

Summary of changes made to Sift rules since 2023

Introduction

Our approach to fraud prevention has been to leverage our solution provider's (Sift) scoring system while minimizing friction to the donor and admin experience. We have rules that automatically reject donations that use foreign prepaid/gift cards, domestic gift cards, are from high-risk/sanctioned countries, and have the highest level of risk as determined by Sift. We manually review all contributions a tier lower than that highest level, and even some lower than that if they have certain risky signals (i.e. an IP/billing country mismatch).

Ultimately, we are not the primary target of people looking to commit merchant fraud. People who steal credit card information do not try to give that money away, they try to buy things on Amazon, etc. Our true rate of fraud is very low, and we only needed to reject ~1,000 donation attempts out of ~6.5 million attempts in August.

From an election integrity standpoint, we take things extremely seriously, and you will find that our rules do an excellent job of preventing inappropriate donations. Risky donors with foreign characteristics are very frequently manually reviewed and we do not accept donations from the riskiest donors at all.

As you will see in the rule changes we have made, we have introduced new rules as we have learned the ways in which we are targeted by fraudsters, but actually, our rules have been historically too *strict*, and we review too many false positive, good donors, and so we have needed to make adjustments to not interfere with the donation process, without opening us up to increased levels of fraud.

Current Rules (as of 9/9/2024)

These rules work in a waterfall manner. A contribution comes into Sift, and then is checked against each rule, in order. This means that a contribution may have characteristics that would, as an example, trigger rule 20 had it made it to it, but if it had triggered rule 12, it would take whatever action in the "Decision Applied" column we have chosen.

Rule Number	Rule Name	Rule Specifics	Decision Applied	Notes
1	Internal Testing Email Whitelist	[REDACTED]	Automatically Accept	In place so ActBlue tech employees can run tests without getting stuck in Sift
2	Foreign Issued Prepaid Cards	[REDACTED]	Automatically Reject	
3	Domestic Gift Cards	[REDACTED]	Automatically Reject	
4	Sift Score 100 Rejection	[REDACTED]	Automatically Reject	The name of the rule needs to be updated, but the rule essentially says we are going to automatically reject what we consider to be the riskiest of all donations regardless of any specific signals
5	High Risk Countries (Review) - IP	[REDACTED]	Manually Review	
6	High Risk Countries (Review) - Billing	[REDACTED]	Manually Review	
7	High Risk Countries (Review) - Credit Card	[REDACTED]	Manually Review	

8	High Risk Countries (Reject) - IP		Automatically Reject	
9	High Risk Countries (Reject) - Billing		Automatically Reject	
10	High Risk Countries (Reject) - Credit Card		Automatically Reject	
11	Naughty Words List		Manually review	Slurs, curses, etc.
12	Griefing Words List		Manually Review	Different from naughty words, more like "Donald Trump", "Joe Biden", etc.
13	Fake Addresses List		Manually Review	
14	X Names List		Manually Review	For when donors try to anonymize themselves by putting "X", "XX", etc. as one of their names. We do proactive outreach to those people asking them to stop.
15	IP/Billing Country Mismatch + Score		Manually Review	This rule captures donors who say their billing address is in one country, but their IP suggests they are donating from another. Could be considered a rule that looks for "foreign" donors, but it isn't that black/white (i.e. if a US citizen is traveling abroad and donates, they aren't a "foreign" donor but may be caught by this rule). There is a score component to this rule so we are reviewing the riskiest donations with this characteristic (but lower than the threshold to be automatically rejected).
16	Multiple Users, Same Device + High Score		Manually Review	
17	■ w/Same Email + Score		Manually Review	

18	Disposable Email Domain + Score	[REDACTED]	Manually Review	
19	ISP is Cloud Provider + Score	[REDACTED]	Manually Review	
20	Chargebacks [REDACTED] in Last Month	[REDACTED]	Manually Review	There is a very clear fraud component to chargebacks but we also use this rule to find donors to whom we want to proactively reach out to about their trouble donating.
21	Chargebacks [REDACTED] + Successful TXNs [REDACTED] in Last Month	[REDACTED]	Manually Review	There is a very clear fraud component to chargebacks but we also use this rule to find donors to whom we want to proactively reach out to about their trouble donating.
22	High Score (c3/c4)	[REDACTED]	Manually Review	This rule flags for manual review the riskiest of contributions, regardless of any specific characteristics (but not risky enough to automatically reject). We allow for slightly more leniency for charitable and civic donations because we accept foreign donations and don't require a billing address (which increases Sift scores).
23	High Score (Political)	[REDACTED]	Manually Review	Same as above, but stricter rules for political contributions.
24	Everything Else	[REDACTED]	Automatically Accept	If a donation does not match any of the previous characteristics/scores, it gets accepted in Sift (this is ~99.8% of volume that flows through Sift).

