

Fraud Detection

Looking for anomalies with data exploration on the IntesaSanPaolo Dataset

Pietro Cappelli, Prof. Marco Zanetti

University of Padova, Physics and Astronomy Department "G. Galilei"
National Institute for Nuclear Physics (INFN), Padova Division

Spoke2-FraudDetection meeting
02/09/2024

pietro.cappelli@unipd.it

Data Exploration



Data exploration

Strategia

Analizzato principalmente il **dataset contenente** eventi **frodi**

- ▶ cartella *punto_2*: tutte le transazioni dei prodotti per gli **utenti** che sono stati **frodati almeno una volta**
- ▶ In particolare le transazioni "*prodotto.bonifico*" perché maggiormente popolati

Analizziamo le transazioni presenti distinguendo eventi **genuini** e **anomali**

Cosa cerchiamo?

- ▶ Variabili **significative** (se esistono)
- ▶ Correlazioni tra le features che descrivono un evento

Data exploration

Dataset informations

Analizzato principalmente il **dataset contenente** eventi **frodi**

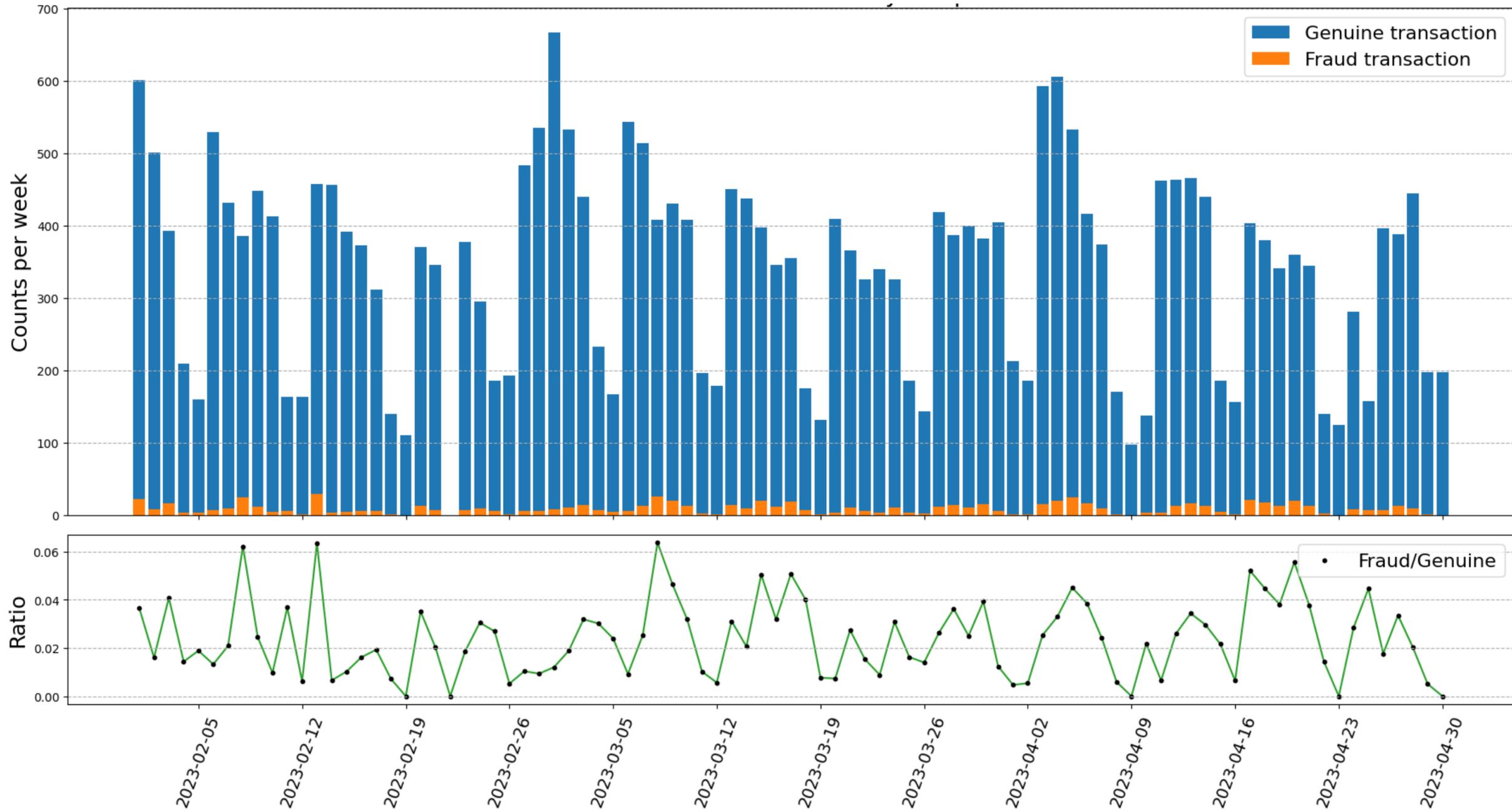
- ▶ cartella *punto_2*: tutte le transazioni dei prodotti per gli **utenti** che sono stati **frodati almeno una volta**
- ▶ In particolare le transazioni "*prodotto.bonifico*" perché maggiormente popolati

Concatenando **tutti i file** relativi al *punto_2*:

- ▶ Dataset contiene **duplicati** (stessa *prodotto_x_isp_chiaveantifrode*) -> teniamo solo **l'ultima entrata**
- ▶ **Dimensione dataset** post pulizia: $N_{\text{tot}} = 90'314$ transazioni,
di cui frodi $N_{\text{fraud}} = 3285$ (3.64%)

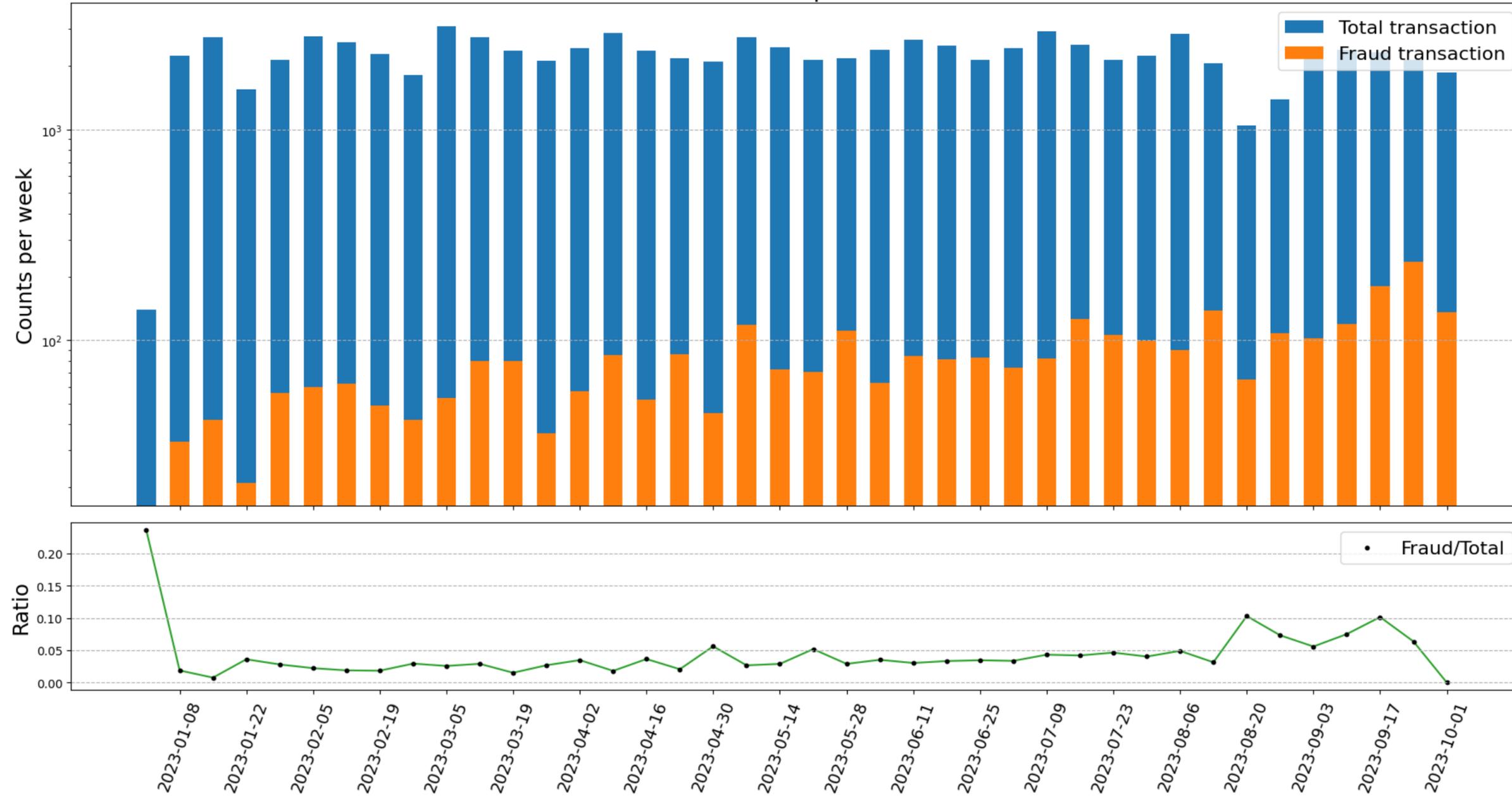
Transaction time distribution

Transazioni per giorno (Feb-Apr)



