Fraud Alert Exploratory Analysis

# 📊 Dataset Overview

Shape: 200000 rows × 24 columns

Columns and Data Types:

|  |  |
| --- | --- |
| Column | Data Type |
| transaction\_id | object |
| user\_id | int64 |
| event\_time | datetime64[ns] |
| amount | float64 |
| currency | object |
| merchant\_id | int64 |
| merchant\_category | object |
| device\_id | object |
| ip\_address | object |
| country | object |
| lat | float64 |
| lon | float64 |
| channel | object |
| payment\_method | object |
| velocity\_1h | int64 |
| velocity\_24h | int64 |
| geo\_distance\_km | float64 |
| behavior\_typing\_speed\_cps | float64 |
| behavior\_navigation\_speed | float64 |
| is\_fraud | int64 |
| fraud\_type | object |
| hour | int32 |
| dayofweek | int32 |
| is\_weekend | int64 |

## 🔍 First 5 Rows

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| transaction\_id | user\_id | event\_time | amount | currency | merchant\_id | merchant\_category | device\_id | ip\_address | country | lat | lon | channel | payment\_method | velocity\_1h | velocity\_24h | geo\_distance\_km | behavior\_typing\_speed\_cps | behavior\_navigation\_speed | is\_fraud | fraud\_type | hour | dayofweek | is\_weekend |
| 8a145a9c-dab4-4c2e-8be0-aae1fd9bd696 | 2978 | 2025-01-01 00:00:00 | 88.81 | INR | 95929 | electronics | dev-2978 | 62.8.18.6 | SM | 1.779572 | 103.835662 | pos | wallet | 9 | 9 | 47.127 | 4.615 | 1.484 | 1 | bot\_burst | 0 | 2 | 0 |
| 25d683fe-1fd2-4cda-a3a3-8cb3d9a6b82e | 146 | 2025-01-01 00:00:00 | 208.02 | INR | 89813 | electronics | dev-146 | 181.178.111.158 | HR | 23.831045 | 52.970005 | web | netbanking | 18 | 18 | 28.081 | 6.701 | 2.127 | 1 | bot\_burst | 0 | 2 | 0 |
| 1cdc34a0-24bb-4a07-a77d-4fa3a9785b11 | 934 | 2025-01-01 00:00:00 | 467.58 | INR | 43058 | electronics | dev-934 | 12.124.150.66 | GN | 20.502734 | 78.819584 | pos | netbanking | 4 | 4 | 26.885 | 5.947 | 2.184 | 0 |  | 0 | 2 | 0 |
| da7de96e-86b4-4232-9231-cffc6bfa50d2 | 1269 | 2025-01-01 00:00:00 | 569.37 | INR | 44711 | travel | fda3f33e-d365-4277-ad0d-efe000672fda | 195.30.13.123 | TO | 23.67316 | 53.529402 | app | upi | 26 | 26 | 20.471 | 7.826 | 2.212 | 0 |  | 0 | 2 | 0 |
| ec7a3d88-e7b5-4854-b02f-d2d48c2ad8ca | 4597 | 2025-01-01 00:00:00 | 255.61 | INR | 38014 | crypto | dev-4597 | 90.92.121.169 | MH | 23.571227 | 54.133372 | web | netbanking | 29 | 29 | 62.999 | 5.768 | 2.604 | 1 | bot\_burst | 0 | 2 | 0 |

## 🔍 Last 5 Rows

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| transaction\_id | user\_id | event\_time | amount | currency | merchant\_id | merchant\_category | device\_id | ip\_address | country | lat | lon | channel | payment\_method | velocity\_1h | velocity\_24h | geo\_distance\_km | behavior\_typing\_speed\_cps | behavior\_navigation\_speed | is\_fraud | fraud\_type | hour | dayofweek | is\_weekend |
| 34d82878-07f4-4b66-bf77-9add0049678c | 3050 | 2025-03-01 23:58:00 | 57.43 | INR | 32370 | food | 2347bee4-57a8-40bd-9897-9f91e9284f64 | 37.223.200.224 | CD | 37.45483 | -95.258165 | app | upi | 0 | 0 | 31.159 | 6.838 | 1.147 | 0 |  | 23 | 5 | 1 |
| 106babab-733b-485c-8e79-041f85ebb8f2 | 3729 | 2025-03-01 23:59:00 | 212.61 | INR | 74603 | crypto | dev-3729 | 55.129.155.123 | SG | 1.463679 | 103.760207 | web | netbanking | 0 | 1 | 46.774 | 6.095 | 1.766 | 0 |  | 23 | 5 | 1 |
| b8d34573-fb90-4746-baa6-0b66bbc31880 | 3620 | 2025-03-01 23:59:00 | 110.04 | INR | 63254 | gaming | dev-3620 | 86.245.9.73 | ZA | 36.585014 | -96.060275 | web | netbanking | 0 | 0 | 33.036 | 4.878 | 2.053 | 0 |  | 23 | 5 | 1 |
| 29a2c683-1afb-4e1f-aa7d-ac2db7911661 | 2223 | 2025-03-01 23:59:00 | 144.56 | INR | 90248 | retail | dev-2223 | 47.149.177.2 | BA | 23.461624 | 54.152331 | web | netbanking | 0 | 0 | 25.777 | 4.319 | 1.452 | 0 |  | 23 | 5 | 1 |
| cc980934-6be1-4545-895e-7b2c48442e2c | 813 | 2025-03-01 23:59:00 | 541.82 | INR | 24104 | crypto | dev-813 | 57.168.64.233 | PS | 55.078861 | -3.369162 | app | netbanking | 0 | 1 | 9.4 | 6.709 | 2.39 | 0 |  | 23 | 5 | 1 |

