**Objective Questions**

1. Are there any tables with duplicate or missing null values? If so, how would you handle them?

**DUPLICATE CHECKING:**  
 **Function used: Distinct**   
 To find the duplicates I used **Distinct** for each and every column

To handle the duplicates we need to use distinct while we execute our further queries.

There are few tables with duplicate values

**OBSERVATION:**  
 **comments table**

**User\_id** - There were total **7488** user id’s returned but when checked for duplicates **77** user id’s returned

**Photo\_id**-There were total **7488** photo id’s returned but when checked for duplicates **257** photo id’s returned.

**Follows table  
Follower\_id -** There were total **7623** follower id’s returned but when checked for duplicates **77** follower id’s returned.

**Followee\_id-** there were total **7623** followee id’s returned but when checked for duplicates **100** followee id’s returned.  
  
 **Likes table**  
**user\_id** - there were total **8782** user id’s returned but when checked for duplicates **77** user id’s returned.  
**Photo\_id**-there were total **8782** photo id’s returned but when checked for duplicates **257** photo id’s returned.

**Photo\_tags table  
photo\_id -** there were total **501** photo id’s returned but when checked for duplicates **190** photo id’s returned.  
**Tag\_id-** there were total **501** tag id’s returned but when checked for duplicates **21** tag id’s returned.

**Photos table   
photo\_id -** there were total **257** photo id’s returned but when checked for duplicates **74** photo id’s returned.  
No duplicates found in tags and users table.  
  
**NULL CHECKING:  
operation used: IS NULL  
  
OBSERVATION:   
 I observed that there are no null values in the tables   
but when we are joining the tables we can observe null values, here we can use COALESCE to give value to the null.**

1. What is the distribution of user activity levels (e.g., number of posts, likes, comments) across the user base?

**Visualization:**

**OBSERVATION:**  It is observed that the users with highest likes and comments are having zero posts as per the query.  
 And the users with highest no of posts are having zero likes and comments as per the query   
 In the above graph we can find the users with moderate number of likes, moderate number of comments and moderate no of posts.

1. Calculate the average number of tags per post (photo\_tags and photos tables).

|  |
| --- |
| **avg\_tags\_per\_post** |
| 2 |

First I have calculated the total number of tags for each post.Then calculated the average of these counts across all posts.  
This means that, on average, each post has 2 tags..

1. Identify the top users with the highest engagement rates (likes, comments) on their posts and rank them.

|  |  |
| --- | --- |
| **UserID** | **username** |
| 96 | Keenan.Schamberger60 |
| 87 | Rick29 |
| 69 | Karley\_Bosco |
| 26 | Josianne.Friesen |
| 63 | Elenor88 |
| 67 | Emilio\_Bernier52 |
| 17 | Norbert\_Carroll35 |
| 15 | Billy52 |
| 2 | Andre\_Purdy85 |
| 35 | Lennie\_Hartmann40 |
| 43 | Janet.Armstrong |
| 22 | Kenneth64 |
| 60 | Sam52 |
| 4 | Arely\_Bogan63 |
| 70 | Erick5 |
| 65 | Adelle96 |

**Visualization:**

**OBSERVATION:** The above mentioned users are having the highest likes and comments on their posts.  
I also observed that there are users with highest likes and comments but with Zero posts.  
Here I have considered only the users with the highest likes and comments which have no of posts.

1. Which users have the highest number of followers and followings?

**Visualization:**

**OBSERVATION:** The above mentioned users are having the highest number of followers and followings  
It is also observed that these users are having highest number of followers than followings.

1. Calculate the average engagement rate (likes, comments) per post for each user.

**Visualization:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| id | username | count\_likes | count\_coments | count\_posts | avg\_engagement\_rate\_per\_post |
| 55 | Meggie\_Doyle | 41 | 34 | 1 | 75 |
| 73 | Jaylan.Lakin | 38 | 35 | 1 | 73 |
| 48 | Granville\_Kutch | 37 | 34 | 1 | 71 |
| 22 | Kenneth64 | 39 | 31 | 1 | 70 |
| 69 | Karley\_Bosco | 36 | 32 | 1 | 68 |
| 94 | Damon35 | 40 | 28 | 1 | 68 |
| 87 | Rick29 | 140 | 132 | 4 | 68 |
| 18 | Odessa2 | 36 | 31 | 1 | 67 |
| 43 | Janet.Armstrong | 180 | 154 | 5 | 66.8 |
| 52 | Zack\_Kemmer93 | 182 | 151 | 5 | 66.6 |

**OBSERAVTION:** The above mentioned users are having highest engagement rate, it means these users are active interacting with the content

From the above graph we observe that user 55 is having highest engagement rate.

