INSTAGRAM USER ANALYTICS

SQL FUNDAMENTALS

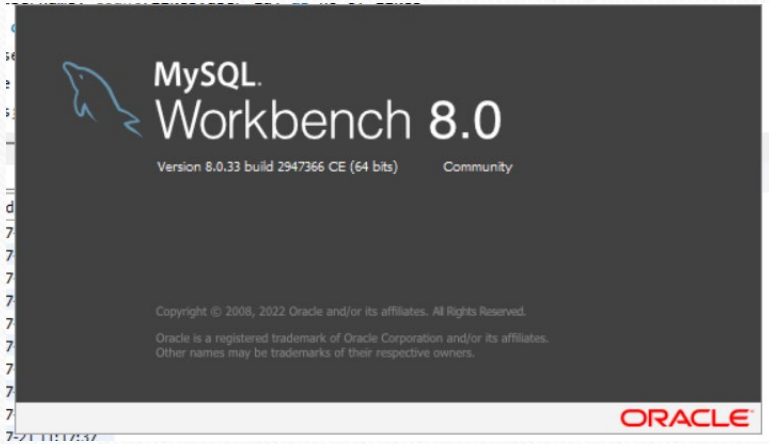
Project by: KIRAN KUMAR CHELLI

Project Description: "INSTAGRAM" is a social media platform where we can share our photos and videos to our friends and public. Now a days it is very much popular that in out of 10 persons 8 are having Instagram account. So in this project we are going to look into the data of Instagram, how can we analyze it, as we have the database of:

1. USERS
2. PHOTOS
3. COMMENTS
4. LIKES
5. FOLLOWS
6. TAGS
7. PHOTO TAGS

Approach: As per the dataset given, I have executed the queries. To find out solution for each question I wrote queries. As the above 7 tables contains the data according to their names, i.e., user table has username and created data and photos table has user id and how many photos were posted likewise all tables have their own specific attributes, which can be used in our data analytics.

Tech Stack Used: In this analytics process I used



to write queries. As per the tutorials I downloaded and installed it.

REPORT

A. Marketing

1. **Rewarding most loyal users**: people who have been using the platform for the longest time.

Task: find the 5 oldest users of Instagram from the database.

Solution: so among all the users we have find the top 5 only, so the query is

Query: select username,created_at from users order by created_at limit 5;

The result is:

USERNMAE	CREATED AT
Darby_Herzog	2016-05-06 00:14:21
Emilio_Bernier52	2016-05-06 13:04:30
Elenor88	2016-05-08 01:30:41
Nicole71	2016-05-09 17:30:22
Jordyn.Jacobson2	2016-05-14 07:56:26

2. **Remind inactive users to start posting**: by sending them promotional emails to post their 1st photo.

Task: find the users who have never posted a single photo on Instagram.

Solution: to find out who have not posted a single photo, we have look into attribute ”image_url” where the values are null. For this the following command was used:-

Aniya_Hackett
Bartholome.Bernhard
Bethany20
Darby_Herzog
David.Osinski47
Duane60
Esmeralda.Mraz57
Esther.Zulauf61
Franco_Keebler64
Hulda.Macejkovic
Jaclyn81
Janelle.Nikolaus81
Jessyca_West
Julien_Schmidt
Kasandra_Homenick
Leslie67
Linnea59
Maxwell.Halvorson
Mckenna17
Mike.Auer39
Morgan.Kassulke
Nia_Haag
Ollie_Ledner37
Pearl7
Rocio33
Tierra.Trantow

Query: select u.username from users u left join photos p on p.user_id=u.id where p.image_url is null order by u.username;

These are the users who have not posted a single photo from the beginning.

3. Declaring Contest Winner: 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.
Task: Identify the winner of the contest and provide their details to the team

Solution: to find out the winner we have to join the tables of user, photos and likes.

Query: select likes.photo_id,users.username, count(likes.user_id) as no_of_likesfrom likes inner join photos on likes.photo_id=photos.idinner join users on photos.user_id=users.id group by likes.photo_id,users.username order by no_of_likes desc limit 1;

The result is:

User Id	USER NAME	Total No.of likes
145	Zack_Kemmer93	48

4.Hashtag Researching: A partner brand wants to know, which hashtags to use in the post to reach the most people on the platform.
Task: Identify and suggest the top 5 most commonly used hashtags on the platform

Solution: To find the solution we must join the tables photo_tags and tags

Query: select t.tag_name,count(p.photo_id) as h_tag from photo_tags p inner join tags t on t.id=p.tag_id group by t.tag_name order by h_tag desc limit 5;

The result is

Tag name	Total tags
smile	59
beach	42
party	39
fun	38
concert	24

5. **Launch AD Campaign:** The team wants to know, which day would be the best day to launch ADs.
Task: What day of the week do most users register on? Provide insights on when to schedule an ad campaign

Solution: To find out which day of the week most no.of users were registered, we use

Query: select dayname(created_at) as day,count(username) as total_registered from users group by 1 order by 2 desc limit 3;

The result is:

Day	Total Registered
Thursday	16
Sunday	16
Friday	15

B. Investor Metrics

1.User Engagement: Are users still as active and post on Instagram or they are making fewer posts
Task: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram.

Solution: To find the solution we have to join user and photos table.
Query: with base as(select u.id as userid,count(p.id) as photoid from users u left join photos p on p.user_id=u.id group by u.id)select sum(photoid) as total_photos,count(userid) as total_users,sum(photoid)/count(userid) as average from base;

The result is:

Total Photos	Total Users	Average
257	100	2.5700

7.Bots & Fake Accounts: The investors want to know if the platform is crowded with fake and dummy accounts

Task: 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).

Solution: To find the solution we have to join user and like tables

Query: with base as(select u.username,count(l.photo_id) as liked from likes l inner join users u on u.id=l.user_idgroup by u.username) select username,liked from base where liked=(select count(*) from photos) order by username;

The result is:

Username	Liked
Aniya_Hackett	257
Bethany20	257
Duane60	257
Jaclyn81	257
Janelle.Nikolaus81	257
Julien_Schmidt	257
Leslie67	257
Maxwell.Halvorson	257
Mckenna17	257
Mike.Auer39	257
Nia_Haag	257
Ollie_Ledner37	257
Rocio33	257

Result

At starting stage this project was bit complicated to me. But I practiced again and again in SQL for better understanding the queries to get the desired result. Thank you for giving me this opportunity to explore myself to understand the Data Analytics.

Thank You