Uncovering Financial Fraud Patterns: An Exploratory Data Analysis

Discovering Hidden Patterns in 636,620 Financial Transactions

Nguyen Tu Nhu / July 2025

The Investigation at a Glance

Objective

Conduct comprehensive exploratory data analysis to uncover **hidden patterns** and **characteristics** in financial transaction fraud

Dataset

- **600K+ transactions** analyzed
- **11 features**
- **5 transaction types** across the financial ecosystem

Key Discovery

Fraudulent transactions exhibit **distinct behavioral patterns** that **differentiate them from legitimate transactions** across multiple dimensions

Analytics Value

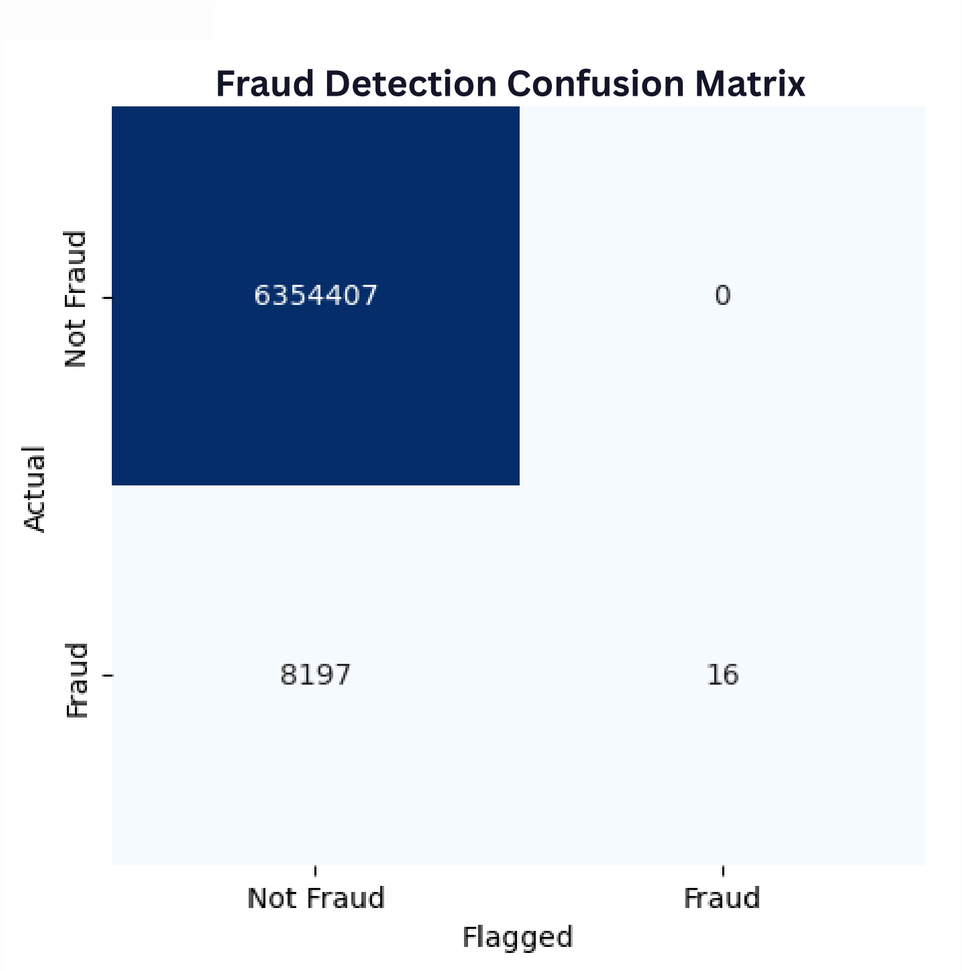
Identified **actionable patterns** and **insights** to inform fraud detection strategies and risk management

The Financial Fraud Landscape

Industry Context

- **Global fraud losses**: $5.1 trillion annually
- **Detection challenges**: High true negative (8197 **VS.** 16 true positive cases)

