
Instagram Data Analysis

Muhammad Asfand Yar



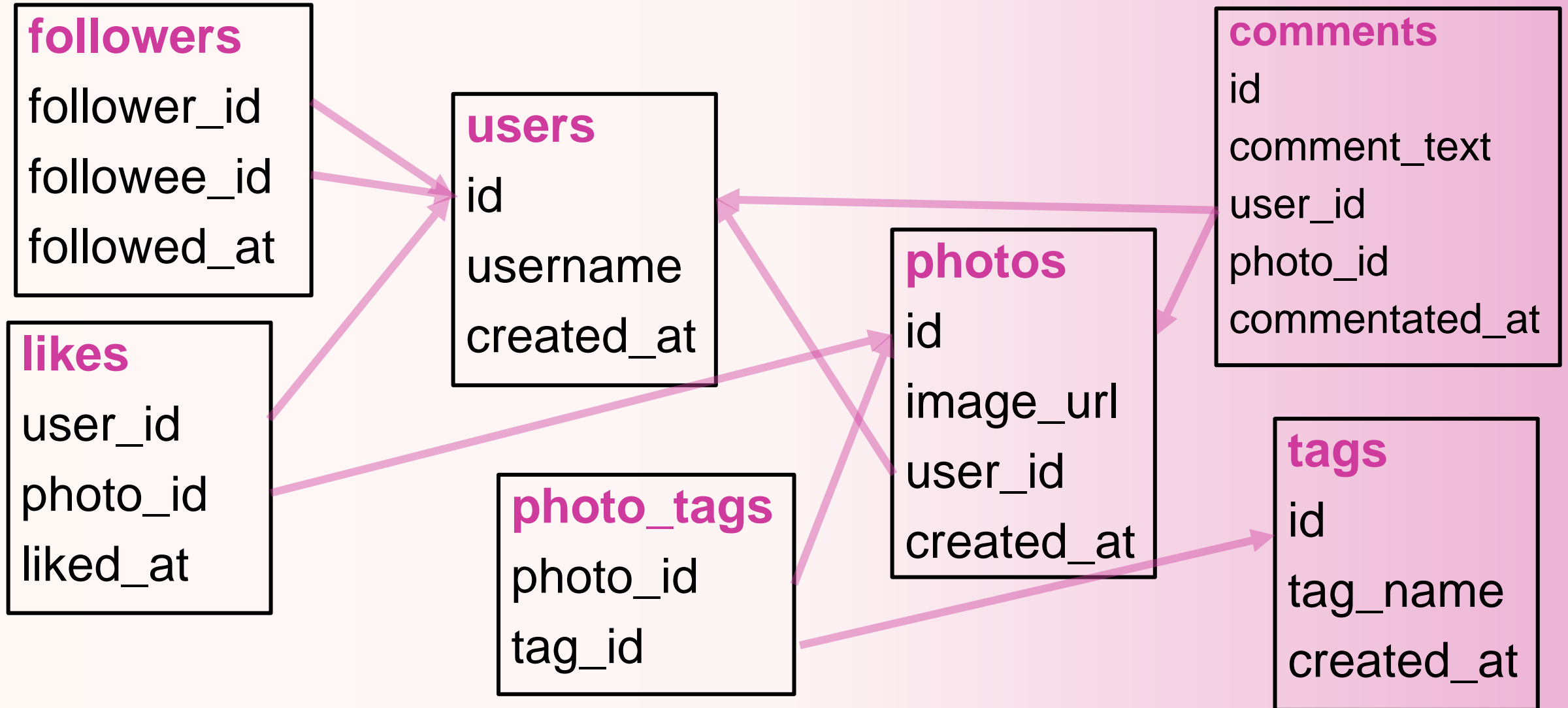
Summary

Utilized SQL to create an Instagram data schema which included tables for users, photos, comments, likes, followers, and hashtags.

Specified data type constraints for each table and created inter-table relationships by specifying primary keys and foreign keys.

After inserting a large set of data into the Instagram database, I analyzed the data using SQL to solve several business cases.

Creating the Data Schema



Case One

Dear Muhammad,

Our advertisers are worried that Instagram may have a bot problem. It is very important for our company and our advertisers that a substantive majority of our users are active users.

I need you to investigate this issue and prepare a report with your analysis.

