NULLFRAUD

Stopping Fraud in Its Tracks

Presented by: Unbeatabolt


Newgen Bao


Kyla Freed


Amy Guan

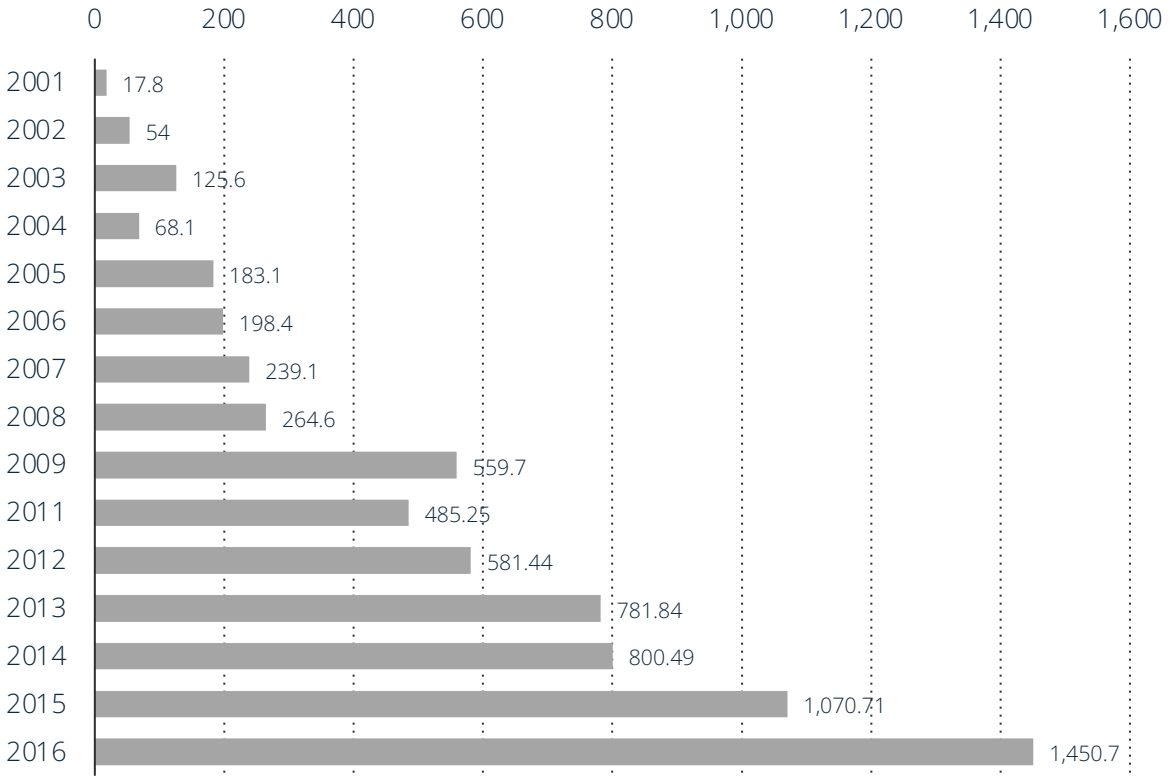

Kezia Rijadi

NullFraud Bank wants to transform how fraud is fought…

NullFraud

To address the need of secure, sustainable, and efficient payment solutions…

Annual amount of financial damage caused by reported cybercrime in U.S. 2001-2016



Source(s): eMarketer; ID 379046

…NullFraud has three key missions in mind.

…making innovation the champion of NullFraud's mission

NullFraud

Question

How can NullFraud reduce fraud, cut down on false positives, and be seen as a pioneer for secure transactions?

Challenges

- Card Not Present Fraud Occurring
- Unreliable risk assessment metric
- Lack of customer-oriented controls

Solutions

STRATEGY: The SPArk of fraud innovation

- **S**cheme Tokenization
- **P**redictive Binary model for internal controls
- **A**I fraud flagging for customers

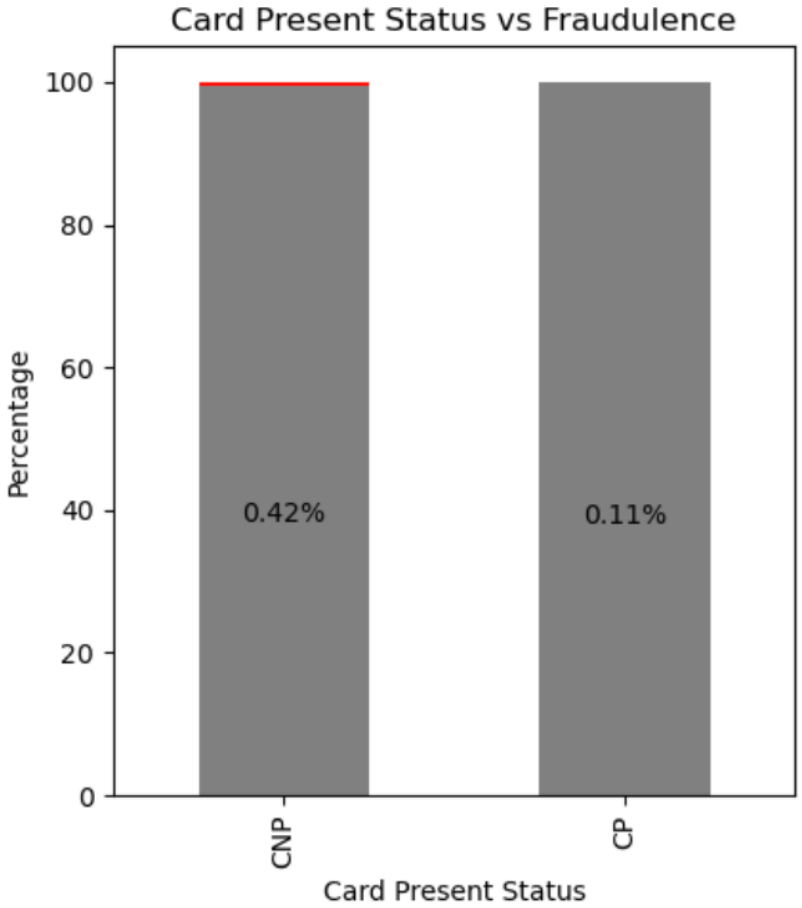
Impact

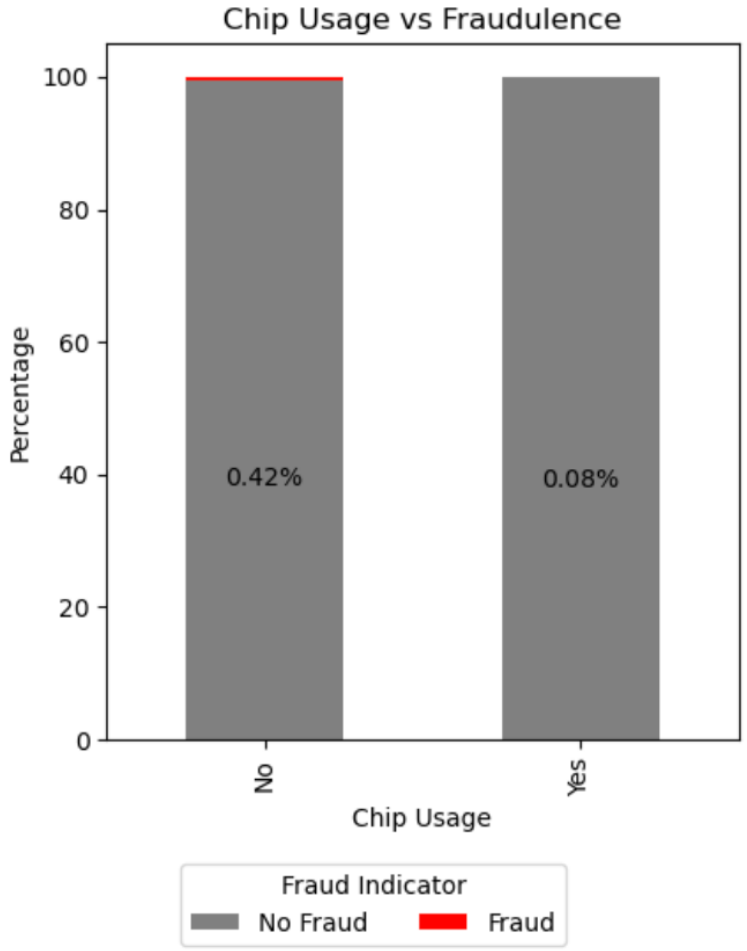
- Protect the valuable customers and the reputation of NullFraud
- Saving NullFraud $12,000 collectively per 100,000 transactions
- High fraud detection while maintaining customer satisfaction

Fraud is impacted by several key factors

NullFraud

Fraudulence is more prevalent when transactions are with no chip and the card is not present…





Which is further confirmed with Chi-Square tests

Chi-Square Test for Independence

Card Present Status - p-value: 5.383848817728829e-21
Chip Usage - p-value: 2.1958704289262313e-24
Cross-border Transaction (Yes/No) - p-value: 6.5051876113018414e-21
Risk Assessment - p-value: 0.0
Payment Method - p-value: 2.0386422879988157e-21
Transaction Value - p-value: 0.0
Merchant Location - p-value: 0.0

Interpretation: All factors, including Card Present Status and Chip Usage, have small p values, making it **statistically significant in predicting fraud**

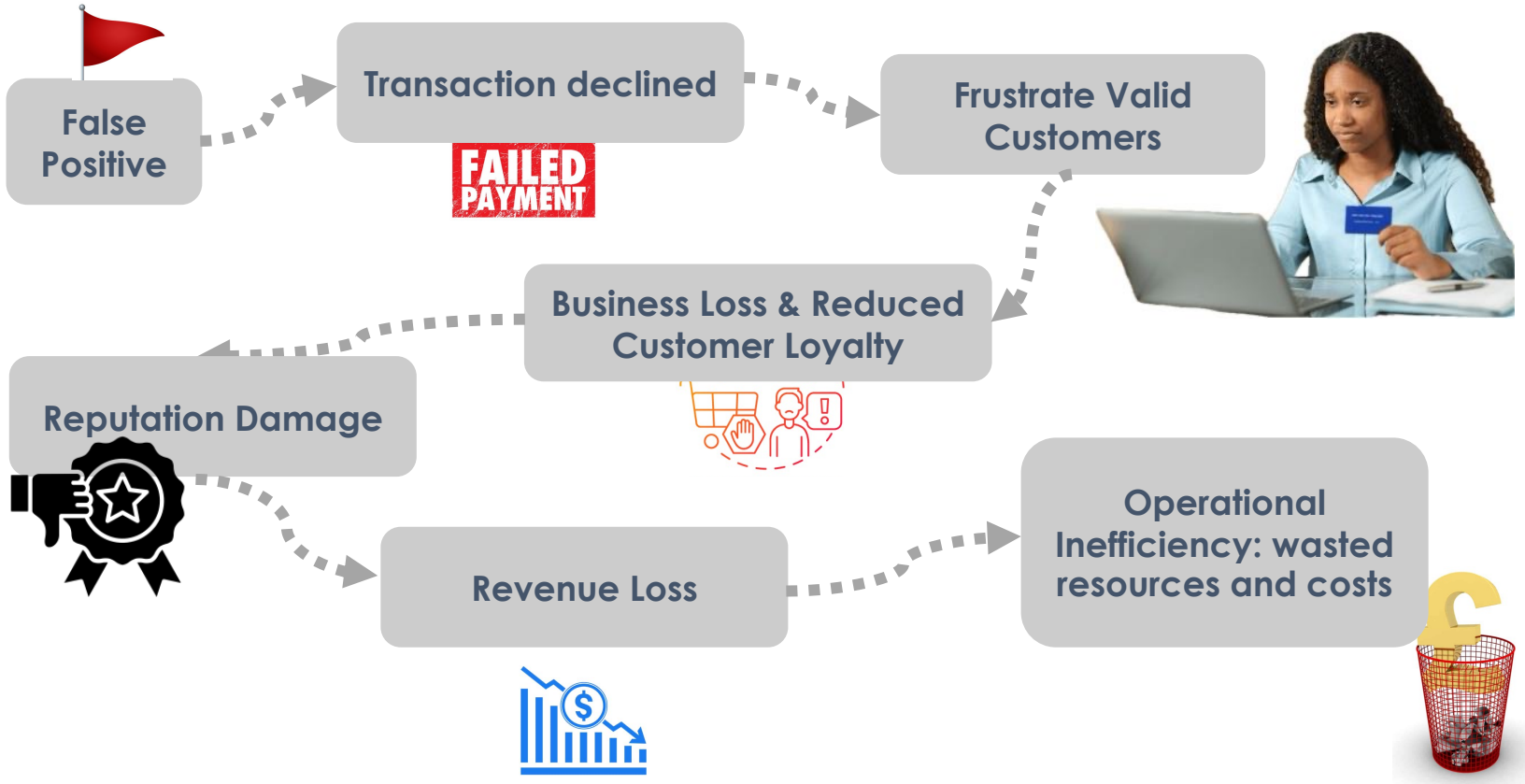
False Positives... Why does it matter?

NullFraud

What causes false positives?

Loose rules in flagging Fraud leads to False Positives

What is the impact of having false positives in fraud detection?

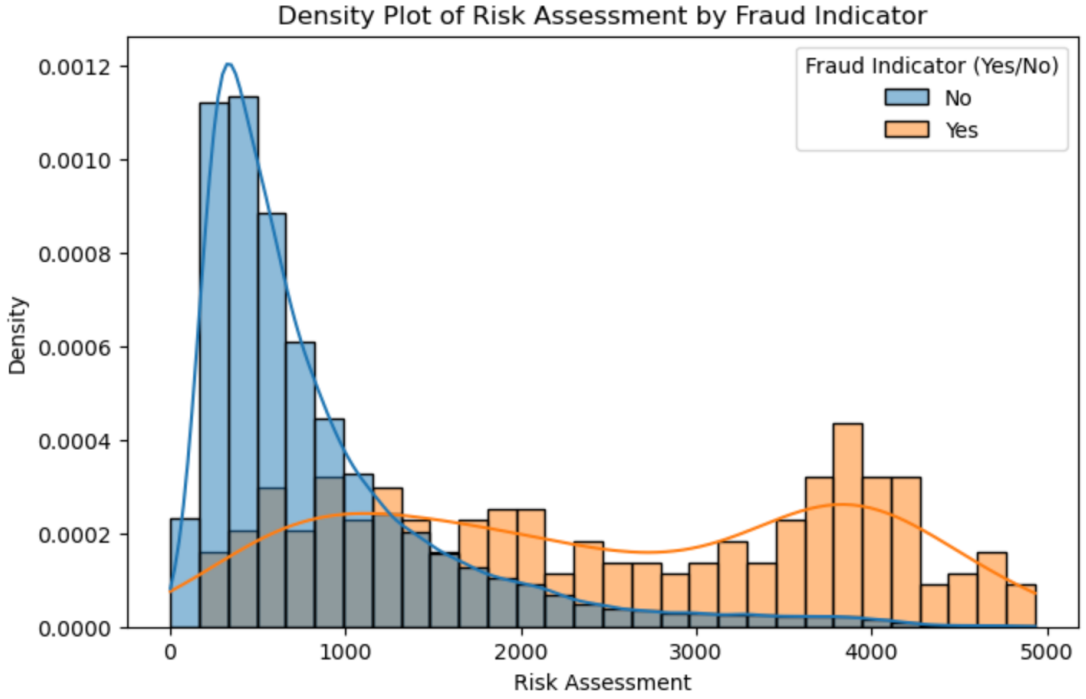


While actual fraud losses represent an estimated 7% of the total cost of fraud, false positive losses amount to 19%!

Is Risk Assessment the best Fraud Indicator?

NullFraud

Sticking to traditional rule-based systems to evaluate modern fraud detection is not sufficient.



Risk assessment is not representative of fraudulence.

Risk assessment is not **interpretable**, making it difficult to refine and adapt against evolving fraud threats.

