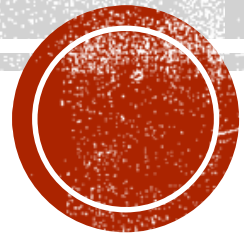
INSTAGRAM USER ANALYTICS

A SQL PROJECT

BY

SIDDHESH PANCHAL



INTRODUCTION

ABOUT THE DATA

This project is about using SQL (Structured Query Language) to perform Instagram user data analysis. The Instagram data was gotten from Trainity. It is a fictitious (imaginary) user data. This data consists of 7 tables; users, photos, follows, likes, tags, photo_tags, and comments tables.

TOOL USED

I made use of MySQL Workbench. As it is easy to use and easy to understand.

TABLES

Users

Field	Type	Null	Key
id	int	NO	PRI
username	varchar(255)	NO	
created_at	timestamp	YES	

Photos

Field	Type	Null	Key
id	int	NO	PRI
image_url	varchar(355)	NO	
user_id	int	NO	MUL
created_dat	timestamp	YES	

Comments

Field	Type	Null	Key
id	int	NO	PRI
comment_text	varchar(255)	NO	
user_id	int	NO	MUL
photo_id	int	NO	MUL
created_at	timestamp	YES	

Follows

Field	Type	Null	Key
follower_id	int	NO	PRI
followee_id	int	NO	PRI
created_at	timestamp	YES	

Likes

Field	Type	Null	Key
user_id	int	NO	PRI
photo_id	int	NO	PRI
created_at	timestamp	YES	

Photo_tags

Field	Type	Null	Key
photo_id	int	NO	PRI
tag_id	int	NO	PRI

Tags

Field	Type	Null	Key
id	int	NO	PRI
tag_name	varchar(255)	NO	UNI
created_at	timestamp	YES	

APPROACH

1. Find the 5 oldest users of the Instagram from the database provided
2. Find the users who have never posted a single photo on Instagram
3. Identify which user's photo has received maximum number of likes.
4. Identify and suggest the top 5 most commonly used hashtags on the platform
5. What day of the week do most users register on?
6. Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users
7. 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).

INSIGHTS

1. Find the 5 oldest users of the Instagram from the database provided

```
89
90  select * from users ORDER BY created_at ASC LIMIT 5;
91
92
93
```

Result Grid | Filter Rows: | Edit: | Export/Impc

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

Out of the 100 users who have been registered on Instagram there are the 5 oldest users of Instagram.

2. Find the users who have never posted a single photo on Instagram

```
99  SELECT *
100 FROM users
101 LEFT JOIN photos  ON users.id = photos.user_id
102 WHERE photos.user_id is null;
103
```

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

id	username	created_at	id	image_url	user_id	created_dat
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

Result 21

Out of the 100 users who have been registered on Instagram there are the users who have not even posted a single photo.

INSIGHTS

3.Identify which user's photo has received maximum number of likes.

```
109
110 •   SELECT
111         users.username,
112         photo_id,
113         COUNT(*) AS most_liked
114     FROM
115         likes
116         INNER JOIN users ON users.id=likes.user_id
117     GROUP BY
118         photo_id
119     ORDER BY
120         most_liked DESC
121     LIMIT
122         1;
```

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

username	photo_id	most_liked
Harley_Lind18	145	48

The 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. This user is the winner of the contest

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

```
95
96 •   SELECT
97         tags.tag_name,
98         COUNT(*) AS mostused
99     FROM
100        photo_tags
101        INNER JOIN tags ON tags.id=photo_tags.tag_id
102    GROUP BY
103        tag_id
104    ORDER BY
105        mostused DESC
106    LIMIT
107        5;
108
```

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

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

Out of all the hashtags used on Instagram these are the top 5 hashtags which are commonly used.

INSIGHTS

5. The team wants to know, which day would be the best day to launch Ads.What day of the week do most users register on?

```
151
152  SELECT MAX(DAYNAME(created_at)) as WeekDay, COUNT(*)
153  FROM users
154  GROUP BY DAYNAME(created_at)
155  ORDER BY count(*) DESC;
156
```

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

WeekDay	COUNT(*)
Thursday	16
Sunday	16
Friday	15
Tuesday	14
Monday	14
Wednesday	13
Saturday	12

Among all the days Thursday is the day when most users register on so Thursday is the best day to launch ads.

