KEY INSIGHTS AND PATTERNS

FRAUD DETECTION ANALYSIS

AN EXPLORATORY DATA ANALYSIS REPORT

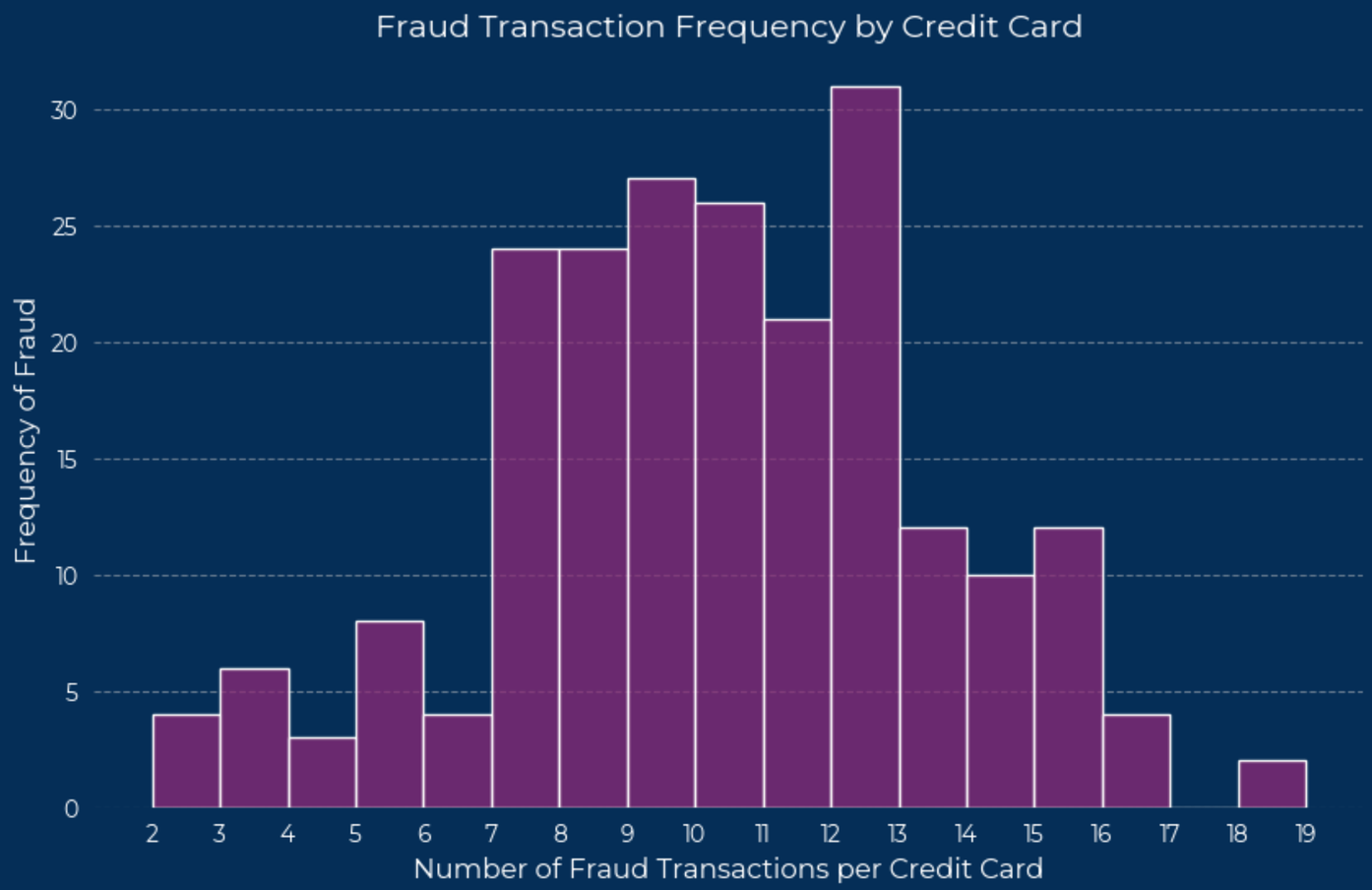
ANNIE MENESES GONZALEZ

WHY FINANCIAL FRAUD DETECTION IS IMPORTANT

Financial fraud is a growing global issue, costing businesses and individuals billions of dollars annually.

