

EDA Report

Generated: 2026-01-30 16:09:53

Dataset Overview

- Data path: data/synthetic_aml_200k_20260130_135951.csv
- Rows used: 10000
- Target column: is_suspicious
- Time column: txn_ts
- Sample fraction: 0.05

Data Quality

- Dataset has 10000 rows and 25 columns.
- Duplicate row ratio: 0.00%.
- Rows limited to 10000 from 200000 for analysis.

Missingness

Column	Missing Count	Missing Rate
txn_id	0	0.00%
account_id	0	0.00%
customer_id	0	0.00%
merchant_id	0	0.00%
counterparty_id	0	0.00%
txn_ts	0	0.00%
txn_amount	0	0.00%
account_balance	0	0.00%
num_txn_1h	0	0.00%
num_txn_24h	0	0.00%
sum_amount_24h	0	0.00%
days_since_last_txn	0	0.00%
velocity_score	0	0.00%
channel	0	0.00%
txn_type	0	0.00%
origin_country	0	0.00%
dest_country	0	0.00%
currency	0	0.00%
device_type	0	0.00%
risk_segment	0	0.00%
is_international	0	0.00%
is_new_beneficiary	0	0.00%
merchant_name	0	0.00%
payment_memo	0	0.00%
is_suspicious	0	0.00%

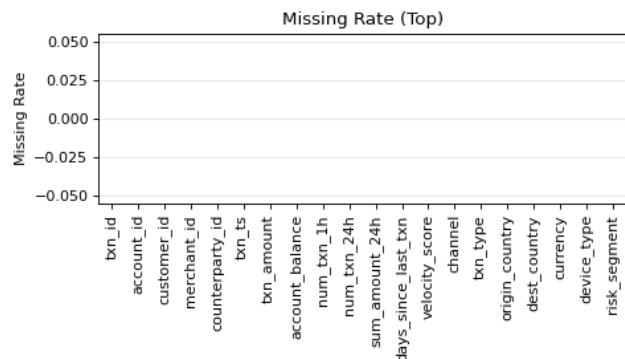
Column Type Classification

Type	Columns
numeric	txn_amount
numeric	account_balance
numeric	num_txn_1h
numeric	num_txn_24h
numeric	sum_amount_24h
numeric	days_since_last_txn
numeric	velocity_score
numeric	is_suspicious
categorical	channel
categorical	txn_type
categorical	origin_country
categorical	dest_country
categorical	currency
categorical	device_type
categorical	risk_segment
categorical	merchant_name
categorical	payment_memo
datetime	txn_ts
boolean	is_international
boolean	is_new_beneficiary
text	txn_id
text	account_id
text	customer_id
text	merchant_id
text	counterparty_id

Outlier Ratio (IQR)

Column	Outlier Ratio
sum_amount_24h	0.090
txn_amount	0.081
account_balance	0.056
velocity_score	0.008
num_txn_24h	0.005
num_txn_1h	0.002
days_since_last_txn	0

Charts



Target

- Target 'is_suspicious' has 2 classes.
- Target rate over time plotted.

Target Distribution

Value	Count	Rate
0	9,404	94.04%
1	596	5.96%

Target Rate by origin_country

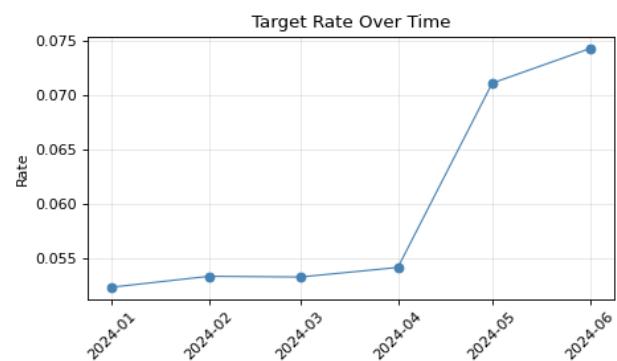
origin_country	Target Rate
PK	8.85%
AE	6.91%
HK	6.50%
NG	6.13%
RU	6.07%
DE	5.96%
US	5.68%
SG	5.55%
BR	5.31%
GB	5.24%

Target Rate by dest_country

dest_country	Target Rate
PK	7.34%
DE	6.62%
RU	6.40%
BR	6.23%
GB	5.87%
HK	5.85%
NG	5.78%
SG	5.53%
US	5.07%
AE	4.86%