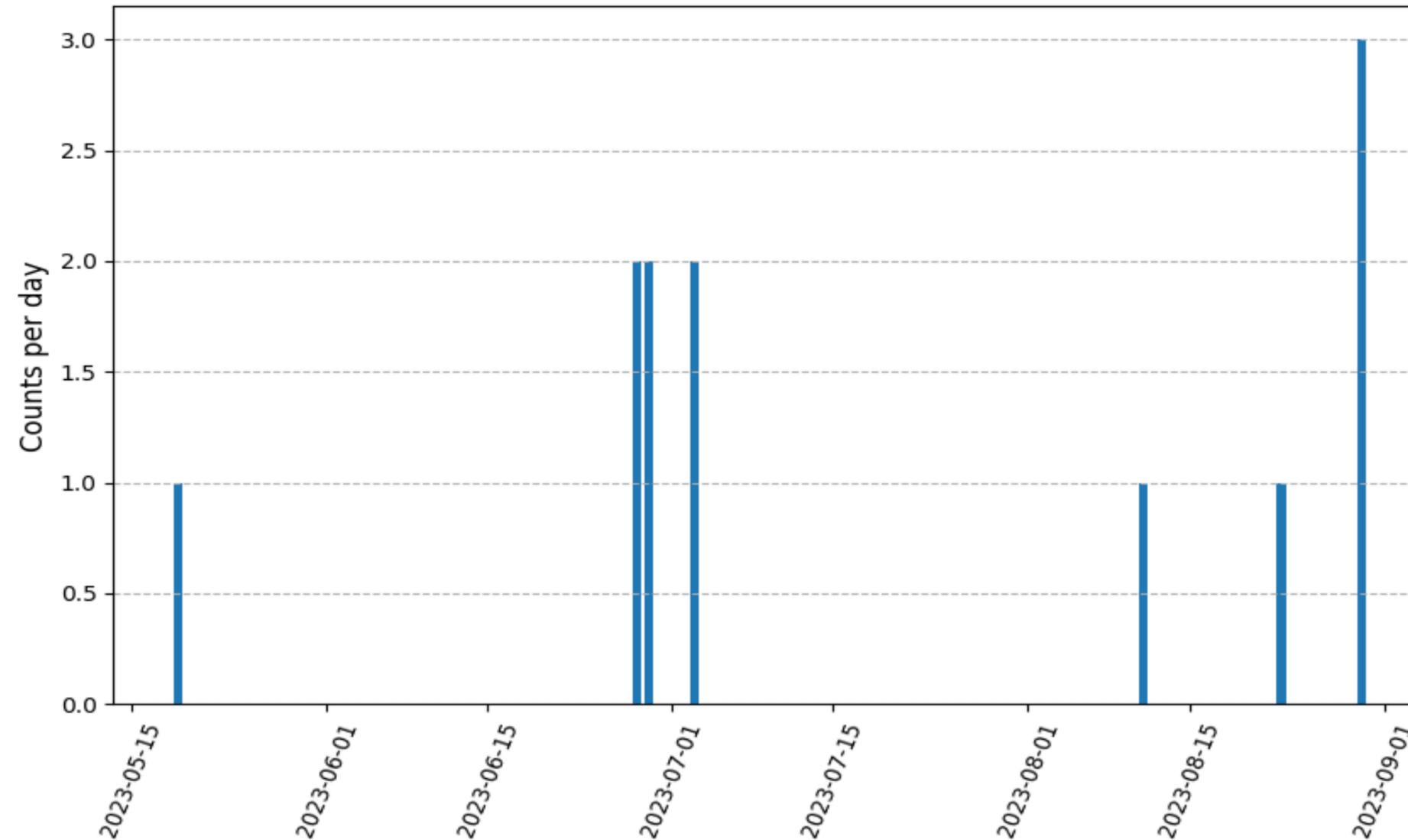
Transaction time distribution

Transazioni per settimana (Gen-Sept)



Transaction time distribution

Utente frodato 12 volte



Data exploration

Features informations

Importo medio (con noise gaussiano std 30%): 1381.1 ± 4.54

- ▶ **Frode:** 2857.6 ± 28.4
- ▶ **Genuine:** 1325.3 ± 4.6

Numero totale di **utenti** 3944 di cui frodati 1991

Numero totale di **iban** 5330 di cui frodati 1770

Numero totale di **iban riceventi** 31917 di cui **frodanti** 1986

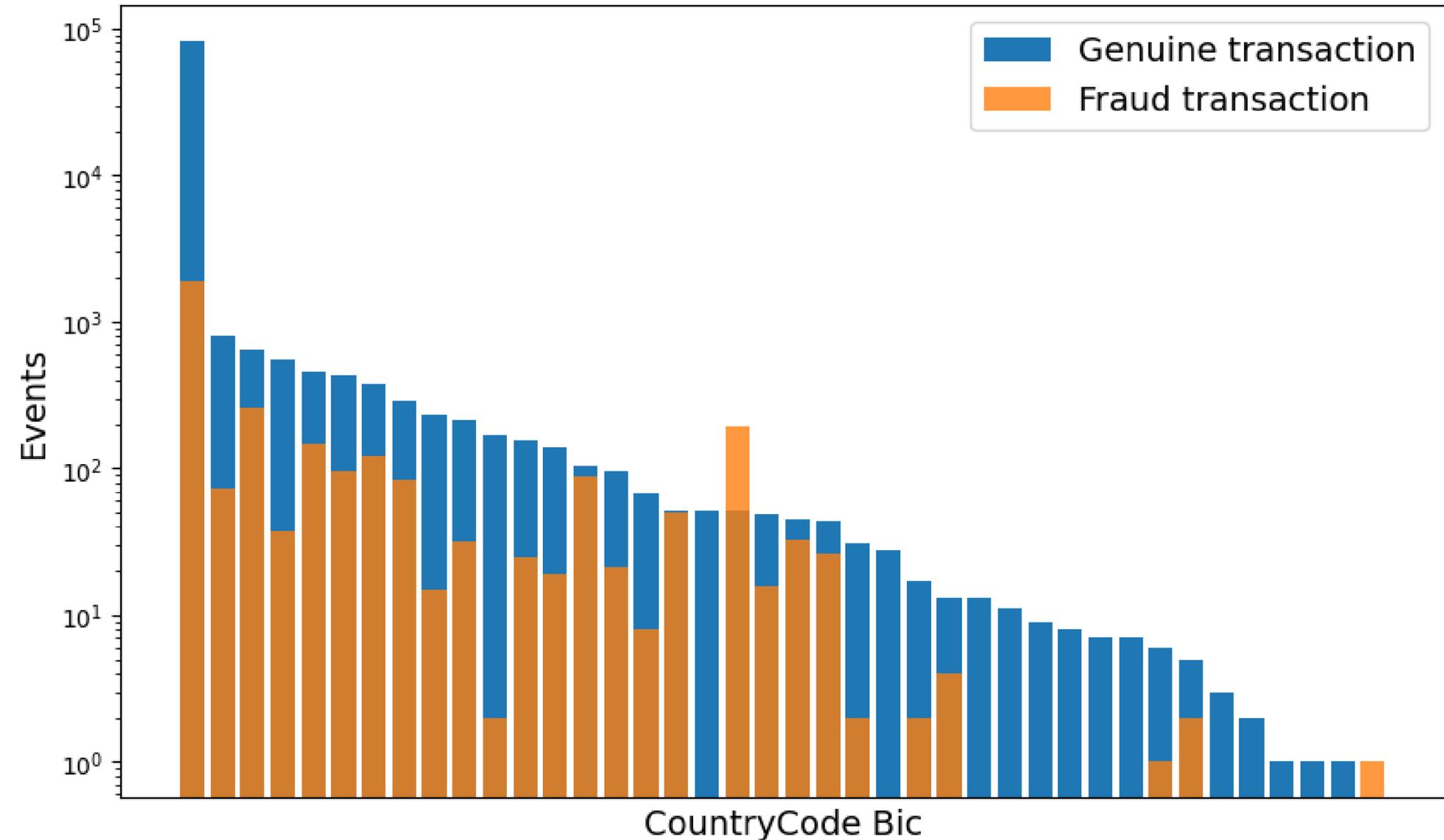
Numero medio di **transazioni per giorno** $N = 22.9$ di cui frodi $N = 1.65$

Features exploration

BIC: Country_Code

Total of 40 CountryCode

- ▶ **26 in common** between genuine and fraud
- ▶ meaning that **1 is only fraud**