Low

- Transactions approved immediately

Medium

- Flagged for review
- Pending transaction

High

- Immediately flagged for fraud
- Suspended account

HIGHER RISK ASSESSMENT

Maintaining reputation is a multilayered task

NullFraud

Why Its Important

(58%) Of consumers rank trust as one of the most important factors in choosing a financial services provider

(53%) Of consumers rank how securely their personal data is protected as one of the most important factors in choosing a financial services provider

(67%) Of consumers who suffered fraud are willing to switch their bank or credit union.

(90%) Of consumers are concerned about the potential of banking or credit fraud as both become more digital

The key customers

How to maintain customers

Communicate trustworthiness to all customers

Prevent fraud from occurring while **maintaining customer friendliness**

If fraud does occur, what can NullFraud offer to **retain customers** and promote loyalty?

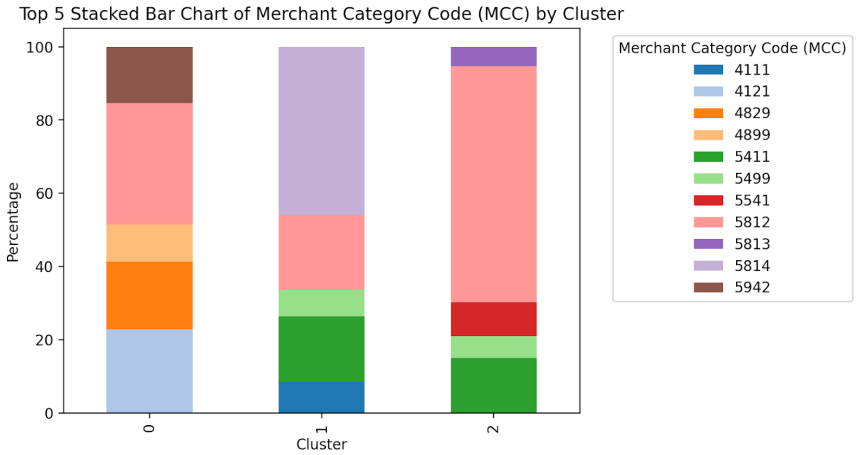
A trust-building solution must address all three layers to maintain a positive customer experience.

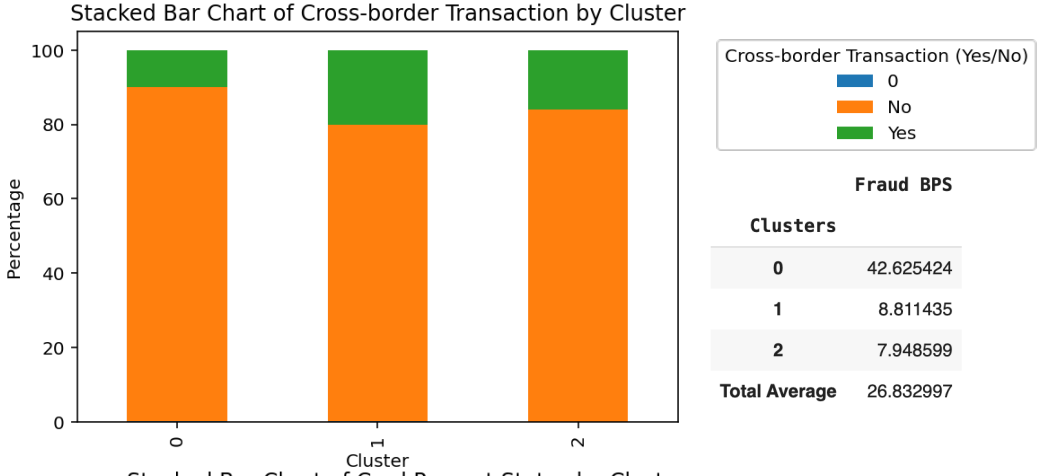
Determining those susceptible to fraud

NullFraud

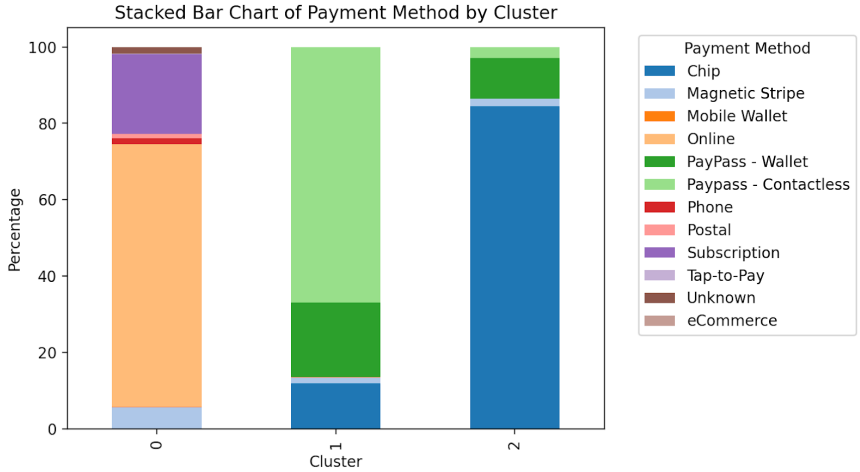
Cluster 0

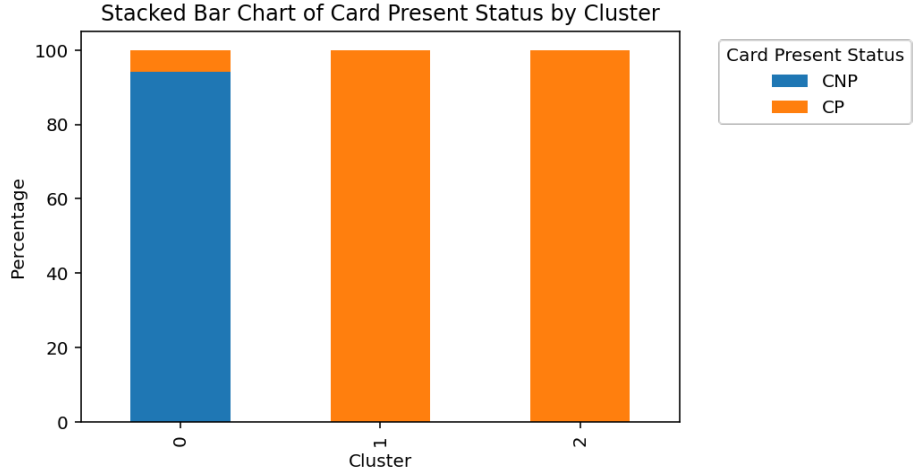
- **High Fraud rates**
- **Not cross-border** transaction
- **Card Not Present**
- **Online** or **subscription** payment
- **No chip** usage
- MCC: **4899** (Cable and other pay television), **4829** (Money Orders – Wire Transfer), and **4121** (Taxicabs and limousines), **5942** (Book Stores and other similar services.)





Clusters	Fraud BPS
0	42.625424
1	8.811435
2	7.948599
Total Average	26.832997





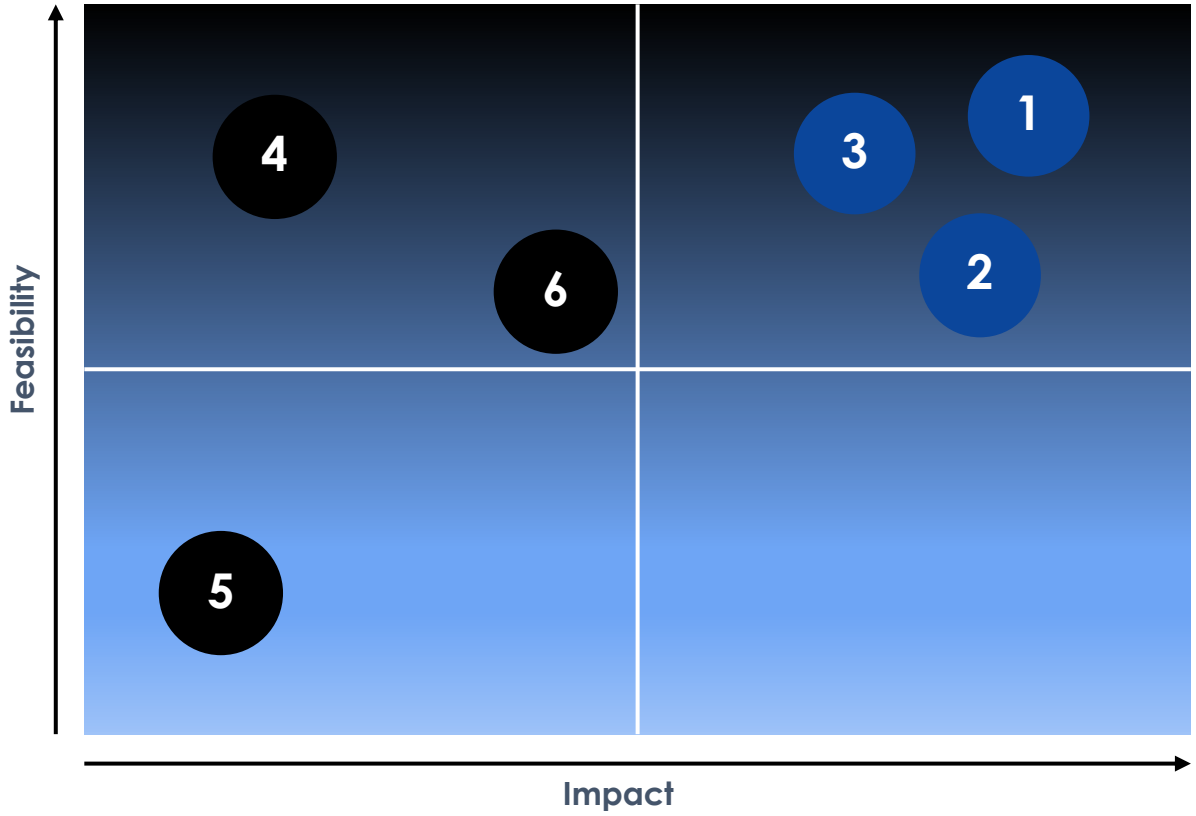
Cluster 0 is a specific segment of transactions much more likely to conduct fraud who must be helped

We conducted a decision matrix to choose an optimal strategy…

NullFraud

CRITERIA CONSIDERED	Measures to avoid CNP fraud	Reducing false positives	Balance between security and customer experience

Analysis of Alternative Strategies

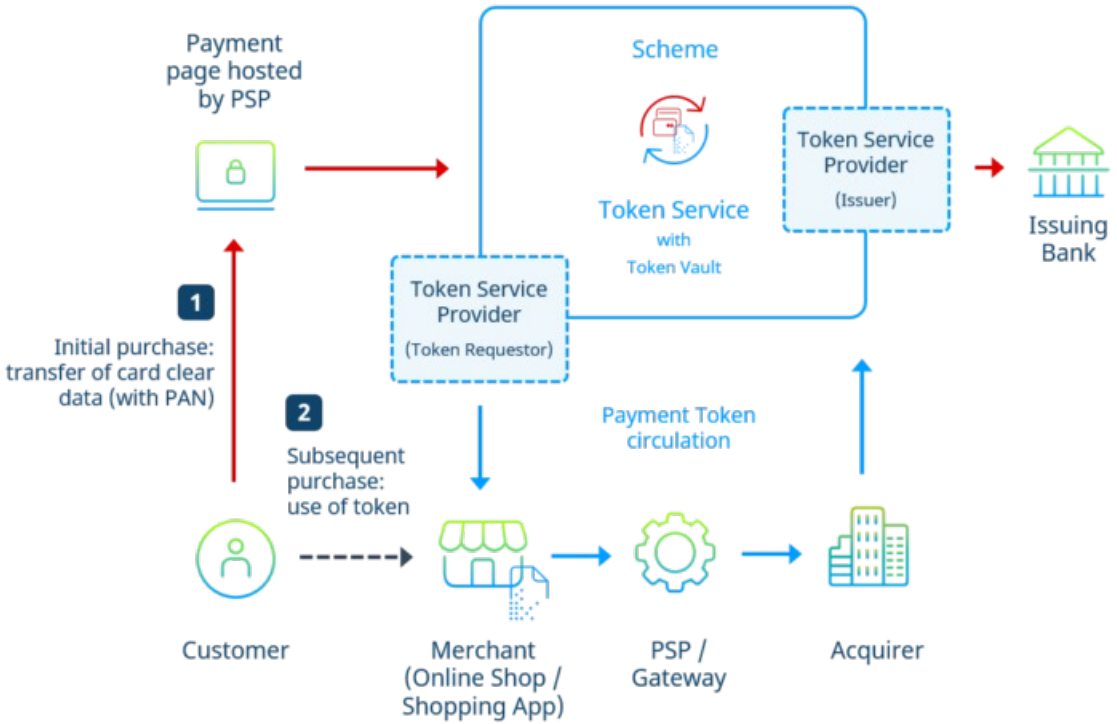


Alternatives Considered

1. Data tokenization
2. Binary predictive model for internal controls
3. Real time fraud flagging for customers
4. Biometrics
5. Consumer Education Programs
6. Geolocation Tracking

Implementing A Scheme Token

NullFraud



What is Tokenization?

A security measure in payment transactions, replaces sensitive data like credit card information with non-sensitive tokens, with scheme tokenization managed by the issuing bank being a focus.

What is Scheme Tokenization?

Scheme tokenization is crucial in mitigating CNP fraud by enabling the card company to serve as the token service, managing the token system, and allowing merchants to link tokens to specific domains and end devices, thereby enhancing protection against theft and misuse in online purchases.

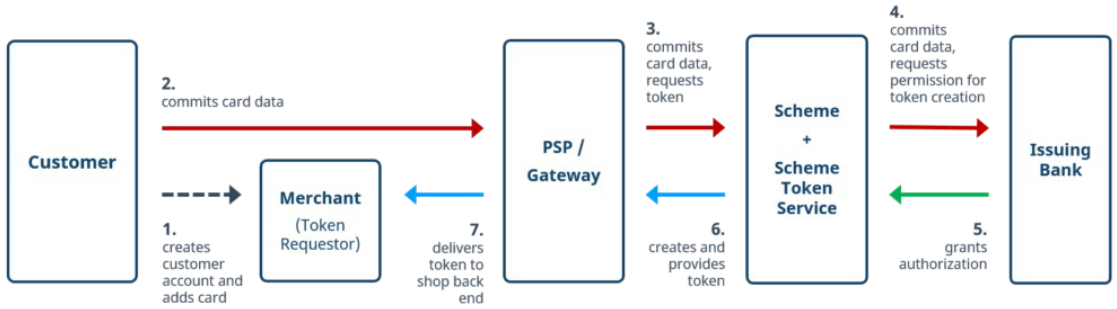
Implementing A Scheme Token

NullFraud

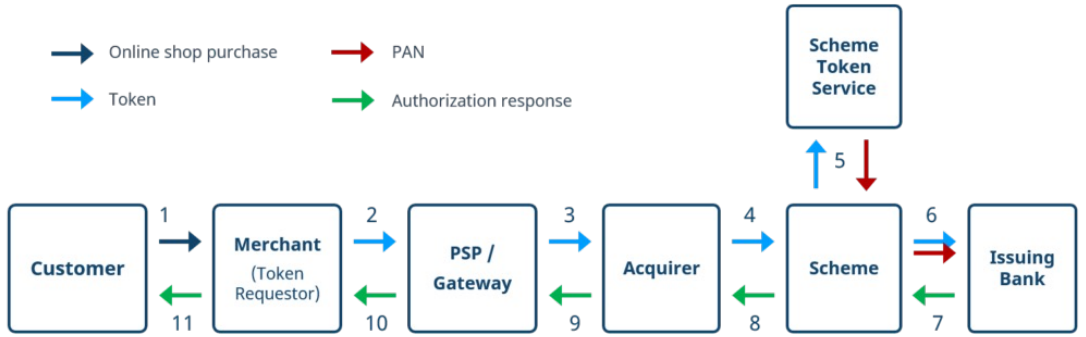
Benefits of Tokenization

- Prevention of data theft during leaks
- Enhances trustworthiness and potentially reduces legal repercussions in case of breaches
- Simplifies data management for merchants
- Allows seamless integration of new payment technologies
- Offers protection against theft and misuse by enabling merchants to link tokens to specific domains and end devices.
- Can be relevant with card present transactions because it can ensure that card information is not stored on POS systems.
- Unlike gateway token systems, the acquirer also only comes into contact with the token. The only points of contact with the PAN are thus the token service provider (as part of the system) and the Scheme and issuer (who have the PAN anyhow).

