# Social Media Analysis

PRESENTED BY ANIKET BOBDE

Problem Statement

YOU ARE HIRED AS A DATA ANALYST AT META AND ASKED TO COLLABORATE WITH MARKETING TEAM. MARKETING TEAMS WANTS TO LEVERAGE INSTAGRAM'S USER DATA TO DEVELOP TARGETED MARKETING STRATEGIES THAT WILL INCREASE USER ENGAGEMENT, RETENTION, AND ACQUISITION. PROVIDE INSIGHTS AND RECOMMENDATIONS TO ADDRESS THE FOLLOWING OBJECTIVES

Agenda

- **DATA QUALITY AND INTEGRITY**

- **KEY METRICS AND INSIGHTS**

- **STRATEGIC RECOMMENDATIONS**

IG Clone
Data Schema
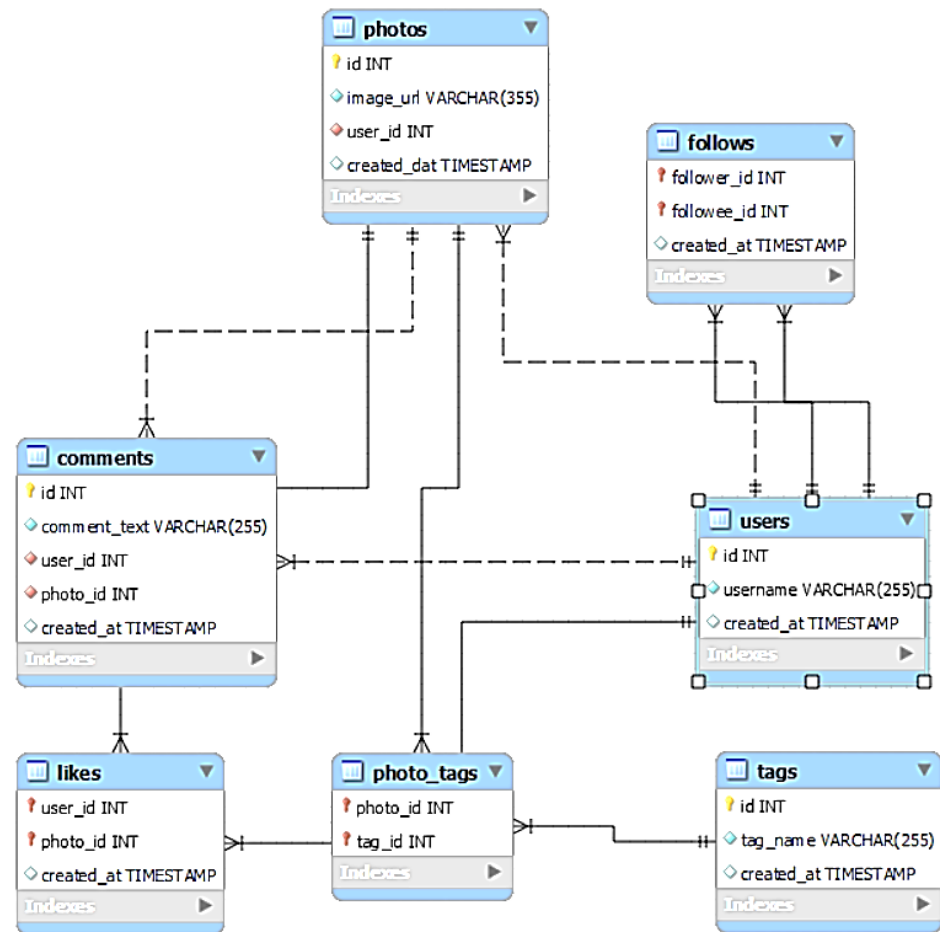
Data Quality & Integrity – Duplicate and Null Checks

