



E SKWE BANKS

Sprint 3: Credit Card Fraud Detection

TEAM PATRICK

HURLY | SEF | TIM | TY



What is Credit Card Fraud?

"a wide-ranging term for theft and fraud committed using a credit card or any similar payment mechanism as a fraudulent source of funds. The purpose is to obtain unauthorized funds from a credit cardholder's account."

-Bangko Sentral ng Pilipinas (BSP)

Objectives

1. Analyze Fraud Behavior in Eskwebanks Credit Card Transactions

2. Provide a Fraud Detection System



CUSTOMER DATA

Customer Details:

SS Number, Credit Card Number, Name, Gender, Address, and Profile

TRANSACTION DATA

Customer Details:

SS Number, Credit Card Number, Name, Gender, and Address

Transaction Details:

Transaction Number, Transaction Date, Product Category, Merchant, and Merchant location

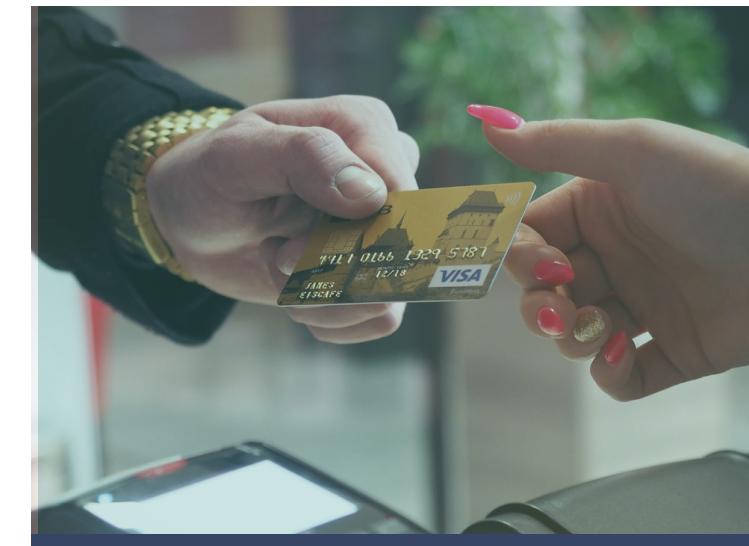
1. Analyze Fraud Behavior in Eskwebanks Credit Card Transactions



