Financial Transaction Anomaly Detection

Team 8: Ching Hsuan (Shawn) Lin, Huawan (Zoey) Zhong,
Aryan Sehgal, Jean Miranda Huarancca, Rodrigo Yeo Cordero

CONTENTS

01

Introduction & Motivation

Introduction & motivation

- **Background:**

  Money laundering undermines the integrity of global financial systems by disguising illicit proceeds as legitimate transactions.

- **Challenge:**

  Manual transaction reviews are becoming inefficient as transaction volumes grow and laundering techniques become more sophisticated.

- **Objective:**

  We aim to utilize machine learning to provide a scalable and proactive solution by detecting hidden patterns and emerging laundering behaviors in real time.

02

Data Sources

Data Source

IBM Anti-Money Laundering Transaction Dataset

LI-Medium_Trans

Group LI: Lower illicit ratio
(more laundering)

HI-Medium_Trans

Group HI: Higher illicit ratio
(less laundering)

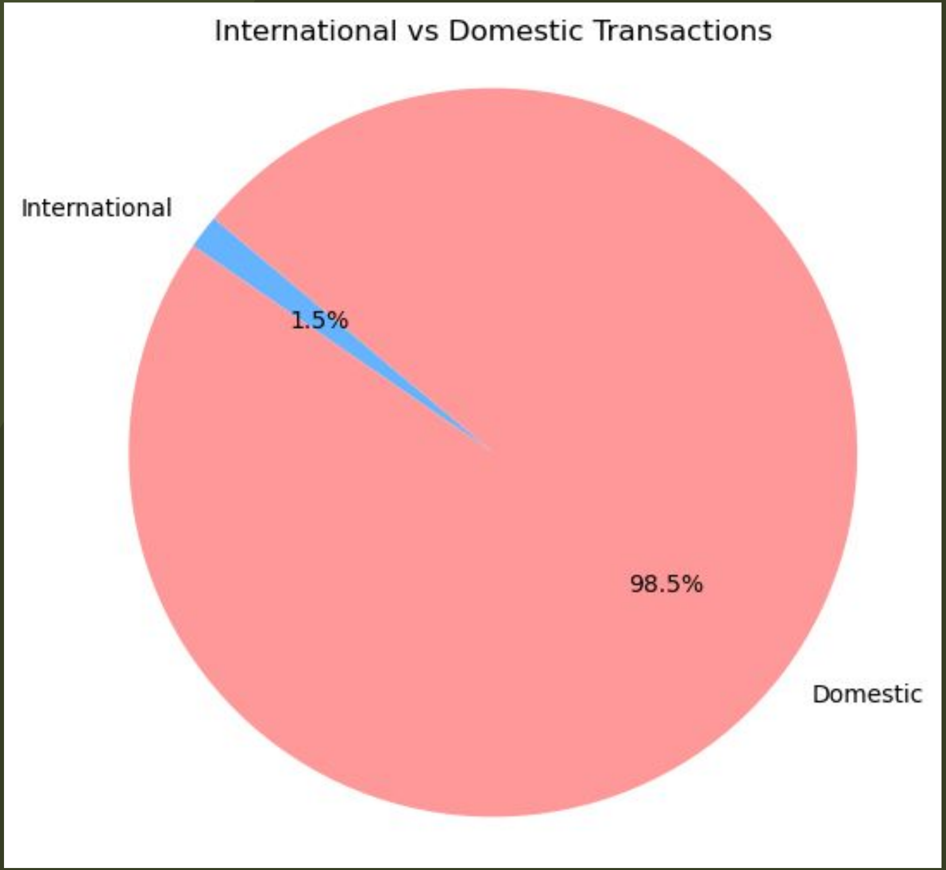
Currency

Payment currency information

03

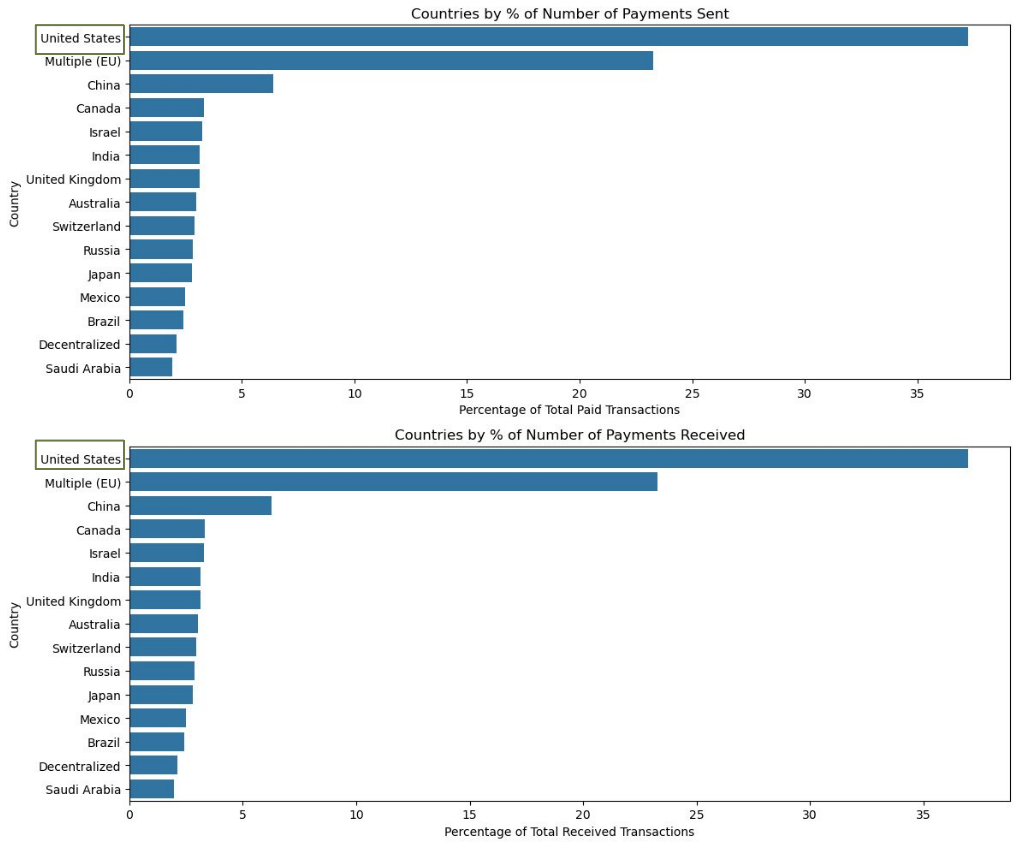
Exploratory Data Analysis

Boarder-cross vs domestic pie chart



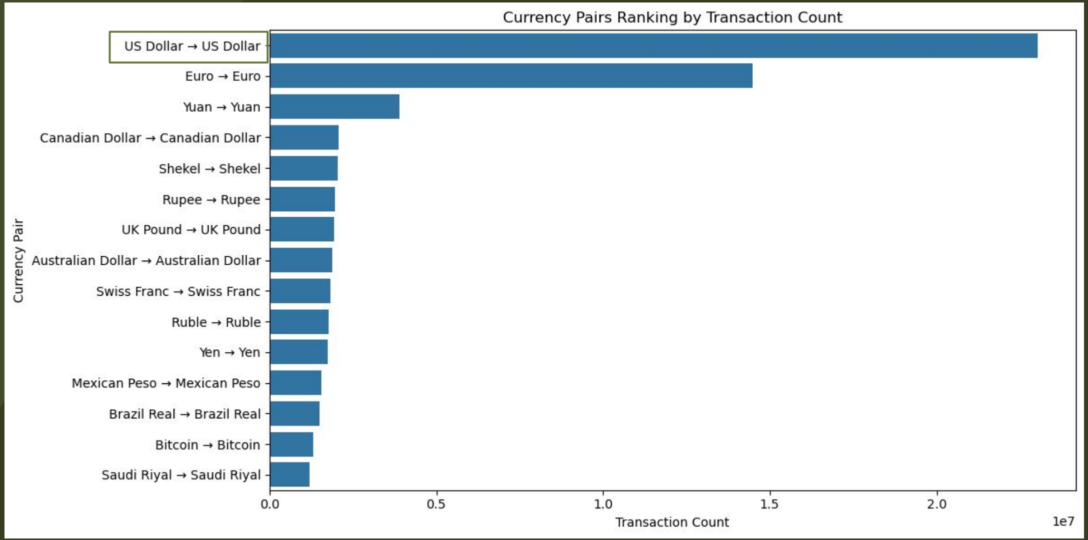
- Majority of transactions are domestic.
- Only 1.5% are international.

