Detecting Scam-Induced Transactions to Mule Accounts

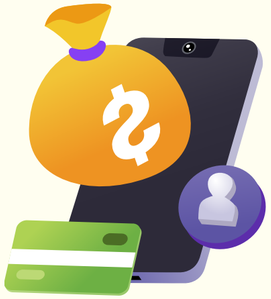
Panitta Jingjit 66102010173
Chanaporn Saikaew 66102010577

Background & Pain Points

- **The Problem:** ปัจจุบันมิจฉาชีพเปลี่ยนรูปแบบจากการเจาะระบบ (Traditional Fraud) มาเป็นการหลอกลวงให้เหยื่อโอนเงินด้วยตนเอง เข้าสู่ "บัญชีม้า"
- **Legacy System Limitations:** ระบบป้องกันภัยแบบเดิมของธนาคารมักตรวจจับจาก "อุปกรณ์ที่ผิดปกติ" หรือ "การเข้าสู่ระบบที่น่าสงสัย" แต่ในกรณี APP Fraud ธุรกรรมเหล่านี้ถูกทำขึ้นโดยตัวลูกค้าเองบนอุปกรณ์เดิมและผ่านการยืนยันตัวตน (เช่น สแกนหน้า/OTP) ทำให้ระบบเดิมไม่สามารถตรวจจับได้
- **Business Impact:** การที่ธนาคารไม่สามารถตรวจจับและระงับเหตุได้ทันท่วงที นำไปสู่ความสูญเสียทางการเงินมหาศาลของลูกค้า (Financial Loss) ทำลายความน่าเชื่อถือของธนาคาร (Reputational Damage) และอาจนำไปสู่การถูกตรวจสอบหรือปรับจากหน่วยงานกำกับดูแล (Compliance Risk)

SMART Objectives

เพิ่มอัตราการตรวจจับทุจริต (Recall) ให้ ≥85% และลดอัตราการแจ้งเตือนผิดพลาด (False Positive Rate) ลงโดยอาศัยการวิเคราะห์พฤติกรรมการโอนเงินออกที่ผิดปกติ (Abnormal outflow behavior)



Questions

- Q1 (Who & What – Behavior Outcome): บัญชีที่มีความเสี่ยงสูง มีสัดส่วนการโอนเงินออก (Outflow Ratio) มากกว่า 80% แตกต่างจากบัญชีปกติอย่างหรือไม่ และสามารถใช้เป็นตัวบ่งชี้พฤติกรรมบัญชีม้าได้หรือไม่?
- Q2 (When & How – Behavior Process): ธุรกรรมที่เกิดขึ้นในช่วงเวลาสั้นหลังจากได้รับเงิน มีรูปแบบความเร็ว (Transaction Velocity) และความถี่แตกต่างจากบัญชีปกติอย่างไร และสะท้อนพฤติกรรมของบัญชีม้าได้หรือไม่?
- Q3 (Where & Why – Context / Risk Factor): ธุรกรรมที่เกิดจากพื้นที่หรืออุปกรณ์ที่มีความเสี่ยงสูง เพิ่มความน่าจะเป็นของการทุจริต (Fraud) ได้หรือไม่?

Data Generation

Overview

This dataset was generated using a structured prompt to simulate realistic Thai banking behavior for early detection of scam-induced transactions.

- Cleaned synthetic dataset (no missing or inconsistent data)
- Designed for behavioral analysis and fraud detection
- Supports BI tools such as Tableau

Dataset Structure

4 Linked Tables:

- Customers
- Accounts
- Devices
- Transactions (Fact Table)

Volume:

- Over 20,000 transaction records
- Proper relational integrity across all tables

Fraud Simulation (Money Mule Scenario)

Rare events (1–3%) modeled using rule-based behavior:

- Rapid fund aggregation from multiple sources
- Immediate draining (>80% of balance)
- Low balance retention (near zero after transfers)
- High transaction frequency in short bursts
- Irregular transaction timing

Data Generation

Key Assumptions

- Mule accounts act as pass-through entities
- Fraudsters may control multiple accounts
- Suspicious activity may involve risky devices or patterns
- Normal users follow stable daily transaction behavior

Analytical Readiness

The dataset enables detection through:

- Outflow ratio (short time windows)
- Transaction frequency analysis
- Time-based anomaly detection
- Behavioral comparison (Normal vs High Risk)

