

Risk Analyst Case – Cloudwalk Inc

Rian Rodrigues de Oliveira

João Pessoa, Paraíba, Brazil

January 2022

Understand the Industry

1. Explain the money flow and the information flow in the acquirer market and the role of the main players.

These are the main players of the acquiring market:

- Issuer (or Issuing Bank): Is the bank who provided the cardholder with the debt or credit card.
- Cardholder: is the person who is using a debt or credit card to purchase goods or services from the merchant.
- Acquirer (or Acquiring Bank): is the bank who provided the merchant with a merchant's account, where the money from the cardholder's purchase will eventually end up.
- Merchant: is the person who is selling goods or services to the cardholder.
- Card Network: the card network acts as the bridge between the acquiring and issuing banks.

A buyer, in this case called a cardholder, goes to a merchant's store to buy goods or services. The cardholder decides to pay through his card (credit or debit). As the card information is read through the terminal, all the relevant information of the purchase is sent to the merchant's bank, in this case known as an Acquiring Bank.

The Acquiring Bank sends this information to the card network, which is controlled by the card brand, for example, Visa or Mastercard. The card network will now request confirmation to the cardholder's bank, known as Issuer or Issuing Bank. If the purchase is valid and the cardholder has funds or credit limit to cover it, the Issuing bank will send the "OK" to the card network, which in turn will send the confirmation to the Acquiring Bank. This flow of information takes about 2-3 seconds. The money is then sent from the cardholder's account in the Issuing bank to the merchant's account in the Acquiring Bank, although it may take a few business days for the merchant to have access to these funds.

2. Explain the difference between acquirer, sub-acquirer and payment gateway and how the flow explained in question 1 changes for these players.

As previously stated, the acquirer is the bank in which the merchant has his merchant's account: where the money from the commerce of his goods or services will end up. A sub-acquirer, if present, will assume the Acquirer's role in processing the transaction and then forwarding it to the card network. However, the sub-acquirer does not have a merchant's account in which the funds will eventually be deposited, and usually does not hold financial liability.

A payment gateway is a platform that will capture the card information and send it forward to be processed. It is a digital equivalent of a card terminal, used in e-commerce. The cardholder can put his information on the gateway purchase goods without ever meeting the merchant directly.

3. Explain what chargebacks are, how they differ from cancellations and what is their connection with fraud in the acquiring world.

A chargeback is a process in which the cardholder requests to his issuing bank be reimbursed of a purchase made after the product in question has already arrived or, in case of a service, has already been used.

To illustrate, let's set the following scenario: Bob, a cardholder, purchases a book on an online retailer. If Bob cancels his order before the book is shipped, it's a **cancellation**. If he received the book, contacts the merchant and claims that, for whatever reason, his purchase was not what he expected, he can request a **refund**. The merchant may or may not accept his request, and may or may not request the retrieval of the merchandise. If, however, Bob receives his book and makes a request to his issuing bank to be reimbursed, it's a **chargeback**.

While the reasons for a chargeback can be many, we can separate chargebacks in three different types. First, criminal fraud, when the purchase was a result of a crime: Bob got his card stolen or cloned, for example. Secondly, merchant error: these happen because of a mistake by the merchant. Maybe the wrong product was sent, maybe it was never sent, maybe the product arrived broken or with defects. Thirdly, the merchant did not make any mistakes, nor did Bob had his information or card stolen from him, but for whatever reason Bob requests his money back: it is known as friendly fraud.

Friendly fraud has many causes, and may not be malicious in nature. It is possible that a cardholder's relative made the purchase without his authorization, or the purchase was a mistake, or the cardholder simply does not wish to contact the merchant to ask for a refund. A friendly fraud can also be requested by malicious cardholders, aiming to have the product without paying for it. It is estimated that the majority of chargeback requests are friendly frauds.

In the case of a chargeback, the merchant will be required to reimburse the cardholder and also pay a number of chargeback fees, the amount of which can be greater than the price of the purchase. Because of that, and because friendly fraud is much more frequent, a merchant would be wise to implement strategies to reduce or, at least, give him the documentation to win chargeback disputes.

