Sprint 3
Calibrated Techniques: Ways to Improve Fraud Detection Systems

Team Patrick
Shawn | Marco | Avie | Bevs

Credit Card Fraud

- Fraud committed using a credit card or any similar form of payment mechanism
- The purpose is to obtain unauthorized funds from the credit cardholder's account

Philippine Outlook
Global Outlook

- ➔ 25% of complaints received were related to credit cards
- ➔ March -May 2020
  - ◆ 98.4% of criminal incidents reported were cyber or online in nature
  - ◆ Losses equivalent to 60.6 M or 54.5% of all total bank losses

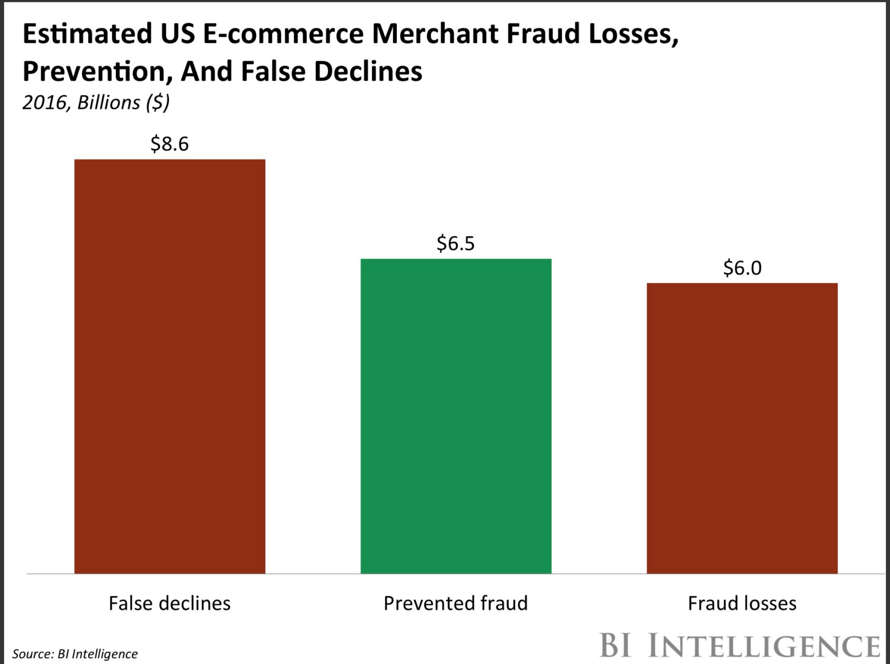
Source: Merchant Savy UK, 2020

Problem Overview

- Fraud Detection System
  - Tags fraudulent transactions
  - Automatic process
- Case study:
  - CEO's card was blocked
  - CEO was on travel