## 📈 Summary Statistics

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Stat | user\_id | event\_time | amount | merchant\_id | lat | lon | velocity\_1h | velocity\_24h | geo\_distance\_km | behavior\_typing\_speed\_cps | behavior\_navigation\_speed | is\_fraud | hour | dayofweek | is\_weekend |
| count | 200000.0 | 200000 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 | 200000.0 |
| mean | 2499.99 | 2025-01-31 00:23:15.015900160 | 296.52 | 55000.38 | 27.51 | 27.55 | 10.06 | 10.38 | 385.54 | 5.01 | 2.0 | 0.53 | 11.49 | 3.03 | 0.28 |
| min | 1.0 | 2025-01-01 00:00:00 | 3.6 | 10000.0 | -0.58 | -98.02 | 0.0 | 0.0 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| 25% | 1248.0 | 2025-01-16 02:28:00 | 95.26 | 32402.0 | 20.3 | -3.71 | 2.0 | 3.0 | 16.03 | 4.18 | 1.59 | 0.0 | 5.0 | 1.0 | 0.0 |
| 50% | 2499.0 | 2025-01-30 23:20:30 | 182.73 | 55035.0 | 23.42 | 53.85 | 7.0 | 8.0 | 23.96 | 5.01 | 1.99 | 1.0 | 11.0 | 3.0 | 0.0 |
| 75% | 3752.0 | 2025-02-14 23:44:15 | 356.84 | 77612.0 | 37.36 | 79.25 | 15.0 | 16.0 | 31.17 | 5.84 | 2.4 | 1.0 | 17.0 | 5.0 | 1.0 |
| max | 5000.0 | 2025-03-01 23:59:00 | 15246.85 | 99998.0 | 57.4 | 105.8 | 62.0 | 62.0 | 15463.44 | 10.23 | 4.5 | 1.0 | 23.0 | 6.0 | 1.0 |
| std | 1444.61 | nan | 365.01 | 26033.11 | 18.13 | 71.39 | 9.45 | 9.58 | 1904.47 | 1.23 | 0.59 | 0.5 | 6.93 | 1.96 | 0.45 |

## 🧼 Missing Values

No missing values detected.

## ⚖️ Fraud Distribution

Fraud Rate: 52.93%

Class Distribution:

|  |  |
| --- | --- |
| Class | Proportion |
| 1 | 52.93% |
| 0 | 47.07% |

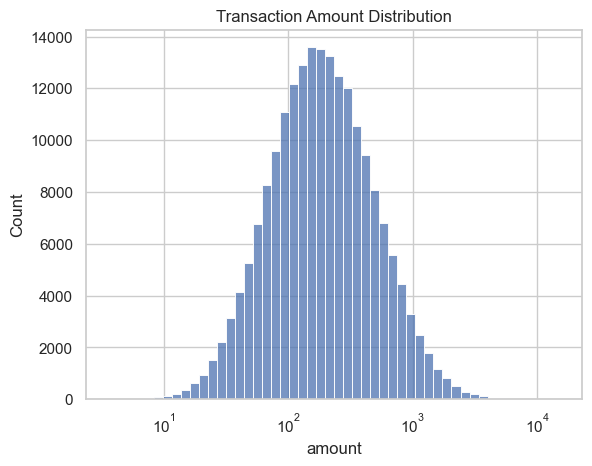
## Amount By Fraud

Amount by Fraud Status  
  
Compares transaction amounts between fraud and non-fraud cases.



