INSTAGRAM USER ANALYTICS

PROJECT 2

Introduction:

Understanding the characteristics of social media platforms is crucial for businesses, influencers, and content providers alike in the digital age, since these platforms dominate the molding of consumer habits and societal narratives. With over a billion active users globally, Instagram is a platform that dominates visual content among the others.

Through the Instagram Analytics project, we set out on a voyage of discovery and analysis with the goal of understanding the complex web of relationships and patterns within Instagram's ecosystem. Our project, which is based on SQL and data analytics, aims to extract useful information from the massive amount of data that is created on this platform every day.

In this project, we are asked to work on certain tasks to get the outcome as per required by the company and the data provided to us.

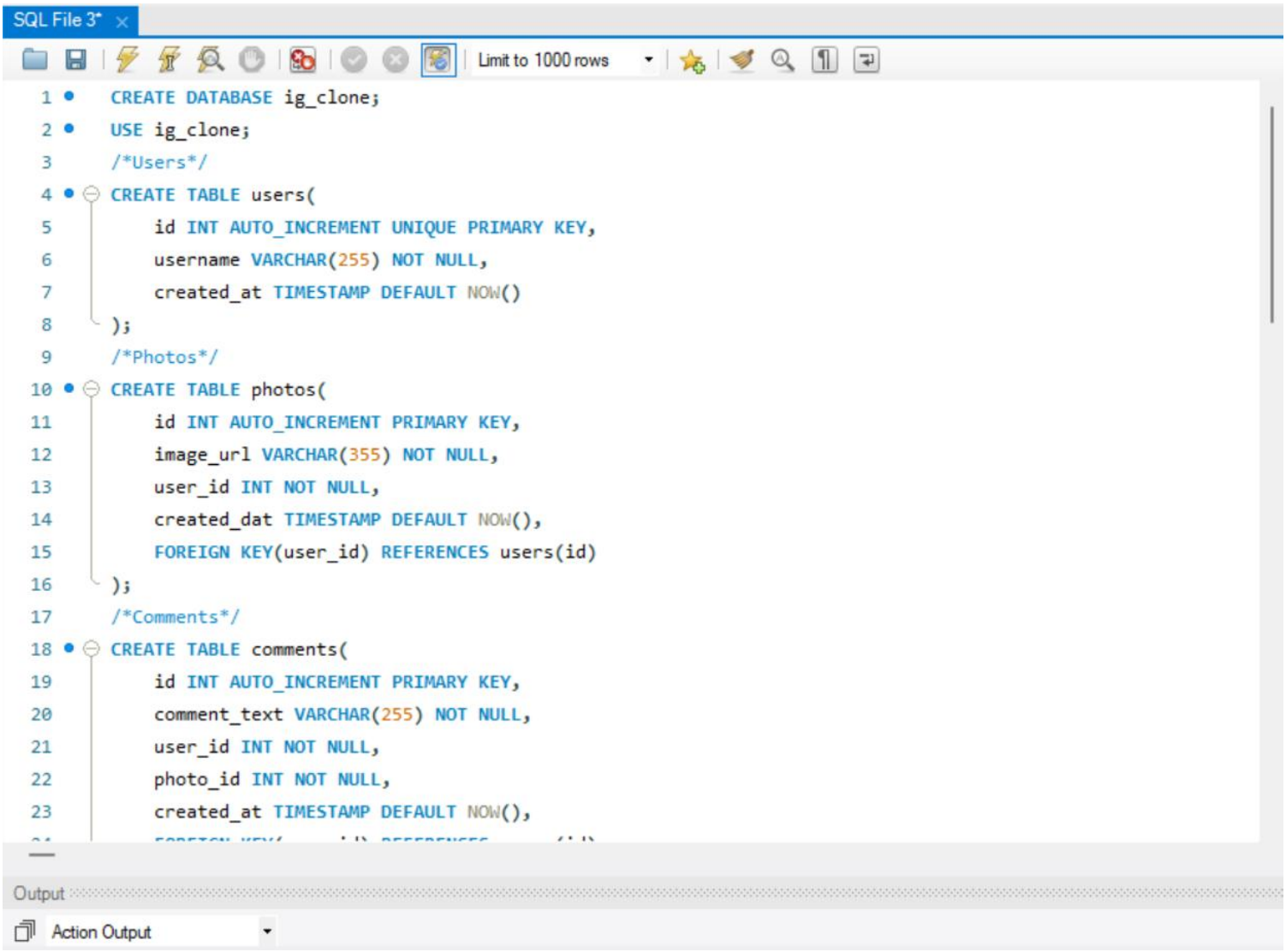
Software used:

