

# What the analysis found

fraud\_200k.csv - run run\_1037d177a96f - part of [REPORT.pdf](#)

Every finding with the tool results and charts behind it, then the plan the analysis followed. The short version is in REPORT.pdf, section 1.

Written by the AI from the tool results, which are shown under each finding (the lines starting T1, T2, ...). Gate 4 checked that every number in the AI's text is one a tool reported.

AI-written: In this PaySim sample, fraud shows up only in two transaction types. All 240 fraudulent transactions are TRANSFER or CASH\_OUT: 0.7% of transfers and 0.2% of cash-outs, against 0.0% for PAYMENT, CASH\_IN and DEBIT. Fraud is typically larger than legitimate activity (median 401,192 against 73,996), but amount works very differently by type. Cash-outs of 1,000,000 and over are mostly fraud (77.8% of 54), while among transfers fraudulent amounts are no larger than legitimate ones (median 371,771 against 485,695). Fraudulent amounts never go above 10,000,000, and the band at exactly that value has a fraud rate of 7.6%. The fraud rate climbs from about 0.1% early in the month to 0.8%-1.0% after step 504. The likely reason is that legitimate volume collapses in the last part of the simulation while fraud continues at a roughly steady pace, so this reflects how the simulation is built rather than a time-of-month habit of fraudsters.

## q1. How common is each transaction type, and what share of each type is fraudulent?

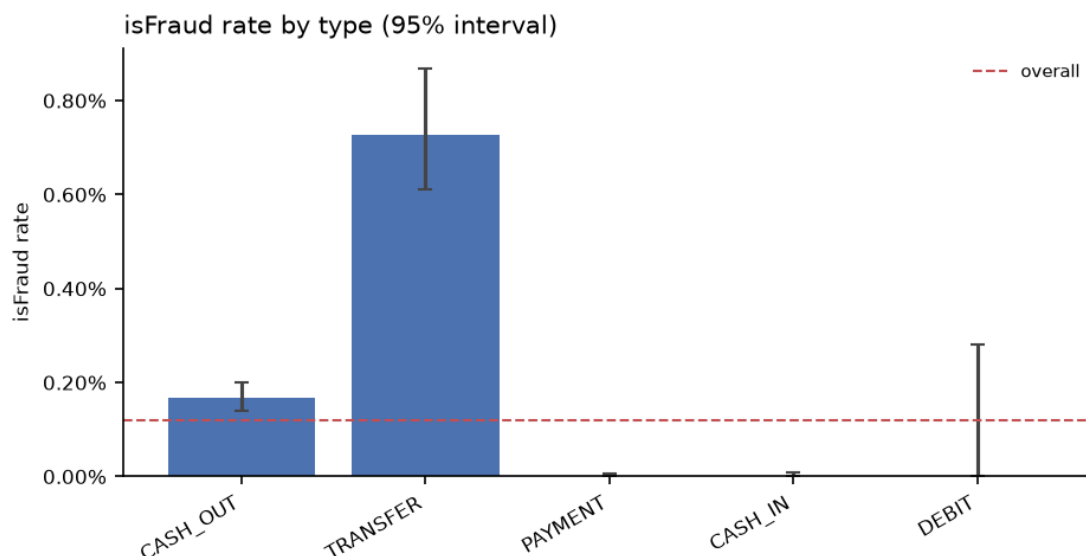
**answered.** CASH\_OUT is 70,084 rows and PAYMENT 67,660, with CASH\_IN 43,991, TRANSFER 16,898 and DEBIT 1,367. Fraud occurs only in TRANSFER (0.7%, interval 0.6% to 0.9%) and CASH\_OUT (0.2%, interval 0.1% to 0.2%). PAYMENT, CASH\_IN and DEBIT show 0.0%.

**Fraud occurs only in transfers and cash-outs; payments, cash-ins and debits carry none.** (strong)

AI-written: TRANSFER has a fraud rate of 0.7% (95% interval 0.6% to 0.9%, n=16,898) and CASH\_OUT 0.2% (0.1% to 0.2%, n=70,084). PAYMENT (n=67,660), CASH\_IN (n=43,991) and DEBIT (n=1,367) show 0.0%. Restricting to TRANSFER and CASH\_OUT still keeps all 240 fraud cases (86,982 rows), which matches the brief's description of fraudsters depleting accounts through transfers and cash-outs.

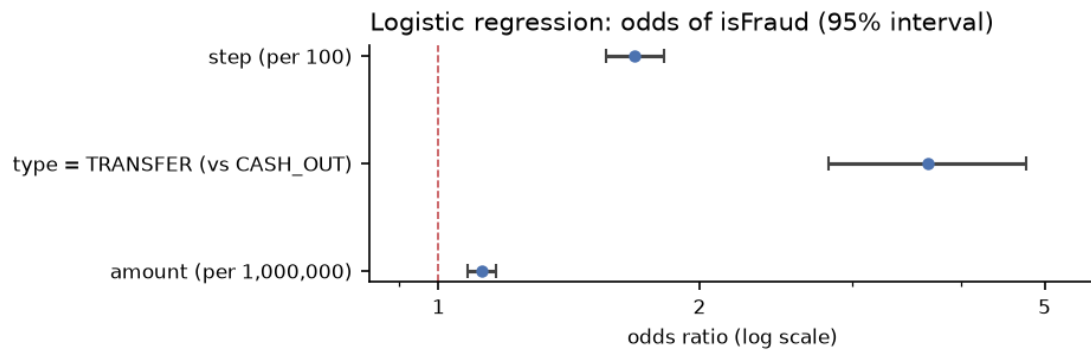
- Caveat: DEBIT is small, so its interval reaches 0.3%.

T1 target\_by\_group: Rate of isFraud by type (overall 0.1%, n=200000): CASH\_OUT: 0.2% (95% interval 0.1% to 0.2%, n=70084); TRANSFER: 0.7% (95% interval 0.6% to 0.9%, n=16898); PAYMENT: 0.0% (95% interval 0.0% to 0.0%, n=67660); CASH\_IN: 0.0% (95% interval 0.0% to 0.0%, n=43991); DEBIT: 0.0% (95% interval 0.0% to 0.3%, n=1367)



T11 logistic\_regression: Within type = TRANSFER or CASH\_OUT: logistic regression of isFraud on type, amount, step: 86982 rows (240 with isFraud = 1), none left out; model vs no model (likelihood ratio)  $p < 0.001$ ; pseudo  $R^2$  0.09; AUC 0.76 on the same rows; of 3 effects, about 0.2 would show  $p < 0.05$  by chance alone

Within type = TRANSFER or CASH\_OUT: Odds ratios (1 = no effect; 95% interval): step x1.68 per 100 (1.56-1.82,  $p < 0.001$ ); type = TRANSFER x3.67 vs CASH\_OUT (2.82-4.77,  $p < 0.001$ ); amount x1.12 per 1,000,000 (1.08-1.16,  $p < 0.001$ )



