Instagram User Analytics

Project Description: This project is about user analysis. I was asked to perform user analysis for Instagram and provide insights to product manager on the questions asked by the management team. First it was required to create the database using dataset provided. Then it was required to perform analysis using SQL to answer the below questions:

A) Marketing:

1. Rewarding Most Loyal Users: Find the 5 oldest users of the Instagram from the database provided
2. Remind Inactive Users to Start Posting: Find the users who have never posted a single photo on Instagram
3. Declaring Contest Winner: Identify the winner of the contest and provide their details to the team
4. Hashtag Researching: Identify and suggest the top 5 most commonly used hashtags on the platform
5. Launch AD Campaign: What day of the week do most users register on? Provide insights on when to schedule an ad campaign

B) Investor Metrics:

1. User Engagement: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users
2. Bots & Fake Accounts: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

Approach: First I gone through all the tables to know all the columns present in each table. Then I saw all the questions and thought of which tables could be used in each question. Then I wrote the sql queries for each question.

Tech-Stack Used: The software used for the project is MySQL workbench 8.0 CE. It is used to run the sql queries and get answers of each question.

Insights:

A) Marketing:

1. Rewarding Most Loyal Users:
   In this, I was required to find 5 oldest users of Instagram.

   Sql query:

   SELECT *
   FROM (
           SELECT
                   id, username, created_at,
                   RANK() OVER(order by created_at) as joining_rank
           FROM
                   users) as ranking
   WHERE
           joining_rank < 6;

   Output:

id	username	created_at	joining_rank
80	Darby_Herzog	2016-05-06 00:14:21	1
67	Emilio_Bernier52	2016-05-06 13:04:30	2
63	Elenor88	2016-05-08 01:30:41	3
95	Nicole71	2016-05-09 17:30:22	4
38	Jordyn.Jacobson2	2016-05-14 07:56:26	5

2. Remind Inactive Users to Start Posting:
   In this, I was required to find users who have never posted a single photo on Instagram.

   Sql query:

```sql
SELECT
    id, username
FROM
    users
WHERE
    id NOT IN (
            SELECT
        user_id
      FROM
         photos);
```

   Output:

id	username
5	Aniya_Hackett
7	Kasandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

3. Declaring Contest Winner:
   In this, I was required to find the details of the user who gets the most likes on a single photo.

   Sql query:

   ```
   CREATE VIEW likes_view AS
      (SELECT
          photo_id, COUNT(photo_id) AS no_of_likes
      FROM
          likes
      GROUP BY photo_id);

   SELECT
      photo_id, user_id, username, no_of_likes
   FROM
      users INNER JOIN
      (SELECT
          photo_id, user_id, no_of_likes
      FROM
          likes_view INNER JOIN
          photos ON likes_view.photo_id = photos.id
      WHERE
          no_of_likes = (SELECT
                  MAX(no_of_likes)
              FROM
                  likes_view)) AS inner_query ON users.id = inner_query.user_id;
   ```

   Output:

photo_id	user_id	username	no_of_likes
145	52	Zack_Kemmer93	48

4. Hashtag Researching: Identify and suggest the top 5 most commonly used hashtags on the platform

Sql query:

```
CREATE VIEW hashtags_view AS
   (SELECT
      tag_id, COUNT(tag_id) AS used_no
   FROM
      photo_tags
   GROUP BY tag_id);

SELECT *
FROM (
      SELECT
            tag_id, tag_name, used_no,
            RANK() OVER(order by used_no desc) as hashtag_rank
      FROM
            hashtags_view inner join
            tags on hashtags_view.tag_id = tags.id) as ranking
WHERE
      hashtag_rank < 6;
```

Output:

tag_id	tag_name	used_no	hashtag_rank
21	smile	59	1
20	beach	42	2
17	party	39	3
13	fun	38	4
5	food	24	5
11	lol	24	5
18	concert	24	5

