FRAUD RISK ANALYSIS: MERCHANTS, USERS AND CHARGEBACKS

DATA-DRIVEN INSIGHTS TO STRENGTHEN FRAUD PREVENTION STRATEGIES

CONTEXT AND **OBJECTIVE**

Analyze transactions data, in order to find suspicious behaviors

Understand patterns and propose an efficient anti-fraud solution for the case

Set and identify fraudsters characteristics

A LITTLE LOOK AT THE **DATABASE**

Transactions Quantity	Merchant Quantity	Chargeback Rate
3.199	1.756	12.2%

Our given database has valuable macro information

With the data, we can understand a little more of what happend

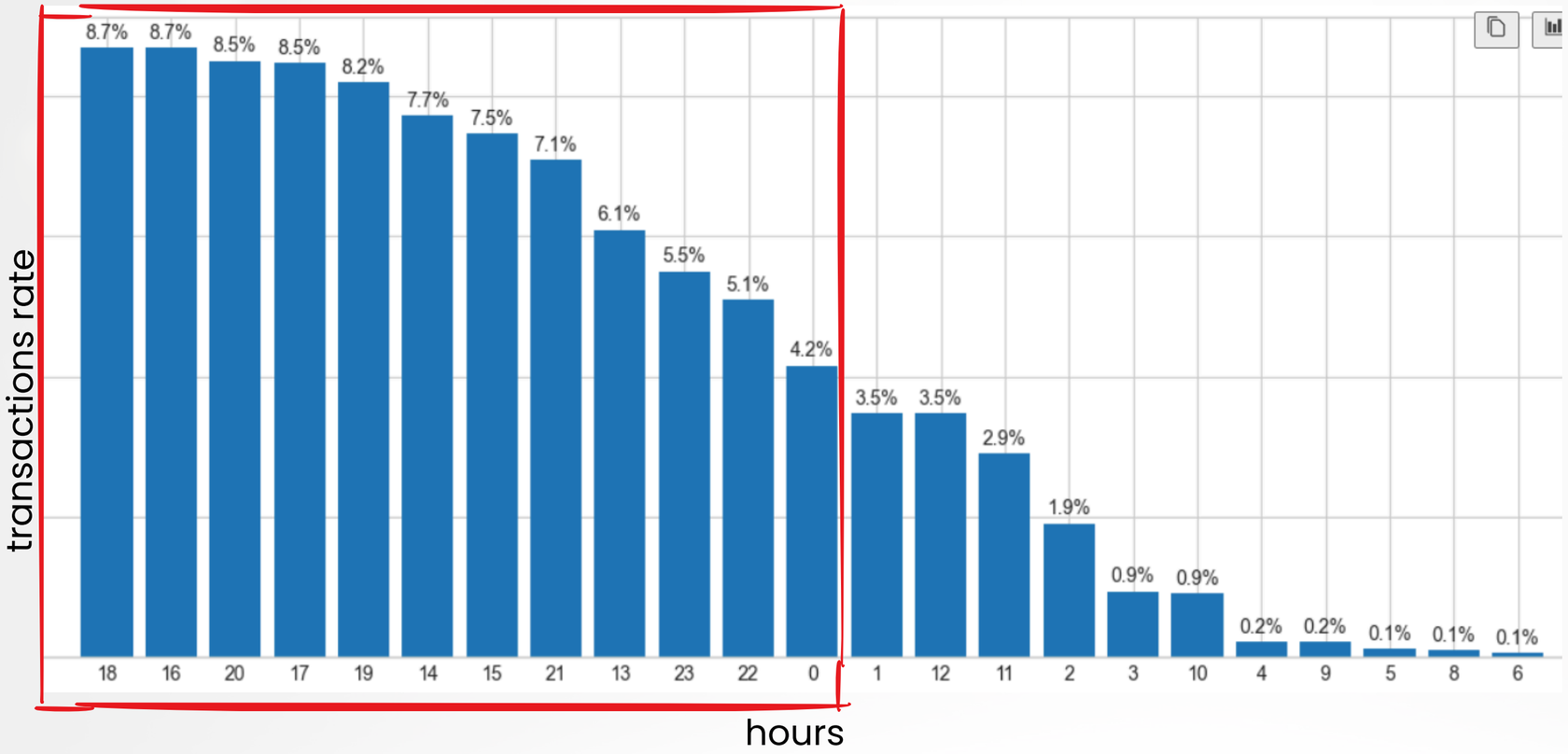
START
QUESTIONING

Based on the initial data, we proposed hypothesis that led our analysis.

When suspicious behavior often occurs?

We assumed that suspicious behavior occurs, in general, at night during non-commercial hours

80% TRANSACTIONS



HOUR
ANALYSIS

Hourly analysis reveals that **80%** of transactions occur at various times.

To better understand this, we classified the hours.

HOUR LABELED

Morning	Afternoon	Evening	Dawn
0	5	6	1

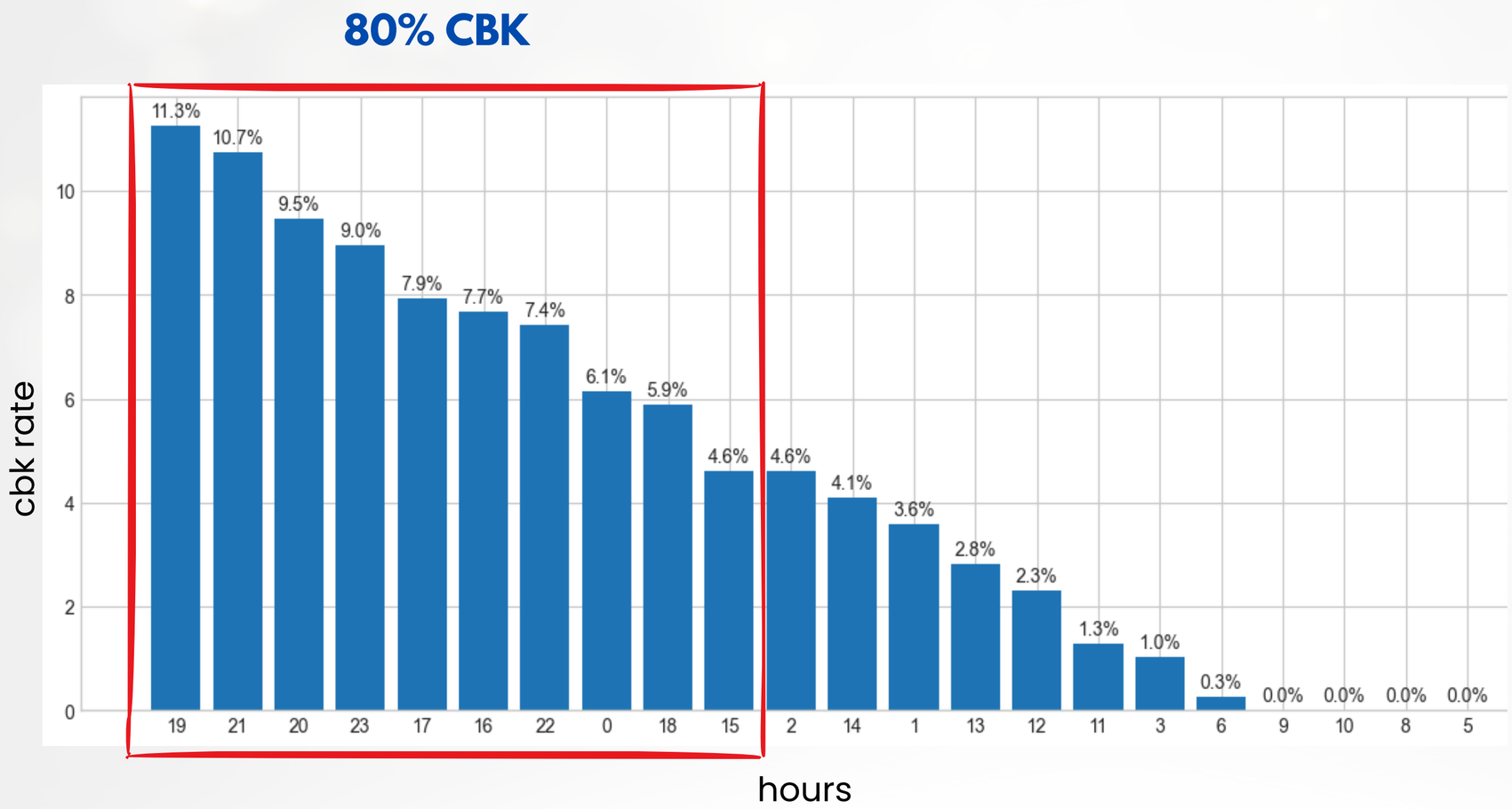
The analysis reveals that 80% of transactions occur in a continuous block that runs from early afternoon (1:00 PM) until midnight.