## q2. Are fraudulent transactions larger than legitimate ones, and how does the fraud rate change across amount ranges?

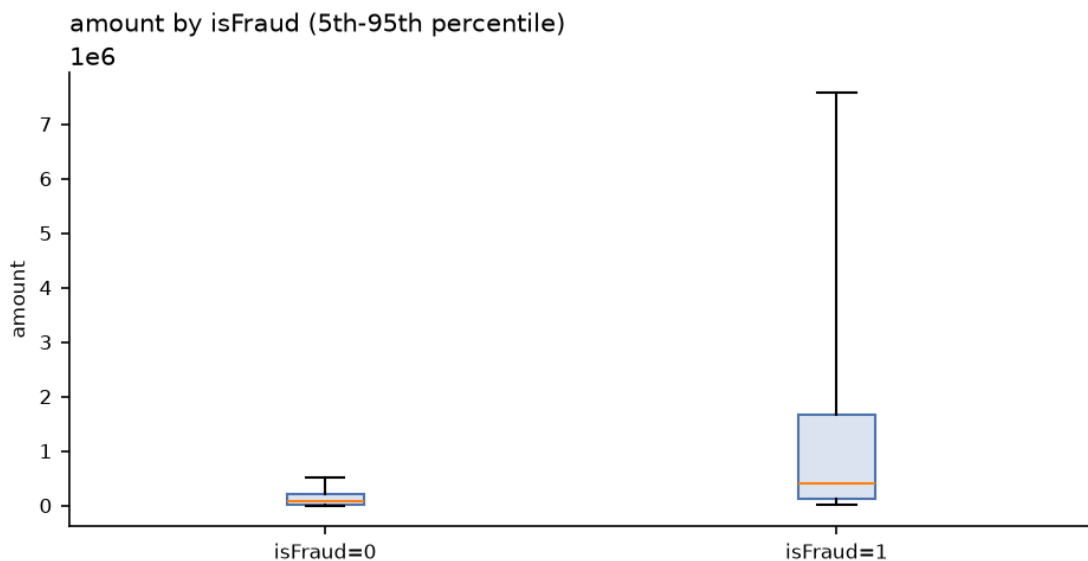
**answered.** Yes. Fraudulent amounts have a median of 401,192 against 73,996 for legitimate ones. The fraud rate rises from 0.0% below 10,000 to 1.8% for 1,000,000 to below 5,000,000 and 4.7% at 5,000,000 and over. Fraud amounts top out at exactly 10,000,000.

**Fraudulent transactions are typically much larger, and the fraud rate climbs steeply with amount. (strong)**

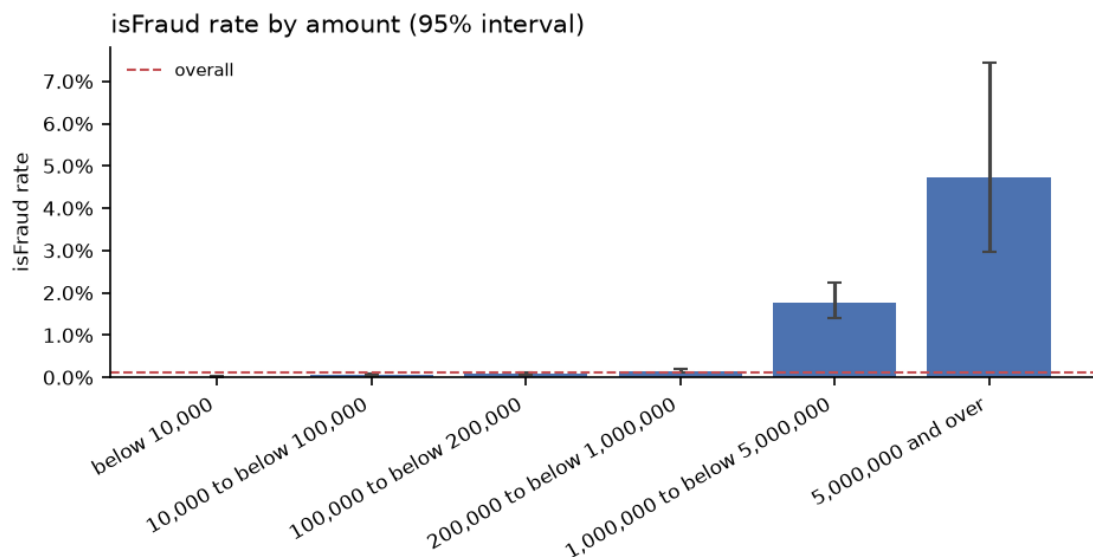
AI-written: The median fraud amount is 401,192 against 73,996 for legitimate transactions (standardised difference +0.74). The fraud rate is 0.0% below 10,000, 0.1% from 10,000 to 200,000 and 0.2% from 200,000 to 1,000,000. It then jumps to 1.8% for 1,000,000 to below 5,000,000 and 4.7% at 5,000,000 and over (interval 3.0% to 7.4%, n=360). Still, 155 of the 240 fraud cases are below 1,000,000, so most fraud is not in the high bands.

- Caveat: Part of this gap reflects type mix: PAYMENT and CASH\_IN are mostly small and carry no fraud.

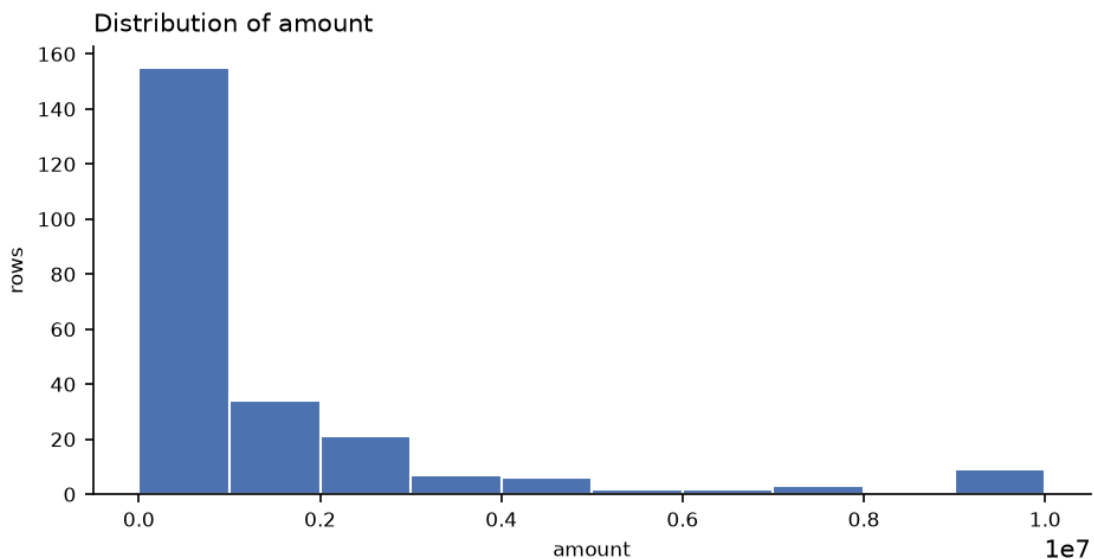
T2 compare\_by\_target: amount by isFraud: mean 178,054 when 0 (n=199760) vs 1,403,726 when 1 (n=240); medians 73,996 vs 401,192; standardised difference +0.74