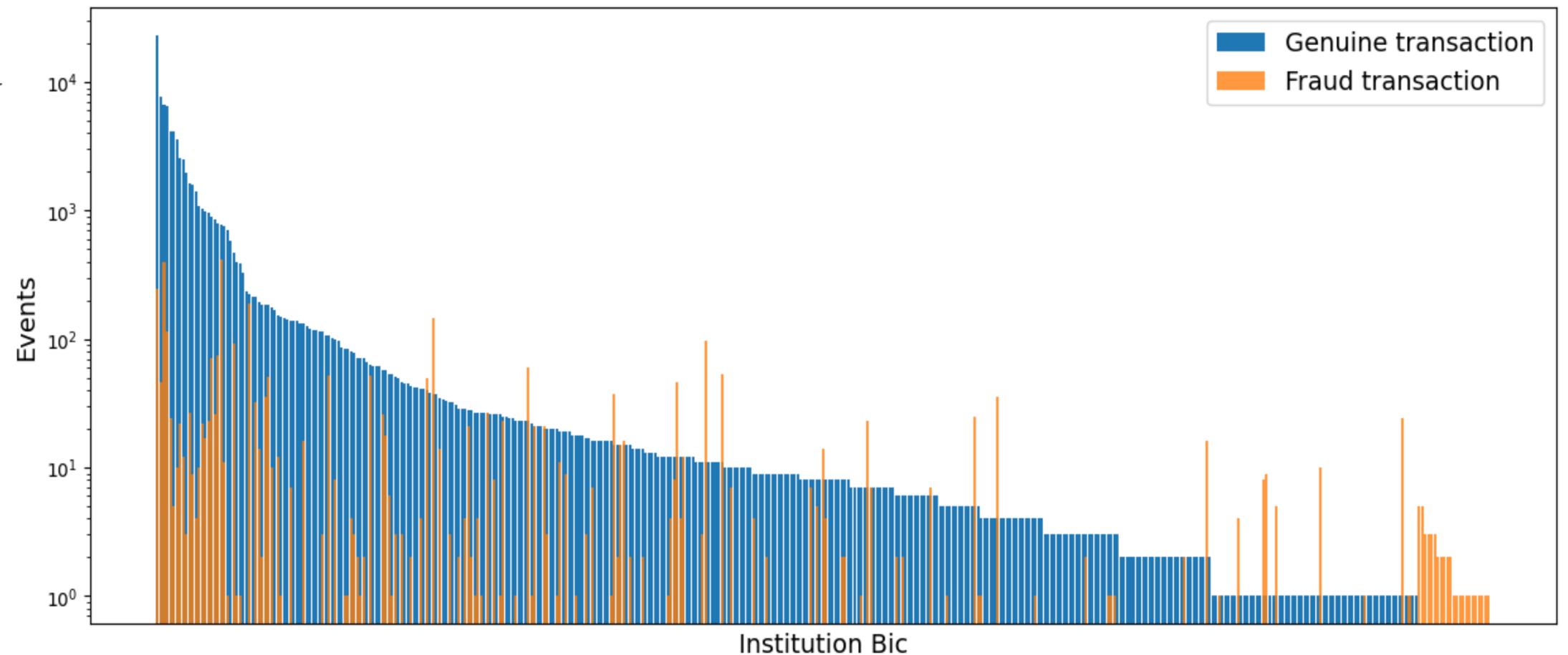


Features exploration

BIC: Institution_Code

Total of 421 InstitutionCode

- ▶ **135 in common** between genuine and fraud
- ▶ meaning that **23 are only fraud**