Country-Level Distribution of Financial Transactions: Paid vs. Received



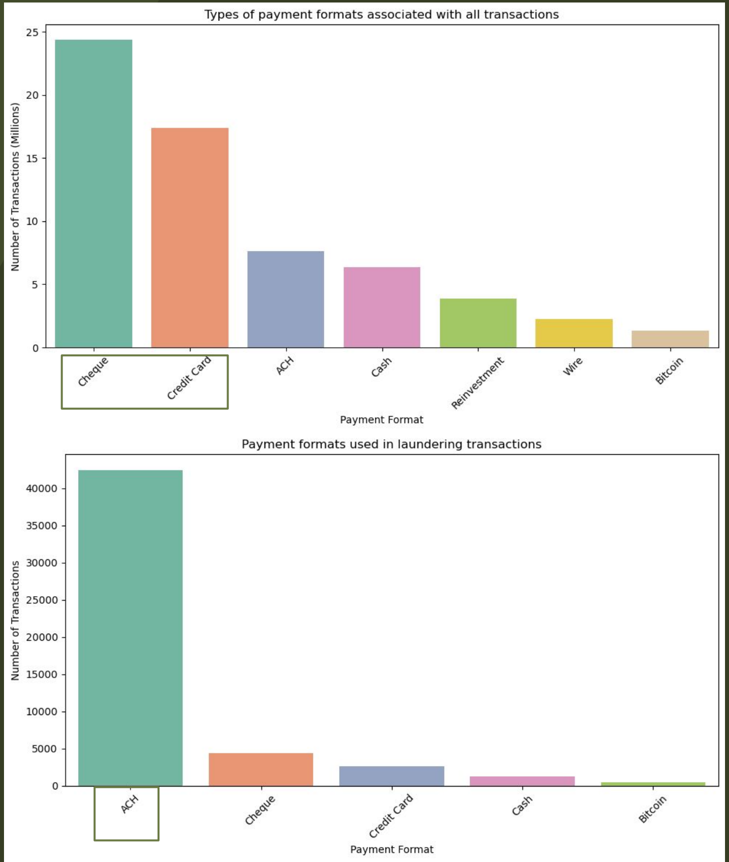
- United States is the dominant player in both sending and receiving transactions

Currency Pairs Ranking by Transaction Count



- The majority of transactions occur within the same currency.
- USD→USD and Euro→Euro dominating the top currency pairs.

Types of payment formats: All vs. Laundering Transactions



- For all transaction, Cheques and Credit Cards lead in payment transaction volume
- ACH is the predominant format used in laundering transactions

04

Modeling Results & Insights

Logistic Regression

Confusion Matrix

	Pred 0	Pred1
Actual 0	16935016	1995639
Actual 1	2651	12644

Accuracy: 0.8945
Recall: 0.8267
Precision: 0.0063

Explain model using LIME

Feature	Avg weight
Currency=AUD	0.158
Region_Pay= Eastern Europe	0.011
Currency=Pound	0.010
Country_recv=UK	0.009
hour	0.007

Random Forest

Filter all the result that is predicted as launder and labelled as launder.