Ensuring Data Integrity Across Tables

-View Describe Outputs for each table

-Schema constraints (PRIMARY KEY, NOT NULL, UNIQUE) ensure high data quality

-Recommended further checks

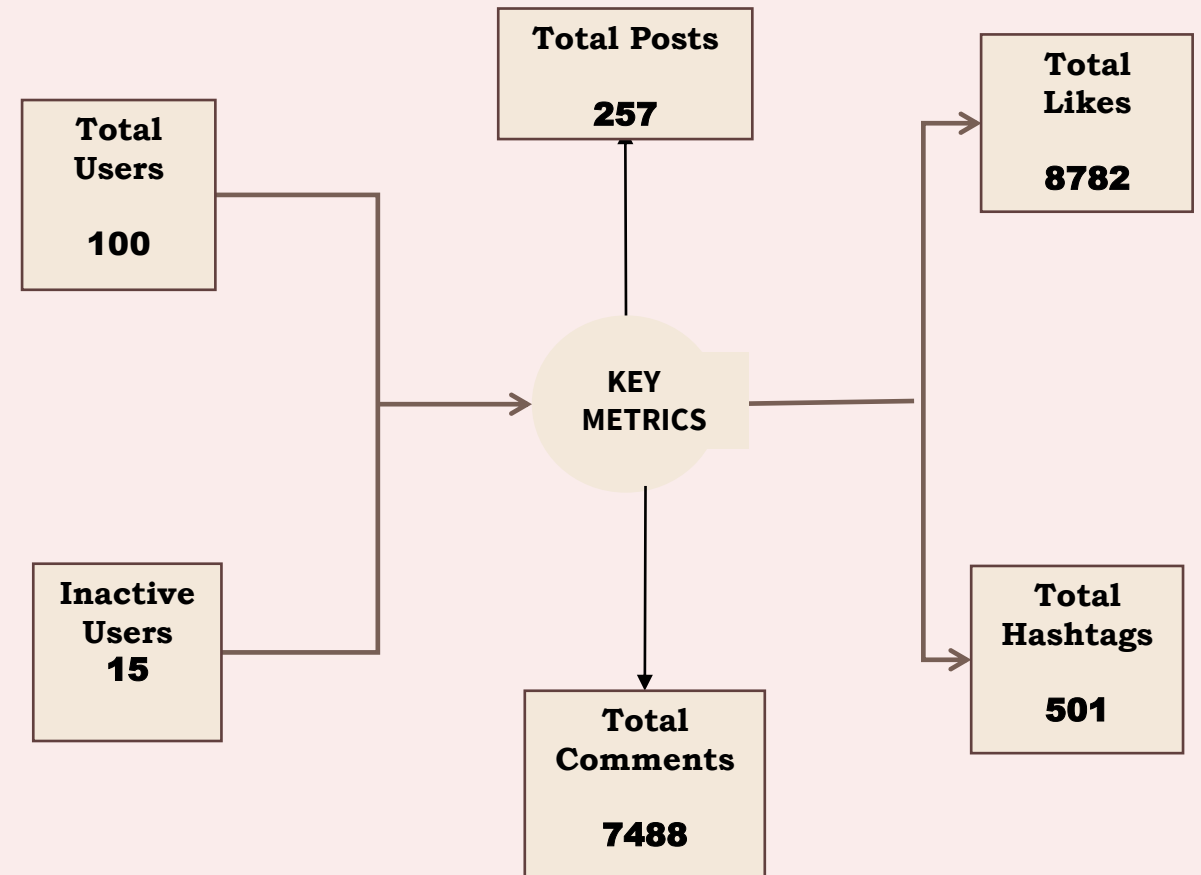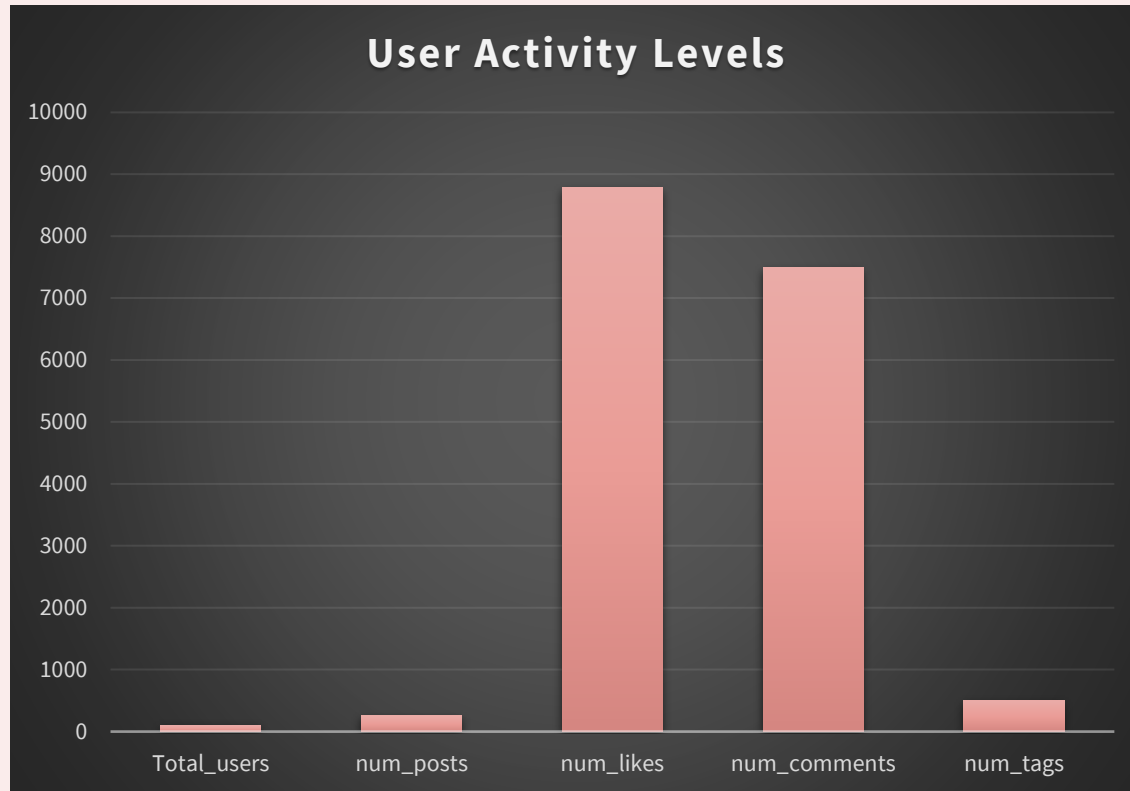| Table	Column	Inherent Constraint	Reason & Need to Check
users	id	PRIMARY KEY (unique, not null)	Duplicates and nulls are prevented by the primary key.
	username	NOT NULL	Nulls are not allowed; however, there's no unique constraint so duplicate usernames could occur if that is undesired. Check for duplicates if uniqueness is a business requirement.
	created_at	DEFAULT CURRENT_TIMESTAMP	Automatically set; check for nulls only if business rules demand a timestamp always be present.
photos	id	PRIMARY KEY	Duplicates and nulls prevented.
	image_url	NOT NULL	Null values are disallowed. No uniqueness constraint is defined so duplicates might exist. Check if duplicate URLs are problematic.
	user_id	NOT NULL, FOREIGN KEY	Must match an existing user. Multiple photos per user are allowed, so duplicates are expected.
	created_dat	DEFAULT CURRENT_TIMESTAMP	As with timestamps, nulls are unlikely but can be checked if needed.
comments	id	PRIMARY KEY	Duplicates and nulls prevented.
	comment_text	NOT NULL	Nulls disallowed; duplicates might not be errors unless business rules require unique comments for a given user/photo.
	user_id	NOT NULL, FOREIGN KEY	Multiple comments by the same user are allowed; just ensure no nulls.
	photo_id	NOT NULL, FOREIGN KEY	Must always reference an existing photo; duplicates allowed.
	created_at	DEFAULT CURRENT_TIMESTAMP	Automatically set.
likes	user_id, photo_id	Composite PRIMARY KEY	The composite PK prevents duplicate (user_id, photo_id) pairs. Nulls are disallowed by NOT NULL.
	created_at	DEFAULT CURRENT_TIMESTAMP	Automatically set.
follows	follower_id, followee_id	Composite PRIMARY KEY	Prevents duplicate follow relationships; null values are not allowed.
	created_at	DEFAULT CURRENT_TIMESTAMP	Automatically set.
tags	id	PRIMARY KEY	No duplicate ids; nulls prevented.
	tag_name	UNIQUE, NOT NULL	Uniqueness is enforced and null values are disallowed.
	created_at	DEFAULT CURRENT_TIMESTAMP	Automatically set.
photo_tags	photo_id, tag_id	Composite PRIMARY KEY	Prevents duplicate associations; nulls are disallowed.