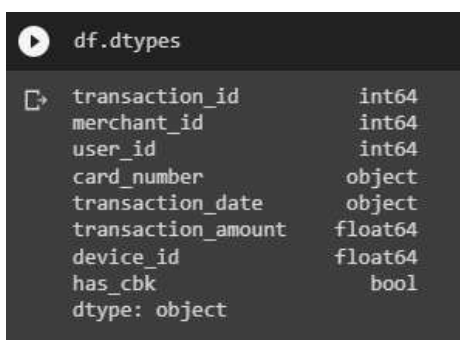
Get your hands dirty

1. Analyze the data provided and present your conclusions.

1.1 Describing the data.

The data provided contains transaction information, between the months of november and december, 2019. There are 3199 rows and 8 columns, or features. The data contains the transaction id, the merchant's id, the user's id, the credit or debit card number, the transaction date and time, the amount of money in the transaction, the id of the device used, and finally a flag that marks if the user has requested chargeback of the transaction. In this dataset, 391 transactions have the chargeback flag. This constitutes approximately a 14% chargeback rate.

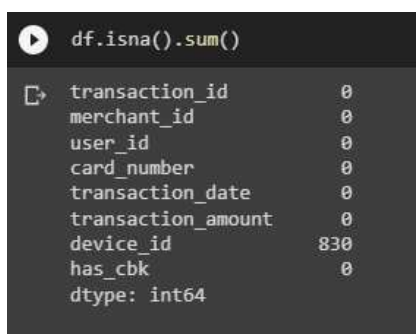
The feature formats are described in the image below:



```
df.dtypes
transaction_id    int64
merchant_id      int64
user_id          int64
card_number      object
transaction_date  object
transaction_amount float64
device_id        float64
has_cbk          bool
dtype: object
```

All the ids are classified as integers, as they should be. The card number is classified as an object due to the dataset saves the numbers, transforming six numbers in the middle to asterisks. The transaction date is incorrectly classified, and we will change it to the proper format before working with it. Transaction amounts and the chargeback flag are correctly formatted.

The database has 830 null values, all in the device_id column. We decided not to remove the rows in which those values where, as it is approximately 25% of all the rows in the dataset and the loss of information would be significant. Instead, we decided to fill the null values with a specific, not yet used id, 999999. The image below illustrates that:



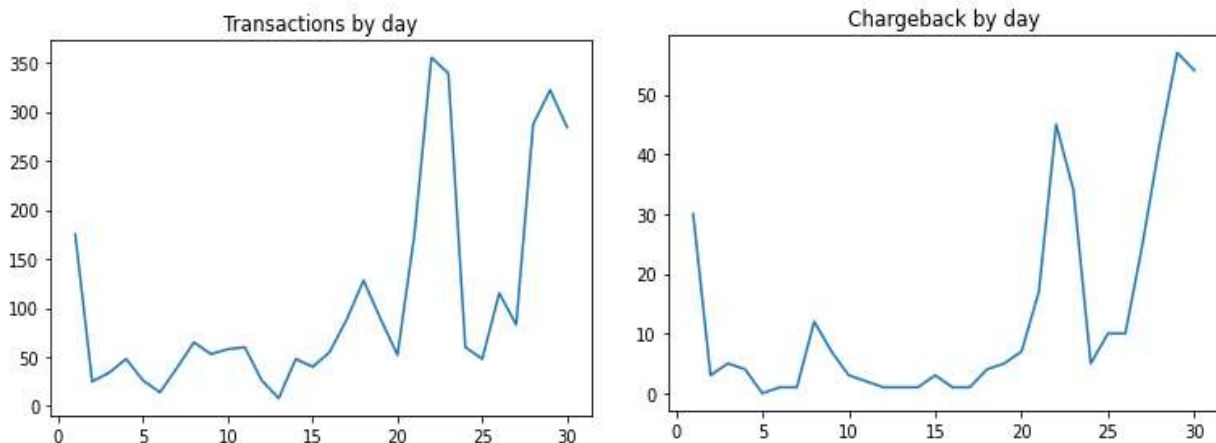
```
df.isna().sum()
transaction_id    0
merchant_id      0
user_id          0
card_number      0
transaction_date  0
transaction_amount 0
device_id        830
has_cbk          0
dtype: int64
```