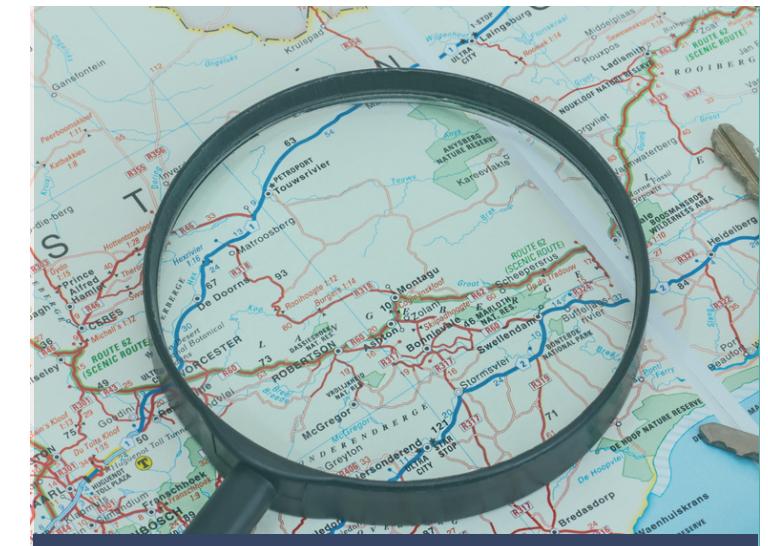
TRENDS



PRODUCTS



MERCHANTS



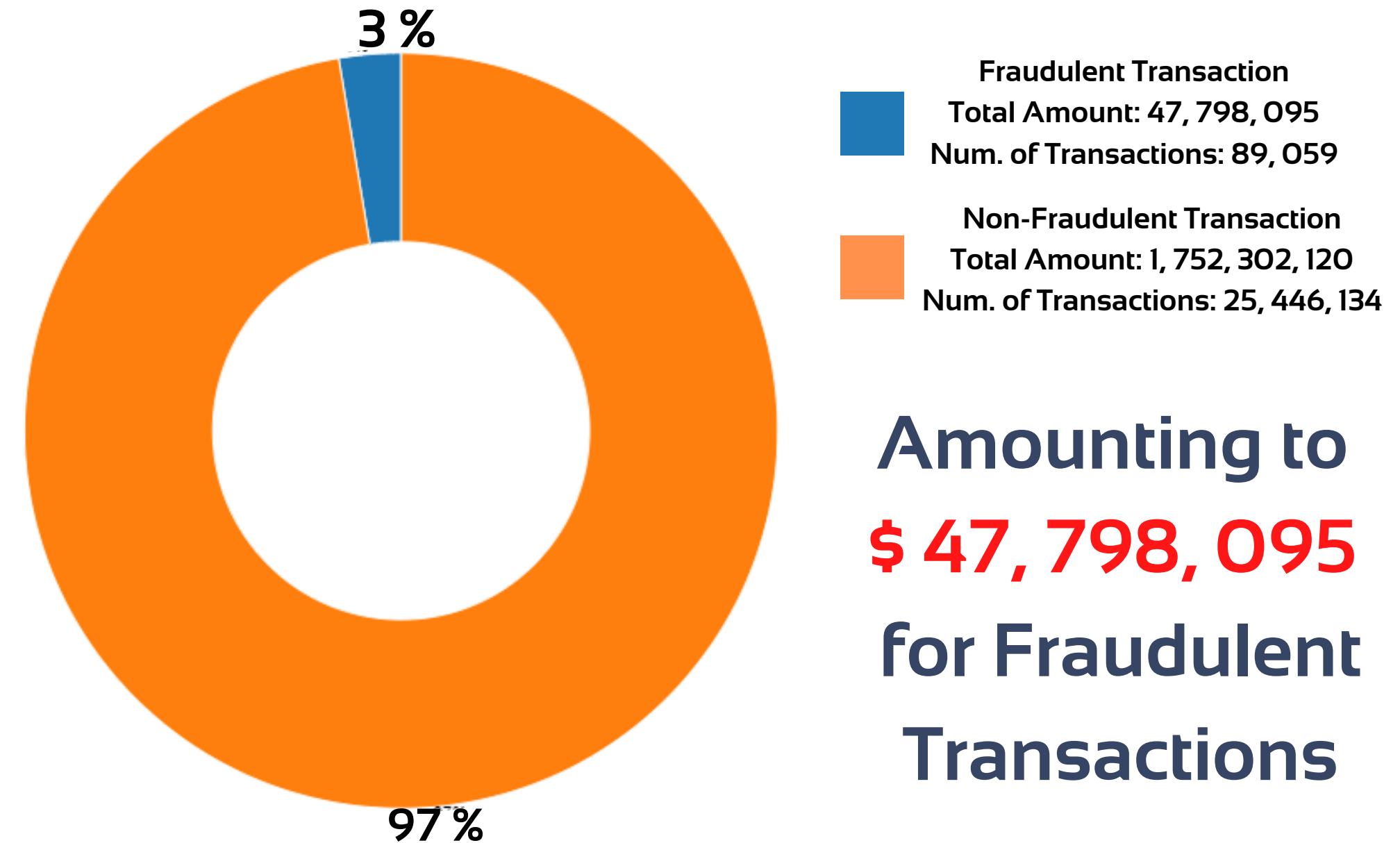
LOCATION



% of Fraud Transactions

**Out of all the transactions,
0.35% is a Fraudulent
Transaction**

Total Amount of the Credit Card Transactions

