Risk Analyst Case

CloudWalk Hiring Process

January 2024

Matias Scherer

Our Agenda

01

Payments Industry Overview

Money flow, information flow and the role of the main players in the payment industry. Main differences between acquirer, sub-acquirer and payment gateway. Chargebacks, cancellation and their connection with fraud in the acquiring world.

02

Chargeback Case

Resolution of a real-world customer problem related to a chargeback dispute.

03

Antifraud Case

Resolution of a real-world fraud prevention task and creation of a simple anti-fraud

Payments Industry Overview

Main Players

Who is who in the payments' industry

01

Issuers

Responsible for issuing cards and for offering credit to the cardholder. It is the entity that directly interacts with the cardholder. It also bears the credit risk.

02

Acquirers

Accredits the store for accepting electronic payment methods, being responsible for capturing, processing, and settling the transaction.

03

Payments Facilitators/ Sub-acquirers

Enables the merchant to accept payments but does not take part in the settlement process. Connects the cardholder to the acquirer, mainly on online transactions.

04

Card Schemes

Systems that facilitate electronic transactions between different parties in the payment ecosystem. They provide the infrastructure, rules, and standards necessary to process payments

Others: Regulators, Clearing Houses, Payment Service Providers, Gateways... and many more!

Payments Flow - Authorisation

A payment transaction occurs in 3 parts: Authorisation, Clearing and Settlement. Here we are looking at the **Authorisation flow**

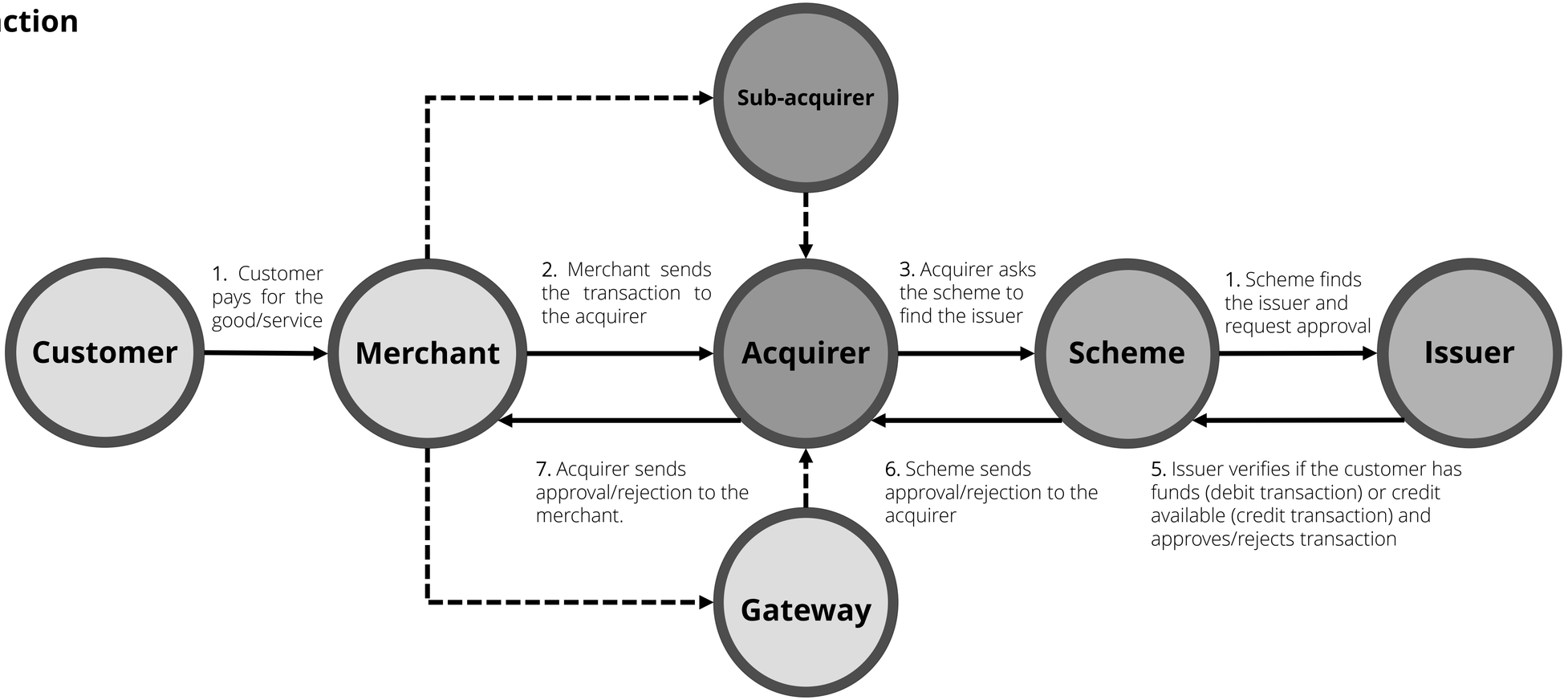
Fees applied in each transaction

Merchant Discount Rate (MDR) = IC + Schema Fee + Acquirer Markup: Total amount charged to the retailer, applied by the acquirer.