Query Outputs for recommended checks

Since primary keys, NOT NULL, and UNIQUE constraints are already defined, no duplicates or nulls are expected in those columns

Table	Check Type	Query Description	Query Output
users	Duplicate Check	SELECT username, COUNT(*) AS cnt FROM users GROUP BY username HAVING cnt > 1;	No rows returned
	Null Check	SELECT * FROM users WHERE username IS NULL;	No rows returned
photos	Duplicate Check	SELECT image_url, COUNT(*) AS cnt FROM photos GROUP BY image_url HAVING cnt > 1;	No rows returned
	Null Check	SELECT * FROM photos WHERE image_url IS NULL OR user_id IS NULL;	No rows returned
comments	Duplicate Check	SELECT user_id, photo_id, comment_text, COUNT(*) AS cnt FROM comments GROUP BY user_id, photo_id, comment_text HAVING cnt > 1;	No rows returned
	Null Check	SELECT * FROM comments WHERE comment_text IS NULL OR user_id IS NULL OR photo_id IS NULL;	No rows returned
likes	Null Check	SELECT * FROM likes WHERE user_id IS NULL OR photo_id IS NULL;	No rows returned
follows	Null Check	SELECT * FROM follows WHERE follower_id IS NULL OR followee_id IS NULL;	No rows returned
tags	Null Check	SELECT * FROM tags WHERE tag_name IS NULL;	No rows returned
photo_tags	Null Check	SELECT * FROM photo_tags WHERE photo_id IS NULL OR tag_id IS NULL;	No rows returned

User Activity Distribution

Key Metrics

User Activity Levels

Bar chart categories: Total_users, num_posts, num_likes, num_comments, num_tags (y-axis 0 to 10000)

Total Users 100

Inactive Users 15

Total Posts 257

KEY METRICS

Total Likes 8782

Total Comments 7488

Total Hashtags 501

Content Tagging Behaviour

AVERAGE TAGS PER POST = 1.95

- USERS TYPICALLY USE ABOUT TWO TAGS PER POST.

- WE CAN PROMPT INITIATIVES TO ENCOURAGE RICHER TAGGING, IMPROVING CONTENT DISCOVERABILITY.