Target Rate by payment_memo

payment_memo	Target Rate
transfer	6.39%
invoice	6.36%
refund	6.04%
cashout	5.88%
fee	5.78%
subscription	5.26%

Charts

Univariate - Summary

Univariate

- Numeric columns analyzed: txn_amount, account_balance, num_txn_1h, num_txn_24h, sum_amount_24h, days_since_last_txn, velocity_score, is_suspicious
- Categorical columns analyzed: origin_country, dest_country, payment_memo, merchant_name, channel, txn_type, device_type, risk_segment, currency, is_international

Numeric Summary Statistics

Column	count	mean	std	min	25%	50%	75%	max
txn_amount	10,000	37.425	47.614	0.430	11.440	22.495	44.450	778.290
account_balance	10,000	3,780.511	2,984.947	248.790	1,846.430	2,964.430	4,747.602	41,182.010
num_txn_1h	10,000	1.289	1.140	0	0	1	2	8
num_txn_24h	10,000	4.213	2.069	0	3	4	6	14
sum_amount_24h	10,000	159.630	254.302	0	36.138	83.150	179.132	4,028.050
days_since_last_txn	10,000	14.490	8.620	0	7	14	22	29
velocity_score	10,000	1.306	0.702	0	0.801	1.264	1.760	4.504
is_suspicious	10,000	0.060	0.237	0	0	0	0	1

Categorical Frequency (Top K)

Top K: origin_country

Category	Count	Rate
US	2,551	25.51%
BR	1,336	13.36%
GB	1,183	11.83%
DE	1,108	11.08%
SG	775	7.75%
HK	677	6.77%
RU	626	6.26%
NG	604	6.04%
PK	576	5.76%
AE	564	5.64%

Top K: dest_country

Category	Count	Rate
AE	1,028	10.28%
DE	1,027	10.27%
SG	1,012	10.12%
BR	1,011	10.11%
PK	1,008	10.08%
NG	1,003	10.03%
RU	1,000	10.00%
GB	988	9.88%
US	966	9.66%

Category	Count	Rate
HK	957	9.57%

Top K: payment_memo

Category	Count	Rate
transfer	1,752	17.52%
fee	1,713	17.13%
cashout	1,666	16.66%
refund	1,657	16.57%
invoice	1,634	16.34%
subscription	1,578	15.78%

Top K: merchant_name

Category	Count	Rate
Bright Mart	2,091	20.91%
Alpha Shop	2,011	20.11%
Delta Foods	2,005	20.05%
Cobalt LLC	1,981	19.81%
Echo Retail	1,912	19.12%

Top K: channel

Category	Count	Rate
card	4,028	40.28%
ach	1,993	19.93%
cash	1,513	15.13%
wire	1,446	14.46%
crypto	1,020	10.20%

Top K: txn_type

Category	Count	Rate
purchase	5,967	59.67%
transfer	2,506	25.06%
cash_out	1,036	10.36%
refund	491	4.91%

Top K: device_type

Category	Count	Rate
ios	4,062	40.62%
android	4,043	40.43%
web	1,895	18.95%

Top K: risk_segment

Category	Count	Rate
low	6,570	65.70%
mid	2,429	24.29%

Category	Count	Rate
high	1,001	10.01%

Top K: currency

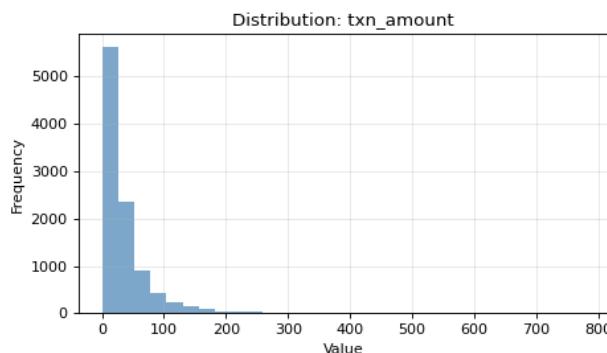
Category	Count	Rate
USD	8,892	88.92%
EUR	1,108	11.08%