Creating A Token



Using A Credential On File (COF) Token

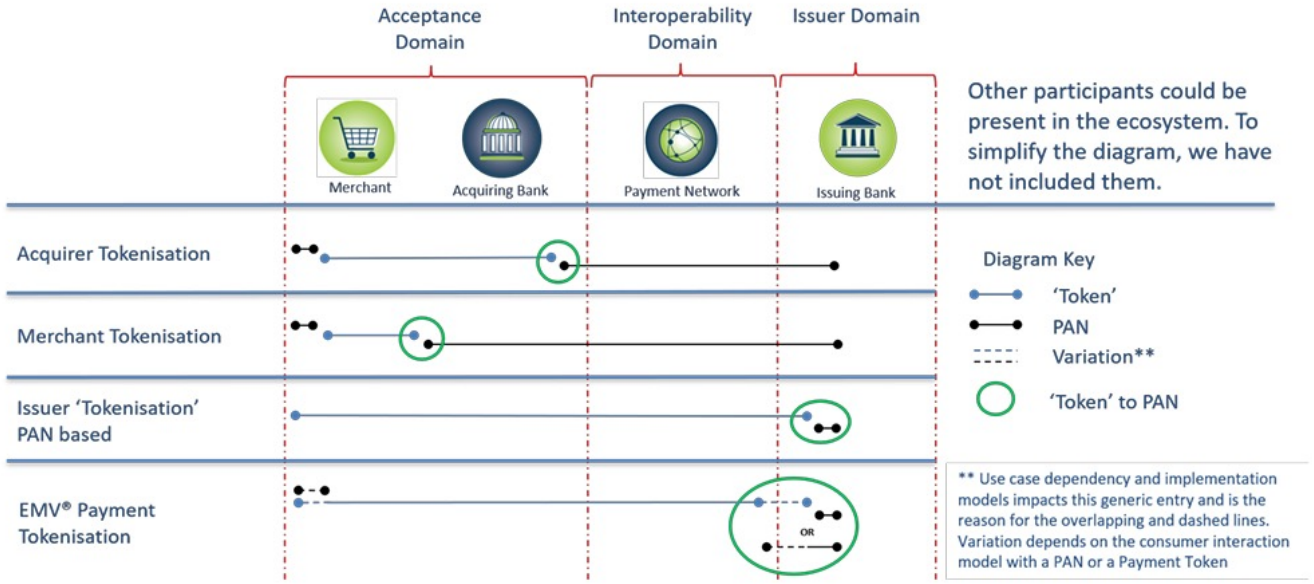


Implementing A Scheme Token

NullFraud

Using EMVCo's EMV Payment Tokenization

EMV Payment Tokenization, a widely trusted product that significantly bolsters security in in-store, e-commerce, and remote payments by substituting the primary account number (PAN) with a unique EMV Payment Token, thereby restricting its usage to specific merchants, devices, or payment scenarios, effectively thwarting fraudsters' access to critical data.



Implementing Binary Classification

NullFraud

1 To enhance fraud detection within the existing system, we recommend implementing binary classification model to complement the limitations of the current risk assessment framework.

Challenge: Using Risk Assessment as the sole indicator creates high reliance on **well-defined rules.**

The results are also not very **representative** or **interpretable**.

Opportunity: Adopting **binary classification** to detect fraud to…

- Reduce False Positives
- Improve interpretability

Binary Classification Methods Performance

Classifier	AUC	AUPRC
Catboost	0.9693	0.8124
Extremely Randomized Trees	0.9348	0.6480
Random Forest	0.9627	0.7859
XGBoost	0.9682	0.7803
Logistic Regression	0.8600	0.2901

Binary classification produces high effectiveness and good balance between TPR and FPR

+

Binary indicator allows for easy interpretation as it directly presents the final decision without the need to define thresholds

AUC: Area Under the Receiver Operating Characteristic Curve
AUPRC: Area Under the Precision-Recall Curve

Determining the Classification Model

NullFraud

2 After experimentation with various binary classification models, the Random Forest Classifier exhibits superior performance and robustness compared to its counterparts.

Starting from the evidence suggested by research, we first adopt Logistic Regression Forward Selection

```
glm(formula = Fraud.Indicator..Yes.No. ~ Risk.Assessment + Acquiring.Institution.ID +
    Payment.Method + Chip.Usage + Card.Present.Status +
Cross.border.Transaction..Yes.No. +
    Merchant.Category.Code..MCC. + Merchant.Identifier, family = "binomial",
    data = data)

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                      8.163e+00  5.399e-01  15.120  < 2e-16 ***
Risk.Assessment                 -5.705e-03  2.612e-04 -21.843  < 2e-16 ***
Acquiring.Institution.ID         9.717e-04  2.389e-04   4.068 4.75e-05 ***
Payment.Method                  -1.026e-01  2.646e-02  -3.878 0.000105 ***
Chip.Usage                       1.415e+00  3.015e-01   4.692 2.70e-06 ***
Card.Present.Status             -1.014e+00  2.782e-01  -3.647 0.000266 ***
Cross.border.Transaction..Yes.No. -3.232e-01  1.414e-01  -2.286 0.022253 *
Merchant.Category.Code..MCC.     -1.403e-03  8.567e-04  -1.638 0.101361
Merchant.Identifier               1.080e-04  7.581e-05   1.424 0.154385
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 3668.7  on 99023  degrees of freedom
Residual deviance: 2980.8  on 99015  degrees of freedom
AIC: 2998.8
```

We decide that the model is not optimal as its coefficients are heavily affected by classifiers, but it highlights specific predictors worthy of further investigation

Random Forest Classifier

Maintains relatively high **precision** & recall.

	Precision	Recall
Not Fraud (0)	0.91	0.91
Fraud (1)	0.91	0.91

↑ Precision

↓ % of FP

- Under sampling performed
- Prioritise **precision > accuracy**
- RF chosen because of categorical variables + **non-linearity**

RF alone is not enough: what can we do to minimize FP while detecting fraud?

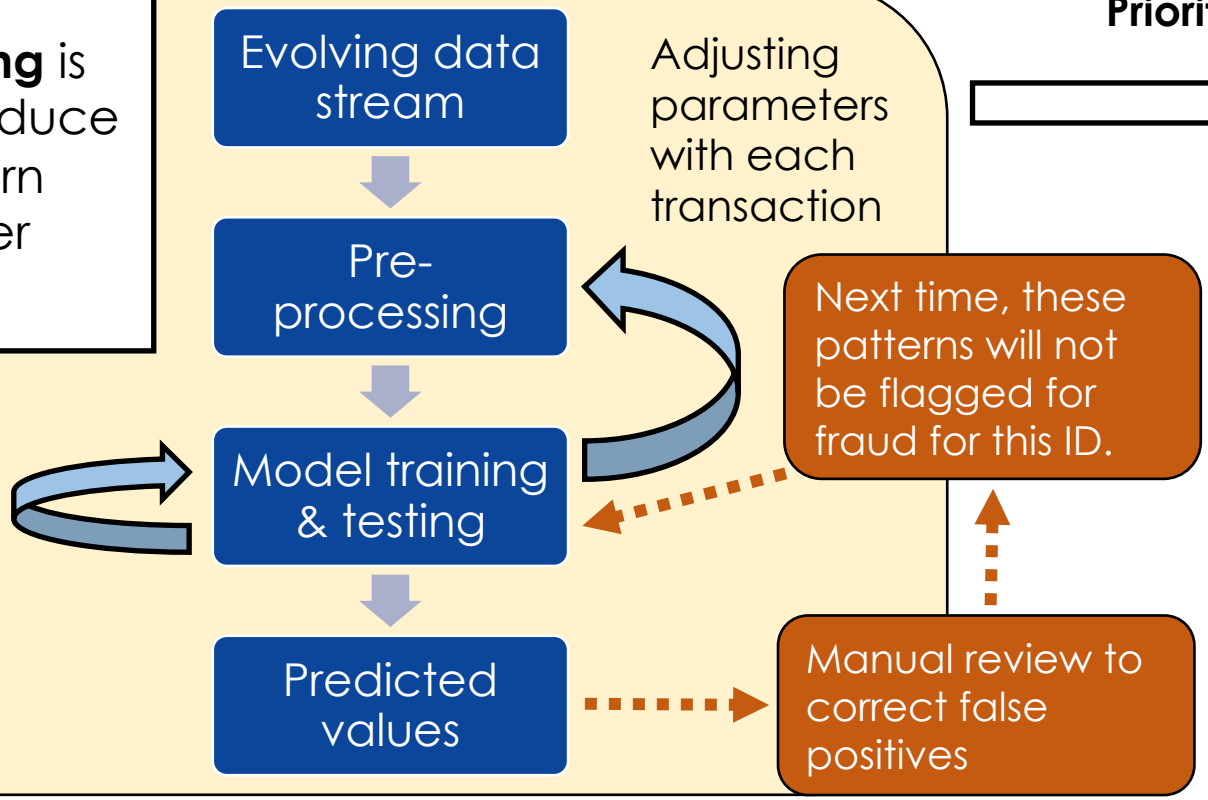
		Predicted: No	Predicted: Yes
Actual	No	48	5
	Yes	5	48

Reducing False Positives

NullFraud

3 Nullfraud maintains retention while curbing fraudulent transactions through implementing incremental learning + a feedback loop.

Incremental learning is implemented to reduce false positives, in turn increasing customer satisfaction.



Priorities for Fraud Detection Architecture

- **Speed + Scalability**
- **Customizable rules**
  - Transaction profiling
  - Testing rule effectiveness
- **Control**
  - Authority management
  - Audit & trace activities
- **Monitoring**
  - Feedback loop
  - Standardized reporting

We maximize resource allocation by developing a model that is constantly up-to-date.

Model Deployment

NullFraud

4 It is crucial to ensure the smooth implementation of the model developed within the financial institution's infrastructure as well as conducting ongoing monitoring and maintenance

- Customization and Configuration
- Data Integration
- Testing and Validation
- Training and Education
- Establishing Protocols
- Monitoring and Maintenance
- Compliance and Regulatory Considerations
- Assessment of Current System

Fine-tuning fraud detection by using not only past data, but also user behaviors.

Historical data

Training initial model based on data already collected

Real-time detection

Robust Fraud Detection

Behavioral Analysis

Anomaly detection

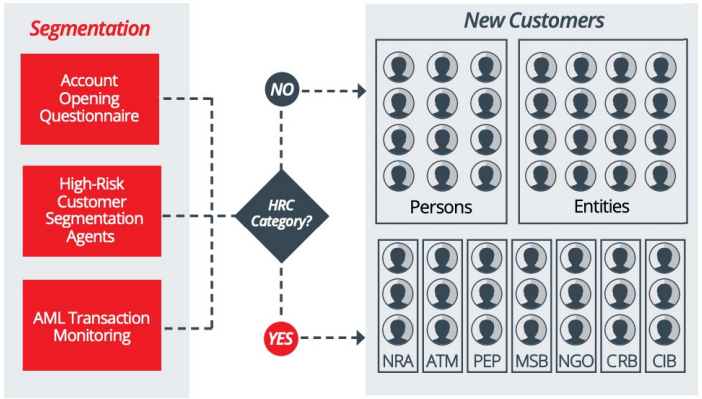
Behavioural patterns: angle at which phone is held, etc.

- Clustering
- Analysis per Card ID

Personalized communication and educational content

NullFraud

Segment high-risk customers automatically based on AI insights

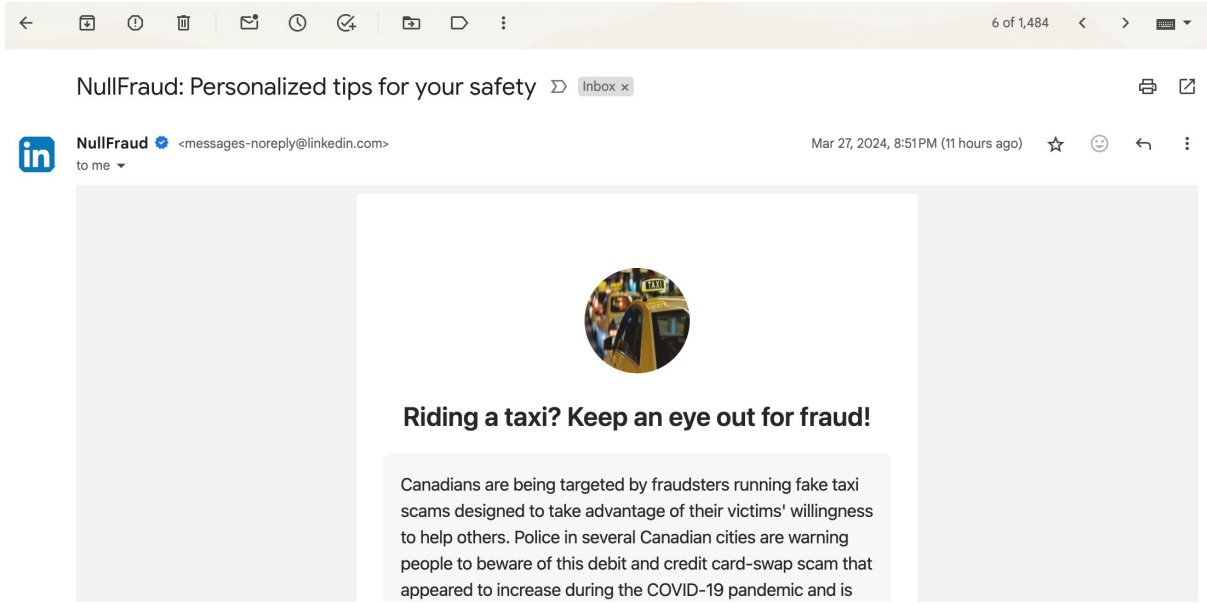


Determine high risk customers using ML and opening questionnaires and segment them in high-risk categories.

Enhanced due diligence reviews to give extra scrutiny to high-risk clients:

- Multiple identification documents
- Establishing the ultimate beneficial owner
- Auditing the beneficial owner's assets
- Performing Know Your Customer checks

Personalizing Anti-fraud Education and Awareness Based on Customer Behavior



Create educational content targeted towards **users conducting higher-risk behaviors.** Ex. If a customer frequently makes online payments to new payees, the bank can **send educational content on the risks associated** with such transactions and how to verify payee authenticity

Real time fraud flagging, giving power to the user

NullFraud

A transaction can be flagged with real-time fraud detection…

Case Study:

NuBank is a fintech online bank with strong roots in Brazil. It launched Intelligent Defenses, "a protection system that is built from ai and understands how each customer behaves, recognizing habits and patterns – when a pattern is broken, it can block the operation."