Followers and Followings Dynamics

HIGHEST FOLLOWERS COUNT = 77

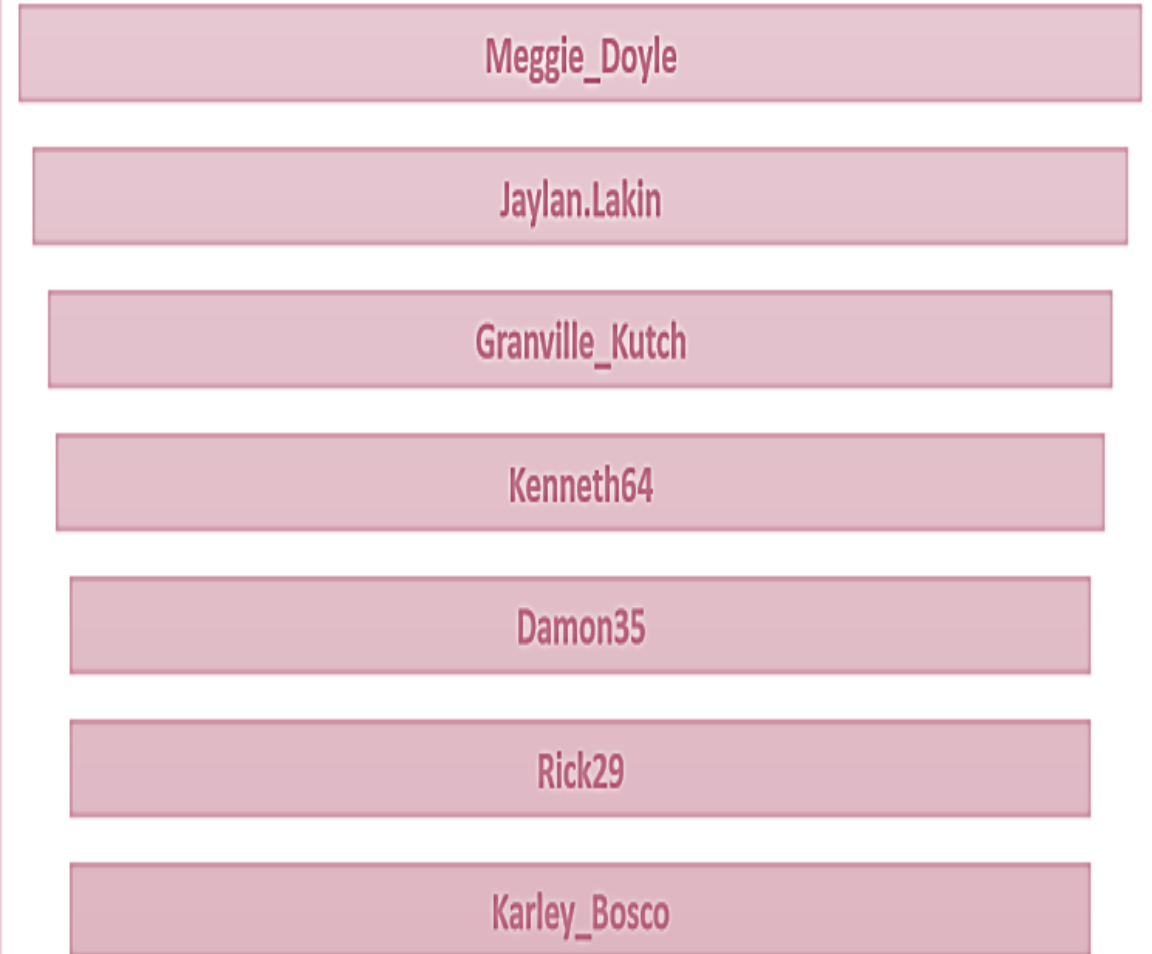**COUNT OF USERS WITH HIGHEST FOLLOWERS = 23**

HIGHEST FOLLOWING COUNT = 99

COUNT OF USERS WITH HIGHEST FOLLOWINGS = 77
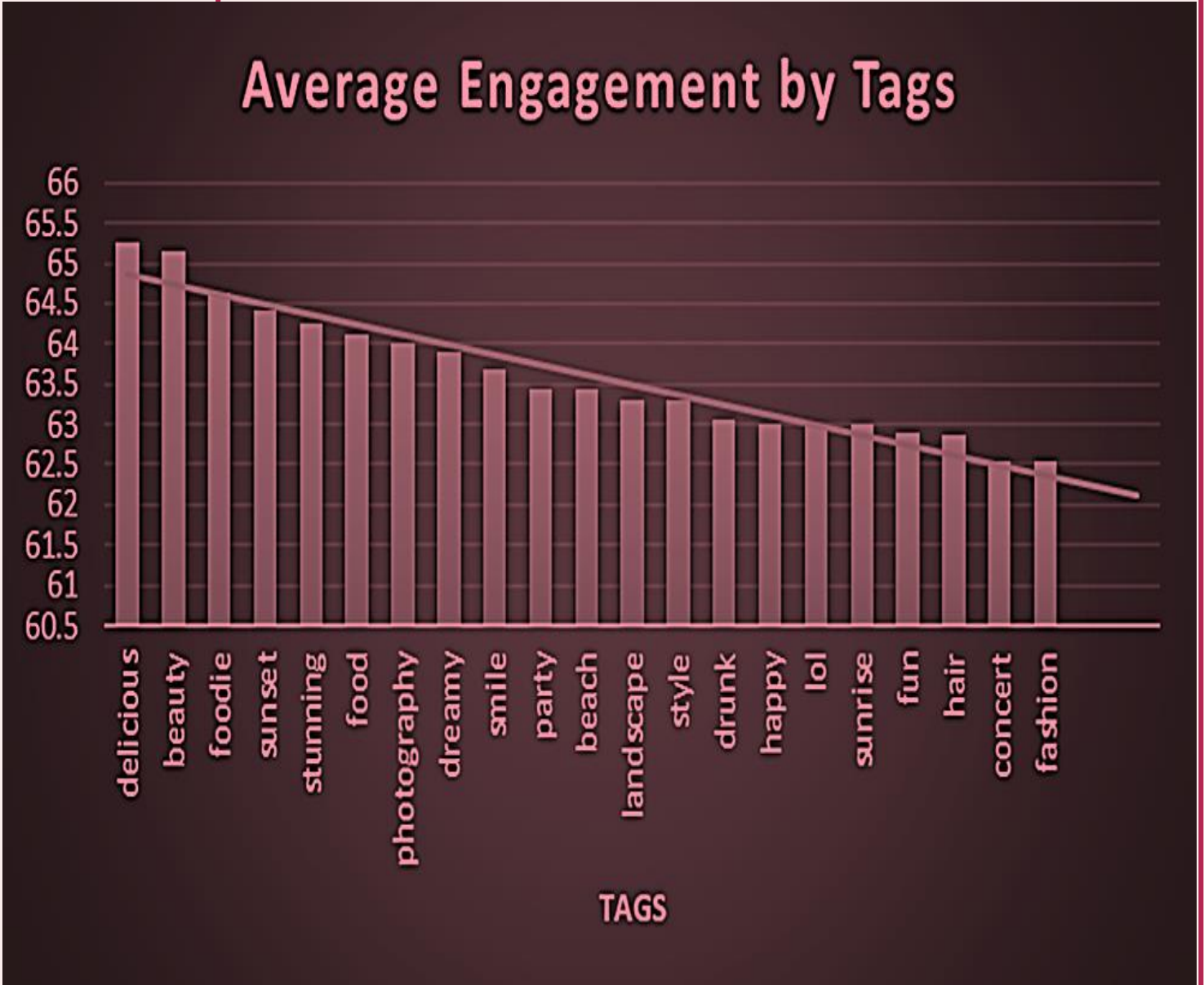
Top Users by Engagement Rate

