Detecting Anomalous Financial Transactions

Using big data to help FinCEN fight financial fraud and money laundering

BY RYAN POSTERNAK

Agenda

1 **BUSINESS UNDERSTANDING**

The high cost of money laundering

2 **DATA UNDERSTANDING**

A model is only as good as the data

3 **THE MODELS**

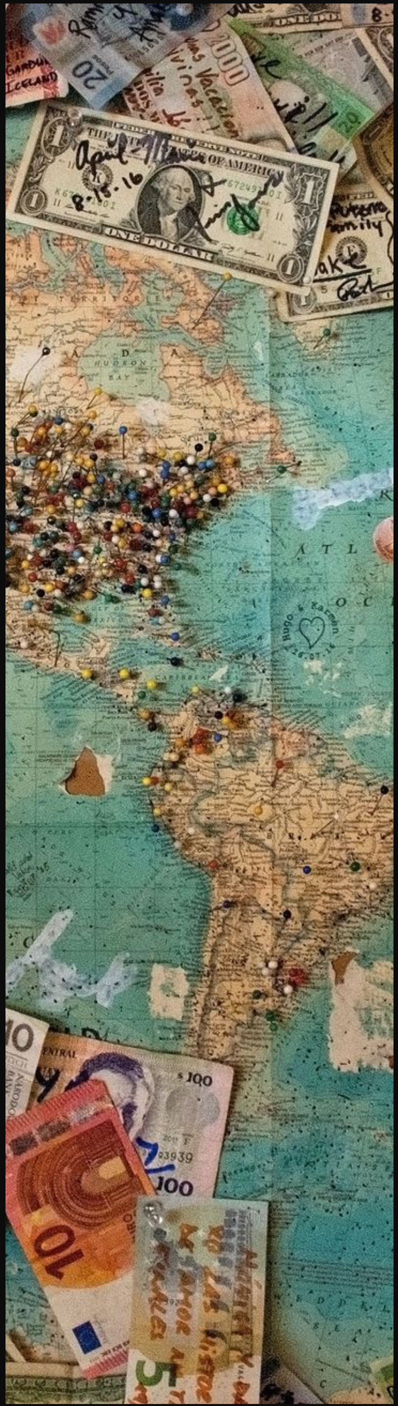
Can we reliably spot anomalous transactions?

4 **ANALYSIS & INFERENCES**

What have we learned?

5 **CONCLUSION & RECOMMENDATIONS**

What are the takeaways?



Business Understanding

"THE ESTIMATED AMOUNT OF MONEY LAUNDERED GLOBALLY IN ONE YEAR IS 2% - 5% OF GLOBAL GDP, OR $800 BILLION - $2 TRILLION USD" — UNITED NATIONS OFFICE ON DRUGS AND CRIME

THEFT AND MAJOR CRIMES

- From stolen pandemic aid to drug dealing and racketeering - criminals go to great lengths to hide the source of their cash

SANCTIONS DIVERSION

- Sanctioned organizations and nation-states will attempt to skirt restrictions through clandestine operations

TERRORIST FINANCING

- Terrorist organizations require capital to carry out their activities

HARMS INDUSTRY AND CONSUMERS

- Banks can suffer reputational damage and face regulatory scrutiny, and costs often pass on to consumers

