Instagram User Analytics

JUNE 27

INSTAGRAM
Authored by: ATTULAYA KUMAR SINGH

PROJECT REPORT

Project Description

I Performed a brief analysis on Instagram users to track how users engage and interact with the site/app in attempt to derive the business insights for marketing, product & development team

These insights will be used by the respective teams to launch new marketing campaign, decide on features to build for the app, track the success of the app by measuring user engagement and improve the experience altogether to help the company grow

"This Project is performed on the ig_clone database"

In this demo project I'm working with the product team and the project manager had asked me to provide insights on the question asked by the demo team

TABLE OF CONTENTS:

PROJECT REPORT 2

Project Description 2

APPROACH 4

TECH-STACK USED:.....5

SQL QUERY along with insights: 6

DATABASE used:ig_clone 6

Sql query used to perform analysis: 6

APPROACH

I started off with learning the concepts required for the projects and did the requirement analysis first then I looked into the database that what data do I have then accordingly I planned which table will be suitable for which question and then went through all the details of the question and planned which tables are to be joined in the process. I took a look into the schema and saw various answers were not directly available and then I had to create a formula for those particular requirements.my approach was simple, look the question perform the requirement analysis and then perform the analysis for that question and provide the insights.

TECH-STACK USED:

- For SQL I’m using MYSQL workbench 8.0
- For graphs I’m using Microsoft office excel spreadsheets (MS office 2019)
- For making report I’m using Microsoft office word (MS office 2019)

SQL QUERY along with insights:

DATABASE used:ig_clone

- **To find the most loyal users :**
  Sql query used to perform analysis:

```
select * from users
order by created_at
limit 5;
```

  here we selected those users who were the oldest on the platform in terms of usage and to find that we extracted the dates on which their accounts were created

id	username	created_at
80	Darby_Herzog	06-05-2016 00:14
67	Emilio_Bernier52	06-05-2016 13:04
63	Elenor88	08-05-2016 01:30
95	Nicole71	09-05-2016 17:30

Insight:We got the 5 most loyal users by running the above query

- **Remind Inactive Users to Start Posting:**
  Sql query used to perform analysis:

```
/*to find users who have never posted*/
select username from users
left join photos on users.id=photos.user_id
where photos.id is null;
```

here we selected all the users from the users table by joining it to the photos table using id in the user table as user_id in the photos table and then extracted the user names of such users for whom the photo id was null

sl.no	username
1	Aniya_Hackett
2	Kasandra_Homenick
3	Jaclyn81
4	Rocio33
5	Maxwell.Halvorson
6	Tierra.Trantow
7	Pearl7
8	Ollie_Ledner37
9	Mckenna17
10	David.Osinski47
11	Morgan.Kassulke
12	Linnea59
13	Duane60
14	Julien_Schmidt
15	Mike.Auer39
16	Franco_Keebler64
17	Nia_Haag
18	Hulda.Macejkovic
19	Leslie67
20	Janelle.Nikolaus81
21	Darby_Herzog
22	Esther.Zulauf61
23	Bartholome.Bernhard
24	Jessyca_West
25	Esmeralda.Mraz57
26	Bethany20

Insight: we got to know that 26 users were inactive and have not posted a single photo also we got their user_name so now we can send them a mail to remind them to upload a photo.

- **Declaring the contest winner:**

  Sql query used to perform analysis:

```
/*declaring the contest winner*/
select
        users.username, photos.id,photos.image_url,count(*) as total_likes
from likes
join photos on photos.id=likes.photo_id
join users on users.id=likes.photo_id
group by photos.id
order by total_likes desc
limit 10;
```

here we combined three tables in the database to get the photo with most likes ,I used the count function on each image url and named the column as totalikes then ordered the images on the basis of toal_likes, after running this query we extracted the top 10 photos on the basis of total likes

username	id	image_url	total_likes_on_phot
Kaley9	30	http://kenny.com	41
Jayson65	61	https://dejon.name	41
Zack_Kemmer93	52	https://hershel.com	41
Tomas.Beatty93	97	ıttps://carolanne.com	40
Alexandro35	13	https://fred.com	40
Javonte83	100	https://brook.com	39
essie_Stanton4	62	https://rigoberto.net	39
Seth46	44	http://golden.org	39
Mike.Auer39	66	http://lionel.net	39
Harley_Lind18	3	http://vicky.biz	38

from the above table we get to know that the photo with id 30 of the user kaley9 has the most likes and she should be awarded as the winner