Pattern recognition: Key to reducing both fraud and customer friction



Our Financial Transaction Universe

Data Characterstics

📊 **Total Transactions**: 636,620
🏦 **Transaction Types**: 5 (CASH_OUT, PAYMENT, CASH_IN, TRANSFER, DEBIT)
💳 **Features**: 11 (amounts, balances, timestamps, fraud flags)
⏰ **Time Range**: 743 sequential steps (hours)
🎯 **Fraud Cases**: 16 confirmed instances
📈 **Flag System**: 8,213 transactions flagged by current system

Feature Overview

Feature	Description	Type
step	Time sequence	Numeric
type	Transaction category	Categorical
amount	Transaction value	Numeric
nameOrig/nameDest	Account identifiers	Categorical
oldbalanceOrg/newbalanceOrig	Origin account balances	Numeric
oldbalanceDest/newbalanceDest	Destination account balances	Numeric
isFlagged	System fraud flag	Binary
isFraud	Confirmed fraud status	Binary

Methodology

Theoretical Approach

Research Foundation

Behavioral Economics
Fraud patterns reflect criminal decision-making processes

Network Analysis
Transaction flows reveal systematic vulnerabilities

Statistical Anomaly Detection
Outliers often indicate fraudulent behavior

Data Analysis Approach

Analysis Framework

Univariate Analysis
Individual feature distributions and characteristics

Multivariate Analysis
Complex pattern interactions

Statistical Techniques

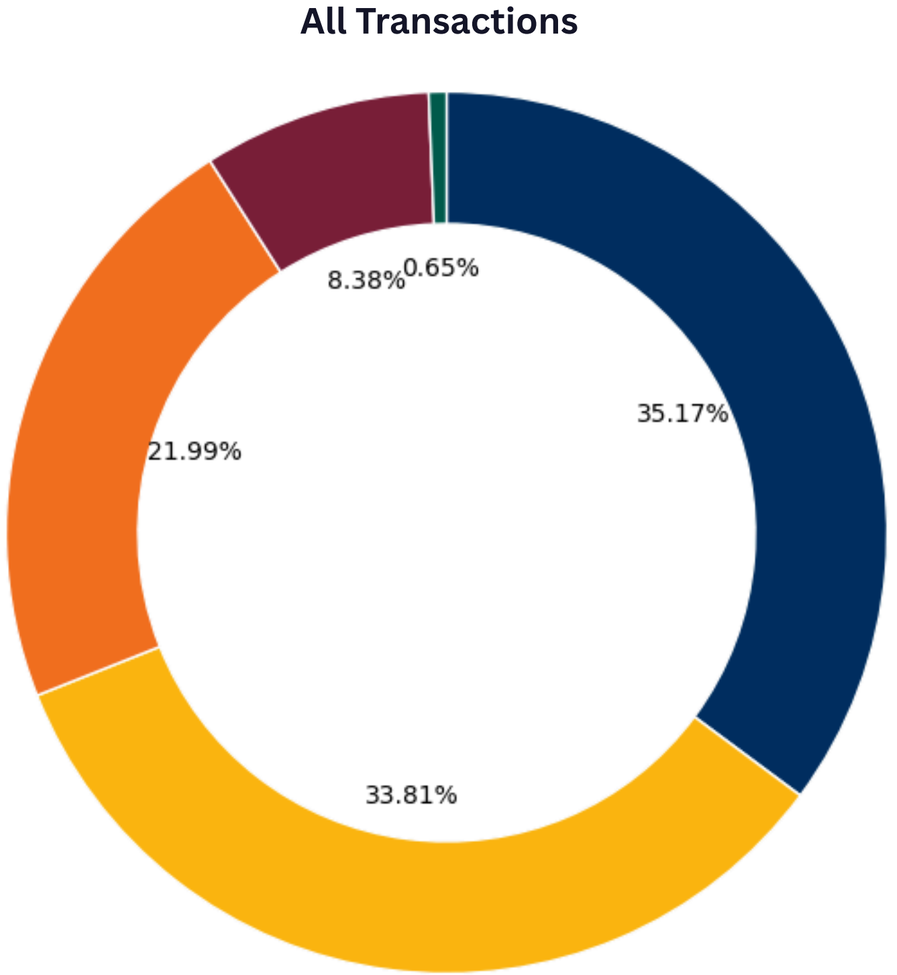
Descriptive Statistics
Distribution

Comparative Analysis
Fraud vs. legitimate transaction contrasts

***Do not include balance of to-merchant transactions**

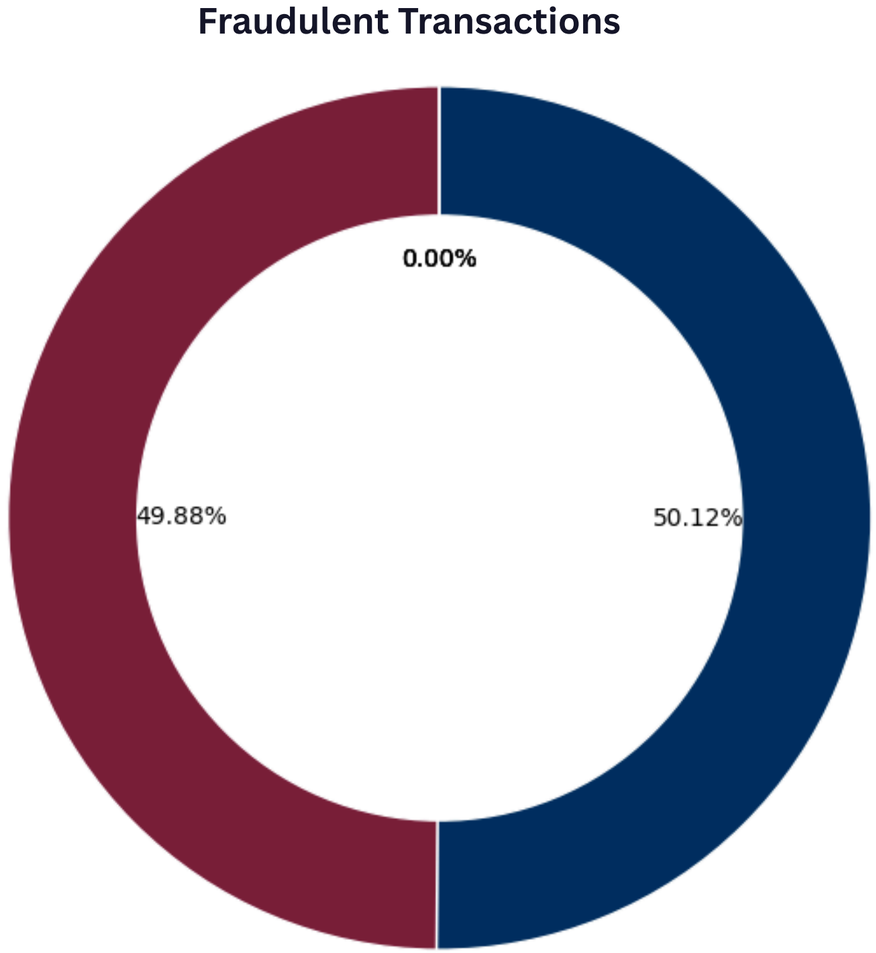
Transaction Types

- CASH_OUT **(35.17%)** - Largest transaction category
- PAYMENT **(33.81%) - Regular merchant payments**
- CASH_IN **(21.99%)** - Money deposits
- TRANSFER **(8.38%)** - Direct account transfers
- DEBIT **(0.65%)** - Smallest category