T3 target\_by\_group: Rate of isFraud by amount (overall 0.1%, n=200000): below 10,000: 0.0% (95% interval 0.0% to 0.0%, n=40692); 10,000 to below 100,000: 0.1% (95% interval 0.0% to 0.1%, n=70375); 100,000 to below 200,000: 0.1% (95% interval 0.1% to 0.1%, n=36332); 200,000 to below 1,000,000: 0.2% (95% interval 0.1% to 0.2%, n=48394); 1,000,000 to below 5,000,000: 1.8% (95% interval 1.4% to 2.2%, n=3847); 5,000,000 and over: 4.7% (95% interval 3.0% to 7.4%, n=360)



T13 distribution: Within isFraud = 1: amount: n=240, missing=0, median 401,192, middle half 127,046-1,657,931, range 0-10,000,000; rows per range: 0-1,000,000: 155; 1,000,000-2,000,000: 34; 2,000,000-3,000,000: 21; 3,000,000-4,000,000: <10; 4,000,000-5,000,000: <10; 5,000,000-6,000,000: <10; 6,000,000-7,000,000: <10; 7,000,000-8,000,000: <10; 8,000,000-9,000,000: <10; 9,000,000-10,000,000: <10

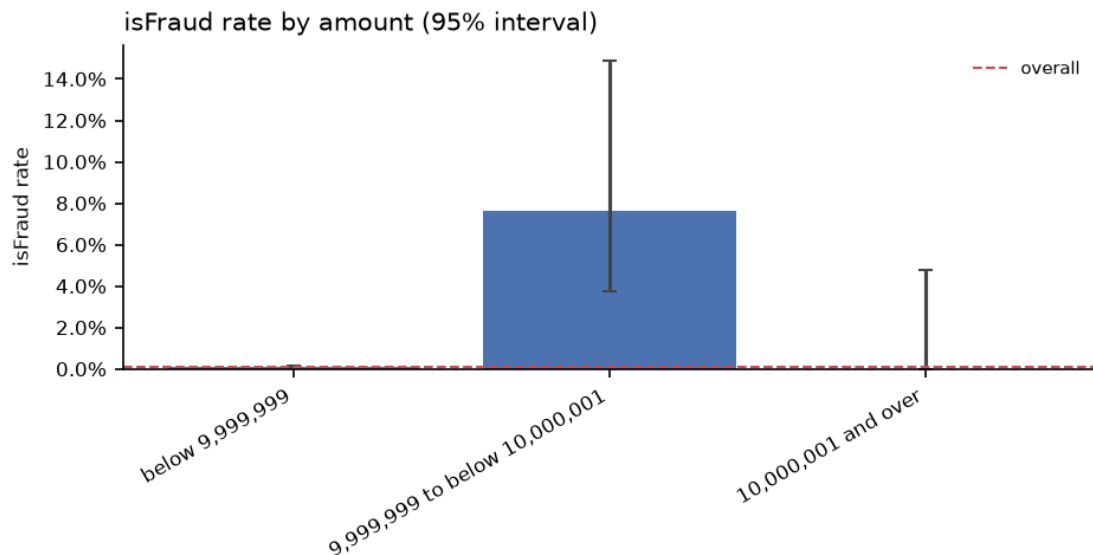


**Fraudulent amounts stop at exactly 10,000,000, a cap visible in the simulation. (moderate)**

AI-written: The largest fraud amount is 10,000,000. Transactions in the band at exactly that value are 7.6% fraudulent (interval 3.7% to 14.9%, n=92). Above 10,000,001 the rate is 0.0% (interval up to 4.8%, n=77).

- Caveat: The groups are small (92 and 77 rows), so the intervals are wide.
- Caveat: This looks like a simulator rule that caps fraud amounts, so it may not carry over to real data.

T13 distribution: Within isFraud = 1: amount: n=240, missing=0, median 401,192, middle half 127,046-1,657,931, range 0-10,000,000; rows per range: 0-1,000,000: 155; 1,000,000-2,000,000: 34; 2,000,000-3,000,000: 21; 3,000,000-4,000,000: <10; 4,000,000-5,000,000: <10; 5,000,000-6,000,000: <10; 6,000,000-7,000,000: <10; 7,000,000-8,000,000: <10; 8,000,000-9,000,000: <10; 9,000,000-10,000,000: <10  
T19 target\_by\_group: Rate of isFraud by amount (overall 0.1%, n=200000): below 9,999,999: 0.1% (95% interval 0.1% to 0.1%, n=199831); 9,999,999 to below 10,000,001: 7.6% (95% interval 3.7% to 14.9%, n=92); 10,000,001 and over: 0.0% (95% interval 0.0% to 4.8%, n=77)



### q3. Does the fraud rate vary over the simulated month (by time step)?

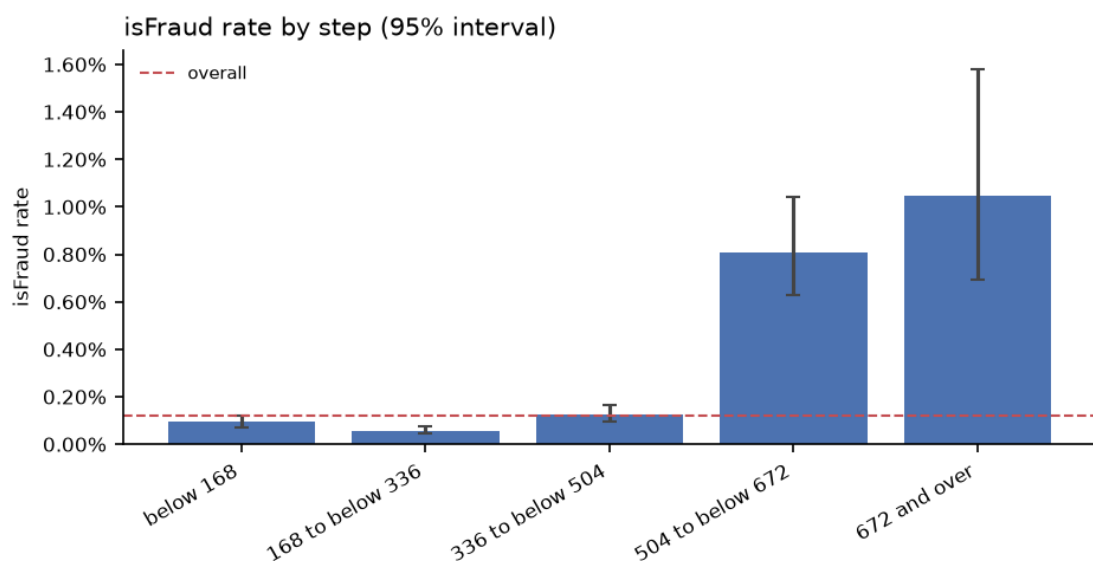
**answered.** The fraud rate is about 0.1% in each of the first three simulated weeks. It rises to 0.8% for steps 504-672 and 1.0% from 672 on. Transaction volume is mostly between 10,188 and 18,027 rows per time range up to step 421, then drops to between 10 and 1,984 rows per range, while fraud counts stay roughly level. The rise in rate mostly reflects fewer legitimate transactions, not more fraud.