Data Understanding

A collaboration between NIST and the NSF

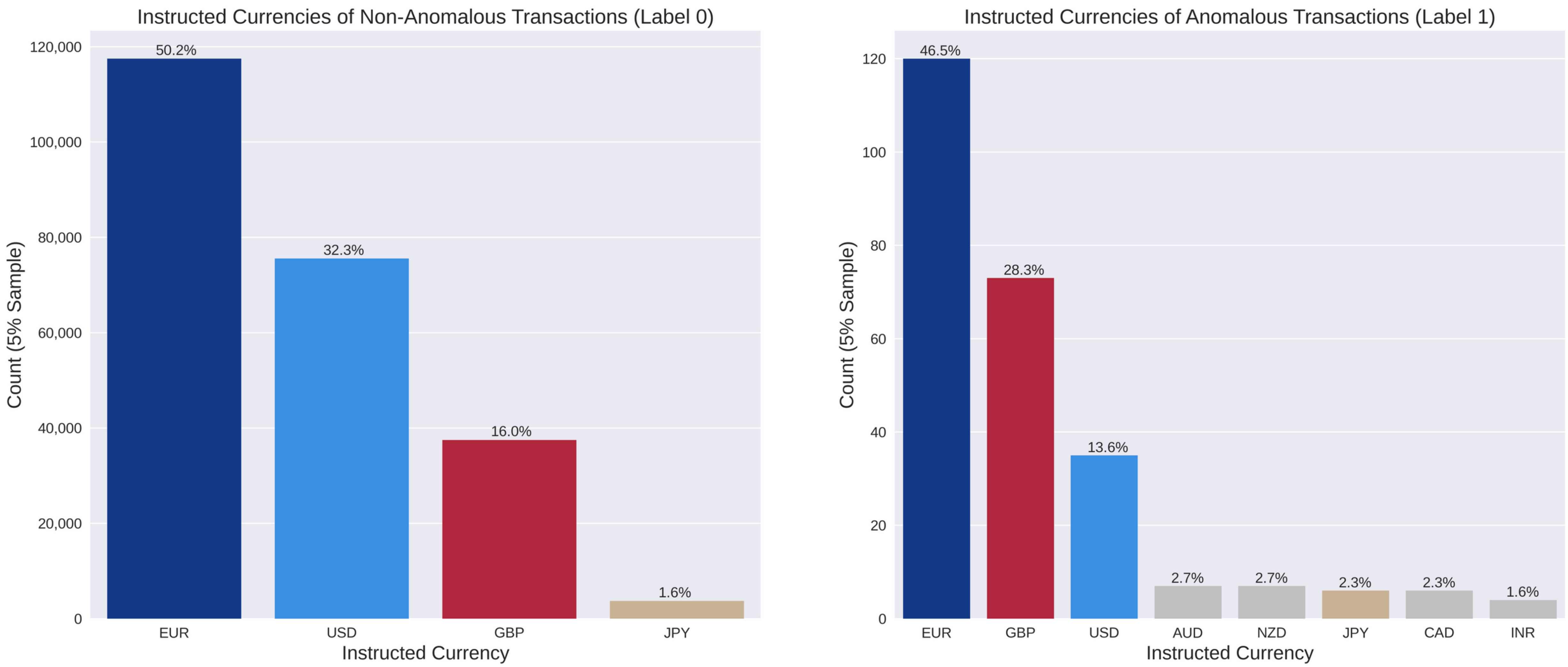
```
Transactions dataframe
----------------------
root
 |-- MessageId: string
 |-- Timestamp: timestamp
 |-- UETR: string
 |-- Sender: string
 |-- Receiver: string
 |-- TransactionReference: string
 |-- OrderingAccount: string
 |-- OrderingName: string
 |-- OrderingStreet: string
 |-- OrderingCountryCityZip: string
 |-- BeneficiaryAccount: string
 |-- BeneficiaryName: string
 |-- BeneficiaryStreet: string
 |-- BeneficiaryCountryCityZip: string
 |-- SettlementDate: integer
 |-- SettlementCurrency: string
 |-- SettlementAmount: double
 |-- InstructedCurrency: string
 |-- InstructedAmount: double
 |-- Label: integer
```