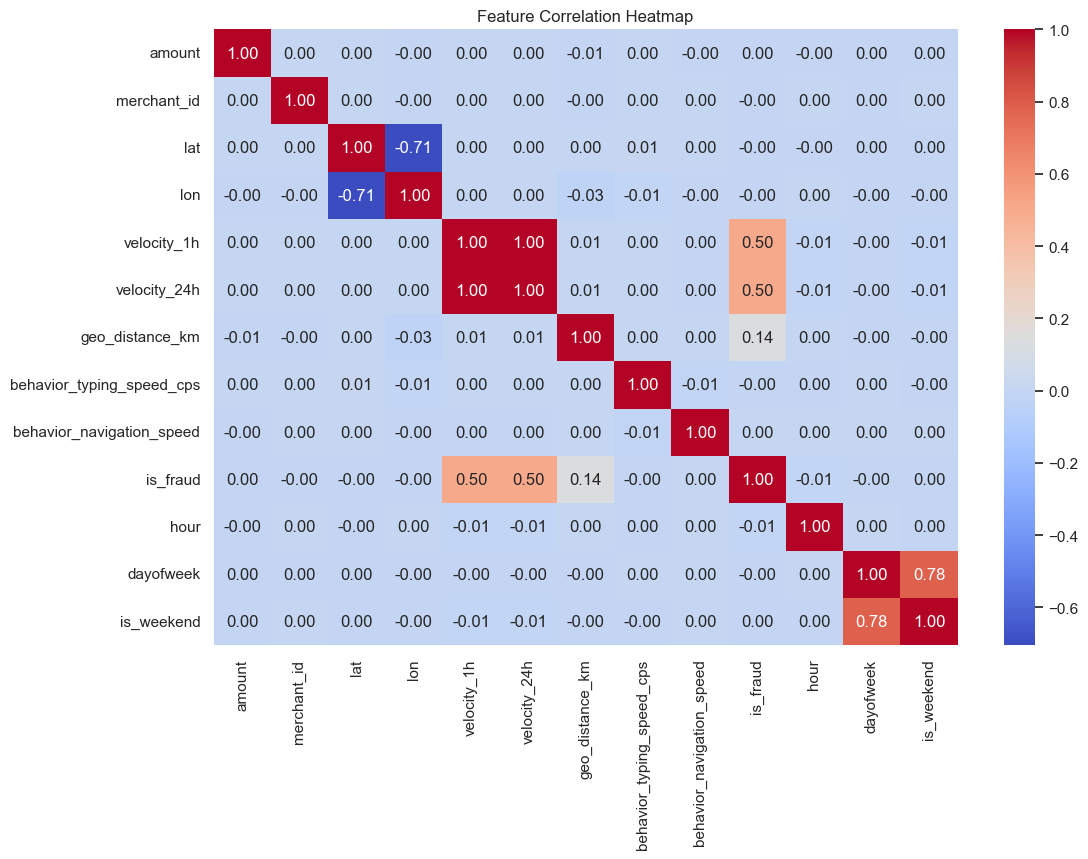
## Amount Distribution

Transaction Amount Distribution  
  
Shows how transaction amounts are distributed. Log scale helps visualize skewed data.



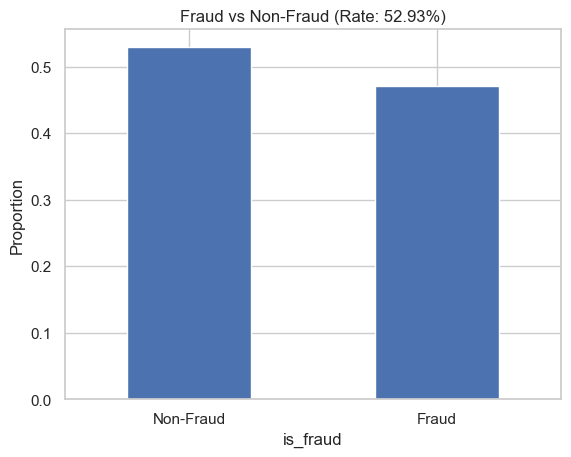
## Feature Correlation Heatmap

Feature Correlation Heatmap  
  
Shows how numeric features correlate with each other. Useful for spotting redundancy or strong fraud signals.



