# INSTAGRAM USER ANALYTICS

**Project Description:** Using Data Analytics to extract useful information like the oldest users, most used hashtags, unactive users, finding the contest winner, the most suitable day for AD campaign in a week, fake accounts, average number of posts made by a user from the Instagram database provided

**Tech-Stack Used**: MYSQL workbench 8.0 CE

**Approach:**

- Create the database in the workbench using the data provided

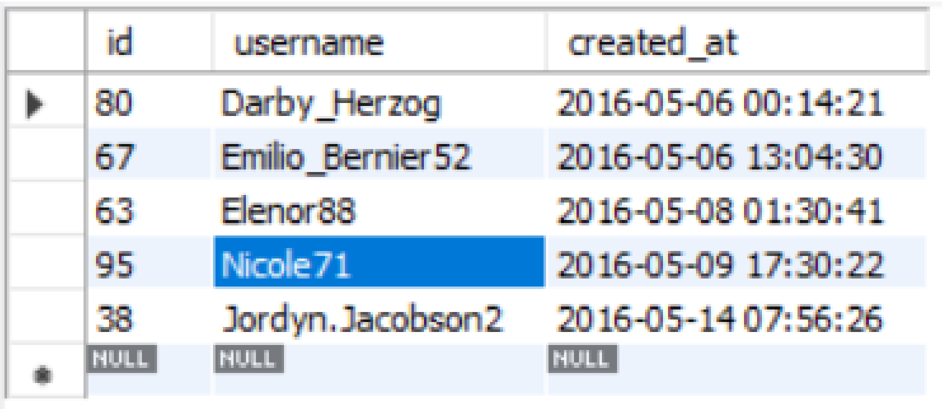
## MARKETING

### Task1: Find the 5 oldest users of the Instagram from the database provided

SQL QUERY:

```
USE ig_clone;

SELECT * FROM users

ORDER BY created_at

LIMIT 5
```

OUTPUT:



| | 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 |

(The rows are sorted by the ascending order of the time the account is created and the top 5 users are taken)

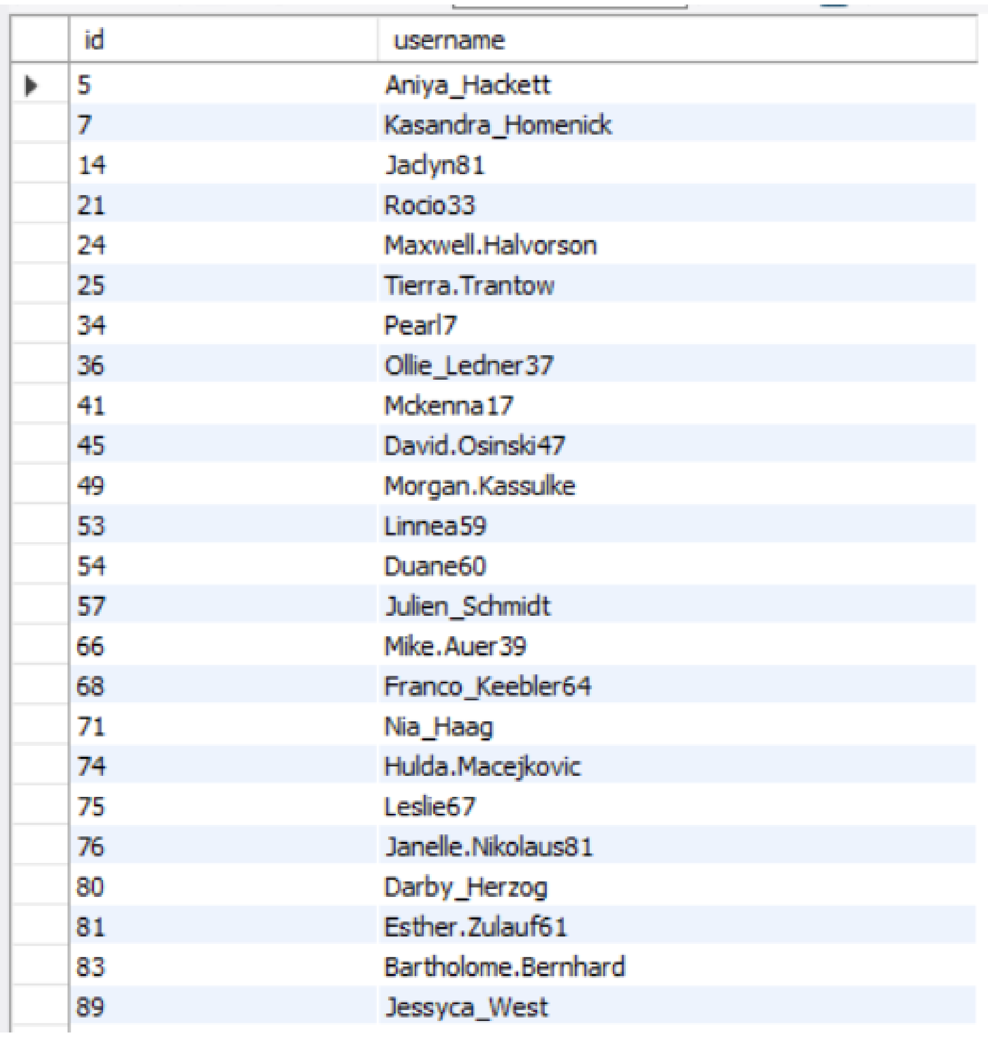
### Task2: Find the users who have never posted a single photo on Instagram