1. Get the list of users who have never liked any post (users and likes tables)

**Visualization:**

|  |  |
| --- | --- |
| **user\_id** | **User\_name** |
| 1 | Kenton\_Kirlin |
| 7 | Kasandra\_Homenick |
| 23 | Eveline95 |
| 25 | Tierra.Trantow |
| 29 | Jaime53 |
| 34 | Pearl7 |
| 45 | David.Osinski47 |
| 49 | Morgan.Kassulke |
| 51 | Mariano\_Koch3 |
| 53 | Linnea59 |
| 58 | Aurelie71 |
| 59 | Cesar93 |
| 64 | Florence99 |
| 68 | Franco\_Keebler64 |
| 74 | Hulda.Macejkovic |
| 77 | Donald.Fritsch |
| 80 | Darby\_Herzog |
| 81 | Esther.Zulauf61 |
| 83 | Bartholome.Bernhard |
| 86 | Delfina\_VonRueden68 |
| 88 | Clint27 |
| 89 | Jessyca\_West |
| 90 | Esmeralda.Mraz57 |

**Observation:**

Above table shows the users who have never liked any post (users and likes tables). It is observed that there are total 23 users who have never liked any post.

1. How can you leverage user-generated content (posts, hashtags, photo tags) to create more personalized and engaging ad campaigns?

**Visualization:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| user\_id | user\_name | tag\_name | hashtag\_engagement | photo\_tag\_engagement | total\_engagement |
| 88 | Clint27 | beach | 6 | 26 | 32 |
| 88 | Clint27 | dreamy | 4 | 26 | 30 |
| 23 | Eveline95 | smile | 5 | 24 | 29 |
| 88 | Clint27 | sunset | 3 | 26 | 29 |
| 88 | Clint27 | sunrise | 3 | 26 | 29 |

**OBSERAVTION:** From the above the above we can observe that tag name Beach is having the highest engagement.With the above we can create more personalized and engaging ad campaigns by Creating Targeted Content  
Utilizing User-Generated Content

Leveraging Influencers

Arranging Interactive Campaigns  
 **Examples:**

1. **Beach Campaign**: "Beach Ready" promotion offering discounts on summer collections and highlighting user photos tagged at various beach locations.
2. **Fun Campaign**: "Capture the Fun" video challenge where users share short clips of their most enjoyable moments using your product, with winners featured in your next ad.
3. **Party Campaign**: "Party with [Brand]" event series promoted on social media, featuring party essentials and tips for the perfect celebration.
4. **Smile Campaign**: "Share Your Best Smile" contest where users post their smiling photos with a specific hash tag for a chance to win a year's supply of dental care products.
5. **Beauty Campaign**: "Beauty Tips" series with influencers providing tutorials and reviews, encouraging users to share their own beauty routines with your products.
6. Are there any correlations between user activity levels and specific content types (e.g., photos, videos, reels)? How can this information guide content creation and curation strategies?

**OBSERAVTION:**there is a weak negative correlation (-0.10) between total tags and total comments, indicating a slight tendency for posts with more tags to have fewer comments, and vice versa.

Strategic Insights:

Balance Tagging and Commenting: Since tags and comments slightly counteract each other, create content that encourages both actions without overwhelming users.

Differentiate Content Types: Use tags on content meant for broad engagement and comments on more discussion-driven posts. For example, tags should be used on promotional content and comments should be encouraged on posts with questions or prompts.

Target Specific Audiences: Identify user segments that prefer tagging versus commenting, and tailor your content strategies accordingly.

.

1. Calculate the total number of likes, comments, and photo tags for each user.

**Visualization:**

**OBSERVATION**:  
  
-It is observed that the users with highest likes and comments are having zero photo tags as per the query.  
  