Changelog this cycle

Key changes highlighted.

Date	Summary of changes made	Notes
10/16/2023	<ul style="list-style-type: none"> Manual Review Countries (IP/Billing/CC Country Matches) [REDACTED] 	Added review and rejection rules for

	<p>[REDACTED]</p> <ul style="list-style-type: none"> Block Countries (IP/Billing/CC Country Matches): [REDACTED] Automatically reject contributions with Sift score of [REDACTED] (monitor heavily and adjust downwards as needed) Rename High Risk Country rules Rename "Chargebacks [REDACTED] in Last Month + Successful TXNs [REDACTED]" -> "Chargebacks [REDACTED] in Last Month + Successful TXNs [REDACTED] in Last Month" Rename "Chargebacks [REDACTED] Last Month" -> "Chargebacks [REDACTED] Last Month" Remove the Sift score component of the Foreign Issued Prepaid Card rule Lower the Sift score component of the IP/billing mismatch rule by 1 Increase Sift score component of the c3/c4 rule to [REDACTED] Flip the chargebacks rules in the waterfall and change the "Chargebacks in Last Month + Successful TXNs [REDACTED]" from less than [REDACTED] successful transactions in a week to less than [REDACTED] successful transactions in a month Split "Naughty Words" list into "XX" names and everything else Always manually review XX names Add score threshold of [REDACTED] to Naughty Words rule Move High Sift Score rules to end of waterfall 	<p>risky/sanctioned countries.</p> <p>Added the rule to automatically reject the riskiest of donations.</p> <p>Removed the score component on this rule so that we review ALL donations with this characteristic.</p> <p>Lowered score in this context means more reviews and a more strict fraud approach.</p> <p>Increased score in this context means less reviews and less false positives reviewed.</p>
4/2-4/2024	<ul style="list-style-type: none"> [REDACTED] Txns in Last Day: Score threshold [REDACTED] → [REDACTED] High Sift Score (c3/c4): Score threshold [REDACTED] → [REDACTED] High Sift Score (political): Score threshold [REDACTED] → [REDACTED] IP/Billing Country mismatch: Score threshold [REDACTED] → [REDACTED] Multiple Users/Same Device: Score threshold [REDACTED] → [REDACTED] Similar emails: Score threshold [REDACTED] → [REDACTED] ISP Cloud Provider: Score threshold [REDACTED] → [REDACTED] Unique CC Bins in past day: Score threshold [REDACTED] → [REDACTED] [REDACTED] Txns in Last Day + Dollar Amount: Score threshold [REDACTED] → [REDACTED] 	<p>Increased score thresholds across the board. This technically means a more lenient approach, but it should be noted that the lower ends of these score ranges included an extreme amount of false positives, so we actually did not increase our exposure to fraud much, if at all.</p> <p>Over 1 month period:</p> <p>Reviews saved: ~13,534 (about 65% of total review volume)</p> <p>Rejections missed: ~14-28 (around 3.2% - 6.4% of total rejection volume)</p>
6/17/2024	<ul style="list-style-type: none"> IP/Billing Country mismatch: Score threshold [REDACTED] → [REDACTED] Multiple Users/Same Device: Score threshold [REDACTED] → [REDACTED] Similar emails: Score threshold [REDACTED] → [REDACTED] ISP Cloud Provider: Score threshold [REDACTED] → [REDACTED] Unique CC Bins in past day: Score threshold [REDACTED] → [REDACTED] [REDACTED] Txns in Last Day + Dollar Amount: Score threshold [REDACTED] → [REDACTED] 	<p>Over 1 month period:</p> <p>Reviews saved: 3684 (about 28% of total review volume)</p> <p>Rejections missed: 0</p>

	<ul style="list-style-type: none"> • [redacted] Txns in Last Day: Score threshold [redacted] → [redacted] 	
7/3/2024	<ul style="list-style-type: none"> • Number of users with <i>similar</i> email -> Number of users with <i>same</i> email • Eliminated the following rules <ul style="list-style-type: none"> ○ Unique BINs ○ Unique Last 4 ○ High Order Velocity #1 ○ High Order Velocity #2 ○ High Order Velocity #3 	<p>Historically, fraudsters on our platform have used the same IP address, email, or credit card, rather than similar ones.</p> <p>These rules were catching almost exclusively false positives. We kept them for a long time to do proactive outreach to these donors, but changed that policy.</p>
8/12/2024	Changed Foreign Issued Prepaid Card rule to automatically reject, rather than send to manual review	
9/9/2024	Added rule to reject all gift card donations	