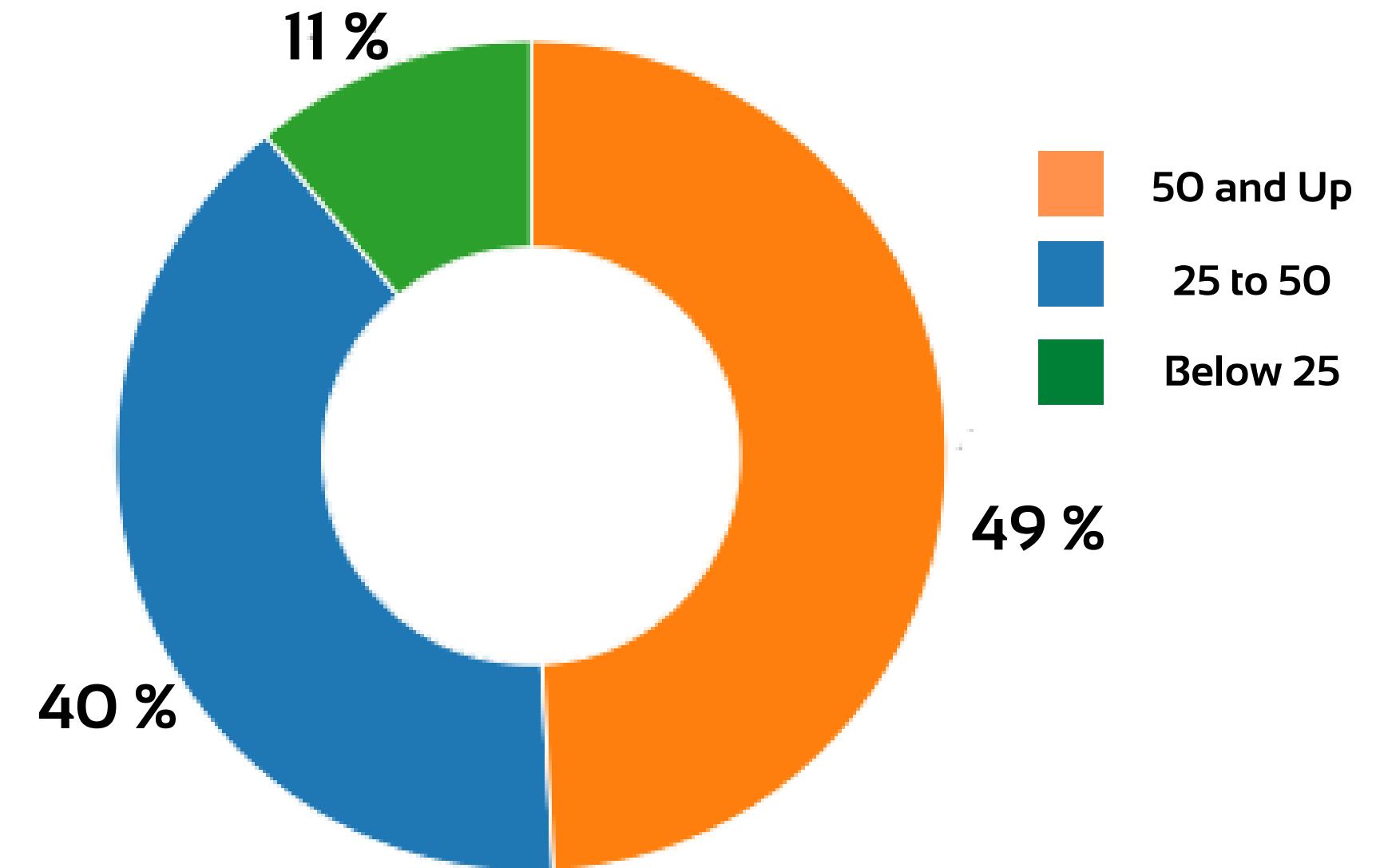




Affected Accounts

**98 % of your customers
have experienced fraud**

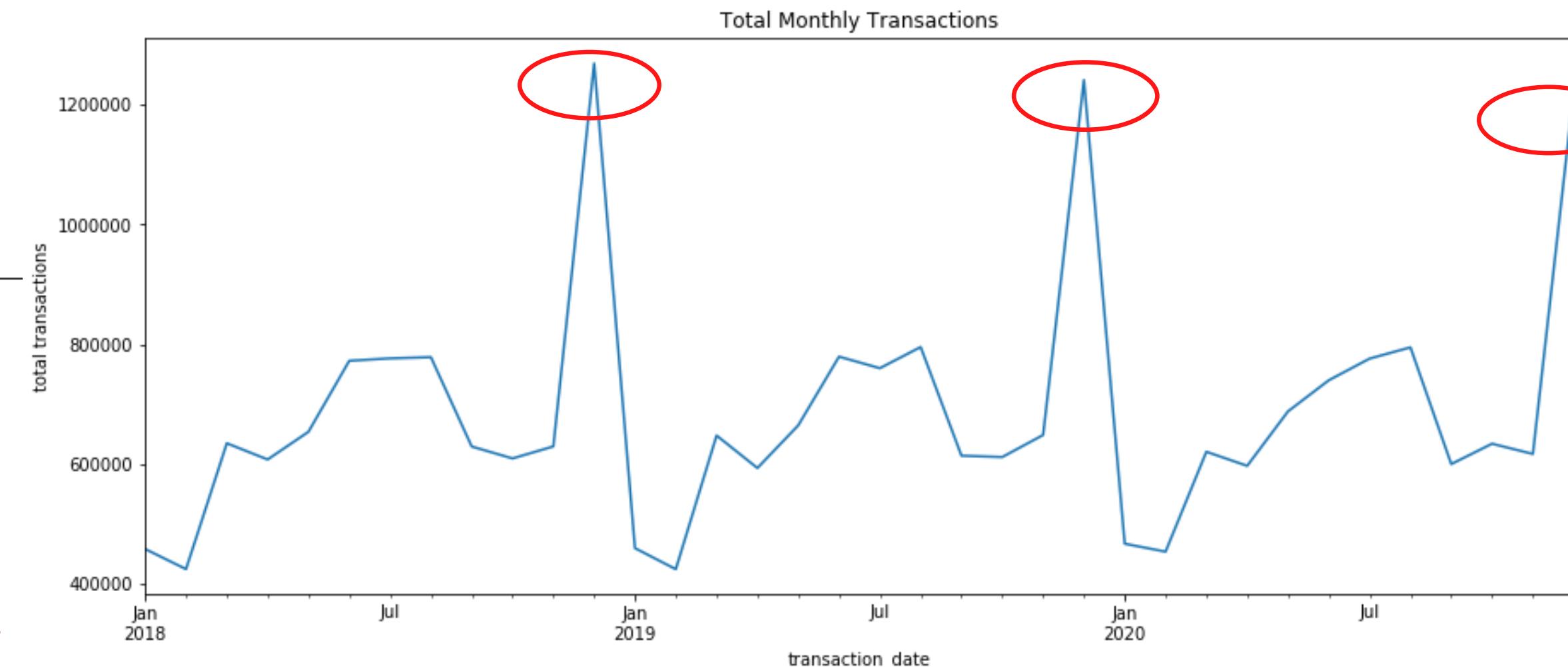
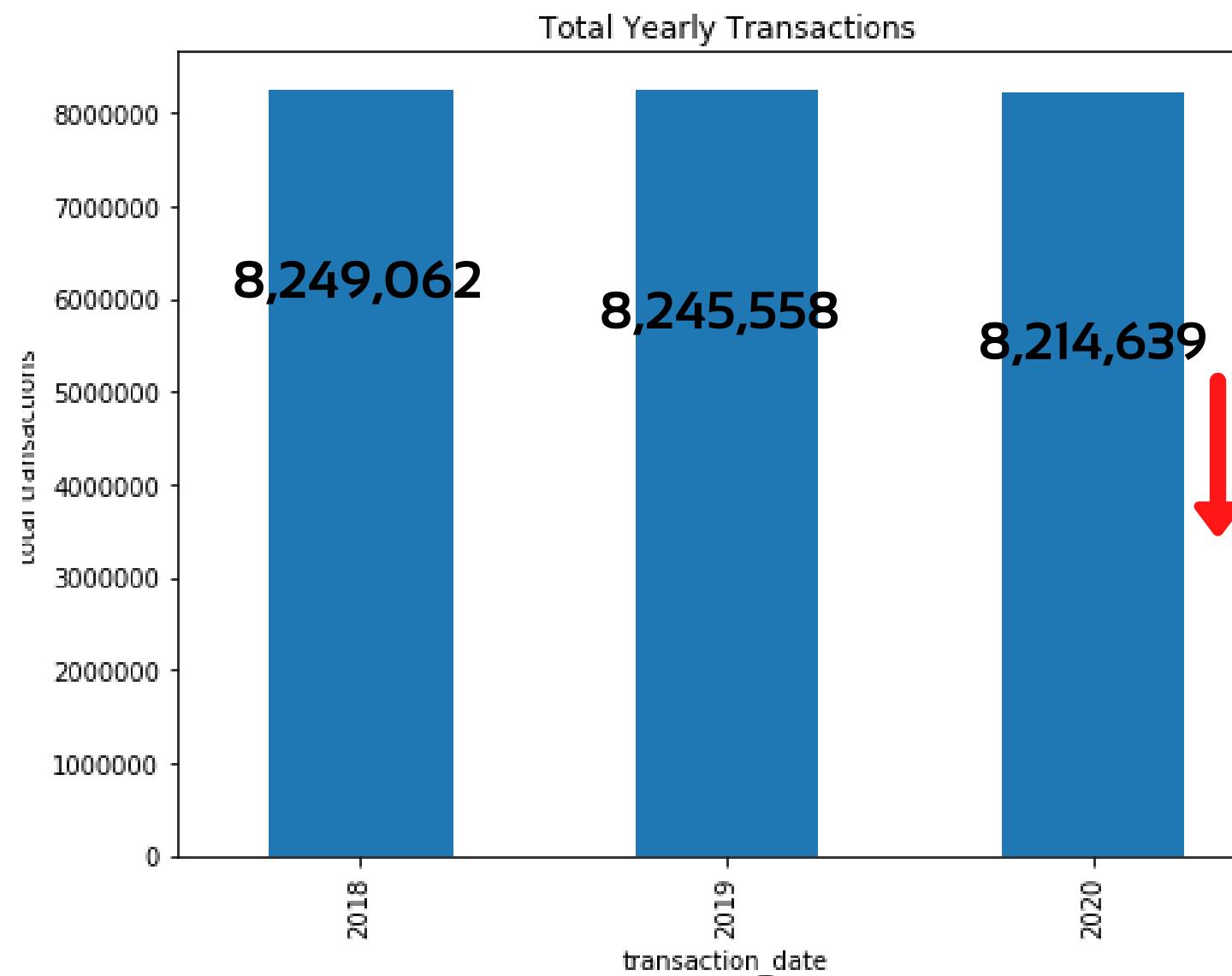
Age profile of customers