Source: NuBank

142 million in net profits

87% increase in revenue from previous year

12% decrease of purchases after blocking

…notifying the customers and giving them the ability to fight fraud hands-on

Suzie tries to buy a gift at a fancy store

↓

Her NullFraud card is declined (a false positive)

↓

She is notified through her favorite channel (ex. text), allowing her to confirm that it is her

↓

The channel shares educational content on common fraudulent behaviors, and asks Suzie to evaluate the situation

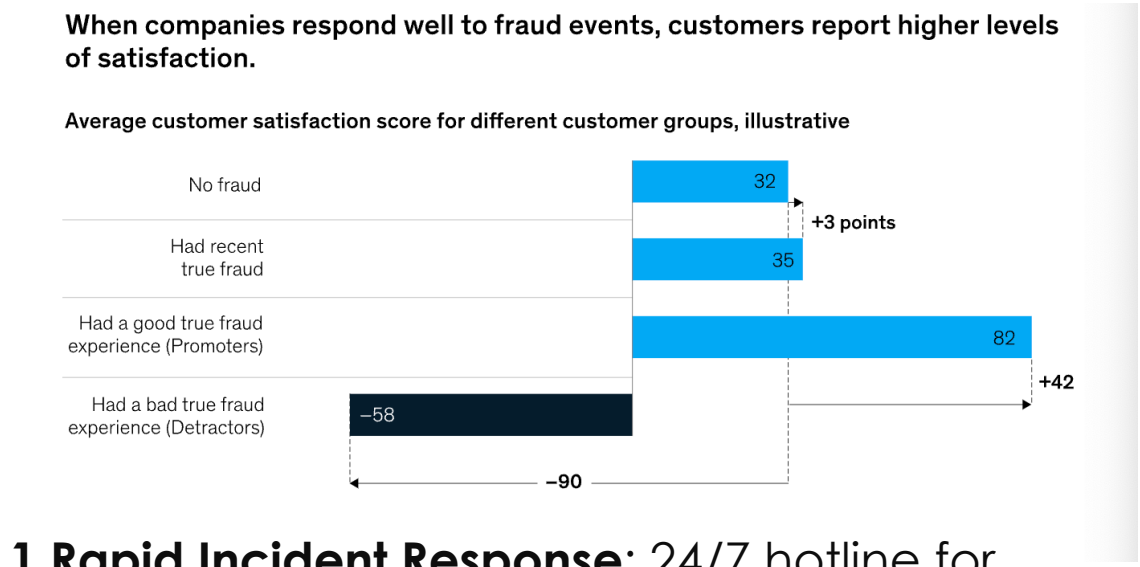
↓

Suzie proceeds, confirms, and finishes their purchase without an issue

Victims of fraud can be taken care of to ensure loyalty

NullFraud

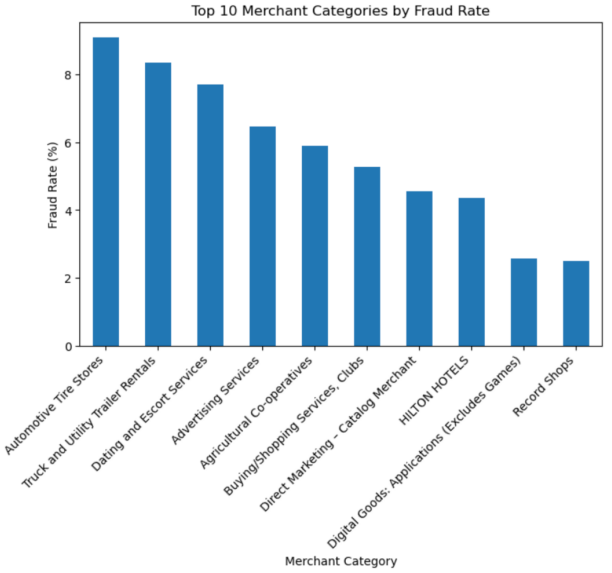
If fraud still does occur, A proper response will improve customer satisfaction

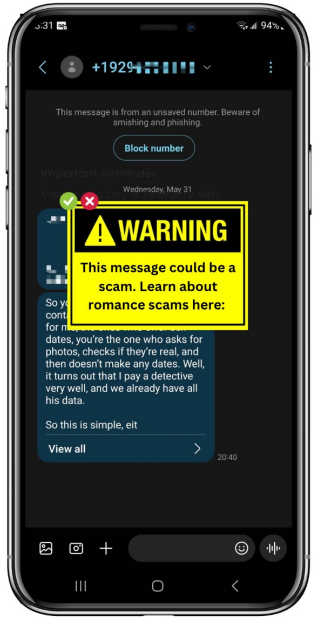


1. **Rapid Incident Response**: 24/7 hotline for customers to report suspected fraud, Quick Freezing of compromised accounts.
2. **Customer Communication**: Inform affected customers promptly and transparently about any security incidents
3. **Financial Reimbursement**: Clear policy for compensating customers for losses due to fraud

Additional controls can be given to consumers who are particularly vulnerable

After the incident, customers must be educated on scams and given "necessary friction" control measures.

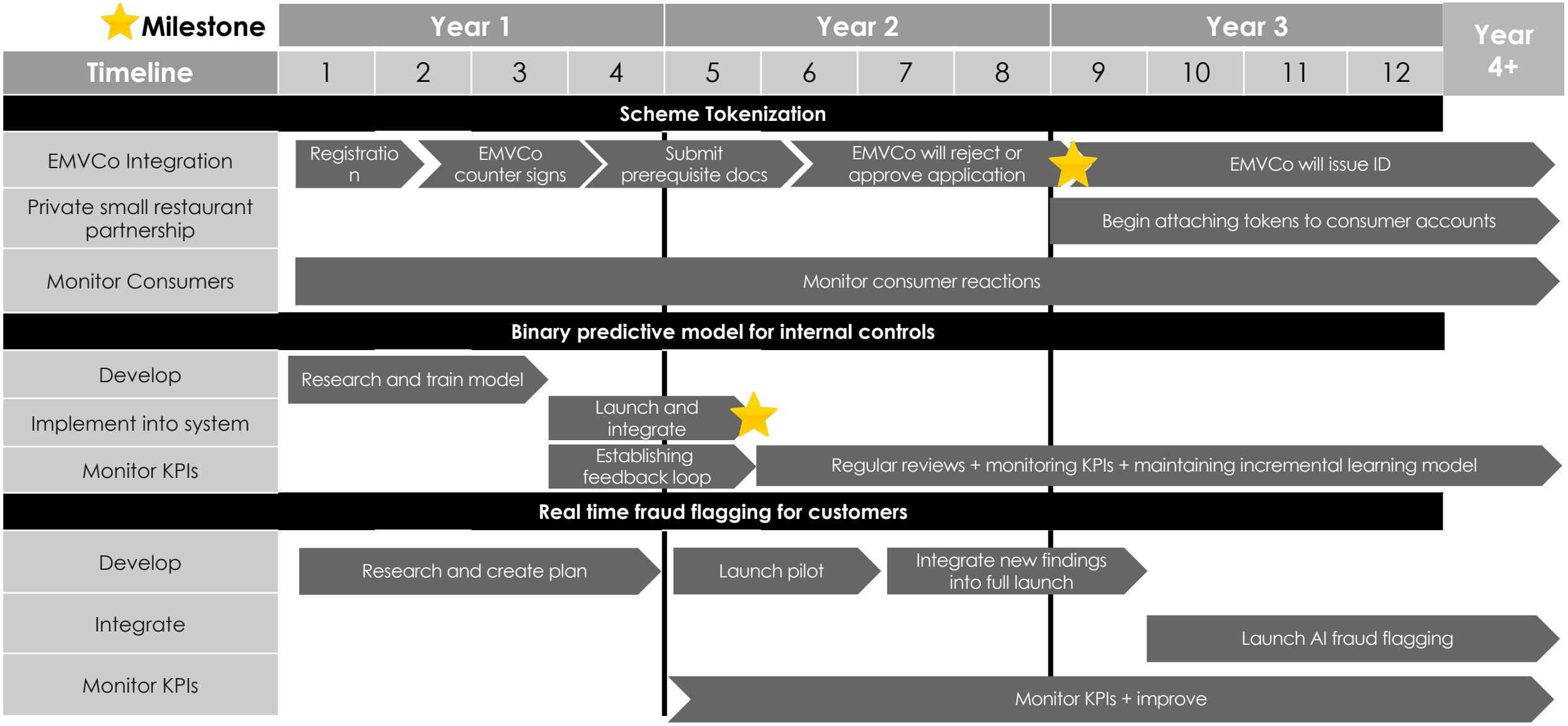




Dating and escort services can have payment prompts to prevent purchase scams on platforms can prevent scams occurring in common areas

The next 3 years for “OUR” future growth, together

NullFraud



Introduction | Analysis | Solutions | **Implementation** | Impact

Key KPIs to track

NullFraud

1.Scheme Tokenization:

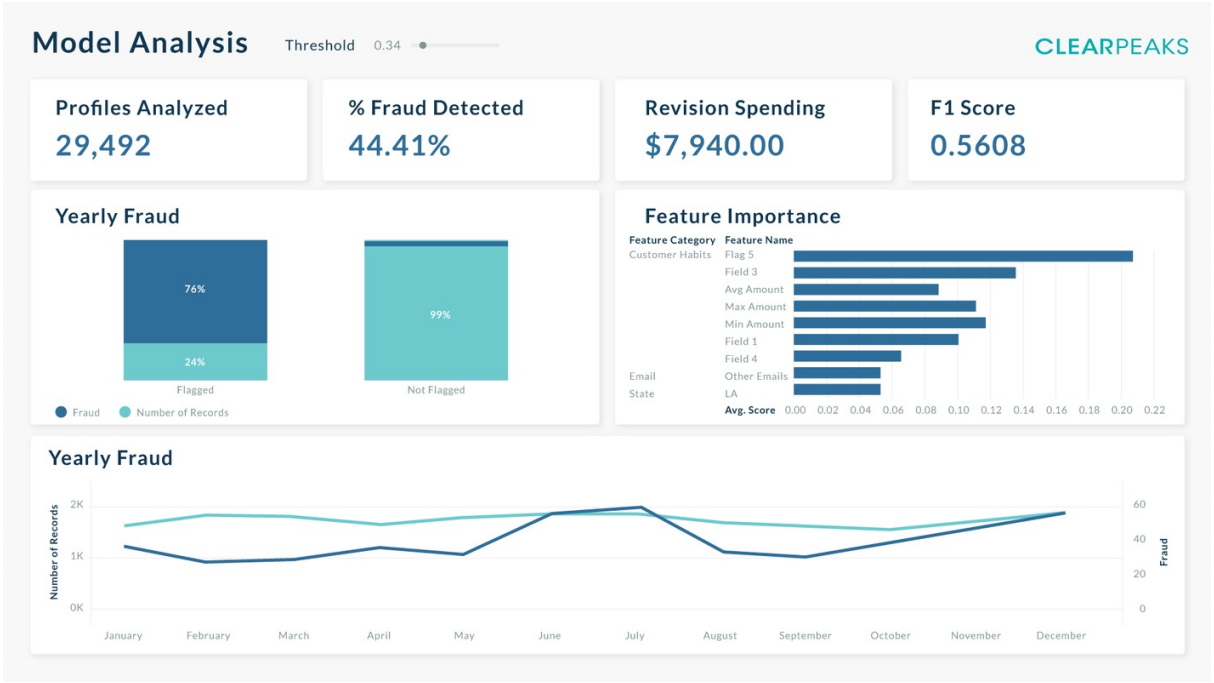
1.Tokenization Adoption Rate:
2. Cost per detection
3.Customer Enrollment in Tokenization
4.Reduction in Card-not-Present (CNP) Fraud:

2.Binary predictive model for internal controls:

- Precision score: % of false positives
- Cost per detection: operational efficiency
- Model drift: monitoring changes that suggest retraining

3.Real time fraud flagging for customers:

1.Customer alert response time
2.Customer resolution satisfaction rate



Summary of Impact

NullFraud

After the implementation, NullFraud will obtain...

Sustainability Impact

Reduce the waste of resources of fraud detection and create sustainable growth of NullFraud business

Cost Impact

Saving NullFraud **$12,000** collectively per 100,000 transactions

Customer Protection

Protect the valuable customers and the reputation of NullFraud with accurate detection

We can create huge impact through continuous development and innovation!

Bivariate Analysis of Merchants

NullFraud

Bivariate Analysis

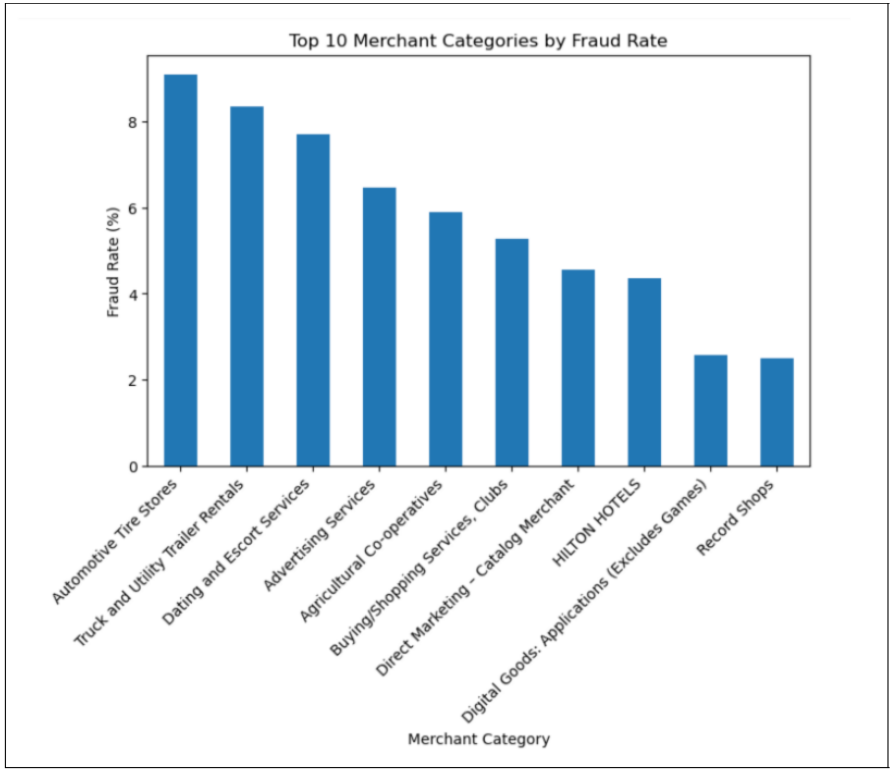
How does each feature correlate to the fraud indicator?

Merchant Category

Merchant Category	
HO JO INN, HOWARD JOHNSON	100.000000
Roofing – Contractors, Sheet Metal Work – Contractors, Siding – Contractors	50.000000
COMFORT INNS	50.000000
LOEWS HOTELS	33.333333
FARIFIELD INN	28.571429
HOMEWOOD SUITES	25.000000
Automotive Tire Stores	9.090909
Truck and Utility Trailer Rentals	8.333333
Dating and Escort Services	7.692308
Advertising Services	6.462585

dtype: float64

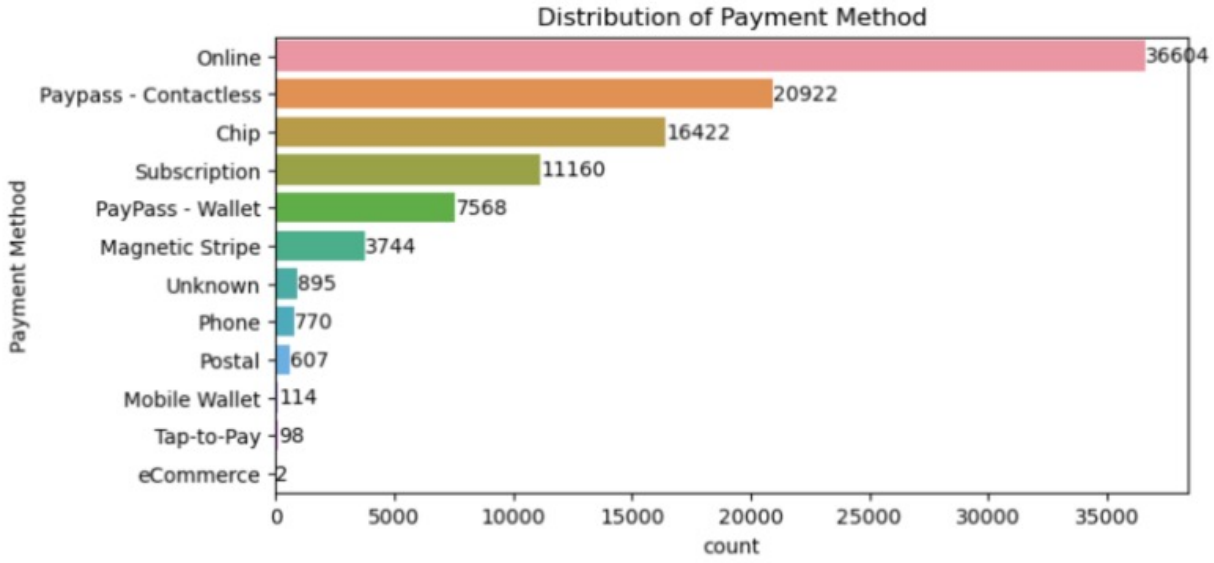
Susceptible to fraud: Inns, hotels, and services.