Target class distribution

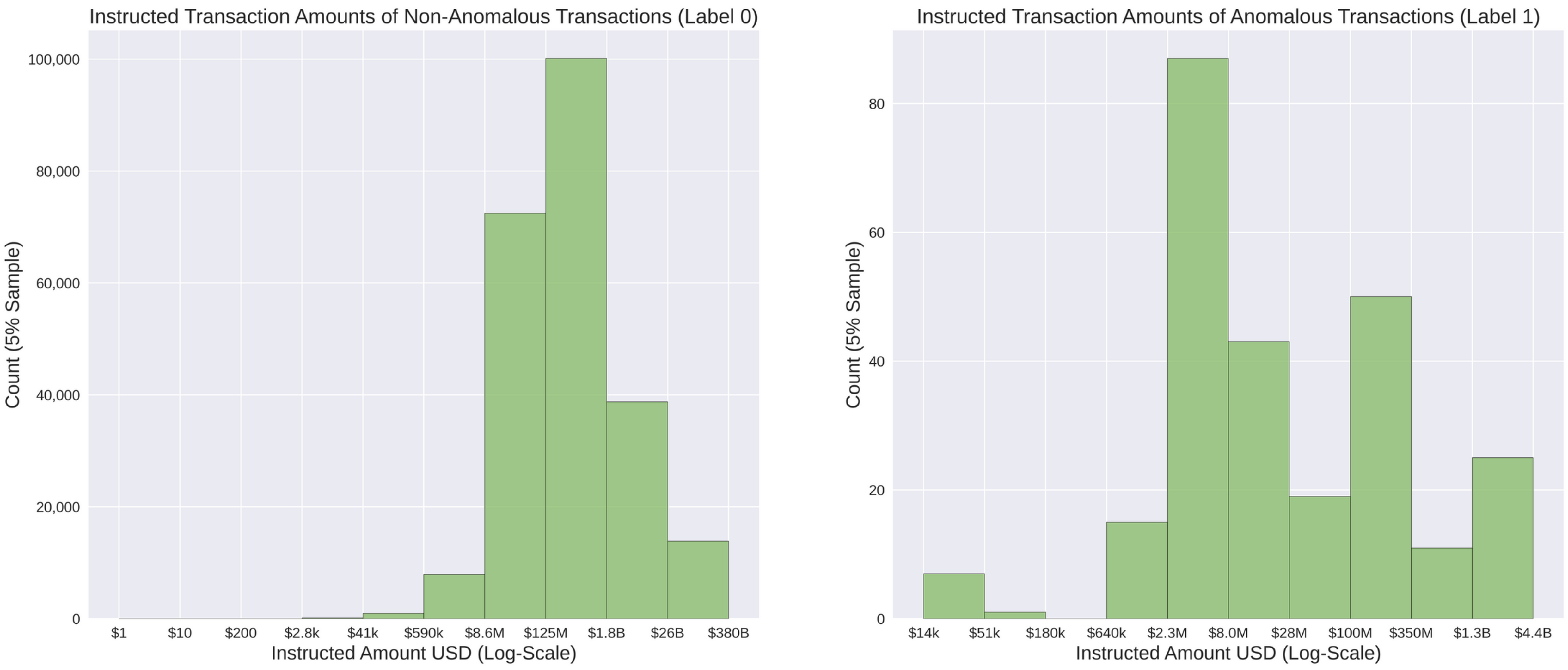
Label	count	percent
Anomalous (1)	4,900	0.104%
Non-anomalous (0)	4,686,825	99.895%

