**Project Description**: This project aims to utilize advanced data management tool for transforming raw user data into actionable insights. Through comprehensive data analysis, we aim to discover essential trends and metrics that can enhance user experience and improve organizational performance.

**Approach:** This project utilized SQL to build a database from the raw data provided, followed by the implementation of sorting and extraction queries to derive the needed insights. These queries helped in generating valuable insights by highlighting key patterns, trends and helped in answering business-related questions from the data.

**Tech-Stack Used:** The Tech-Stack used is MYSQL Workbench Version 8.0.38.MySQL’s ease of setup and lower resource usage are advantageous for small to medium projects.

**Insights:**

* **Marketing Analysis**:

1. Individuals who have utilized the platform for the longest duration.

|  |  |  |
| --- | --- | --- |
| id | username | created\_at |
| 80 | Darby\_Herzog | 5/6/2016 0:14 |
| 67 | Emilio\_Bernier52 | 5/6/2016 13:04 |
| 63 | Elenor88 | 5/8/2016 1:30 |
| 95 | Nicole71 | 5/9/2016 17:30 |
| 38 | Jordyn.Jacobson2 | 5/14/2016 7:56 |

**Query:** SELECT \*

FROM users

ORDER BY created\_at ASC

LIMIT 5;

1. Users who have not posted any photographs on Instagram.

|  |
| --- |
| username |
| Aniya\_Hackett |
| Kasandra\_Homenick |
| Jaclyn81 |
| 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 |
| Bethany20 |

**Query**: SELECT username

FROM users

LEFT JOIN

photos

ON users.id = photos.user\_id

WHERE

photos.id IS NULL;

1. The user with the highest number of likes on a single photograph.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | username | photo\_id | most\_liked | image\_url |
| 52 | Zack\_Kemmer93 | 145 | 48 | <https://jarret.name> |

**Query**: WITH most\_liked\_photos AS

(SELECT photo\_id , COUNT(photo\_id) AS most\_liked FROM likes

GROUP BY photo\_id)

SELECT users.id,username,photo\_id,most\_liked,image\_url

FROM photos

INNER JOIN most\_liked\_photos

ON photos.id = most\_liked\_photos.photo\_id

INNER JOIN users

ON photos.user\_id = users.id

ORDER BY most\_liked DESC

LIMIT 1;

1. The five most frequently used hashtags on Instagram

|  |  |
| --- | --- |
| tag\_name | frequency |
| smile | 59 |
| beach | 42 |
| party | 39 |
| fun | 38 |
| food | 24 |

**Query**: WITH popular\_tag\_id AS

(SELECT tag\_id,COUNT(tag\_id) AS frequency FROM photo\_tags

GROUP BY tag\_id)

SELECT tags.tag\_name,popular\_tag\_id.frequency

FROM popular\_tag\_id

INNER JOIN tags

ON popular\_tag\_id.tag\_id = tags.id

ORDER BY frequency DESC

LIMIT 5;

1. The day of the week on which the highest number of users register on Instagram.

|  |  |
| --- | --- |
| day\_ | frequency |
| Thursday | 16 |
| Sunday | 16 |
| Friday | 15 |
| Tuesday | 14 |
| Monday | 14 |
| Wednesday | 13 |
| Saturday | 12 |

**Query**: SELECT DAYNAME(created\_at) AS day\_,

COUNT(DAYNAME(created\_at)) AS frequency

FROM users

GROUP BY day\_

ORDER BY frequency DESC;

* **Investor Metrics**:

1. Average number of posts per user

|  |
| --- |
| Average |
| 2.57 |

**Query**: SELECT round((SELECT COUNT(id)

FROM photos)

/ (SELECT COUNT(id)

FROM users),2)

AS Average;

1. Bots and fake accounts

|  |  |  |  |
| --- | --- | --- | --- |
| id | username | created\_at | no\_of\_likes |
| 5 | Aniya\_Hackett | 12/7/2016 1:04 | 257 |
| 14 | Jaclyn81 | 2/6/2017 23:29 | 257 |
| 21 | Rocio33 | 1/23/2017 11:51 | 257 |
| 24 | Maxwell.Halvorson | 4/18/2017 2:32 | 257 |
| 36 | Ollie\_Ledner37 | 8/4/2016 15:42 | 257 |
| 41 | Mckenna17 | 7/17/2016 17:25 | 257 |
| 54 | Duane60 | 12/21/2016 4:43 | 257 |
| 57 | Julien\_Schmidt | 2/2/2017 23:12 | 257 |
| 66 | Mike.Auer39 | 7/1/2016 17:36 | 257 |
| 71 | Nia\_Haag | 5/14/2016 15:38 | 257 |
| 75 | Leslie67 | 9/21/2016 5:14 | 257 |
| 76 | Janelle.Nikolaus81 | 7/21/2016 9:26 | 257 |
| 91 | Bethany20 | 6/3/2016 23:31 | 257 |

**Query**: WITH fake AS

(SELECT user\_id,COUNT(photo\_id)

AS no\_of\_likes

FROM likes

GROUP BY user\_id

HAVING no\_of\_likes = 257)

SELECT id, username, created\_at, no\_of\_likes

FROM users

INNER JOIN fake

ON users.id = fake.user\_id;

**Alternate query:**

WITH fake AS

(SELECT user\_id,COUNT(photo\_id)

AS no\_of\_likes

FROM likes

GROUP BY user\_id

HAVING no\_of\_likes

= (SELECT COUNT(\*) FROM photos))

SELECT id, username, created\_at, no\_of\_likes

FROM users

INNER JOIN fake

ON users.id = fake.user\_id;

**Results**:

* I have acquired substantial information regarding users on the platform, including the most popular day of the week for user registrations and the ratio of loyal users to fraudulent users and of active to inactive users. This information can help in identifying and removing fake accounts, implementing targeted strategies to engage inactive users, and rewarding active users for their participation.
* Gathering data on popular hashtags and the average number of posts per user enables the platform to identify emerging trends and patterns. This approach helps in anticipating future events and allows the platform to implement well-informed, strategic actions based on historical and current data.