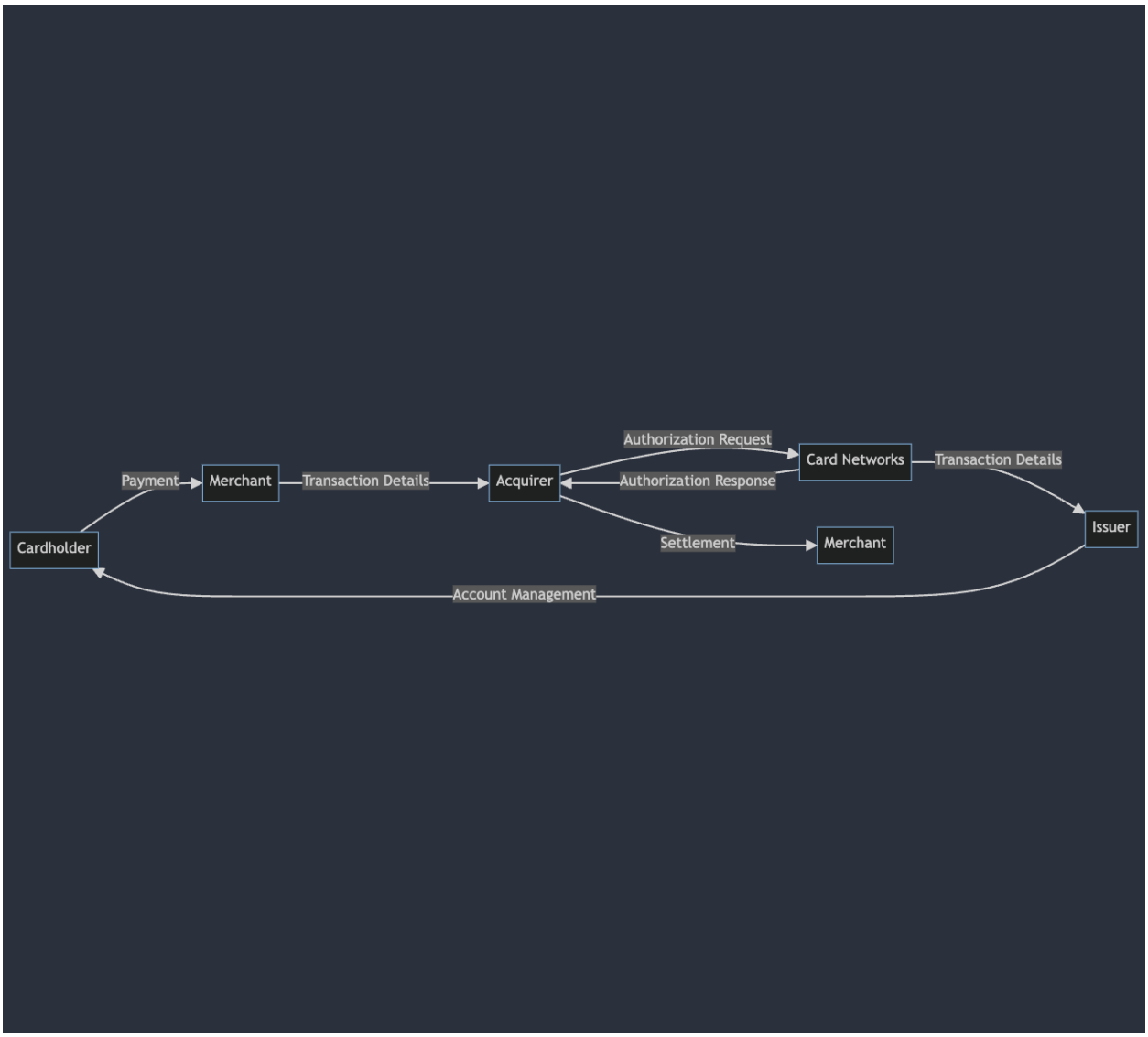
3.1 Understand the Industry

Answers

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

There are several key players involved in both the money flow and information flow. Below is a chart that represents the flow of a payment:



In this chart:

- The money flow starts with the cardholder making a payment to the merchant using a payment card.
- The transaction details are sent from the merchant to the acquirer.
- The acquirer sends an authorization request to the card networks, which then communicate with the issuer to verify the transaction details and the cardholder's account status.