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

```
178
179  select avg(numpost)
180  from (
181  SELECT
182      users.username,
183      COUNT(*) AS numpost
184  FROM
185      photos
186      INNER JOIN users ON users.id=photos.user_id
187  GROUP BY
188      photos.user_id
189  ORDER BY
190      numpost DESC
191      ) as counts
```

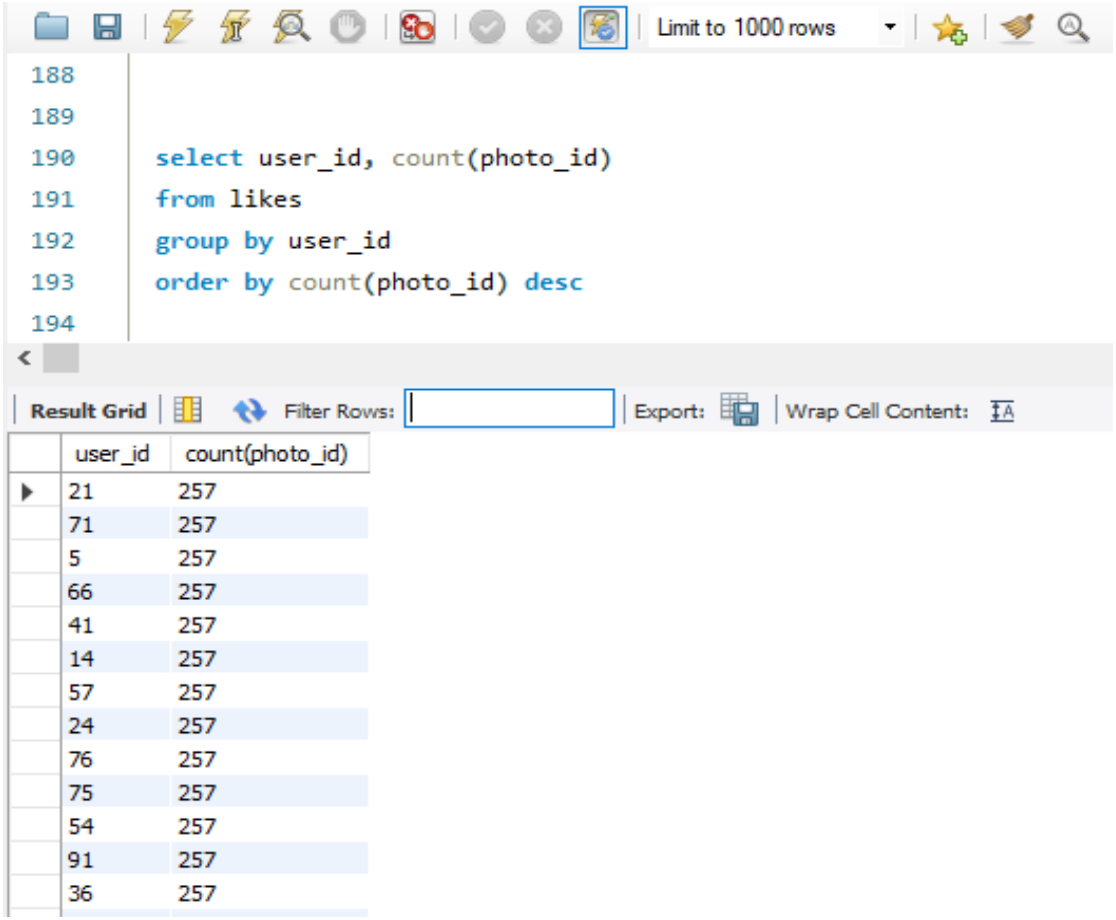
Result Grid | Filter Rows: | Export: | Wrap Cell Cont

avg(numpost)
3.4730

total number of photos on Instagram/total number of users=257/100. Therefore on an average 1 user post around 3 photos .

INSIGHTS

8.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).



user_id	count(photo_id)
21	257
71	257
5	257
66	257
41	257
14	257
57	257
24	257
76	257
75	257
54	257
91	257
36	257

There are total 257 post till now on Instagram and as we can see these are the user id of users who have liked every single photo on Instagram which is practically not possible . Therefore they are bots.

CONCLUSION

- By doing this project, I made use of group by clause.
- I made use of inner join all through because I was focusing majorly on the user table.
- Understanding the entire data took a lot of time as this was my first project.
- SQL might seem weird for starters, but it is really a straight-forward language if we understand the problem or question we are trying to answer.