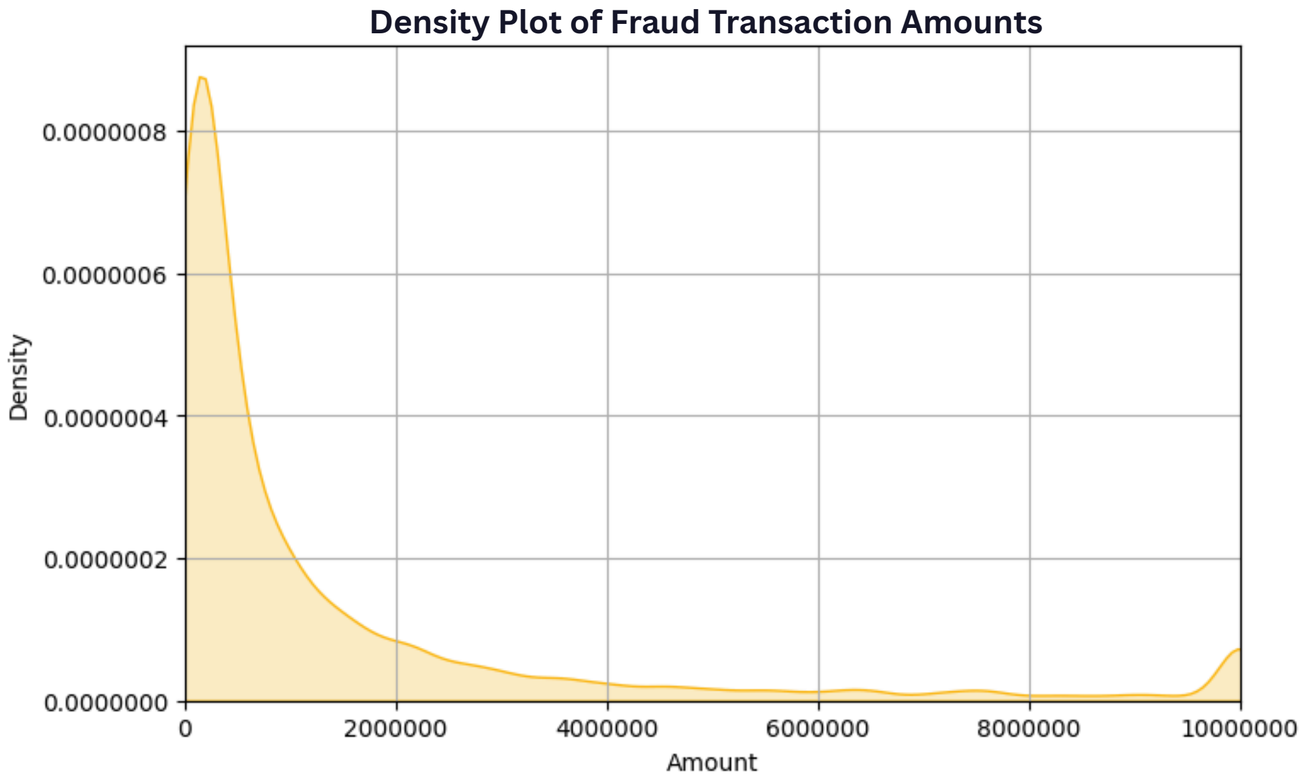


Transaction Types

Channel Preference
Fraudsters exclusively target money-moving transactions
Risk Concentration
2/ 5 transaction types account for 100% of fraud
Behavioral Consistency
Zero fraud in merchant-mediated transaction
(no data provided)



Transaction Amount



Range

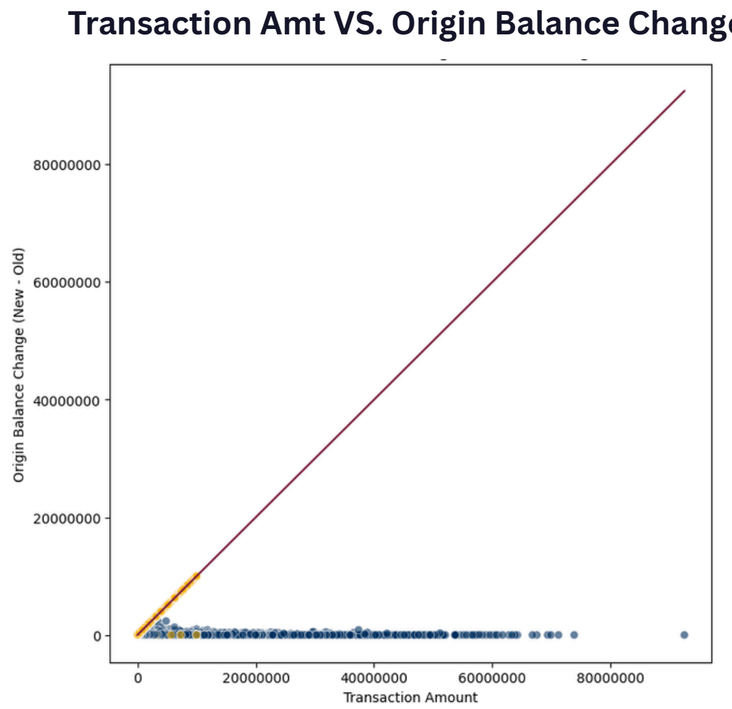
Wide variation from small to large amounts