Top K: is_international

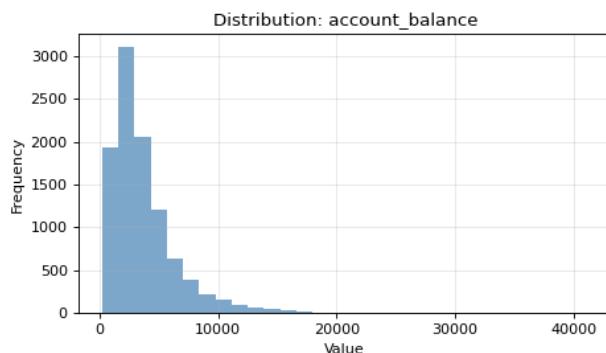
Category	Count	Rate
True	8,963	89.63%
False	1,037	10.37%

Univariate Plots

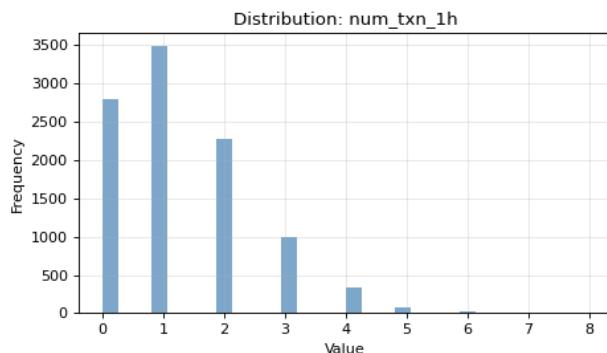
txn_amount distribution



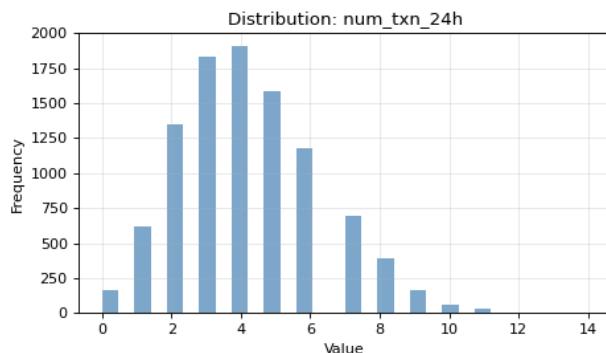
account_balance distribution



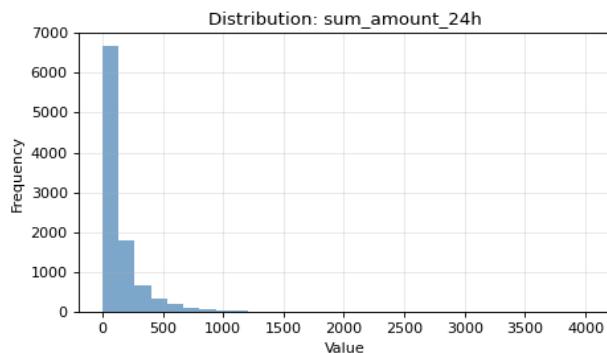
num_txn_1h distribution



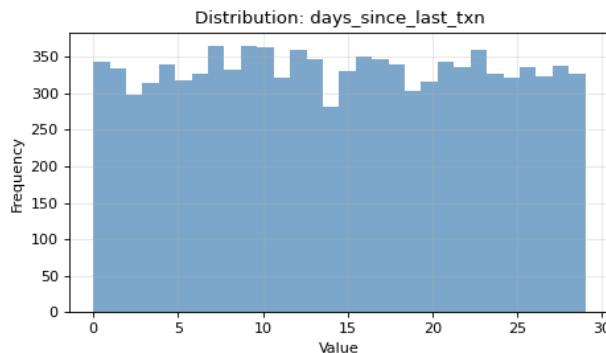
num_txn_24h distribution