5.5M Transactions
530k Matching Accounts
91 Features

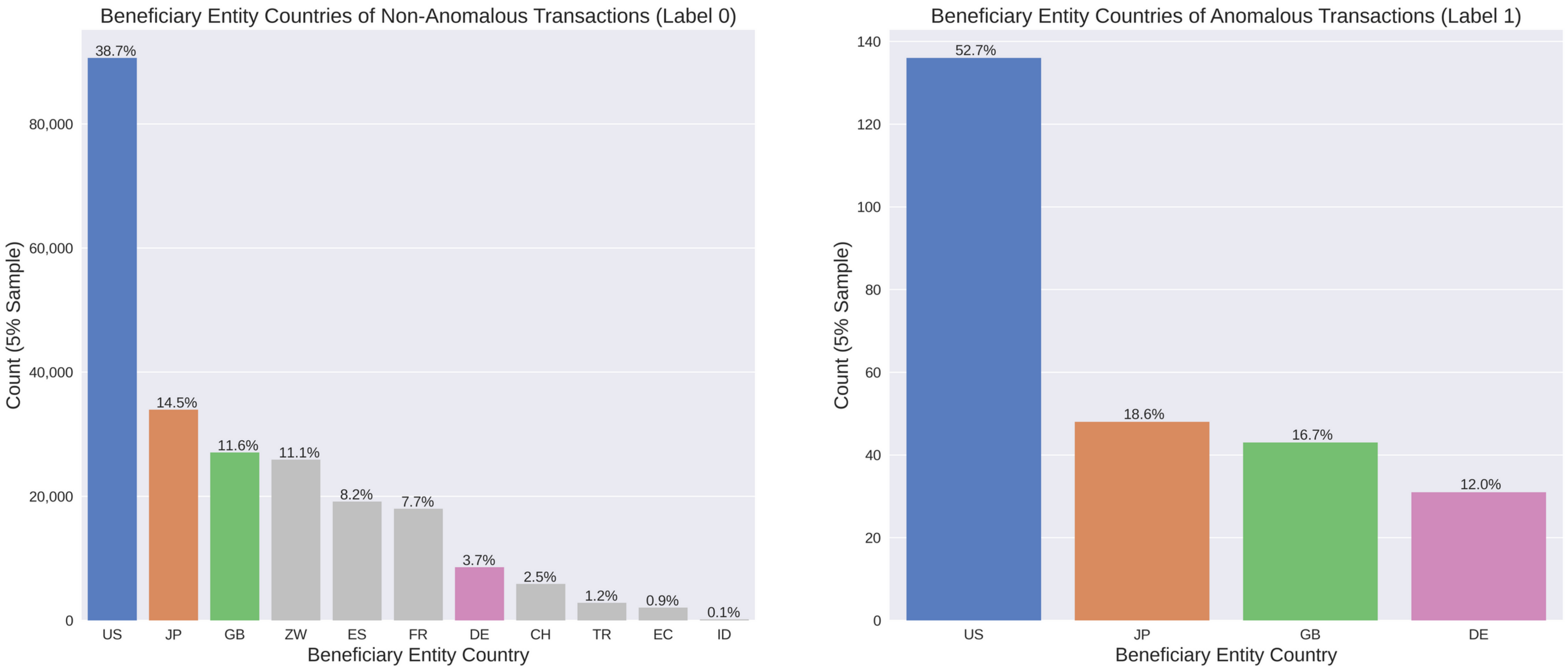
```
Bank accounts dataframe
-----------------------
root
 |-- Bank: string
 |-- Account: string
 |-- Name: string
 |-- Street: string
 |-- CountryCityZip: string
 |-- Flags: integer
```



INSTRUCTED TRANSACTION CURRENCIES



INSTRUCTED TRANSACTION AMOUNTS (LOG-SCALE USD)