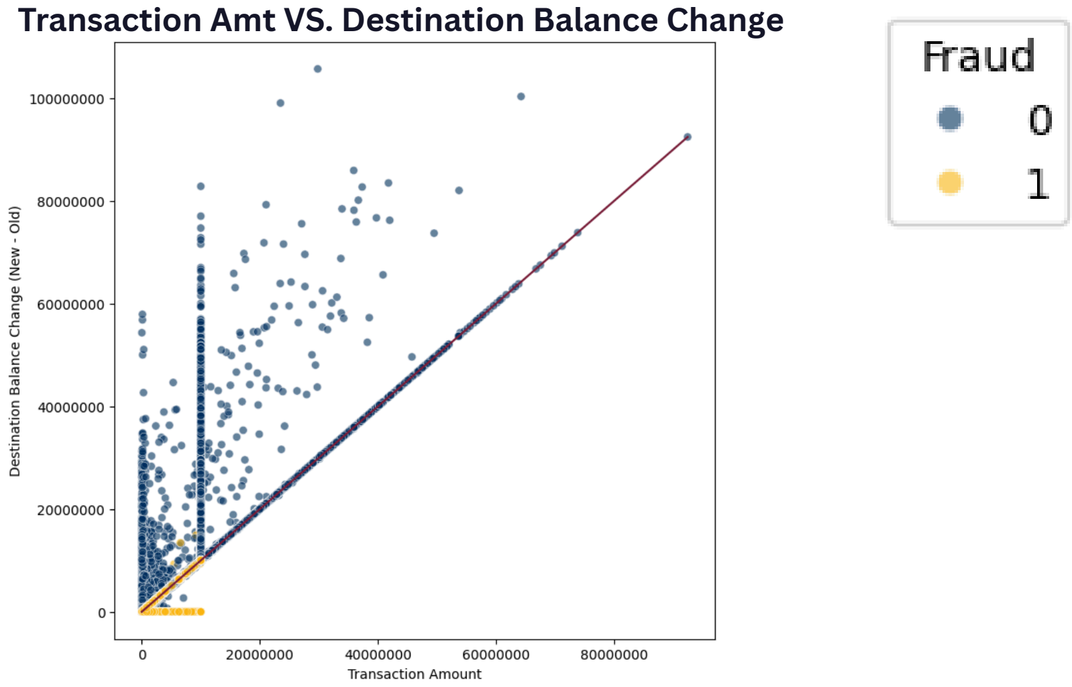
Mode

- Mostly under 2M $
- High proportion of **transactions with amount as 0**
- Significant probability of fraud transactions with nearly 10,000,000 $

All transactions with 0$ is fraud without being flagged

Transaction Amount





Legitimate

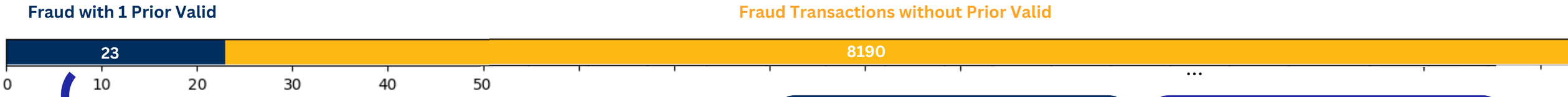
- Most people withdraw all money from their account
- Many received more than transaction amount