TRENDS

Credit card transactions have decreased annually for the past 3 years

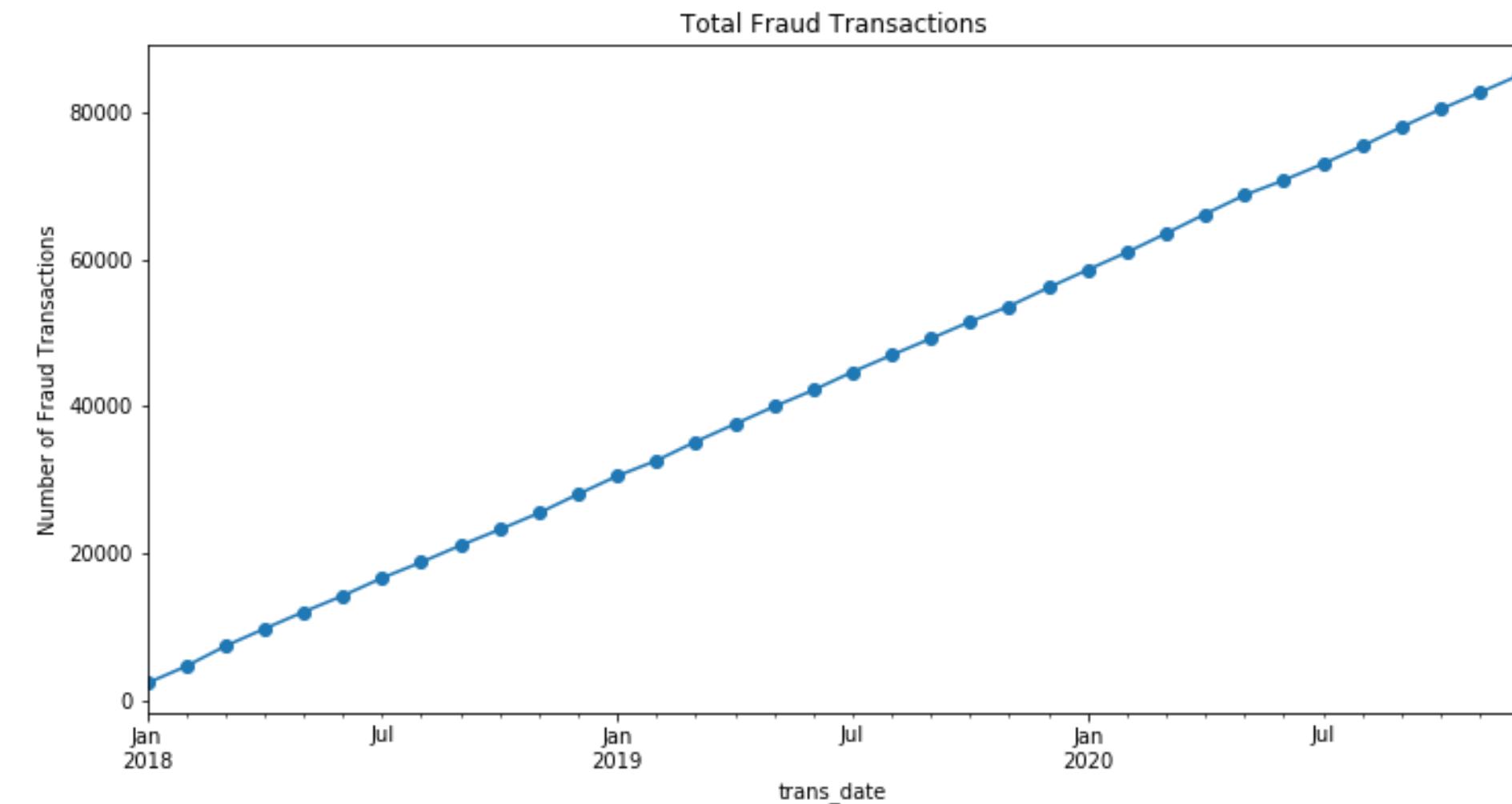
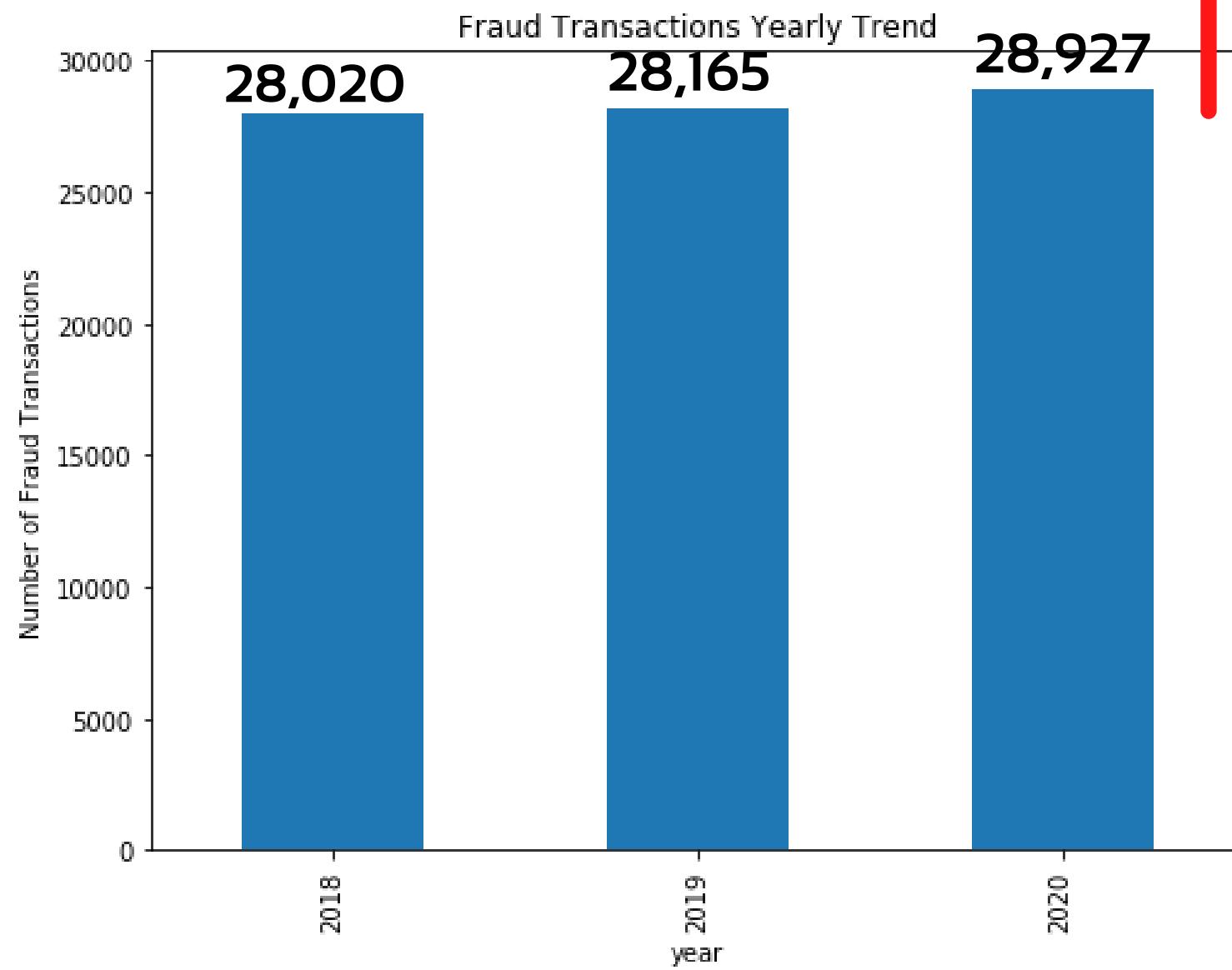