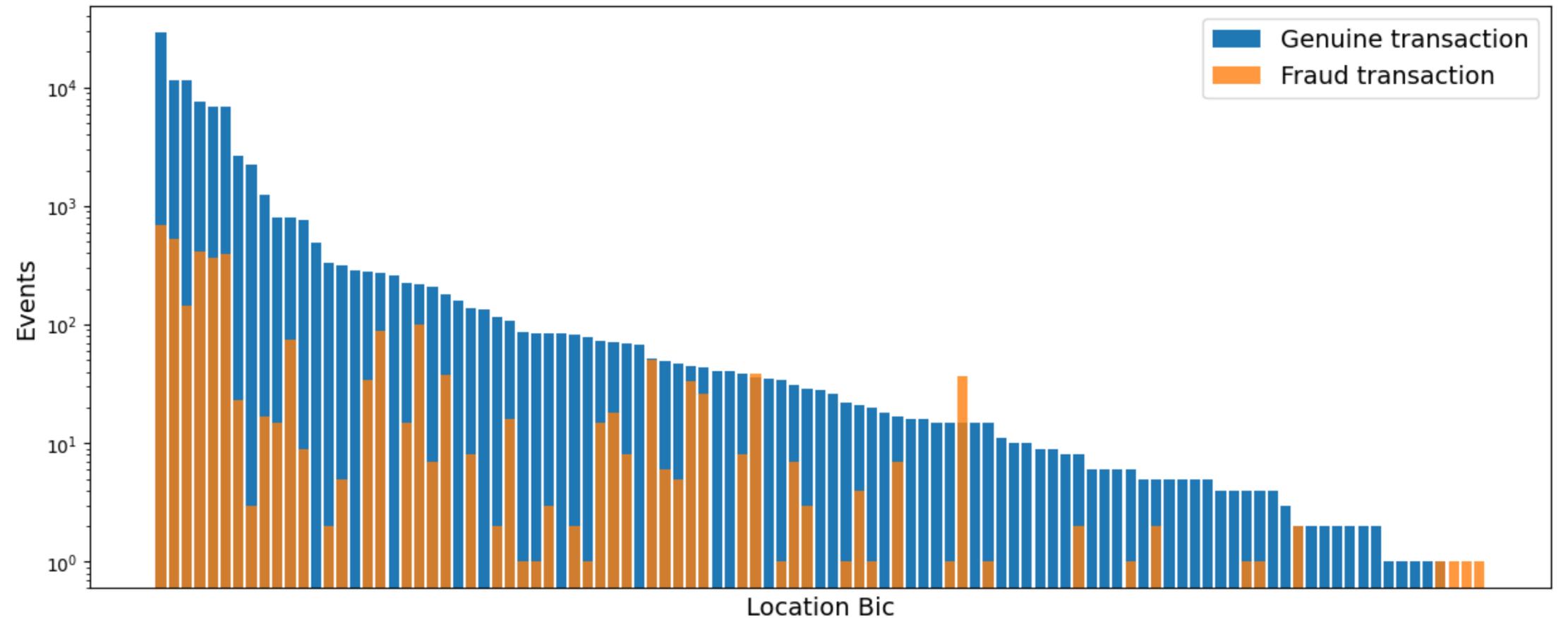


Features exploration

BIC: Location_Code

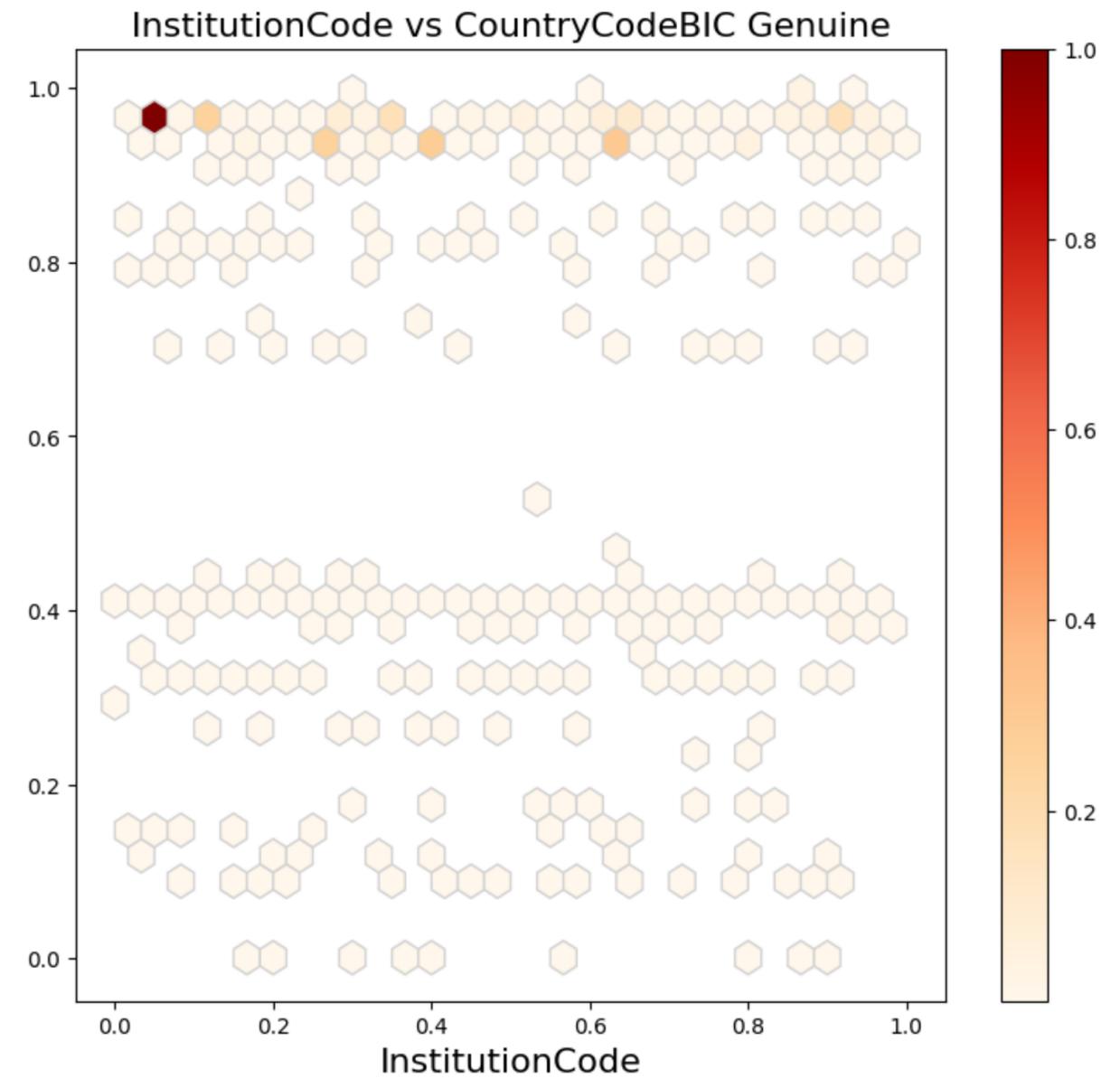
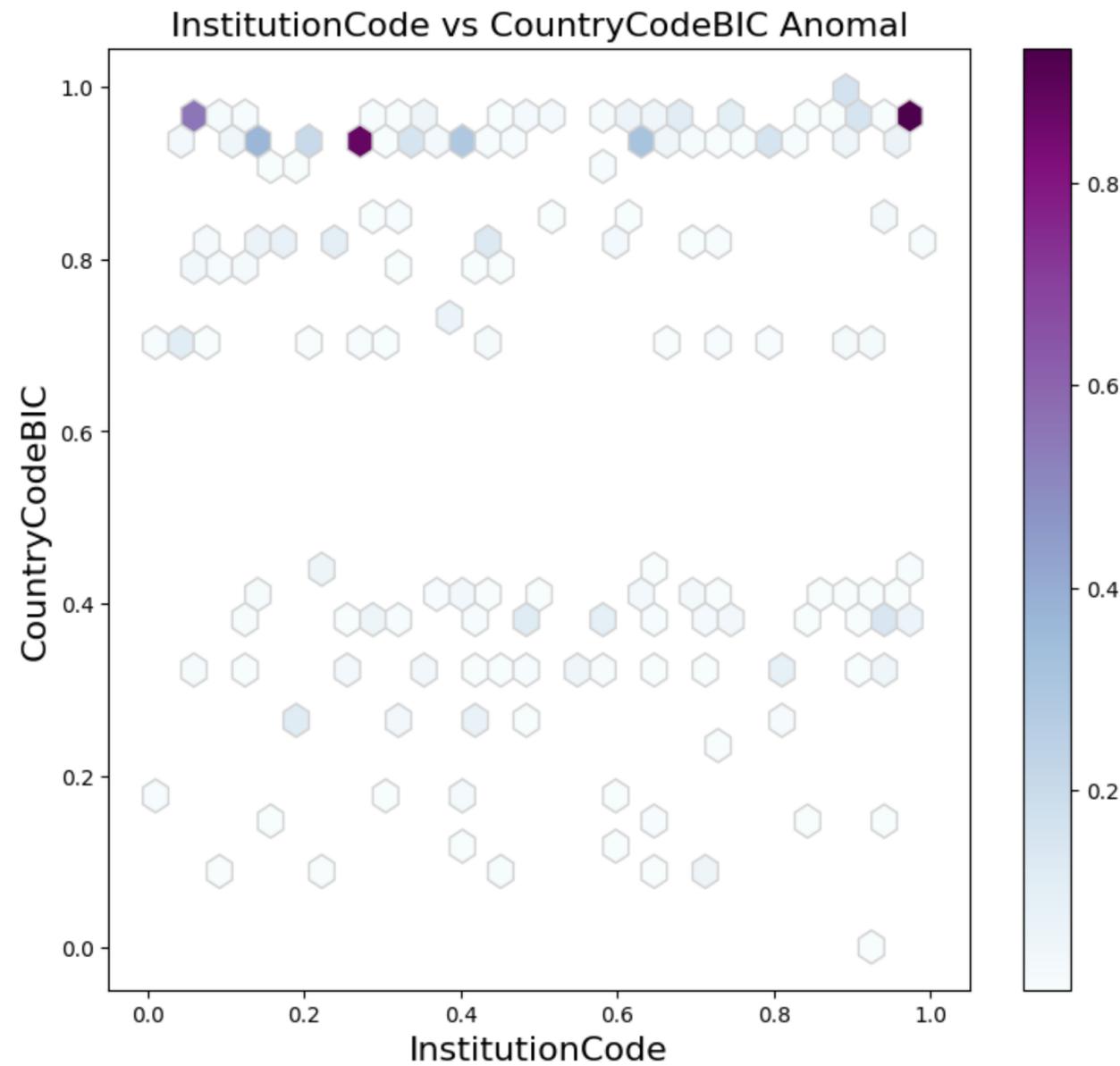
Total of 103 LocationCode

- ▶ **55 in common** between genuine and fraud
- ▶ meaning that **3 are only fraud**



Features exploration

BIC: CountryCode vs InstitutionCode

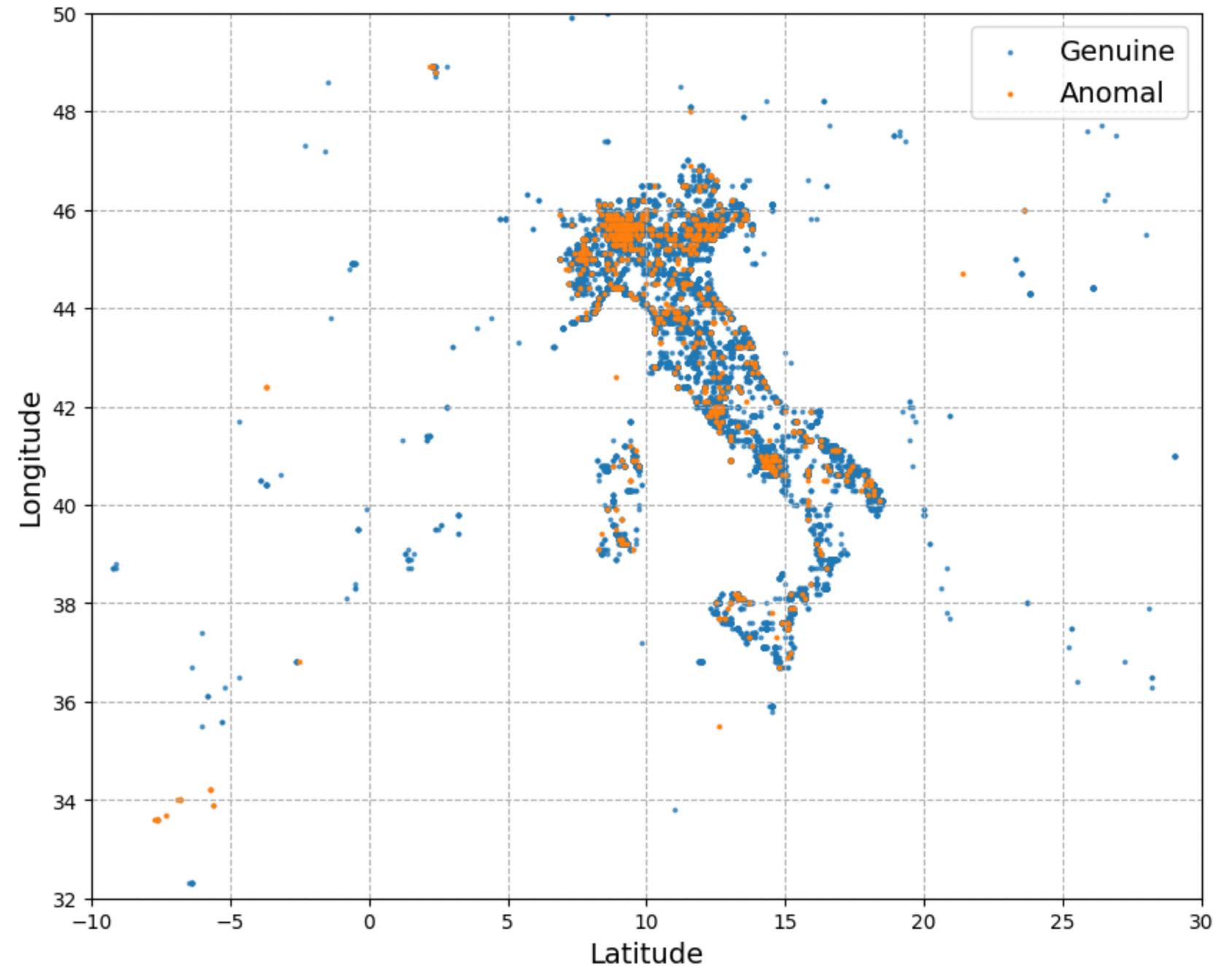


Features exploration

Posizione

Ratio di transazioni che **NON** hanno una posizione:

- ▶ genuine 54%
- ▶ fraudolente 64%



Frodatori

Iban riceventi

Ci sono **iban riceventi** che si ripetono nelle frodi?