Fraudulent transactions not only lead to direct financial losses but also

- undermine trust in financial institutions,
- damage reputations,
- and create legal complications.



INTRODUCTION

OBJECTIVES OF THIS PROJECT:

- ✓ Perform **data cleaning & preprocessing**
- ✓ Conduct **EDA** (to uncover fraud patterns)
- ✓ Provide **business insights** (for fraud prevention)

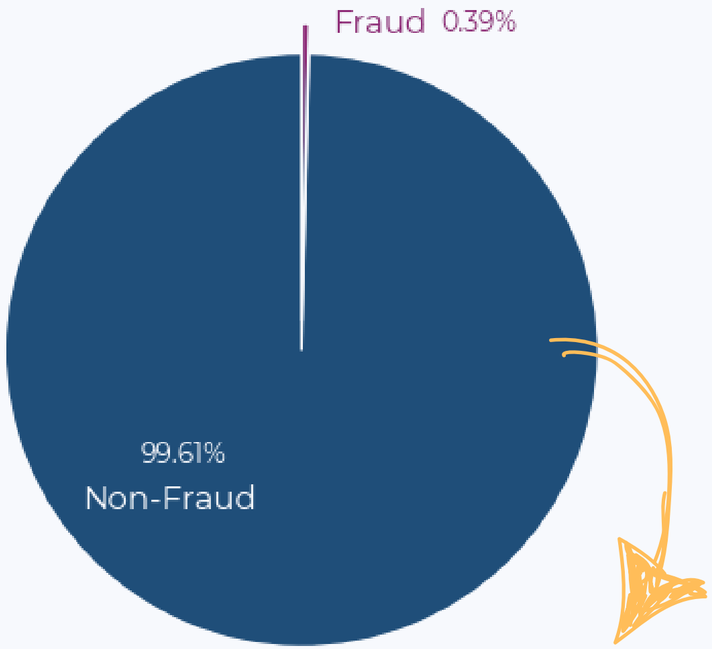
DATASET SUMMARY:

- Source: Kaggle - Credit Card Transactions Fraud Detection Dataset
- 555,719 transactions
- 23 features, including transaction time, merchant, category, amount, location, and fraud labels

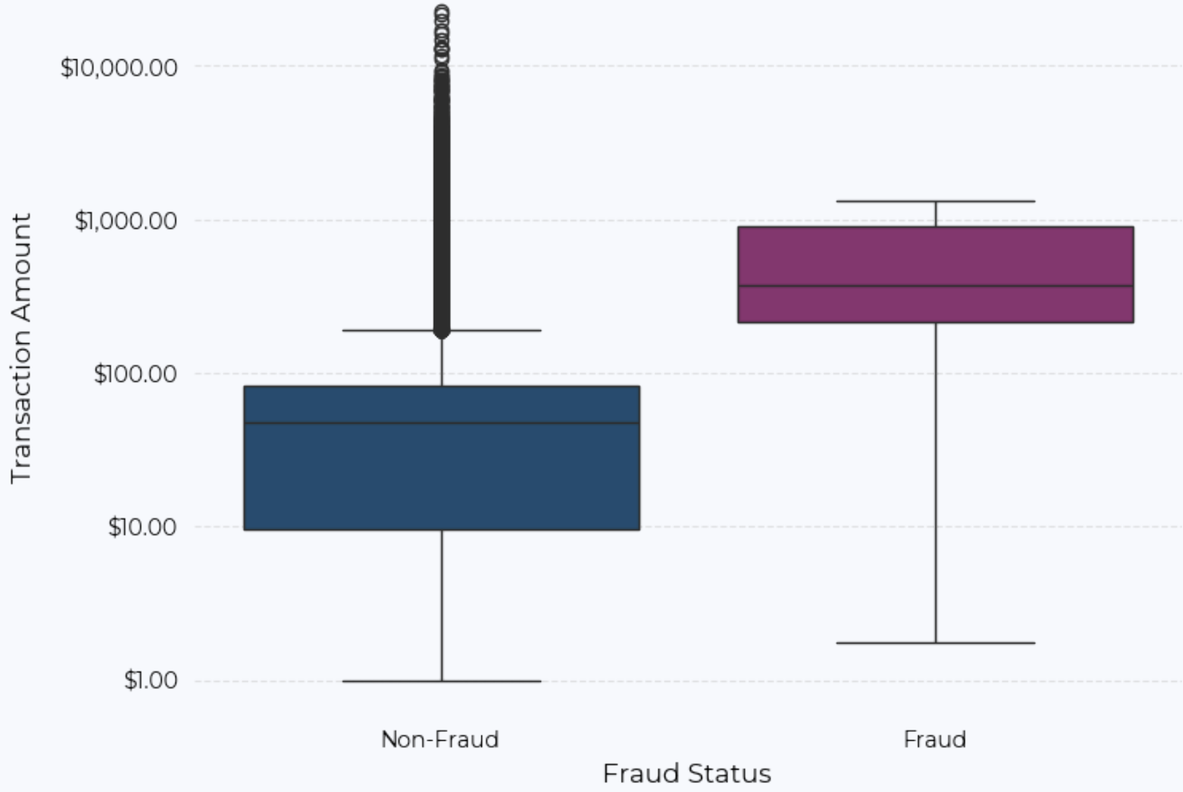
GOAL:

1. Understand the behavior of fraudsters to improve detection strategies.
2. Identify high risk factors.
3. Provide data-driven insights to improve fraud detection strategies.

FRAUD VS. NON-FRAUD DISTRIBUTION

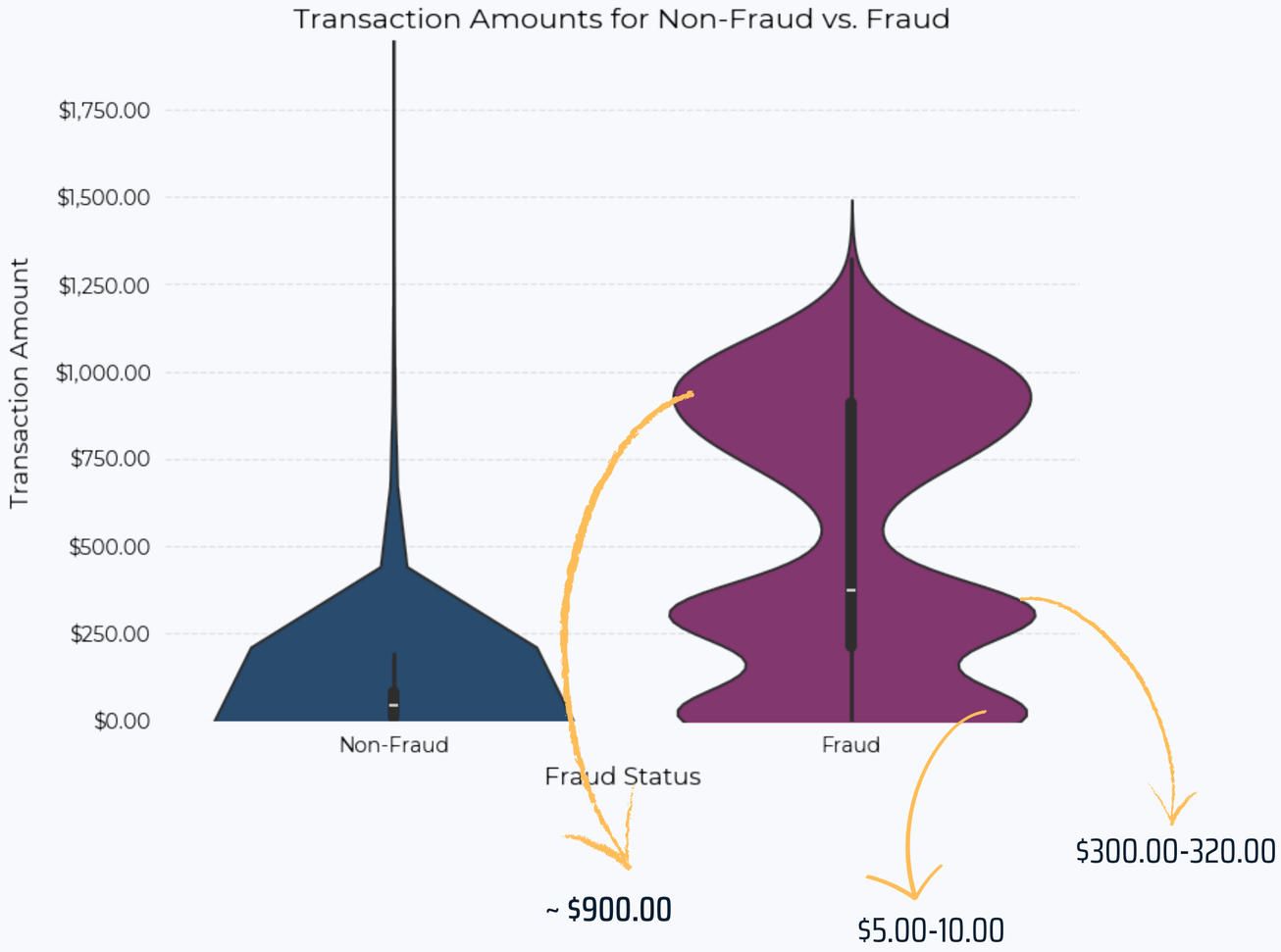


- Fraudulent transactions appear to have a higher central tendency (median).
- This suggests that, on average, **fraudulent** transactions tend to be **larger in value,** as fraudsters may target larger amounts to maximize potential gains.
- This insight guide the banks in focusing more on large transactions for early fraud detection.
- The presence of many outliers in non-fraudulent transactions increases the difficulty of distinguishing between legitimate high-value transactions and fraudulent ones.