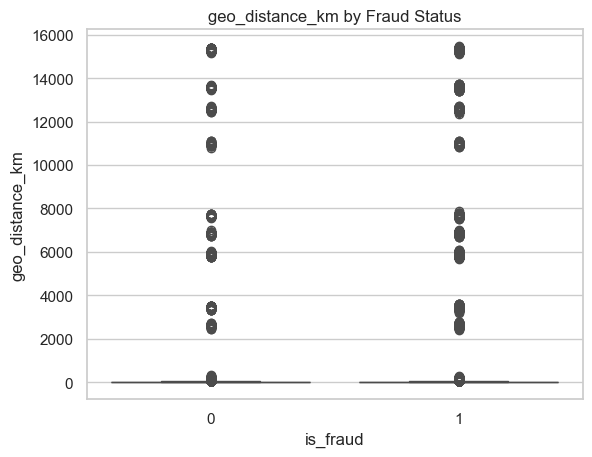
## Fraud Distribution

Fraud vs Non-Fraud Distribution  
  
Shows the proportion of fraudulent vs non-fraudulent transactions. Useful to understand class imbalance.



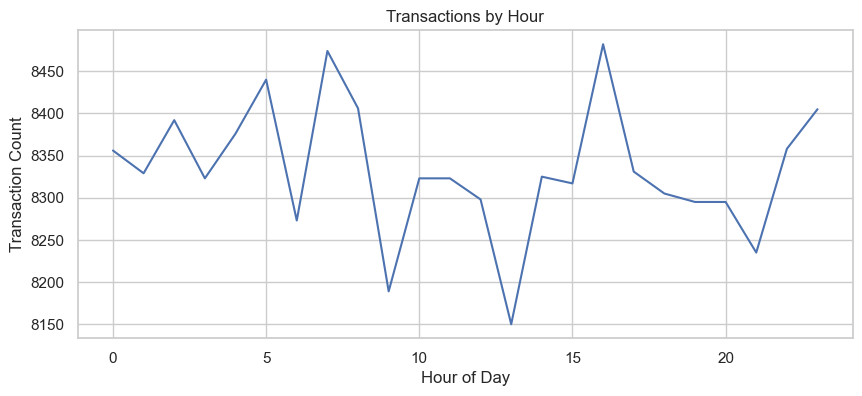
## Geo Distance Km By Fraud

geo\_distance\_km by Fraud Status  
  
Shows how geo distance km differs between fraud and non-fraud transactions.



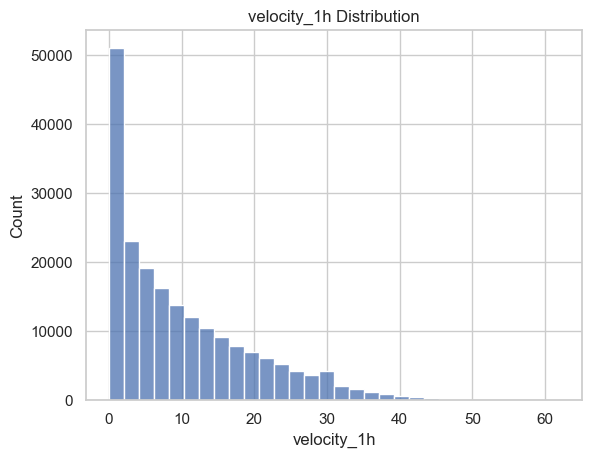
## Transactions By Hour

Transactions by Hour  
  
Shows transaction volume by hour of day. Useful for identifying peak fraud windows.



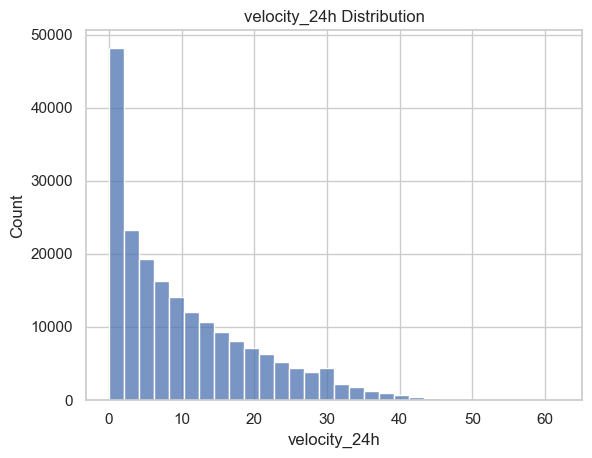
## Velocity 1H Distribution

velocity\_1h Distribution  
  
Shows how velocity 1h varies across transactions.



## Velocity 24H Distribution

velocity\_24h Distribution  
  
Shows how velocity 24h varies across transactions.



# 🧾 Summary Insights

✅ Exploratory Analysis Complete  
Fraud Rate: 52.93%  
Missing Columns: []  
Consider engineering more behavioral features or fixing velocity merge issues.