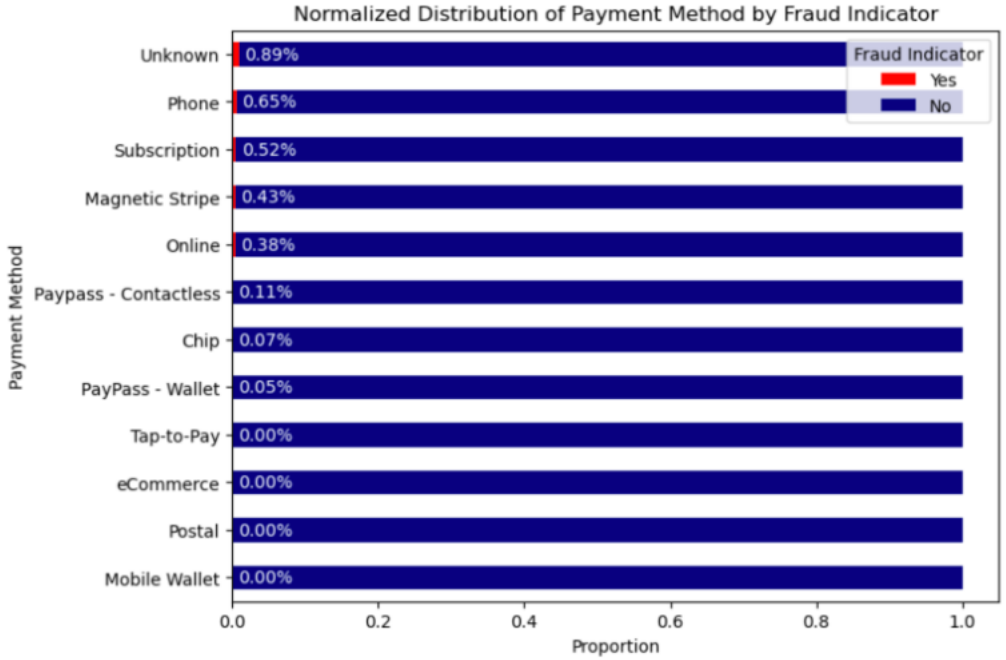


Appendix

Payment Method Bivariate Analysis

NullFraud





Merchant Location analysis

NullFraud

Merchant Location (May be insignificant)

```
Merchant Location
LBN    38.888889
SGP     7.389163
IND     6.896552
ECU     5.000000
PER     4.761905
ISL     3.125000
DOM     3.030303
DEU     2.439024
BRA     2.208202
ITA     1.016260
dtype: float64
```

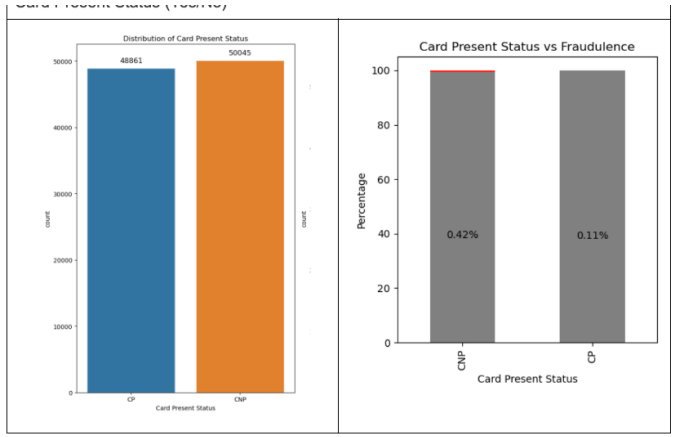
From previous visualizations, we know that 0.267931% of transactions are flagged for fraud. Similarly, we calculate the average per merchant location to compare with these high fraudulence locations.

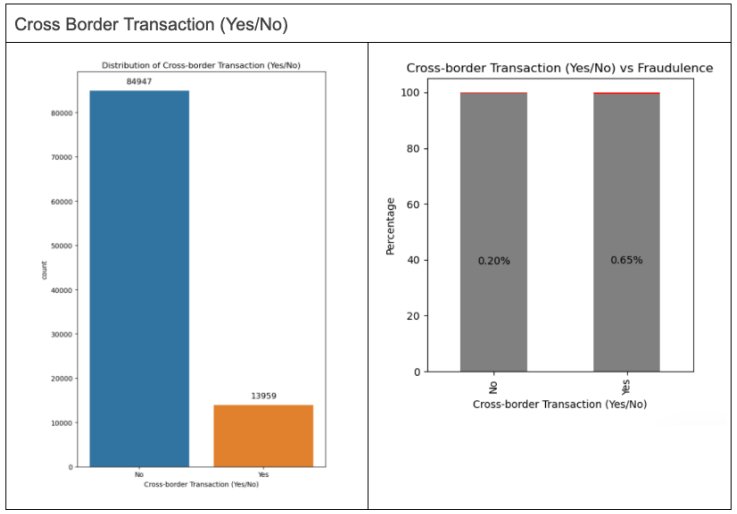
```
# Average pct of fraudulence per merchant location
avg_fraud_rate_all = fraud_rate.mean()
print(f"Average % of fraudulence per merchant location: {avg_fraud_rate_all:.6f}%")
```

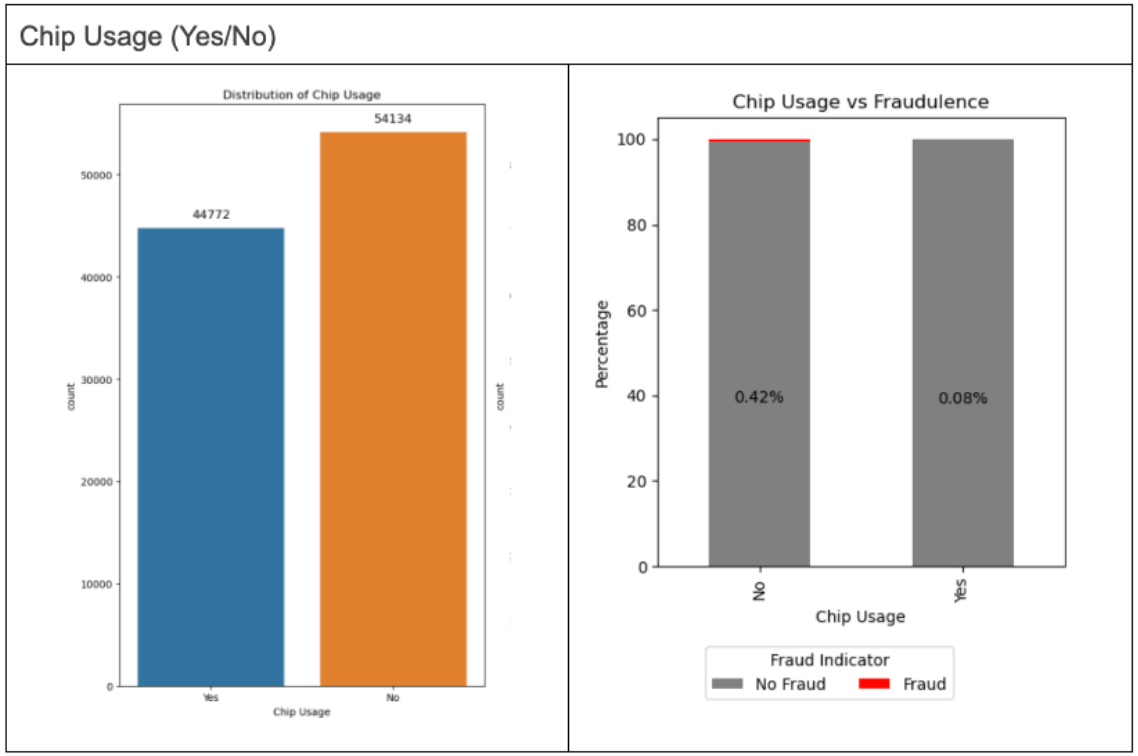
```
Average % of fraudulence per merchant location: 1.182484%
```

Merchant Location analysis

NullFraud



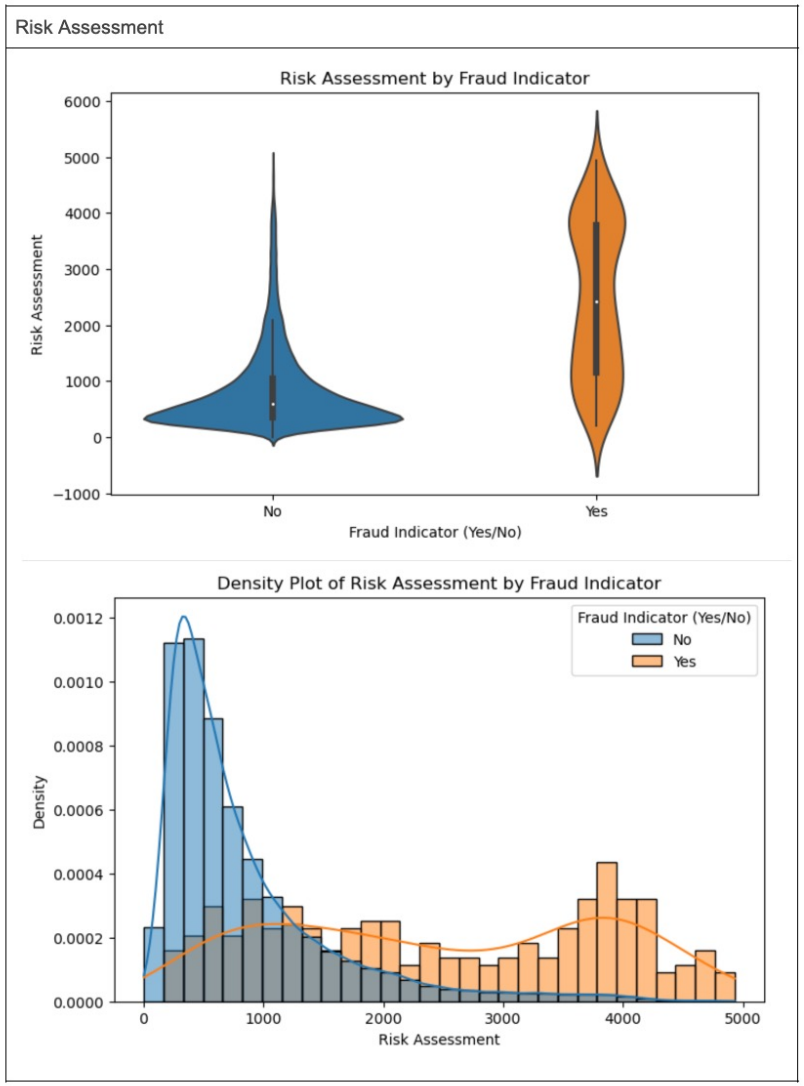


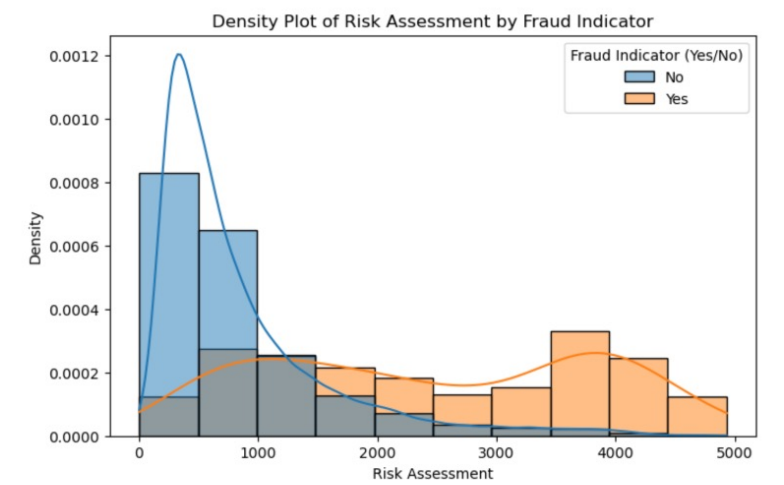


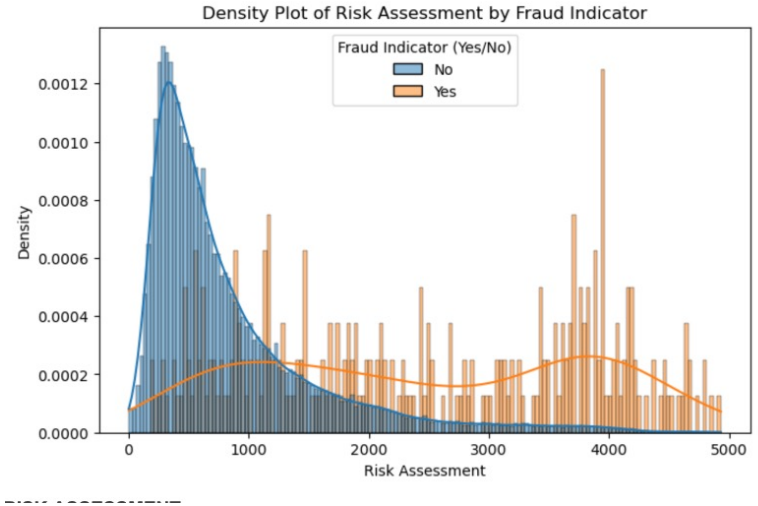
Appendix

Risk Assessment Analysis

NullFraud







Appendix

Risk Assessment Analysis

NullFraud

RISK ASSESSMENT

Distribution Shape: The 'Risk Assessment' for non-fraudulent transactions (No) is tightly concentrated around lower values with a narrow distribution, indicating that most non-fraudulent transactions have a low-risk assessment score. The distribution for fraudulent transactions (Yes) is wider, suggesting a greater variability in the risk assessment scores for fraudulent transactions.

Range: The range of 'Risk Assessment' scores for fraudulent transactions is broader than for non-fraudulent ones, indicating that fraudulent transactions can have a wide range of risk scores, but tend to have higher scores on average.

Median and Quartiles: The median (indicated by the white dot) for fraudulent transactions is higher than for non-fraudulent ones, which aligns with the expectation that transactions deemed more risky are more likely to be fraudulent.

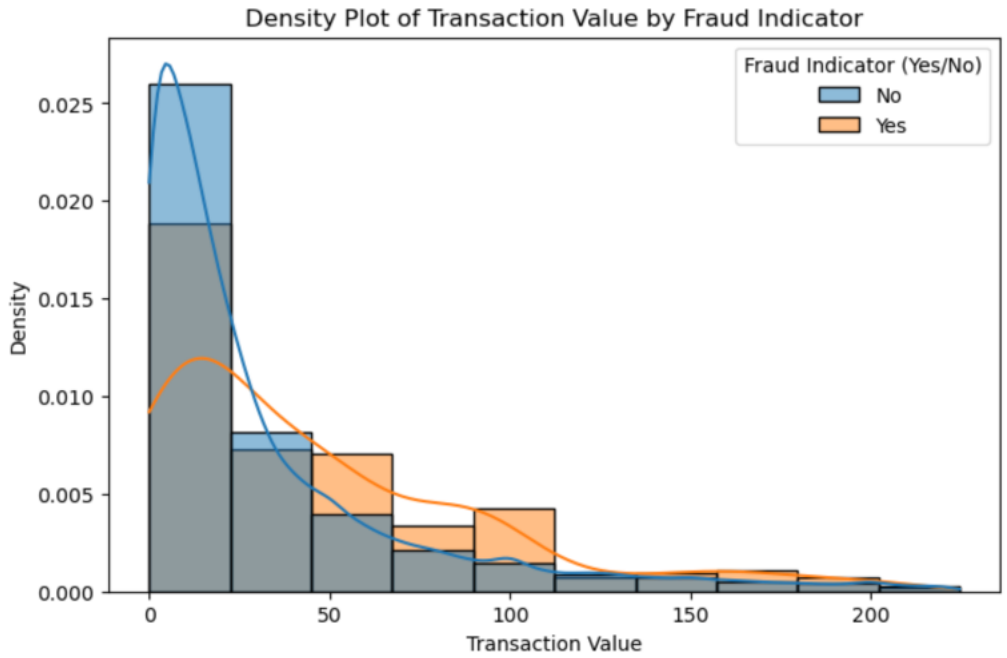
Outliers: The plot for non-fraudulent transactions shows fewer outliers compared to the plot for fraudulent transactions, indicating that most non-fraudulent transactions adhere to a lower risk profile.