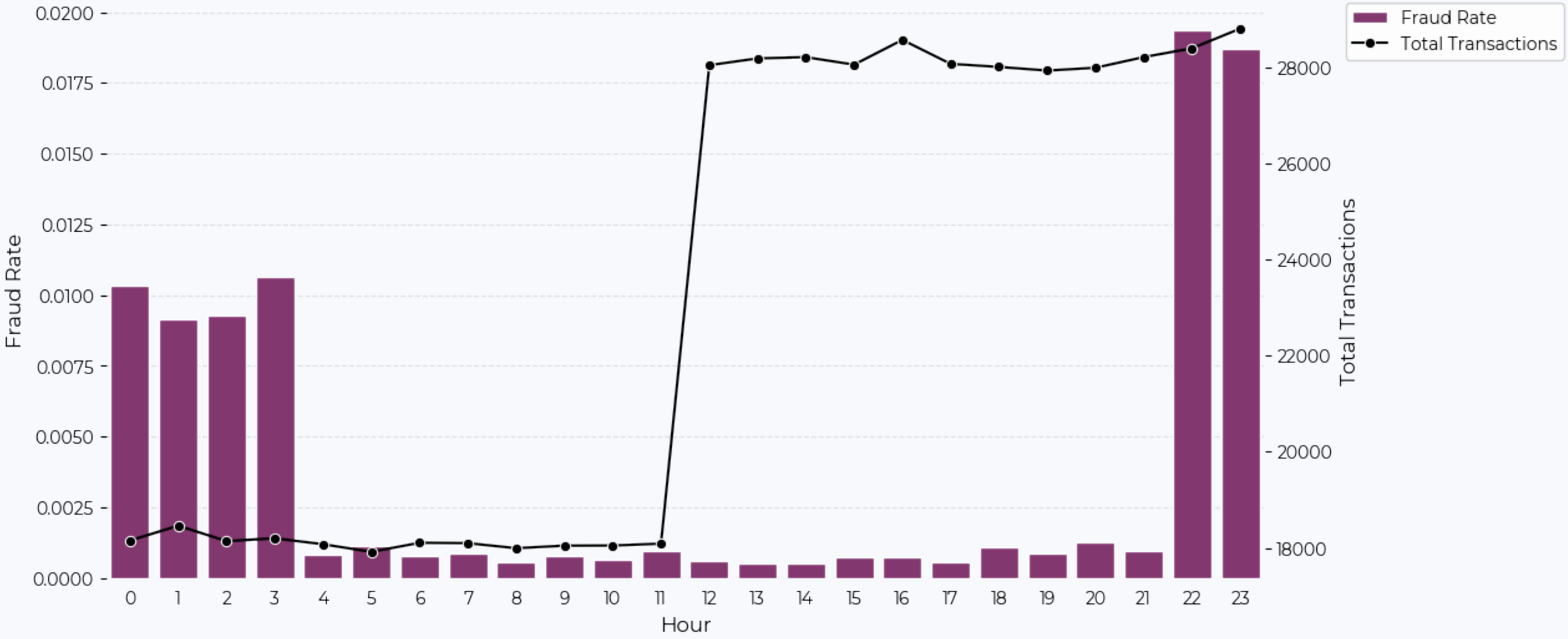
- Fraud cases are extremely rare**,** with a significant imbalance between fraudulent and non-fraudulent transactions.
- **Fraud Detection Challenge**: The large imbalance poses a significant challenge for detecting fraudulent activities.
- **Risk-Based Scoring Models**: Given this imbalance, banks must rely on risk-based scoring models that take into account multiple factors such as transaction history, user behavior, and anomalies. These models are designed to predict and identify high-risk transactions that may involve fraud, even when fraud cases are rare.