1.2 Data Analysis

1.2.1 Plotting the chargebacks through date and time.

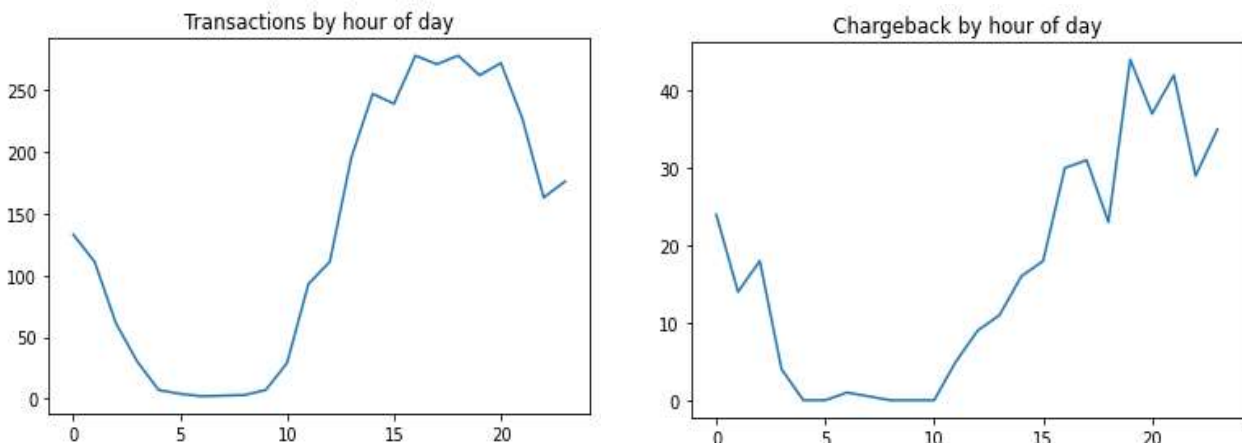
We will begin by using the transaction date and time to analyze the behavior of the chargeback transactions. We may be able to obtain valuable information through this analysis. First, we obtain the hour and day values of the transaction and plot the chargebacks through them.

The images below graph the amount of transactions and chargebacks by day:



As we can see, the majority of transactions happened on the period from day 20 to day 24, and then in the period between days 26 to 30. The chargebacks tend to follow this trend.

The following graphs show the number of transactions and chargebacks by hour of day:



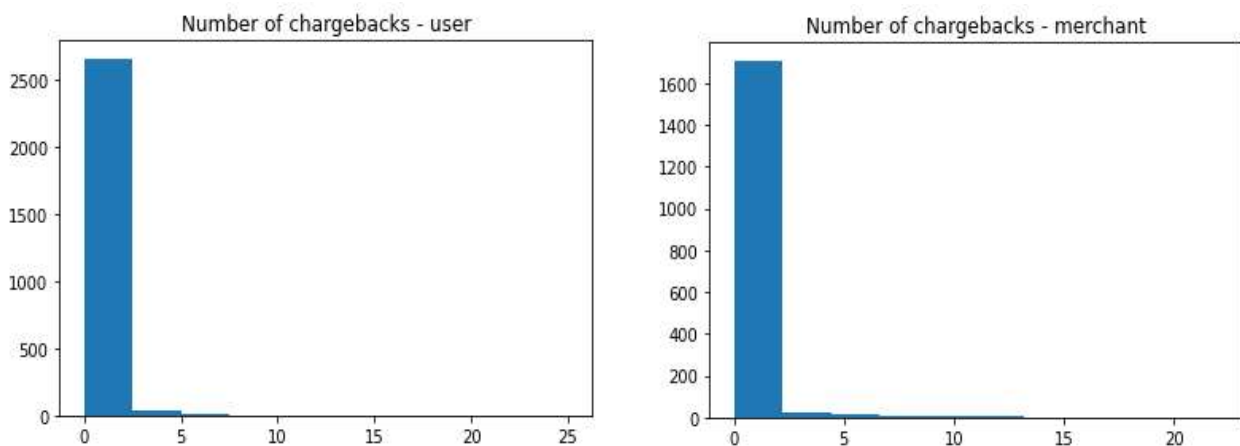
We can see that the number of transactions and chargebacks grows steadily after around 10 o'clock. Visually, we cannot perceive any meaningful difference between normal transactions and transactions that become chargebacks by looking at the date and hour: the amount of chargebacks rise because the number of transactions rise.