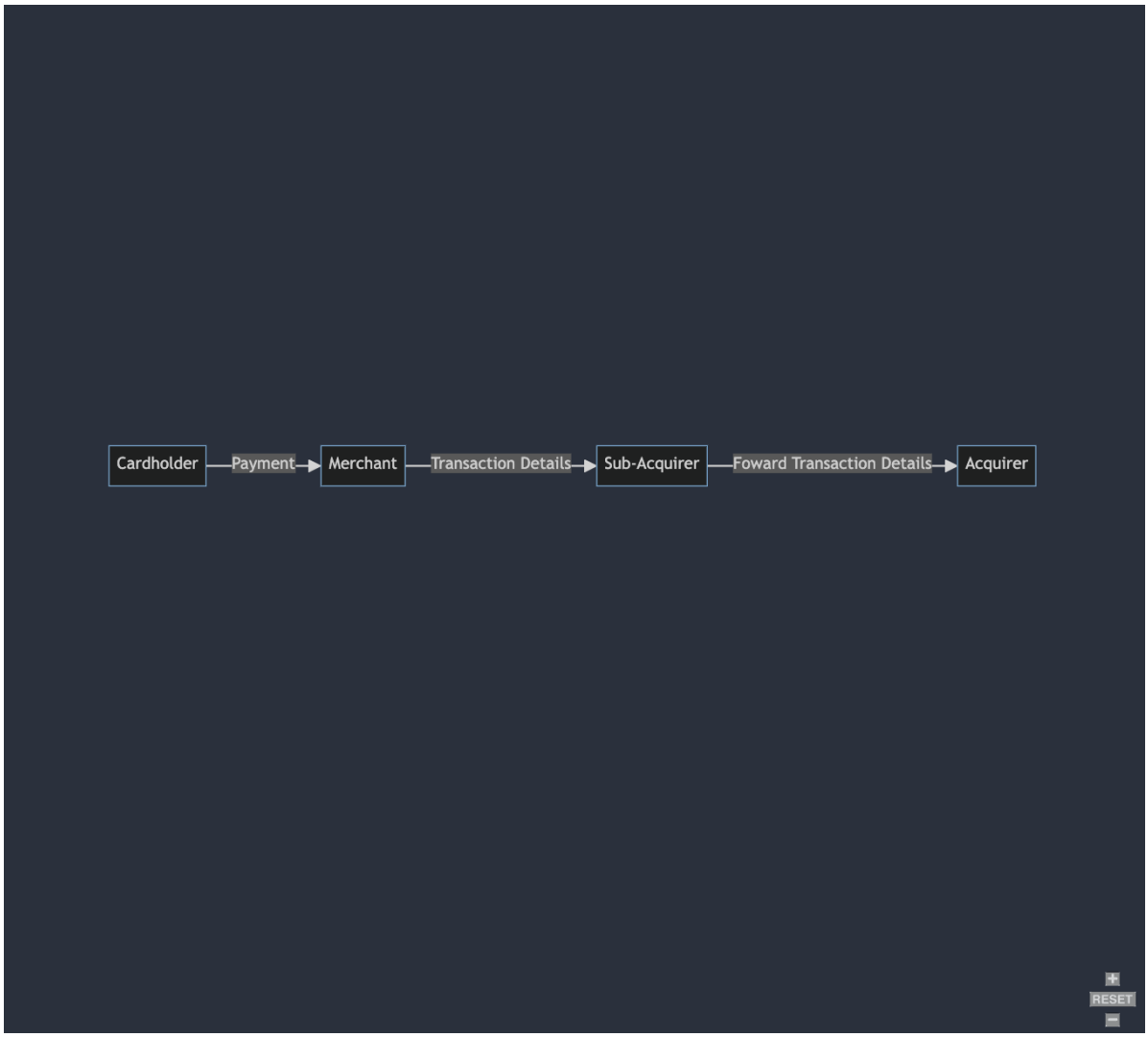
- Upon receiving the authorization response, the acquirer informs the merchant whether the transaction is approved or declined.
- After the transaction is approved, the settlement process begins, where funds are transferred from the issuer to the acquirer and then to the merchant.
- Additionally, transaction details may be sent from the card networks to the issuer for account management purposes, such as updating the cardholder's account balance or transaction history.