SQL QUERY:

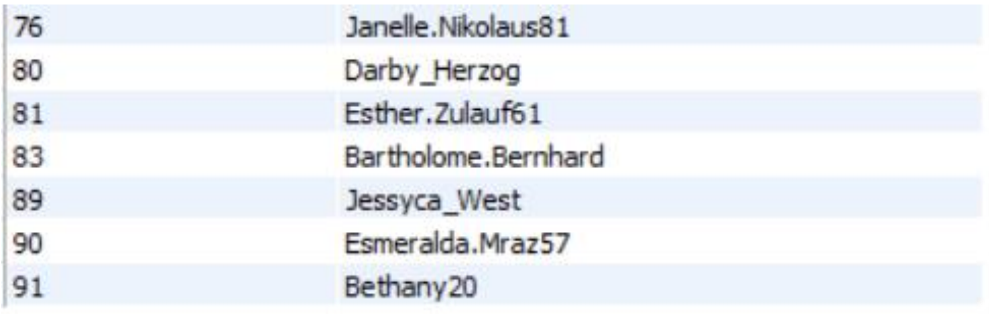
```
USE ig_clone;

SELECT U.id, U.username FROM users U
```

```
LEFT JOIN photos P
 ON U.id = P.user_id
WHERE P.user_id IS NULL
```

OUTPUT:



| id | username |
|---|---|
| 5 | Aniya_Hackett |
| 7 | Kasandra_Homenick |
| 14 | Jaclyn81 |
| 21 | Rocio33 |
| 24 | Maxwell.Halvorson |
| 25 | Tierra.Trantow |
| 34 | Pearl7 |
| 36 | Ollie_Ledner37 |
| 41 | Mckenna17 |
| 45 | David.Osinski47 |
| 49 | Morgan.Kassulke |
| 53 | Linnea59 |
| 54 | Duane60 |
| 57 | Julien_Schmidt |
| 66 | Mike.Auer39 |
| 68 | Franco_Keebler64 |
| 71 | Nia_Haag |
| 74 | Hulda.Macejkovic |
| 75 | Leslie67 |
| 76 | Janelle.Nikolaus81 |
| 80 | Darby_Herzog |
| 81 | Esther.Zulauf61 |
| 83 | Bartholome.Bernhard |
| 89 | Jessyca_West |



| | |
|---|---|
| 76 | Janelle.Nikolaus81 |
| 80 | Darby_Herzog |
| 81 | Esther.Zulauf61 |
| 83 | Bartholome.Bernhard |
| 89 | Jessyca_West |
| 90 | Esmeralda.Mraz57 |
| 91 | Bethany20 |

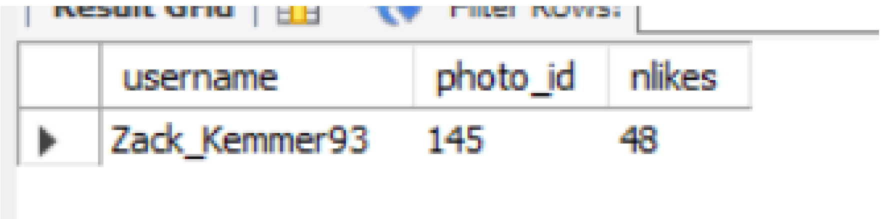
(The users table and the photos table are joined using left join and the null values are selected)

