PROJECT 2

INSTAGRAM USER ANALYTICS

SQL FUNDAMENTALS

Analyzed by:

Parag Jyoti Nath

PROJECT OVERVIEW

- To-do :
  - Analyze user interactions and engagement with the Instagram app.
- Goal :
  - Extract valuable insights to help the business grow.
  - The marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience.

TECH-STACK USED

- In this project, SQL and MySQL Workbench is being used as the tool to analyze Instagram user data and answer questions posed by the management team. The insights will help the product manager and the rest of the team to make informed decisions about the future direction of the Instagram app.

PROJECT APPROACH

MARKETING ANALYSIS :

- Loyal User Reward
- Inactive User Engagement
- Contest Winner Declaration
- Hashtag Research
- Ad Campaign Launch

INVESTOR METRICS :

- User Engagement
- Bots & Fake Accounts

MARKETING ANALYSIS

1. LOYAL USER REWARD

- The marketing team wants to reward those who have been using the platform for the longest time.

  Task: Identify the five oldest users on Instagram from the provided database.

```
1   use ig_clone;
2
3 • select * from users
4   order by created_at asc
5   limit 5;
```

Result Grid | Filter Rows:

	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
*	NULL	NULL	NULL

2. INACTIVE USER ENGAGEMENT

- The team wants to encourage inactive users to start posting by sending them promotional emails.
  Task: Identify users who have never posted a single photo on Instagram.

```
 9 • select * from users
10   left join photos
11   on users.id = photos.user_id
12   where image_url is null;
```

26 rows returned on next page

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

id	username	created_at	id	image_url	user_id	created_dat
5	Aniya_Hackett	2016-12-07 01:04:39	NULL	NULL	NULL	NULL
7	Kasandra_Homenick	2016-12-12 06:50:08	NULL	NULL	NULL	NULL
14	Jaclyn81	2017-02-06 23:29:16	NULL	NULL	NULL	NULL
21	Rocio33	2017-01-23 11:51:15	NULL	NULL	NULL	NULL
24	Maxwell.Halvorson	2017-04-18 02:32:44	NULL	NULL	NULL	NULL
25	Tierra.Trantow	2016-10-03 12:49:21	NULL	NULL	NULL	NULL
34	Pearl7	2016-07-08 21:42:01	NULL	NULL	NULL	NULL
36	Ollie_Ledner37	2016-08-04 15:42:20	NULL	NULL	NULL	NULL
41	Mckenna17	2016-07-17 17:25:45	NULL	NULL	NULL	NULL
45	David.Osinski47	2017-02-05 21:23:37	NULL	NULL	NULL	NULL
49	Morgan.Kassulke	2016-10-30 12:42:31	NULL	NULL	NULL	NULL
53	Linnea59	2017-02-07 07:49:34	NULL	NULL	NULL	NULL
54	Duane60	2016-12-21 04:43:38	NULL	NULL	NULL	NULL
57	Julien_Schmidt	2017-02-02 23:12:48	NULL	NULL	NULL	NULL
66	Mike.Auer39	2016-07-01 17:36:15	NULL	NULL	NULL	NULL
68	Franco_Keebler64	2016-11-13 20:09:27	NULL	NULL	NULL	NULL
71	Nia_Haag	2016-05-14 15:38:50	NULL	NULL	NULL	NULL
74	Hulda.Macejkovic	2017-01-25 17:17:28	NULL	NULL	NULL	NULL
75	Leslie67	2016-09-21 05:14:01	NULL	NULL	NULL	NULL
76	Janelle.Nikolaus81	2016-07-21 09:26:09	NULL	NULL	NULL	NULL
80	Darby_Herzog	2016-05-06 00:14:21	NULL	NULL	NULL	NULL
81	Esther.Zulauf61	2017-01-14 17:02:34	NULL	NULL	NULL	NULL
83	Bartholome.Bernhard	2016-11-06 02:31:23	NULL	NULL	NULL	NULL
89	Jessyca_West	2016-09-14 23:47:05	NULL	NULL	NULL	NULL
90	Esmeralda.Mraz57	2017-03-03 11:52:27	NULL	NULL	NULL	NULL
91	Bethany20	2016-06-03 23:31:53	NULL	NULL	NULL	NULL

3. CONTEST WINNER DECLARATION

- The team has organized a contest where the user with the most likes on a single photo wins.
  Task: Determine the winner of the contest and provide their details to the team.

Step 1 - Find the photo ID with the most number of likes:

```
14 •  select photo_id, count(user_id) as likecount
15    from likes
16    group by photo_id
17    order by likecount desc;
```

Result Grid

photo_id	likecount
145	48
127	43
182	43
123	42
30	41

Step 2 - Use the previous query as a subquery to find the details of the user with the most likes on single photo.

```
19 • select users.id,  users.username, users.created_at,
20   photos.id as photo_id, photos.image_url,
21   photos.created_dat
22   from users
23   inner join photos on users.id = photos.user_id
24   where photos.id = (
25       select photo_id
26       from likes
27       group by photo_id
28       order by count(user_id) desc
29       limit 1
30       );
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

id	username	created_at	photo_id	image_url	created_dat
52	Zack_Kemmer93	2017-01-01 05:58:22	145	https://jarret.name	2024-07-04 12:32:40

4. HASHTAG RESEARCH

- A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
  Task: Identify and suggest the top five most commonly used hashtags on the platform.