**The fraud rate rises late in the month mainly because legitimate volume collapses while fraud keeps a steady pace.** (moderate)

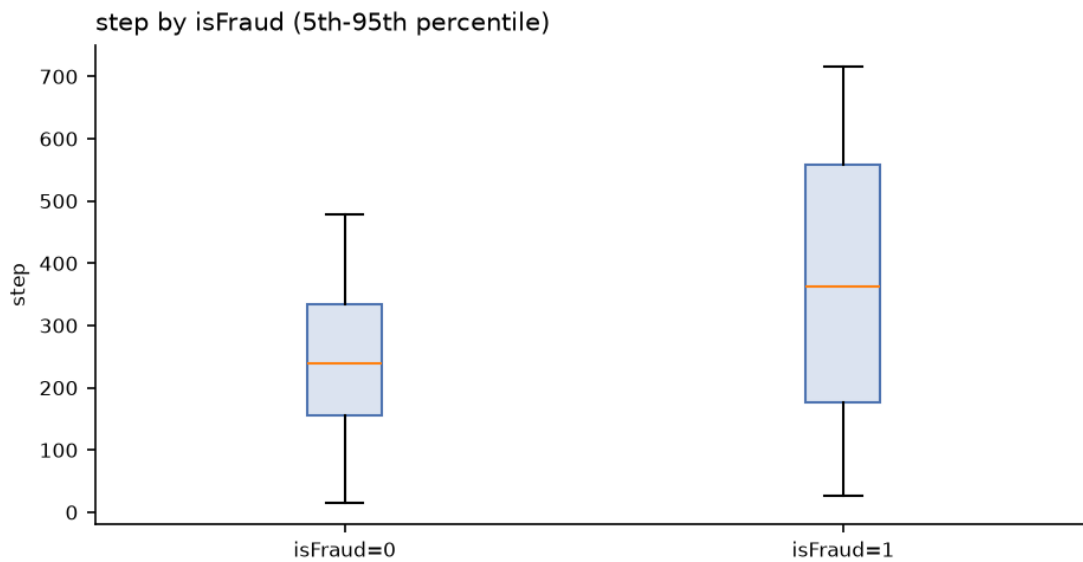
AI-written: The fraud rate is about 0.1% before step 504, then 0.8% for steps 504-672 (n=7,173) and 1.0% from 672 on (n=2,101). Total volume is mostly between 10,188 and 18,027 rows per time range up to step 421, then falls to between 10 and 1,984 rows per range. Fraud counts stay at around 10 or fewer per range across the whole month. The rise holds within both types: CASH\_OUT goes from 0.1% to 1.4% and TRANSFER from 0.5% to 4.6% after step 504.

- Caveat: Fraud counts per time range are mostly reported as fewer than 10, so 'steady' is a rough reading.
- Caveat: This is likely a feature of the PaySim simulation, which also has a near-empty stretch around steps 50-75. It should not be read as a real-world timing pattern.

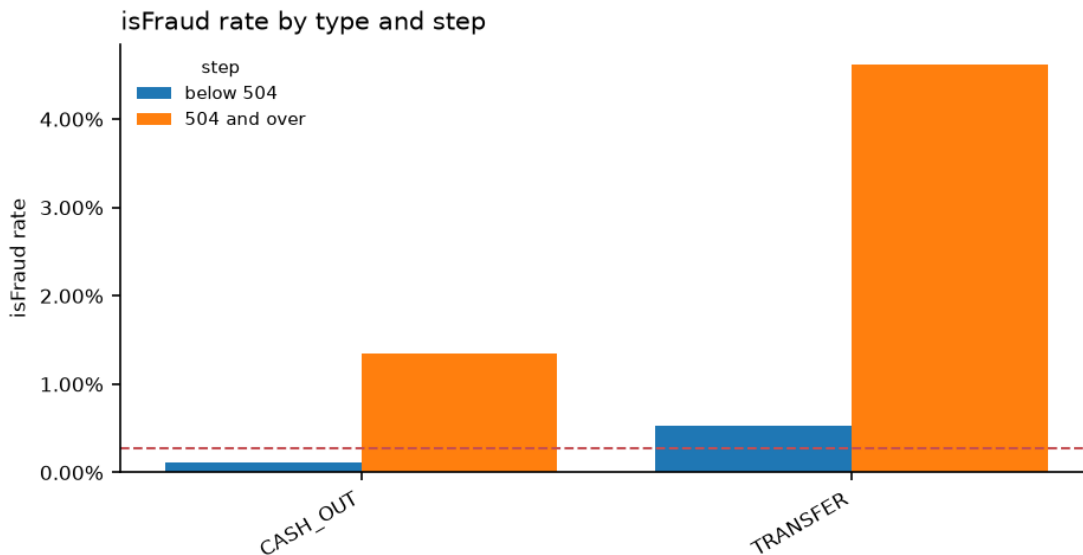
T4 target\_by\_group: Rate of isFraud by step (overall 0.1%, n=200000): below 168: 0.1% (95% interval 0.1% to 0.1%, n=60424); 168 to below 336: 0.1% (95% interval 0.0% to 0.1%, n=89841); 336 to below 504: 0.1% (95% interval 0.1% to 0.2%, n=40461); 504 to below 672: 0.8% (95% interval 0.6% to 1.0%, n=7173); 672 and over: 1.0% (95% interval 0.7% to 1.6%, n=2101)



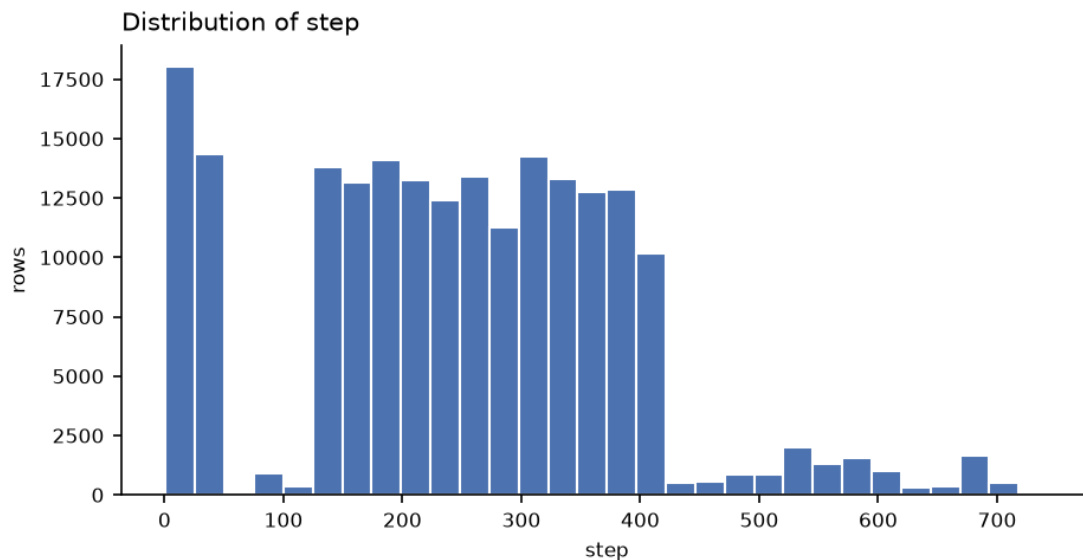
T5 compare\_by\_target: step by isFraud: mean 243 when 0 (n=199760) vs 366 when 1 (n=240); medians 239 vs 364; standardised difference +0.67