Interchange Fee (IF): Fee determined by the schema that depends on various factors (card type, market segment, channel of the transaction etc.). Is paid by the **acquirer** to the **issuer** and represents the largest share of the **MDR.**

Card Scheme Fee: Represents a small share of the **MDR**, paid both by the **acquirer** and the **issuer** to the card scheme.

Other fees: Fees included in the **MDR** that are paid by the **acquirer** to other intermediaries, such as **payment service providers** and **sub-acquirers.**



Payments Flow - Settlement

A payment transaction occurs in 3 parts: Authorisation, Clearing and Settlement. Here we are looking at the Settlement flow

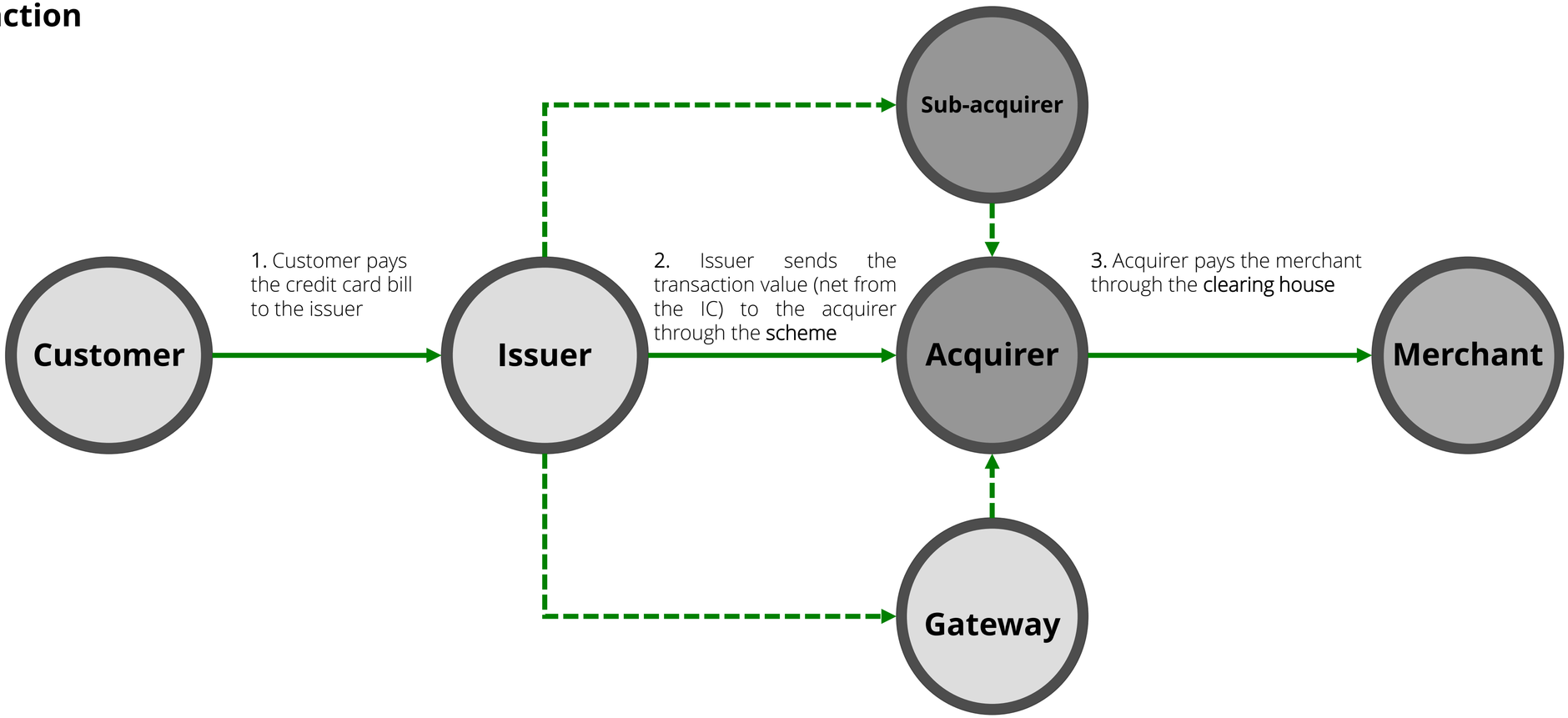
Fees applied in each transaction

Merchant Discount Rate (MDR) = IC + Schema Fee + Acquirer Markup: Total amount charged to the retailer, applied by the acquirer.

Interchange Fee (IF): Fee determined by the schema that depends on various factors (card type, market segment, channel of the transaction etc.). Is paid by the acquirer to the issuer and represents the largest share of the MDR.

Card Scheme Fee: Represents a small share of the MDR, paid both by the acquirer and the issuer to the card scheme.

Other fees: Fees included in the MDR that are paid by the acquirer to other intermediaries, such as payment service providers and sub-acquirers.



Acquirers X Sub-acquirers X Gateways

What's the difference between the acquirers and these intermediaries created mainly to facilitate online shopping?

Acquirers

Facilitate financial transactions conducted with credit and debit cards by communicating **directly** with banks and card networks.