Activity peaks predominantly in the evening, suggesting a strong after-hours pattern.

The complete absence of significant transactions in the morning indicates a window of low activity.

But, those informations alone, don’t give us many insights

So we applied the same analysis to Chargebacks



HOUR **ANALYSIS (CBK)**

The analysis reveals that **80%** of chargebacks are also concentrated within a specific set of hours.

Therefore, these hours should be classified for clearer monitoring

CBK HOUR LABELED

Morning	Afternoon	Evening	Dawn
0	3	6	1

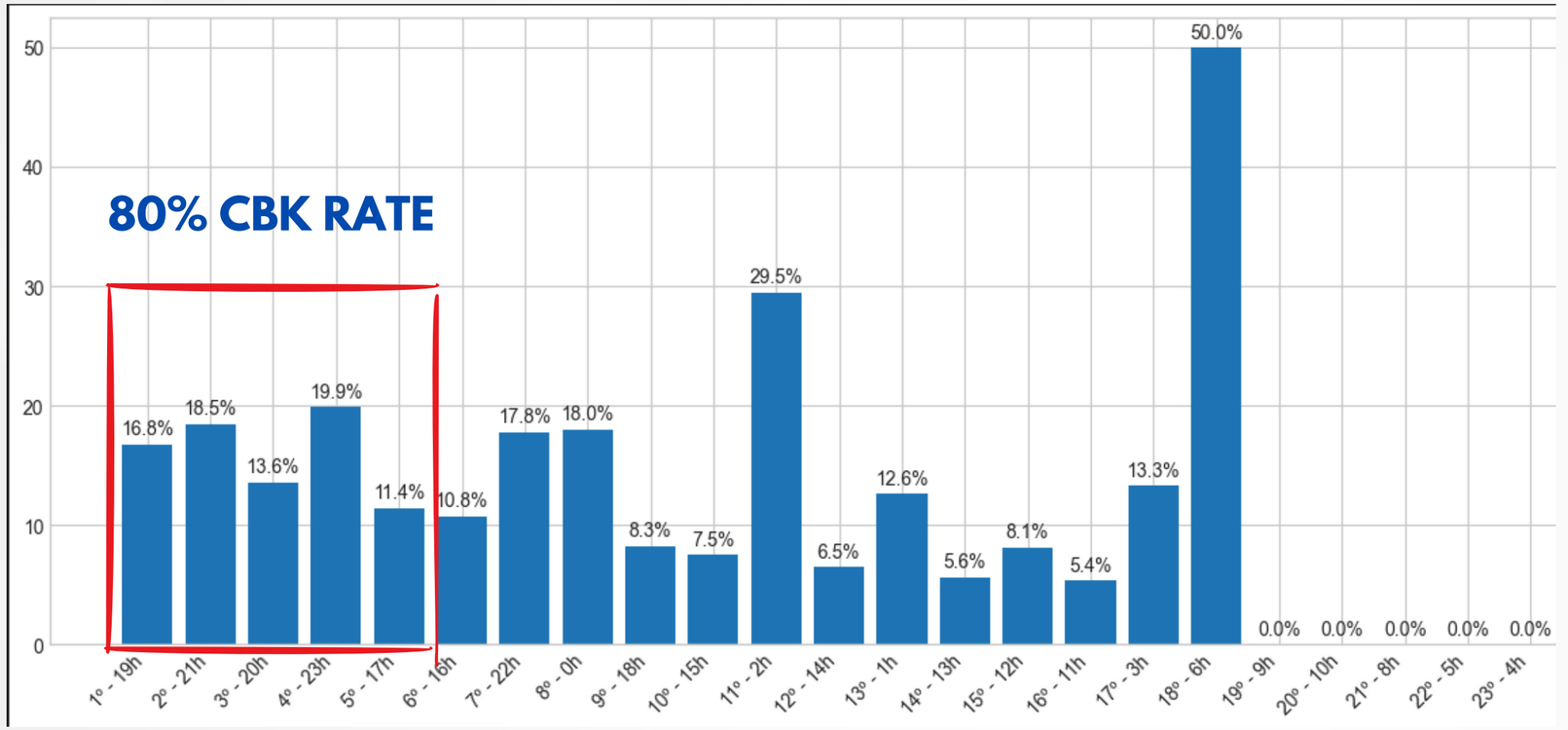
Similar to the general transaction pattern, chargeback analysis also shows a strong concentration in the afternoon and evening.

However, the chargeback pattern is even more skewed toward the end of the day, with a high-risk block beginning at 3:00 PM and intensifying massively after 6:00 PM.

To add more depth

to our findings, the next step is to analyze the hourly chargeback rates, both by transaction count and by monetary value.

