotional emails to post their 1st photo.

SELECT * FROM users WHERE id NOT IN (SELECT user_id FROM photos)		
id	username	created_at
5	Aniya_Hackett	2016-12-07 01:04:39
7	Kasandra_Homenick	2016-12-12 06:50:08
14	Jaclyn81	2017-02-06 23:29:16
21	Rocio33	2017-01-23 11:51:15
24	Maxwell.Halvorson	2017-04-18 02:32:44
25	Tierra.Trantow	2016-10-03 12:49:21
34	Pearl7	2016-07-08 21:42:01
36	Ollie_Ledner37	2016-08-04 15:42:20
41	Mckenna17	2016-07-17 17:25:45
45	David.Osinski47	2017-02-05 21:23:37
49	Morgan.Kassulke	2016-10-30 12:42:31
53	Linnea59	2017-02-07 07:49:34
54	Duane60	2016-12-21 04:43:38
57	Julien_Schmidt	2017-02-02 23:12:48
66	Mike.Auer39	2016-07-01 17:36:15
68	Franco_Keebler64	2016-11-13 20:09:27
71	Nia_Haag	2016-05-14 15:38:50
74	Hulda.Macejkovic	2017-01-25 17:17:28
75	Leslie67	2016-09-21 05:14:01
76	Janelle.Nikolaus81	2016-07-21 09:26:09
80	Darby_Herzog	2016-05-06 00:14:21
81	Esther.Zulauf61	2017-01-14 17:02:34
83	Bartholome.Bernhard	2016-11-06 02:31:23
89	Jessyca_West	2016-09-14 23:47:05
90	Esmeralda.Mraz57	2017-03-03 11:52:27
91	Bethany20	2016-06-03 23:31:53

The users who have never posted a single photo on Instagram as per database used.

Q3. Marketing team started a contest and the user who gets the most likes on a single photo will win the contest now they wish to declare the winner.

```
SELECT * FROM users WHERE id = (SELECT user_id FROM insta.photos where
id = (SELECT photo_id FROM insta.likes GROUP BY photo_id ORDER BY
COUNT(photo_id) DESC limit 1))
```

id	username	created_at
52	Zack_Kemmer93	2017-01-01 05:58:22

Zack Kemmer is the winner of the contest

Other Approach:
Approach 1
Step 1:

```
SELECT COUNT(photo_id) AS Likes, photo_id FROM likes GROUP BY photo_id
ORDER BY Likes DESC limit 5
```

Likes	photo_id
48	145
43	127
43	182
42	123
41	61

Step 2:

```
SELECT * FROM users WHERE id = (SELECT user_id FROM photos where id =
145)
```

id	username	created_at
52	Zack_Kemmer93	2017-01-01 05:58:22

Approach 2 (via JOIN)
Step 1:

```
SELECT COUNT(l.photo_id) AS Likes, l.photo_id, p.user_id FROM likes l
JOIN photos p ON l.photo_id = p.id
GROUP BY photo_id ORDER BY Likes DESC LIMIT 1
```

Likes	photo_id	user_id
48	145	52

Step 2:

```
SELECT * FROM users WHERE id = 52
```

id	username	created_at
52	Zack_Kemmer93	2017-01-01 05:58:22

Q4. A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform?

```
SELECT t.tag_name, COUNT(pt.tag_id) AS tag_usage
FROM tags t
INNER JOIN photo_tags pt ON pt.tag_id = t.id
GROUP BY pt.tag_id
ORDER BY tag_usage DESC
LIMIT 5
```

tag_name	tag_usage
smile	59
beach	42
party	39
fun	38
concert	24

The top 5 most commonly used hashtags on the platform are smile, beach, party, fun, concert.

Other Approach:
Step 1:

```
SELECT COUNT(tag_id) AS Tags, tag_id FROM photo_tags GROUP BY tag_id
ORDER BY Tags DESC limit 5
```

Tags	tag_id
59	21
42	20
39	17
38	13
24	18

Step 2:

```
SELECT id, tag_name FROM tags WHERE id IN (21,20,17,13,18)
```

id	tag_name
13	fun
17	party
18	concert
20	beach
21	smile

Q5. Marketing team wants to know, which day would be the best day to launch Ads?

```
SELECT DAYNAME(created_at) AS day, COUNT(*) AS no_of_regs FROM users
GROUP BY day ORDER BY no_of_regs DESC
```

day	no_of_regs
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Sunday and Thursday of the week most users have register thus would be the best ways to launch campaign. If not Friday should be the last choice.

These where some of the marketing team challenges. Now our investors wants to know if Instagram is performing well and is not becoming redundant like Facebook, they want to assess the app on the following grounds

Q6. Investors wants to know the User Engagement i.e. the users still as active and post on Instagram or they are making fewer posts?

```
SELECT AVG(Posts) AS total_post_per_user FROM (SELECT COUNT(user_id) AS
Posts FROM photos GROUP BY user_id) AS t
```

total_post_per_user
3.473

3.5 times average user posts on Instagram.

```
SELECT COUNT(p.id)/(SELECT COUNT(u.id) FROM users u) AS absolute_avg
FROM photos p
```

absolute_avg
2.57

Absolute Average = total photos / total users.

Q7. Investors wants to know if the platform is crowded with fake and dummy accounts?

```
SELECT COUNT(user_id) AS likes_per_user, username
FROM users u
JOIN likes l ON l.user_id = u.id
GROUP BY user_id
HAVING likes_per_user=(SELECT COUNT(id) FROM photos)
ORDER BY likes_per_user DESC
```

likes_per_user	username
257	Aniya_Hackett
257	Jaclyn81
257	Rocio33
257	Maxwell.Halvorson
257	Ollie_Ledner37
257	Mckenna17
257	Duane60
257	Julien_Schmidt
257	Mike.Auer39
257	Nia_Haag
257	Leslie67
257	Janelle.Nikolaus81
257	Bethany20

These users (bots) who have liked every single photo on the site since any normal user would not be able to do this, thus these are fake accounts.

Other Approach:

Step 1:

```
SELECT COUNT(id) AS total_photos FROM photos
```

total_photos
257

Step 2:

```
SELECT COUNT(user_id) AS likes_per_user, user_id FROM likes GROUP BY
user_id HAVING likes_per_user='257' ORDER BY likes_per_user DESC
```

likes_per_user	user_id
257	5

257	14
257	21
257	24
257	36
257	41
257	54
257	57
257	66
257	71
257	75
257	76
257	91

Step 3:

SELECT * FROM insta.users WHERE id IN (5,14,21,24,36,41,54,57,66,71,75,76,91)

id	username	created_at
5	Aniya_Hackett	2016-12-07 01:04:39
14	Jaclyn81	2017-02-06 23:29:16
21	Rocio33	2017-01-23 11:51:15
24	Maxwell.Halvorson	2017-04-18 02:32:44
36	Ollie_Ledner37	2016-08-04 15:42:20
41	Mckenna17	2016-07-17 17:25:45
54	Duane60	2016-12-21 04:43:38
57	Julien_Schmidt	2017-02-02 23:12:48
66	Mike.Auer39	2016-07-01 17:36:15
71	Nia_Haag	2016-05-14 15:38:50
75	Leslie67	2016-09-21 05:14:01
76	Janelle.Nikolaus81	2016-07-21 09:26:09
91	Bethany20	2016-06-03 23:31:53