Improving WhatsApp Groups to Reduce Spam and Increase Engagement

Case Prompt:

How would you improve WhatsApp groups to reduce spam and make them more engaging?

Step 1: Clarifying Questions

Candidate Question: Are we targeting a specific age group or demographic, or is this improvement for all users in general?

Interviewer Answer:
 For this case, assume we are targeting the general WhatsApp user base globally. You can choose to focus on a specific segment if you believe it will lead to more meaningful insights. WhatsApp groups vary from small family or friends groups to large community and business groups. If you want to narrow your focus, focusing on large community groups where spam is more frequent is reasonable.

Candidate Question: Can I focus on community groups, given that these tend to have more spam compared to family or work groups?

Interviewer Answer:
 Yes, focus on community groups. These tend to experience the highest spam volume and engagement drop-off. You can assume groups have 50+ members and 1–2 admins.

Candidate Question: Apart from existing admin controls to mute group members, is there any other context I should consider before analyzing solutions?

Interviewer Answer:
 Yes. Key points to consider are:

1. Many community group members become passive readers. Around 60 to 70 percent of messages are forwards or promotional.

2. WhatsApp's goal is to improve perceived group value so users are less likely to mute or leave. Engagement means meaningful interactions, not just message volume.

3. Constraints: WhatsApp prefers privacy-conscious and lightweight solutions. Avoid public feeds or features that read message content in the cloud.

Step 2: Defining Spam and Engagement

Spam:

- Spam is any content in a group that is
  - **irrelevant** to the group's purpose
  - **benefits** only the **sender**
  - **disrupts** meaningful **interaction**
- This includes
  - overly **promotional** messages
  - **forwards** with **no** added **context**
  - posts that lead members to leave, mute, or disengage
- Spam can be **measured** through signals such as
  - **messages** receiving **no reactions**
  - **causing** members to **exit** the **group**
  - being **frequently forwarded** without relevance

Engagement:

- Engagement is the extent to which group members actively participate and derive value from the content shared.
- It reflects meaningful **interaction**, including **reading**, **responding**, **reacting**, or **contributing** to discussions.
- Engagement can be **quantified** by metrics such as the
  - number of **active members per week**
  - **reactions** per message
  - reply **frequency**
  - member **retention** over time

Step 3: Identifying Pain Points

Members want to engage in meaningful groups but often face irrelevant or spammy messages. Their options are limited to muting or leaving.

Admins want to bring like-minded people together but currently cannot effectively moderate content. Existing controls are limited to deleting messages or removing members.

To address these issues:

- Members can add personal filters for certain keywords or block messages from specific users to reduce exposure to irrelevant content.

- Admins can restrict certain keywords in the group, limit messages per user per day or per hour, and block users who repeatedly send irrelevant content. Admins can also monitor engagement to identify which posts are valued by the group.

Step 4: Prioritization and Success Metrics

The primary metric to focus on is **group retention**, aiming for a 10% increase, because retaining members is a stronger indicator of value than raw engagement volume.

Prioritization is given to **admin-level features**, as admins have the most leverage in shaping group quality, and not all members may have the knowledge or interest to customize filters themselves.

Admin feature proposals include:

- Daily or weekly engagement insights to identify top-performing and least-performing messages, allowing admins to take informed moderation actions.

- Restricting individual members or all members from sending excessive messages per day.

- Highlighting or pinning top-performing posts to reward positive engagement.

Step 5: Phased Solution and Summary

Phase 1:
 Admins can restrict certain keywords in new or existing groups. For example, political keywords in a hobby-focused group are blocked to maintain relevance.

Phase 2:
 Admins can limit the number of messages per user per day or per hour to control spam.

Phase 3:
 Admins can view the top five posts with the highest and lowest engagement. Based on this, admins can restrict certain types of posts or highlight the most valuable content.

  Optionally, admins can pin top-performing posts of the week to reward meaningful contributions.

Success Metrics:

- % **increase** in **group retention** over months.

- % **increase** in **engagement** among retained members.

Impact:
 By empowering admins with lightweight moderation tools and engagement visibility, irrelevant or spammy content is reduced. This improves retention, makes group participation more meaningful, and naturally increases engagement.

Step 6: Risks and Validation

Potential risks include **admin overload** or excessive controls making **moderation tedious**, and **over-restriction** potentially **discouraging** member **contributions**.

Pilot testing in select community groups before a global rollout is **recommended**, tracking retention, engagement, and admin satisfaction to validate effectiveness.