- ▶ Ci sono 1986 iban nelle frodi
- ▶ Un totale di 403 iban in comune tra frodi e genuine, mentre 1583 compaiono solamente in transazioni fraudolente

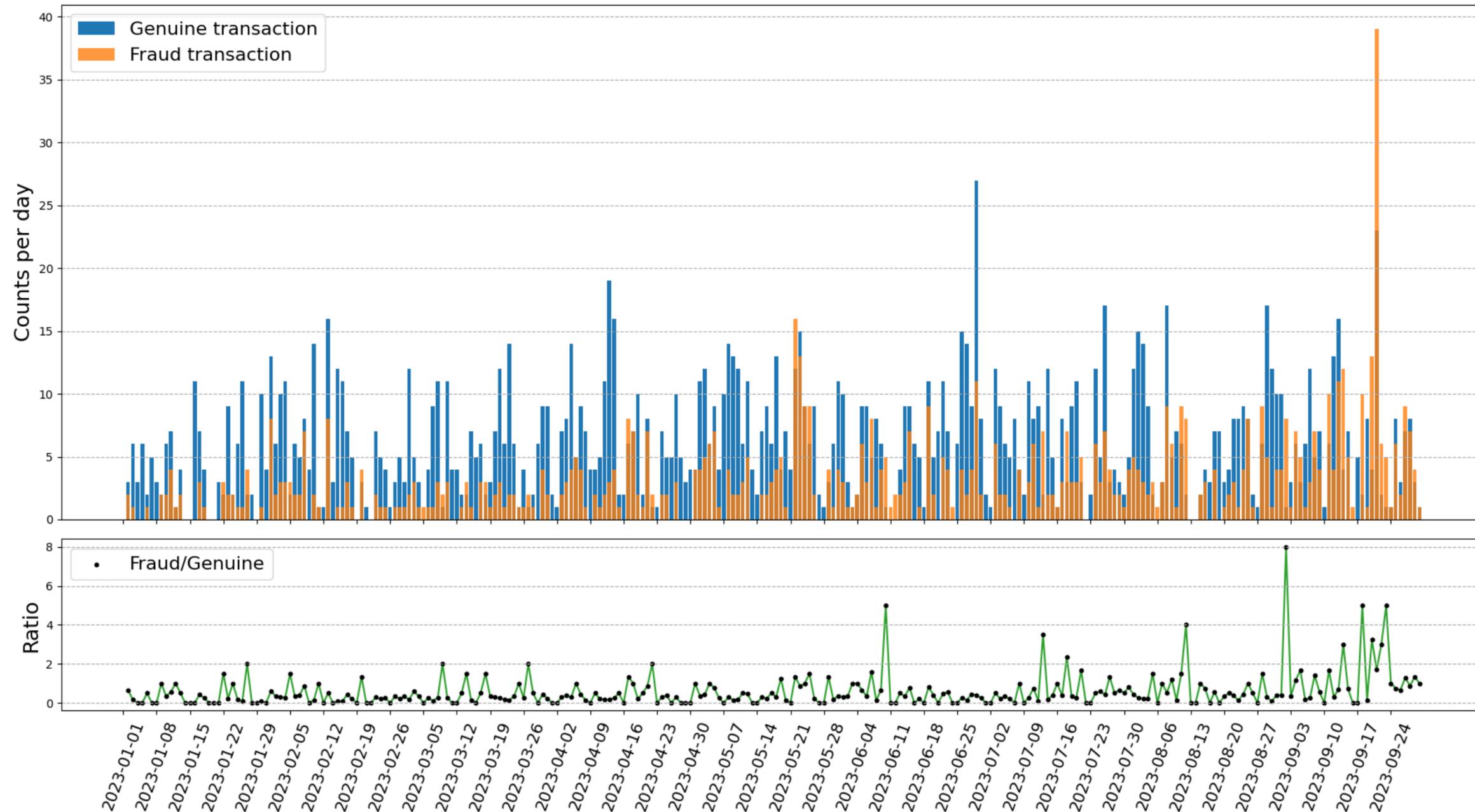
==> ci sono iban fraudolenti che ricevono anche transazioni genuine!

I 403 iban sono coinvolti in 2456 transazioni (di cui 32% fraud and 68% genuine)

Gli altri iban sono coinvolti in 2490 transazioni (100% fraud)

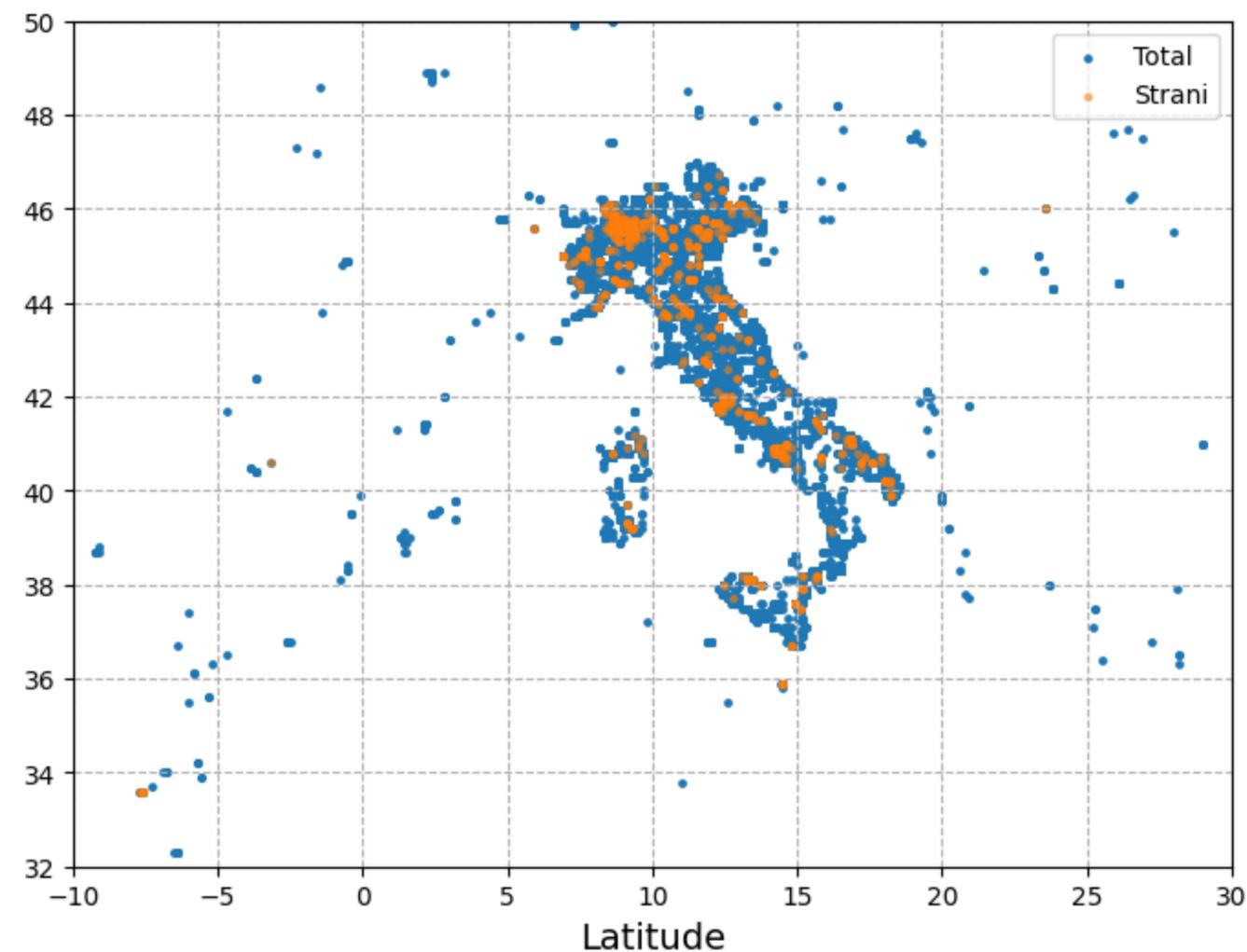
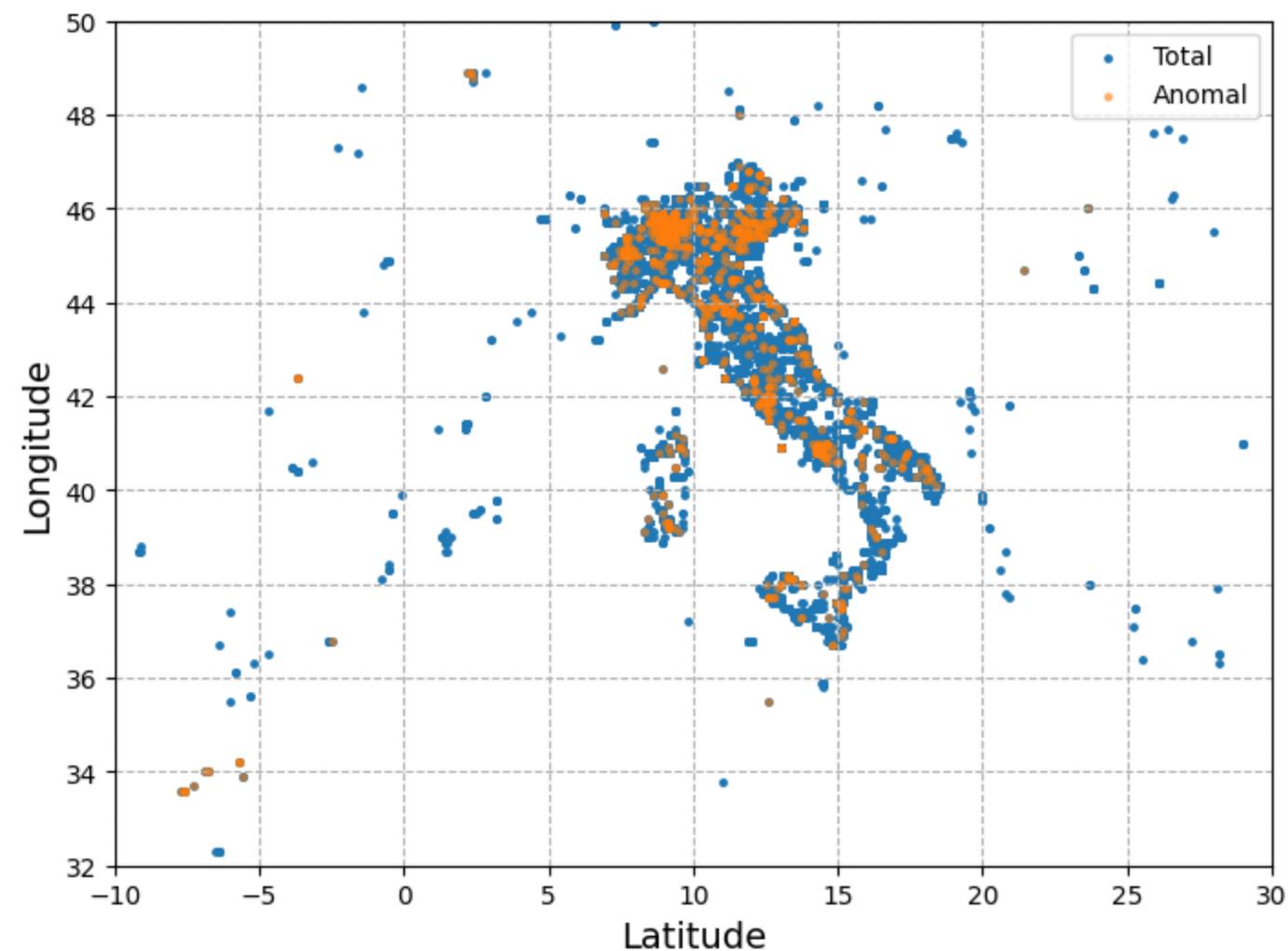
Frodatori

