Instagram User Analytics

By: Rahul Patil

Project Description:

- User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) in an attempt to derive business insights for marketing, product & development teams.
- These insights are then used by teams across the business to launch a new marketing campaign, decide on features to build for an app, track the success of the app by measuring user engagement and improve the experience altogether while helping the business grow.
- In this project we are going to analyze Instagram user data which help to improve the performance of application,improve the user engagement and attract new investment.

◆ Approach

- To successfully carry out this project we are going to use **SIX STEP** of Data Analysis Process i.e (Ask,Prepare,Process,Analyze,Share,Act)
- Ask step include asking right set of question which justify goal and give motivation to carry out analysis
- We have following set of question (reasons) to justify goal of this project .
  - Find the 5 oldest users of the Instagram from the database provided.
  - Find the users who have never posted a single photo on Instagram.
  - Identify the winner of the contest and provide their details to the team.
  - Identify and suggest the top 5 most commonly used hashtags on the platform
  - What day of the week do most users register on? Provide insights on when to schedule an ad campaign.
  - Provide how many times does average user posts on Instagram.
  - Provide data on fake and real account.

Approach

- Prepare: We have users data store in RDBMS in different table like user,photos,likes,tags
- This step include selecting right data,tools, data source to make project successful
- Process: Data we have is already in process format ,i.e data store in RDBMS is process,clean and useful data.
- Analyze: We are using SQL to analyze data to draw insight.
- Share: Most of solution management looking for is number for each set of question,we can share result in the form of numbers likes(top 5 users,top 5 hashtag use,total real users)
- Act: Step include taking decision base on insight opt from this project.

Tech-Stack Used

- Data is stored in RDBMS (Relational Database Management System) in different table
- We could use jupyter notebook and programming language to carry out this project. Which requires programming skills as well further extra step such as connecting to database,loading data from DBMS to notebook.
- While RDBMS come with inbuilt data query language called SQL(Structure Query Language) which can successfully carry out entire analysis for this project.
- Tech we are going to use for this project is Postgresql which is RDBMS

Insights

1. People who have been using the platform for the longest time?

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

id [PK] integer	username character varying (255)	created_at timestamp without time zone
11	Justina.Gaylord27	2017-05-04 16:32:15.577
6	Travon.Waters	2017-04-30 13:26:14.496
85	Milford_Gleichner42	2017-04-30 07:50:51.28
19	Hailee26	2017-04-29 18:53:39.65
24	Maxwell.Halvorson	2017-04-18 02:32:43.597

Top 5 oldest user

◆ Insights

2. Remind Inactive Users to Start Posting: By sending them promotional emails to post their 1st photo?

```
select users.id, username from users left join photos on users.id=photos.user_id where photos.image_url is null
```

id [PK] integer	username character varying (255)
25	Tierra.Trantow
66	Mike.Auer39
89	Jessyca_West
57	Julien_Schmidt
34	Pearl7
71	Nia_Haag
83	Bartholome.Bernhard
91	Bethany20
21	Rocio33
5	Aniya_Hackett
24	Maxwell.Halvorson
68	Franco_Keebler64
80	Darby_Herzog

76	Janelle.Nikolaus81
81	Esther.Zulauf61
90	Esmeralda.Mraz57
74	Hulda.Macejkovic
41	Mckenna17
54	Duane60
36	Ollie_Ledner37
53	Linnea59
49	Morgan.Kassulke
45	David.Osinski47
75	Leslie67
14	Jaclyn81
7	Kasandra_Homenick

There are 26 users who have never posted a single photo on Instagram

◆ Insights

3. User who gets the most likes on a single photo will win the contest now they wish to declare the winner.

```
select users.username,photos.id as photo_id,photo_like.total_likes
from photos
inner join
    (select photo_id, count(*) as total_likes from likes group by photo_id) as photo_like
on photos.id=photo_like.photo_id
join users
on users.id=photos.user_id
order by
    total_likes desc limit 5
```

username character varying (255) 🔒	photo_id integer 🔒	total_likes bigint 🔒
Zack_Kemmer93	145	48
Adelle96	182	43
Malinda_Streich	127	43
Seth46	123	42
Annalise.McKenzie16	52	41

Top 5 user who gets the most likes on a single photo can be declare winner

Insights

4. Which are commonly use hashtag on platform which help partner brand to reach their audience?

```sql
select
    tags.tag_name,tags.id,tag_counts.total_count
from tags
JOIN
    (select count(*) as total_count ,tag_id from photos_tags group by tag_id order by total_count desc) as tag_counts
on tags.id=tag_counts.tag_id
order by
    total_count desc limit 10;
```

	tag_name character varying (255)	id [PK] integer	total_count bigint
1	smile	21	59
2	beach	20	42
3	party	17	39
4	fun	13	38
5	concert	18	24

6	lol	11	24
7	food	5	24
8	hair	15	23
9	happy	12	22
10	dreamy	10	20

Above top 10 most commonly used hashtags on the platform this could help partner brand to reach their audience

Insights

5. Which day would be the best day to launch ADs?

```
select * from users;

select
    count(*) as total_account_created,to_char(created_at,'day') as "Day"
from
    users
group by "Day" order by total_account_created desc;
```

total_account_created bigint	Day text
16	thursday
16	sunday
15	friday
14	tuesday
14	monday
13	wednesday
12	saturday

- Most of new user added on thursday and sunday
- We can use this 2 days to schedule ADs campaign

Insights

6. how many times does average user posts on Instagram.

```sql
SELECT users.username,COUNT(photos.image_url)
FROM users
JOIN photos ON users.id = photos.user_id
GROUP BY users.id
ORDER BY 2 DESC;
```

username character varying (255)	count bigint
Eveline95	12
Clint27	11
Cesar93	10
Delfina_VonRueden68	9
Aurelie71	8
Jaime53	8
Donald.Fritsch	6
Zack_Kemmer93	5
Josianne.Friesen	5
Janet.Armstrong	5
Travon.Waters	5
Florence99	5

❖ Some user are highly active as they are posting 10-12 post.

```sql
SELECT count(*) as total_users,(SELECT COUNT(*)FROM photos) as total_post,
  ROUND((SELECT COUNT(*)FROM photos)/(SELECT COUNT(*) FROM users),2)as avg_post_per_user from users;
```

❖ While per user average post is 2

total_users bigint	total_post bigint	avg_post_per_user numeric
100	257	2.00

◆ Insights

7. Provide data on real user and fake(boat) accounts.

```
select
    user_id,count(photo_id) as Total_post_like
from
    likes
group by
    user_id
having
    count(photo_id)=(select count(*) from photos)
```

There are 13 account who had like every single post as normal user won't do this

This is mostly a (bots) or fake account.

	user_id integer 🔒	total_post_like bigint 🔒
2	71	257
3	36	257
4	14	257
5	66	257
6	75	257
7	41	257
8	57	257
9	24	257
10	21	257
11	76	257
12	5	257
13	91	257

Result

❖ In this project we understand how to use SIX step process of Data analysis.

❖ Project help to understand how to use SQL joins as data store in different table for simple data analytics job .

❖ Key project insight are :

- How is user engagement on platform
- Thursday and sunday would be most favourable day to launch any AD campaign
- There are 13 fake account out of 100 i.e 13% fake account which need to regulate using implementing new police to increase real user on platform