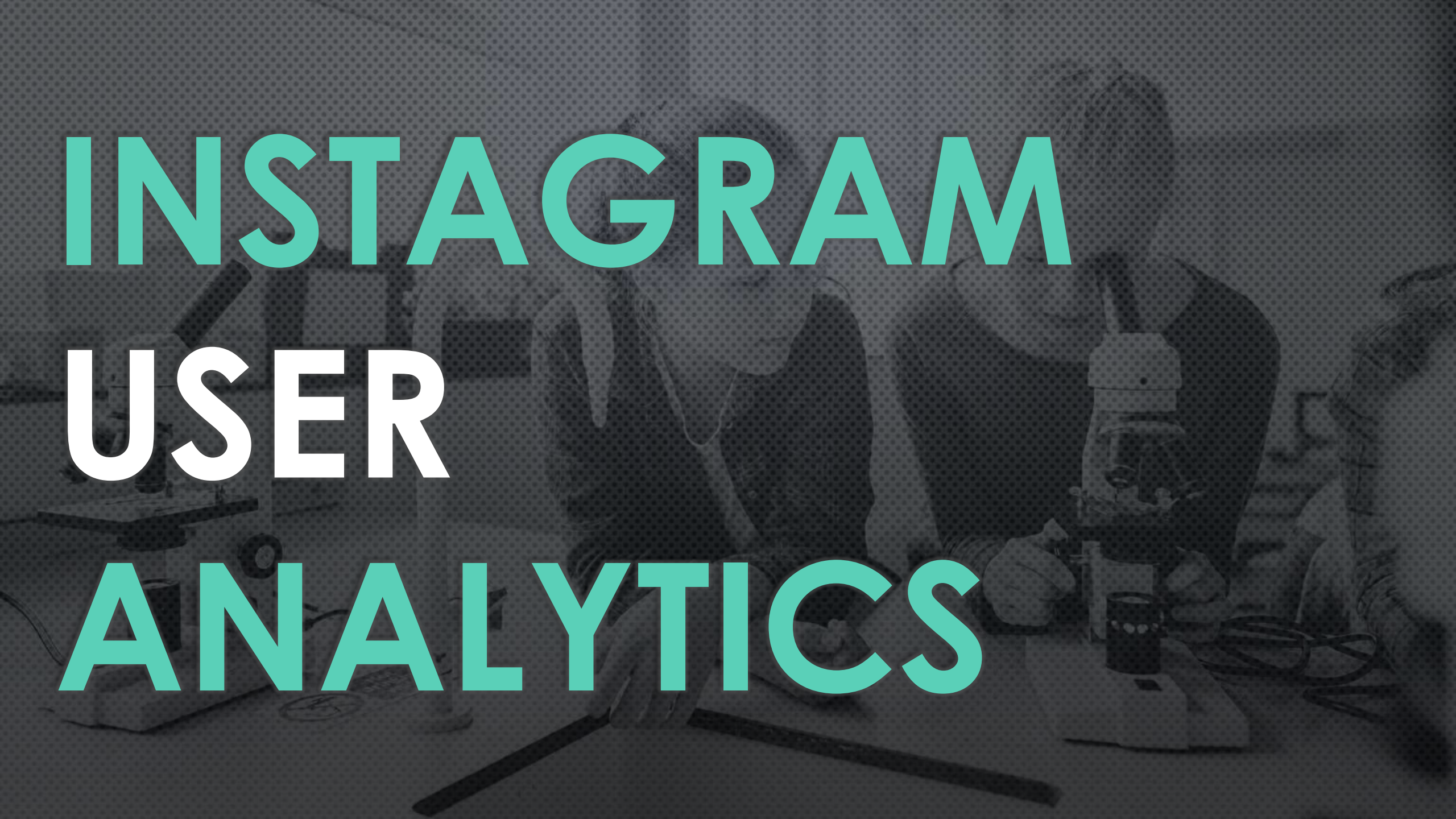
INSTAGRAM
USER
ANALYTICS

PROJECT DESCRIPTION

- As a Data Analyst collaborating with the Instagram Product team, my primary responsibility is about understanding user interactions and engagement within the Instagram app.
- The aspect of my role is to extract meaningful insights from the Instagram user data. These insights will offer actionable information for various teams within the business.
- The Product team can then make informed decisions about introducing new features and the development of enhancements to improve overall user experience.
- The goal is to empower the product manager and the entire team with actionable insights that will shape the future development and user experience in the app