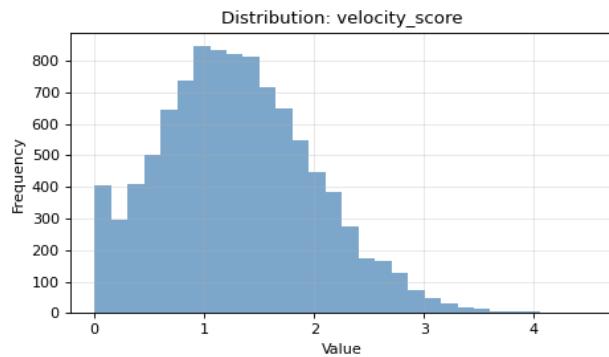
sum_amount_24h distribution



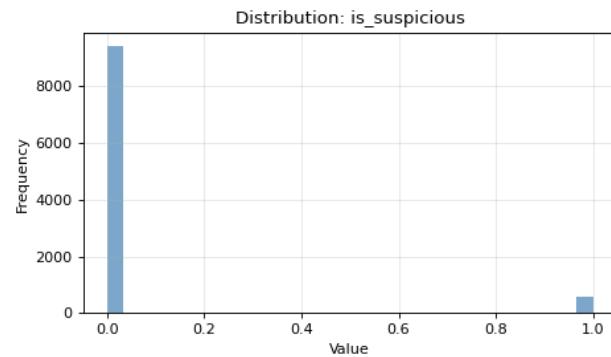
days_since_last_txn distribution



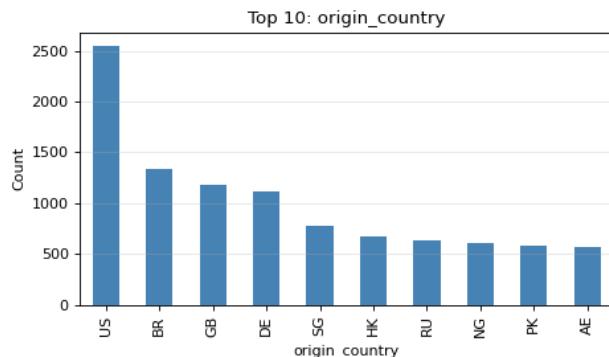
velocity_score distribution



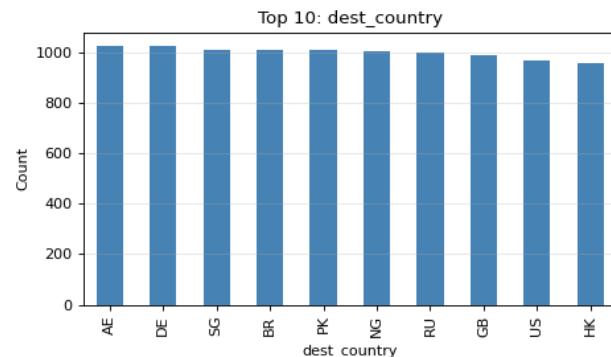
is_suspicious distribution



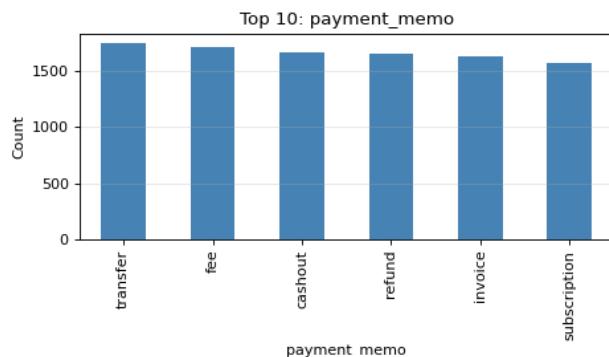
origin_country top 10



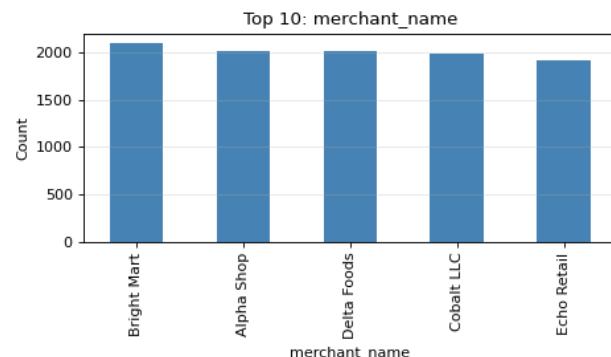
dest_country top 10



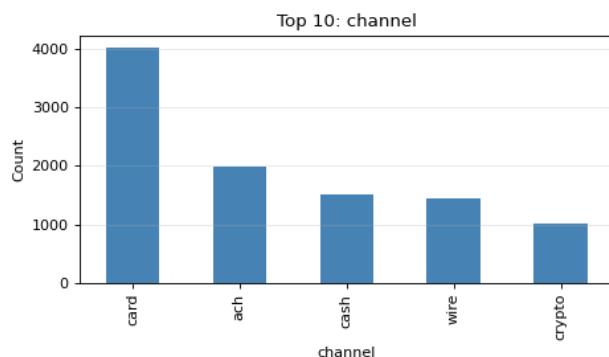