This will reveal which hours are proportionally riskier, not just those with high volume



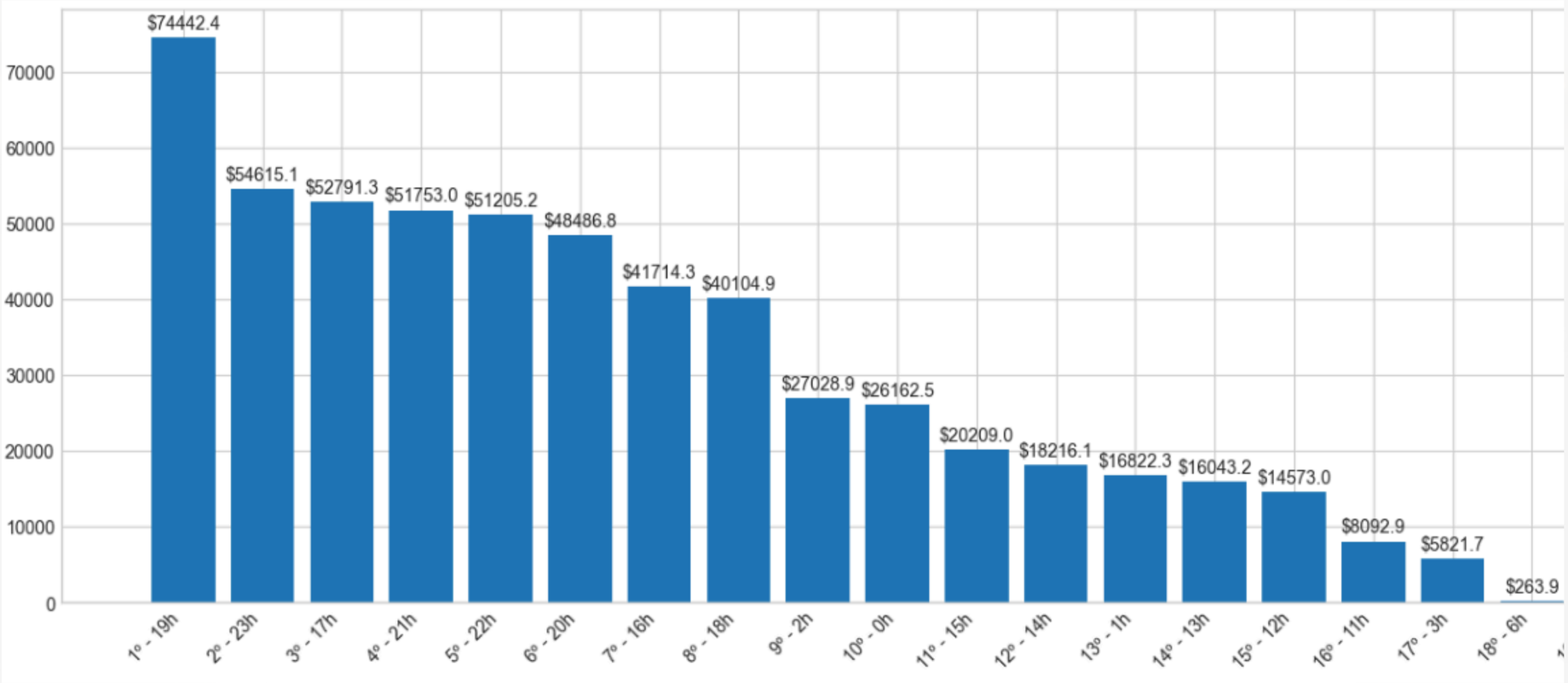
HOUR **ANALYSIS (CBK RATE)**

In this analysis, we check the CBK rate per hour, **ordered by the amount of CBK.**

We can see that **80%** of the CBK rate is concentrated at night.

But, in addition to the frequency of chargebacks, it is crucial to analyze their monetary value.

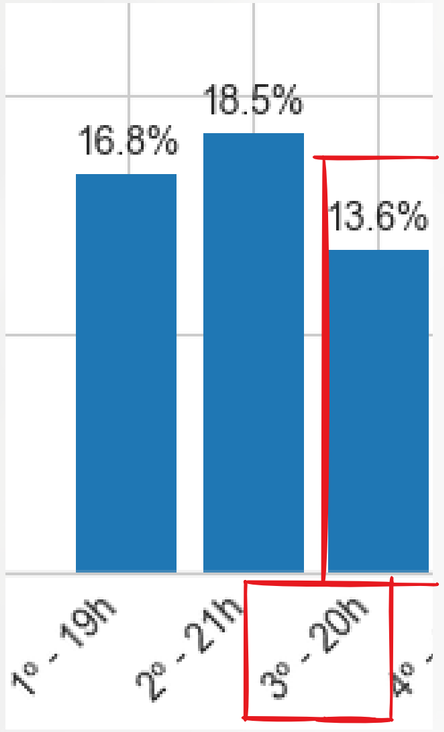
HOUR
ANALYSIS (CBK MONETARY)



While the chargeback rate reveals the frequency of fraudulent incidents, the monetary rate exposes their severity.

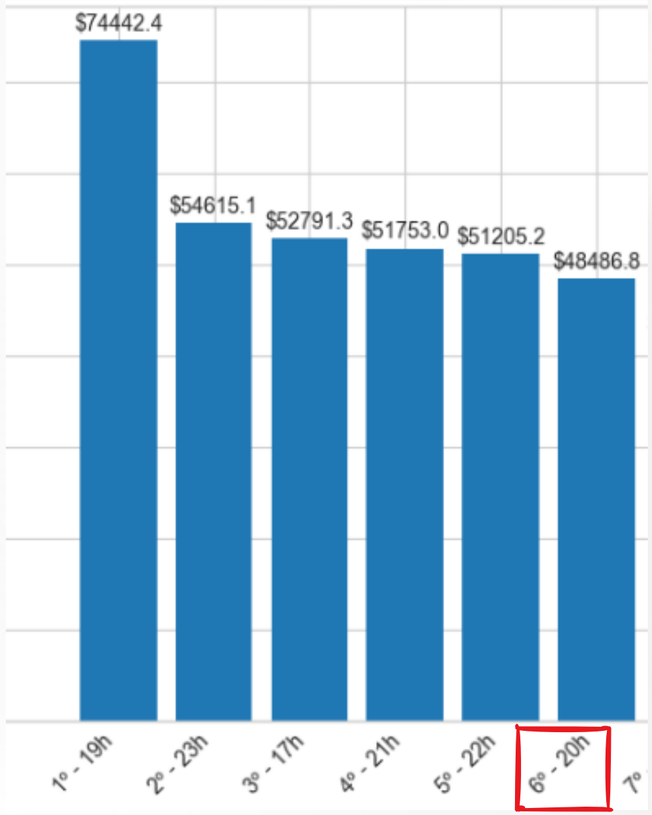
Analyzing both is essential for a complete understanding of our financial risk exposure.

Why it works?



For example, while the 8 PM hour ranks 3rd for the number of chargeback incidents ...

... it drops to 6th place when ranked by total financial loss



TIME
ANALYZING

When suspicious behavior often occurs?

By cross-analyzing transactions and chargebacks, we partially validated the hypothesis that suspicious activity occurs outside of business hours.

However, the key insight is more subtle:

Fraud follows volume, not a vacuum. Fraudsters focus their activities during peak periods of legitimate user traffic—late afternoon and evening.

This may suggest a deliberate strategy to camouflage itself among the high volume of transactions.