Pros:

1. Transaction fees, which are the lowest among other payment methods.
2. It is not difficult to install and operate an acquiring system.

Cons:

1. Need to separately hire an anti-fraud system.

Sub-acquirers

Intermediary agents that communicate merchants with acquirers. In the case of sub-acquirers, **the funds from the sale flow through them, and they deduct the fees before passing the money on to the merchants**.

Pros:

1. Ability to integrate various functionalities into the system, such as multiple credit and debit options, anti-fraud systems, and reconciliation tools.

Cons:

1. Integrated services are not included in the gateways, requiring separate payments and contracts for each.
2. Acquirer needs to be contracted individually.

Gateways

Used in software format and process various types of payments, not only with cards. Bank slips and direct debits are examples of payments processed by this system. **As gateways only process payments and store data, they need to send this information to an acquirer.**

Pros:

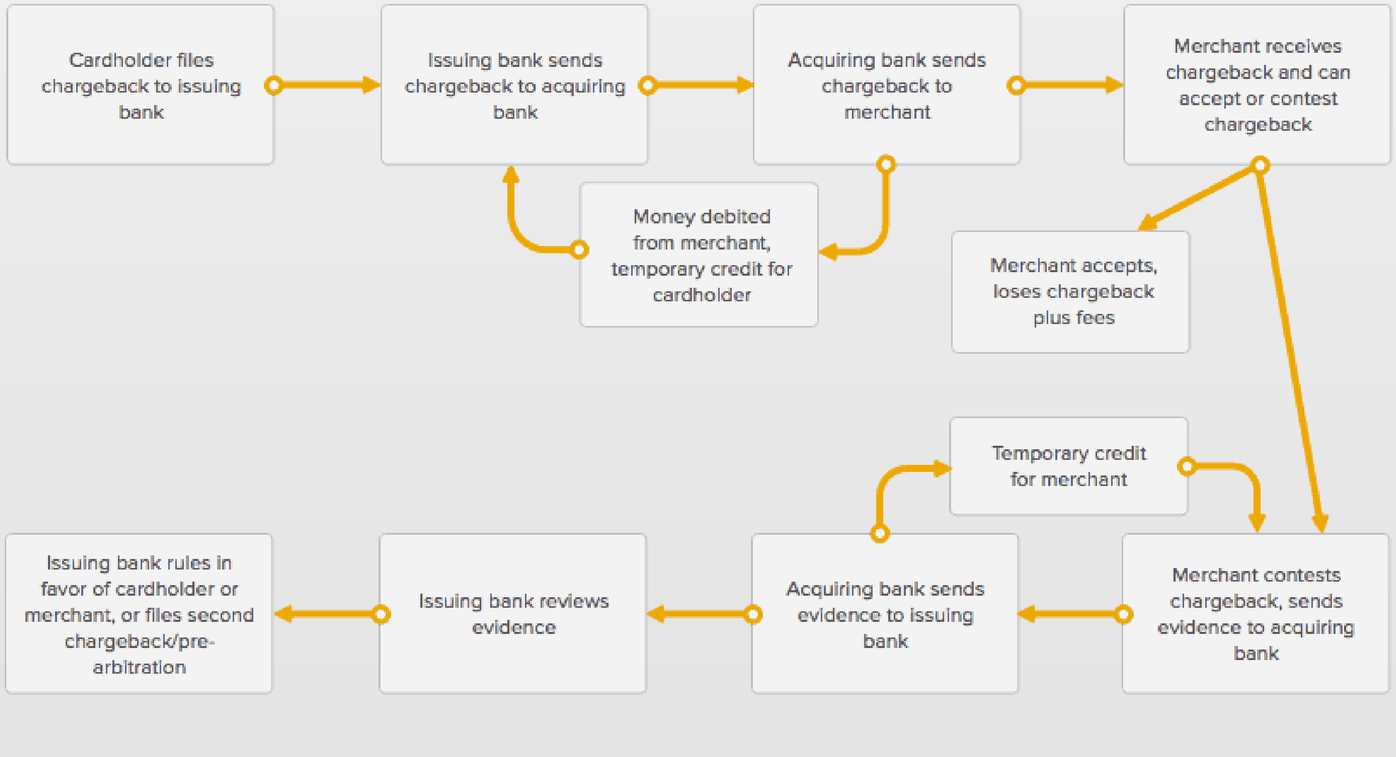
1. Integration with other services, such as acquirers and anti-fraud systems, without the need to pay extra for it, as these functionalities are already included with them.

Cons:

1. Higher transaction fees compared to other payment methods.

Chargebacks

What is it and what does it have to do with fraud?



What is it?

A chargeback is a process where a cardholder disputes a transaction and **requests a refund from their card-issuing bank**. This can occur for various reasons, such as unauthorized transactions, billing errors, or dissatisfaction with the purchased goods or services.

How it differs from a cancellation?

Chargebacks differ from cancellations in that cancellations typically involve a **mutual agreement between the buyer and the seller** to terminate a transaction before it is completed. Chargebacks, on the other hand, occur after the transaction has already taken place, and the cardholder seeks intervention from their bank to reverse the charge.