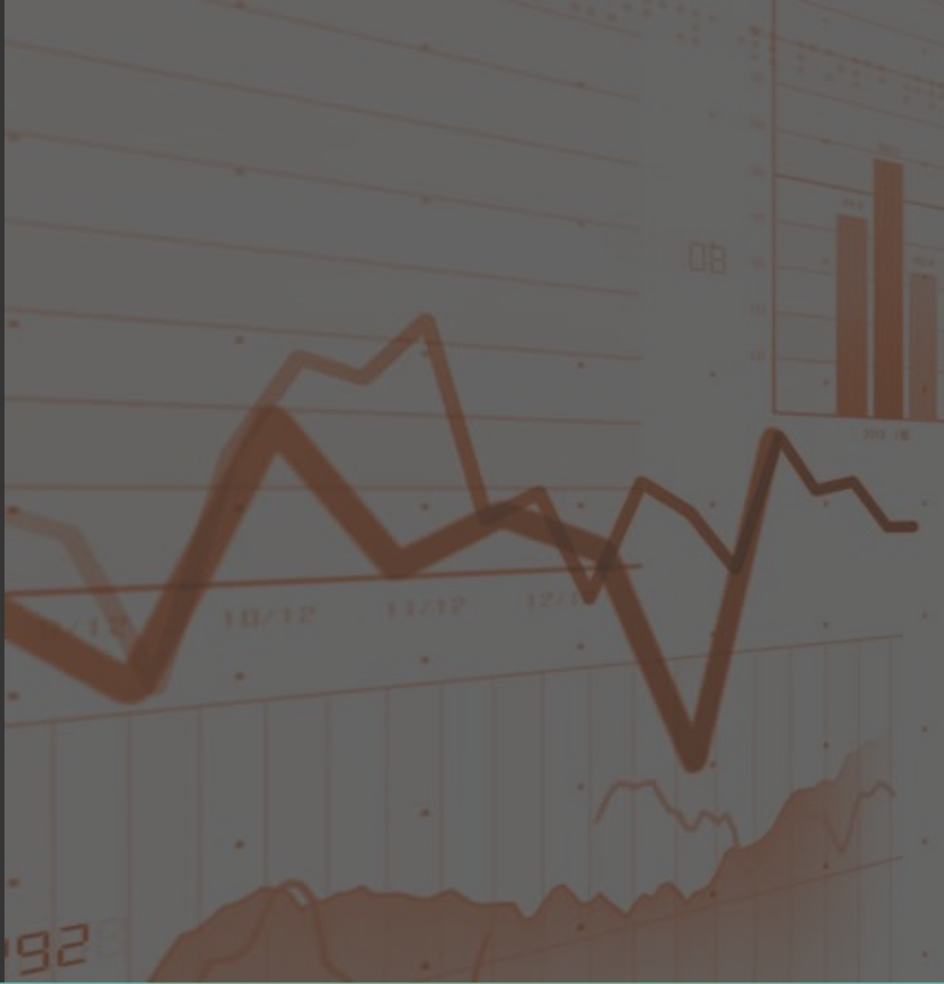
The Case of False Positives

- Occurs when merchants or financial institutions decline legitimate orders

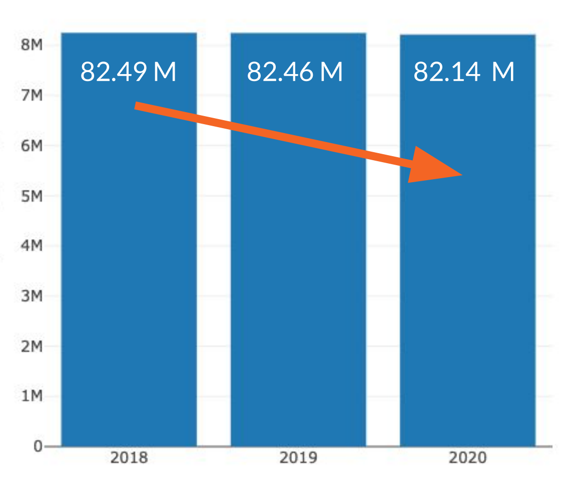


Objectives

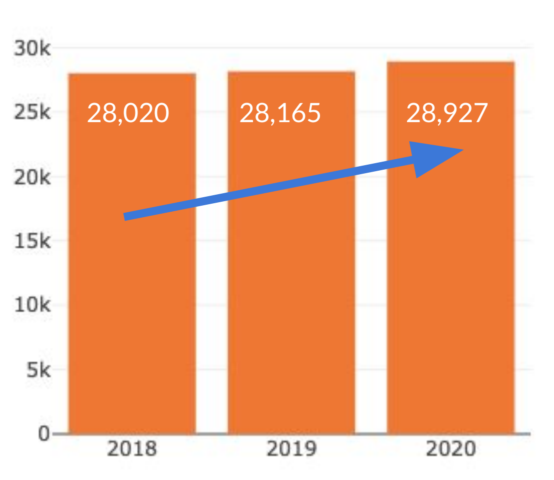
1. Review current fraud detection system