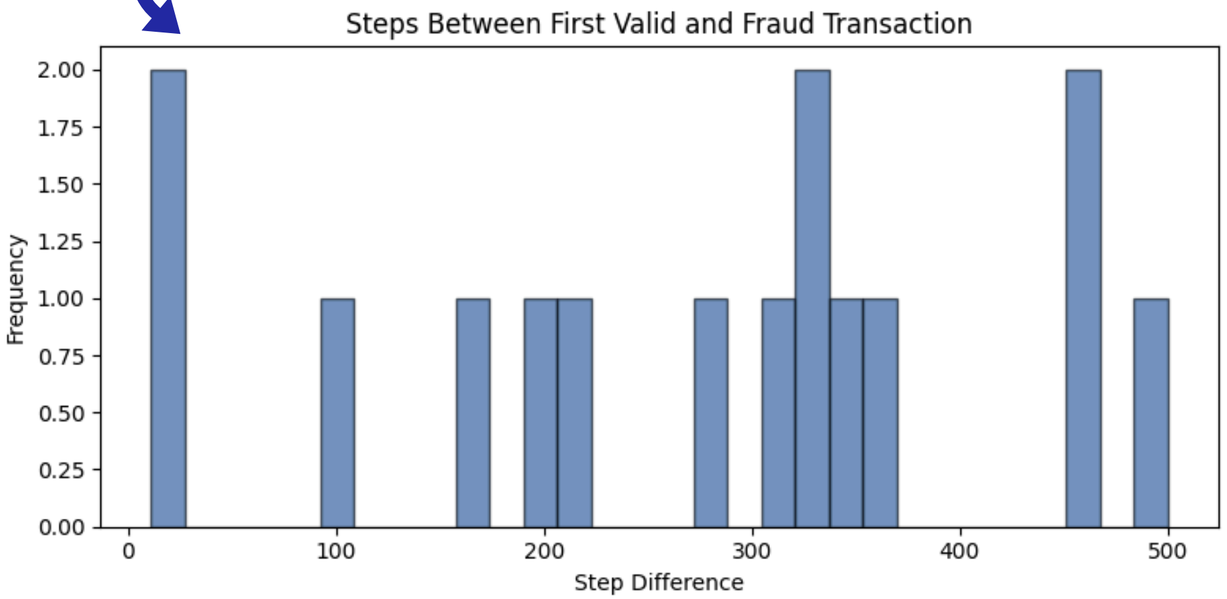
Fraudulent

- Some **transactions without balance change**
- In some cases **money was sent but never credited to the destination**

Customers

What Did They Do ?





23 accounts

Had legitimate transaction history

21 transactions

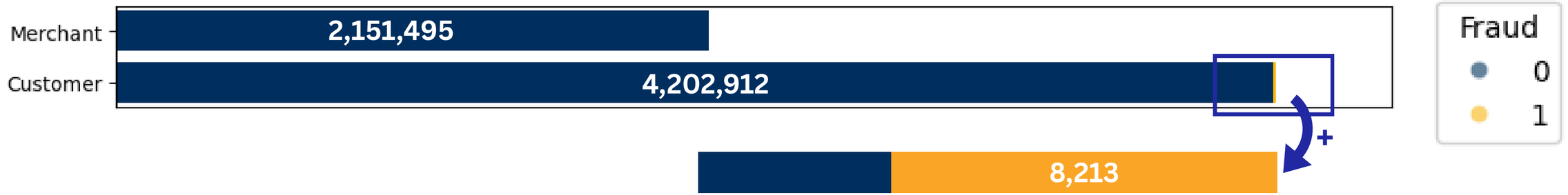
Had unmatched balance change with destination balance change

Fraud occurrence across time steps

- Early Steps (1-200): 25% of fraud cases
- Mid-Range (201-500): 50% of fraud cases
- Late Steps (501+): 25% of fraud cases

Customers

Who Are They ?