What does it have to do with fraud?

In the acquiring world, chargebacks are closely connected to fraud because they often arise from fraudulent activities, such as **unauthorized card usage or identity theft**. There are also frauds initiated by cardholders who request a chargeback for a legitimate transaction. In these cases, there may be occurrences of **friendly fraud**, in which a cardholder deliberately disputes a legitimate charge with the intention of evading payment.

Chargeback Case

Chargeback Case

Helping a costumer to win a chargeback dispute

Problem:

A client sends you an email asking for a chargeback status. You check the system, and see that we have received his defense documents and sent them to the issuer, but **the issuer has not accepted our defense**. **They claim that the cardholder continued to affirm that she did not receive the product, and our documents were not sufficient to prove otherwise**. You respond to our client informing that the issuer denied the defense, and the next day he emails you back, extremely angry and disappointed, claiming the product was delivered and that this chargeback is not right. **Considering that the chargeback reason is "Product/Service not provided", what would you do in this situation?**

Solution:

- As an acquirer, we don't have the authority to settle the chargeback dispute in favor of our client, but we can **guide and assist him in presenting their defense to the card network and issuing bank.**
- The first step is **to ensure that the customer has already provided all relevant documents** to prove to the issuer that the delivery of the product/service was completed:
    - **Order details:** invoices, tax invoices, receipts, and all related data;
    - **Customer information:** name, address, phone and/or mobile number, email, social media interactions, and any other personal data collected;
    - **Proof of delivery signed by the buyer:** product tracking number, merchandise registration ID, or any other evidence that the customer received the product delivery or accepted the provided service.

- After an **internal mapping** of which of these documents have been sent by the merchant and which ones are still missing, I would write a message to him **listing the documents that could still be submitted and reassuring him that we are going to help him** in every possible way.

- Simultaneously, I would conduct an **analysis based on internal data** to identify in the transaction history of the cardholder that filed the chargeback the presence or absence of a **suspicious behavior**, which can be identified in situations such as:

    - Different billing and shipping addresses.

    - Multiple orders to a single address but with different cards for each.

    - Numerous orders with different delivery addresses, all originating from the same IP address.

- Finally, I would **send to the issuer** the (i) new evidence provided by the merchant and (ii) the evidence collected internally based on the cardholder's transaction history (if any indications of fraudulent behavior are found).

- At each stage I would notify the customer about the progress of the chargeback dispute (that can eventually escalate to the card network).

Antifraud Case

Suspicious behaviors

Identifying fraud patterns in transaction history

- The database contains 3,199 transaction records, of which 391 (12%) received chargebacks..
- A total of 2,704 distinct customers and 1,756 distinct merchants were identified.
- Suspicious Behaviors investigated:
    1. Customers with a high number of chargebacks (in absolute terms and proportionally to the total number of transactions) and associated with many cards (especially if the card numbers show little variation).
    2. Customers with a high number of transactions in a short period of time.
    3. Cards associated with many customers.
    4. Merchants with a high number of chargeback transactions.
    5. Merchants with a high number of transactions with values unusually above their businesses' average.
- It was then possible to identify (i) customers who are likely having their personal data or cards used by fraudsters in scams and (ii) merchants that are being targeted by many fraudulent transactions.

Suspicious behaviors

Identifying fraud patterns in transaction history

- 1st Behavior: Customers with a high number of chargebacks (in absolute terms and proportionally to the total number of transactions): Some customers (identified by user_id) stood out due to a high number of chargeback transactions, which can be divided into 2 groups.:
  - I. High fraud risk: Customers with numerous transactions (5+), many cards, and a high chargeback rate (chargeback transactions/total transactions >= 60%).
  - II. Moderate fraud risk: Customers with a low number of transactions (4-) and a high chargeback rate (>= 60%).
  - When delving into the transaction history of customers exhibiting the most suspicious behavior, the use of various cards with similar numbering is observed, increasing the probability of fraud..

I. 19 users

(11750, 91637, 79054 etc.)

user_id	number_cards	number_transactions	cbk_true	cbk_false	percentage_cbk
11750	31	31	25	6	81%
91637	22	22	19	3	86%
79054	15	17	15	2	88%
96025	10	14	13	1	93%
78262	10	13	12	1	92%
75710	7	10	10	0	100%
7725	5	7	7	0	100%
17929	5	6	6	0	100%
21768	4	6	6	0	100%
3584	4	6	4	2	67%
67519	4	6	4	2	67%
83722	4	6	4	2	67%
28218	4	5	5	0	100%
71424	4	5	5	0	100%
86411	3	5	5	0	100%
99396	3	5	5	0	100%
27657	3	5	4	1	80%
69588	3	5	4	1	80%
42677	3	5	3	2	60%

II. 119 users

(11065, 17807, 27555 etc.)

user_id	number_cards	number_transactions	cbk_true	cbk_false	percentage_cbk
11065	3	4	4	0	100%
17807	3	4	4	0	100%
27555	3	4	4	0	100%
30874	3	4	4	0	100%
50643	3	4	4	0	100%
58905	3	4	4	0	100%
76819	3	4	4	0	100%
81152	3	4	4	0	100%
18227	2	4	3	1	75%
31819	2	4	3	1	75%
89615	2	4	3	1	75%
92034	2	4	3	1	75%
4651	2	3	3	0	100%
5541	2	3	3	0	100%
6761	2	3	3	0	100%
19820	2	3	3	0	100%
40493	2	3	3	0	100%
44531	2	3	3	0	100%
53850	2	3	3	0	100%

Suspicious behaviors

Identifying fraud patterns in transaction history

- 2nd Behavior: Customers with a high number of transactions in a short period of time: Some customers (identified by user_id) stood out due to a high number of transactions (3+) in a short period of time (time between one transaction and another <= 45 minutes).