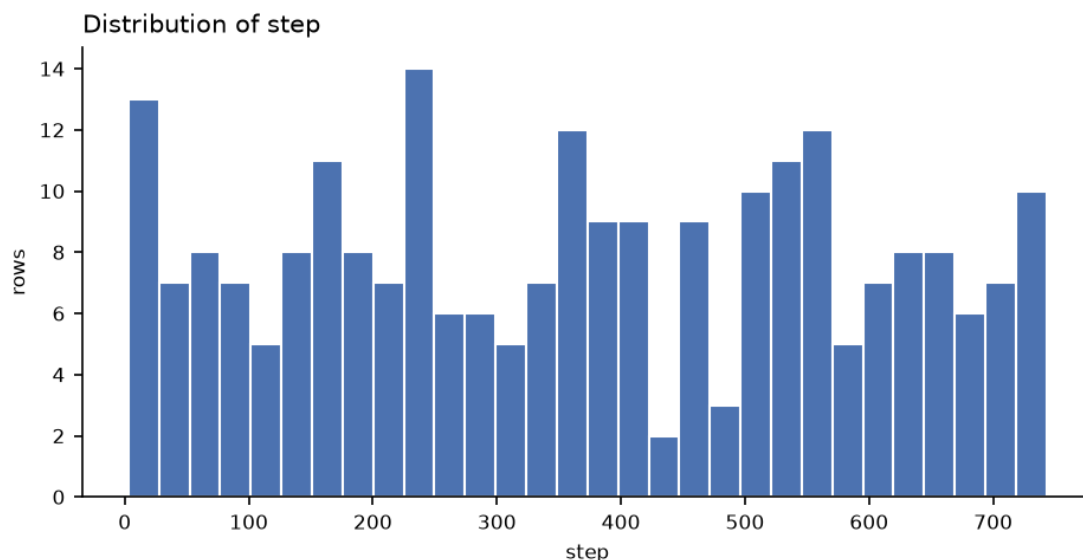
T12 target\_by\_two\_groups: Within type = TRANSFER or CASH\_OUT: Rate of isFraud by type and step (overall 0.3%): CASH\_OUT / below 504: 0.1% (95% interval 0.1% to 0.1%, n=67049); CASH\_OUT / 504 and over: 1.4% (95% interval 1.0% to 1.8%, n=3035); TRANSFER / below 504: 0.5% (95% interval 0.4% to 0.6%, n=16053); TRANSFER / 504 and over: 4.6% (95% interval 3.4% to 6.2%, n=845)



T15 distribution: step: n=200000, missing=0, median 239, middle half 156-335, range 1-743; rows per range: 1-25.7: 18027; 25.7-50.5: 14326; 50.5-75.2: <10; 75.2-99.9: 926; 99.9-125: 352; 125-149: 13824; 149-174: 13149; 174-199: 14105; 199-224: 13264; 224-248: 12396; 248-273: 13417; 273-298: 11274; 298-323: 14241; 323-347: 13326; 347-372: 12778; 372-397: 12871; 397-421: 10188; 421-446: 520; 446-471: 546; 471-496: 880; 496-520: 872; 520-545: 1984; 545-570: 1323; 570-595: 1545; 595-619: 1019; 619-644: 301; 644-669: 367; 669-694: 1631; 694-718: 529; 718-743: 10



T16 distribution: Within isFraud = 1: step: n=240, missing=0, median 364, middle half 177-559, range 3-743; rows per range: 3-27.7: 13; 27.7-52.3: <10; 52.3-77: <10; 77-102: <10; 102-126: <10; 126-151: <10; 151-176: 11; 176-200: <10; 200-225: <10; 225-250: 14; 250-274: <10; 274-299: <10; 299-324: <10; 324-348: <10; 348-373: 12; 373-398: <10; 398-422: <10; 422-447: <10; 447-472: <10; 472-496: <10; 496-521: 10; 521-546: 11; 546-570: 12; 570-595: <10; 595-620: <10; 620-644: <10; 644-669: <10; 669-694: <10; 694-718: <10; 718-743: 10



#### q4. Within the transaction types where fraud occurs (for example TRANSFER and CASH\_OUT), do fraudulent amounts differ from legitimate ones?

**answered.** It depends on the type. Within CASH\_OUT, fraud is much larger (median 455,710 against 146,796), and cash-outs of 1,000,000 and over are mostly fraudulent. Within TRANSFER, fraudulent amounts are not larger (median 371,771 against 485,695), and the fraud rate stays around 0.4% to 1.2% across amount bands below 5,000,000.

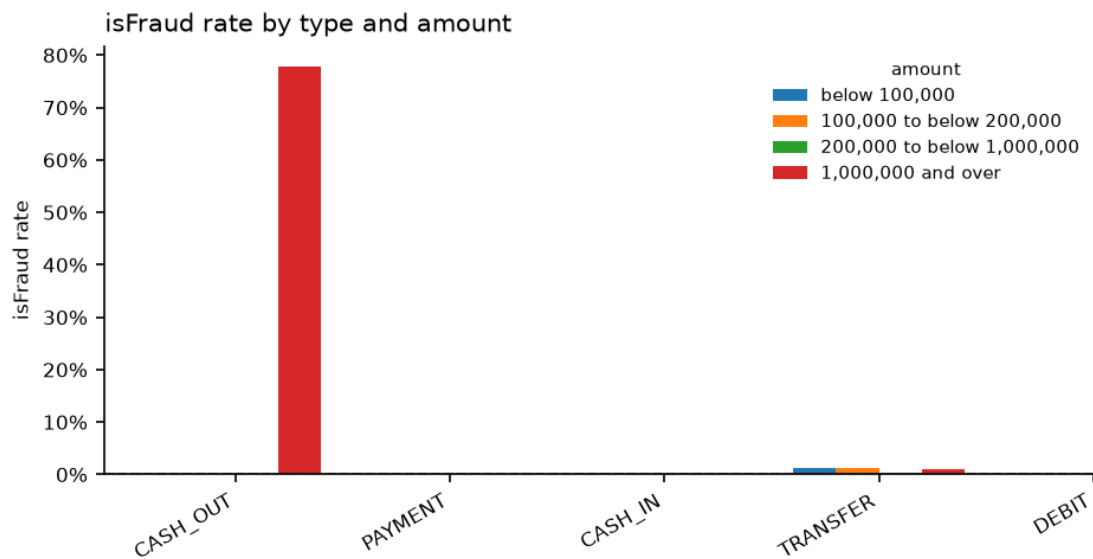
**Large amounts mark fraud among cash-outs, but not among transfers. (strong)**