Transaction Values

NullFraud

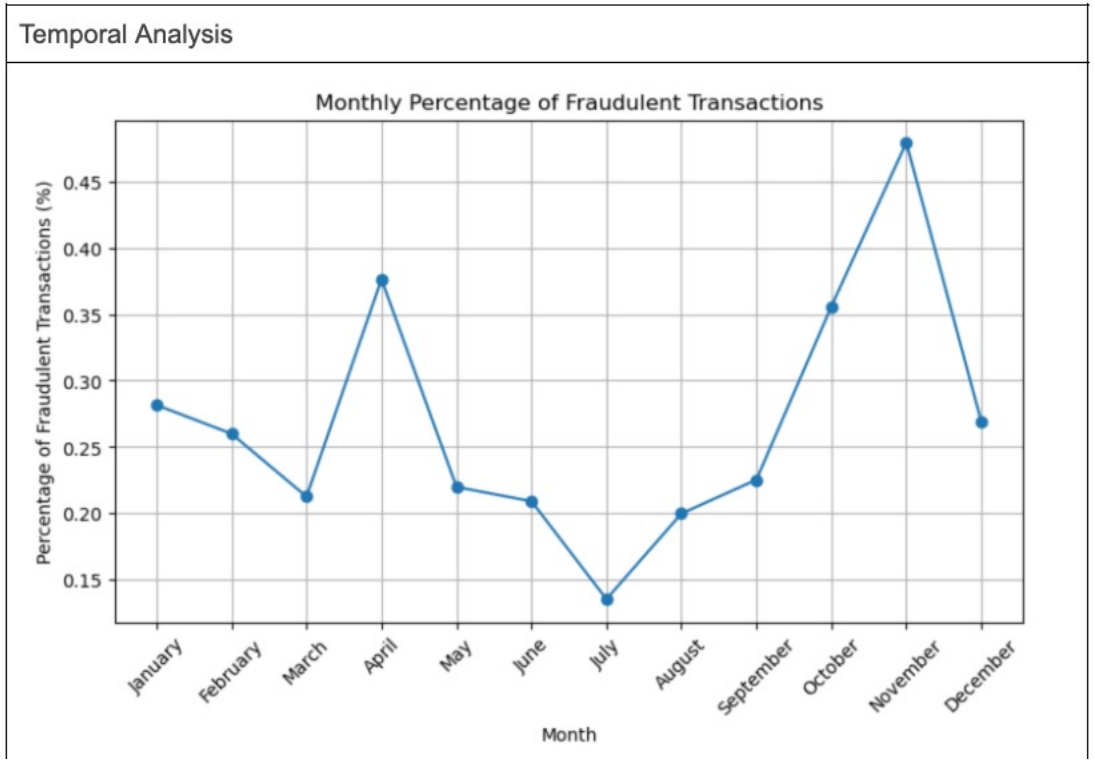
Transaction Values

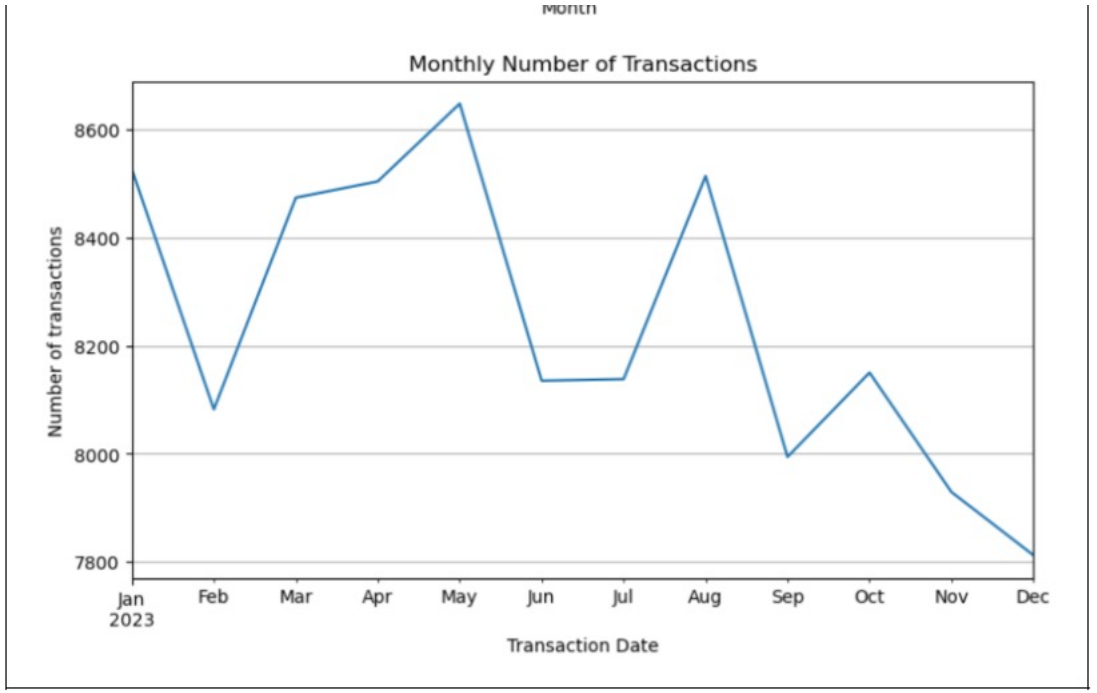
Transaction values seemingly does not reflect upon the fraudulence of a particular transaction in a linear relationship.



Temporal Analysis

NullFraud





Appendix

Cramer's V Test

NullFraud

Cramer's V-Test

Is the occurrence of fraud independent of a particular feature, and if so, by what degree

Chi-Square Test for Independence

Card Present Status - p-value: 5.383848817728829e-21
Chip Usage - p-value: 2.1958704289262313e-24
Cross-border Transaction (Yes/No) - p-value: 6.5051876113018414e-21
Risk Assessment - p-value: 0.0
Payment Method - p-value: 2.0386422879988157e-21
Transaction Value - p-value: 0.0
Merchant Location - p-value: 0.0

So what do these p-values tell us?

1. **'Card Present Status', 'Chip Usage', 'Cross-border Transaction', 'Payment Met** display **very small p-values**. These extremely small numbers suggest that the likeli of observing the data if the null hypothesis were true (no association) is extremely lo practical terms, these results indicate that there is a very strong statistical significance and a likely association between each of these categories and the occurence of fraud. The distribution of fraud indicators varies significantly across the different levels of these categorical variables.
2. **'Risk Assessment', 'Transaction Value', 'Merchant Location'** display a **p-value of 0** (likely due to rounding). This suggests a certain statistical association between these variables and the 'Fraud Indicator (Yes/No)'.

This is where the V Test comes in–we estimate the strength of the correlations.

Cramer's V Test

Card Present Status: 0.029724385577123456
Chip Usage: 0.032245345011337606
Cross-border Transaction (Yes/No): 0.02966066849583715
Risk Assessment: 0.3032596563816111
Payment Method: 0.03390723967705102
Transaction Value: 0.298090525518079
Merchant Location: 0.14172863438503566
Merchant Category Code (MCC): 0.13796010489253901

How do we interpret these values?

1. ***WEAK ASSOCIATIONS:*** **Card Present Status (0.0297), Chip Usage (0.0322), Cross-border Transaction (0.0298)**, and **Payment Method (0.0339)** have very low Cramér's V values, suggesting these variables have very weak associations with the variable they were compared against. These factors might not be strong predictors on their own for the variable of interest in your analysis.
2. ***MODERATELY WEAK ASSOCIATIONS:*** **Merchant Location (0.1417)** and **Merchant Category Code (MCC) (0.1379)** have slightly higher but still relatively low Cramér's V values. There is a weak association with the variable they were compared against, indicating they have a bit more influence than the previously mentioned variables but still a limited predictive power.
3. ***MODERATELY STRONG ASSOCIATIONS:*** **Risk Assessment (0.3033)** and **Transaction Value (0.2983)** stand out with the highest Cramér's V values among those listed, indicating a moderate association with the variable they were compared against. This suggests that these variables have a more substantial relationship and could be more significant predictors in your analysis.

SOM analysis

NullFraud

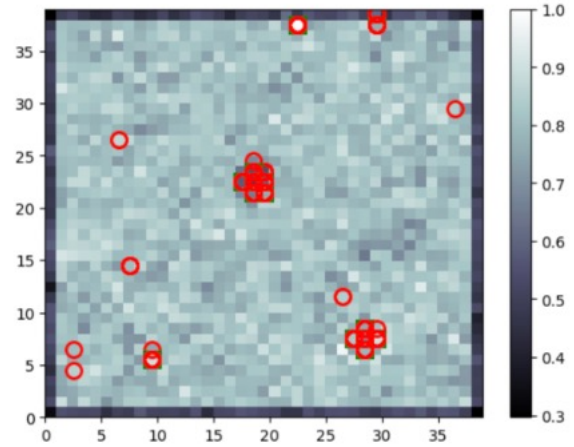
Self-Organized Map (SOM) Analysis

The Self-Organized Map is a **clustering** algorithm that leverages neural networks to iteratively group similar data points together. It has been proven to produce the highest accuracies between clustering algorithms when predicting credit fraud (G et al., 2018).

References:

G, A., K, M., Reddy, B. K. K., Iyengar, N. Ch. S. N., & Caytiles, R. D. (2018). Analyzing the performance of Various Fraud Detection Techniques. *International Journal of Security and Its Applications*, *10*(5), 21–36. https://doi.org/10.14257/ijsia.2018.12.5.03

SOM U-Matrix



Map of all the neurons (clusters) where each cluster is a square; clusters can have very little transactions or very many according to how they arrange themselves. Similar transactions will naturally group together.

How to interpret this:

- ➔ Darker shades: boundaries between different clusters.
- ➔ Lighter shades: clusters of similar data points.
- ➔ **Areas with dense clusters of red circles**: common patterns or features among fraudulent transactions. These neurons have learned to represent the "profile" of fraudulent activity within your dataset.
- ➔ **Isolated red circles:** indicate outliers or less common types of fraudulent transactions.
- ➔ **Neurons with mixed red and green shapes**: represent ambiguous areas or transitional zones between fraudulent and non-fraudulent transactions.

Clustering

NullFraud

Cluster	Features
(17, 22)	Fraud Rate: 0.113234253361641 Feature Analysis: Payment Method mode: 2.9999999999999996 Merchant Location mode: 127.0 Card Present Status mode: 0.0 Chip Usage mode: 0.0 Cross-border Transaction (Yes/No) mode: 0.0 Merchant Category mode: 134.0 Transaction Value mean: 70.5966053314461 Transaction Value median: 22.52 Risk Assessment mean: 1282.073130455296 Risk Assessment median: 877.0000000000002
(22, 37)	Fraud Rate: 0.10582010582010581 Feature Analysis: Payment Method mode: 2.9999999999999996 Merchant Location mode: 46.0 Card Present Status mode: 0.0 Chip Usage mode: 0.0 Cross-border Transaction (Yes/No) mode: 1.0 Merchant Category mode: 331.0 Transaction Value mean: 75.0776402116402 Transaction Value median: 15.889999999999999 Risk Assessment mean: 1901.989417989418 Risk Assessment median: 1717.0
(9, 5)	Fraud Rate: 0.17276014463640015 Feature Analysis: Payment Method mode: 2.9999999999999996 Merchant Location mode: 46.0 Card Present Status mode: 0.0 Chip Usage mode: 0.0 Cross-border Transaction (Yes/No) mode: 1.0 Merchant Category mode: 331.0 Transaction Value mean: 103.15316191241463 Transaction Value median: 17.0 Risk Assessment mean: 1866.0498192044997 Risk Assessment median: 1647.0

Summative Analysis

- Predominant payment method (mode approximately 3) → Online
- Card not present (0.0), and no chip usage (0.0), suggesting a pattern of remote transactions without physical card verification.
- Risk assessments notably high across all clusters, with means over 1000 and medians also high, reflecting the transactions' perceived riskiness.
- Mode for "Cross-border Transaction (Yes/No)" alternates between 0.0 and 1.0 across clusters; both domestic and cross-border transactions prone to fraud.
- Transaction values vary, with means ranging from approximately 70 to 103, and medians significantly lower.

Random forest + gradient boosting

NullFraud

Training using Random Forest & Gradient Boosting

Due to the fact that we have many categorical variables and it is not computationally viable to OneHotEncode some of these variables (such as Merchant Location), we choose to select between the **Random Forest Classification** or **Gradient Boosting** models, as it handles non-linear relationships and interactions between categorical features well.

Random Forest Classification

	precision	recall	f1-score	support
0	0.91	0.91	0.91	53
1	0.91	0.91	0.91	53
accuracy			0.91	106
macro avg	0.91	0.91	0.91	106
weighted avg	0.91	0.91	0.91	106

Accuracy: 0.9056603773584906

Feature Importances:
Risk Assessment: 0.3948
Transaction Value: 0.2168
Merchant Category Code (MCC): 0.1224
Merchant Category: 0.1167 -> Highly correlated, may have overfit
Payment Method: 0.0491
Cross-border Transaction (Yes/No): 0.0428
Chip Usage: 0.0352
Card Present Status: 0.0223

Random forest + gradient boosting

NullFraud

Training using Random Forest & Gradient Boosting

Due to the fact that we have many categorical variables and it is not computationally viable to OneHotEncode some of these variables (such as Merchant Location), we choose to select between the **Random Forest Classification** or **Gradient Boosting** models, as it handles non-linear relationships and interactions between categorical features well.

Random Forest Classification

	precision	recall	f1-score	support
0	0.91	0.91	0.91	53
1	0.91	0.91	0.91	53
accuracy			0.91	106
macro avg	0.91	0.91	0.91	106
weighted avg	0.91	0.91	0.91	106

Accuracy: 0.9056603773584906

Feature Importances:
Risk Assessment: 0.3948
Transaction Value: 0.2168
Merchant Category Code (MCC): 0.1224
Merchant Category: 0.1167 -> Highly correlated, may have overfit
Payment Method: 0.0491
Cross-border Transaction (Yes/No): 0.0428
Chip Usage: 0.0352
Card Present Status: 0.0223

Gradient Boosting (CatBoost)

CatBoost Classification Report:

	precision	recall	f1-score	support
0	0.83	0.80	0.82	56
1	0.79	0.82	0.80	50
accuracy			0.81	106
macro avg	0.81	0.81	0.81	106
weighted avg	0.81	0.81	0.81	106

Accuracy: 0.8113207547169812

Random Forest has a higher accuracy of 90.5%, compared to the 81.1% accuracy of CatBoost. The Random Forest also leads in precision, recall, and F1-scores for both classes. This suggests that Random Forest is suited for the task at hand, as it is able of not only identifying correct instances but also reducing the chances of false alarms and misses.