17 users...
(11750, 91637, 79054 etc.)

user_id	number_cards	number_transactions	cbk_true	cbk_false	percentage_cbk	repeated_transactions_under_45_minutes
11750	31	31	25	6	81%	8
91637	22	22	19	3	86%	9
79054	15	17	15	2	88%	8
96025	10	14	13	1	93%	8
75710	7	10	10	0	100%	7
56877	6	9	5	4	56%	5
9853	6	9	4	5	44%	3

... Including 6 users that did not exhibit behaviors 1.1 or 1.2
(56877, 9853, 49106 etc.)

Clients/Behaviors	1.1 Customers with numerous transactions (5+), multiple cards, and a high chargeback rate (chargeback transactions/total transactions >= 60%).	1.2 Customers with few transactions (4-) and a high chargeback rate (chargeback transactions/total transactions >= 60%).	2 Customers with numerous transactions (3+) in a short period of time (time between one transaction and another <= 45 minutes):
56877	NO FRAUD	NO FRAUD	FRAUD
9853	NO FRAUD	NO FRAUD	FRAUD
49106	NO FRAUD	NO FRAUD	FRAUD
77959	NO FRAUD	NO FRAUD	FRAUD
40779	NO FRAUD	NO FRAUD	FRAUD
76837	NO FRAUD	NO FRAUD	FRAUD

Suspicious behaviors

Identifying fraud patterns in transaction history

- 3rd Behavior: Cards associated with many customers: Some cards (identified by card_number) stood out for being associated with a high number of customers, and they can be divided into 2 groups:
  - I. High fraud risk: Cards associated with many customers (3+).
  - II. Moderate fraud risk: Cards associated with 2 customers and sharing the first 6 digits (similar numbering) with another card associated with 2 customers.

I. 2 cards

(496045****1160 and 550209******6420)**

Unique Cards	Number of users
496045******1160	4
550209******6420	3

II. 13 cards

(406655***4572, 5162992*****1671 etc.)**

Unique Cards	Number of users
406655******4572	2
406655******5763	2
406655******7343	2
516292******1671	2
516292******1745	2
516292******2831	2
544731******3506	2
544731******3609	2
544731******8590	2
550209******1795	2
550209******6408	2
606282******1376	2
606282******4880	2

Suspicious behaviors

Identifying fraud patterns in transaction history

- 4th Behavior: Merchants with a high number of chargebacks (proportionate to the total number of transactions): Some merchants (identified by merchant_id) stood out for a high chargeback rate (chargeback transactions/total transactions >= 60% for merchants with 10 transactions or more).

12 merchants
(17275, 4705, 53041 etc.)

merchant	number_transactio	number_cbk_tru	number_cbk_true %
17275	30	22	73%
4705	22	19	86%
53041	19	14	74%
1308	15	15	100%
77130	15	13	87%
91972	14	11	79%
42356	12	8	67%
44927	11	11	100%
55854	11	9	82%
29214	10	9	90%
65330	10	8	80%
73271	10	10	100%

Suspicious behaviors

Identifying fraud patterns in transaction history

- 5th Behavior: Merchants with a high number of transactions with values unusually above their businesses' average: Some merchants (identified by merchant_id) stood out due to a significant number of transactions (above 20%) with unusually high values (transactions at least 1 standard deviation above their businesses' sample mean of the transaction amount, considering merchants with 10 transactions or more).

... Including 1 merchant that did not exhibit behavior 4
(24340)

Clients/Behaviors	4	5
	Merchants with a high chargeback rate (chargeback transactions/total transactions >= 60% for merchants with 10 transactions or more)	Merchants with a substantial percentage (above 20%) of transactions with unusually high values (transactions with at least 1 standard deviation above the sample mean of the transaction amount, considering merchants with 10 transactions or more):
24340	NO FRAUD	FRAUD

4 Merchants...
(1308, 73271, 29214 etc.)

merchant_id	average_transactio	std_deviation_transaction	number_transactions	number_cbk_true	number_transactions_above_avg+1_std_de	number_transactions_above_avg+1_std_dev %
1308	2,301	346	15	15	3	20.0%
73271	1,458	982	10	10	2	20.0%
29214	583	95	10	9	2	20.0%
24340	456	225	10	0	3	30.0%

Actions

Based on the observed behaviors, how to act to reduce the number of frauds?

Based on the spreadsheet analysis, which actions can be taken?