Total number of credit card transactions peak during the Christmas holiday (December)



TRENDS

However, despite the decrease in total credit card transactions, **fraud transactions been increasing** for the past 3 years



We can also see that fraud transactions display a **linear growth** from Jan 2018 to Dec 2020

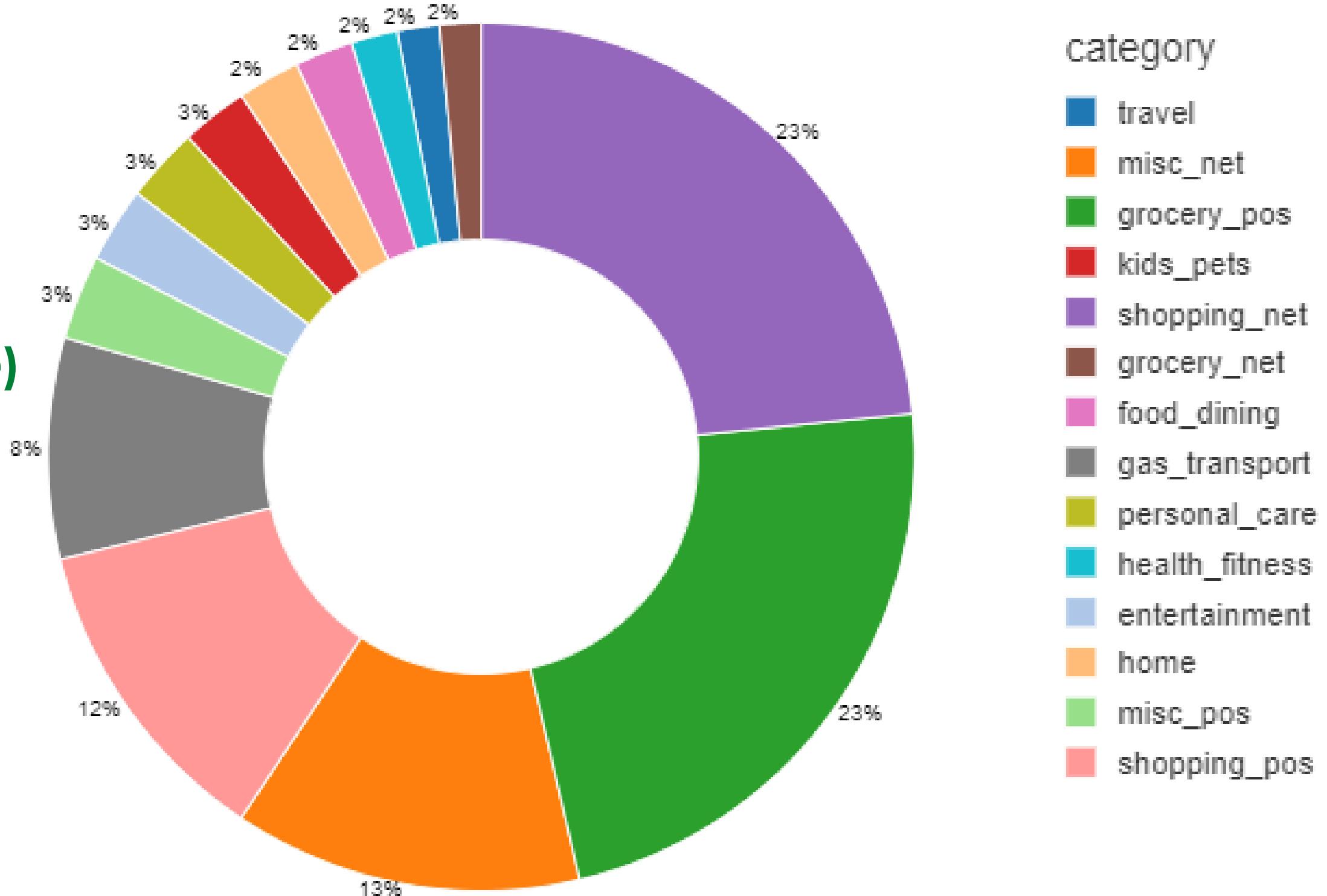


PRODUCTS

Highest affected in terms of number of transactions:

Online Shopping
Grocery Shopping (in-store)