BENEFICIARY ENTITY COUNTRIES

Modeling Results

As measured by test data F1 scores

0.019
BASELINE LOGISTIC REGRESSION

0.019
GBT model 1
Downweighted 10x

0.026
Naive Bayes model 1
Downweighted 100x

0.265
SVC model 1
Downweighted 10x

0.265
Logistic regression model 1
Unsampled

0.487
Random forest model 1
Oversampled 10x

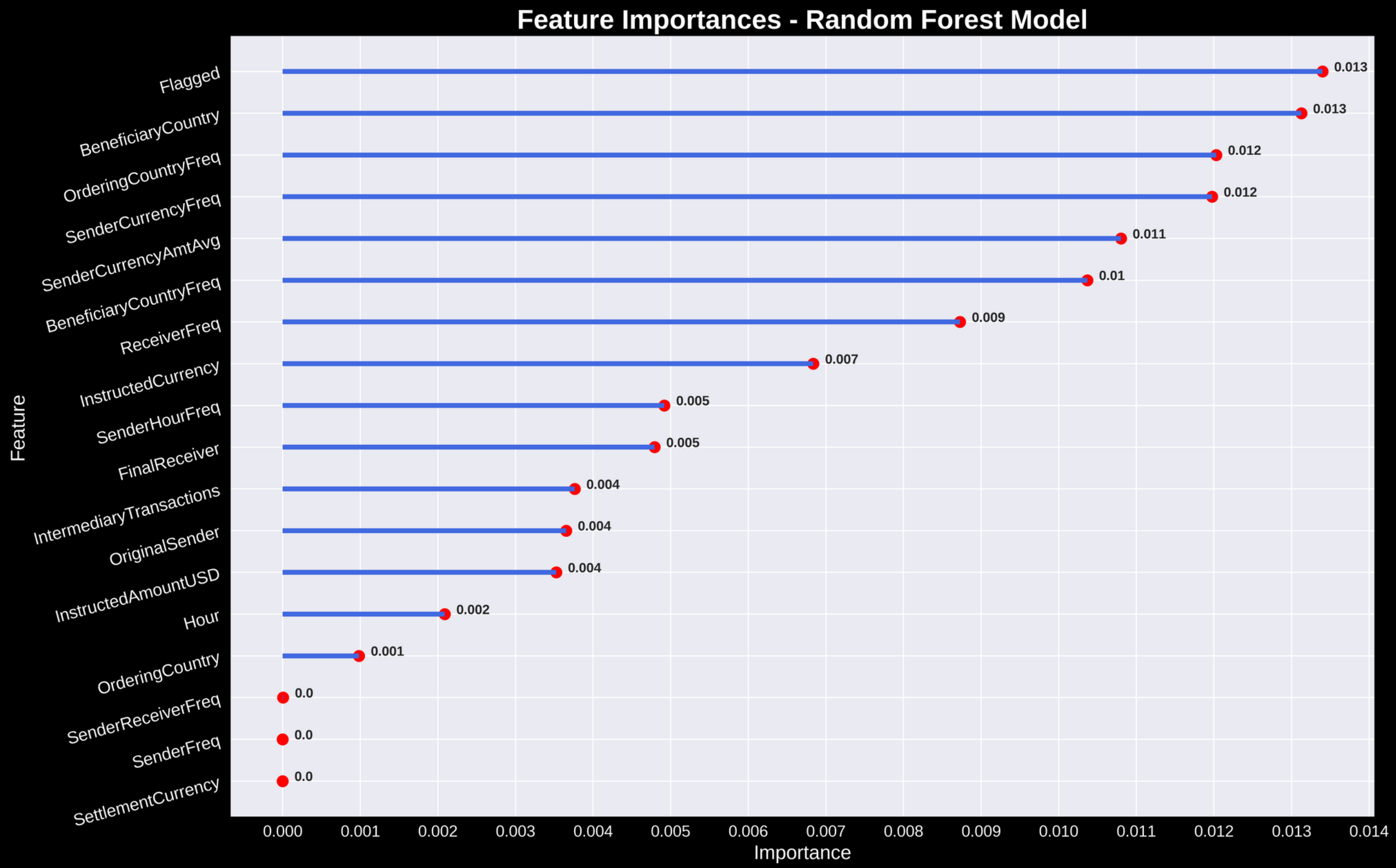
0.555
Decision tree, models 1 & 2
Model 1: Unsampled
Model 2: Oversampled 10x

0.567
Multilayer perceptron model 4
Downsampled 50%

CONFUSION MATRIX

TRUE

	PREDICTED 0	PREDICTED 1
0	704K	22
1	451	310



Analysis

Conclusion & Recommendations

IT'S A TRADE OFF

Catching more anomalous transactions = more false positives

USE MODEL TO FOCUS RESOURCES

Knowing where to look is half the battle

OBTAIN MORE POSITIVE CLASS OBSERVATIONS

Models could be improved with more data

Thank you

Email: rposternak@yahoo.com
GitHub: github.com/rjpost20
LinkedIn: linkedin.com/in/ryanposternak