Kind regards,

Manager

Approach

First, I need to determine what constitutes bot like activity and determine the proportion of users that are bots on Instagram.

From my judgment I will constitute an account that likes more than 10% of the posts on Instagram and has zero uploads as a bot account.

It is reasonable to believe that no human would be so active as to like 10% of the posts on Instagram, but at the same time not upload anything.

The thought process behind this judgment of what constitutes a bot account must be conveyed to the manager.

SQL Analysis

```
121 • CREATE VIEW bots AS
122     SELECT
123         users.id AS 'Account ID Number',
124         users.username AS 'List of Bot Accounts'
125     FROM users
126     INNER JOIN likes ON users.id = likes.user_id
127     LEFT JOIN photos ON users.id = photos.user_id
128     WHERE image_url IS NULL
129     GROUP BY users.id
130     HAVING COUNT(*) > ((SELECT COUNT(*) FROM photos) * 0.10);
```

Gives us a list of bot accounts

```
96 • SELECT
97     (COUNT(*) / (SELECT COUNT(*) FROM users)) * 100 AS
98     'Percentage of Bot Accounts'
99 FROM bots;
```

13% of Instagram accounts are bots

Conclusion

Dear Manager,

From my analysis I have found that about **13% of the accounts** on Instagram are what I believe to be bot accounts.

In my analysis, a bot account was one which had zero posts and had liked more than 10% of the posts on Instagram.

From my findings **I do not believe that Instagram has a bot problem**, as the accounts are a small proportion and can be identified and deleted.

An **Excel file has been attached to this email, listing the bot accounts and their account ID numbers**. I recommend not including these accounts in our active user total, this will give advertisers more confidence when advertising on our platform.

Kind regards,

Muhammad Asfand Yar

	A	B
1	Account ID Number	List of Bot Accounts
2	5	Aniya_Hackett
3	14	Jaclyn81
4	21	Rocio33
5	24	Maxwell.Halvorson
6	36	Ollie_Ledner37
7	41	Mckenna17
8	54	Duane60
9	57	Julien_Schmidt
10	66	Mike.Auer39
11	71	Nia_Haag
12	75	Leslie67
13	76	Janelle.Nikolaus81
14	84	D... ..

Case Two

Dear Muhammad,

One of our biggest advertisers, Luis Vuitton, reached out to us and is wondering when is the best time to display promotional ads that will bring new users to their Instagram page.

They also want to know what are the best hashtags to use in their posts.

Our business relationship with Luis Vuitton is very important, as they bring in substantial users and ad revenue to our company through the ads they run on our platform.

I would like you to prepare a report with your analysis and send it to me by this Friday.

Thank you,

Manager

Approach

- First, I need to find the day of the week and month of the year when new users join Instagram the most.
- I will then visualize this data in Excel by using a bar graph.
- Then, I need to find the hashtag that is most popular on Instagram posts.
- I will then present the list of hashtags from most popular to least popular in on a CSV file.

SQL Analysis

Gives us a list of the 30 most used hashtags on Instagram

```
103 • SELECT
104     tags.tag_name AS 'Hashtag',
105     COUNT(*) AS 'Total Number of Times Used'
106 FROM tags
107 INNER JOIN photo_tags ON tags.id = photo_tags.tag_id
108 GROUP BY tags.id
109 ORDER BY COUNT(*) DESC LIMIT 30;
```

Excel Visualization

	A	B
1	Hashtag	Total Number of Times Used
2	smile	59
3	beach	42
4	party	39
5	fun	38
6	food	24
7	lol	24
8	concert	24
9	hair	23
10	happy	22
11	beauty	20
12	dreamy	20
13	sunset	19
14	fashion	19
15	drunk	19

most_used_hashtags

Louis Vuitton should select the most relevant hashtags from this list to use on their posts.

