Instagram User Analytics

Project Description:

This project focuses on analyzing user engagement with Instagram to gain valuable business insights for the marketing, product, and development teams. By utilizing SQL queries on a cloned Instagram dataset in a MySQL database, I was able to extract information on various metrics, such as identifying the oldest and inactive users, contest winners, frequently used hashtags, registration patterns, average posts per user, and total user and photo counts. Additionally, the use of SQL queries enabled the detection of bots and fake accounts.

Approach:

To execute this project, MySQL workbench was utilized. Initially, SQL queries were employed to create a database using the given raw data. Subsequently, multiple SQL operations and queries were executed based on the management team's provided scenarios to derive useful insights. These insights aided in improving the overall user experience and facilitated business growth.

Tech-Stack Used:

MySQL Workbench v8.0 is used a database management tool for firing SQL queries and getting desired insights.

Insights:

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 the Instagram from the database provided

Query:

SELECT username, created at

FROM ig clone.users

ORDER BY created at ASC

LIMIT 5;

Output:

	username	created_at	
•	Darby_Herzog	2016-05-06 00:14:21 2016-05-06 13:04:30	
	Emilio_Bernier52		
	Elenor88	2016-05-08 01:30:41	
	Nicole71	2016-05-09 17:30:22	
	Jordyn, Jacobson 2	2016-05-14 07:56:26	

The above output shows the 5 oldest users of Instagram based on their data and time of account creation

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

Query:

SELECT username

FROM ig_clone.users

WHERE id

NOT IN (SELECT user id FROM ig clone.photos);

Output:

username
Kasandra_Homenick
Jaclyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna 17
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39
Franco_Keebler64
Nia_Haag
Hulda.Macejkovic
Leslie67
Janelle.Nikolaus81
Darby_Herzog
Esther.Zulauf61
Bartholome.Bernhard
Jessyca_West
Esmeralda.Mraz57
Bethany20

The above output shows the list of users which have registered but haven't posted any photo on Instagram.

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

Query:

SELECT users.username, photos.id, count(*) AS amount of likes

FROM ig clone.likes

JOIN ig clone.photos ON ig clone.photos.id=ig clone.likes.photo id

JOIN ig clone.users ON ig clone.users.id=ig clone.likes.photo id

GROUP BY ig_clone.photos.id

ORDER BY amount_of_likes DESC

LIMIT 1;

Output:

	username	id	amount_of_likes
•	Kaley9	30	41

The above shows the username of the winner of the contest with highest amount of likes on a single photo accounting to 41 likes.

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

Query:

SELECT tag_name, COUNT(tag_name) AS frequency

FROM ig clone.tags

JOIN ig clone.photo tags ON ig clone.tags.id = ig clone.photo tags.tag id

GROUP BY ig clone.tags.id

ORDER BY frequency DESC

LIMIT 5;

Output:

	tag_name	frequency
١	smile	59
	beach	42
	party	39
	fun	38
	concert	24

The above output shows the top 5 most commonly used hashtags based on the frequency of use.

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

Query:

SELECT date format(created at, '%W') AS 'Day', COUNT(*) AS 'No of Registrations'

FROM ig_clone.users

GROUP BY 1

ORDER BY 2 DESC;

Output:

	Day	No of Registrations
١	Thursday	16
	Sunday	16
	Friday	15
	Tuesday	14
	Monday	14
	Wednesday	13
	Saturday	12

The above output shows the day of week consisting of user registrations. Since most numbers of registrations happened on Sunday and Thursday, thus, these days can be considered as a best day to conduct AD campaign.

B. Investor Metrics:

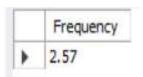
1. User Engagement: Are users still as active and post on Instagram or they are making fewer posts

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

Query:

SELECT ROUND((SELECT COUNT(*)FROM ig_clone.photos)/(SELECT COUNT(*)FROM ig_clone.users),2) AS 'Frequency';

Output:



The above output shows the frequency of posts per user on Instagram

2. 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).

Query:

SELECT id, username, COUNT(users.id) As No of likes

FROM ig_clone.users

JOIN ig clone.likes ON ig clone.users.id = ig clone.likes.user id

GROUP BY ig clone.users.id

HAVING No of likes = (SELECT COUNT(*) FROM ig clone.photos);

Output:

	id	username	No_of_likes
•	5	Aniya_Hackett	257
	14	Jaclyn81	257
	21	Rocio33	257
	24	Maxwell.Halvorson	257
	36	Ollie_Ledner37	257
	41	Mckenna 17	257
	54	Duane60	257
	57	Julien_Schmidt	257
	66	Mike. Auer 39	257
	71	Nia_Haag	257
	75	Leslie67	257
	76	Janelle.Nikolaus81	257
	91	Bethany20	257

The above output shows the users (bots) which have liked every single photo i.e. 257 likes on 257 photos.

Result:

The primary objective of this project was to comprehend the necessary specifications and provide valuable insights to both the Marketing team and Investors. By analyzing the Instagram database using SQL queries, I generated results that can facilitate modifications to the marketing strategy, identify areas where improvements are needed, and determine where we are falling short.