The key players are represented by the boxes

1. CardHolder: The customer who initiates the whole flow with a payment
2. Merchant: The business or entity that sells goods or services and accepts payment cards as a form of payment from cardholders.
3. Acquirer(Also referred to as an acquiring bank or merchant acquirer): A financial institution or company that processes credit or debit card transactions on behalf of merchants. They play a important role in the whole process as they:
    a. Establishing and maintain the merchant accounts, which allow businesses to accept card payments
    b. Facilitate the Authorization and settlement of the transactions between the Business and the Card Issuer
    c. Assumes the risk of chargebacks, frauds, and other disputes over the card transactions
4. Cards Networks: These are organizations such as Visa, Mastercard, American Express, and Discover that act as intermediaries between card issuers and acquirers. They facilitate the flow of transaction information and set the rules and standards for card transactions.
5. Issuer: The actual financial institution that are responsible for the cardholder account. Providing credit or funds managements through the issued card

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

- Acquirer as explained before are the companies that are responsible for the processing of the cards transactions on Behalf of merchants. Some examples in Brazil: Cielo, Rede, Stone
- Sub-Acquirers are companies that sit between the Acquirer and the merchants, most of the times bringing more capillarity to the business, which facilitate part of the process of getting merchants for the Acquirers. Some examples in Brazil: Stone, PagSeguro



- Payments Gateway are service providers that facilitates the secure transmission of transaction data between merchants and acquirers. They encrypt and transmit securely the needed informations through the internet. In the flow presented into the first answer, they would be connected to the merchants sites/point-of-sales and act altogether as sub-acquirers. They could too integrate multiple Adquirers into the solution provided, bringing a lot of value for the merchants. Examples: PagSeguro, MercadoPago

Is important to note that a lot of companies act in multiple positions of the payment flow, acting sometimes as Acquirer, Sub-Acquirer and Payment Gateway.

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

A chargeback could be resumed as payment return to a cardholder when any of the follow occurs:

- Cardholders disputes of transactions
- Issuing Bank Fraud detections

They generally occurs when cardholders are billed wrongly or do not recognize transactions(fraudulent transactions) or even when they dispute the quality of services/products. The initiator of chargebacks are the Issuing Banks, in both cited cases

Cancelations are a form of payment return too, but they occur in totally different cases, cancelations generally are initiated by the merchant or caldholder when the transaction was not fully approved or settled.

3.2 Get your hands dirty

Disclaimer: I'm not used to doing data analysis, I am a backend engineer who didn't worked in fraud risk before, I tried my best to absorve the most I could while making the challenge to get up with this. In the process I had to learn how to deal with pandas and jupyter notebook.

1. Analyze the data provided and present your conclusions (consider that all transactions are made using a mobile device).