Appendix

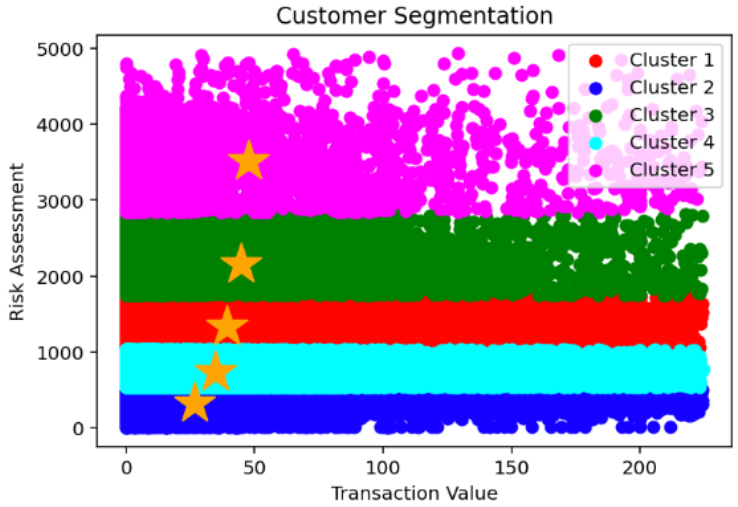
Cluster analysis of Continuous Variables

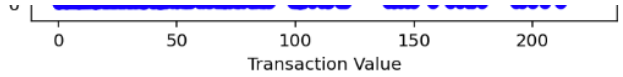
NullFraud

I conducted segmentation, classification, and prediction analyses, and created a draft interaction dashboard with Tableau that allows the freedom to drill down each trait.

1. KMeans/KModes Clustering Analysis

- Risk Assessment vs. Transaction Value



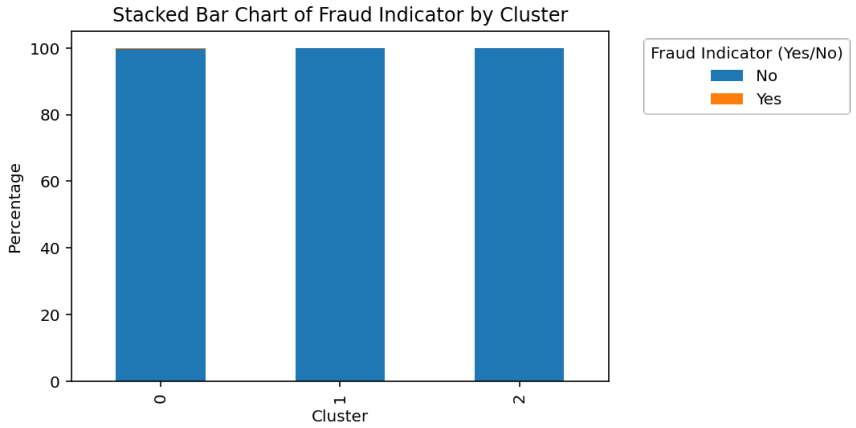


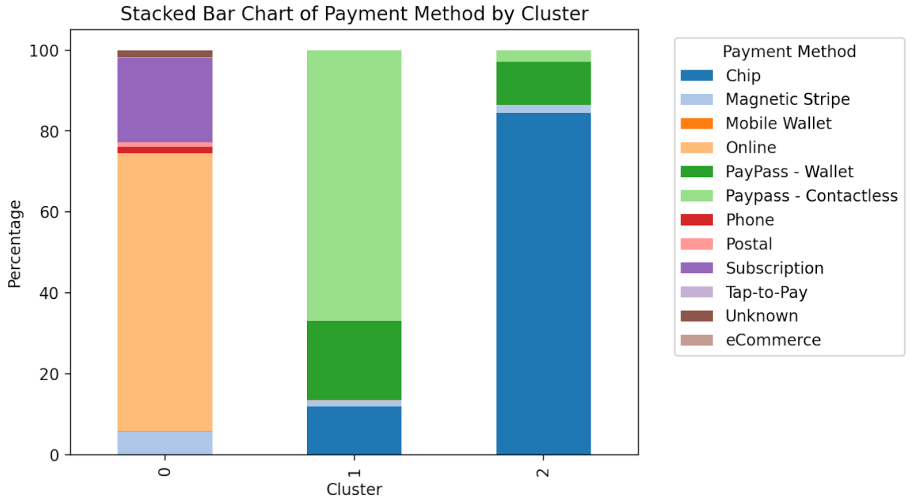
From the above segmentation, we can see that transaction value is not good for segmenting the clients.
For all 5 segments, the cluster means are below $50. There is a weak trend of larger transaction value associated with higher risk assessment, but it's not significant and the cluster with the highest assessed risk doesn't agree with the trend.

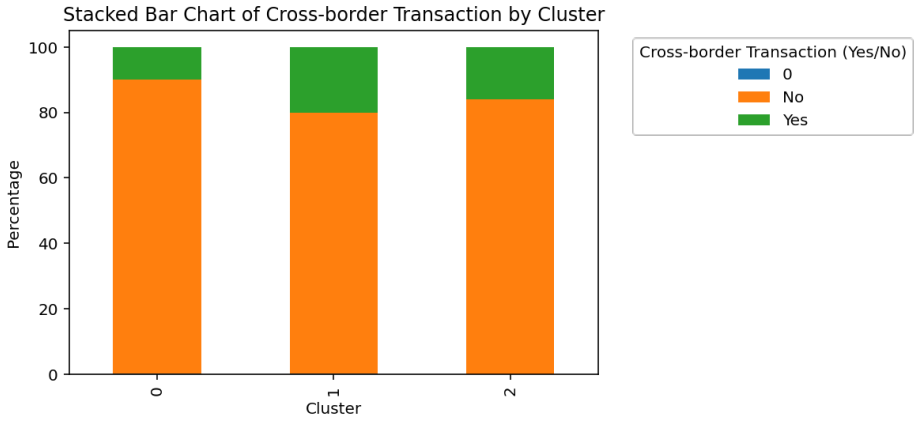
Clustering Analysis of Categorical Variables

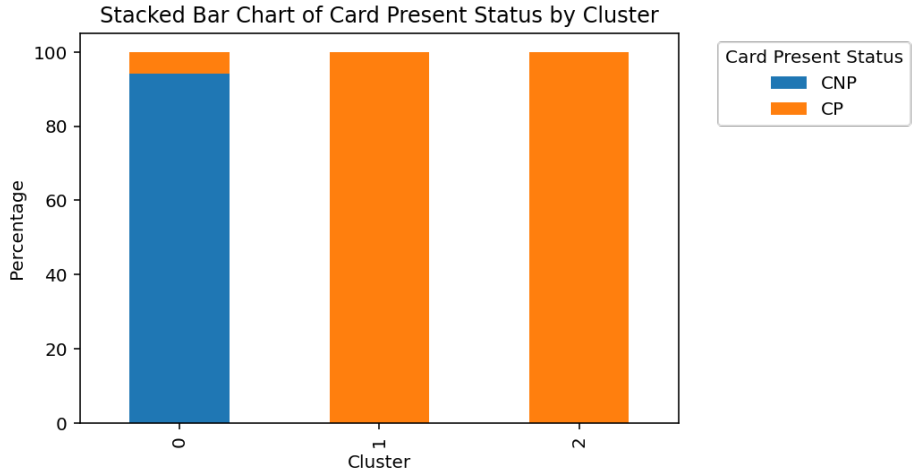
NullFraud

Clusters	Fraud BPS
0	42.625424
1	8.811435
2	7.948599
Total Average	26.832997





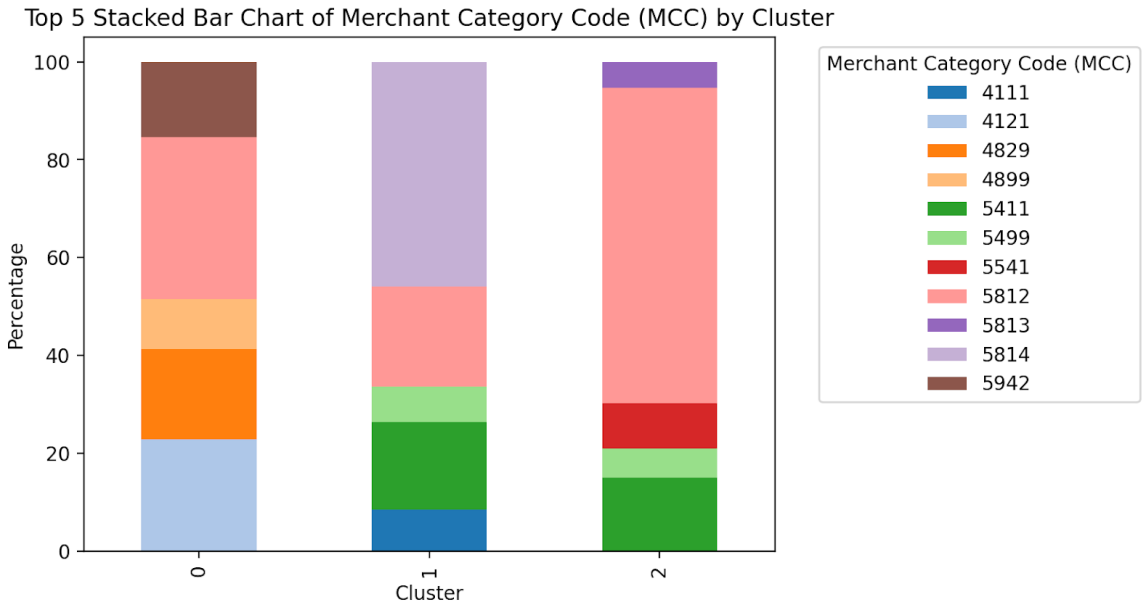


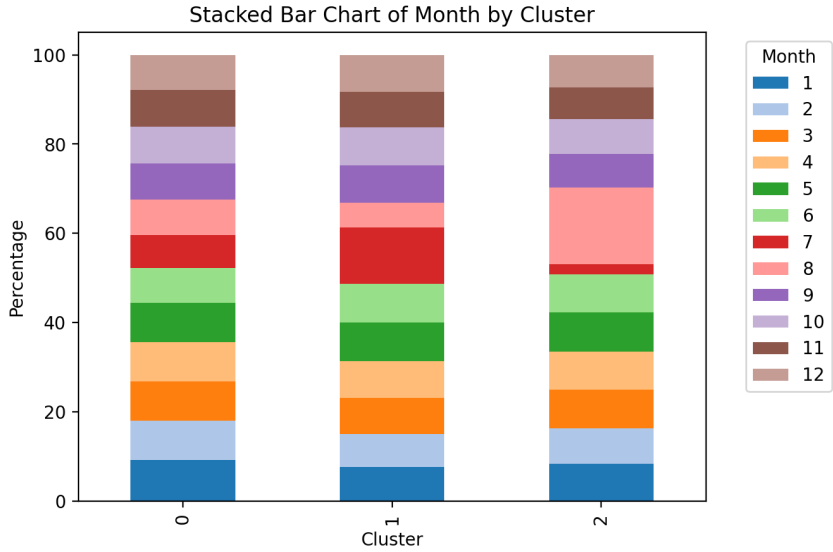


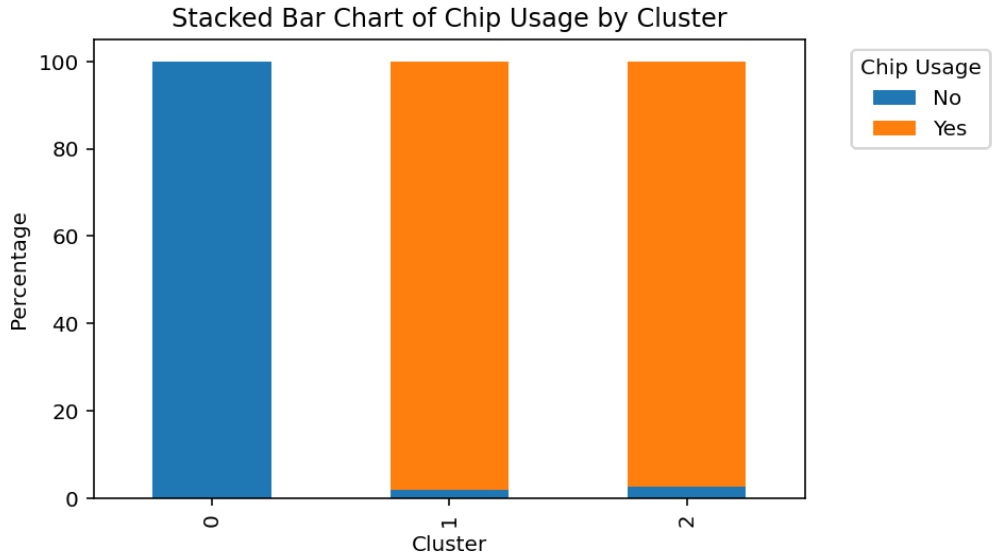
Appendix

Clustering Analysis of Categorical Variables

NullFraud







Appendix

Clustering Analysis of Categorical Variables

NullFraud

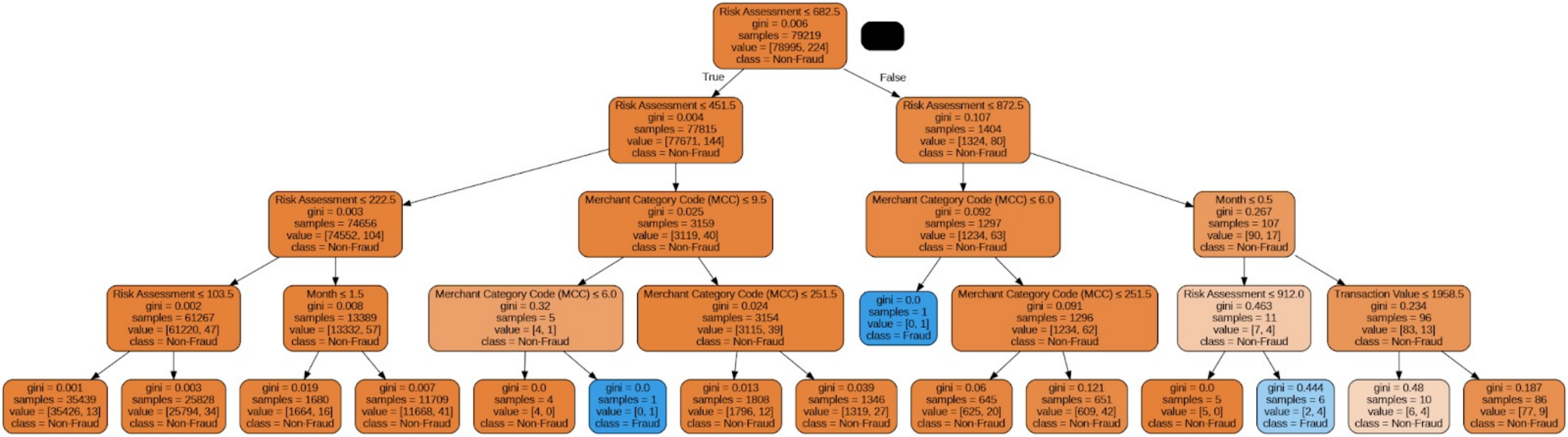
- **Cluster 0 profile:**
- Not cross-border transaction
- CNP
- Online or subscription payment
- No chip usage
- MCC: 4899 (Cable and other pay television), 4829 (Money Orders – Wire Transfer), and 4121(Taxicabs and limousines), 5942 (Book Stores and other similar services.)

Limitations:

- Unable to cluster the quantitative variables and categorical variables together
- Clustering provides a general profile of each segment with the transactions of the traits that are the closest to the centroids, but it's not a classification/predictive model.