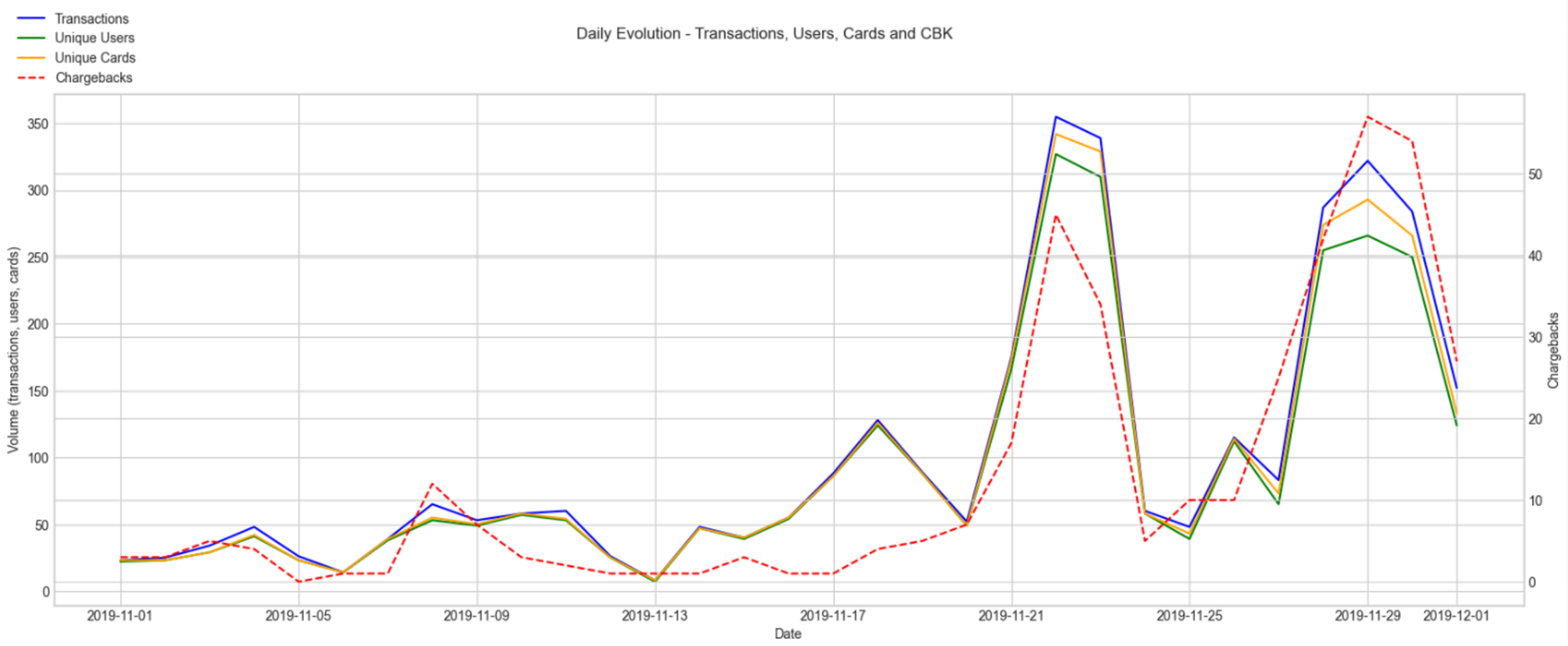
RISK FOLLOWS VOLUME:
FROM DAILY TO SEASONAL PATTERNS

In our previous analysis, we discovered a clear pattern: 80% of chargebacks occur during the busiest times of the day – late afternoon and evening.

Now, when we analyze the daily evolution throughout the month, we see this same principle repeat itself on a macro scale.

Now, when we analyze daily trends throughout the month, we see this same principle repeat itself on a macro scale.

This chart shows that days with peak transaction numbers are also days with peak chargeback numbers.



Note the strong correlation between the transactions line (blue) and the chargebacks line (red).

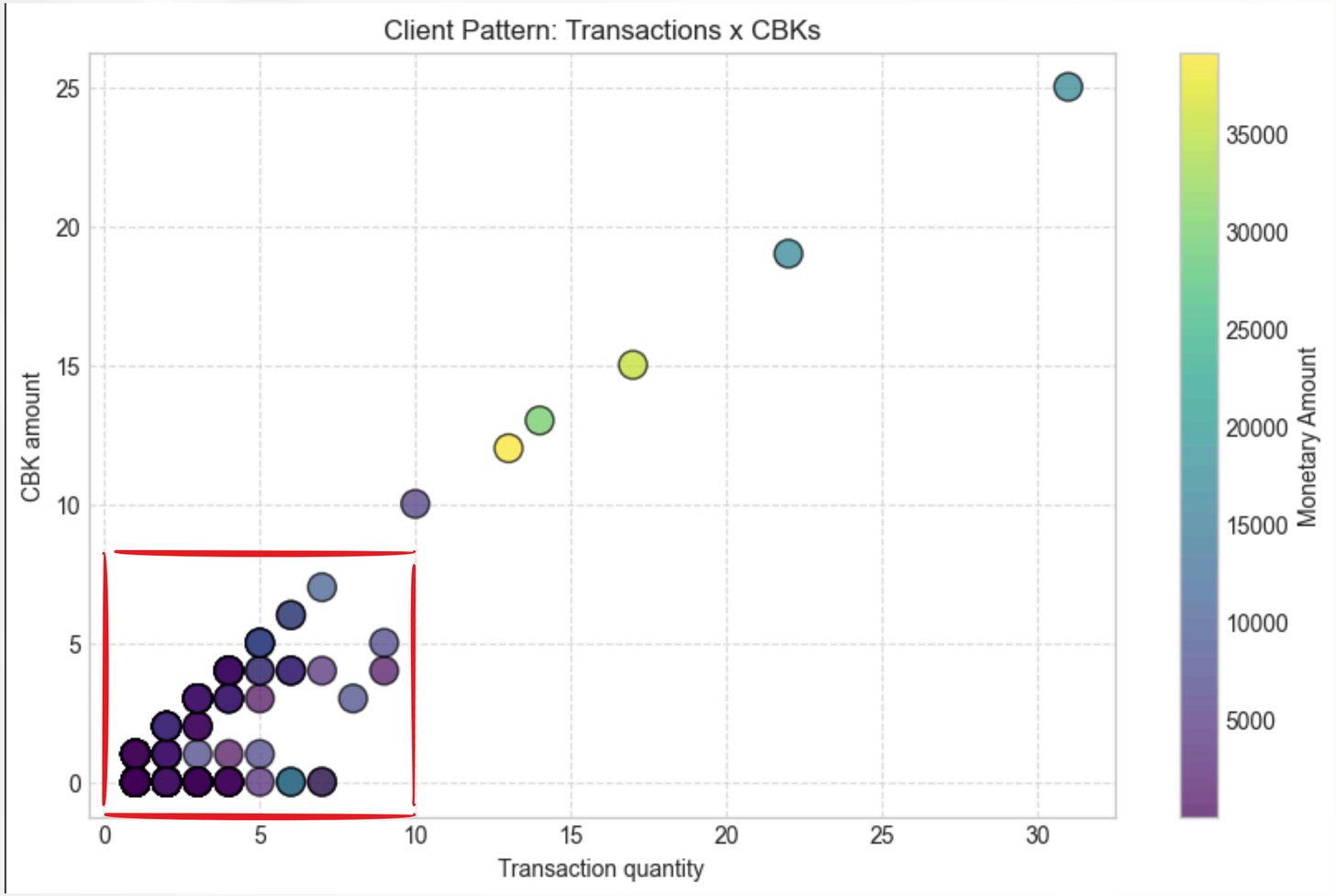
When sales volume spikes, such as during the peaks around **November 22nd** and especially **November 29th** (Black Friday Week), the chargeback volume rises correspondingly.