The Samples file contains 3199 entries, but of these, only 2369 have device_id filled in, representing 74.05% of the data.

```
]: transactions_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3199 entries, 0 to 3198
Data columns (total 8 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   transaction_id      3199 non-null   object 
 1   merchant_id         3199 non-null   object 
 2   user_id             3199 non-null   object 
 3   card_number         3199 non-null   object 
 4   transaction_date    3199 non-null   object 
 5   transaction_amount  3199 non-null   float64
 6   device_id           2369 non-null   object 
 7   has_cbk             3199 non-null   bool   
dtypes: bool(1), float64(1), object(6)
memory usage: 178.2+ KB
```

The analyses mentioned here will be done on the total file, so as not to lose a significant amount of data in terms of analysis.

Merchants

There are 1756 merchants within the data, but only 514 have more than one transaction.

```
[32]: grouped_merchants = transactions_data.groupby('merchant_id').size()
      grouped_merchants.sort_values(ascending=False)
```

```
[32]: merchant_id
      49205    73
      17275    30
      79698    22
      4705     22
      53041    19
               ..
      44519     1
      44464     1
      44379     1
      44336     1
      99799     1
      Length: 1756, dtype: int64
```

```
[35]: grouped_merchants[grouped_merchants > 1].sort_values(ascending=False)
```

```
[35]: merchant_id
      49205    73
      17275    30
      79698    22
      4705     22
      53041    19
               ..
      48376     2
      47252     2
      46666     2
      46609     2
      5557      2
      Length: 514, dtype: int64
```

As I searched around, I found out that 1% is an acceptable chargeback ratio for merchants. Our datasample brings a lot of merchants with high percentages of chargeback. Maybe it's because we have a small sample, but that got me to think in put this as a pre-determinated Rule in our service.

Some numbers:

[53]:

merchant_id	total_count	has_cbk
17275	30	73.333333
4705	22	86.363636
53041	19	73.684211
77130	15	86.666667
1308	15	100.000000
91972	14	78.571429
99510	12	58.333333
42356	12	66.666667
44927	11	100.000000
55854	11	81.818182
73271	10	100.000000
62052	10	30.000000
29214	10	90.000000
65330	10	80.000000

From the top 20 merchants of the datasample, 9 has high percentage(assumed 10% as high for our datasample) of charge_backs:

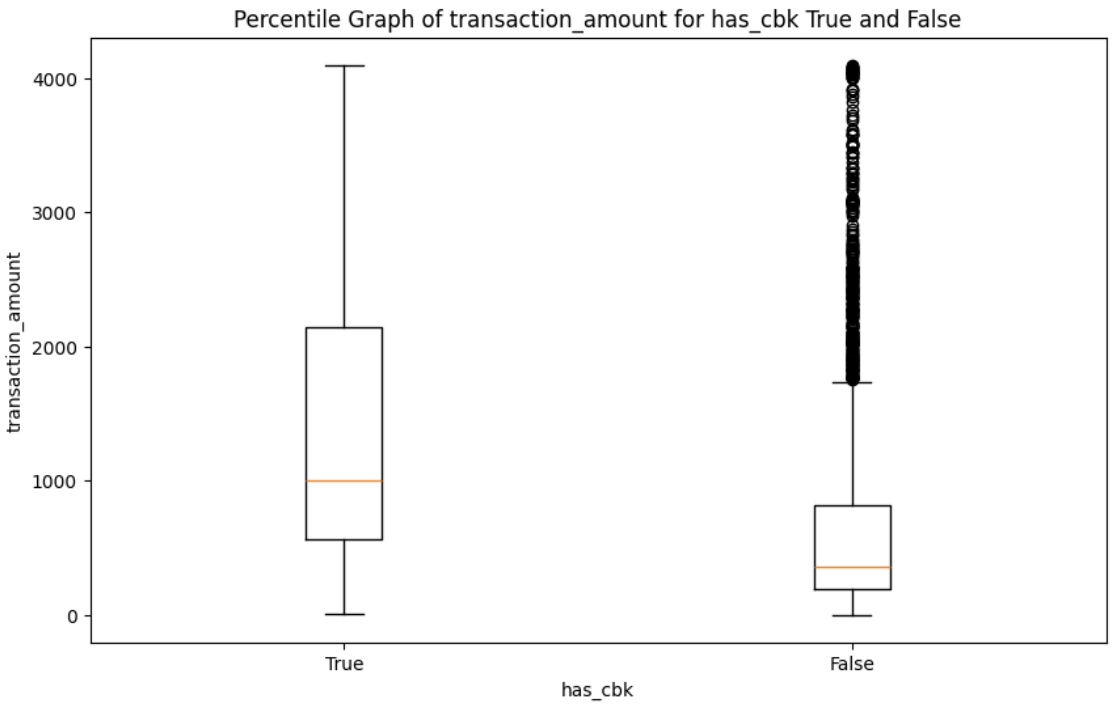
```
[79]: top_20_merchants.sort_values(by='has_cbk', ascending=False)
```