- Engagement rate is defined as (total likes + total comments) per post.

Users with Top Engagement

- Meggie_Doyle
- Jaylan.Lakin
- Granville_Kutch
- Kenneth64
- Damon35
- Rick29
- Karley_Bosco

Hashtag Engagement

tag_name	avg_engagement
delicious	65.27
beauty	65.15
foodie	64.64
sunset	64.42
stunning	64.25
food	64.13
photography	64
dreamy	63.9
smile	63.69
party	63.44
beach	63.43
landscape	63.29
style	63.29
drunk	63.05
happy	63
lol	63
sunrise	63
fun	62.89
hair	62.87
concert	62.54
fashion	62.53

Average Engagement by Tags

Strategic Recommondations
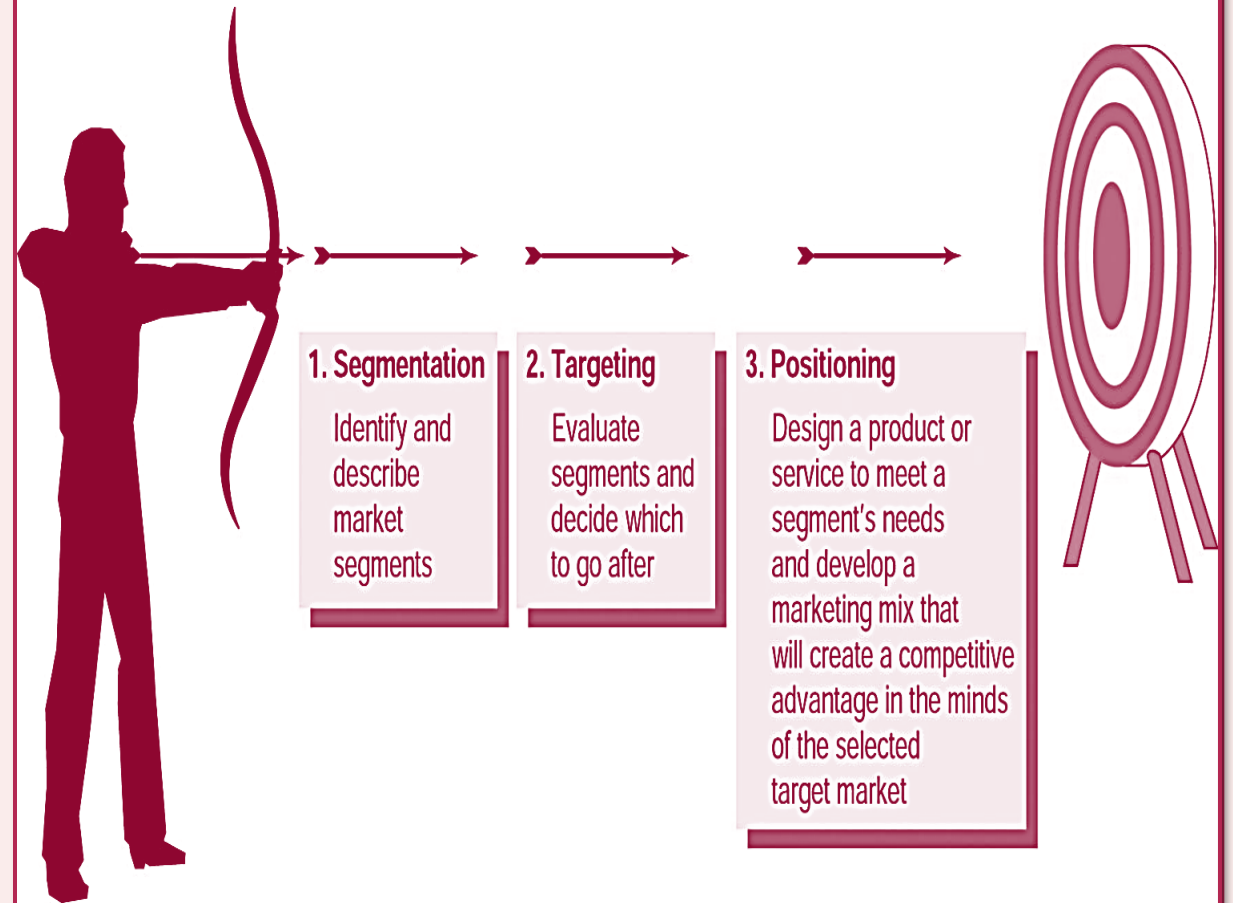
Targeted Marketing

Segmentation Criteria:

- **Activity Levels:** Active vs inactive users.