2. Recommend ways to make the system better



Overview of Fraudulent Transactions



Total credit card transactions (2018-2020)

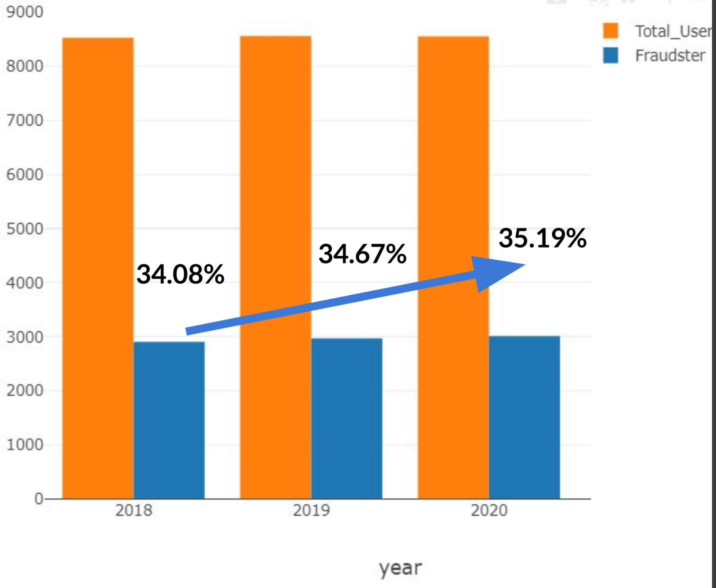


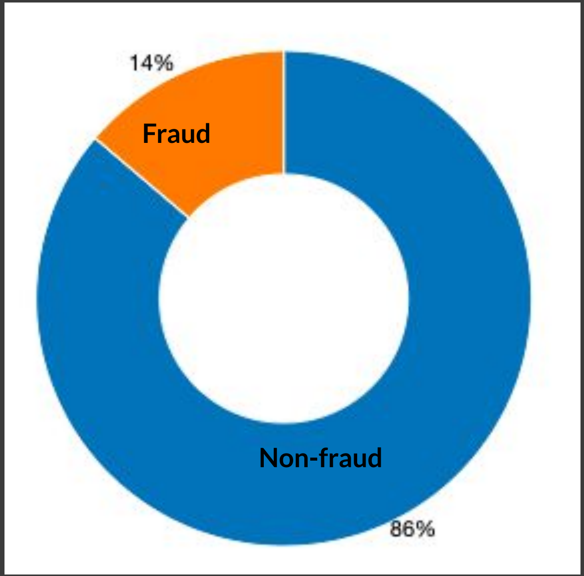
Fraudulent transactions (2018-2020)