START
QUESTIONING

Moving beyond the timing of fraud, our analysis then shifted.

How the users behave?

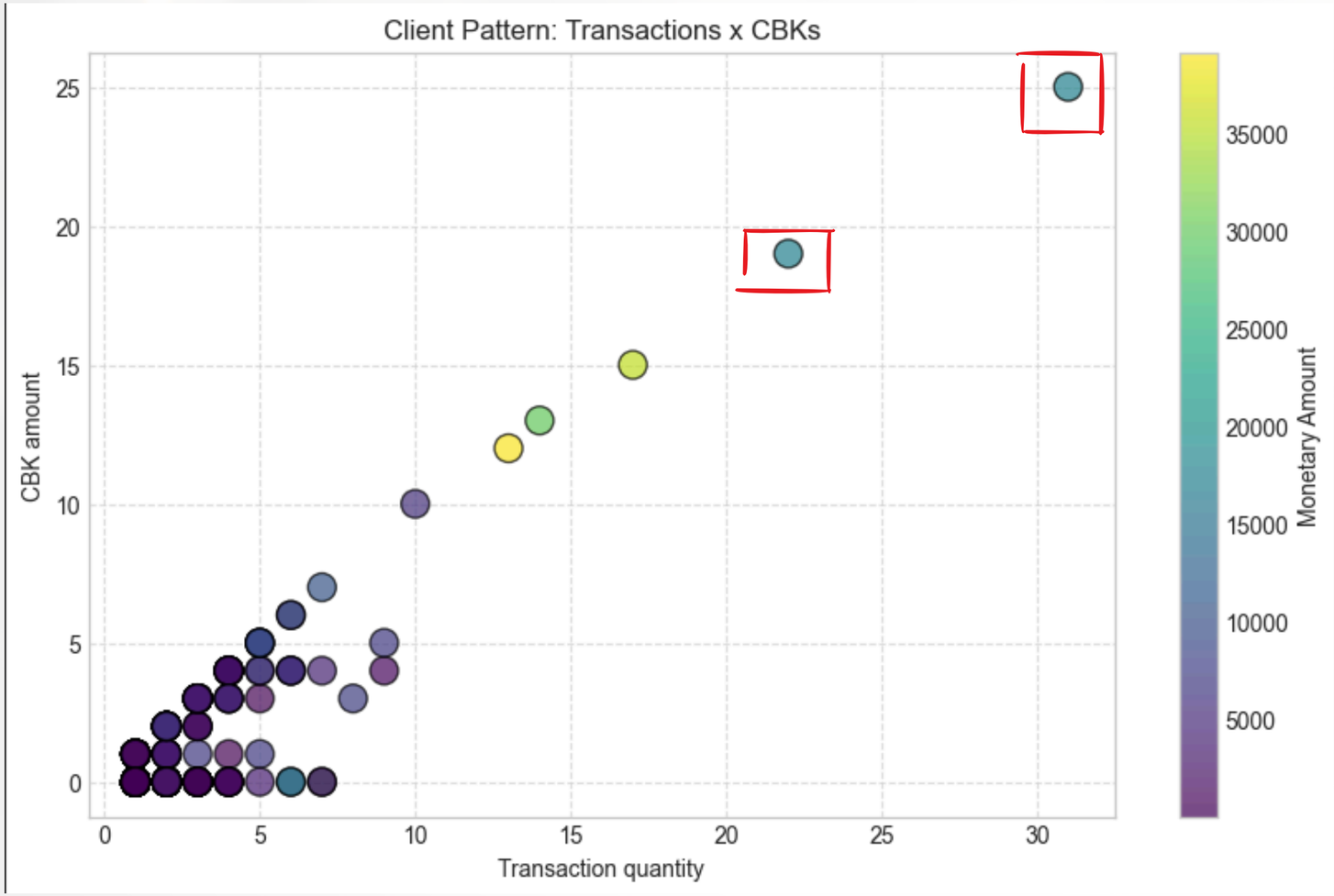
We want here, to get some pattern from the users



DEFAULT BEHAVIOR

By visualizing users on a graph that intersects the number of transactions with the number of chargebacks, we identify a clear pattern.

The vast majority of our base is concentrated in the lower left corner, characterizing the typical customer profile: low transaction volume and, consequently, minimal or no chargebacks.