Pattern Characteristics

1. **Customer Vulnerability:** Higher fraud rates in personal accounts
2. **Merchant Protection:** Business accounts show resilience

Underlying Factors

1. **Security Awareness**: Variable across customer segments
2. T**ransaction Verification**: Varying authentication requirements
3. **Operational Patterns**: Different usage behaviors

What Lead to Faulty Fraud Detection ?

Analysis of 8197 fraud transactions reveals:

Flagged

- All are TRANSFER (16 cases)
- Amount varies from 0 to 10M $

Not Flagged

- 4K TRANSFER + 4K CASH OUT transactions
- Amount: peak frequency at 0 and 10M $ (min and max amount)

Data Limitations

- **Synthetic Data:** May not capture real-world complexity
- **Sample Size:** Limited fraud cases (16)
- **Temporal Scope:** Fixed time window analysis

Analytical Constrains

- **Pattern Stability**: Fraud patterns may evolve over time
- **External Factors**: Economic and regulatory influences not captured

Business Intelligence Insights

🚨 Risk Concentration Areas

- **High-Risk Channels**

CASH_OUT and TRANSFER transactions

- **Vulnerable Populations**

Individual customer accounts

- **Detection Gaps**

Current system over-flagging legitimate transactions

🔧 Pattern-Based Opportunities

- **Channel Monitoring**

Enhanced scrutiny for specific transaction types

- **Balance Verification**

Real-time balance change validation

- **Customer Protection**

Targeted security measures for individual accounts

📚 Strategic Applications

- **Risk Scoring**: Multi-factor pattern-based assessment
- **Customer Education**: Targeted awareness for high-risk scenarios
- **System Optimization**: Reduce false positives while maintaining detection accuracy

Next Steps & Considerations

Further Analysis Opportunities

- **Advanced Clustering:** Unsupervised learning for pattern discovery
- **Network Analysis:** Transaction flow pattern exploration
- **Behavioral Modeling:** Customer transaction pattern profiling

Implementation Priorities

- **Balance Verification:** Immediate implementation opportunity
- **Channel-Specific Monitoring:** Risk-based transaction scrutiny
- **Customer Segmentation:** Targeted protection strategies

Validation Steps

- **Pattern Testing**

Validation with additional datasets

- **Real-world Application**

Pilot implementation programs

- **Continuous Monitoring**

Pattern evolution tracking

Conclusions

🎯 What We've Accomplished

1. Identified **distinct fraud patterns** across multiple dimensions
2. Revealed **system performance characteristics** and optimization opportunities
3. Provided **evidence-based insights** for fraud detection enhancement
4. Demonstrated **analytical proficiency** in financial data exploration

💡 The Key Takeaway

Every dataset tells a story - through careful exploration, we can uncover the hidden patterns that inform better decision-making and risk management

🚀 The Value

These patterns provide the foundation for building more effective, efficient, and customer-friendly fraud detection systems in our digital financial ecosystem.

🤔 Ready to Discuss:

- Pattern validation methodologies
- Implementation strategies
- Extended analysis opportunities
- Real-world application scenarios

💼 Professional Focus

Data Analysis | Pattern Recognition | Financial Analytics | Business Intelligence

Thank you for your time

📧 Contact Information

Email: nguyentunhu0886@gmail.com
LinkedIn: www.linkedin.com/in/nguyễn-tú-như-959b35353
GitHub: https://github.com/nguyentunhu
Portfolio: https://nguyentunhu.github.io/portfolio/

Appendix

📚 Research References

- https://www.researchgate.net/publication/388415209_Identifying_Patterns_in_Financial_Transactions_to_Combat_Fraud_in_Real_Time
- https://www.researchgate.net/publication/379654286_Fraud_Detection_in_Financial_Transactions
- https://arxiv.org/abs/2308.14215
- https://www.researchgate.net/publication/390426026_Financial_Statement_Manipulation_in_the_Digital_Age_The_Role_of_AI_and_Blockchain_in_Prevention

🔗 Technical Resources

- GitHub Repository: https://github.com/nguyentunhu/Synthetic-Financial-Dataset-EDA
- Dataset URL: https://www.kaggle.com/datasets/ealaxi/paysim1