-The users with the highest no of photo tags have zero likes and comments as per the query   
  
-In the above graph some of the users are mentioned with the number of likes,number of comments and number of hast tags.

1. Rank users based on their total engagement (likes, comments, shares) over a month.

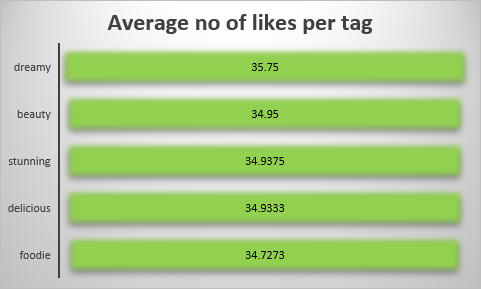
**Visualization:**

**OBSERVATION**:  
 Results shows users ranked by their total engagement over June 2024.  
  
By ranking users based on their engagement, we can identify which users are the most active and influential on the platform. These users are valuable because they are more likely to interact with and spread content, thereby increasing its reach.

1. Retrieve the hashtags that have been used in posts with the highest average number of likes. Use a CTE to calculate the average likes for each hashtag first.

|  |  |  |
| --- | --- | --- |
| tag\_id | tag\_name | avg\_likes |
| 10 | dreamy | 35.75 |
| 8 | beauty | 34.95 |
| 9 | stunning | 34.9375 |
| 7 | delicious | 34.9333 |
| 6 | foodie | 34.7273 |

**Visualization:**



**Observation:** Using CTE we have calculated the Average likes for each hashtag first then after that we retrieve the hashtags that have been used in posts with the highest average number of likes.

It is observed that **‘dreamy, beauty, stunning, delicious, foodie'** tag names are having highest average likes.

1. Retrieve the users who have started following someone after being followed by that person

**Visualization:**

|  |
| --- |
| USER\_ID |
| Andre\_Purdy85 |
| Harley\_Lind18 |
| Arely\_Bogan63 |
| Aniya\_Hackett |
| Travon.Waters |
| Tabitha\_Schamberger11 |
| Gus93 |
| Presley\_McClure |
| Justina.Gaylord27 |
| Dereck65 |
|  |

**Observation:**

Above list shows the users who have started following someone after being followed by that person.It is observed that most of the users are having the same number of followers.  
 This can be useful for:  
 -Understanding Mutual Engagement  
 -Community Insights  
 -**User Engagement Analysis  
 -**Targeted Communication  
 -Community Management

**Subjective Question:**

1. Based on user engagement and activity levels, which users would you consider the most loyal or valuable? How would you reward or incentivize these users?

**Visualization:**

**Recommendations:**

* Using Sql query we found out the users which are most loyal or valuable based on user engagement and activity levels.
* These are the top 10 loyal users with high engagement activity.
* We can reward or incentivize these users by providing access to exclusive content, early access to new features, or sneak peeks of upcoming products, send personalized thank-you messages or notes appreciating their engagement and loyalty, send them branded merchandise or gifts as a token of appreciation.

1. For inactive users, what strategies would you recommend to re-engage them and encourage them to start posting or engaging again?

**Recommendations:**

* The strategies that can be recommended to re-engage the inactive users can be Send personalized emails with subject lines like "We Miss You!" or "Come Back for Something Special!".
* Use engaging and personalized push notifications or in-app messages to remind users of what they’re missing out on.
* Highlight recent activity, popular posts, or community achievements to spark their interest.
* Use data-driven algorithms to recommend content, posts, or users they might be interested in based on their past behavior. Highlight trending posts or popular content in their areas of interest.
* Highlight the benefits of being active in the community and any new, engaging discussions or events.

1. Which hashtags or content topics have the highest engagement rates? How can this information guide content strategy and ad campaigns?