- I. I would reach out to all merchants involved in transactions with suspicious behaviors and notify them about suspicious cards (card_number) and consumers (user_id), so they can address the issues with their customers who are being targeted by fraudsters.
- II. For the most impacted merchants with significant relevance to our TPV, I would coordinate with the sales and customer service teams to guide them on fraud prevention practices.
- III. Temporarily, I would block transactions associated with cards, user_ids, or establishments linked to many suspicious transactions.
- IV. I would collaborate with regulatory bodies and players from other links in the chain, within regulatory limits (such as LGPD), to identify complex networks of fraudsters.
- V. I would incorporate the latest patterns into our internal fraud prevention models.

What other data should we consider to find patterns of possible fraudulent behavior?

- Geolocation Data: Information about the location of transactions, such as IP address, physical location of the device, or merchants' location.
- Merchant Data: Identify the business category of the business (fashion retail, digital goods etc.).
- Device Data: Further information about the device used in the transaction, such as device type, operating system, and device activity history.
- Card Data: In addition to the card digits provided in the spreadsheet, consider also the usage history of that card, its expiration date, card tier ("gold", "platinum" etc.), issuer etc.

Actions

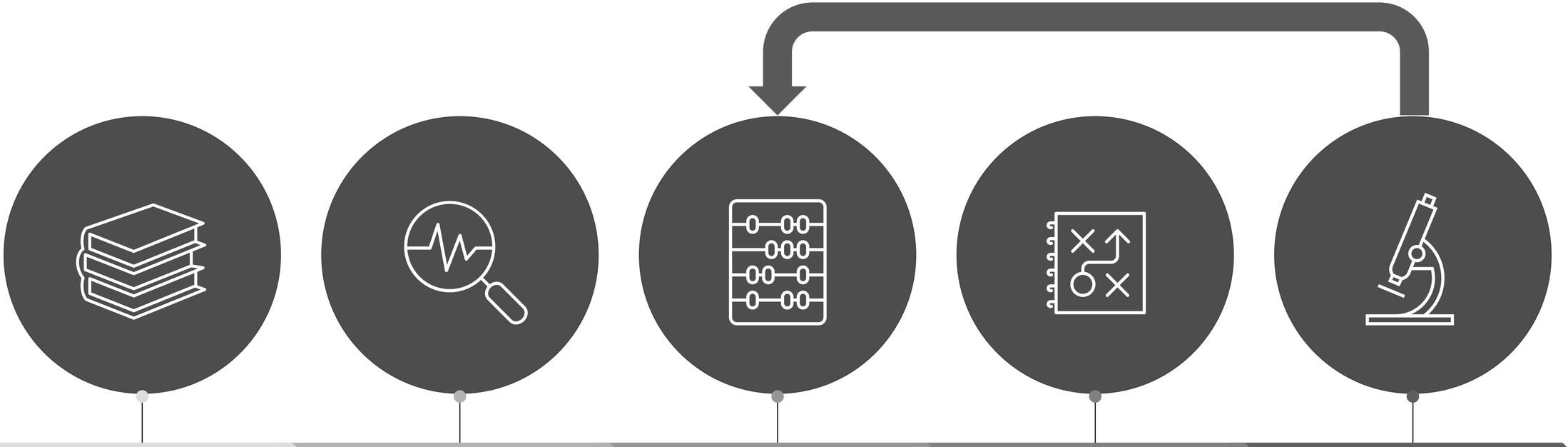
Based on the observed behaviors, how to act to reduce the number of frauds?

Considering the conclusions, what can we <u>further suggest</u> in order to prevent frauds and/or chargebacks?

- 1. Enhance Customer Verification: (I) Implement stronger customer verification measures, especially for users with a high number of chargebacks. (II) Demand multi-factor authentication or additional verification steps for transactions involving suspicious users.
- 2. Limit Transaction Frequency: (I) Set limits on the number of transactions a customer or card can make within a specific time frame.
- 3. Transaction Blocking Rules: (I) Develop dynamic rules to temporarily block transactions associated with cards, user_ids, or merchants linked to many suspicious activities.
- 4. Educate Merchants: (I) Provide ongoing education and training to merchants, especially those most impacted and with the biggest TPV on recognizing and preventing fraudulent transactions and proactive measures to reduce the risk of chargebacks.
- 5. Periodic Fraud Analysis: (I) Conduct regular and systematic analyses of transaction data to identify new trends and patterns and use the findings to continually refine and update fraud prevention strategies.

Developing an anti-fraud model!

Timeline of the modeling process



Importing libraries and dataset

Libraries:
- Data wrangling (pandas, numpy)
- DataViz (matplotlib, plotly, seaborn)
- Statistics and ML: sklearn, imblearn

Dataset:
transactions_database.csv

Exploring the dataset

- df.shape
- df.info()
- df.describe()
- df.isnull().sum()
...