For the last 3 years:

- gradual decrease in total credit card transactions
- gradual increase in fraudulent transactions

Overview of Fraudulent Transactions





For the last 3 years, ~35% of total users experience fraud annually

~14% of the most recent user transactions are fraudulent

Client Profile data

ssn | credit card number | account number | name | sex | address | profession | birthdate

Spending Behavior data

transaction number | transaction date and time | product category | amount | merchant | merchant location

Amount

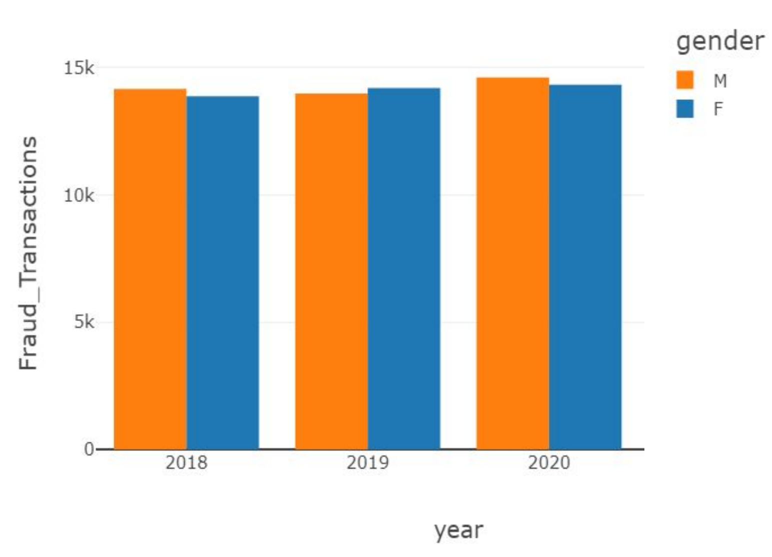
Geospatial Occurrence

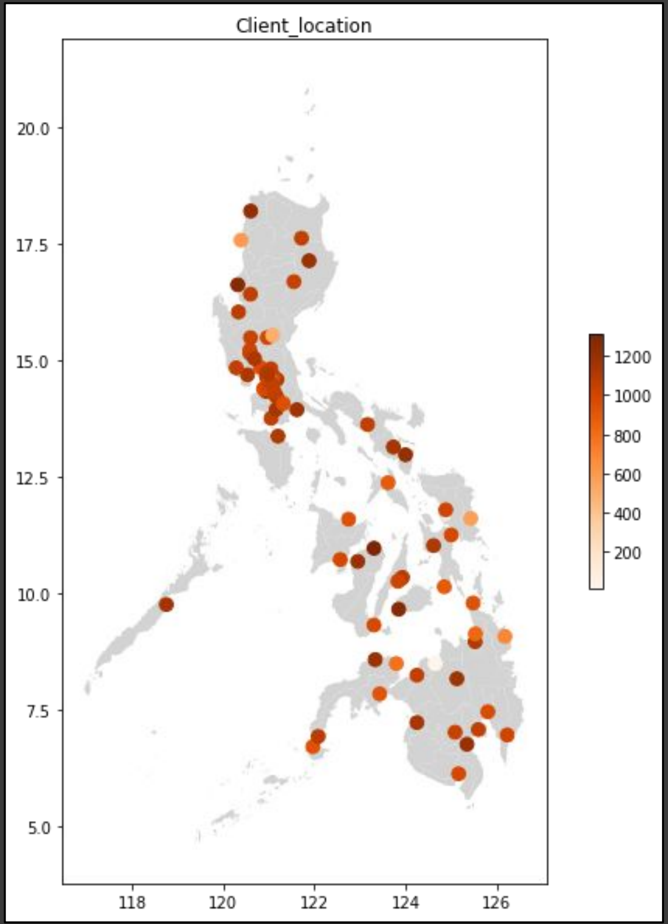
Merchants

Product Category

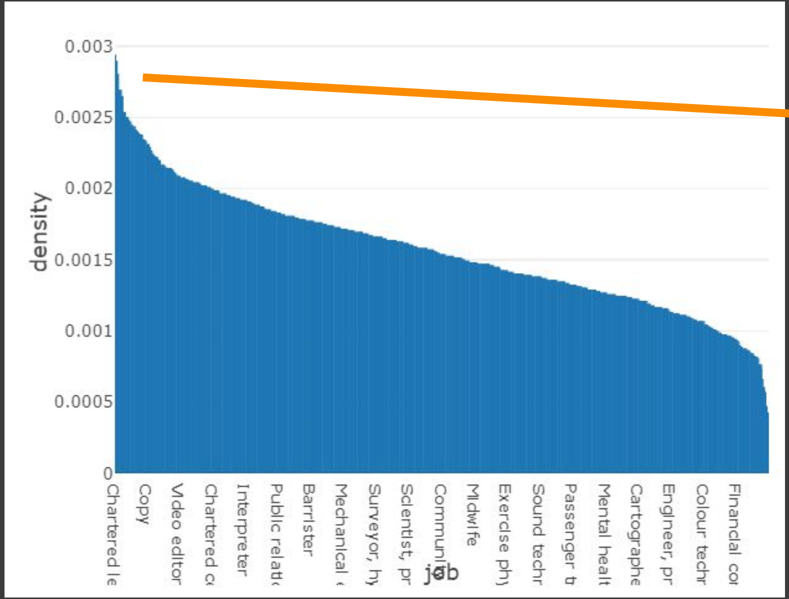
Time Period

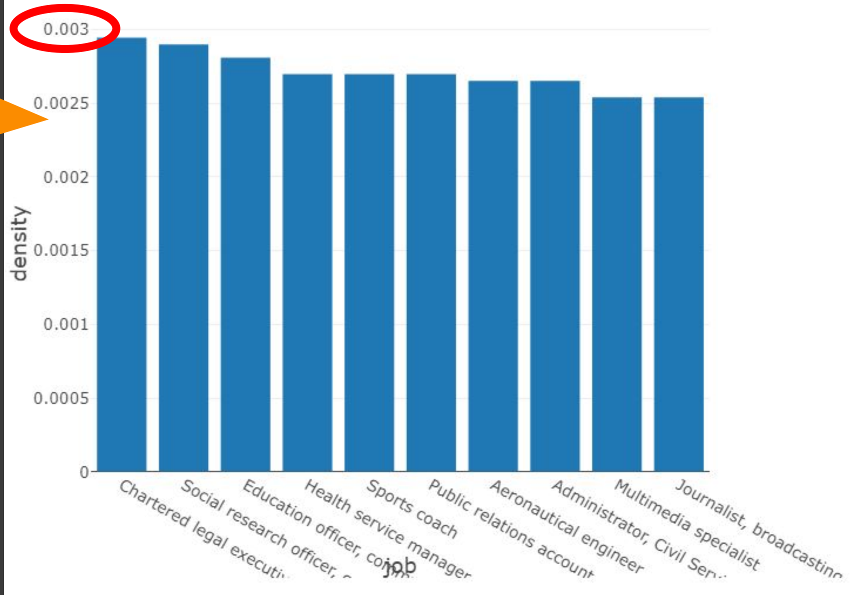
Fraudulent Transactions: Client Profile (Sex, Address)