Transactions affected by Fraud





PRODUCTS

Transactions affected by Fraud

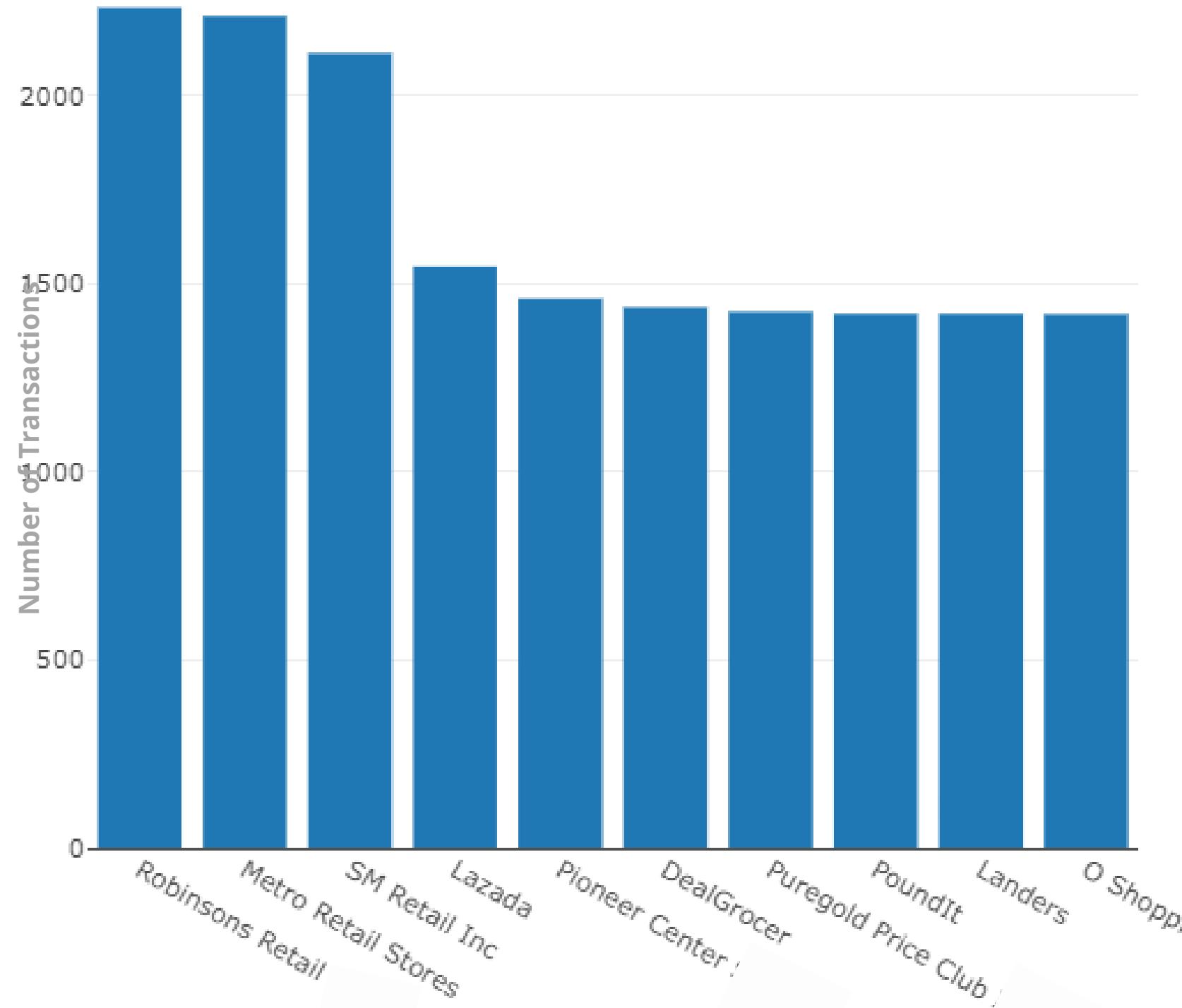


Highest affected in terms of amount:
Online Shopping amounting
with **\$ 20.82 Million**