Iban riceventi frodatori e genuini



Frodatori

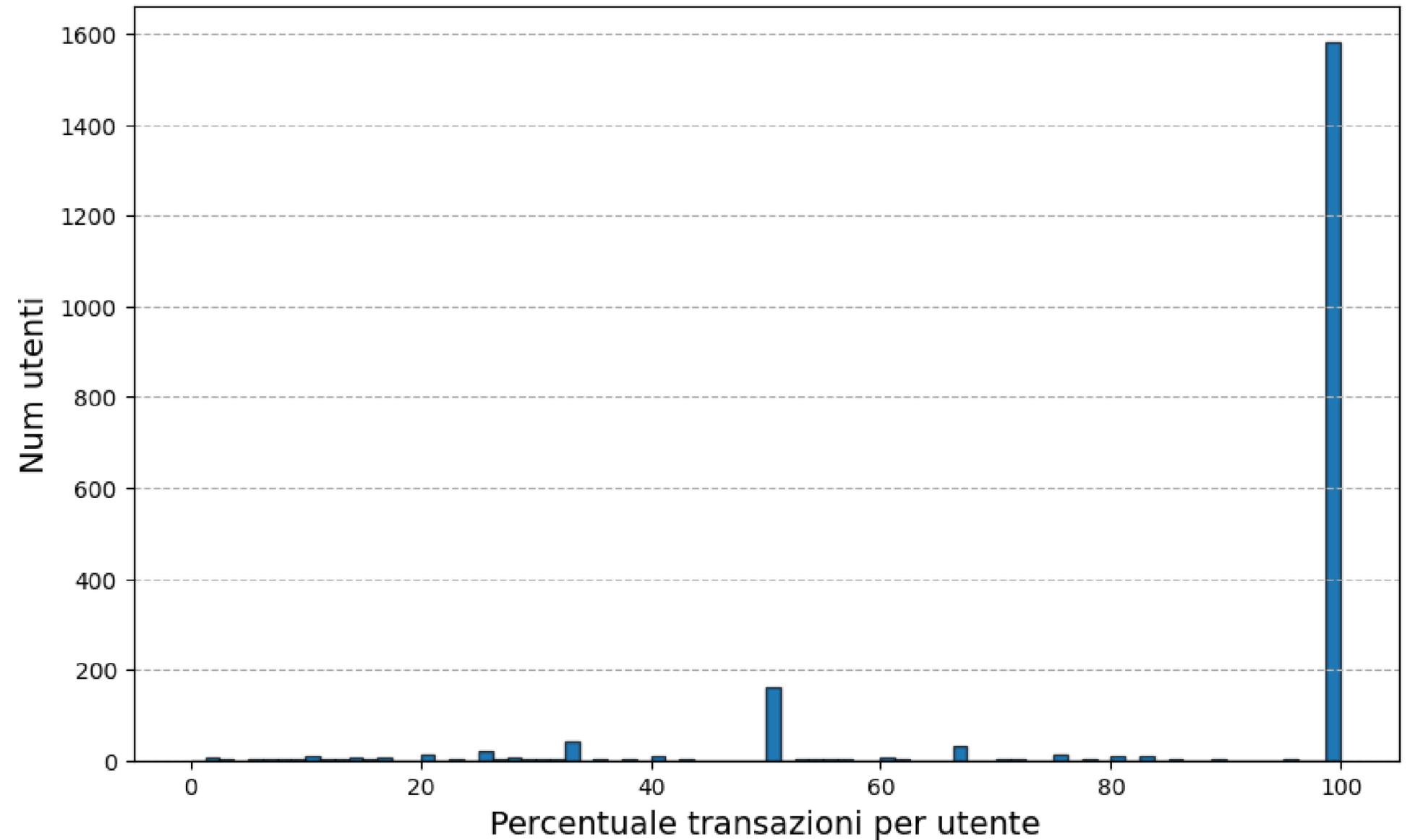
Iban riceventi frodatori e genuini



Frodi per utente

Ratio N_{frodi}/N_{tot} per iban sospetti

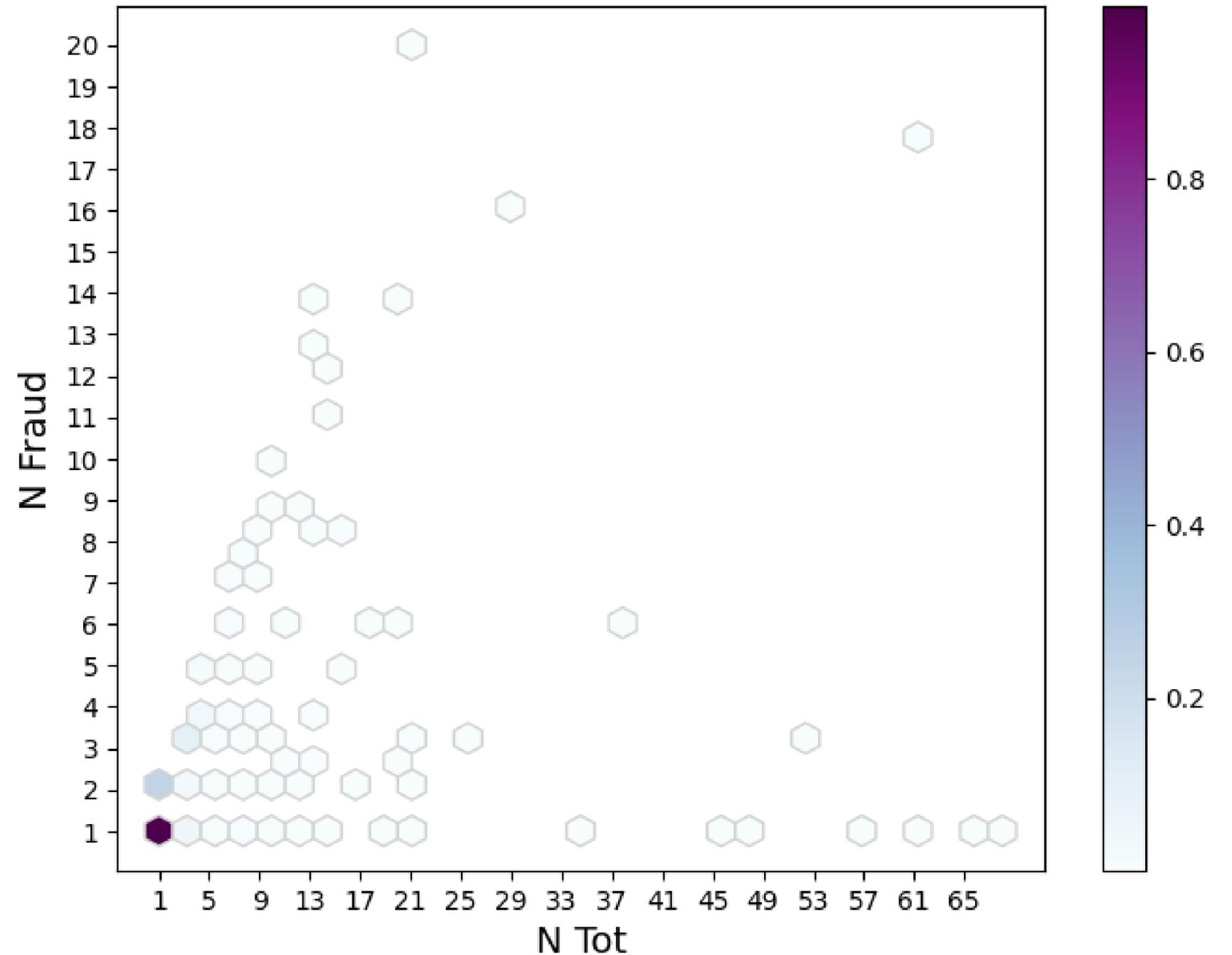
- ▶ Isoliamo gli **iban** (riceventi) che hanno **frodato almeno una volta**
- ▶ Consideriamo tutte le transazioni in cui sono coinvolti
- ▶ Calcoliamo rapporto N_{frodi}/N_{tot} per singolo iban sospetto