MYSQL Workbench.

Setting up the workbench:

We are provided with a dataset which has all the required information for the project to be executed smoothly. Henceforth,we have to first copy the dataset from the word file and then paste it in the workbench.
For that, we first connect to the MYSQL Server by entering the passcode and then click on the new sql query page. There we can paste the copied information from the dataset.
Execute the query and look for errors. If found any, look for the solutions and resolve them. Once error free,we can proceed further.

Some screenshots are attached for reference.(no errors were detected):



SQL File 3*

Limit to 1000 rows

```
24      FOREIGN KEY(user_id) REFERENCES users(id),
25      FOREIGN KEY(photo_id) REFERENCES photos(id)
26  );
27  /*Likes*/
28  CREATE TABLE likes(
29      user_id INT NOT NULL,
30      photo_id INT NOT NULL,
31      created_at TIMESTAMP DEFAULT NOW(),
32      FOREIGN KEY(user_id) REFERENCES users(id),
33      FOREIGN KEY(photo_id) REFERENCES photos(id),
34      PRIMARY KEY(user_id,photo_id)
35  );
36  /*follows*/
37  CREATE TABLE follows(
38      follower_id INT NOT NULL,
39      followee_id INT NOT NULL,
40      created_at TIMESTAMP DEFAULT NOW(),
41      FOREIGN KEY (follower_id) REFERENCES users(id),
42      FOREIGN KEY (followee_id) REFERENCES users(id),
43      PRIMARY KEY(follower_id,followee_id)
44  );
45  /*Tags*/
46  CREATE TABLE tags(
```

Output

Action Output

SQL File 3*

Limit to 1000 rows

```
47      id INTEGER AUTO_INCREMENT PRIMARY KEY,
48      tag_name VARCHAR(255) UNIQUE NOT NULL,
49      created_at TIMESTAMP DEFAULT NOW()
50  );
51  /*junction table: Photos - Tags*/
52  CREATE TABLE photo_tags(
53      photo_id INT NOT NULL,
54      tag_id INT NOT NULL,
55      FOREIGN KEY(photo_id) REFERENCES photos(id),
56      FOREIGN KEY(tag_id) REFERENCES tags(id),
57      PRIMARY KEY(photo_id,tag_id)
58  );
59  INSERT INTO users (username, created_at) VALUES ('Kenton_Kirlin', '2017-02-16 18:22:10.846'), ('Andre_Purd
60
61  INSERT INTO photos(image_url, user_id) VALUES ('http://elijah.biz', 1), ('https://shanon.org', 1), ('http:
62
63  INSERT INTO follows(follower_id, followee_id) VALUES (2, 1), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (2, 8
64
65  INSERT INTO comments(comment_text, user_id, photo_id) VALUES ('unde at dolorem', 2, 1), ('quae ea ducimus'
66
67  INSERT INTO likes(user_id,photo_id) VALUES (2, 1), (5, 1), (9, 1), (10, 1), (11, 1), (14, 1), (19, 1), (21
68
69  INSERT INTO tags(tag_name) VALUES ('sunset'), ('photography'), ('sunrise'), ('landscape'), ('food'), ('foo
```

Output

Action Output

```
71  INSERT INTO photo_tags(photo_id, tag_id) VALUES (1, 18), (1, 17), (1, 21), (1, 13), (1, 19), (2, 4), (2, 3
72
73
74
75
```