MERCHANTS

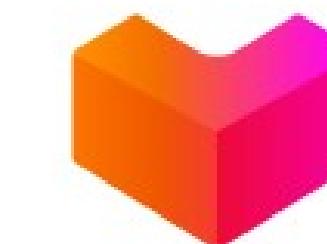
Merchants affected by Fraud



Retail Stores



Online Stores



Deal Grocer



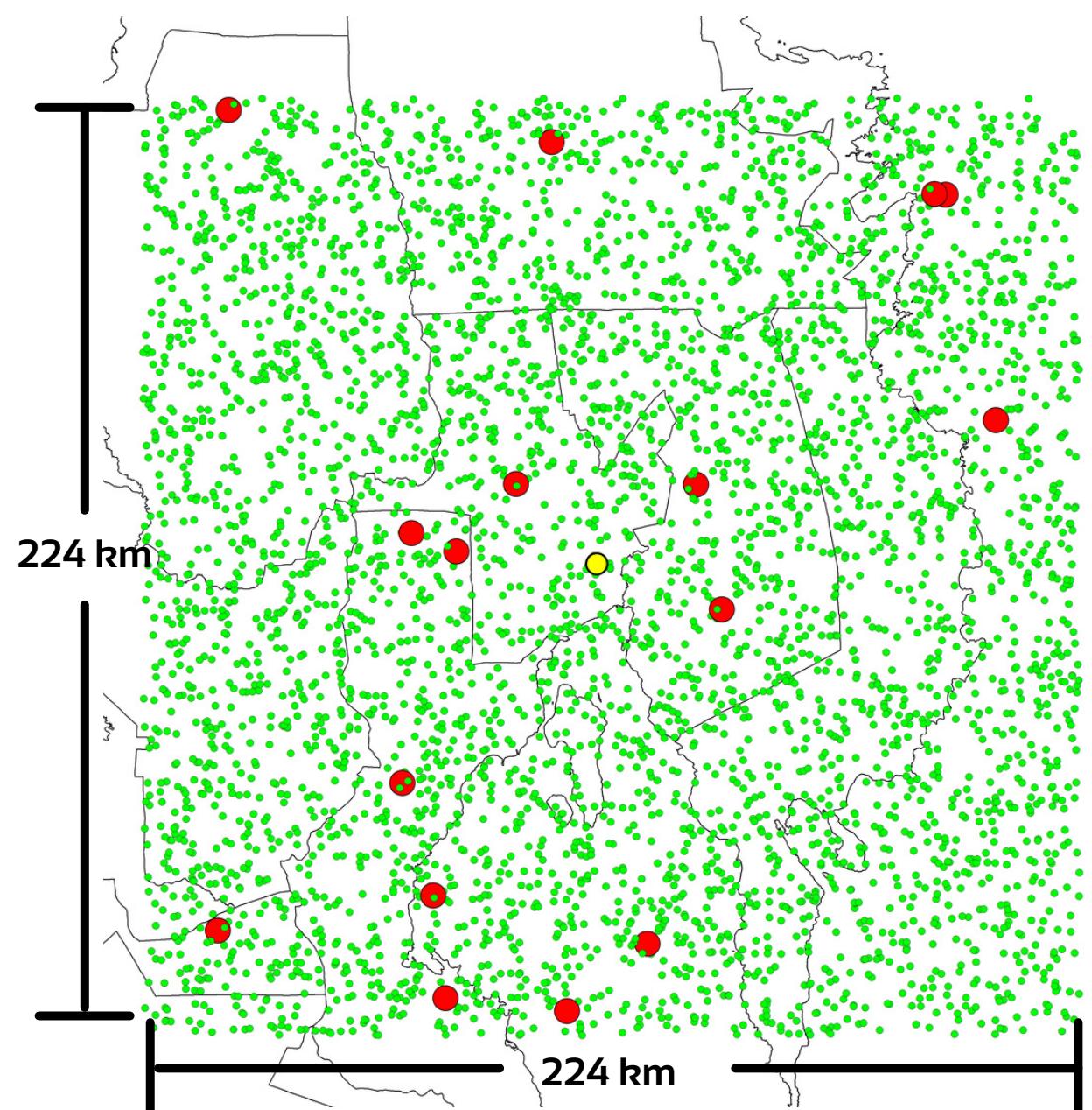
Lazada





LOCATION

Sample Transactions of a User



- Home of user
- Fraud transaction
- Non - Fraud transaction

User transactions are
within 224 x 224 km



Geographic distribution of
Fraud Transactions based
on Merchant location

Dense clustering in
Luzon Area
**(mainly with many
existing cities)**

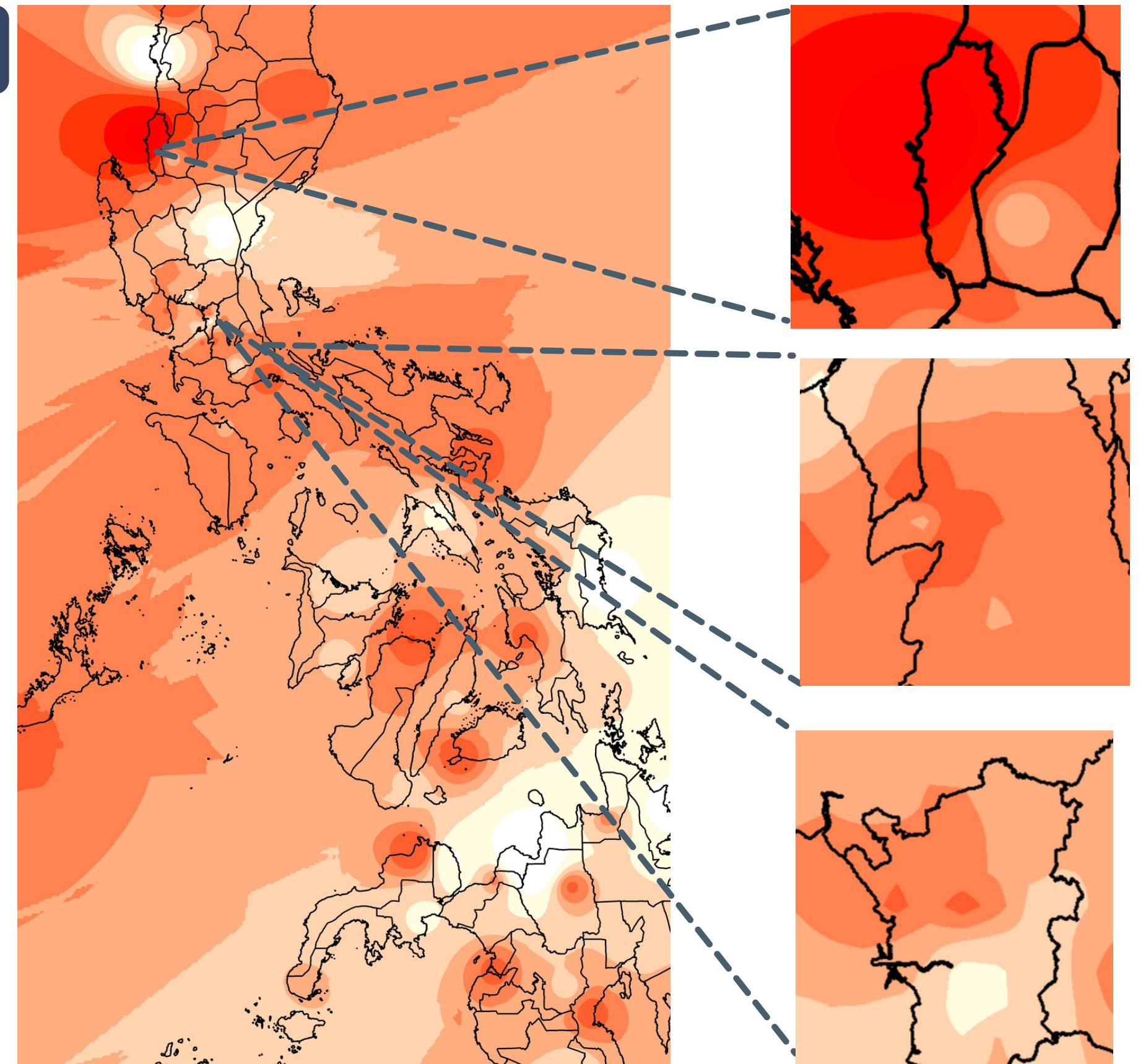


LOCATION

Cities affected
by Fraud

Legend

Sum Amount of Fraudulent Transactions	
932,127 - 1,281,096	
717,759 - 932,126	
613,068 - 717,758	
573,185 - 613,067	
538,288 - 573,184	
468,494 - 538,287	
323,921 - 468,493	
9,846 - 323,920	