payment_memo top 10



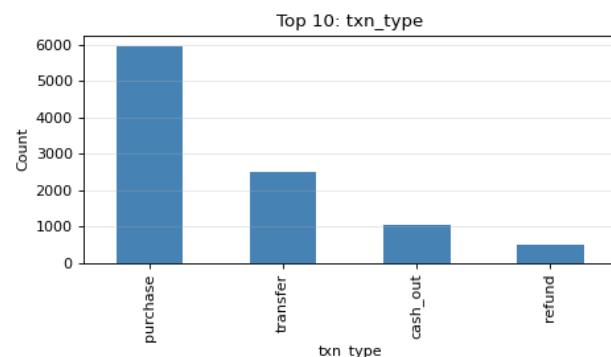
merchant_name top 10



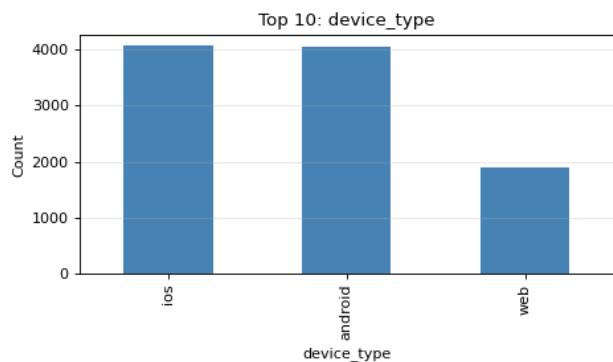
channel top 10



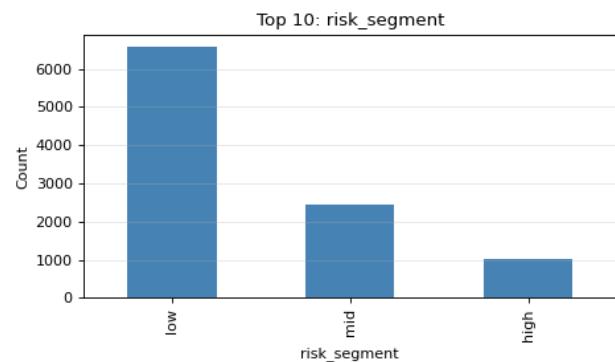
txn_type top 10



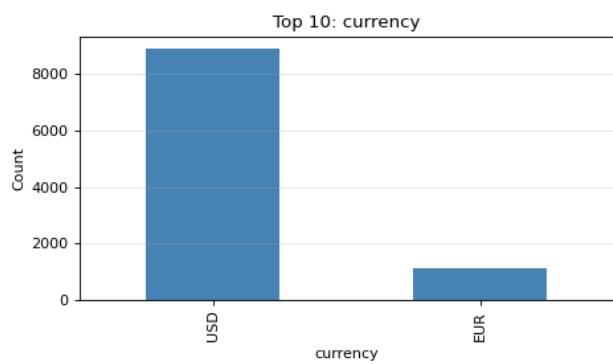
device_type top 10



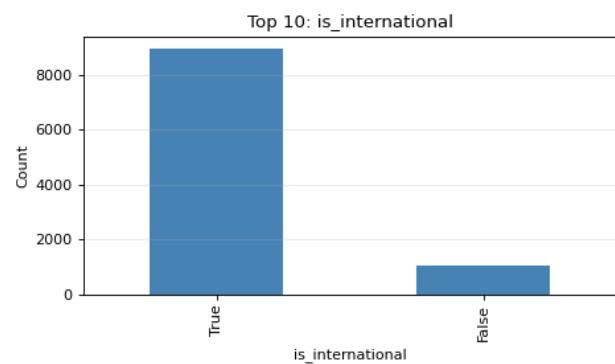
risk_segment top 10



currency top 10



is_international top 10



Bivariate Target

- Bivariate target analysis completed.