Output

Action Output

#	Time	Action	Message	Duration / Fetch
1	10:03:04	CREATE DATABASE ig_clone	1 row(s) affected	0.016 sec
2	10:03:04	USE ig_clone	0 row(s) affected	0.000 sec
3	10:03:04	CREATE TABLE users(id INT AUTO_INCREMENT UNIQUE PRIMARY KEY, usema...	0 row(s) affected	0.063 sec
4	10:03:04	CREATE TABLE photos(id I	row(s) affected	0.078 sec
5	10:03:04	CREATE TABLE comments(i	row(s) affected	0.063 sec
6	10:03:04	CREATE TABLE likes(user_i	row(s) affected	0.062 sec
7	10:03:04	CREATE TABLE follows(follc	row(s) affected	0.047 sec
8	10:03:04	CREATE TABLE tags(id INT	row(s) affected	0.046 sec
9	10:03:04	CREATE TABLE photo_tags(photo_id INT NOT NULL, tag_id INT NOT NULL, FORE...	0 row(s) affected	0.063 sec
10	10:03:04	INSERT INTO users (username, created_at) VALUES ('Kenton_Kirlin', '2017-02-16 18:...	100 row(s) affected Records: 100 Duplicates: 0 Warnings: 0	0.015 sec
11	10:03:04	INSERT INTO photos(image_url, user_id) VALUES ('http://elijah.biz', 1), ('https://shan...	257 row(s) affected Records: 257 Duplicates: 0 Warnings: 0	0.016 sec
12	10:03:04		7623 row(s) affected Records: 7623 Duplicates: 0 Warnings: 0	0.281 sec
13	10:03:04		7488 row(s) affected Records: 7488 Duplicates: 0 Warnings: 0	0.359 sec
14	10:03:05		8782 row(s) affected Records: 8782 Duplicates: 0 Warnings: 0	0.343 sec
15	10:03:05	INSERT INTO tags(tag_name) VALUES ('sunset'), ('photography'), ('sunrise'), ('landsca...	21 row(s) affected Records: 21 Duplicates: 0 Warnings: 0	0.016 sec
16	10:03:05	INSERT INTO photo_tags(photo_id, tag_id) VALUES (1, 18), (1, 17), (1, 21), (1, 13), (1...	501 row(s) affected Records: 501 Duplicates: 0 Warnings: 0	0.016 sec

CREATE TABLE photos(
id INT AUTO_INCREMENT PRIMARY KEY,
image_url VARCHAR(355) NOT NULL,
user_id INT NOT NULL,
created_dat TIMESTAMP DEFAULT NOW(),
FOREIGN KEY(user_id) REFERENCES users(id)
)

PROVIDED TASK:

- A) Marketing Analysis:

- Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
  Your Task: Identify the five oldest users on Instagram from the provided database.
- Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.
  Your Task: Identify users who have never posted a single photo on Instagram.
- Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.
  Your Task: Determine the winner of the contest and provide their details to the team.
- Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
  Your Task: Identify and suggest the top five most commonly used hashtags on the platform.
- Ad Campaign Launch: The team wants to know the best day of the week to launch ads.
  Your Task: Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.
- B) Investor Metrics:
- User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.
  Your 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.
- Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.
  Your Task: Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

SOLUTION:

1. Loyal User Reward: For this task,we have to display the name of the users who have been using the platform for the longest time now.

```
72
73 •    select username,created_at,
74      rank() over(order by created_at) as oldest_customers
75      from users
76      limit 5;
```

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

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

Output: The above output lists the name of the oldest users and the date when they created their account,also ranking them as per their account creation date.This information would be useful for the task assigned as we can reward them for being the most LOYAL USERS.

2. Inactive User Engagement: This task requires us to list out those users who have not been active on their account for the longest time.