SQL Analysis

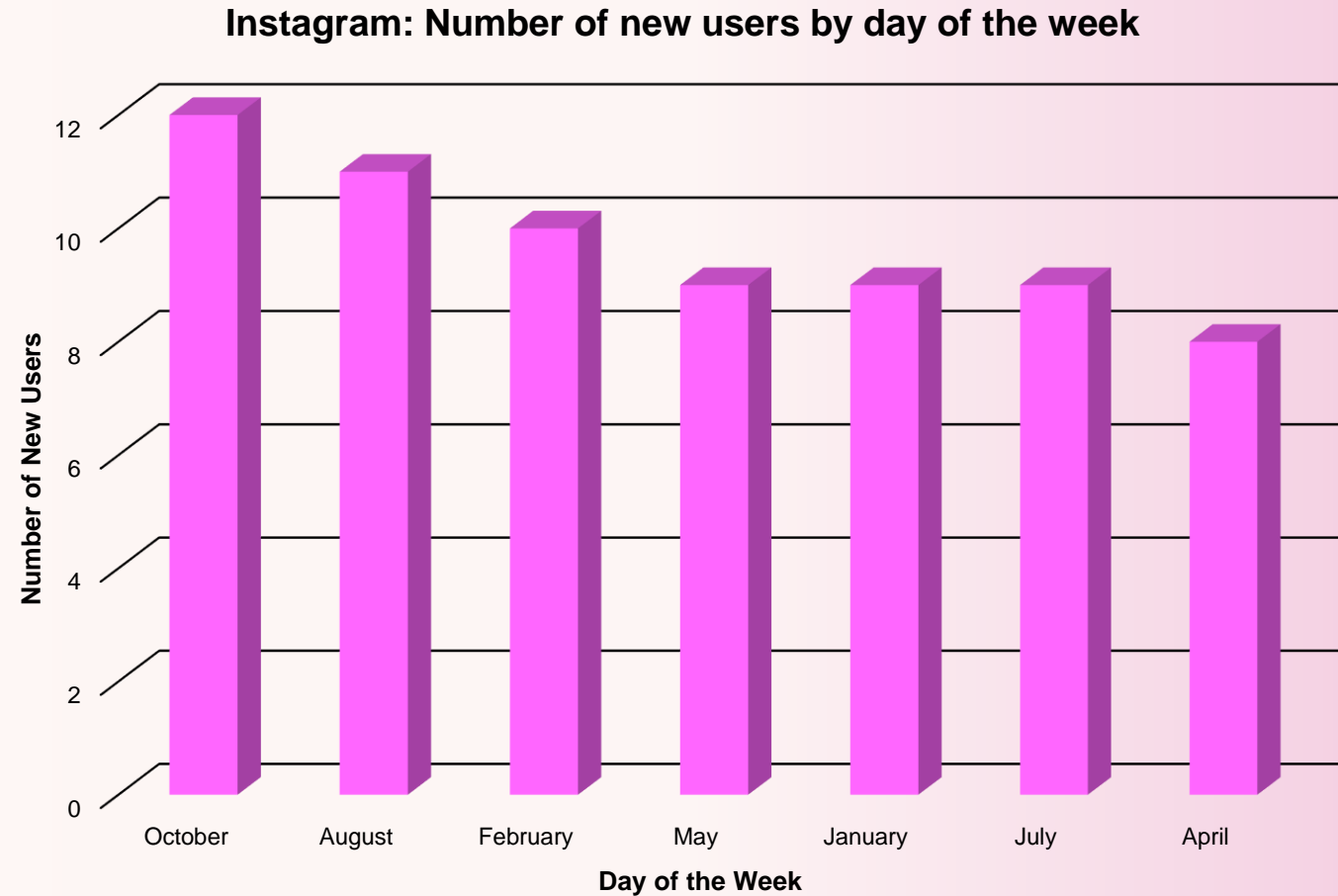
Gives us the days in which most users sign-up to Instagram

```
77 • SELECT
78     (DATE_FORMAT(created_at, '%W')) AS DAY,
79     COUNT(*) AS 'Number of Sign-Ups by Day'
80 FROM users
81 GROUP BY DAY;
```