## Task3: Identify the winner of the contest and provide their details to the team

SQL QUERY:

```
USE ig_clone;
SELECT  users.username, likes.photo_id, COUNT(likes.user_id) AS nlikes FROM likes
JOIN photos
 ON photos.id = likes.photo_id
JOIN users
 ON photos.user_id = users.id
GROUP BY photo_id
ORDER BY nlikes DESC
LIMIT 1
```

OUTPUT:



| username | photo_id | nlikes |
|---|---|---|
| Zack_Kemmer93 | 145 | 48 |

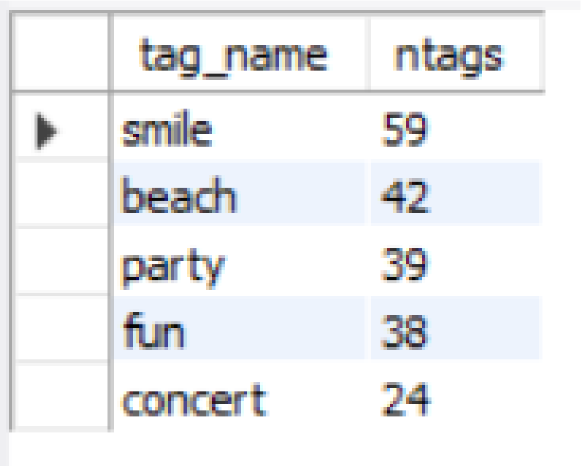
(The likes, users, photos tables are joined and the number of users liked a photo is grouped and the user who posted the photo with the highest number of likes is selected as the winner)

## Task4: Identify and suggest the top 5 most commonly used hashtags on the platform

SQL QUERY:

```
USE ig_clone;
SELECT tag_name, COUNT(user_id) AS ntags FROM photo_tags pt
JOIN tags t
 ON pt.tag_id = t.id
JOIN photos p
 ON pt.photo_id = p.id
JOIN users u
 ON p.user_id = u.id
GROUP BY tag_name
ORDER BY ntags DESC
LIMIT 5
```

OUTPUT:



| | tag_name | ntags |
|---|---|---|
| ▶ | smile | 59 |
| | beach | 42 |
| | party | 39 |
| | fun | 38 |
| | concert | 24 |

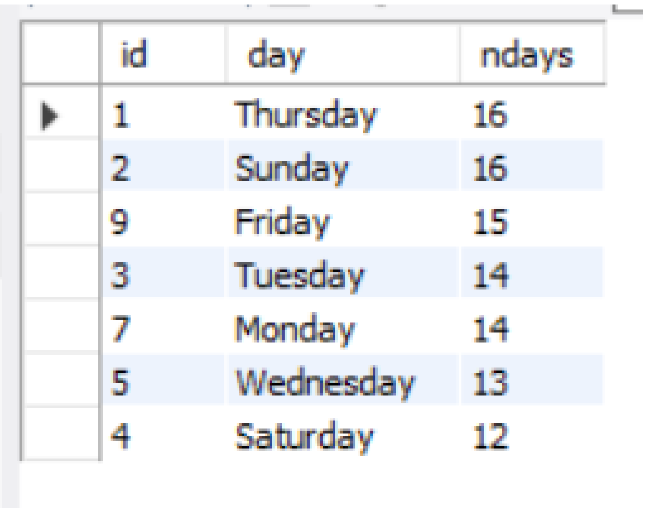
(The photo tags, photos, tags and users tables are joined the togs used by the most users are grouped and sorted in descending order and top 5 are selected)

### Task5: What day of the week do most users register on? Provide insights on when to schedule an ad campaign

SQL QUERY:

```
USE ig_clone;

SELECT id, DAYNAME(created_at) AS day, COUNT(id) AS ndays

FROM users

GROUP BY day

ORDER BY ndays DESC
```

OUTPUT:



| | id | day | ndays |
|---|---|---|---|
| ▶ | 1 | Thursday | 16 |
| | 2 | Sunday | 16 |
| | 9 | Friday | 15 |
| | 3 | Tuesday | 14 |
| | 7 | Monday | 14 |
| | 5 | Wednesday | 13 |
| | 4 | Saturday | 12 |

(The created_at column is modified to the day of that date and then grouped by the number of times the day occurred)

People are more active and registered on Thursday and Sunday in the week.