- **Engagement Rates:** High vs low interaction users.

- **Content Preferences:** Users engaging with specific types of posts (e.g., photos vs videos).

- **Demographics:** Age, location, gender-based segmentation

Implementation:

- Use data-driven insights to create personalized recommendations (e.g., curated hashtags).

- Tailor marketing campaigns focusing on each segment's preferences.

Target Marketing Process:
Linking Customer Needs to Marketing Action

1. Segmentation
Identify and describe market segments

2. Targeting
Evaluate segments and decide which to go after

3. Positioning
Design a product or service to meet a segment's needs and develop a marketing mix that will create a competitive advantage in the minds of the selected target market

Loyal and Valuable Users

- **Definition**: Total Engagement over last month >200

Reward Strategies:

- Offer exclusive perks such as early access to features, discounts, or badges.

- Create personalized campaigns highlighting their contributions (e.g., "Top Contributor of the Month").

- Provide opportunities for collaboration, such as co-creating content or becoming brand ambassadors.
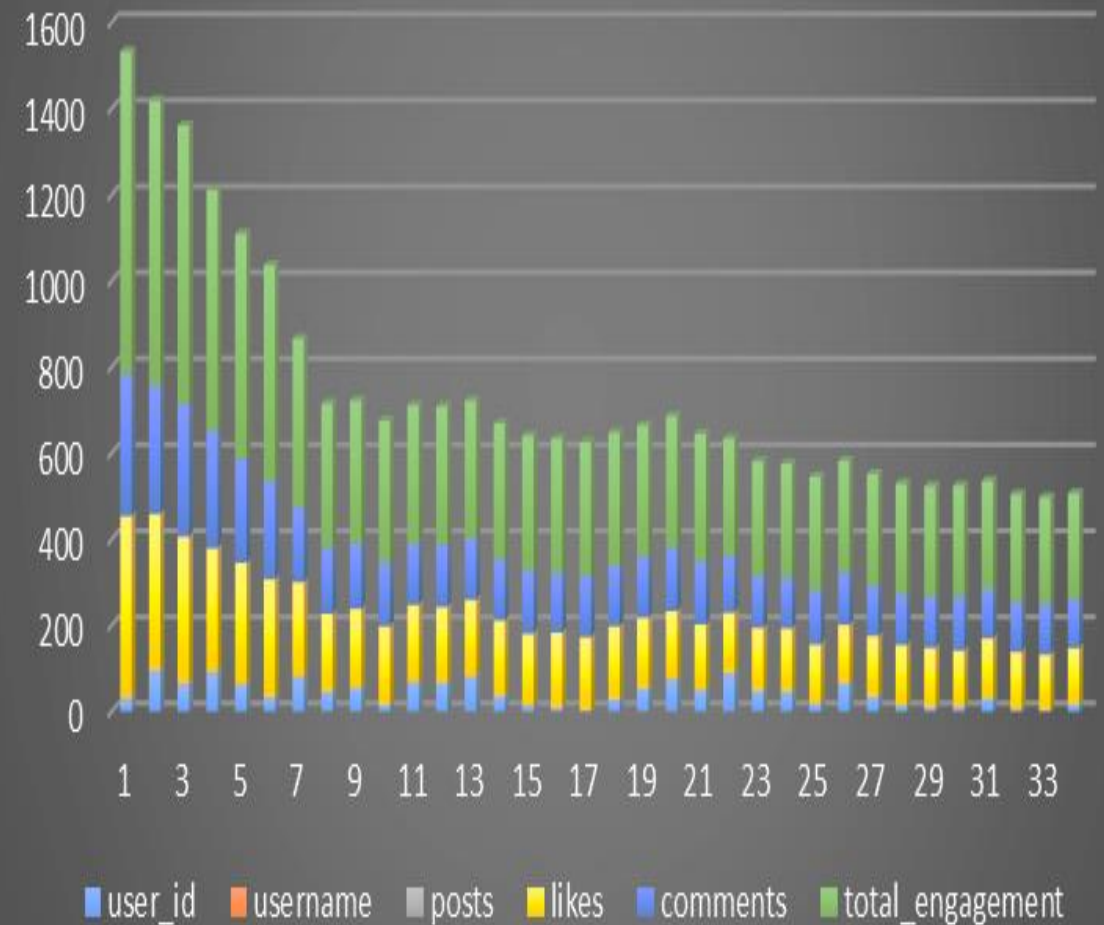
Inactive Users

- **Definition**: Total engagement is 0

Re-engagement Strategies:

1. **Personalized Outreach:** Send tailored messages with recommendations based on their past activity.

2. **Gamification:** Introduce rewards for engagement milestones (e.g., badges for first post after inactivity).

3. **Humor and Creativity:** Use playful content like jokes or memes to spark interest3.

4. **Exclusive Offers:** Provide incentives such as free trials, discounts, or giveaways for returning users.

5. **Targeted Campaigns:** Highlight trending topics or hashtags to encourage participation.