5. Launch AD Campaign: What day of the week do most users register on?
   Provide insights on when to schedule an ad campaign

   Sql query:

   ```
   CREATE VIEW users_registered_view AS
      (SELECT
          DAYOFWEEK(created_at) as Day_of_week,
          COUNT(DAYOFWEEK(created_at)) AS Users_registered
      FROM
          users
      GROUP BY DAYOFWEEK(created_at)
      ORDER BY Day_of_week);

   SELECT
      Day_of_week, Users_registered AS Most_users_registerd
   FROM
      users_registered_view
   WHERE
      Users_registered = (SELECT
           MAX(Users_registered)
         FROM
           users_registered_view);
   ```

   Output:

Day_of_week	Most_users_registerd
1	16
5	16

B) Investor Metrics:

1. User Engagement: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users

   Sql query:

   ```
   CREATE VIEW posts_view AS
      (SELECT
         user_id, COUNT(id) AS no_of_posts
      FROM
         photos
      GROUP BY user_id
      ORDER BY user_id);

   SELECT
         id, username, no_of_posts
      FROM
         users
            LEFT JOIN
         posts_view ON users.id = posts_view.user_id;

   CREATE VIEW posts_per_user AS
      (SELECT
         id, username, no_of_posts
      FROM
         users
            LEFT JOIN
         posts_view ON users.id = posts_view.user_id);

   SELECT * FROM posts_per_user;

   SELECT
      COUNT(id) AS no_of_users,
   ```

SUM(no_of_posts) AS total_no_of_posts,
    SUM(no_of_posts) / COUNT(id) AS `total number of photos on Instagram/total number of users`
FROM
    posts_per_user;

Output:

id	username	no_of_posts
1	Kenton_Kirlin	5
2	Andre_Purdy85	4
3	Harley_Lind18	4
4	Arely_Bogan63	3
5	Aniya_Hackett	NULL
6	Travon.Waters	5
7	Kasandra_Homenick	NULL
8	Tabitha_Schamberger11	4
9	Gus93	4
10	Presley_McClure	3
11	Justina.Gaylord27	5
12	Dereck65	4
13	Alexandro35	5
14	Jaclyn81	NULL
15	Billy52	4
16	Annalise.McKenzie16	4
17	Norbert_Carroll35	3
18	Odessa2	1
19	Hailee26	2
20	Delpha.Kihn	1
21	Rocio33	NULL
22	Kenneth64	1
23	Eveline95	12
24	Maxwell.Halvorson	NULL
25	Tierra.Trantow	NULL

id	username	no_of_posts
26	Josianne.Friesen	5
27	Darwin29	1
28	Dario77	4
29	Jaime53	8
30	Kaley9	2
31	Aiyana_Hoeger	1
32	Irwin.Larson	4
33	Yvette.Gottlieb91	5
34	Pearl7	NULL
35	Lennie_Hartmann40	2
36	Ollie_Ledner37	NULL
37	Yazmin_Mills95	1
38	Jordyn.Jacobson2	2
39	Kelsi26	1
40	Rafael.Hickle2	1
41	Mckenna17	NULL
42	Maya.Farrell	3
43	Janet.Armstrong	5
44	Seth46	4
45	David.Osinski47	NULL
46	Malinda_Streich	4
47	Harrison.Beatty50	5
48	Granville_Kutch	1
49	Morgan.Kassulke	NULL
50	Gerard79	3
51	Mariano_Koch3	5
52	Zack_Kemmer93	5
53	Linnea59	NULL
54	Duane60	NULL
55	Meggie_Doyle	1
56	Peter.Stehr0	1
57	Julien_Schmidt	NULL
58	Aurelie71	8
59	Cesar93	10
60	Sam52	2
61	Jayson65	1
62	Ressie_Stanton46	2
63	Elenor88	4
64	Florence99	5
65	Adelle96	5
66	Mike.Auer39	NULL
67	Emilio_Bernier52	3
68	Franco_Keebler64	NULL
69	Karley_Bosco	1
70	Erick5	1
71	Nia_Haag	NULL
72	Kathryn80	5
73	Jaylan.Lakin	1
74	Hulda.Macejkovic	NULL
75	Leslie67	NULL