hour	count
12	518
11	516
13	501
16	481
15	459

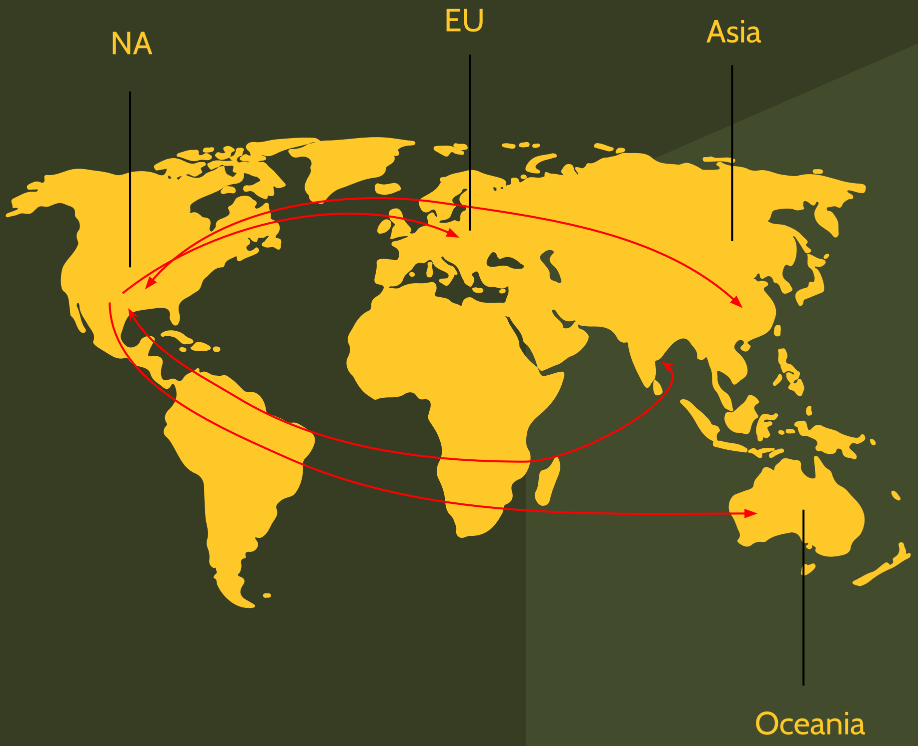
Payment Format	count
ACH	8172

Accuracy: 0.9171
Recall: 0.8016
Precision: 0.0078

- Peak laundering activities in the afternoon.
- ACH(Automated Clearing House) is the primary channel for laundering transactions.

Isolation Forest

Anomaly Score	Payment Currency	Amount Pay	Flow Direction	Receiving Currency	Amount Received
0.1230	USD	2.24 B	NA→ Asia	Yen	236.09 B
0.1146	USD	132.81 M	NA→Oceania	AUD	187.63 M
0.1113	USD	453.46 M	NA→ Asia	Rupee	33.30 B
0.1091	USD	243.10 M	NA→ EU	Euro	207.46 M
0.1073	Rupee	35.32 B	Asia→ NA	USD	480.89 M



Unsupervised segmentation of transactions using K-Means

Goal: Identify distinct transaction behavior patterns using unsupervised clustering (K-Means).

Cluster	Profile name	Avg amount	% Cross-Border Transactions	Laundering rate
0	Global movers	$930K	67%	0.03%
1	Big domestic players	$7M	2%	1.78%
2	Mid-level Operators	$500K	0.40%	0.93%
3	Everyday senders	$220K	0.18%	0.39%

05

Conclusion

What we found

Key laundering patterns were identified: ACH transfers were the main laundering channel, with suspicious activity peaking at midday and midnight; overall, transactions were mostly domestic and dominated by the United States.

Supervised models successfully prioritized laundering detection: Logistic Regression and Random Forest models were tuned to maximize recall, ensuring most laundering activities were identified, even at the cost of higher false positives.

Isolation Forest successfully uncovered hidden anomalies, highlighting high-risk, large-value cross-border transactions, particularly between North America, Asia and Oceania (e.g., Yen → USD, AUD → USD, CNY → USD).

K-Means clustering revealed four groups: **Global movers**, **big domestic players**, **mid-level operators** and **everyday senders**, providing a strong basis for risk segmentation

Strategy Recommendation

- Deploy real time ACH anomaly scoring with lower thresholds on any non-USD or cross-border.
- Increase alert sensitivity during **midday** and **midnight** transaction peaks.
- Implement a **tiered alert system** to prioritize high-value cross-border transactions and known high-risk regions.
- Run monthly targeted audits per cluster and track suspicious activity to recalibrate thresholds and enrich training labels.
- Enrich features with **dynamic behavioral metrics** (e.g., transaction velocity) to improve future detection performance.

THANK YOU