Gives us the months in which most users sign-up to Instagram

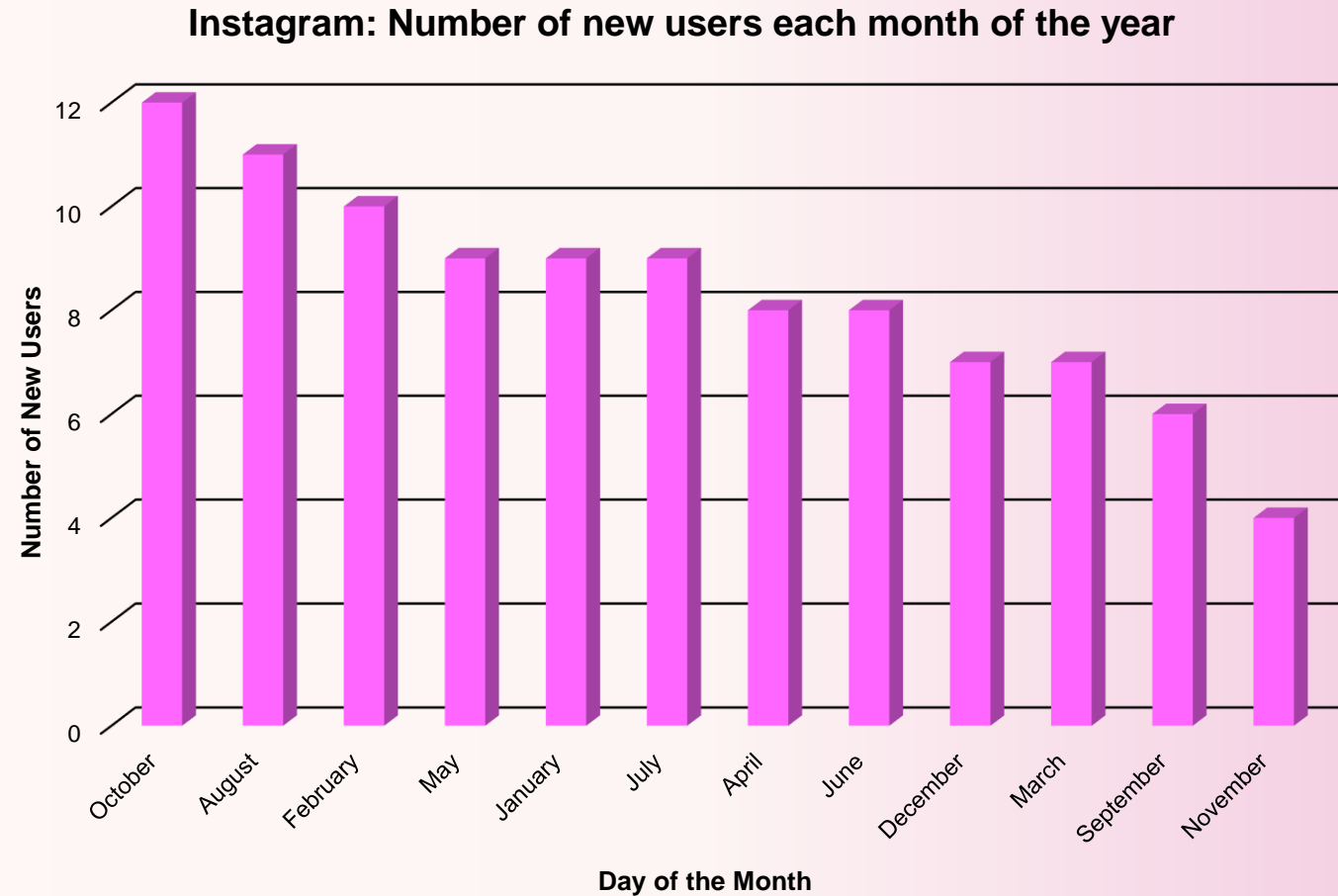
```
83 • SELECT
84     (DATE_FORMAT(created_at, '%M')) AS DAY,
85     COUNT(*) AS 'Number of New User by Month'
86 FROM users
87 GROUP BY DAY;
```

Excel Visualization



From this data we can see that the best day of the week to post promotional ads for new users is either on Thursday or Sunday.

Excel Visualization



From this data we can see that the best month of the year to post promotional ads for new users is in October.

Conclusion

Dear Manager,

From my analysis, **I have found the 30 most used hashtags on Instagram.** Please send the CSV file attached in this email to Louis Vuitton.

I would advise Luis Vuitton to use one of the listed hashtags to maximize the reach and discovery of their posts.

Also from my analysis, I have found that **Thursday** and **Sunday** are the days of the week with the most new users, and **October** as the month of the year with the more new users.

To maximize their reach, I would advise Luis Vuitton to **run their promotional adds for new users on every Thursday and Sunday in the month of October.**

Please let me know if you have any questions.

Kind regards,

Muhammad Asfand Yar