1.2.2 Working with the user's and merchant's information

As our attempt to identify trends through date and time was not successful, we will attempt to obtain information by performing a deep dive the two main players transaction: the user and the merchant.

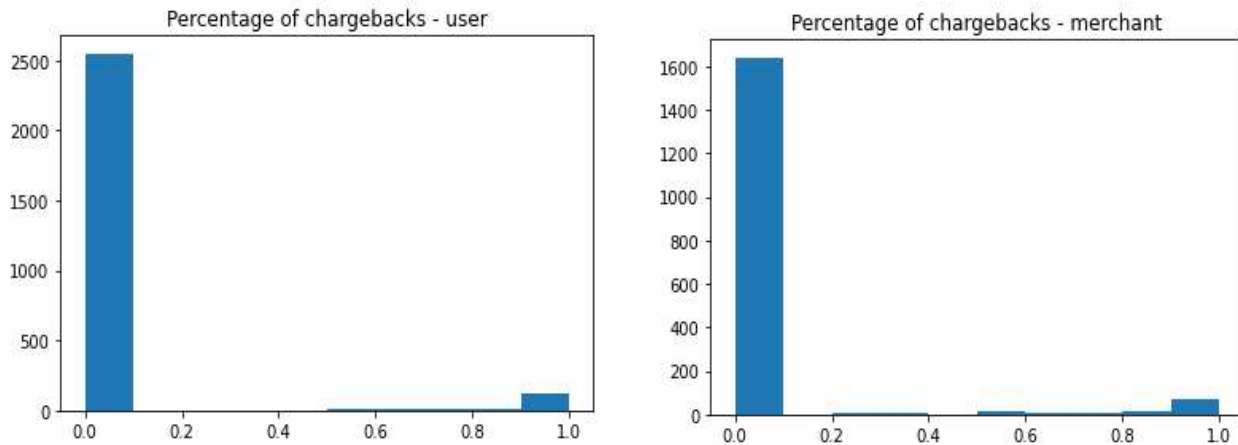
With the information from the previous database, we create a new database by grouping information of the users: each row will contain the user id, the amount of transactions the user has made, the sum and mean of all his transaction amounts, and the sum and mean of the amount of chargebacks the user has requested. The mean of the chargebacks will give us a decimal number between 1 and 0, where 1 means 100% of the user's transactions have been chargebacks. This will allow us to identify potentially problematic users in our database.

We will repeat the same process with the merchant's information, and then compare the results to see if we gain any insight from the new data. The histograms below compare the number of chargebacks by user and by merchants.



Most users, therefore, have very few chargeback requests, however, some users have up to 25 of those. While small in number, we can assume these users have malicious intent in their chargeback requests. The merchants have a bigger spread of chargeback requests than the users, however the number of merchants with few chargeback requests are the clear majority.