- **Hashtag Researching:**

  Sql query used to perform analysis:

```
SELECT
        tags.tag_name,
        COUNT(*) AS total
FROM photo_tags
   JOIN tags
     ON photo_tags.tag_id= tags.id
GROUP BY tags.id
ORDER BY total DESC
Limit 5;
```

Here we joined the tables photo_tags and tags and extracted the tagnames and counted them on the photos used and named the column as total then we grouped the data on the basis of tags and got the top 5 hashtags used

tag_name	total
smile	59
beach	42
party	39
fun	38

Insight:These are the top 5 hash tags that people use on their Instagram post and by using these tags the companies can reach more people

- **Launch AD Campaign:**

  Sql query used to perform analysis:

```
/*users joined on the day*/
SELECT
    DAYNAME(created_at) AS day,
    count(*) as total
FROM users
GROUP BY day
ORDER BY total DESC
LIMIT 3;
```

Here we used the users table in the database and used the dayname() function by taking the created at as the parameter for the dayname() function and used the count function by grouping the data on the basis of dayname

day	total
Thursday	16
Sunday	16
Friday	15

Insight:The results came as out of the 100 users most people joined Instagram on thursdays

And Sundays(16 people each day) this means to launch an ad campaign these are the best suited days

- **User Engagement:**

  Sql query used to perform analysis:

  ```
  /*to find average post per user per day*/
  SELECT
        (SELECT COUNT(*) FROM photos) / (SELECT COUNT(*) FROM users) AS
  average;
  ```

Here we divided the total number of photos by the total number of users which gave us the average post per user

Insight:The result came out to be 2.57 post per user

- **Bots & Fake Accounts:**

Sql query used to perform analysis:

```
/*to identify the bot account*/
SELECT
   photos.user_id,
   COUNT(photos.user_id) AS users_in_photos,
   COUNT(likes.user_id) AS users_in_likes,
   users.username
FROM
   photos
      INNER JOIN
   likes ON photos.user_id = likes.user_id
      JOIN
   users ON photos.user_id = users.id
GROUP BY photos.user_id;
```

Here we joined likes ,photos and user table and got the user id and user name of those users who have liked all the photos as it is not possible for a human to like all the photos

Insight:we found that 64 users in this dataset were bots

user_id	users_in_photos	users_in_likes	username
2	376	376	Andre_Purdy85
3	316	316	Harley_Lind18
4	279	279	Arely_Bogan63
6	410	410	Travon.Waters
8	316	316	Tabitha_Schamberger11
9	340	340	Gus93
10	261	261	Presley_McClure
11	445	445	Justina.Gaylord27
12	308	308	Dereck65
13	465	465	Alexandro35
15	336	336	Billy52
16	412	412	Annalise.McKenzie16
17	234	234	Norbert_Carroll35
18	82	82	Odessa2
19	180	180	Hailee26
20	87	87	Delpha.Kihn
22	91	91	Kenneth64
26	470	470	Josianne.Friesen
27	79	79	Darwin29
28	308	308	Dario77
30	162	162	Kaley9
31	88	88	Aiyana_Hoeger
32	364	364	Irwin.Larson
33	385	385	Yvette.Gottlieb91
35	184	184	Lennie_Hartmann40
37	84	84	Yazmin_Mills95
38	170	170	Jordyn.Jacobson2
39	89	89	Kelsi26
40	85	85	Rafael.Hickle2
42	261	261	Maya.Farrell
43	430	430	Janet.Armstrong
44	344	344	Seth46
46	352	352	Malinda_Streich
47	380	380	Harrison.Beatty50
48	75	75	Granville_Kutch
50	243	243	Gerard79
52	425	425	Zack_Kemmer93
55	78	78	Meggie_Doyle
56	81	81	Peter.Stehr0
60	172	172	Sam52
61	83	83	Jayson65
62	176	176	Ressie_Stanton46
63	332	332	Elenor88

Results:

In this project I learned the basic functions in SQL and how to think analytically and what steps to follow while performing an analysis this project taught me new ways to approach the problem especially the last question was a bit tricky and hard to think off but after a few hours of brainstorming I got to how to extract that data I have noted the insights that i got while I was performing the analysis and I have mentioned it after every query that I performed.