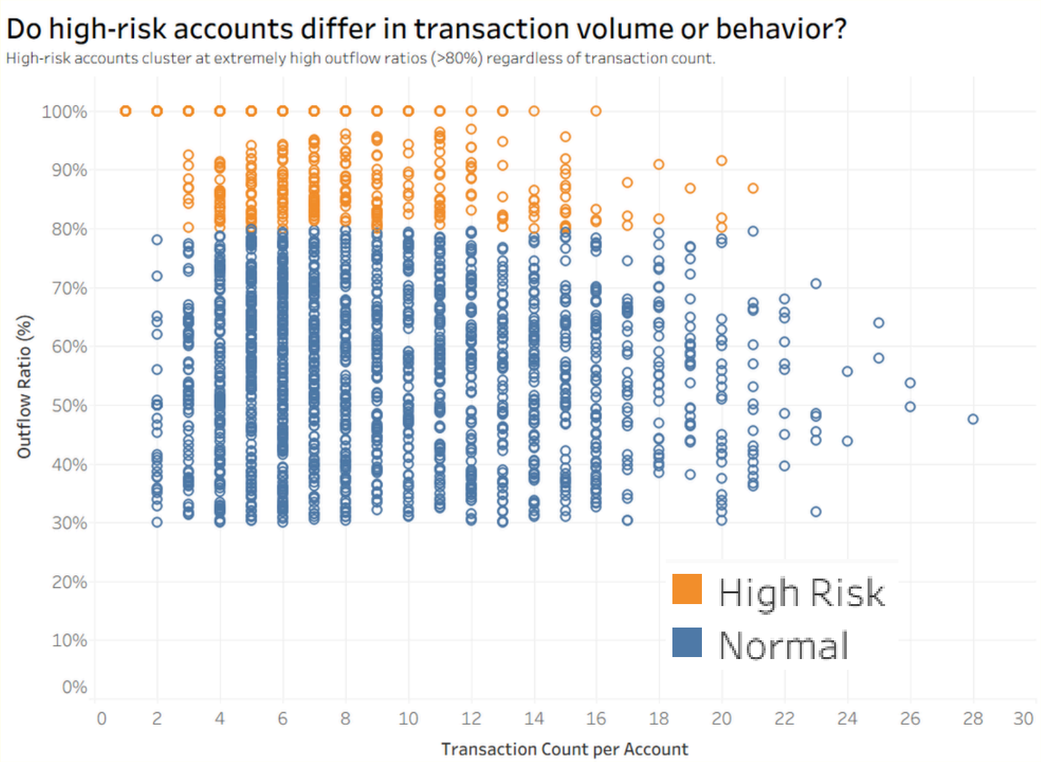
Insight Goal

Detect suspicious transaction patterns early to enable timely intervention and reduce financial losses.

Data Quality & Cleaning

- **Data Integrity & Quality Check:** ตรวจสอบความครบถ้วนของข้อมูล (Missing Values) และกำจัดข้อมูลซ้ำซ้อนโดยใช้ transaction_id เป็น Unique Identifier เพื่อป้องกันความคลาดเคลื่อนในการวิเคราะห์เชิงสถิติ
- **Structural & Logical Correction:** ระบุและแก้ไขข้อมูลที่ผิดตรรกะทางธุรกิจ เช่น ยอดเงินติดลบ หรือรูปแบบวันเวลา (Timestamp) ที่ไม่อยู่ในขอบเขตการวิเคราะห์ เพื่อคงความถูกต้องของโครงสร้างข้อมูล (Data Integrity)
- **Standardization & Consistency:** ปรับมาตรฐานรูปแบบตัวแปร (เช่น ชื่อธนาคาร, ประเภทธุรกรรม) ให้เป็นรูปแบบเดียวกันทั้งชุดข้อมูล และกำหนด Data Type ให้ถูกต้อง (Datetime, Numeric) เพื่อความพร้อมในการประมวลผล
- **Strategic Outlier & Signal Preservation:** "ไม่ลบค่า Outliers ทิ้ง เนื่องจากค่าที่สูงผิดปกติ (เช่น การโอนเงินถี่หรือยอดสูง) คือ สัญญาณหลัก (Fraud Signals) ที่ใช้บ่งชี้พฤติกรรมบัญชีม้า โดยมุ่งเน้นการขจัด Noise แต่ยังคงรักษาพฤติกรรมสำคัญของธุรกรรมทุจริตไว้ครบถ้วน

Exploratory Data Analysis



Key Insight

- บัญชีที่มีความเสี่ยงสูงกระจุกตัวที่ Outflow Ratio 80–100%
- ปริมาณธุรกรรม (Transaction Volume) ไม่สามารถแยก fraud ได้ชัดเจน
- รูปแบบพฤติกรรมการโอนเงิน เป็นตัวบ่งชี้สำคัญมากกว่าปริมาณ