Numeric vs Target (Binned)

Column	Bin	Target Rate
txn_amount	(0.429, 9.728]	6.00%
txn_amount	(9.728, 17.486]	6.05%
txn_amount	(17.486, 29.084]	5.80%
txn_amount	(29.084, 52.39]	5.64%
txn_amount	(52.39, 778.29]	6.31%
account_balance	(248.789, 1641.066]	6.25%
account_balance	(1641.066, 2481.73]	6.25%
account_balance	(2481.73, 3541.17]	5.70%
account_balance	(3541.17, 5292.994]	6.35%
account_balance	(5292.994, 41182.01]	5.25%
num_txn_1h	(-0.001, 1.0]	5.71%
num_txn_1h	(1.0, 2.0]	5.69%
num_txn_1h	(2.0, 8.0]	7.49%
num_txn_24h	(-0.001, 2.0]	5.82%
num_txn_24h	(2.0, 4.0]	6.18%
num_txn_24h	(4.0, 5.0]	5.22%
num_txn_24h	(5.0, 6.0]	6.62%
num_txn_24h	(6.0, 14.0]	5.86%
sum_amount_24h	(-0.001, 29.35]	5.85%
sum_amount_24h	(29.35, 60.96]	6.25%
sum_amount_24h	(60.96, 110.294]	5.55%
sum_amount_24h	(110.294, 218.724]	5.90%
sum_amount_24h	(218.724, 4028.05]	6.25%
days_since_last_txn	(-0.001, 6.0]	6.42%
days_since_last_txn	(6.0, 11.0]	6.53%
days_since_last_txn	(11.0, 17.0]	6.11%
days_since_last_txn	(17.0, 23.0]	4.96%
days_since_last_txn	(23.0, 29.0]	5.79%
velocity_score	(-0.001, 0.697]	5.70%
velocity_score	(0.697, 1.077]	6.15%
velocity_score	(1.077, 1.446]	5.69%
velocity_score	(1.446, 1.891]	5.71%
velocity_score	(1.891, 4.504]	6.56%
is_suspicious	(-0.001, 1.0]	5.96%

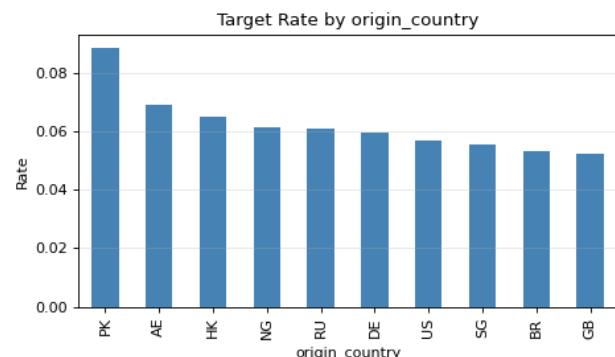
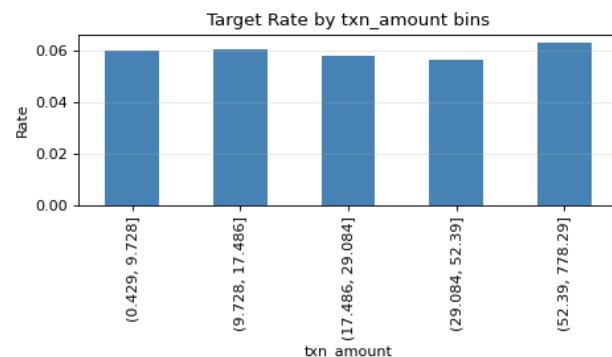
Categorical vs Target