Frodi per utente

Ratio N_{frodi}/N_{tot} per iban sospetti

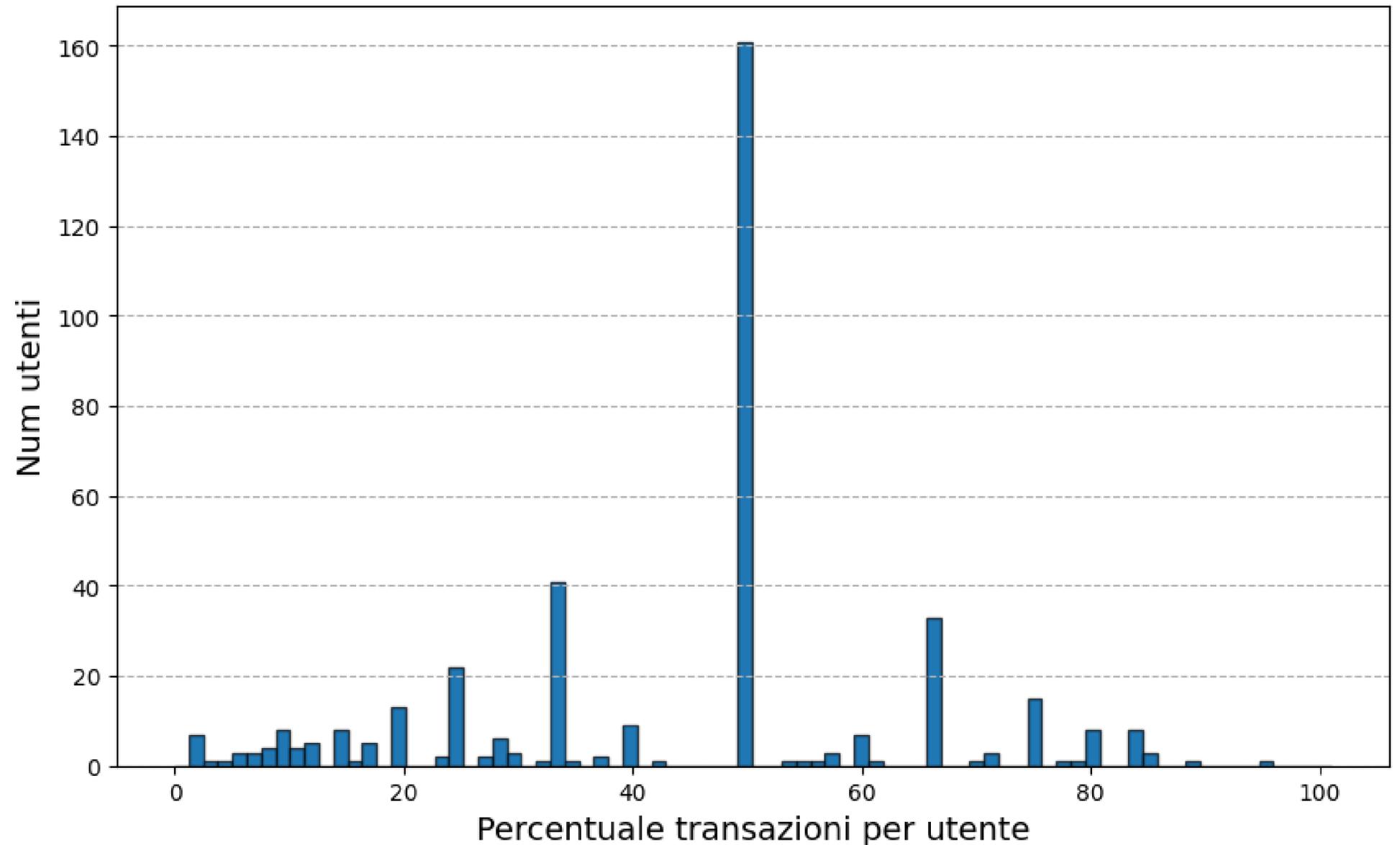
- ▶ Isoliamo gli **iban** (riceventi) che hanno **frodato almeno una volta**
- ▶ Consideriamo tutte le transazioni in cui sono coinvolti
- ▶ Istogramma 2D N_{frodi} vs N_{tot} per singolo iban sospetto



Frodi per utente

Ratio N_{frodi}/N_{tot} per iban sospetti

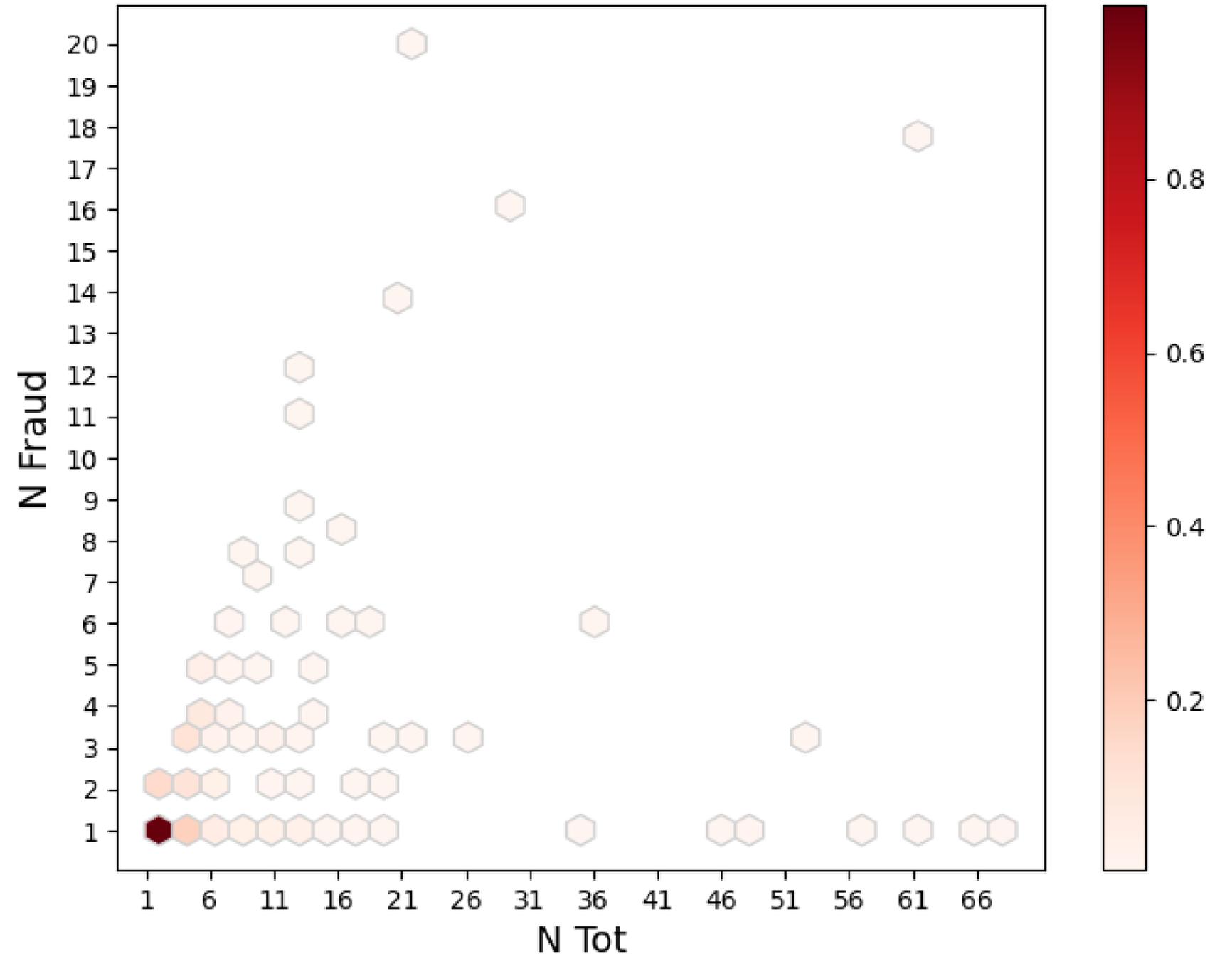
- ▶ Isoliamo gli **iban** (riceventi) che hanno **frodato almeno una volta**
- ▶ Consideriamo tutte le transazioni in cui sono coinvolti
- ▶ **Escludiamo i casi 100% frodi**
- ▶ Calcoliamo rapporto N_{frodi}/N_{tot} per singolo iban sospetto



Frodi per utente

Ratio N_{frodi}/N_{tot} per iban sospetti

- ▶ Isoliamo gli **iban** (riceventi) che hanno **frodato almeno una volta**
- ▶ Consideriamo tutte le transazioni in cui sono coinvolti
- ▶ **Escludiamo i casi 100% frodi**
- ▶ Istogramma 2D N_{frodi} vs N_{tot} per singolo iban sospetto



Frodi per utente

Iban sospetti

- ▶ Isoliamo gli **iban** (riceventi) che hanno **frodato almeno una volta**
- ▶ Consideriamo tutte le transazioni in cui sono coinvolti
- ▶ Calcoliamo rapporto N_{frodi}/N_{tot} per singolo iban sospetto
- ▶ Isoliamo quei iban che hanno un ratio $N_{frodi}/N_{tot} < 10\%$ or $=50\%$ or $>80\%$