SQL File 3*

```
78  select username,created_dat
79  from users
80  left join photos
81  on users.id=photos.user_id
82  where photos.id is null;
```

Result Grid | Filter Rows:

username	created_dat
Aniya_Hackett	NULL
Kasandra_Homenick	NULL
Jadyn81	NULL
Rocio33	NULL
Maxwell.Halvorson	NULL
Tierra.Trantow	NULL
Pearl7	NULL
Ollie_Ledner37	NULL
Mckenna17	NULL
David.Osinski47	NULL
Morgan.Kassulke	NULL
Linnea59	NULL
Duane60	NULL
Julien_Schmidt	NULL
Mike.Auer39	NULL
Franco_Keebler64	NULL
Nia_Haag	NULL
Hulda.Macejkovic	NULL
Leslie67	NULL
Janelle.Nikolaus81	NULL

Result 2

Darby_Herzog	NULL
Esther.Zulauf61	NULL
Bartholome.Bernhard	NULL
Jessyca_West	NULL
Esmeralda.Mraz57	NULL
Bethany20	NULL

Output: The above list displays the names of those account holders who have not been active. They have not posted a single photo on their handle and hence the company shall send them emails requesting them to start posting photos.

3. Contest Winner Declaration: For this task, we have to organize a contest wherein the user with the most likes on his post wins.

```
84 •  select likes.photo_id, users.username, photos.image_url,
85    count(*) as total_likes
86    from photos
87    inner join likes
88    on likes.photo_id=photos.id
89    inner join users
90    on photos.user_id=users.id
91    group by photo_id
92    order by total_likes desc
93    limit 1;
```

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

photo_id	username	image_url	total_likes
145	Zack_Kemmer93	https://jarret.name	48

Output: The above output shows the name of the person who has the most likes on his post during the contest along with the link to the post and the number of likes(48) the post has.

4. Hashtag Research: For this task, we have to search for the current most trendy hashtags so that a partner brand may use them to increase their post reach.

```
94
95 •  select tags.tag_name,
96     count(*) as popular_tags
97     from photo_tags
98     left join tags
99     on photo_tags.tag_id=tags.id
100    group by tags.id
101    order by popular_tags desc
102    limit 5;
103
104
```

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

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

Output: The above output list shows the most popular hashtags along with the number of times they have been used(for eg, smile has been used 59 times).This will help the partner brand to accordingly increase their reach by using the most suitable hashtags for their posts.

5. Ad Campaign Launch: For this task, the team wants to know the day of the week when the post reach is high.This is so that they can begin their post campaign on that particular day.

```
103
104 •  select dayname(created_at) as Day,
105    count(*) as total_registers
106    from users
107    group by Day
108    order by total_registers desc
```

Result Grid | Filter Rows: | Export:

Day	total_registers
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Output: The output shows the days of the week and the number of reach it has on that day. For eg, as is been shown, Thursday and Sunday have a reach of 16.This information may be used by the team to launch their campaign on the day when the reach is high.

INVESTOR METRICS TASK:

1. User Engagement: For this task,we have to figure out if users are still active on their accounts and posting photos. The investor wants to know the average user is to post ratio.

```
110 •  select
111    (select count(*) from photos)
112    /
113    (select count(*) from users)
114    as average_posts;
115
```

Result Grid | Filter Rows:

average_posts
2.5700

Output: The output shows the average post per user on instagram which turns out to be 2.57. This result can be shown to the investor and he can derive the necessary information.

2. Bots & Fake Accounts: For this task,we have to derive the information whether the platform has more number of bots and dummy accounts.

```
116 •  select users.id,users.username, count(*) as total_likes
117    from users
118    join likes
119    on likes.user_id=users.id
120    group by likes.user_id
121    having total_likes=(select count(*) from photos as total_photos);
```

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

id	username	total_likes
5	Aniya_Hackett	257
14	Jadyn81	257
21	Rocio33	257
24	Maxwell.Halvorson	257
36	Ollie_Ledner37	257
41	Mckenna17	257
54	Duane60	257
57	Julien_Schmidt	257
66	Mike.Auer39	257
71	Nia_Haag	257
75	Leslie67	257
76	Janelle.Nikolaus81	257
91	Bethany20	257

Output: As per the above output, we can see names of all those accounts who have liked every post on instagram which is not possible for a normal user. Hence,all the above names listed are bots.

CONCLUSION:

We have successfully completed all the required task with the output attached as per the requirement.