AI-written: Within CASH\_OUT, fraud amounts have a median of 455,710 against 146,796 (standardised difference +0.76). Cash-outs of 1,000,000 and over are 77.8% fraudulent (interval 65.1% to 86.8%, n=54), against 0.2% or less below that amount. Within TRANSFER, fraudulent amounts are if anything smaller (median 371,771 against 485,695; standardised difference +0.23 on means). The TRANSFER fraud rate is 1.2% below 100,000, 0.4% from 200,000 to 1,000,000 and 1.0% at 1,000,000 and over.

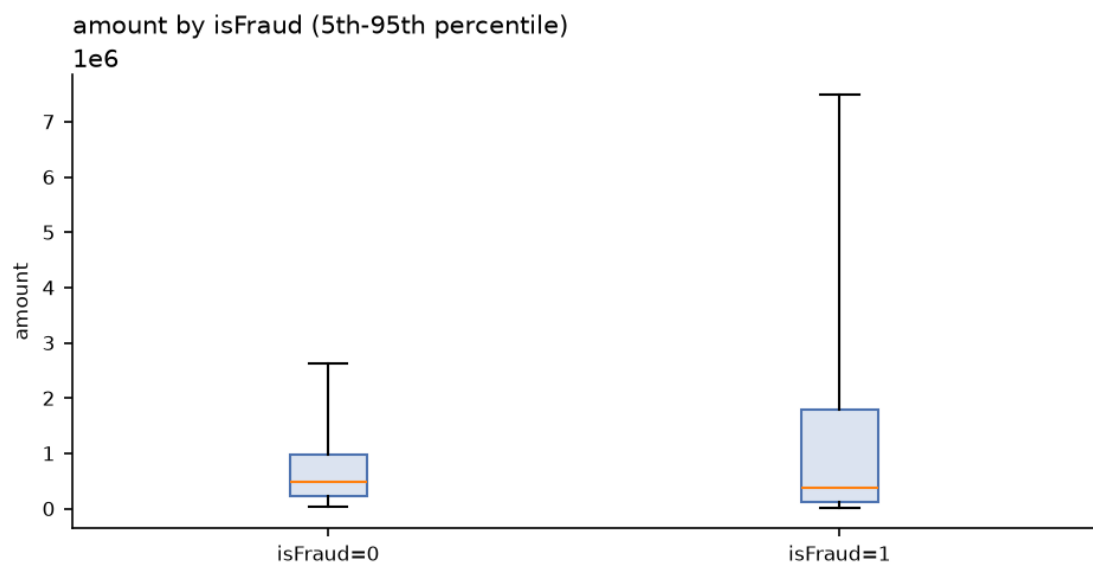
- Caveat: The CASH\_OUT group at 1,000,000 and over is only 54 rows, so the exact share is imprecise. 9 rows in the larger CASH\_OUT bands were too few to show.

T6 target\_by\_two\_groups: Rate of isFraud by type and amount (overall 0.1%): CASH\_OUT / below 100,000: 0.1% (95% interval 0.1% to 0.2%, n=24472); CASH\_OUT / 100,000 to below 200,000: 0.0% (95% interval 0.0% to 0.1%, n=20967); CASH\_OUT / 200,000 to below 1,000,000: 0.2% (95% interval 0.1% to 0.2%, n=24591); CASH\_OUT / 1,000,000 and over: 77.8% (95% interval 65.1% to 86.8%, n=54); PAYMENT / below 100,000: 0.0% (95% interval 0.0% to 0.0%, n=67632); PAYMENT / 100,000 to below 200,000: 0.0% (95% interval 0.0% to 12.1%, n=28); CASH\_IN / below 100,000: 0.0% (95% interval 0.0% to 0.0%, n=15655); CASH\_IN / 100,000 to below 200,000: 0.0% (95% interval 0.0% to 0.0%, n=15655);

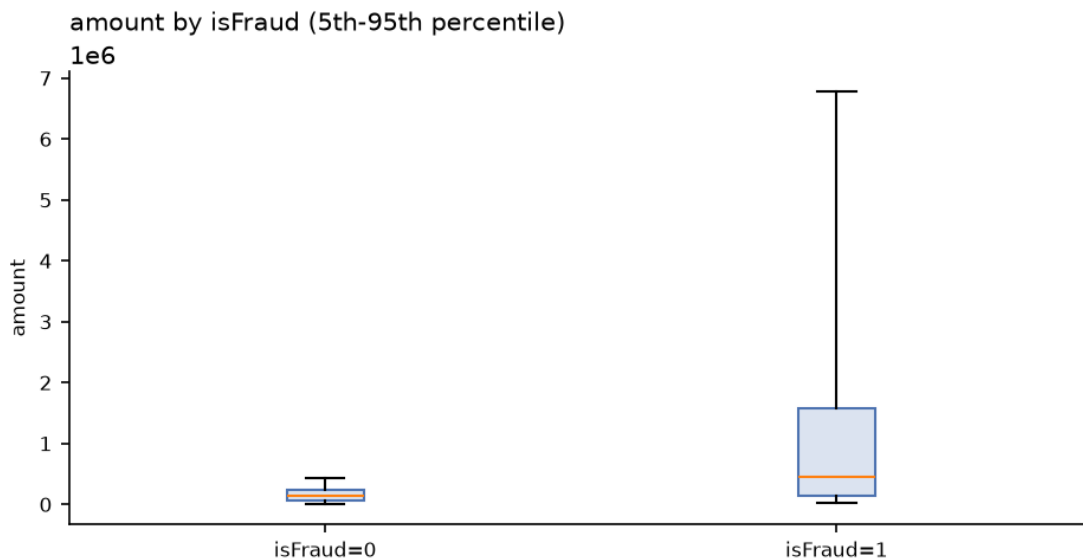
n=13410); CASH\_IN / 200,000 to below 1,000,000: 0.0% (95% interval 0.0% to 0.0%, n=14922); TRANSFER / below 100,000: 1.2% (95% interval 0.8% to 1.8%, n=1946); TRANSFER / 100,000 to below 200,000: 1.2% (95% interval 0.8% to 1.8%, n=1923); TRANSFER / 200,000 to below 1,000,000: 0.4% (95% interval 0.3% to 0.5%, n=8880); TRANSFER / 1,000,000 and over: 1.0% (95% interval 0.8% to 1.4%, n=4149); DEBIT / below 100,000: 0.0% (95% interval 0.0% to 0.3%, n=1362); 9 row(s) in combinations too small to show