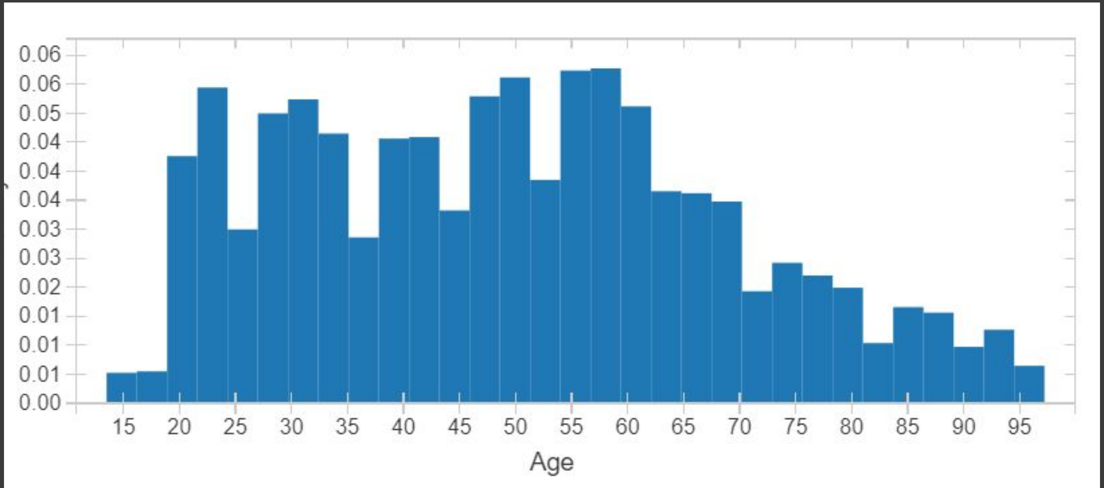


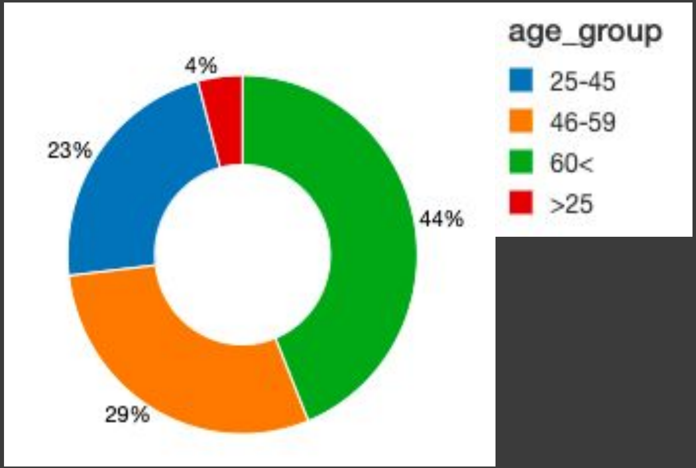
Fraudulent Transactions: Client Profile (Job)



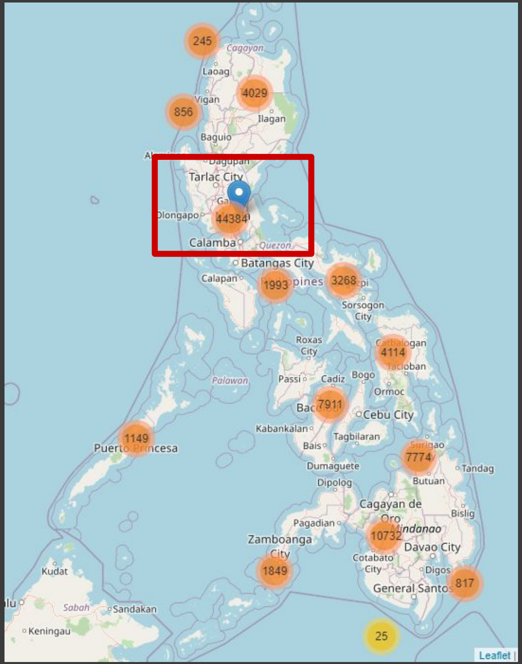


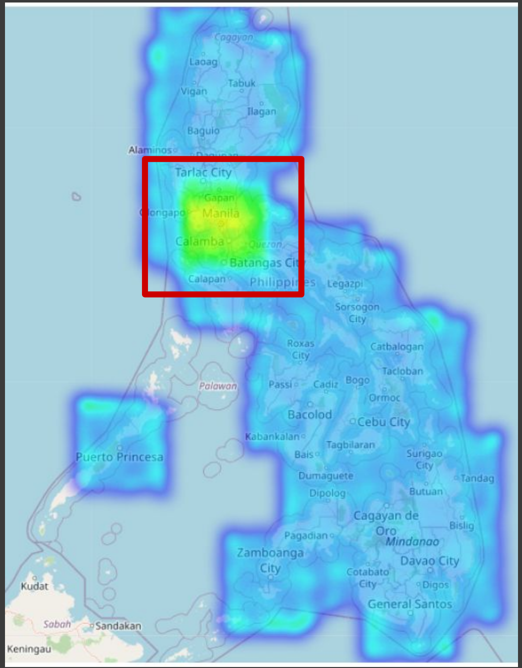
Fraudulent Transactions: Client Profile (Age)





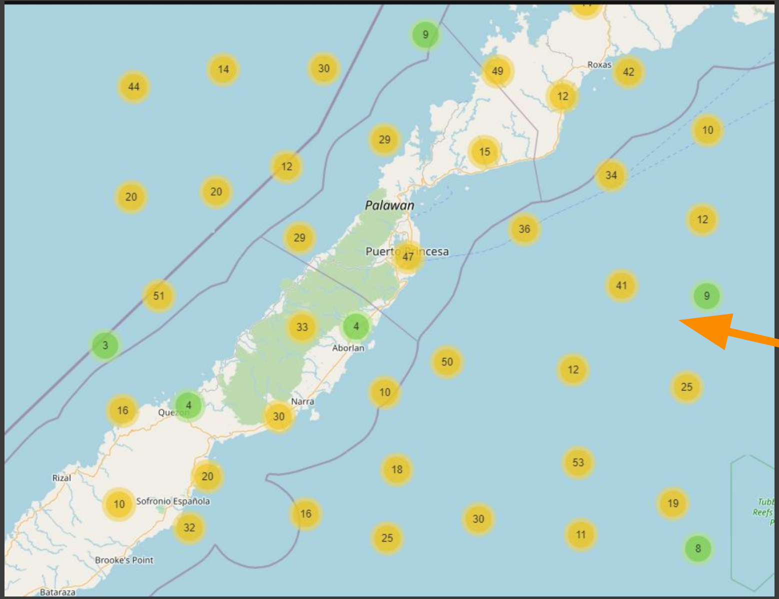
Fraudulent Transactions: Geospatial Occurrence