User Activity Status
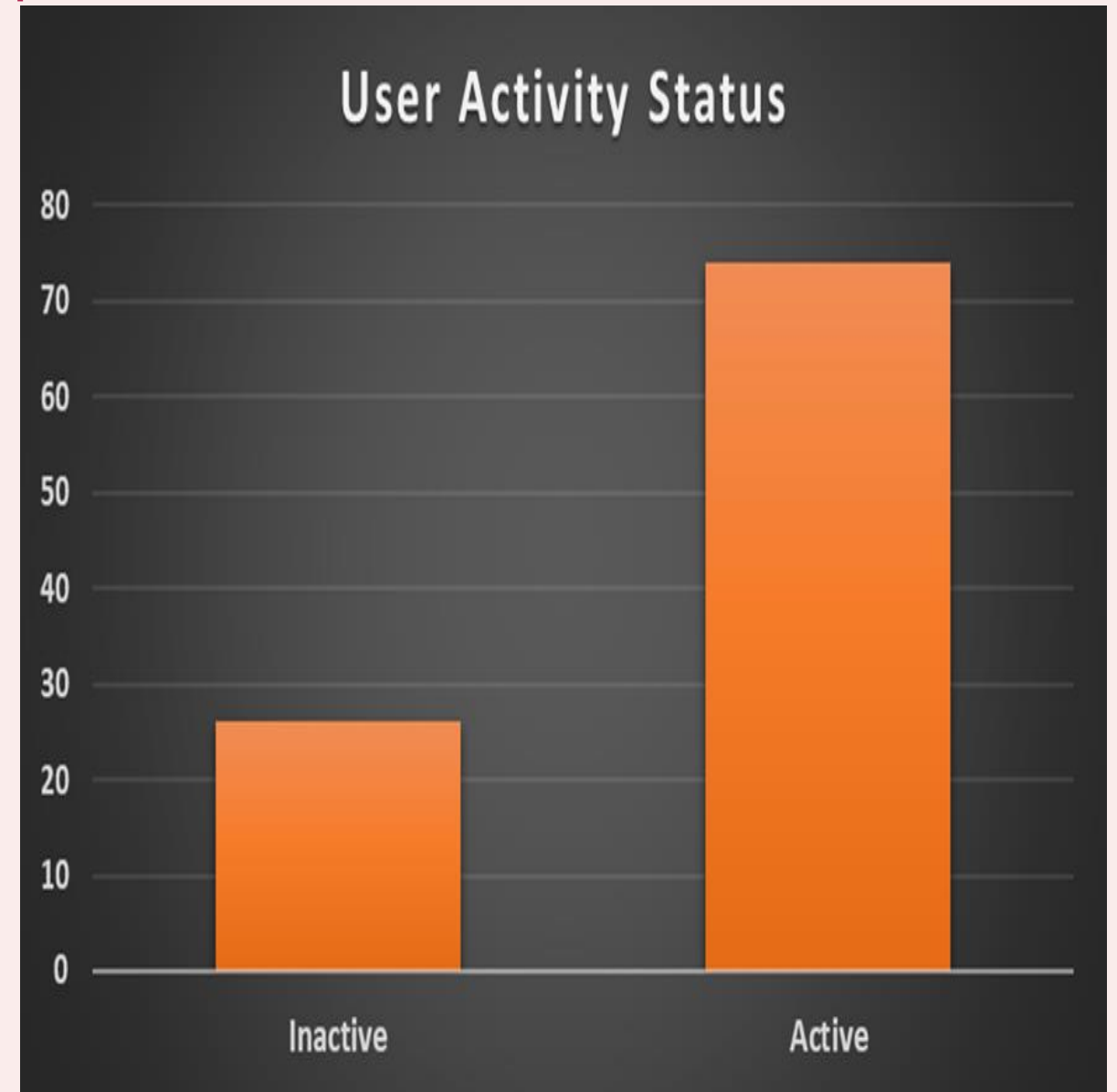
Hashtag Content Strategy

Insights from Hashtags:

- Popular hashtags consistently drive high engagement like *"delicious," "beauty," "stunning,"* and *"delicious"*

- Posts using niche-specific hashtags tend to attract targeted audiences.

Guiding Content Strategy:

- Focus on creating content around trending hashtags to maximize visibility.

- Use high-engagement hashtags in ad campaigns to reach active users.

- Encourage user-generated content with branded hashtags to amplify reach

tag_name	avg_engagement
delicious	65.27
beauty	65.15
foodie	64.64
sunset	64.42
stunning	64.25
food	64.13
photography	64
dreamy	63.9
smile	63.69
party	63.44
beach	63.43
landscape	63.29
style	63.29
drunk	63.05
happy	63
lol	63

Posting Time Patterns

Insights from posting data

- Optimal posting times are between 12 p.m. and 1 p.m., particularly on Friday

Marketing Implications:

- Schedule posts during peak hours for maximum visibility.

- Use demographic data (e.g., age groups) to tailor content themes (e.g., travel for younger users).

- Create location-specific campaigns leveraging regional trends.

post_hour	post_day	avg_engagement
13	Friday	63.3074

Influencer Marketing Campaigns

Ideal Candidates:

- Users with high follower counts and engagement

- Definition:

   Influence Score = followers * engagement_rate

Collaboration Approach:

- **Outreach:** Send personalized invitations explaining mutual benefits.

- **Incentives:** Offer compensation through monetary rewards or exclusive perks.

- **Content Co-Creation:** Collaborate on posts featuring your product/service.

- **Analytics:** Monitor campaign performance to refine future strategies.