**Visualization:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| tag\_name | likes\_count | comment\_count | post\_count | engagement\_rate |
| smile | 77 | 1725 | 59 | 30.5424 |
| beach | 77 | 1216 | 42 | 30.7857 |
| party | 77 | 1151 | 39 | 31.4872 |
| fun | 77 | 1089 | 38 | 30.6842 |
| food | 77 | 727 | 24 | 33.5 |
| lol | 77 | 691 | 24 | 32 |
| concert | 77 | 676 | 24 | 31.375 |
| hair | 77 | 652 | 23 | 31.6957 |
| happy | 77 | 625 | 22 | 31.9091 |
| beauty | 77 | 604 | 20 | 34.05 |
| dreamy | 77 | 563 | 20 | 32 |
| sunset | 77 | 574 | 19 | 34.2632 |
| drunk | 77 | 551 | 19 | 33.0526 |
| fashion | 77 | 548 | 19 | 32.8947 |
| landscape | 77 | 505 | 17 | 34.2353 |
| sunrise | 76 | 497 | 17 | 33.7059 |
| style | 77 | 490 | 17 | 33.3529 |
| photography | 77 | 472 | 16 | 34.3125 |
| stunning | 77 | 469 | 16 | 34.125 |
| delicious | 77 | 455 | 15 | 35.4667 |
| foodie | 77 | 329 | 11 | 36.9091 |

The top hast tags are :  
-smile  
-beach  
-party  
-fun

**Recommendation:**

* Above hashtags or content topics have the highest engagement rates this information can guide content strategy and ad campaigns.
* Create content that centers around the identified high-engagement hashtags or topics. This ensures that the content resonates well with the audience and drives higher interaction.
* Encourage users to create and share content using these popular hashtags to increase organic reach and engagements.
* Use high-engagement hashtags in your ad campaigns to target users who are more likely to interact with those topics. This can improve ad relevance and effectiveness.
* Partner with influencers who frequently use these high-engagement hashtags to amplify your campaign’s reach and authenticity

1. Are there any patterns or trends in user engagement based on demographics (age, location, gender) or posting times? How can these insights inform targeted marketing campaigns?

**Recommendations:**

* Identify which age groups have the highest engagement rates. Tailor content and ad campaigns to appeal to these age groups. For instance, if younger users are more engaged, use trends and language that resonate with them.
* Determine which locations have the highest engagement. Focus marketing efforts on these regions. Use localized content, offers, and promotions to increase relevance and engagement.
* Analyze engagement differences between genders. Create gender-specific campaigns if there's a significant difference. For example, if female users are more engaged, develop campaigns that cater to their interests and preferences.
* Find out the times of day when engagement is highest. Schedule posts and ad campaigns during peak engagement hours to maximize visibility and interaction. Use insights to inform the timing of email campaigns and social media posts.

1. Based on follower counts and engagement rates, which users would be ideal candidates for influencer marketing campaigns? How would you approach and collaborate with these influencers?

**Visualization:**

|  |  |  |  |
| --- | --- | --- | --- |
| user\_id | user\_name | follower\_count | total\_engagement |
| 96 | Keenan.Schamberger60 | 76 | 252 |
| 87 | Rick29 | 76 | 246 |
| 26 | Josianne.Friesen | 76 | 244 |
| 63 | Elenor88 | 76 | 243 |
| 69 | Karley\_Bosco | 76 | 243 |
| 15 | Billy52 | 76 | 241 |
| 67 | Emilio\_Bernier52 | 76 | 241 |
| 2 | Andre\_Purdy85 | 76 | 240 |
| 17 | Norbert\_Carroll35 | 76 | 240 |
| 43 | Janet.Armstrong | 76 | 239 |
|  |  |  |  |

**Recommendation:**

* According to follower counts and engagement rates the above users will be ideal candidates for influencer marketing campaigns.
* Ensure the influencer's content aligns with your brand values and target audience. Review their posts to see if their style and audience demographics match your brand.
* Send a personalized message expressing genuine interest in their content and explaining why you believe they would be a great fit for your campaign.
* Provide detailed information about the campaign, including objectives, timeline, key messages, and expectations.
* Allow influencers some creative freedom to ensure authenticity. Their followers trust their voice and content style.

1. Based on user behavior and engagement data, how would you segment the user base for targeted marketing campaigns or personalized recommendations?

**Recommendations:**

* Based on user behaviour and engagement data, we can segment the user base where engagement is greater than 300 we can specify it as “highly engaged” ,where engagement is between 150 to 300 then “moderately engaged” or else “low engaged”.
* For highly engaged user we can offer exclusive content or early access to new features, encourage them to become brand ambassadors or influencers, provide personalize recommendations based on their interaction history.
* For moderately engaged user we can send notifications and updates to keep them engaged, provide personalized content suggestions to increase engagement.
* For low engaged users we can offer tutorials or tips to help them engage more with the platform, highlight popular or trending content to spark interest.