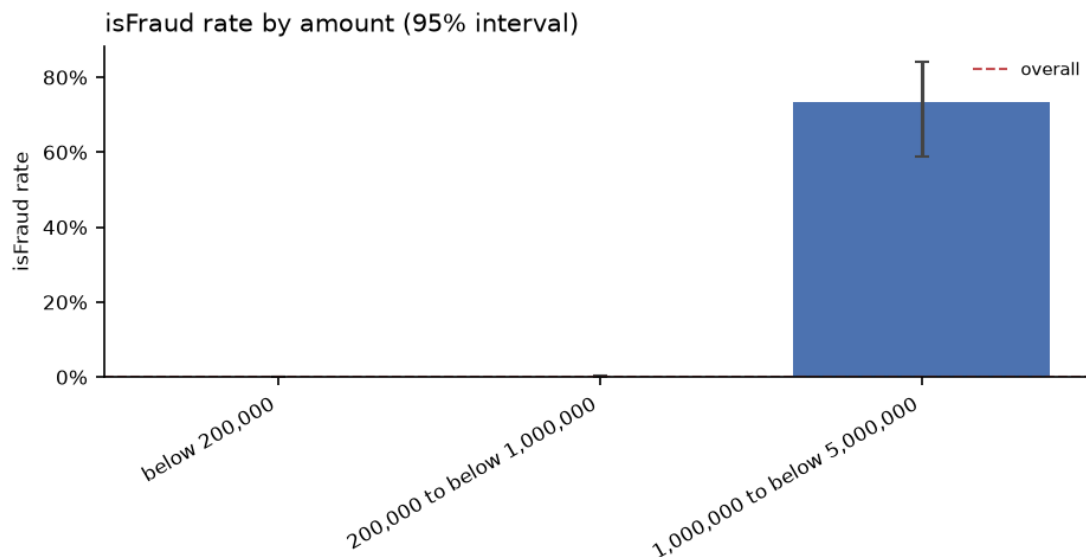
T7 compare\_by\_target: Within type = TRANSFER: amount by isFraud: mean 900,145 when 0 (n=16775) vs 1,359,626 when 1 (n=123); medians 485,695 vs 371,771; standardised difference +0.23



T17 compare\_by\_target: Within type = CASH\_OUT: amount by isFraud: mean 173,826 when 0 (n=69967) vs 1,450,088 when 1 (n=117); medians 146,796 vs 455,710; standardised difference +0.76



T18 target\_by\_group: Within type = CASH\_OUT: Rate of isFraud by amount (overall 0.2%, n=70084): below 200,000: 0.1% (95% interval 0.1% to 0.1%, n=45439); 200,000 to below 1,000,000: 0.2% (95% interval 0.1% to 0.2%, n=24591); 1,000,000 to below 5,000,000: 73.3% (95% interval 59.0% to 84.0%, n=45); 9 row(s) in groups too small to show



#### q5. Taken together, which of type, amount and time step still go with fraud when the others are held fixed?

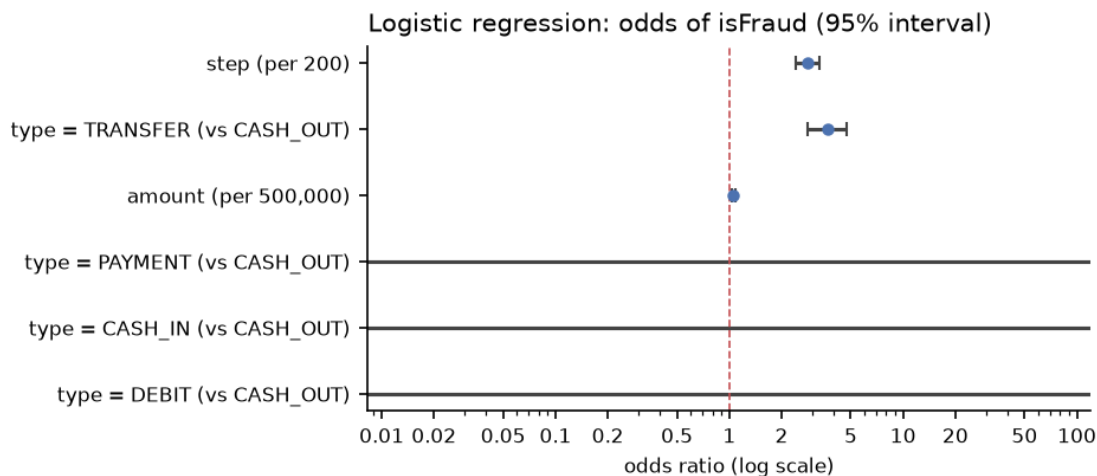
**answered.** Within the two types that carry fraud, all three features stay associated with fraud. TRANSFER goes with x3.67 the odds of CASH\_OUT, each 100 steps later with x1.68, and each extra 1,000,000 with x1.12. The full model including the zero-fraud types did not settle on stable estimates, so those types get no usable estimate.

**Among transfers and cash-outs, type, timing and amount each stay associated with fraud when the others are held fixed. (strong)**

AI-written: In a model limited to TRANSFER and CASH\_OUT (86,982 rows, 240 fraud), TRANSFER goes with x3.67 the odds of fraud against CASH\_OUT (2.82-4.77). Each 100 steps later goes with x1.68 (1.56-1.82), and each extra 1,000,000 with x1.12 (1.08-1.16), all  $p < 0.001$ . The model clearly beats no model ( $p < 0.001$ ; AUC 0.76 on the same rows). In the full model over all types, the zero-fraud types got x0.00 with intervals reaching infinity.

- Caveat: The full model (T8) warned: the model did not settle on stable estimates, often because a category has only one outcome, and effects with very wide intervals are unreliable. Only the restricted model (T11) is used for conclusions.
- Caveat: Amount enters as a straight line. This understates the sharp jump among cash-outs of 1,000,000 and over and averages the opposite patterns in the two types.
- Caveat: The step effect largely mirrors the late-month volume drop (F4).
- Caveat: AUC is measured on the same rows the model was fitted on.