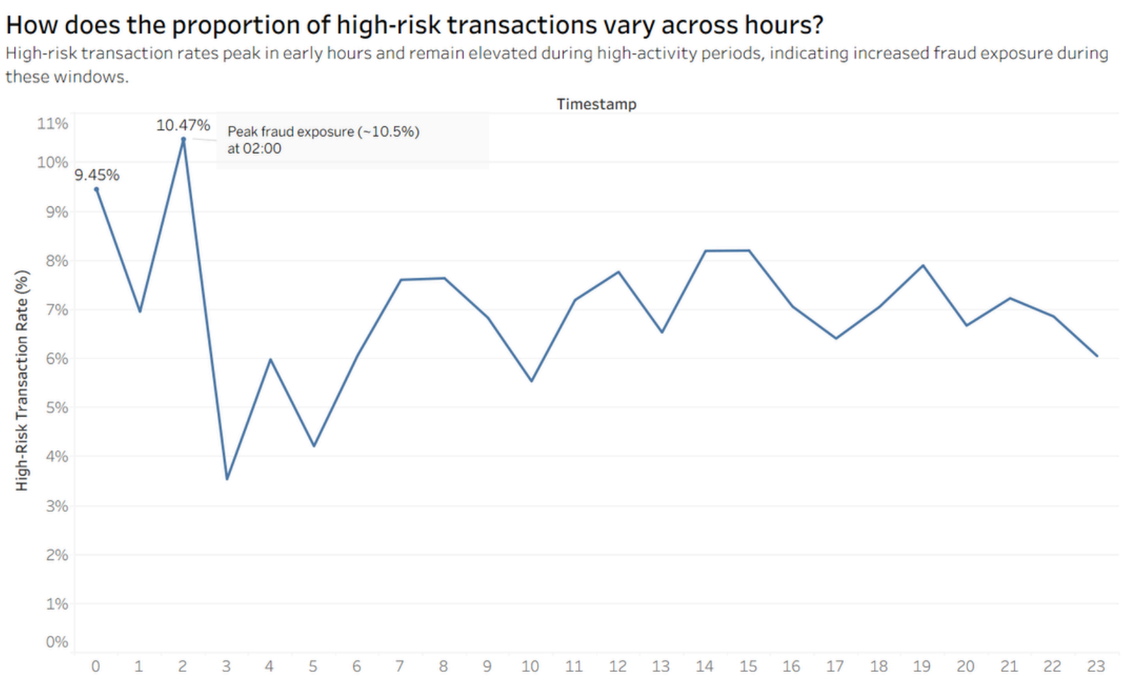
Interpretation

- บัญชีที่มีความเสี่ยงสูงมักมีพฤติกรรม รับเงินแล้วโอนออกทันที (Rapid Drain)
- พฤติกรรมผิดปกติเกิดจาก รูปแบบการไหลของเงิน (Money Flow Pattern) ไม่ใช่จำนวนธุรกรรม
- สะท้อนลักษณะของ บัญชีม้า (Mule Accounts) ที่เคลื่อนย้ายเงินอย่างรวดเร็ว

Business Value

- ช่วยตรวจจับ fraud ได้ แม้ไม่ได้มีปริมาณธุรกรรมสูง
- สนับสนุนการพัฒนา behavior-based detection

Exploratory Data Analysis



Key Insight

- ธุรกรรมในช่วงเวลา 02:00 น. มีสัดส่วน High-Risk สูงสุด (~10.5%)