Fraud transactions mostly occurs on **NCR Area and nearby provinces**

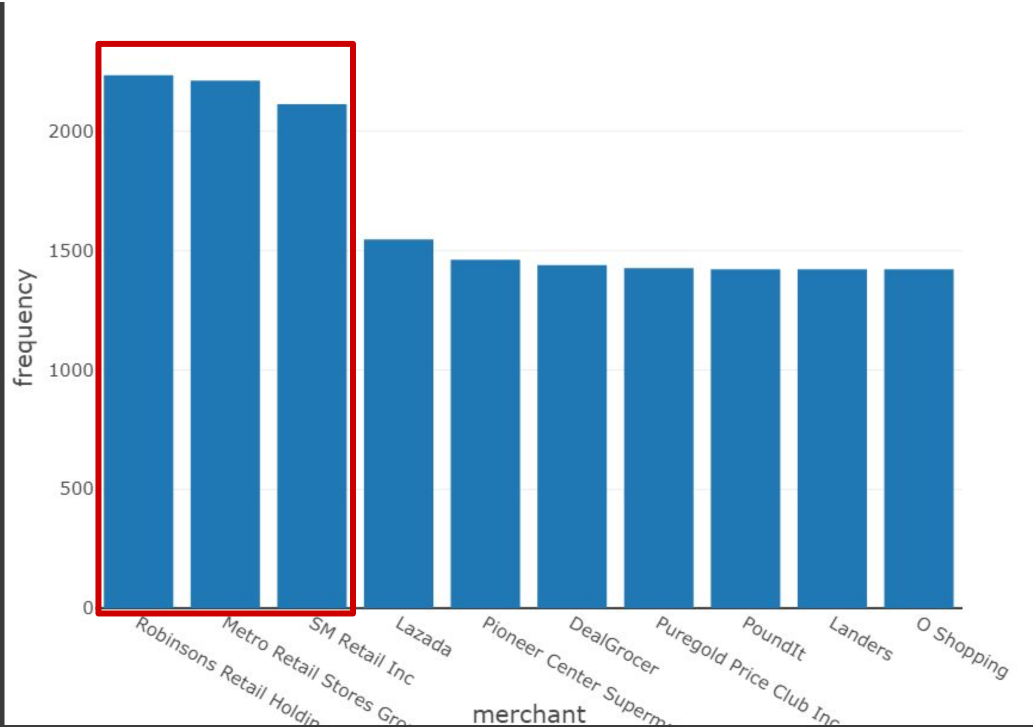
Fraudulent Transactions: Geospatial Occurrence



Fraud transactions mostly occurs on **NCR Area and nearby provinces**

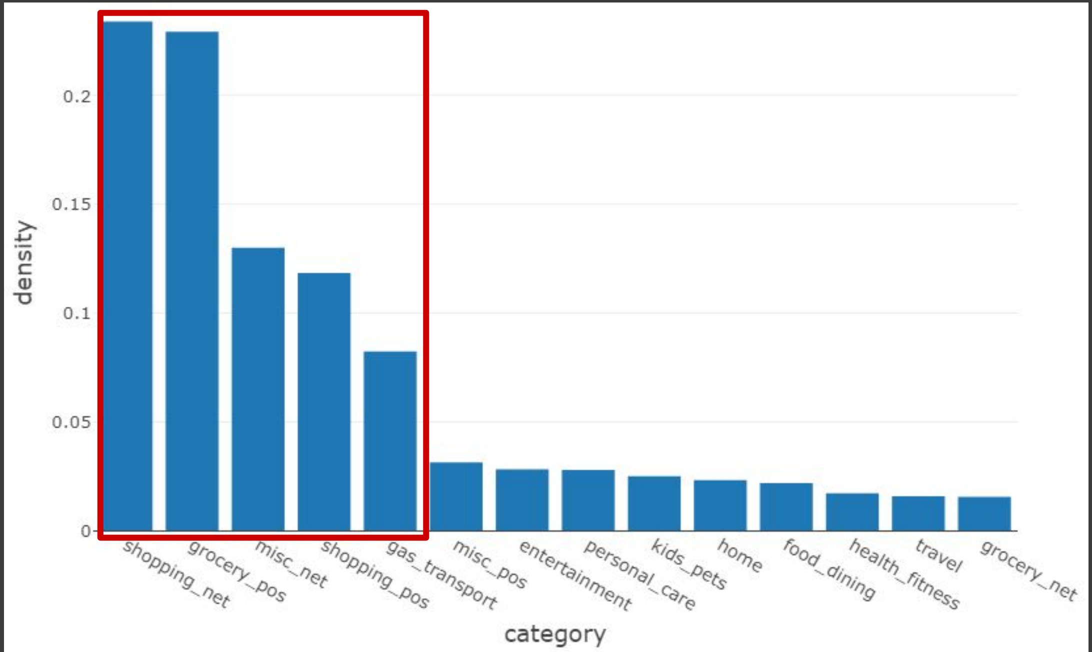
However, geospatial points are inaccurate. Many locations are outside land areas