Column	Category	Target Rate
origin_country	PK	8.85%
origin_country	AE	6.91%

Column	Category	Target Rate
origin_country	HK	6.50%
origin_country	NG	6.13%
origin_country	RU	6.07%
origin_country	DE	5.96%
origin_country	US	5.68%
origin_country	SG	5.55%
origin_country	BR	5.31%
origin_country	GB	5.24%
dest_country	PK	7.34%
dest_country	DE	6.62%
dest_country	RU	6.40%
dest_country	BR	6.23%
dest_country	GB	5.87%
dest_country	HK	5.85%
dest_country	NG	5.78%
dest_country	SG	5.53%
dest_country	US	5.07%
dest_country	AE	4.86%
payment_memo	transfer	6.39%
payment_memo	invoice	6.36%
payment_memo	refund	6.04%
payment_memo	cashout	5.88%
payment_memo	fee	5.78%
payment_memo	subscription	5.26%
merchant_name	Bright Mart	6.93%
merchant_name	Echo Retail	6.85%
merchant_name	Cobalt LLC	5.65%
merchant_name	Delta Foods	5.54%
merchant_name	Alpha Shop	4.82%
channel	cash	6.61%
channel	card	6.16%
channel	ach	5.82%
channel	crypto	5.69%
channel	wire	5.12%
txn_type	refund	6.52%
txn_type	purchase	6.07%
txn_type	transfer	5.91%
txn_type	cash_out	5.21%
device_type	web	6.97%
device_type	ios	5.93%
device_type	android	5.52%
risk_segment	high	6.39%
risk_segment	low	6.01%

Column	Category	Target Rate
risk_segment	mid	5.64%
currency	USD	5.96%
currency	EUR	5.96%
is_international	False	7.52%
is_international	True	5.78%

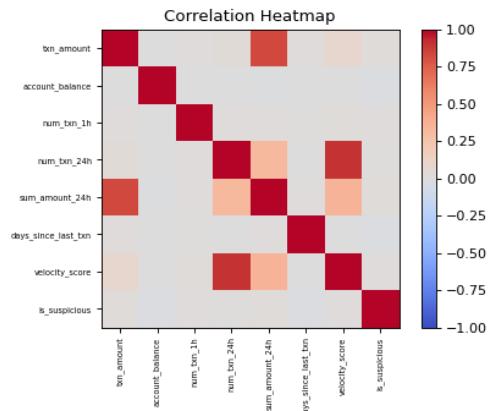
Charts



Feature Vs Feature

- No highly correlated feature pairs detected.

Charts



Time Drift

- Time series and drift analysis completed.

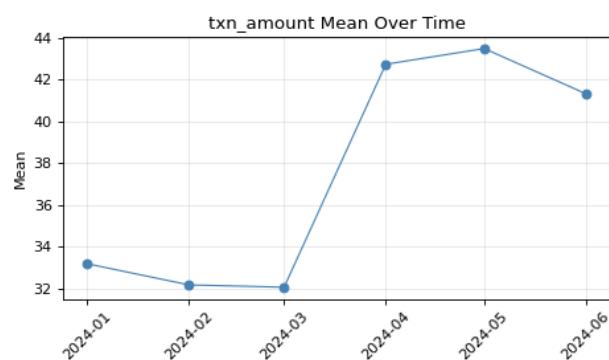
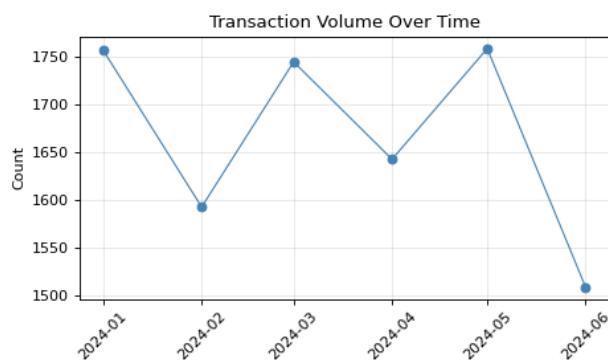
PSI Drift (Numeric)

Column	PSI
sum_amount_24h	0.129
txn_amount	0.086
velocity_score	0.065
num_txn_24h	0.062
num_txn_1h	0.023
days_since_last_txn	0.004
account_balance	0.004
is_suspicious	0

Categorical Drift (Total Variation)

Column	Drift Score
channel	0.103
origin_country	0.082
dest_country	0.025
txn_type	0.018
payment_memo	0.015
currency	0.012
merchant_name	0.011
device_type	0.010
is_international	0.009
risk_segment	0.005

Charts



Summary

- Heavy-tailed numeric features: ['sum_amount_24h', 'txn_amount', 'is_suspicious', 'account_balance'].
- Next steps: consider feature engineering, handling missing values, and monitoring drift.

Skipped EDA Sections

No sections were skipped.

Run Configuration

Key	Value
data_path	data/synthetic_aml_200k_20260130_135951.csv
rows_original	200,000
rows_used	10,000
target_col	is_suspicious
time_col	txn_ts
time_parse_ratio	1
sample_frac	0.050