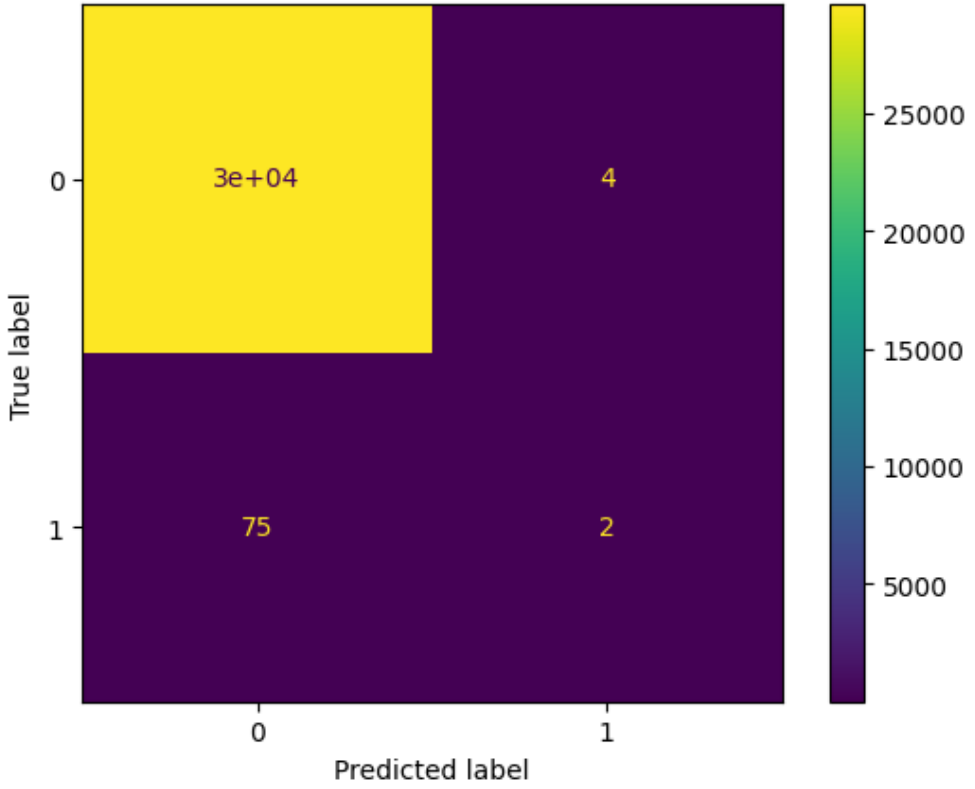
Decision Tree/ Random Forest Classification Analysis

NullFraud



Decision Tree/ Random Forest Classification Analysis

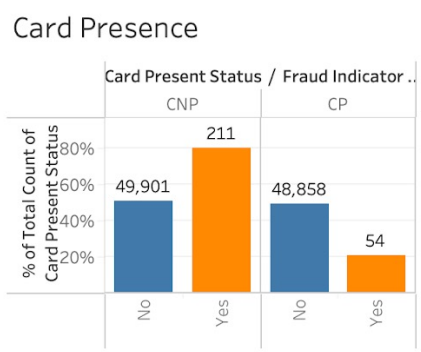
NullFraud

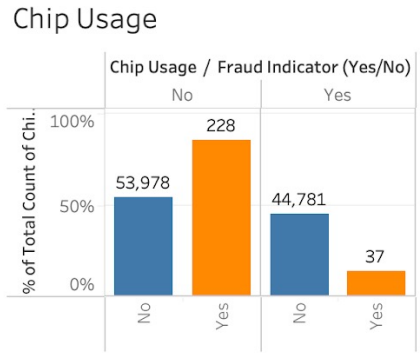


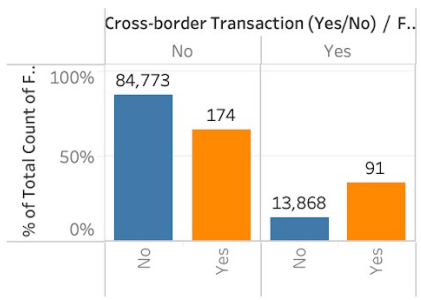
- Random Forest
  - Model Score: 0.9974
  - Confusion Matrix:
    - True Negative: 29627 - Predict No Fraud Actually No Fraud
    - True Positive: 3 - Predict Fraud Actually Fraud
    - False Negative: 74 - Predict No Fraud Actually Fraud
    - False Positive: 4 - Predict Fraud Actually No Fraud

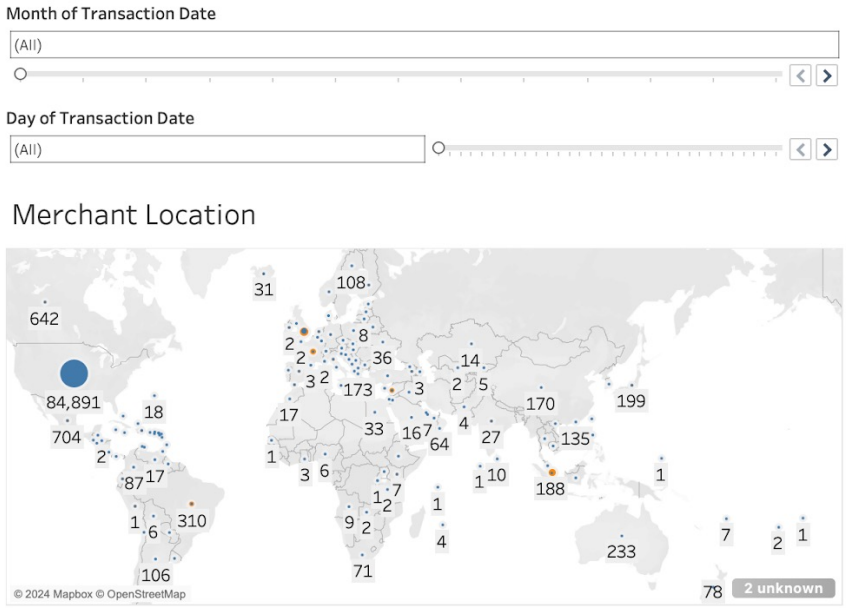
Tableau Dashboard

NullFraud









Payment Method

Payment Method	Fraud Indicator (Yes.. No	Yes
Online	36.97%	52.83%
Subscription	11.25%	21.89%
Paypass - Contactless	21.19%	8.30%
Magnetic Stripe	3.78%	6.04%
Chip	16.63%	4.53%
Unknown	0.90%	3.02%
Phone	0.78%	1.89%
PayPass - Wallet	7.66%	1.51%
eCommerce	0.00%	
Mobile Wallet	0.12%	
Postal	0.62%	
Tap-to-Pay	0.10%	

MCC

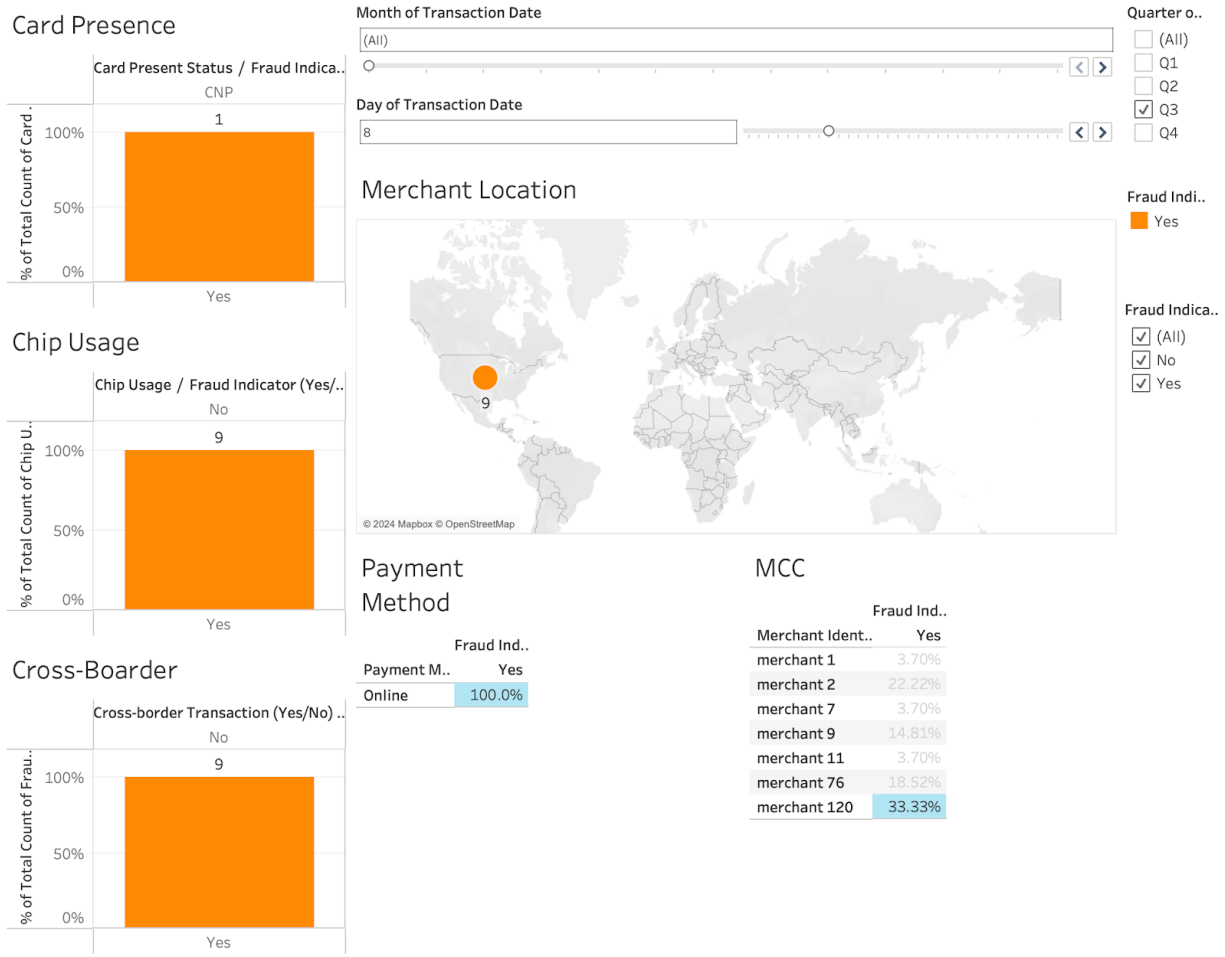
Merchant Ident..	Fraud Indicator (Yes.. No	Yes
merchant 1	33.63%	25.00%
merchant 2	8.90%	14.58%
merchant 5	5.23%	4.17%
merchant 7	6.77%	2.08%
merchant 9	6.16%	8.33%
merchant 11	7.09%	8.33%
merchant 46	12.43%	6.25%
merchant 76	7.85%	12.50%
merchant 120	7.25%	18.75%
merchant 302	4.68%	

- The main feature is that you can drill down to any level of individual trait (filter the entire dashboard) by clicking onto the filters / categories on the chart
- Example 1: Frauds using Online Payment in Q3 paid to Merchant 120

Appendix

Tableau Dashboard

NullFraud

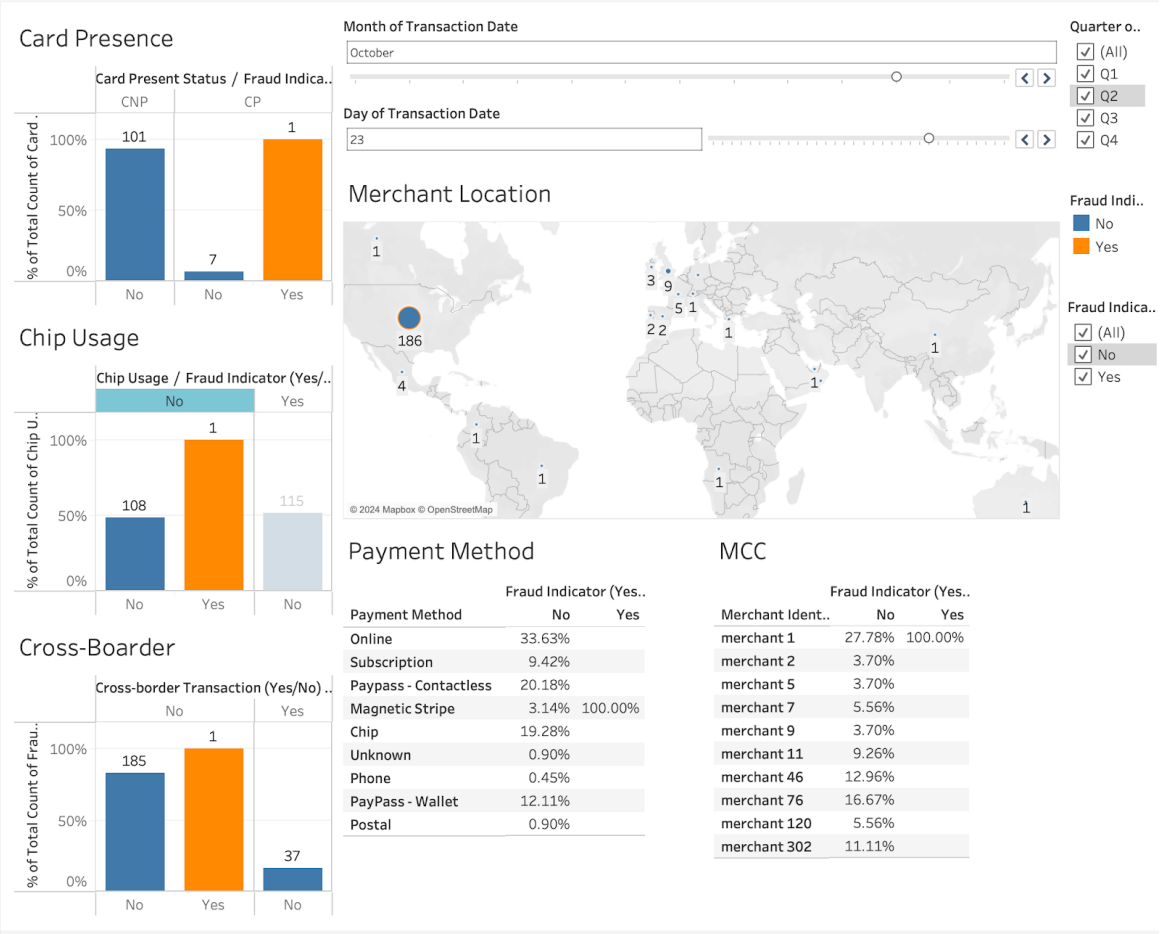


Example 1: Frauds using Online Payment in Q3 paid to Merchant 120

Appendix

Tableau Dashboard

NullFraud



Payment Method

Payment Method	Fraud Indicator (Yes.. No	Yes
Online	33.63%	
Subscription	9.42%	
Paypass - Contactless	20.18%	
Magnetic Stripe	3.14%	100.00%
Chip	19.28%	
Unknown	0.90%	
Phone	0.45%	
PayPass - Wallet	12.11%	
Postal	0.90%	

MCC

Merchant Ident..	Fraud Indicator (Yes.. No	Yes
merchant 1	27.78%	100.00%
merchant 2	3.70%	
merchant 5	3.70%	
merchant 7	5.56%	
merchant 9	3.70%	
merchant 11	9.26%	
merchant 46	12.96%	
merchant 76	16.67%	
merchant 120	5.56%	
merchant 302	11.11%	

Example 2: All transactions on October 23rd that had no chip usage

Appendix

Implementing A Scheme Token

NullFraud

TOKENIZATION ENCRYPTS...

Credit card Information ⇒ Non-sensitive tokens

Bank checks:
1. Transaction details
2. Account balance
3. Fraud alerts

Token Service Provider
(Also Nullfraud)

Token Request & Generation → Secure Storage → Token Transmission → Decrypting Token

Bank's Token Vault ensures secure PAN storage

Authorization code

Merchant

Customer Checkout → Payment Preprocessing → Token verification → Approval & completion

Tokenization secures transaction end-to-end, substantially reducing fraud risk.

NullFraud's Solution: The SPArk of Fraud Innovation

Question: How can NullFraud reduce fraud, cut down on false positives, and be seen as a pioneer for secure transactions?

Challenges:	Solutions:
How do we prevent Card Not Present fraud from occurring with internal controls?	**S**cheme Tokenization
How do we reduce the number of false positives in the predictive model?	**P**redictive ML model with binary outputs
How do we help customers fight fraud themselves while avoiding too much friction?	**A**I fraud flagging and real-time authentication for high-risk customers