Preparing Data

- Converting DateTime column to DateTime format
- Performing calculations to create new cathegorical variables from the original ones.
- Dealing with Null values
- Dividing the df in features (X) and target variables (Y)
- Separating training and testing dataset
- Rebalancing data using SMOTENC

Applying models

- Replacing string cathegoric values for numeric ones
- Applying machine learning models: Random Forest, Decision Tree, Ada Boost and Gradient Boosting

Evaluating the models

- Accuracy
- Recall
- Precision
- F1 score
- Confusion Matrix
- Comparing results
- Decision: train models with other datasets?

Developing an anti-fraud model!

Challenges and solutions

Main challenges:

- Unlike most datasets used in machine learning models applied to credit card fraud prevention available online, the dataset we had at our disposal had **many identifying attributes** ("user_id," "merchant_id," "device_id," "card_number") and **only one continuous numeric variable** ("transaction_amount"), in addition to the date column ("transaction_date") and the binary categorical variable "has_cbk" (target variable).

- As is common in credit card transaction datasets, there is a **strong imbalance** between occurrences of the target variable (many more legitimate transactions than fraudulent ones), requiring (i) **rebalancing the training dataset** in the data preparation stage to include records identified as fraud and (ii) **combining different metrics in the model evaluation stage** (besides solely accuracy, for example).

Solutions:

- **Creation of categorical variables that** can be (i) used in the SMOTE rebalancing method and (ii) inputed into predictive models. They are:

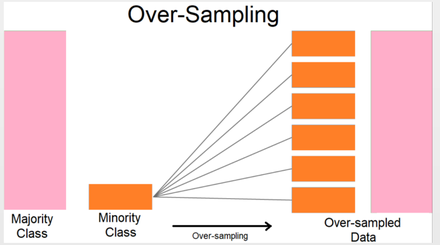
1. **user_transactions_category, card_transactions_category,** and **device_transactions_category**: Count the number of transactions associated with each user, card, and device, respectively, in the sample. They can take on the values "low," "medium," and "high" depending on the number of occurrences (later replaced by 0, 1, and 2).

2. **last_user_transaction_category** and **last_card_transaction_category:** Quantify the time of the last transaction associated with a user or card, respectively. They can take on the values "recent," "medium," and "distant" depending on the time elapsed between transactions (later replaced by 0, 1, and 2)..

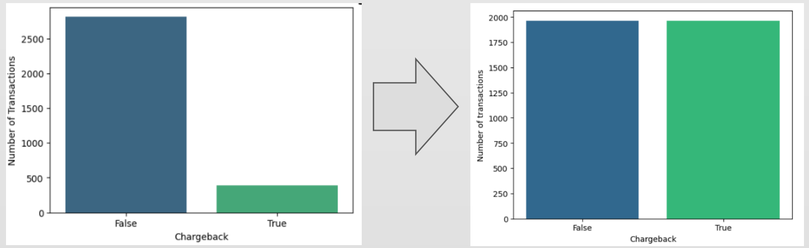
3. **transaction_profile_merchant:** Quantify whether the transaction amount is above the mean + 1 * standard deviation for the establishment. It can take on the values "typical" and "atypical" (later replaced by 0 and 1).

4. **similar_card_in_the_day:** Quantify whether the first 6 numbers of the 'card_number' variable are found in any other transaction made within 24 hours. It can take on the values "yes" and "no" (later replaced by 0 and 1).

- **Rebalancing the training dataset** using the SMOTE oversampling method:



- In our dataset, we had **only 12%** of the records marked as "TRUE" in the target variable ("has_cbk"). After applying SMOTENC (which can handle both numerical and categorical features) **only on the training dataset**, this number reaches 50%.



Developing an anti-fraud model!

Evaluating the results and of 4 different models applied to 3 different dataframes versions.

Glossary

ML Models:

Random Forest (RF): The idea is to train multiple decision trees (uncorrelated) using samples from the dataset and make predictions based on the most frequent outcomes.

Decision Tree (DT): Decision tree builds classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

AdaBoost (AB): AdaBoost is built using a similar concept of Random Forest, but instead of trees, the algorithm uses stumps. A stump is a decision tree consisting of a root node + leaf nodes.

Gradient Boosting (GB): Gradient boosting is one of the variants of ensemble methods where you create multiple weak models and combine them to get better performance as a whole.

Model Evaluation Metrics:

Accuracy - (True Negatives + True Positives)/(True Negatives+False Positives+True Positives+False Negatives): Accuracy represents the number of correctly classified data instances over the total number of data instances, so **it's not the best metric for unbalanced datasets.**

Precision – (True Positives)/(True Positives + False Positives): Precision is a good measure to determine, when the costs of False Positive is high.

Recall – (True Positives)/(True Positives + False Negatives): Recall is the model metric we use to select our best model when there is a high cost associated with False Negative.

F1 Score – 2*((Precision*Recall)/(Precision+Recall)): F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall AND there is an uneven class distribution (large number of Actual Positives).

