instagram data analysis

By-Ankit Gaurav

PROJECT DESCRIPTION

GENERAL OVERVIEW

Data set have Details related to Instagram users and their activity. I am using SQL to find some results related to user activity such as how many post at an average user touch what is total number of photos uploaded per user et cetera. I have used SQL workbench on MacBook to come to these conclusions.

SQL version used 9.2.0

Loyal user reward

Task

Find five oldest users

SQL query use

```
USE ig_clone;
SELECT *
FROM users;
SELECT *
FROM USERS
ORDER BY created_at ASC
LIMIT 5;
```

Result

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

Insights

These are the five person which of the oldest account on the platform. I have use SQL query in which I have ordered the ascending format, meaning, the date will come first, hence, it will be the oldest, and then I have used limit five that means it will not displaymore than five

Inactive user Engagement

Task

user that not posted a single photo

SQL query use

```
SELECT users.username,users.id
FROM users
LEFT OUTER JOIN photos
ON users.id=
photos.user_id
Where photos.id is null;
;
```

Result

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

Insights

In this, I needed to find the user that have not posted a single photo for this. I joined user table with photo table are you left? Join meaning user table is on the left. Hence all the ID from user will come but from photo table only those will come which are in the photos as well as user table. So user ID that have not posted a single photo gets none in the joint table as all of them are displaced.

Contest winner declaration

Task

Find user with most like on a single photo

SQL query use

```
select *
from users
where id =(
select user_id
from photos
where id=(
select photo_id
from likes
group by photo_id
order by count(*) desc limit 1));
```

Result

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

Insights

In this, I needed to find the user which has got most like on a single photo so I have used listed query here. First, I got the photo ID of the photo which has got maximum like from like table and then I am match that ID to photo table ID to get the user ID which have created that photo and then I have taken the user ID to find the name by equating it with ID from user table

Hashtag Research

Task

Find the top five, most used hashtag

SQL query use

```
SELECT p.tag_id,count(p.tag_id),t.tag_name
from tags t
join photo_tags p
on t.id = p.tag_id
group by p.tag_id
order by count(p.tag_id) DESC
LIMIT 5;
```

Result

tag_id	count(p.tag_...	tag_name
21	59	smile
20	42	beach
17	39	party
13	38	fun
18	24	concert

Insights

I have used to join to join photo tag and tag table, then have grouped them by tag ID from photo tag to get the top five most hashtags which are given in the table which can in marketing to make the ad viral.

Ad campaign launch

Task

Determine the day on which most user register
to the platform

SQL query use

```
select dayname(created_at) as weekday,count(*)
from users
group by weekday
order by count(*) DESC LIMIT 7;
```

Result

weekday	count...
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Insights

I have used the day name to get the week name for each date and then used to group them and then limit seven. I could have also printed limit one, but I thought as there are only seven things. I need to print it all all seven in descending order to check, which one is the highest Hair Tuesday and Sunday both have equal. But I think the best day would be Thursday as after Thursday Friday comes which is third highest so that would be very useful, but after Sunday Monday comes, which is not that high

User engagement

Task

Find the average number of post per user and photo per user

SQL query use

```
SET @TOTALPHOTO =(SELECT COUNT(*) from photos);
SET @TOTALUSER =(SELECT COUNT(*) from USERS);
SET @TOTALCOMMENT =(SELECT COUNT(*) from
COMMENTS);
SELECT
(@TOTALPHOTO+@TOTALCOMMENT)/@TOTALUSER AS
POST_PER_USER, @TOTALPHOTO/@TOTALUSER AS
PHOTO_PER_USER
```

Result

POST_PER_USER	PHOTO_PER_USER
77.4500	2.5700

Insights

Used variable to do calculations, first, I have considered comment and the photo posted by the user as a post together, so have calculated total comment and total photos and divided them by total user to get post per user and then total photos divided by total user to get photo user

Bots and fake account

Task

Find the user who have liked all the photos

SQL query use

```
SET @TOTAL =(SELECT COUNT(*) from photos);
SELECT USER_ID ,username , (COUNT(*)/@TOTAL)*100 as
likepercentage
FROM LIKES
join users
on users.id=likes.user_id
GROUP BY USER_ID
having likepercentage =100;
```

Result

USER_ID	username	likepercenta...
5	Aniya_Hackett	100.0000
14	Jaclyn81	100.0000
21	Rocio33	100.0000
24	Maxwell.Halvorson	100.0000
36	Ollie_Ledner37	100.0000
41	Mckenna17	100.0000
54	Duane60	100.0000
57	Julien_Schmidt	100.0000
66	Mike.Auer39	100.0000
71	Nia_Haag	100.0000
75	Leslie67	100.0000
76	Janelle.Nikolaus81	100.0000
91	Bethany20	100.0000

Insights

I use total variable to find total photos and then used count to get total likes by user, using group and then multiplying it by hundred to get total l0 percentage

Result

From this project, I have learnt installation of SQL and use of various queries to get to a certain output as per the requirement.

- I have used join
- I have also used variables
- I have also used nested query
- I have also used group by order by and limit
- I am also used analytical skill to come up with the day on which advertisement should be launched.

Thank you!