We will now look the chargeback rate, presented as a decimal number:



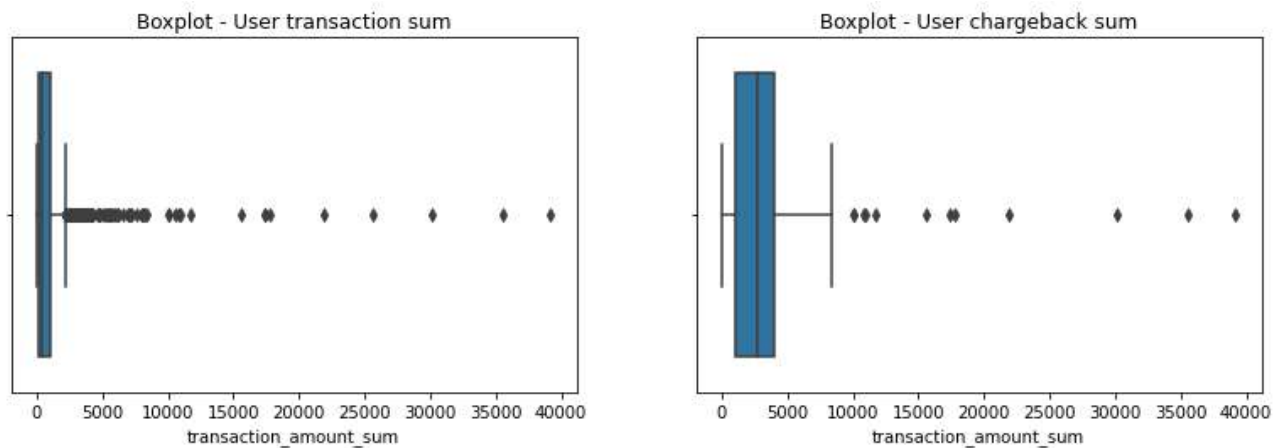
Therefore, it can be concluded that the overwhelming majority of chargeback requests are done by users between 0 to 10% of chargeback percentages. We can safely assume that most of these users have little to no malicious intent in their requests. However, we can clearly see that there are some malicious users with a percentage of chargebacks over fifty percent, a portion of which has between 90 to 100% of their transactions having chargeback requests. It would not be a bad idea to deny a transaction request of any user that surpasses a 10% chargeback rate.

Similarly, the majority of merchants have between 0 to 10% of a chargeback rate, but there is a small group with rates of 50% and more. There is a possibility these merchants commit an above average rate of error, engage in misleading and false advertising, and/or deny any refund request done by users. It would be prudent to deny transactions to these merchants aswell.

1.3 Chargeback and outliers

We should now analyse the transaction outliers, compare them to the chargeback outliers, and see if we can obtain any insight or additional information based on this analysis. It stands to reason that a friendly or criminal fraud would tend to be on the higher price tag, and we would like to verify if that's the case.

The following graphs show the boxplot of all transactions and the boxplot of only the transactions that had chargeback, respectively.



From this, we can infer that the chargebacks do, in fact, have a higher amount of money transacted than the overall data. It may be due to malicious intent, such as criminal or friendly fraud, or it may simply be because clients would be more demanding of those purchases with higher prices.

In conclusion, we have identified that the majority of chargeback cases are not malicious in nature. The chargebacks over the days and hours follows the transactions: in this dataset, there is no significant day nor hour in which chargebacks more likely to occur. The majority of users request a chargeback around 10% of the time or less, although there is a visible group of users that are malicious in nature, with a chargeback rate over 50%, some reaching 100%. Most merchants also have low chargeback rates, but there is a minority with a high rate. As such, while we would benefit from protecting ourselves from criminal and friendly fraud, it would be more beneficial to dedicate resources to reducing the chargeback rates of those clients who request them sparingly, for that is the greater mass.

1. 3. 2 In addition to the spreadsheet data, what other data would you look at to try to find patterns of possible frauds?

Data regarding the user(cardholder): Age, gender, number of children and/or financial dependents, social status (married, single, etc), occupation, income, property, place of residence, financial history. These features could help in finding patterns of frauds, from criminal frauds to friendly fraud. For example, it is more likely that unauthorized purchases are made in families with many children, big purchases by people who don't have a matching income are probably criminal fraud, etc.

Data regarding the merchant his goods and his industry: The size of the business, the reach of the business, what sector of the economy he belongs to. It would be possible that some businesses and industries are more likely to be preyed by fraud, criminal or otherwise. Knowing the rate of chargebacks in a sector of the economy can help in identifying trends of fraud.