```
33 •  select tags.tag_name, tags.id, count(tag_name) as tag_count
34    from tags
35    inner join photo_tags
36    on tags.id = photo_tags.tag_id
37    group by tag_name
38    order by tag_count desc;
```

Result Grid | Filter Rows:

tag_name	id	tag_count
smile	21	59
beach	20	42
party	17	39
fun	13	38
concert	18	24

5. AD CAMPAIGN LAUNCH

- The team wants to know the best day of the week to launch ads.
  Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

```
40 •  select dayname(created_at) as day_name,
41    count(dayname(created_at)) as accounts_created
42    from users
43    group by day_name
44    order by accounts_created desc;
```

Result Grid | Filter Rows:

day_name	accounts_created
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Insights :

1. Both Thursdays and Sundays are equally popular for most registrations on Instagram. The audience interest level is well balanced throughout the week. It is best to launch ad campaigns on both Thursdays and Sundays for peak audience engagement.

2. Also, Fridays are almost equally popular. Hence Fridays can also be used to launch ad campaigns to engage even more audience.

3. Afternoons on Sundays and evenings on Thursday and Fridays are likely to be peak time when users are free to use Instagram. Hence ad campaigns should be scheduled during these times for maximum engagement.

INVESTOR METRICS

1. USER ENGAGEMENT

- Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
  Task: Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

Step 1 - Find total posts per user.

```
51 •  select username, count(photos.user_id) as posts_per_user
52    from users
53    inner join photos
54    on users.id = photos.user_id
55    group by username;
```

Result Grid | Filter Rows:

username	posts_per_user
Kenton_Kirlin	5
Andre_Purdy85	4
Harley_Lind18	4
Arely_Bogan63	3
Travon.Waters	5
Tabitha_Schamberger11	4
Gus93	4
Presley_McClure	3
Justina.Gaylord27	5
Dereck65	4
Alexandro35	5
Billy52	4
Annalise.McKenzie16	4
Norbert_Carroll35	3

1. USER ENGAGEMENT

Step 2 - Use the previous subquery to find the average post per user.

```
57 •  select avg(post_count.posts_per_user) as avg_posts
58    from (
59        select username, count(photos.user_id) as posts_per_user
60        from users
61        inner join photos
62        on users.id = photos.user_id
63        group by username
64    ) as post_count;
```

Result Grid

	avg_post
▶	3.4730

1. USER ENGAGEMENT

Step 3 - Divide total posts by total users

```
66 •  select
67        (select count(*) from photos) / (select count(*) from users)
68        as post_per_user;
```

Result Grid

	post_per_user
▶	2.5700

Conclusion : As you can see, the average post per user is significantly different from what we obtained by dividing the total posts by the total users. Hence, we must always be careful to apply the correct way to use our data.

2. BOTS & FAKE ACCOUNTS

- Investors want to know if the platform is crowded with fake and dummy accounts. **Task**: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

Step 1 - Count the total photos liked by each user.

```
70 •  select user_id, count(*) as total_photos_liked
71    from likes
72    group by user_id;
```

Result Grid | Filter Rows:

user_id	total_photos_liked
2	94
3	79
4	93
5	257
6	82
8	79
9	85
10	87
11	89
12	77
13	93
14	257
15	84
16	103
17	78

- Step 2 - Use the previous query as a subquery to check if any user has liked exactly the same number of photos as the total photos present on Instagram, i.e. liked every single photo as any user can like each photo only once.

```
74 •  select *
75    from users
76    join (
77        select user_id, count(*) as total_photos_liked
78        from likes
79        group by user_id
80    ) likes on users.id = likes.user_id
81    where likes.total_photos_liked = (
82        select count(*) from photos
83    );
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

id	username	created_at	user_id	total_photos_liked
5	Aniya_Hackett	2016-12-07 01:04:39	5	257
14	Jaclyn81	2017-02-06 23:29:16	14	257
21	Rocio33	2017-01-23 11:51:15	21	257
24	Maxwell.Halvorson	2017-04-18 02:32:44	24	257
36	Ollie_Ledner37	2016-08-04 15:42:20	36	257
41	Mckenna17	2016-07-17 17:25:45	41	257
54	Duane60	2016-12-21 04:43:38	54	257
57	Julien_Schmidt	2017-02-02 23:12:48	57	257
66	Mike.Auer39	2016-07-01 17:36:15	66	257
71	Nia_Haag	2016-05-14 15:38:50	71	257
75	Leslie67	2016-09-21 05:14:01	75	257
76	Janelle.Nikolaus81	2016-07-21 09:26:09	76	257
91	Bethany20	2016-06-03 23:31:53	91	257

RESULTS

- Through these tasks, we are able to provide actionable insights to both the marketing team and investors as :
  - For the marketing team, identifying loyal users, engaging inactive users, declaring contest winners, researching hashtags, and determining the best days for ad campaigns allows us to find strategies that likely increases user engagement and revenue.
  - For the investors, understanding user engagement levels and identifying potential bot accounts has shown the transparency of the platform and allows us to focus on finding strategies to minimize bot activity as it can become a KPI in the future.
- The Instagram User Analytics has shown us the importance of data-driven decision-making to help a business grow. The insights gained from these tasks have benefited the current stakeholders and has set a foundation for business strategy and investment decisions in the future.

THANK YOU