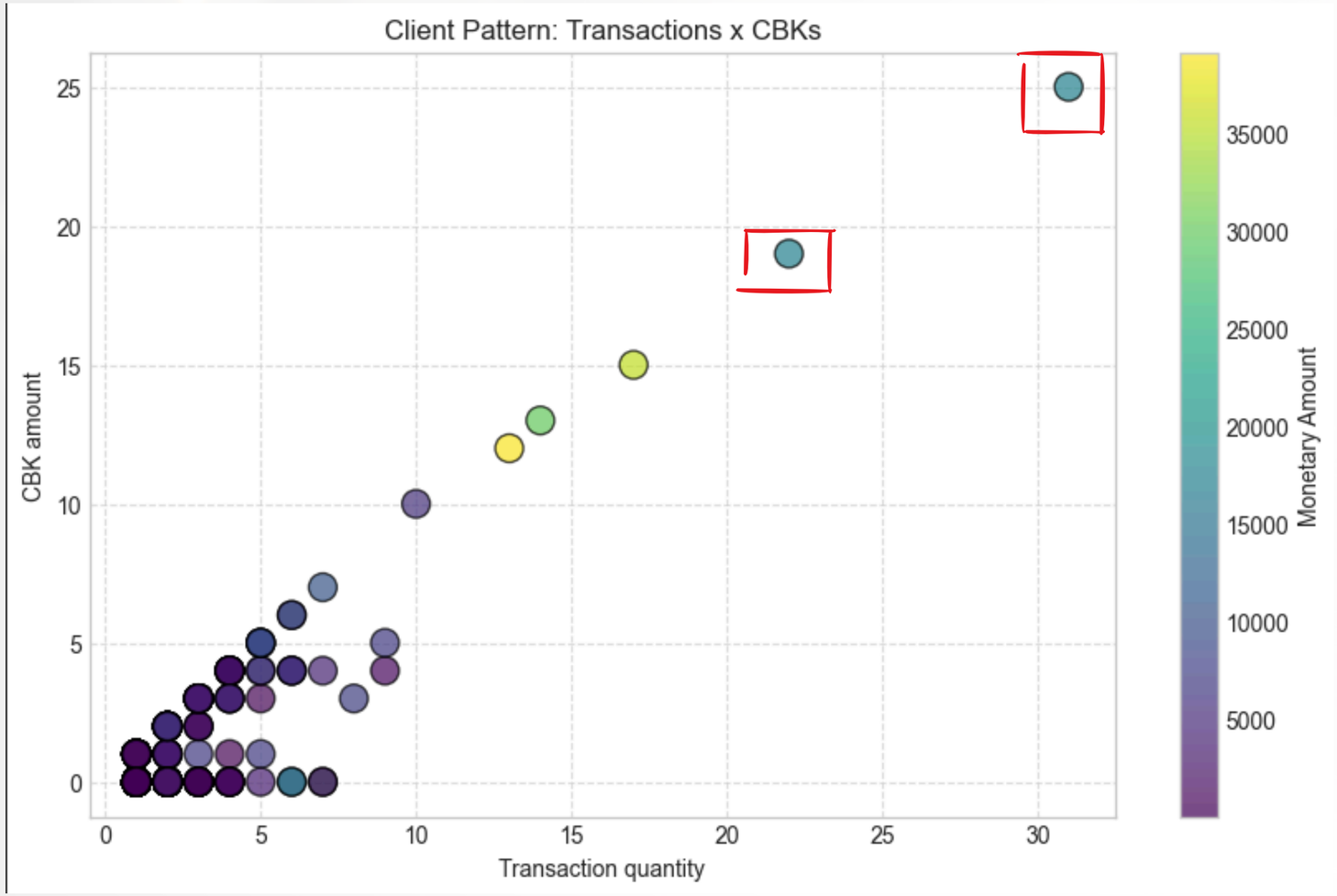


HIGH RISK ANOMALIES

Our attention turns to isolated points that deviate from this main cluster.

Users located further to the right (high transaction volume) and/or higher (high chargeback volume) represent atypical and highly suspicious behavior.

The combination of these two factors, especially in the upper right quadrant, is a classic indicator of fraudulent accounts that warrant immediate investigation.



POSSIBLE FRAUD CLASSIFIER

The color scale, which represents monetary value, adds a crucial third dimension that allows us to qualify the risk. With it, we can generate hypotheses about the type of fraud:

High Impact Fraud: We observed that the most suspicious points (top-right) also tend to have the hottest color (yellow), indicating that the accounts with the most fraudulent behavior are also those that generate the greatest financial loss.

Card Testing Hypothesis: On the other hand, a user with many transactions/chargebacks (shifted to the right/up), but with low monetary value ("cold" color, purple/blue), would be a strong candidate for card testing.

The Conclusion

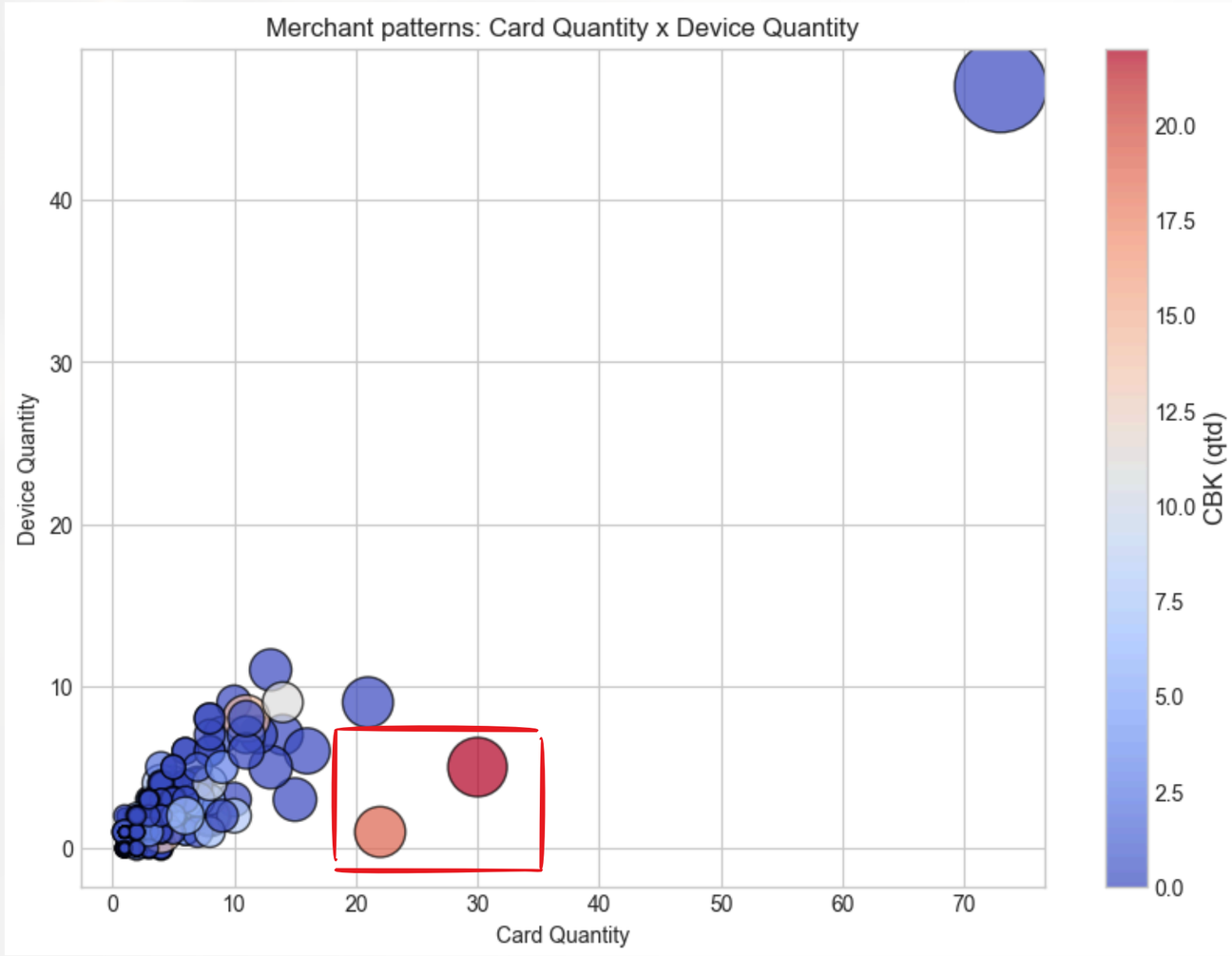
we can see from the graph is that potential fraudsters are some of our base points. And that, in this case, they can generate a large financial loss.

START
QUESTIONING

Moving beyond the timing of fraud, and users patterns ...

Where are the suspicious behaviors?