San Fernando City in La Union

Sum Amount of Fraud: 1,282,851

Num. of Users: 250

Num. of Transaction: 2,393

Biñan City in Laguna

Sum Amount of Fraud: 704,264

Num. of Users: 136

Num. of Transaction: 1,295

Quezon City in Metro Manila

Sum Amount of Fraud: 693,673

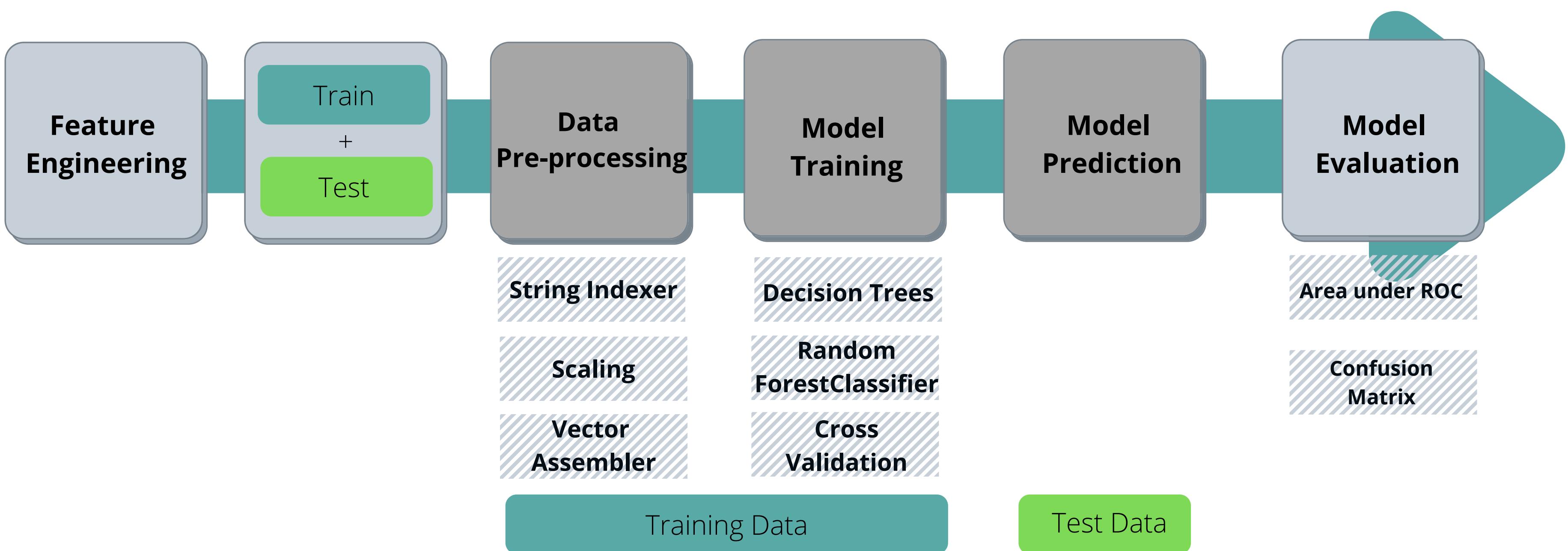
Num. of Users: 133

Num. of Transaction: 1,272



2. Provide a Fraud Detection System

ML Pipeline





2. Provide a Fraud Detection System

Unbalanced Dataset

Model: *Decision Trees*
Area under ROC: **0.9381**

Precision (1) : **1.0**
Recall (1) : **0.49**

	precision	recall	f1-score	support
0	1.00	1.00	1.00	5090529
1	1.00	0.49	0.66	17576
accuracy			1.00	5108105
macro avg	1.00	0.75	0.83	5108105
weighted avg	1.00	1.00	1.00	5108105

Balanced Dataset

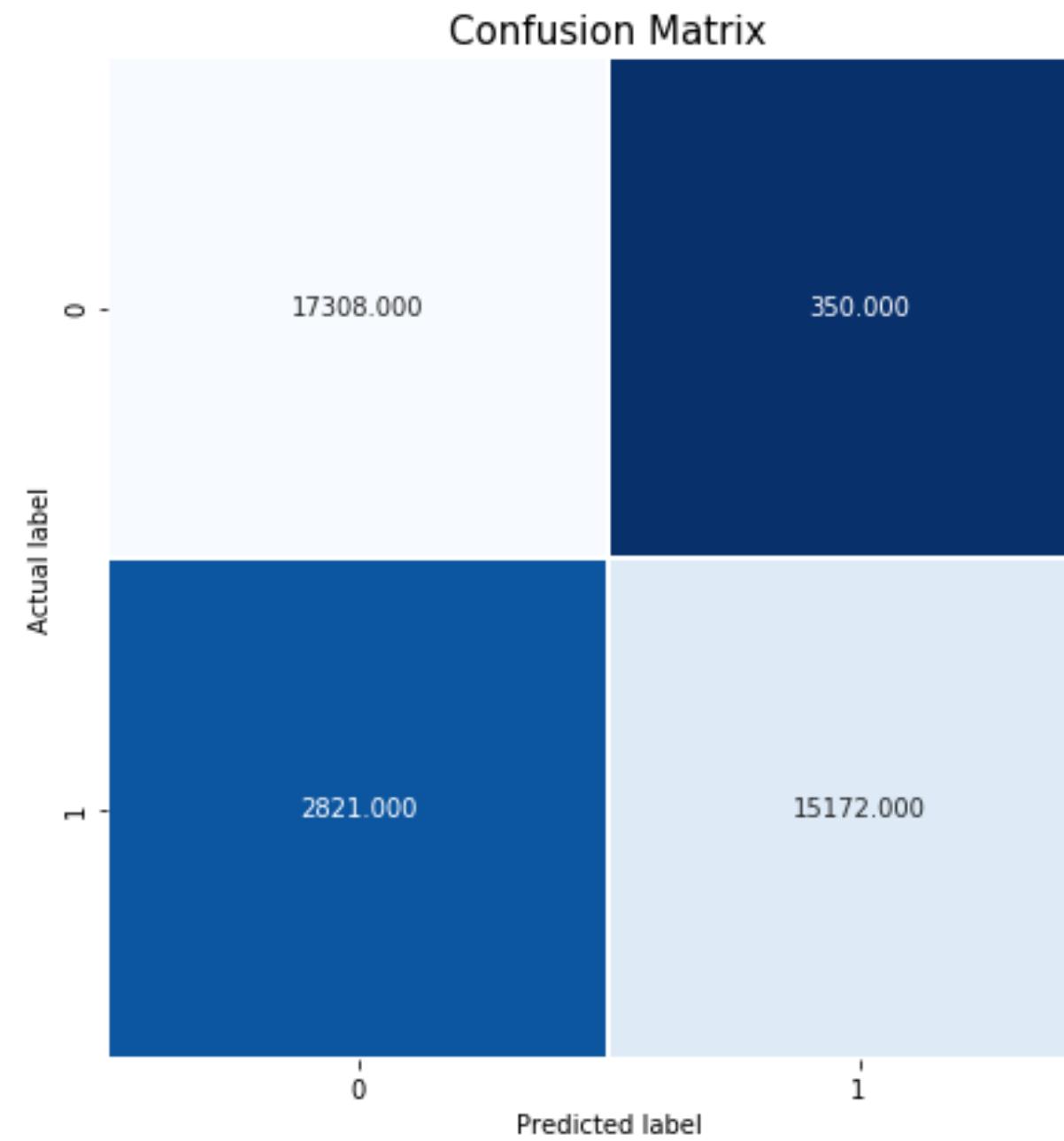
Model: *Random Forest*
Area under ROC: **0.98707**

Precision (1) : **0.98**
Recall (1) : **0.84**

	precision	recall	f1-score	support
0	0.86	0.98	0.92	17658
1	0.98	0.84	0.91	17993
accuracy			0.91	35651
macro avg	0.92	0.91	0.91	35651
weighted avg	0.92	0.91	0.91	35651



2. Provide a Fraud Detection System



Best Model Results

	precision	recall	f1-score	support
0	0.86	0.98	0.92	17658
1	0.98	0.84	0.91	17993
accuracy			0.91	35651
macro avg	0.92	0.91	0.91	35651
weighted avg	0.92	0.91	0.91	35651

Feature Importances

Amt: 0.68	Merchant: 0.12
Age_Group: 0.02	Category : 0.16
Job: 0.06	City: 0.002
Haversine_Dist: 0.0	



Conclusion

Random Forest Classification was the best performing model

Precision: 98%

9.8 out of 10 transactions is correctly classified as fraud

Recall: 84%

8.4 out of 10 fraud transactions is detected by the model

Recommendations

Explore if a customer's spending capacity and previous transactions can be used as features

Try out more models and other feature combinations

Test this model for deployment using streaming data



ESKWEBANKS

The future of work is data. Eskwebanks creates opportunities for people and teams to thrive in the future of work through data skills in banking.