Frodi per utente

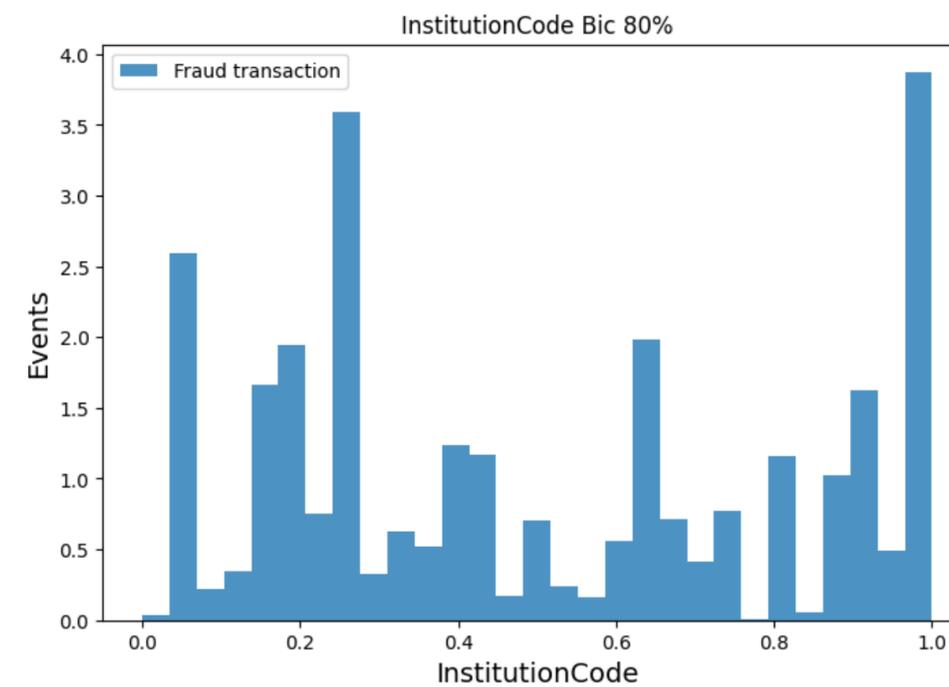
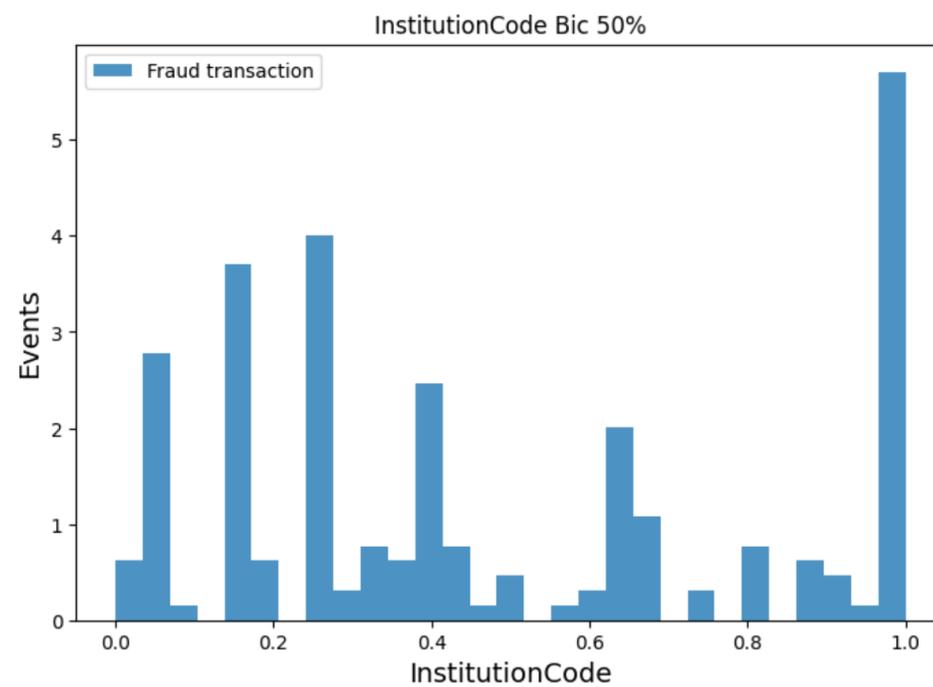
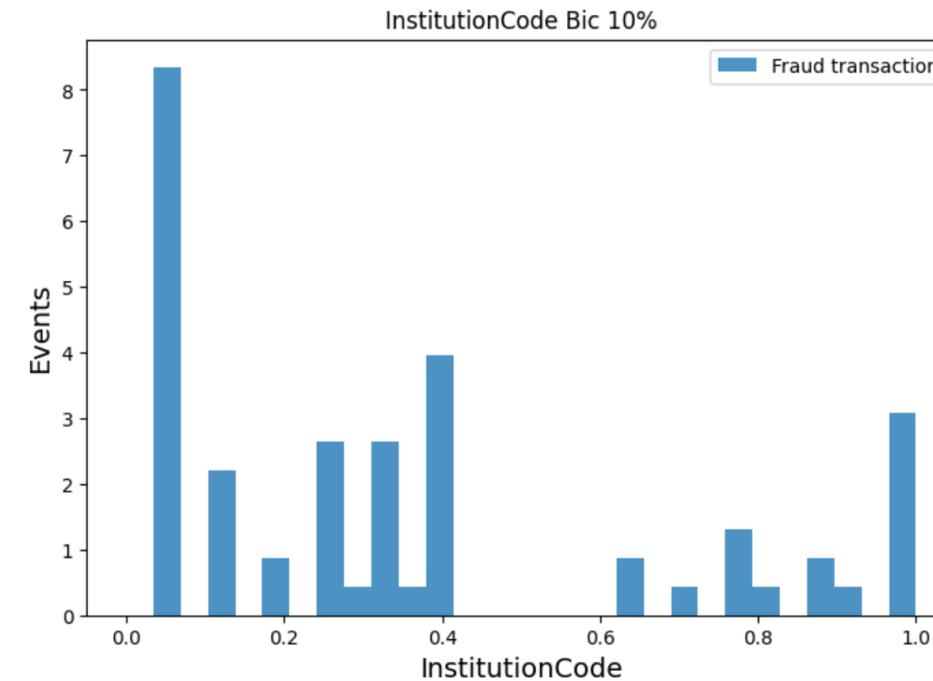
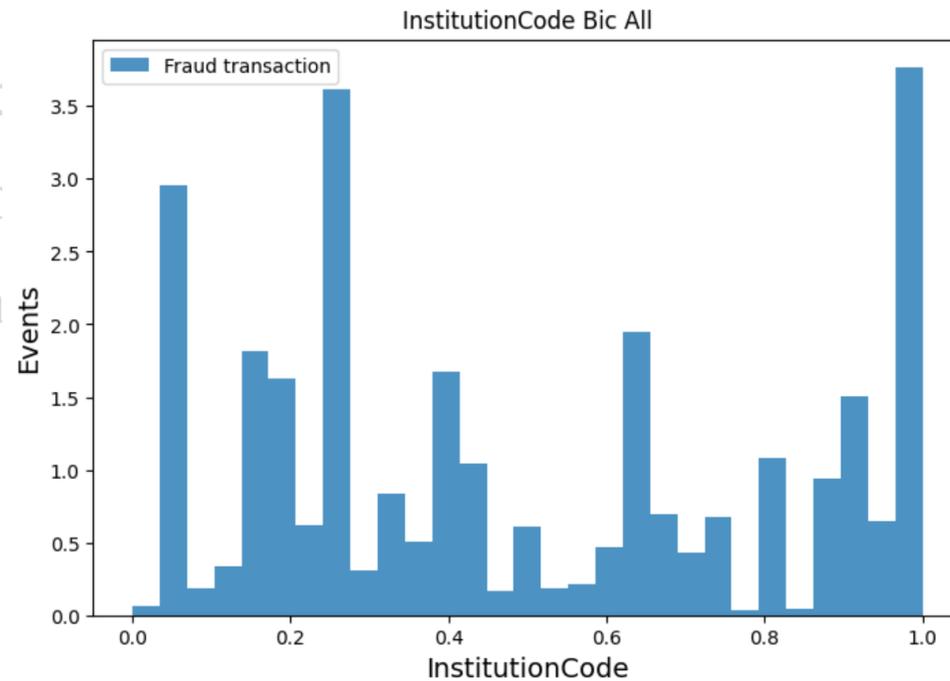
Iban sospetti

- ▶ Isoliamo gli **iban** (riceventi) che hanno **frodato almeno una volta**
- ▶ Consideriamo tutte le transazioni in cui sono coinvolti
- ▶ Calcoliamo rapporto N_{frodi}/N_{tot} per singolo iban sospetto
- ▶ Isoliamo quei iban che hanno un ratio $N_{frodi}/N_{tot} < 10\%$ or $=50\%$ or $>80\%$
- ▶ **Importo medio:**
 - ▶ **Ratio < 10%:** 5053.2 €
 - ▶ **Ratio = 50%:** 1915.3 €
 - ▶ **Ratio > 80%:** 2855.2 €

Frodi per utente

Iban sospetti

- ▶ Isoliamo gli iban
- ▶ Consideriamo
- ▶ Calcoliamo ra
- ▶ Isoliamo quei



Conclusioni