Users with Top Influencer Score

- Meggie_Doyle
- Jaylan.Lakin
- Granville_Kutch
- Kenneth64
- Rick29
- Karley_Bosco
- Damon35
- Odessa2
- Janet.Armstrong
- Aracely.Johnston98

Brand Ambassadors

Identifying Brand Ambassadors:

- Look for users with high follower counts and consistent engagement rates.

- Prioritize those actively using branded hashtags.

- Selection based on Influencer score or celebrities aligning with engagement thesis

Promotion Strategies:

- Invite ambassadors to exclusive events or collaborations.

- Provide branded merchandise for promotional use.

- Encourage them to share authentic testimonials about the platform.

Optimization using Ad Data
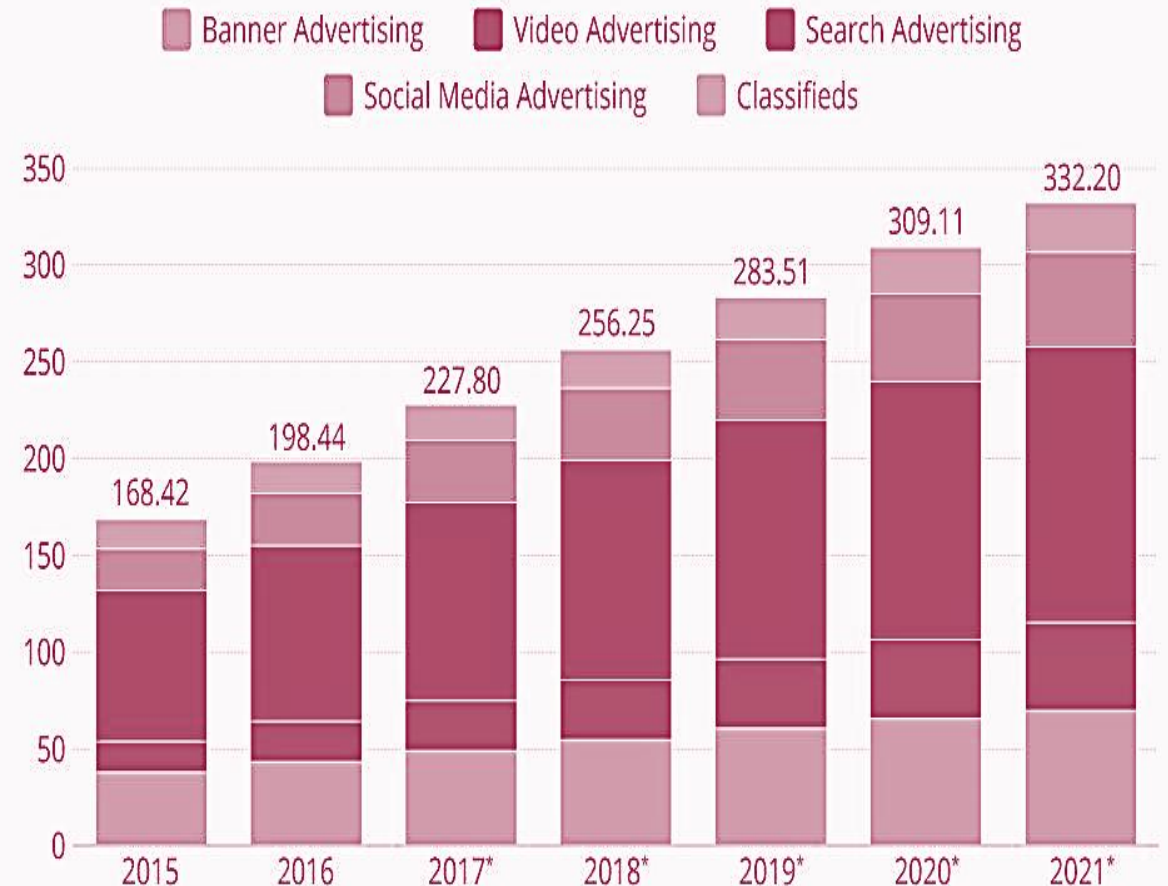
Measuring Effectiveness:

- Track key metrics like click-through rate (CTR), conversion rate, and return on ad spend (ROAS).

- Analyze user behavior post-click (e.g., time spent on landing pages).

Optimization Strategies:

- Refine targeting by analyzing demographic performance.

- A/B test different ad creatives and formats.

- Focus budget allocation on high-performing channels.

What's Ad Growth Going to Look Like in the Digital Arena?

Digital advertising revenue worldwide from 2015 to 2021, by format (in billion U.S. dollars)

Legend:
- Banner Advertising
- Video Advertising
- Search Advertising
- Social Media Advertising
- Classifieds

Year	Total
2015	168.42
2016	198.44
2017*	227.80
2018*	256.25
2019*	283.51
2020*	309.11
2021*	332.20

* forecast

@StatistaCharts Source: Statista Digital Market Outlook

statista

Conclusion

Insights

- Reliable data (no duplicates/nulls).

- Top users and hashtags drive engagement.

- Loyal users and influencers are assets.

- Inactive users can be re-engaged.

Recommendations

- **Reward loyal users.**

- **Use trending hashtags.**

- **Partner with influencers.**

- **Segment for personalization.**

- **Optimize ads with data.**

Data is the new oil.
And
Analytics is the new Engine.

Thank You

Aniket Bobde

Aniketbobde.iitkgp@gmail.com