

## Question Type: Analytics

**Duration:** 20 Minutes

**Difficulty:** Hard

**Domain:** Risk

### Problem

Facebook's integrity team ensures that users are safeguarded from spams across platforms - Facebook, Instagram and Whats App. How would you measure spam on the Facebook ecosystem?

## Solution

Facebook's integrity team ensures that users are safeguarded from spams. How would you measure spam on Facebook?

**[Candidate]** To address this question, I need to define the meaning of spam on Facebook. A spam could be a post with misleading, disturbing or sexual content generated by a fake account. I know that on Facebook, a post that violates community policy is removed. Perhaps, I can count the number of posts removed per month.

**[Interviewer]** Okay, but how would you determine that a post is a spam?

**[Candidate]** I know that users can report a post should the content concerns them. I can use the count of flags as a proxy for spam. The higher count the more likely that a post is a spam. I can apply a decision rule that a post with greater than five flags is a spam. Then, count the number of spams or the proportion of spam posts per month.

**[Interviewer]** But, what if a spam doesn't have any reports from other users? How would you count those cases?

**[Candidate]** For those cases, I can think of a spam being a clickbait, meaning that a post contains misleading content that directs users to websites with malicious intentions (i.e. malware, fake goods, sexual predator).

**[Interviewer]** Okay, how would you determine that a post is a clickbait?

**[Candidate]** I am assuming that Facebook has a review team. A sample of posts can be sent to the review team and classified as spam or ham.

**[Interviewer]** Think about billions of posts published on Facebook daily. Given that a spam is a low-probability event, a sample of hundreds of thousands of posts sent to the review team may not contain a single spam. That means that the number of spam reported on a month would be 0, which is not accurate. Can you think of a better solution?

**[Candidate]** Hmm... I think I have an approach. First of all, can I assume that Facebook has a users activity data that tracks timestamps of when a user exit the platform upon clicking a post and re-entered the platform?

**[Interviewer]** Yes.

**[Candidate]** Great, I can use the timestamps as a proxy for spam. The assumption is that the shorter the difference in timestamps, the more likely that the post is a clickbait. The reason is that a user may have entered the external website but returned to the platform immediately upon realizing that the site content is misleading from the post.

I can compute the average of time delta across clicks per post. Now, I have a continuous distribution of average time delta across billions of posts. I would send a sample of posts with the time delta average less than the 10th percentile. The assumption is that the sample with a cutoff will improve detection rate.

---

### Interviewer Comments

The candidate offered a solution to measure spam using proxy and sub-sampling techniques. Although the solution improves detection rate of spam, the estimation is based on a fraction of the population, not the entire population. Hence, the estimation is bias. While continuing to use the proxy measure, method could be improved using stratified sampling.

Based on the time-delta proxy, bucket the posts into five groups as shown in the table below. The thresholds can be defined based on percentiles. Assume that the top one percentile has a threshold at  $\leq 0.1$ . The top ten percentile has a threshold at 1.0. Create your cutoffs based on top 1%, 10%, 30%, 60% and 100%. Based on the cutoffs, from high to low spam-density buckets, the proportion is, respectively, 1%, 9%, 20%, 30%, 40%.

In each bucket, vary the sample rate, but ensure that the total sample size is 1M given the review capacity. To start, set the sample rate such that the sample size for each bucket is the same at 200,000.

Buckets	Proxy Thresholds	Post Proportions	Post Counts	Sample Rate	Sample Size
High	$t \leq 0.1$	1%	1,000,000	20.00%	200,000
Medium-High	$0.1 < t \leq 1.0$	9%	9,000,000	2.22%	200,000
Medium	$1.0 < t \leq 3.0$	20%	20,000,000	1.00%	200,000
Low-Medium	$3.0 < t \leq 5.0$	30%	30,000,000	0.67%	200,000
Low	$t < 5.0$	40%	40,000,000	0.50%	200,000
	Total	100%	100,000,000	100.00%	1,000,000

Suppose the review team returns the following spam count per each bucket:

Buckets	Post Proportions	Sample Size	Spam Count	Spam Rate
High	1%	200,000	100,000	0.50
Medium-High	9%	200,000	30,000	0.15
Medium	20%	200,000	5,000	0.025
Low-Medium	30%	200,000	3,000	0.015
Low	40%	200,000	1,000	0.005

Using the bucket proportion and spam rate, apply the law of total probability to estimate the true spam rate in December 2019.

$$(0.01 * 0.50) + (0.09 * 0.15) + (0.20 * 0.025) + (0.30 * 0.015) + (0.40 * 0.005) = 0.03$$

---

## Interviewer Assessment

In the statistics section, a candidate is assessed based on correctness and soundness of statistical methodology, product sense and communication. For each dimension the candidate is rated in the following scale: (5) superior, (4) good, (3) adequate, (2) marginal, (1) not competent.

Assessments	Rating	Comments
<b>Statistical Methodology</b>	4	The candidate proposed a decent idea to use a proxy to classify a post as a spam. However, there are some flawed attempts that required hints. When the candidate was asked how a clickbait spam is counted, the candidate proposed random sampling across the entire population which is flawed. With a hint that the approach is flawed, the candidate managed to offer an improved solution using proxy. However, the approach can be improved further using stratified sampling.
<b>Product Sense</b>	4	The candidate demonstrated decent product sense in understanding how a post could be flagged on the Facebook platform. She mentioned post-flagging, user activity data and clickbaits that could help measure spam. However, the solution appears to be specific to the Facebook platform, not the sister apps such as Instagram and Whats App. This suggests that the candidate may not be familiar with those product apps.
<b>Communication</b>	5	The candidate responded with competence and clarity. She made sure her assumptions are clarified before offering solutions. Her conversation style felt natural, not a monologue. She made sure to correct her methodology once she listened to a hint provided.