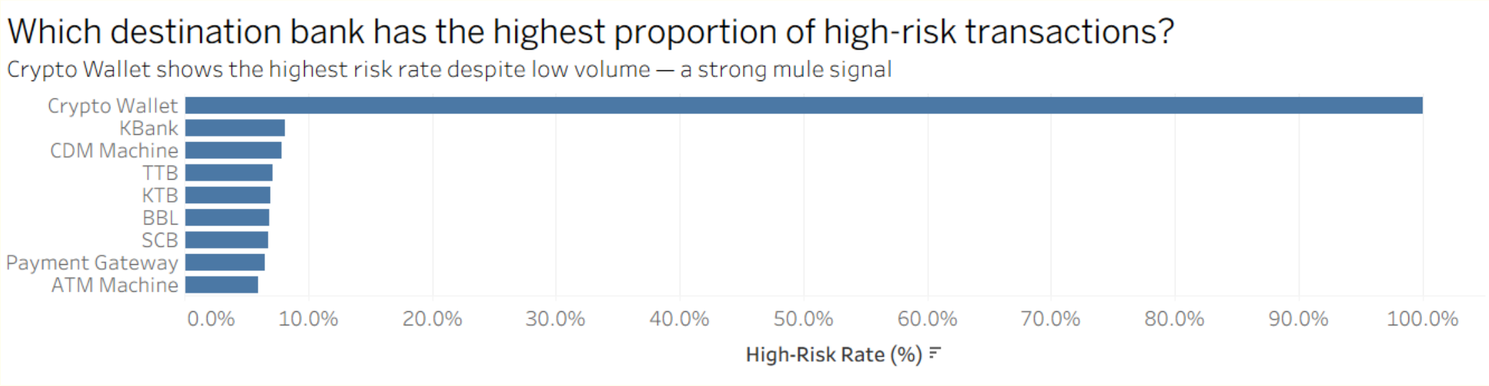
Interpretation

- พฤติกรรมทุจริตเกิดทั้งในช่วงเวลาที่คนใช้งานน้อย
- แสดงให้เห็นว่า ธุรกรรมที่มีความเสี่ยงสามารถแฝงอยู่ในพฤติกรรมปกติได้
- ชี้ให้เห็นถึง ความสัมพันธ์ระหว่างช่วงเวลาและโอกาสเกิดการทุจริต

Business Value

- สามารถใช้กำหนด ช่วงเวลาที่ต้องเฝ้าระวัง (High-Risk Hours)
- เพิ่มประสิทธิภาพในการ ตรวจจับ fraud แบบเชิงเวลา (time-based monitoring)
- ช่วยจัดสรรทรัพยากรในการป้องกัน fraud ได้เหมาะสมยิ่งขึ้น

EDA & Visualizations



Key Insight

- Crypto Wallet มีสัดส่วนธุรกรรมความเสี่ยงสูงสุด (~100%)
- ช่องทางอื่นมีความเสี่ยงต่ำกว่าชัดเจน

Interpretation

- บ่งชี้พฤติกรรมบัญชีม้า / การฟอกเงิน
- ความเสี่ยงกระจุกตัวในบางปลายทาง

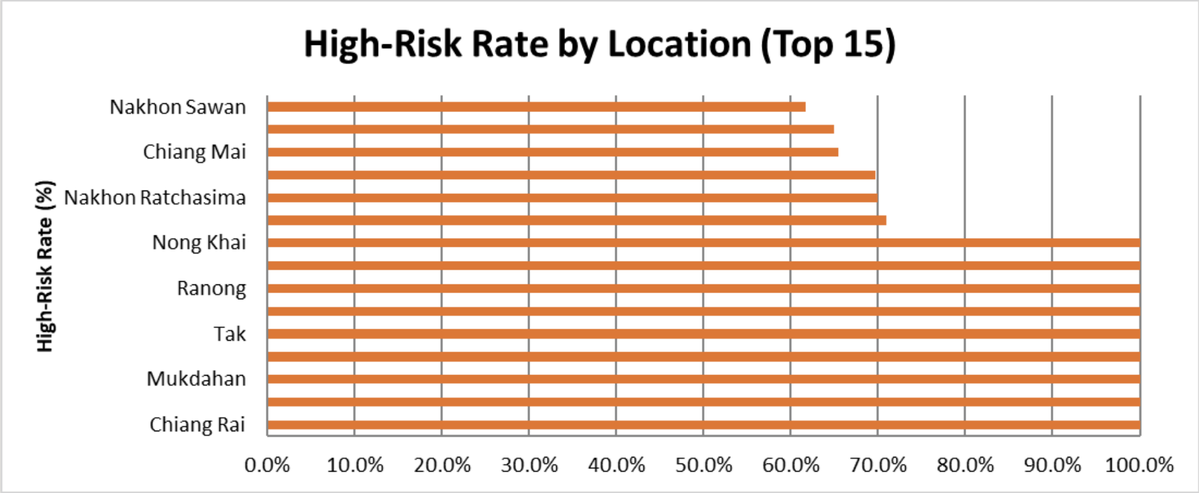
Imbalance Handling

- ใช้ % แทนจำนวน
- ลด bias จากข้อมูลที่ไม่สมดุล

Business Value

- โฟกัสตรวจสอบช่องทางเสี่ยงสูง
- เพิ่มประสิทธิภาพการตรวจจับ fraud

Exploratory Data Analysis



Key Insight

- หลายพื้นที่มี High-Risk Rate สูงถึง 100% เช่น Chiang Rai, Mukdahan, Tak
- จังหวัดหลักอย่าง Nakhon Ratchasima และ Chiang Mai ก็มีความเสี่ยงสูง (>60%)
- สะท้อนว่า ความเสี่ยงไม่ได้กระจาย แต่กระจุกตัวในบางพื้นที่