T8 logistic\_regression: logistic regression of isFraud on type, amount, step: 200000 rows (240 with isFraud = 1), none left out; model vs no model (likelihood ratio)  $p < 0.001$ ; pseudo  $R^2$  0.19; AUC 0.90 on the same rows; of 6 effects, about 0.3 would show  $p < 0.05$  by chance alone  
Odds ratios (1 = no effect; 95% interval): step x2.84 per 200 (2.43-3.31,  $p < 0.001$ ); type = TRANSFER x3.67 vs CASH\_OUT (2.82-4.77,  $p < 0.001$ ); amount x1.06 per 500,000 (1.04-1.08,  $p < 0.001$ ); type = PAYMENT x0.00 vs CASH\_OUT (0.00-inf,  $p = 0.994$ ); type = CASH\_IN x0.00 vs CASH\_OUT (0.00-inf,  $p = 0.995$ ); type = DEBIT x0.00 vs CASH\_OUT (0.00-inf,  $p = 0.996$ )



T11 logistic\_regression: Within type = TRANSFER or CASH\_OUT: logistic regression of isFraud on type, amount, step: 86982 rows (240 with isFraud = 1), none left out; model vs no model (likelihood ratio)  $p < 0.001$ ; pseudo  $R^2$  0.09; AUC 0.76 on the same rows; of 3 effects, about 0.2 would show  $p < 0.05$  by chance alone  
Within type = TRANSFER or CASH\_OUT: Odds ratios (1 = no effect; 95% interval): step x1.68 per 100 (1.56-1.82,  $p < 0.001$ ); type = TRANSFER x3.67 vs CASH\_OUT (2.82-4.77,  $p < 0.001$ ); amount x1.12 per 1,000,000 (1.08-1.16,  $p < 0.001$ )

## Limitations

- The data is synthetic (PaySim) and observational, so these are associations within the simulation, not causes or real-world fraud behaviour.
- Fraud is rare (240 of 200,000 rows), so some subgroups have wide intervals, such as large cash-outs (54 rows) and amounts of exactly 10,000,000 (92 rows).
- The balance columns were excluded as the brief instructs, and account names were not analysed, so balance-draining patterns and repeat accounts are not examined.
- Timing is only available as simulation step; there is no clock time or weekday.

## Worth looking at next

- Within transfers, what does distinguish fraud from legitimate transfers, if not the amount, using only the allowed columns?
- Are fraudulent transfers and cash-outs paired in time (a transfer followed by a cash-out), as the PaySim fraud scenario suggests?
- How do fraud counts compare with legitimate counts per simulated day, once volume is taken into account?
- Does the jump in cash-out fraud sit at a particular amount between 1,000,000 and 5,000,000?

## Gate 4 checks

| Check                  | Result      | Details                                                                                                                                                               |
|------------------------|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| every_question_handled | pass        | Every plan question is answered or marked infeasible.                                                                                                                 |
| evidence_exists        | pass        | Every finding rests on tool calls that ran.                                                                                                                           |
| numbers_traceable      | pass        | Every number in the write-up is one the tools reported.                                                                                                               |
| no_overclaiming        | pass        | No finding claims more than its evidence shows.                                                                                                                       |
| model_warnings         | <b>FLAG</b> | T8 logistic_regression: The model did not settle on stable estimates, often because a category has only one outcome. Effects with very wide intervals are unreliable. |

# The plan behind the analysis

Written by the AI from your brief and a summary of the clean data (never rows), checked by Gate 3 and approved automatically. To change it, edit `edit/analysis_plan.vaml` and run the analysis again: your version wins.

AI-written: Describe what sets fraudulent transactions apart from legitimate ones in this 200,000-transaction sample. The comparison covers transaction type, amount and timing, and the results describe association only.

How the brief was read:

AI-written: I read 'characteristics' as which kinds of transactions, which amounts and which times of the simulated month go with a higher share of fraud. The four balance columns are left out, as the brief instructs. This is a description of the sample, not a fraud-detection model.

## Questions

| #  | Question                                                                                                                                | Tools                                                  | Status  |
|----|-----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|---------|
| q1 | How common is each transaction type, and what share of each type is fraudulent?                                                         | distribution,<br>target_by_group                       | planned |
| q2 | Are fraudulent transactions larger than legitimate ones, and how does the fraud rate change across amount ranges?                       | distribution,<br>compare_by_target,<br>target_by_group | planned |
| q3 | Does the fraud rate vary over the simulated month (by time step)?                                                                       | distribution,<br>target_by_group,<br>compare_by_target | planned |
| q4 | Within the transaction types where fraud occurs (for example TRANSFER and CASH_OUT), do fraudulent amounts differ from legitimate ones? | target_by_two_groups,<br>compare_by_target             | planned |
| q5 | Taken together, which of type, amount and time step still go with fraud when the others are held fixed?                                 | logistic_regression                                    | planned |

## Left out of the analysis

- `oldbalanceOrg` (brief): The brief excludes it: balances of fraudulent transactions are annulled, so they distort fraud analysis.
- `newbalanceOrig` (brief): The brief excludes it: balances of fraudulent transactions are annulled, so they distort fraud analysis.
- `oldbalanceDest` (brief): The brief excludes it: balances of fraudulent transactions are annulled, so they distort fraud analysis.
- `newbalanceDest` (brief): The brief excludes it: balances of fraudulent transactions are annulled, so they distort fraud analysis.
- `isFlaggedFraud` (planner): Only 1 of 200,000 rows is flagged, so it cannot support any comparison. It is also a rule on large transfers, tied to the target by construction.

## Limitations

- The data is observational and synthetic (PaySim simulation), so the findings describe associations in the simulation and do not show cause and effect or real-world fraud behaviour.
- Fraud is very rare (240 of 200,000, 0.12%), so rates for small groups have wide intervals and some subgroup comparisons may be too thin to report.
- The balance columns are excluded as the brief instructs, so any pattern in account balances is not examined.
- `isFlaggedFraud` has a single flagged row and is not used, so the existing flagging rule cannot be evaluated.
- Account names are personal data and are not analysed, so repeat offenders or recipient patterns cannot be examined.
- Only step (hours) is available for timing, with no clock time or weekday, and types with zero fraud may make regression estimates unstable.

## Gate 3 checks

| Check                              | Result | Details                                                            |
|------------------------------------|--------|--------------------------------------------------------------------|
| <code>columns_exist</code>         | pass   | Every column exists.                                               |
| <code>methods_exist</code>         | pass   | Every method is a tool.                                            |
| <code>target_as_briefed</code>     | pass   | The target is the brief's and can be analysed.                     |
| <code>task_type_fits_target</code> | pass   | The task type fits the target.                                     |
| <code>usable_columns</code>        | pass   | No identifier, personal, free-text or excluded column is analysed. |
| <code>exclusions_listed</code>     | pass   | Every excluded column is listed with its reason.                   |

| Check                  | Result      | Details                                                                                                                                                                                                                                                                                                                                                                                                  |
|------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| methods_fit_target     | pass        | Every method can run on this target.                                                                                                                                                                                                                                                                                                                                                                     |
| leaky_columns_excluded | pass        | No column that gives the answer away is used.                                                                                                                                                                                                                                                                                                                                                            |
| feasible               | pass        | Every question has enough rows, or is marked infeasible.                                                                                                                                                                                                                                                                                                                                                 |
| methods_fit_question   | <b>NOTE</b> | q1 is a segment question, but distribution is meant for distribution or data_quality questions; q2 is a relationship question, but distribution is meant for distribution or data_quality questions; q3 is a relationship question, but distribution is meant for distribution or data_quality questions; q4 is a segment question, but compare_by_target is meant for relationship or drivers questions |
| something_to_do        | pass        | At least one question can be answered.                                                                                                                                                                                                                                                                                                                                                                   |