APPROACH

- Understanding the Schema: The first step is to examine the structure of the table holding the user and engagement data.
- Identifying the Key tables: Identification of the primary key from each of the tables of Users, likes, comments, photos, tags etc.
- Checking for null values: Before the analysis, it is necessary to check for null values in the given tables
- Visually Appealing: The SQL Queries need to be properly formatted so that they can be understood by any user.

A) MARKETING ANALYSIS

1) LOYAL USER REWARD: IDENTIFY THE FIVE OLDEST USERS ON INSTAGRAM FROM THE PROVIDED DATABASE

```
89    -- TASK 01 LOYAL USER REWARD  (VICTOR SHAH)
90    SELECT
91        *
92    FROM
93        users
94    ORDER BY created_at ASC
95    LIMIT 5;
```

Result Grid | Filter Rows: | Edit:

id	username	created_at
80	Darby_Herzog	2016-05-06 00:14:21
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson2	2016-05-14 07:56:26
NULL	NULL	NULL

2) INACTIVE USER MANAGEMENT: IDENTIFY USERS WHO HAVE NEVER POSTED A SINGLE PHOTO ON INSTAGRAM

```
97      -- TASK 02 INACTIVE USER ENGAGEMENT  (VICTOR SHAH)
98  •   SELECT u.username
99      FROM users AS u
100     LEFT JOIN photos AS p ON u.id = p.user_id
101     WHERE p.image_url IS NULL;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Co

username
Aniya_Hackett
Kasandra_Homenick
Jadyn81
Rocio33
Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
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

3) CONTEST WINNER DECLARATION: DETERMINE THE WINNER OF THE CONTEST AND PROVIDE THEIR DETAILS TO THE TEAM

```
103     -- TASK 03 CONTEST WINNER DECLARATION  (VICTOR SHAH)
104 •   SELECT
105         u.id, u.username, l.photo_id, COUNT(l.photo_id) AS likes
106     FROM
107         likes AS l
108             JOIN
109         photos AS p ON l.photo_id = p.id
110             JOIN
111         users AS u ON p.user_id = u.id
112     GROUP BY l.photo_id
113     ORDER BY COUNT(l.photo_id) DESC
114     LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fe

id	username	photo_id	likes
52	Zack_Kemmer93	145	48

4) HASHTAG RESEARCH: IDENTIFY AND SUGGEST THE TOP FIVE MOST COMMONLY USED HASHTAGS ON THE PLATFORM

```sql
-- TASK 04 HASHTAG RESEARCH  (VICTOR SHAH)
SELECT
    p.tag_id, t.tag_name, COUNT(p.tag_id) AS frequency
FROM
    photo_tags AS p
        JOIN
    tags AS t ON p.tag_id = t.id
GROUP BY tag_id
ORDER BY COUNT(tag_id) DESC
LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tag_id	tag_name	frequency
21	smile	59
20	beach	42
17	party	39
13	fun	38
18	concert	24

5) AD CAMPAIGN LAUNCH: DETERMINE THE DAY OF THE WEEK WHEN MOST USERS REGISTER ON INSTAGRAM

