IDA findings

EDA findings

Col basic stats, counts

Datetime: genuine trades -peak 7am-4pm; fraud: random distribution

Merchant category/MCCs: evenly distributed

Type of card used: investigate

Card not present: online more subject to fraud

Ip: blacklist IP

Correlation matrix, Before transformation, numerical cols amount has something to do with fraud

New features

Amount: mostly below 2000; fraud has way lower median (below 100) but wider value range (longer tail, larger outlier)

Transformation logic

country\_code\_date\_sec\_since\_midnight

IP investigation🡪IP address country, country code date

Difference between genuine and fraud transactions

customer\_id\_distinct\_ip in n days

more than two customer – likely hacker hacking multiple customer with same IP

seconds\_since\_last\_ip : seconds since the same IP last appeared for each customer

date🡪GB weekdate, GB holiday, GB weekend

transaction\_date\_day transaction\_date\_month

trading interval by customer id: customer\_id\_sec\_since\_last\_trade

established\_period\_in\_day: merchant history

ma\_amount by customer id: 1 week 2 weeks 1 month, 2 months

transaction\_date\_sec\_since\_midnight

customer\_id\_sec\_since\_last\_trade : trades frequency

established\_period\_in\_day: merchant newly established?

merchant\_category: one-hot

type\_of\_credit\_card\_used\_with\_mer\_id

is\_merchant\_card + is\_merchant\_mismatch + is\_merchant\_card\_and\_mismatch (intersact feature)