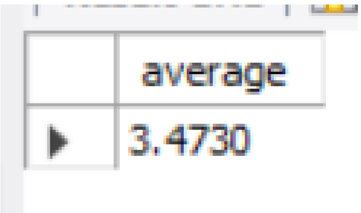
# Investor Metrics

Task1: Provide how many times does average user posts on Instagram. Also, provide the total number of photos on Instagram/total number of users

SQL QUERY1:

```
USE ig_clone;

SELECT COUNT(photos.id)/COUNT(DISTINCT photos.user_id) AS average

FROM photos
```

OUTPUT:



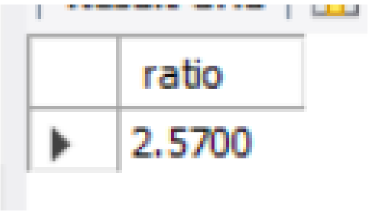
| | average |
|---|---|
| ▶ | 3.4730 |

(The number of photos divided by number of users who posted those photos gives the average number of photos posted by users)

SQL QUERY2:

```
USE ig_clone;

SELECT  COUNT(DISTINCT photos.id)/COUNT(DISTINCT users.id) AS ratio

FROM users , photos
```

OUTPUT:



| | ratio |
|---|---|
| ▶ | 2.5700 |

(The ratio of total number of photos and total number of users is taken)

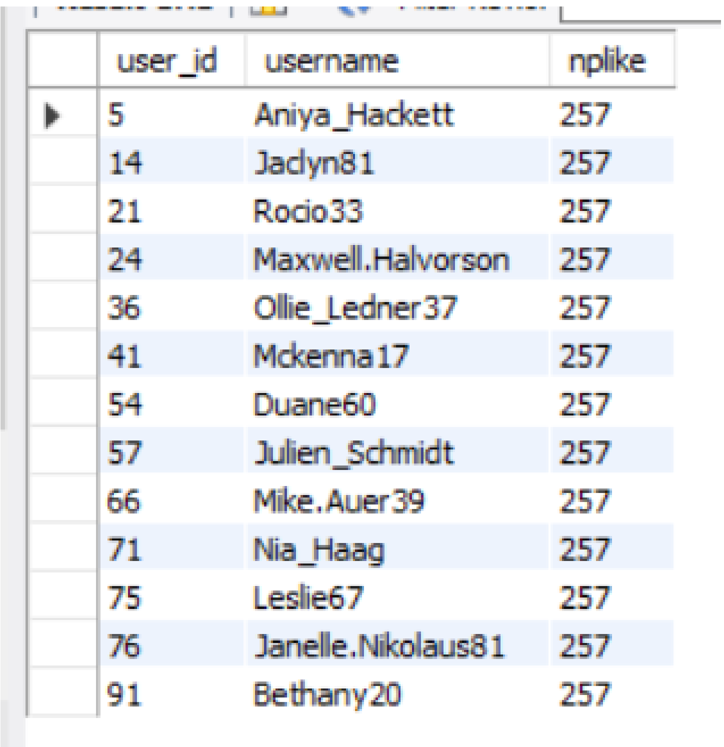
Task2: Provide data on users (bots) who have liked every single photo on the site (since any normal user would not be able to do this).

SQL QUERY:

```
USE ig_clone;

SELECT likes.user_id, users.username, COUNT(likes.photo_id) AS nplike

FROM likes
```

```
JOIN users
 ON likes.user_id = users.id
GROUP BY user_id
HAVING nplike = (SELECT COUNT(id) FROM photos)
```

OUTPUT:



| user_id | username | nplike |
|---|---|---|
| 5 | Aniya_Hackett | 257 |
| 14 | Jaclyn81 | 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 |

(The likes and users tables are joined and the uses are grouped by the number of likes they made and the users who have liked total number of photos posted are considered as fake accounts)

**Result:**

The project helped me in getting a clarity of the concepts I learned and it helped me understand the real world application of data analytics.