[79]:

merchant_id	total_count	has_cbk
1308	15	100.000000
44927	11	100.000000
77130	15	86.666667
4705	22	86.363636
91972	14	78.571429
53041	19	73.684211
17275	30	73.333333
42356	12	66.666667
99510	12	58.333333
29744	11	9.090909
76725	12	8.333333
42178	15	0.000000
4129	14	0.000000
26981	16	0.000000
36929	13	0.000000
17348	16	0.000000
26765	18	0.000000
29931	12	0.000000
79698	22	0.000000
49205	73	0.000000

Transactions amount

Now let's look at the values of transactions. Firstly I plotted the percentile graphs for transaction_amount:



It could be seen that the transaction_amount for chargebacked transactions is a lot more that the ones for non-chargebacked. For example let's see the 90th percentile of non-chargebacked vs the 70th of chargebacked:

has_cbk

```
[97]: cbk_true_percentile = cbk_true['transaction_amount'].quantile(0.7)
      cbk_false_percentile = cbk_false['transaction_amount'].quantile(0.9)

      print("70th percentile of 'transaction_amount' in cbk_true:", cbk_true_percentile)
      print("90th percentile of 'transaction_amount' in cbk_false:", cbk_false_percentile)

      70th percentile of 'transaction_amount' in cbk_true: 2022.18
      90th percentile of 'transaction_amount' in cbk_false: 1787.8170000000014
```

We can also compare the median of the two groups, which shows a high difference:

```
[98]: cbk_true_median = cbk_true['transaction_amount'].median()
      cbk_false_median = cbk_false['transaction_amount'].median()

      print("median of 'transaction_amount' in cbk_true:", cbk_true_median)
      print("median of 'transaction_amount' in cbk_false:", cbk_false_median)

      median of 'transaction_amount' in cbk_true: 999.47
      median of 'transaction_amount' in cbk_false: 360.315
```

So this shows that fraudulent transactions tends to have higher values in comparison with the non-fraudulent. Which shows that a anti-fraud system should look from great deviations of transactions values from the norm, specially consideration history of transactions by user.

I assume that learning the daily and weekly patterns of users/merchants could help with this calculation to prevent chargebacks

Users

A lot of Users are reincident with chargebacks, not only that but into the top 10 users, all of them had more than 37% of they transactions chargebacked:

```
[100]: user_id_with_cbk_percentages = transactions_data.groupby('user_id')['has_cbk'].agg(total_count='count', has_cbk=lambda x: (x.sum() / len(x))

       user_id_with_cbk_percentages.head(10)
```

[100]:

user_id	total_count	has_cbk
11750	31	80.645161
91637	22	86.363636
79054	17	88.235294
96025	14	92.857143
78262	13	92.307692
75710	10	100.000000
56877	9	55.555556
9853	9	44.444444
11452	8	37.500000
7725	7	100.000000

From this datasample, we could conclude that rules that penalize reicident Users are helpfull to the company.

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

frauds?

- Geolocation Data:
  - Transactions made from diferente geolocations than the User home address
- IP addresses
- Merchant Categories with high cbk percentages