Dataframe version: 1

Variables:
user_transactions_category,
card_transactions_category,
device_transactions_category,
last_user_transaction_category,
last_card_transaction_category,
transaction_profile_merchant
transaction_amount

	Metrics	Results RF	Results DT	Results AB	Results GB
0	Accuracy	0.795	0.785	0.866	0.865
1	Precision	0.328	0.311	0.465	0.462
2	Recall	0.664	0.638	0.741	0.733
3	F1_score	0.439	0.418	0.571	0.567

Dataframe version: 2

Variables:

user_transactions_category,
last_user_transaction_category,
transaction_profile_merchant,
transaction_amount

	Metrics	Results RF	Results DT	Results AB	Results GB
0	Accuracy	0.773	0.772	0.861	0.852
1	Precision	0.298	0.300	0.453	0.437
2	Recall	0.647	0.664	0.707	0.776
3	F1_score	0.408	0.413	0.552	0.559

Dataframe version: 3

Variables:

user_transactions_category,
last_user_transaction_category, last
transaction_profile_merchant,
similar_card_in_the_day,
transaction_amount

	Metrics	Results RF	Results DT	Results AB	Results GB
0	Accuracy	0.796	0.789	0.850	0.858
1	Precision	0.337	0.333	0.429	0.451
2	Recall	0.716	0.750	0.733	0.793
3	F1_score	0.459	0.462	0.541	0.575

Developing an anti-fraud model!

Picking one out of 4 different models and one of the 3 different dataframes versions

Glossary

ML Models:

Random Forest (RF): The idea is to train multiple decision trees (uncorrelated) using samples from the dataset and make predictions based on the most frequent outcomes.

Decision Tree (DT): Decision tree builds classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

AdaBoost (AB): AdaBoost is built using a similar concept of Random Forest, but instead of trees, the algorithm uses stumps. A stump is a decision tree consisting of a root node + leaf nodes.

Gradient Boosting (GB): Gradient boosting is one of the variants of ensemble methods where you create multiple weak models and combine them to get better performance as a whole.

Model Evaluation Metrics:

Accuracy - (True Negatives + True Positives)/(True Negatives+False Positives+True Positives+False Negatives): Accuracy represents the number of correctly classified data instances over the total number of data instances, so **it's not the best metric for unbalanced datasets.**

Precision – (True Positives)/(True Positives + False Positives): Precision is a good measure to determine, when the costs of False Positive is high.

Recall – (True Positives)/(True Positives + False Negatives): Recall is the model metric we use to select our best model when there is a high cost associated with False Negative.

F1 Score – 2*((Precision*Recall)/(Precision+Recall)): F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall AND there is an uneven class distribution (large number of Actual Positives).

And the chosen ones were:

Model: Gradient Boosting - Dataframe version: 3

	Metrics	Results RF	Results DT	Results AB	Results GB
0	Accuracy	0.796	0.789	0.850	0.858
1	Precision	0.337	0.333	0.429	0.451
2	Recall	0.716	0.750	0.733	0.793
3	F1_score	0.459	0.462	0.541	0.575

Why?

Although Gradient Boosting in version 3 of the DataFrame has lower accuracy and precision compared to AdaBoost and the Gradient Boosting in version 1 of the DataFrame, it had the best performance in metrics considered most important for the business context: Recall and F1 Score.

This is because it was adopted the premise that **the damage caused by authorizing a fraudulent transaction (considering the entire emotional and financial impact on the consumer and the merchant) is greater than blocking a legitimate transaction.**

Access all the resources on my GitHub page

Excel file with some analysis, CSV with the transactions-data and the 3 Python versions of the 4 ML models

Excel file

Analysis used to solve questions 3.1, 3.2 and 3.3

CSV file

Dataframe uploaded to solve question 3.4

Jupyter Notebooks - Python

Fraud Prevention Model – Dataframe version 1

Fraud Prevention Model – Dataframe version 2

Fraud Prevention Model – Dataframe version 3

References

Resources used in the case resolution

- Payments 4.0: As forças que estão transformando o mercado brasileiro: https://www.amazon.com.br/Payments-4-0-transformando-mercado-brasileiro-ebook/dp/B08P5X9VTC
- Data Science do zero: noções fundamentais com Python: https://www.amazon.com.br/Data-Science-Do-Zero-Fundamentais/dp/8550811769
- O que acontece se meu estabelecimento tiver uma venda contestada?: https://ajuda.infinitepay.io/pt-BR/articles/3944818-o-que-acontece-se-meu-estabelecimento-tiver-uma-venda-contestada
- O que é chargeback? E como ele influencia no seu negócio?: https://www.infinitepay.io/blog/o-que-e-chargeback-e-como-ele-influencia-no-seu-negocio?utm_source=central-ajuda&utm_medium=referral
- O que é chargeback e quando contestar uma compra feita com o InfiniteCard?: https://ajuda.infinitepay.io/pt-BR/articles/6820527-o-que-e-chargeback-e-quando-contestar-uma-compra-feita-com-o-infinitecard
- Imbalanced Learn Documentation: https://imbalanced-learn.org/dev/references/generated/imblearn.over_sampling.SMOTENC.html#imblearn.over_sampling.SMOTENC
- Accuracy, Precision, Recall or F1?: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

Thanks!

Call me:

+55 11 99706 9331

Email me:

matias.scherer@usp.br

Follow me:

https://www.linkedin.com/in/matias-scherer-/
https://github.com/matiascherer