76	Janelle.Nikolaus81	NULL
77	Donald.Fritsch	6
78	Colten.Harris76	5
79	Katarina.Dibbert	1
80	Darby_Herzog	NULL
81	Esther.Zulauf61	NULL
82	Aracely.Johnston98	2
83	Bartholome.Bernhard	NULL
84	Alysa22	2
85	Milford_Gleichner42	2
86	Delfina_VonRueden68	9
87	Rick29	4
88	Clint27	11
89	Jessyca_West	NULL
90	Esmeralda.Mraz57	NULL
91	Bethany20	NULL
92	Frederik_Rice	3
93	Willie_Leuschke	2
94	Damon35	1
95	Nicole71	2
96	Keenan.Schamberger60	3
97	Tomas.Beatty93	2
98	Imani_Nicolas17	1
99	Alek_Watsica	3
100	Javonte83	2

no_of_users	total_no_of_posts	total number of photos on Instagram/total number of users
100	257	2.5700

2. Bots & Fake Accounts: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this)

Sql query:

SELECT
   user_id, username
FROM
   (SELECT
      user_id, username, COUNT(photo_id) AS posts_liked
   FROM
      likes
   INNER JOIN users ON likes.user_id = users.id
   GROUP BY user_id) AS likes_no
WHERE
   posts_liked = (SELECT
         COUNT(DISTINCT id) AS total_posts
      FROM
         photos);

Output:

user_id	username
5	Aniya_Hackett
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
36	Ollie_Ledner37
41	Mckenna17
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
71	Nia_Haag
75	Leslie67
76	Janelle.Nikolaus81
91	Bethany20

Results:

A) Marketing:

1. Rewarding Most Loyal Users:
   Darby_Herzog, Emilio_Bernier52, Elenor88, Nicole71, Jordyn.Jacobson2
   with user ids 80, 67, 63, 95, 38 respectively are five oldest users of
   Instagram.

2. Remind Inactive Users to Start Posting:
   Aniya_Hackett, Kasandra_Homenick, Jaclyn81, Rocio33,
   Maxwell.Halvorson, Tierra.Trantow, Pearl7, Ollie_Ledner37, Mckenna17
   David.Osinski47, Morgan.Kassulke, Linnea59, Duane60, Julien_Schmidt,
   Mike.Auer39, Franco_Keebler64, Nia_Haag, Hulda.Macejkovic, Leslie67,
   Janelle.Nikolaus81, Darby_Herzog, Esther.Zulauf61,
   Bartholome.Bernhard, Jessyca_West, Esmeralda.Mraz57, Bethany20 are
   the users who have never posted a single photo on Instagram

3. Declaring Contest Winner:
   Zack_Kemmer93 with user id 52 is winner of the contest by getting most
   likes on a single photo.

4. Hashtag Researching:
   smile, beach, party, fun, food are the five most used hashtags on the
   platform. Also, lol and concert are used same number of times as food.

5. Launch AD Campaign:
   Most users register on day 1 and day 5 of week i.e., Sunday and
   Thursday. According to me, ad campaign could be launched on the day
   when most number of likes are created which mean on the day when
   posts engagement are highest.

B) Investor Metrics:

1. User Engagement:
   Total number of photos on Instagram/total number of users = 2.57

2. Bots & Fake Accounts:
   Aniya_Hackett, Jaclyn81, Rocio33, Maxwell.Halvorson, Ollie_Ledner37, Mckenna17, Duane60, Julien_Schmidt, Mike.Auer39, Nia_Haag, Leslie67 Janelle.Nikolaus81, Bethany20 have liked each and every photo on platform.