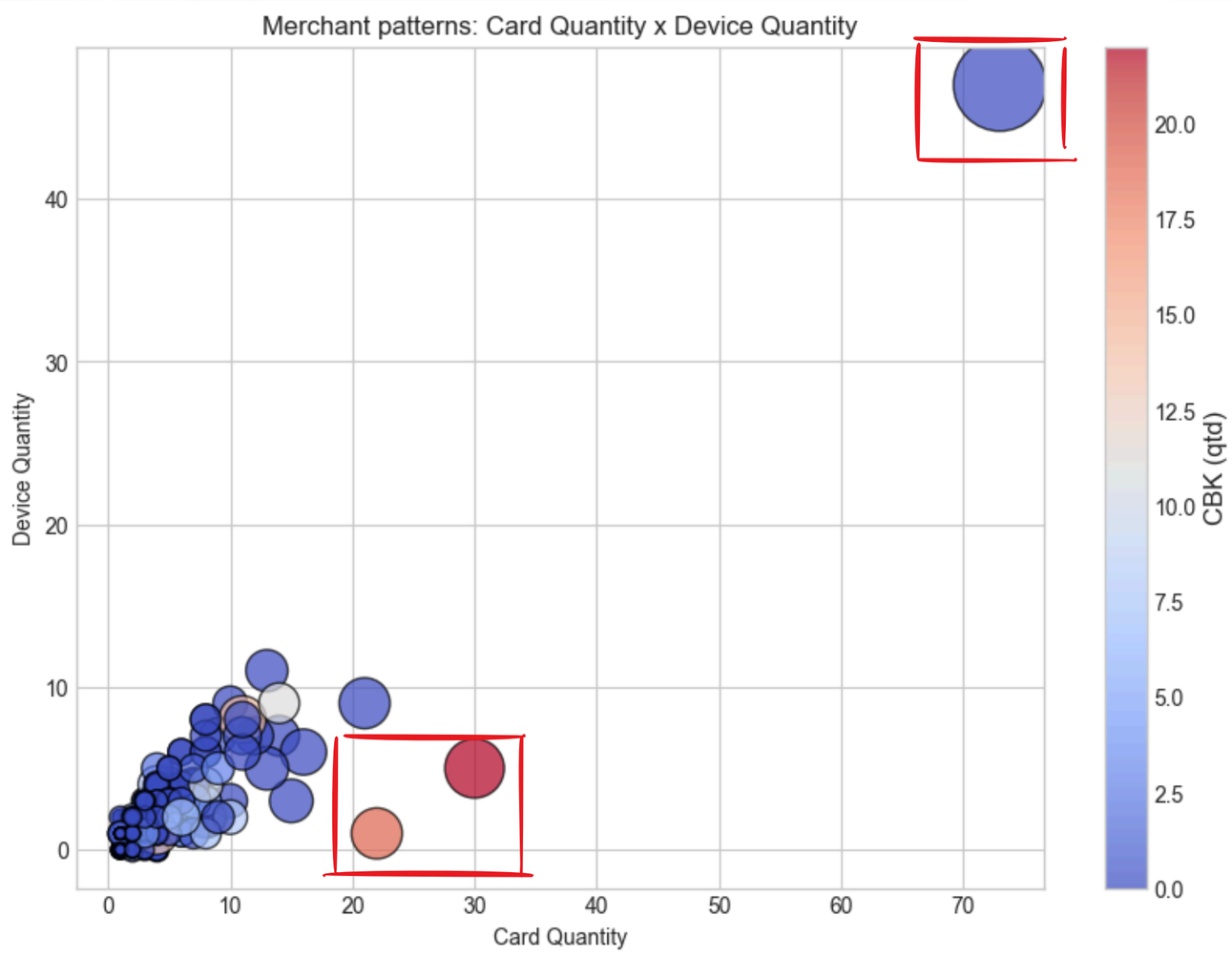
Where do these suspicious behaviors tend to occur?



MERCHANT ANALYSIS

This graph shows merchant behavior in relation to the number of different cards, devices and chargebacks.

We can quickly identify that the reddest spots are those that demonstrate more suspicious behavior, as they indicate a high amount of CBK.

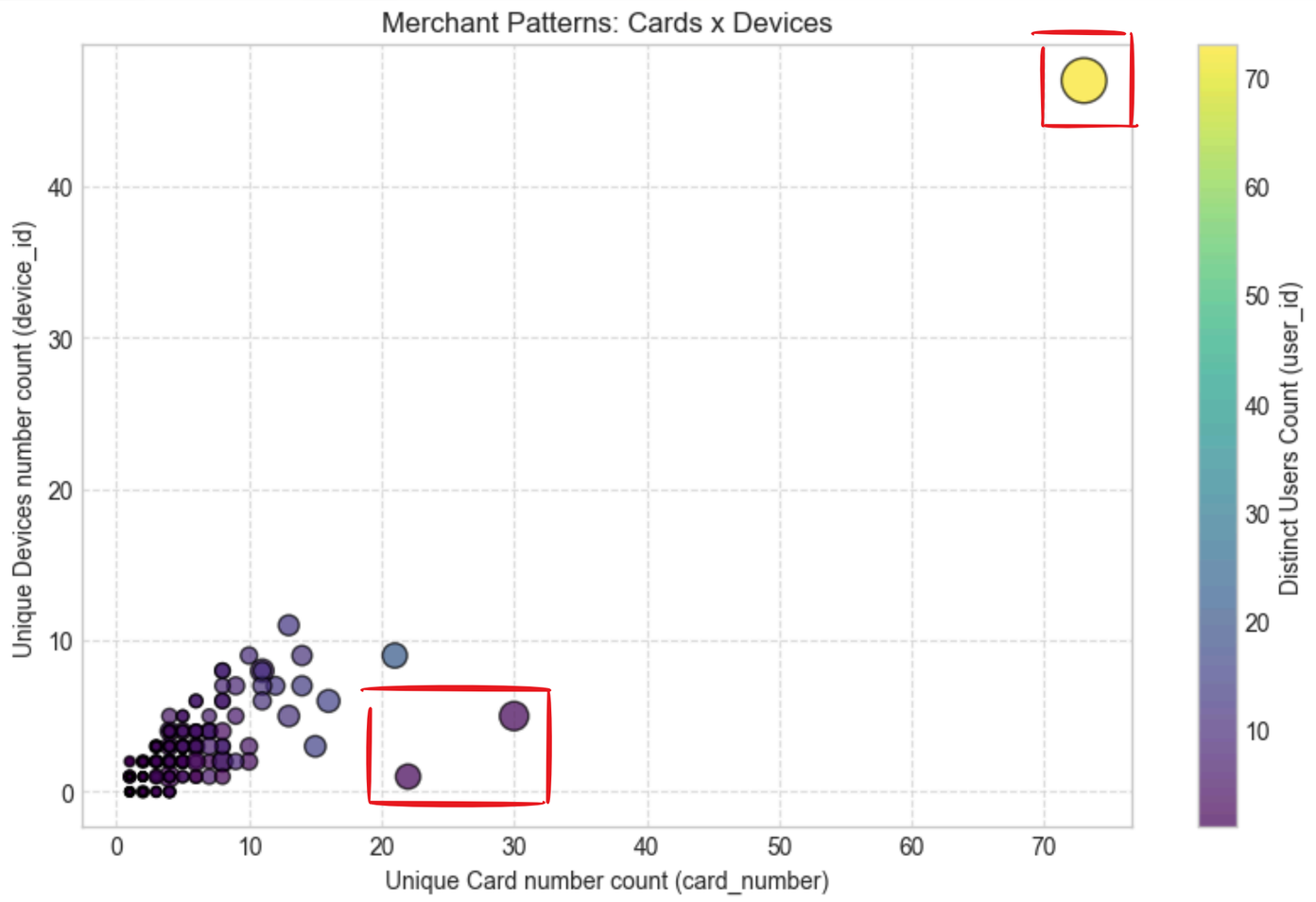


MERCHANT ANALYSIS

We can also immediately identify an outlier, where a large number of distinct cards and devices are concentrated, in addition to a high number of transactions.