Fraudulent Transactions: Most Common Merchants



Robinsons, Metro Retail Stores Groceries, and SM Retail Inc. are merchants with the highest occurrence of fraud

Fraudulent Transactions: Product Category

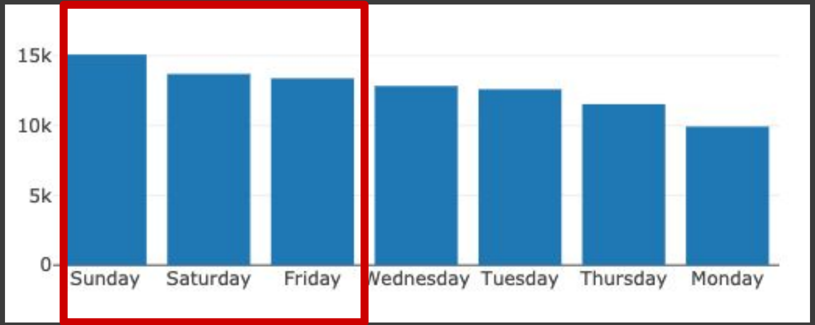


~80% of total fraud transactions came from the following categories

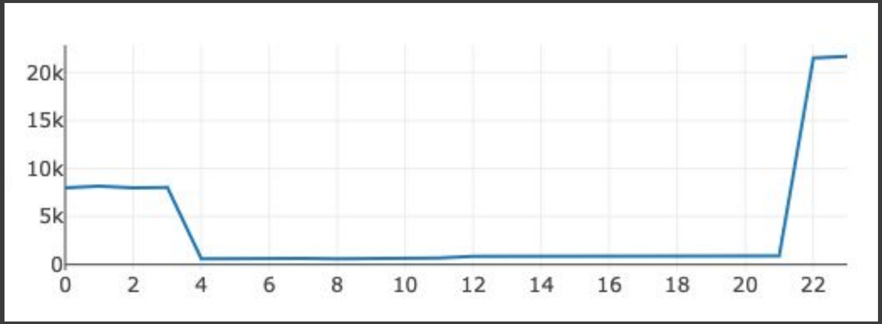
- Online Shopping (23.4%)
- Groceries (22.9%)
- Misc. Online Transactions (13%)
- Shopping (11.8%)
- Gas and Transport (8.2%)

Fraudulent Transactions: Time Period of Occurrence

- Fraudulent transactions occur most often during the weekend (Friday-Sun) and at 10PM - 3AM



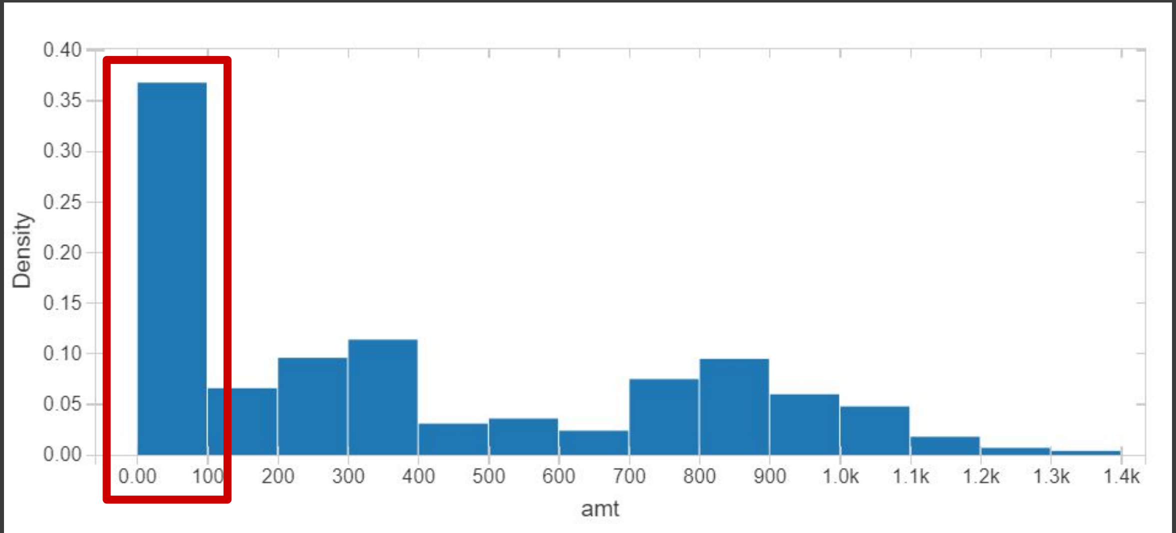
Total number of fraudulent transactions per day of the week