Interpretation

- ธุรกรรมในบางพื้นที่มี แนวโน้มเป็น fraud สูงกว่าปกติอย่างชัดเจน
- บ่งชี้ถึง ความสัมพันธ์ระหว่างพื้นที่กับโอกาสเกิดการทุจริต
- อาจสะท้อนถึง เครือข่ายบัญชีม้าในบางพื้นที่

Business Value

- ใช้กำหนด พื้นที่เฝ้าระวัง (High-Risk Locations)
- เพิ่มประสิทธิภาพในการ ตรวจจับ fraud เชิงพื้นที่ (location-based monitoring)

Confusion Matrix

สัญญาณ	เงื่อนไข	คะแนน	ทำไมถึงสงสัย
Outflow สูง	Ratio > 0.6	+40	โอนเงินออกมากผิดปกติ
Outflow สูงมาก	Ratio > 0.8	+30	เพิ่มน้ำหนักถ้ายิ่งสูง
ยอดเงินต่ำ	Balance < 10,000	+20	เงินไม่เหลือ = ล้างบัญชี
ทำรายการกลางคืน	Night flag > 0.1	+10	พฤติกรรมหลบเลี่ยง

Count of account_id	Predict Labels		
Actual Label	0	1	Grand Total
0	100.00%	0.00%	100.00%
1	0.00%	100.00%	100.00%
Grand Total	**24.60%**	**75.40%**	**100.00%**

Findings and Insights

1. Money Flow Anomaly (พฤติกรรมการไหลของเงิน)
   - Account Draining Effect: บัญชี High-risk มี Outflow Ratio สูง (80–100%) สะท้อนพฤติกรรม “รับแล้วโอนออกหมด”
   - Low Volume, High Impact: ไม่จำเป็นต้องทำธุรกรรมบ่อย แต่มีสัดส่วนการโอนออกสูง → แยก fraud ได้ดีกว่าการดู Volume
2. Temporal & Channel Patterns (เวลาและช่องทาง)
   - High-Risk Time: ความเสี่ยงสูงสุดช่วง ตี 2 และยังคงสูงในช่วงนอกเวลาปกติ
   - High-Risk Channels: กระจุกในช่องทางที่โอนเร็วและติดตามยาก เช่น Crypto Wallet, CDM, Payment Gateway
3. Geographic Concentration (พื้นที่)
   - Fraud Clustering: ธุรกรรมเสี่ยงกระจุกในบางพื้นที่ โดยเฉพาะ จังหวัดชายแดน/พื้นที่ที่มีสแกมสูง
   - บางพื้นที่มี High-Risk Rate สูงมาก (ใกล้ 100%) สะท้อนแหล่งปฏิบัติการหรือเครือข่ายมิจฉาชีพ

Recommendations

Real-time Rule-based Alerting: ตั้งค่าตรวจจับธุรกรรมผิดปกติ (เช่น Outflow >80%,ยอดเงินคงเหลือต่ำ และการทำธุรกรรมกลางคืน) และใช้ **"Step-up Authentication"** (แจ้งเตือนให้ยืนยันตัวตน) เพื่อป้องกันเงินออกแทนการล็อคบัญชีทันที

Risk-based Account Limits: กำหนดวงเงินโอนเริ่มต้นสำหรับ "บัญชีเปิดใหม่" ให้ต่ำกว่าปกติจนกว่าจะผ่านเกณฑ์ความน่าเชื่อถือ เพื่อลดความเสี่ยงจากบัญชีม้าใช้แล้วทิ้ง

High-Risk Destination Monitoring: เพิ่มการตรวจสอบเข้มงวดและแจ้งเตือนสำหรับปลายทางที่มีความเสี่ยงสูง

Expected Impacts

Business Outcome: ลดความสูญเสียทางการเงิน และเพิ่มความน่าเชื่อถือของธนาคารผ่านกระบวนการป้องกันเชิงรุก (Proactive Prevention)

Operational: ลดภาระงานตรวจสอบของเจ้าหน้าที่ธนาคารด้วยระบบกรองอัตโนมัติ (Rule-based)

Customer Experience: เพิ่มความปลอดภัยสูงสุดโดยที่ลูกค้าปกติยังใช้งานได้ราบรื่น (Minimal Disruption)