However, there is no high CBK rate.

Therefore, a new variable must be added to understand whether there is a risk in these cases.



MERCHANT ANALYSIS

In this view, we can pinpoint merchants with higher risk: few users (color scale), but many distinct cards and few devices.

This highlights merchants where risk is concentrated in a small user base.

On the other hand, some outliers may look unusual but are not high risk, since their large number of cards, users, and devices suggest a more normal behavior.

What could increase the analysis?

Our initial database was quite macro information, besides that, we were able to extract valuables insights that led us to further possible solutions.

On the other hand, it is possible to increase the analysis with more information that could be extremely useful and handy.

IP Address: The IP address of the device that performed the transaction.
User Registration Date: The date the user account was created
Card Country Issuer: Country of card issuance
Country: IP derived from country
Billing Address: Billing address associated with the card
Shipping Address: Product delivery address

With this information, we would be able to carry out more detailed analyses with a greater capacity to indicate a potential risk, such as crossing divergence between the card issuance location and transaction IP, recent user accounts and number of devices, etc.

KEY FRAUD ANALYSIS
TAKEAWAYS

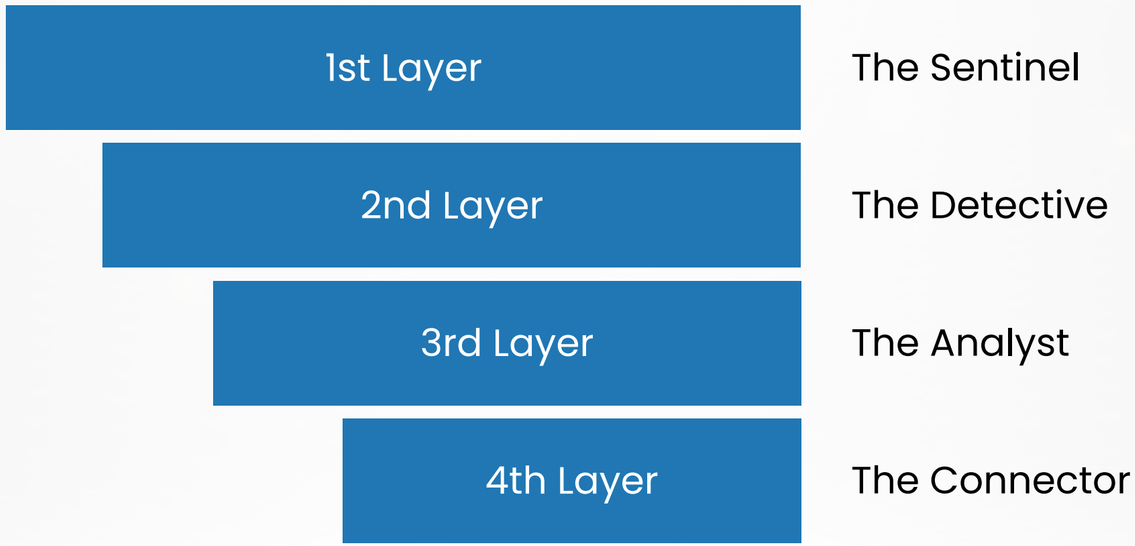
This exploratory analysis has yielded several critical insights into the fraudulent activity within this dataset:

- **Peak Activity Hours:** The vast majority of transactions, both legitimate and fraudulent, are concentrated in the afternoon and evening, with activity peaking from approximately 1 PM until midnight.
- **Fraud Follows Volume:** Chargeback incidents directly correlate with overall transaction volume. This suggests a deliberate camouflage strategy where fraudsters hide their activity within periods of high legitimate traffic.
- **Seasonal Spikes:** The correlation between transaction and chargeback volumes extends to a macro level. Both metrics experience significant spikes during major commercial events, such as the Black Friday period in the last week of November.
- **High-Risk User Profiles:** We have identified a distinct segment of users characterized by an unusually high chargeback-to-transaction ratio. This behavior serves as a primary indicator for flagging suspicious accounts.
- **Significant Financial Impact:** The monetary losses attributed to these chargebacks are substantial, representing 23.13% of the total transacted value and underscoring the severity of the financial risk.
- **Geographic Concentration:** Evidence suggests that fraudulent activity may be geographically clustered. This poses a direct risk to specific merchants who could face penalties or sanctions from payment networks due to high chargeback rates at their establishments.

POSSIBLE **SOLUTIONS**

A Multi-Layered Fraud Prevention Strategy.

With the insights extracted from this analysis, we propose a layered anti-fraud solution, seeking to increase security.



THE SENTINEL

1st Layer

The "sentinel" system, which directs rules based on its findings:
"Block/review users with CBK rate > X%",

"Alert merchants in cities with suspicious activity",

"Review overnight transactions above Y amount".

THE DETECTIVE

2nd Layer

The "detective," a machine learning model designed to find complex patterns that simple rules miss.

It learns from historical data to predict the likelihood of fraud in a new transaction in real time.

THE ANALYST

3rd Layer

The "Analyst," a system that focuses on the individual, not just the transaction.

It creates a profile of "normal" behavior for each customer (common locations, average prices, purchase times) and detects deviations.

These deviations can be used to create verification and validation methods (Face ID, 2FA).

THE CONNECTOR

4th Layer

The “Connector”, a network analysis, maps hidden connections between users, devices, and IP addresses to identify coordinated fraud rings.

OUR ACTION PLAN: **TOWARDS A SAFER PLATFORM**

Phase 1

Implement Alerting Rules: Leverage identified patterns (hourly activity, user/merchant CBK rates, transaction values) to create immediate monitoring and flagging rules.

Data Enrichment Initiative: Prioritize engineering efforts to collect the 3 most critical fields for enhanced analysis: card_bin, ip_address, and billing/shipping addresses.

Phase 2

Develop Machine Learning PoC (Proof of Concept): Initiate a pilot project to build our first risk scoring model and demonstrate its superior detection capabilities over static rules.

Deep Dive into Network Analysis: Investigate existing clusters of suspicious users to validate the potential for dismantling entire fraud networks.

Phase 3

Implement Layered Defense Strategy: Fully integrate ML, rules, behavioral analysis, and network intelligence into a unified, automated anti-fraud platform.

BY INVESTING in a smart, data-driven anti-fraud strategy, we don't just mitigate losses. We build a safer, more secure platform, and, above all, we strengthen the trust of our customers and partners.