Total number of fraudulent transactions per hour

Fraudulent Transactions: Most Common Amount

- **Majority of fraudulent transactions costs PHP 100 or less**



Common Features of Fraudulent Transactions

Time Occurence

10PM-3AM

Day Occurrence

Friday to Sunday

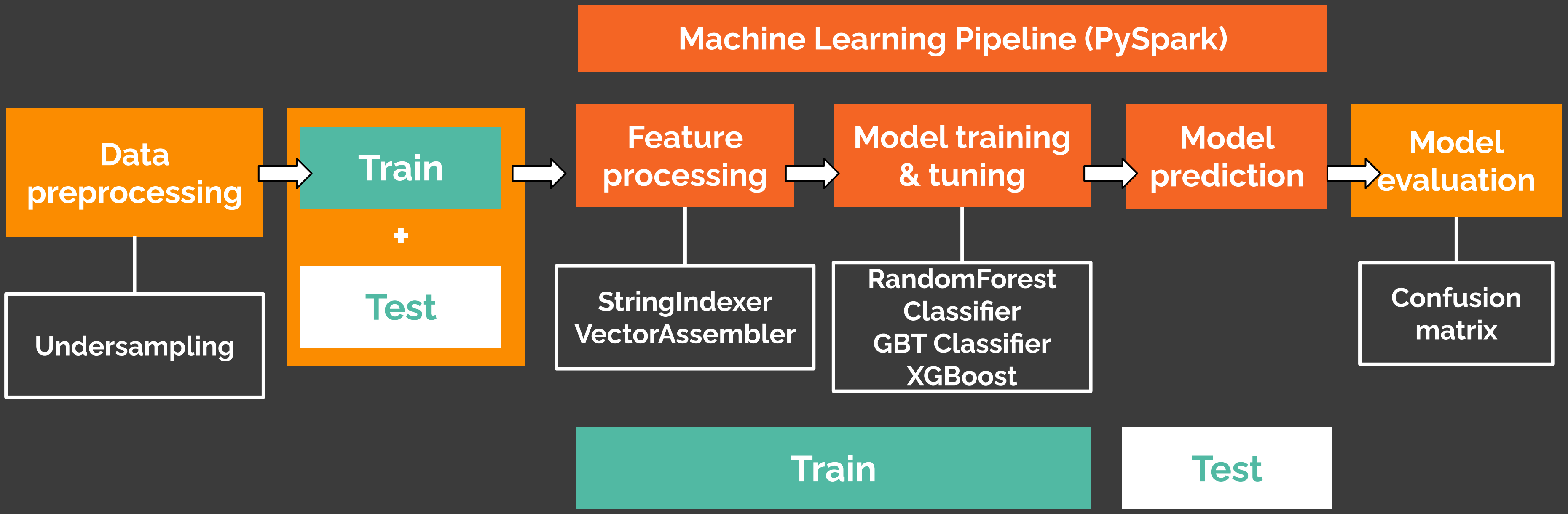
Transaction Amount

PHP 100.00 and below

Product Category

Online Shopping, Groceries, Online Transactions, Shopping, Gas and Transport

Fraud Detection System Version 1.0



Fraud Detection System Version 1.0 Performance

Accuracy:

How many Fraud and Legitimate Transactions were predicted accurately out of total transactions

Precision:

How many of those predicted as Legitimate are actually Legitimate

*single class metric

Recall:

How many Fraud cases are accurately identified as Fraud

*single class metric

Fraud Detection System Version 1.0

GBTClassifier: XGBoost

	Precision	Recall
Non-fraud	97%	97%
Fraud	97%	97%

Random Forest Classifier

	Precision	Recall
Non-fraud	94%	97%
Fraud	96%	94%

Fraud Detection System Version 1.0

GBTClassifier: XGBoost

```
gbt = GBTClassifier(labelCol='label', featuresCol="features", maxBins=250)
```

Feature Importance

Amount	0.6518	Age	0.0039
Time	0.1489	Gender	0.0019
Merchant	0.1192	Merchant Distance	0.0
Category	0.0742	Day of the Week	0.0

Confusion Matrix

TP: 17284	FP: 474
FN: 470	TN: 17368

Fraud Detection System Version 2.0

Conclusion

- False positives must be reduced in order for fraud detection mechanisms to be cost effective
- Issues with false-positives can be summarized into 3 categories:
    1. Identity-related
    2. Technical
    3. Structural
- Consumer spending behavior VARY GREATLY
- Feature engineering is (almost) everything
- Need for a better model that can provide a deeper analysis of spending behavior

Fraud Detection System Version 2.0

Recommendations

- Identity-related (historical data + transaction data)
  - biometrics, IP address, conflicting billing + shipping information, updated card information...
- Technical (bank)
  - local domains, smart routing
- Structural (Version 2.0):
  - Increase processing power for further hyperparameter tuning
  - Extract better features
  - Increase number of features
  - Consistent updating and review