TRANSACTION AMOUNT ANALYSIS

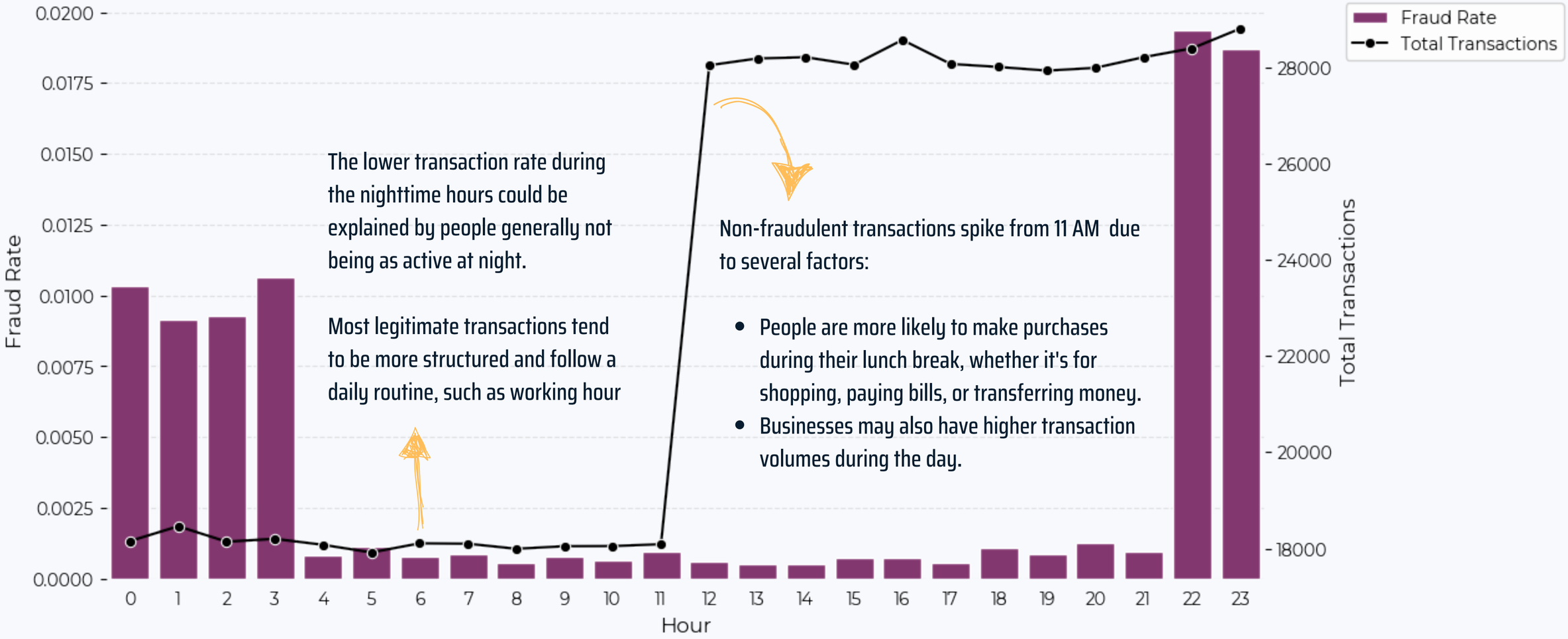


- The multiple peaks in fraud transactions suggest that fraudsters may have different strategies depending on the transaction amount.
  - Some may opt for smaller, less detectable transactions, possibly for testing or avoiding detection
  - while others target medium amounts could be an indication of fraudsters targeting moderate-value transactions for more substantial payouts while not raising immediate suspicion.
  - Higher amounts, indicating fraudsters targeting higher-value transactions, possibly because the potential gain justifies the risk of detection.
- The shape of the violin plot for non-fraud transactions , suggesting that small-to-medium value transactions are more common for legitimate purchases.