```sql
-- TASK 05 AD CAMPAIGN LAUNCH  (VICTOR SHAH)
SELECT d, COUNT(d)
FROM
    (SELECT
        CASE WEEKDAY(created_at)
                WHEN 0 THEN 'MONDAY'
                WHEN 1 THEN 'TUESDAY'
                WHEN 2 THEN 'WEDNESDAY'
                WHEN 3 THEN 'THURSDAY'
                WHEN 4 THEN 'FRIDAY'
                WHEN 5 THEN 'SATURDAY'
                WHEN 6 THEN 'SUNDAY'
            END AS d
    FROM
        users) AS sub1
GROUP BY d
ORDER BY COUNT(d) DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Con

d	count(d)
THURSDAY	16
SUNDAY	16
FRIDAY	15
TUESDAY	14
MONDAY	14
WEDNESDAY	13
SATURDAY	12

THE BEST DAY TO SCHEDULE AN AD CAMPAIGN FOR THE REGISTRATION OF INSTAGRAM USERS WILL BE BOTH **THURSDAY** AND **SUNDAY** DUE TO LARGE COUNT IN THE REGISTRATION OF NEW USERS

B) INVESTOR METRICS

1) USER ENGAGEMENT: CALCULATE THE AVERAGE NUMBER OF POSTS PER USER ON INSTAGRAM. ALSO, PROVIDE THE TOTAL NUMBER OF PHOTOS ON INSTAGRAM DIVIDED BY THE TOTAL NUMBER OF USERS

```
138    -- TASK 06 USER ENGAGEMENT PART I (VICTOR SHAH)
139
140    SELECT
141        AVG(number_of_posts) AS average
142    FROM
143        (SELECT
144            user_id, COUNT(user_id) AS number_of_posts
145        FROM
146            photos
147        GROUP BY user_id) AS sub1;
148    SELECT
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

average
3.4730

```
149    -- TASK 06 USER ENGAGEMENT PART II (VICTOR SHAH)
150
151    SELECT
152        (SELECT
153                COUNT(*)
154            FROM
155                photos) / COUNT(*) AS division
156    FROM
157        users;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Co

division
2.5700

2) BOTS AND FAKE ACCOUNTS: IDENTIFY USERS (POTENTIAL BOTS) WHO HAVE LIKED EVERY SINGLE PHOTO ON THE SITE, AS THIS IS NOT TYPICALLY POSSIBLE FOR A NORMAL USER

```
159    -- TASK 07 BOTS AND FAKE ACCOUNTS  (VICTOR SHAH)
160
161    SELECT
162        user_id, u.username as bots, COUNT(user_id) as likes
163    FROM
164        likes AS l
165            JOIN
166        users AS u ON l.user_id = u.id
167    GROUP BY user_id
168    HAVING COUNT(user_id) = (SELECT
169            COUNT(*)
170        FROM
171            photos);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

user_id	bots	likes
5	Aniya_Hackett	257
14	Jadyn81	257
21	Rocio33	257
24	Maxwell.Halvorson	257
36	Ollie_Ledner37	257
41	Mckenna17	257
54	Duane60	257
57	Julien_Schmidt	257
66	Mike.Auer39	257
71	Nia_Haag	257
75	Leslie67	257
76	Janelle.Nikolaus81	257
91	Bethany20	257

TECH-STACK USED

- MySQL Workbench(8.0.34): This is the primary interactive development environment for SQL queries. It enables efficient query building, execution and debugging for data analysis

- SQL commands:
  1. DDL commands: These commands were used for the creation of the database and the multiple tables such as users, likes, comments and photos.
  2. DML commands: These commands were used for the insertion of the data into the records of the table.
  3. DQL commands: The select query with where, order by, group by clauses helped for the further analysis of the data from the table.

INSIGHTS

- While analysing the tables we were able to figure out the oldest customers of the Instagram app. For these customers to remain with the application the best way would gift them perks.
- The identification of bot and fake accounts will help in removal of these accounts to improve security and privacy within the application.
- We were able to understand how different users engage with the Instagram app.
- The Identification of the most popular content and trends such as hashtags.
- Targeted ad campaigns to different user groups.

RESULTS

- Remembering to adapt these queries on specific database schemas.
- These learned insights helped me understand specific business questions which were addressed by SQL queries.
- Learning about the SQL clauses such as the join clauses and sub-queries. The importance of order by and group by and many more.
- We were able to segment the user database of the Instagram app into various categories from old to new accounts and real to fake accounts.
- Achieving the ability to learn and write SQL queries to execute different business questions.


THANK
YOU