BRAND • MARKETING • DIGITAL
FIELD MARKETING BY MARRIOTT BCV

Twitter's Fight Against Platform Manipulation

Background:

- In 2018, Twitter initiated a massive effort to fight increasing "platform manipulation", in order to limit abuse of the platform, improve the collective health of conversations, and publicly reinforce their commitment to the open exchange of information on their platform.

What changes did they make?

- They launched a drastic effort which includes a variety of new measures and policies, but the focus of their effort is to **proactively identify problematic accounts and behavior** (whereas previously, they only reviewed if a report was submitted).

- In order to help strongly enforce this across a global network, Twitter implemented machine learning tools and new technology that detect potential violations and **automatically** take action.
    - Once an account is suspected by Twitter's detection tools, the account is automatically locked, the owner of the account is notified, asked to validate the account, and reset their password.
    - **If the owner of the account does not comply, the account is suspended and then terminated.**

How does that effect our Twitter followers?

- Twitter also actively takes steps to reduce the visibility of suspicious accounts in Tweet and account metrics.
- When an account is suspended, it is put into a read-only state. Since these accounts can't engage with others or Tweet, and as a consequence of suspicious behavior, **Twitter now removes it from follower figures and engagement counts** (until the account is able to properly verify).
- As a result, Twitter users will likely notice their account metrics dramatically change more regularly.
- While these are dramatic changes, Twitter reinforces this as an important shift in how we display Tweet and account information to ensure that malicious actors aren't able to artificially boost an account's credibility permanently by inflating metrics like number of followers.

What are the numbers?

- From January - June 2019, Twitter has challenged **97,123,667 accounts**
- From January - March 2019, over **100,000 accounts have been suspended**
- From a report in October 2019, Twitter reported a **105% increase** in accounts actioned by Twitter (locked or suspended for violating the Twitter Rules) compared to the previous reporting period of January-June 2019.
- According to Yoel Roth, Twitter's head of site integrity, "**8-10 million accounts a week are challenged automatically**, and more than **3/4 of those accounts wind up removed automatically from the service".**
- In September 2019, Twitter disclosed another 100,000 accounts suspended for suspicious activity

Resources:

- https://blog.twitter.com/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html
- https://blog.twitter.com/en_us/topics/company/2019/twitter-transparency-report-2019.html
- https://blog.twitter.com/en_us/topics/company/2019/health-update.html
- https://blog.twitter.com/en_us/topics/company/2019/info-ops-disclosure-data-september-2019.html
- https://blog.twitter.com/en_us/topics/company/2019/health-update.html
- https://transparency.twitter.com/en/platform-manipulation.html
- https://www.fastcompany.com/90331696/twitter-is-automatically-removing-about-10-accounts-every-second
- https://techcrunch.com/2019/09/20/twitter-discloses-another-10000-accounts-suspended-for-fomenting-political-discord-globally/