1. If data on ad campaigns (impressions, clicks, conversions) is available, how would you measure their effectiveness and optimize future campaigns?

**Recommendations:**

Measuring the effectiveness of ad campaigns (impressions, clicks, conversions) is crucial for optimizing future campaigns  
**Key metrics:  
Impressions:** Track how many times the ad is displayed to users.  
**Clicks:** Measure the number of times users click on the ad to visit the website or landing page.  
**Conversions:** Track actions that align with the campaign goals, such as sign-ups, purchases, or downloads.

**Calculate Performance Metrics:**

**Click-Through Rate (CTR):** CTR = (Clicks / Impressions) \* 100%. Indicates how engaging the ad is.  
**Conversion Rate:** Conversion Rate = (Conversions / Clicks) \* 100%. Measures how effective an ad is at converting clicks into desired actions.  
  
**Attribution models**:Use attribution models (e.g., first-click, last-click, multi-touch) to understand which ad interactions contributed most to conversions.

1. How can you use user activity data to identify potential brand ambassadors or advocates who could help promote Instagram's initiatives or events?

**Visualization:**

**Recommendations:**

So here we already know who are our top users with the user engagement levels data which has been already explained .  
-so the first step is to identify the potential brand ambassadors or advocates.

**We can identify them by analyzing the engagement data**

-For example the users who consistently engage with your content. Look for those who frequently like, comment, share, or save posts related to your brand.  
-Users who regularly leave thoughtful or enthusiastic comments can be strong advocates.  
-We can also choose the users who are actively promoting the brand by using specific hash tags.  
-We need to choose the influencers with highest followers so the reach will be more.  
-Identify users who participate in brand-related events, contests, or challenges. This indicates enthusiasm and loyalty to the brand.  
-We can also focus on these kind of users  
-Users with the highest average likes and comments on brand-related posts  
-Users who frequently share or save your content.  
-Users who often mention or tag your brand in their posts.

**This is how we can promote**

* Contact potential ambassadors with personalized messages highlighting their engagement and influence.
* We can look for users with high engagement, influence, frequent activity, and strong brand affinity.
* Offer incentives such as exclusive access, discounts, or monetary compensation.
* Involve them in collaborative campaigns, allowing them to co-create content.

1. How would you approach this problem, if the objective and subjective questions weren't given?

If the objective and subjective questions were not given.

1.Understand the data   
 · **Examine the Structure**: I will for the check the columns, data types, and a few sample rows to get an idea of what the data set contains.  
 · Checking for the data types of all the columns.

· I will refer to the schema in MYSQL server.

2.**Data Cleaning**:  
 · **Handle Missing Values**: Identify and address any missing values.

· **Check** **Duplicates**: Check for the duplicate rows and values.  
 If duplicates were found, using **Distinct** for the further queries to get distinct values.  
 And also checking for null values, if null values found I will use **coalesce** function wherever it is required.  
  
 3.Writing SQL queries to retrieve the data using   
 · Basic queries such as SELECT, WHERE, JOIN, GROUP BY, ORDER BY

· Advance queries such as window functions,CTEs, Sub queries.

4.writing a query to get the most active users to reward them

5.Writing a query to know what day of the week most users register on

6.Finding the users who have liked every single photo.

7.I want to target the inactive users so we can encourage them to re engage with an email campaign

8.I want to know which hash tag has been the most popular one.  
   
 9.I want to reward the users who have been for the longest period.  
  
 10.With the active user data, I will make the most loyal and active users as the brand ambassadors

1. Assuming there's a "User\_Interactions" table tracking user engagements, how can you update the "Engagement\_Type" column to change all instances of "Like" to "Heart" to align with Instagram's terminology?

**UPDATE** User\_Interactions

**SET** Engagement\_Type = 'Heart'

**WHERE** Engagement\_Type = 'Like';

**Recommendations:** We can use update function to update the "User\_Interactions" table where we can use a filter “where” to update like to heart.