  - This reinforces the challenge in fraud detection where high-value legitimate transactions might look similar to fraudulent ones.
- A robust fraud detection system must account for this variation in transaction behavior, focusing on recognizing the different fraud strategies while minimizing false positives for legitimate transactions.

FRAUD TRANSACTIONS OVER TIME



FRAUD TRANSACTIONS OVER TIME



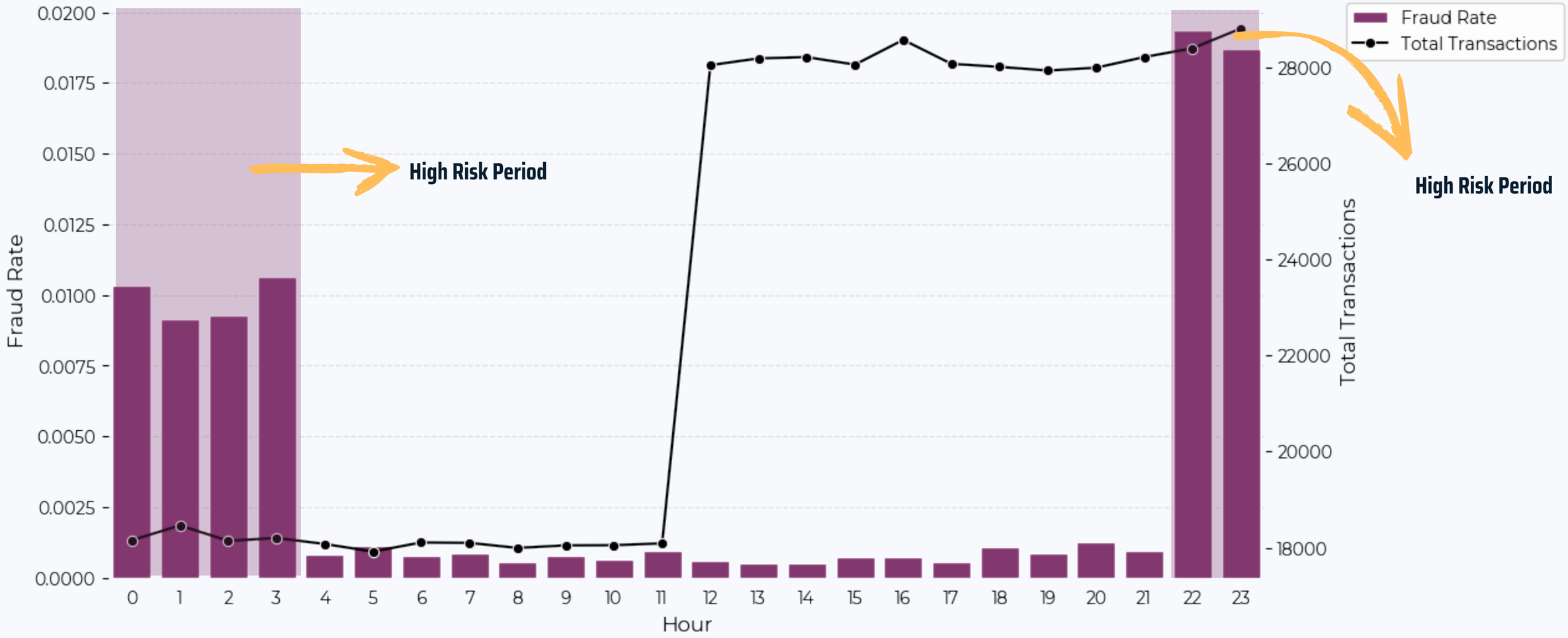
The lower transaction rate during the nighttime hours could be explained by people generally not being as active at night.

Most legitimate transactions tend to be more structured and follow a daily routine, such as working hour

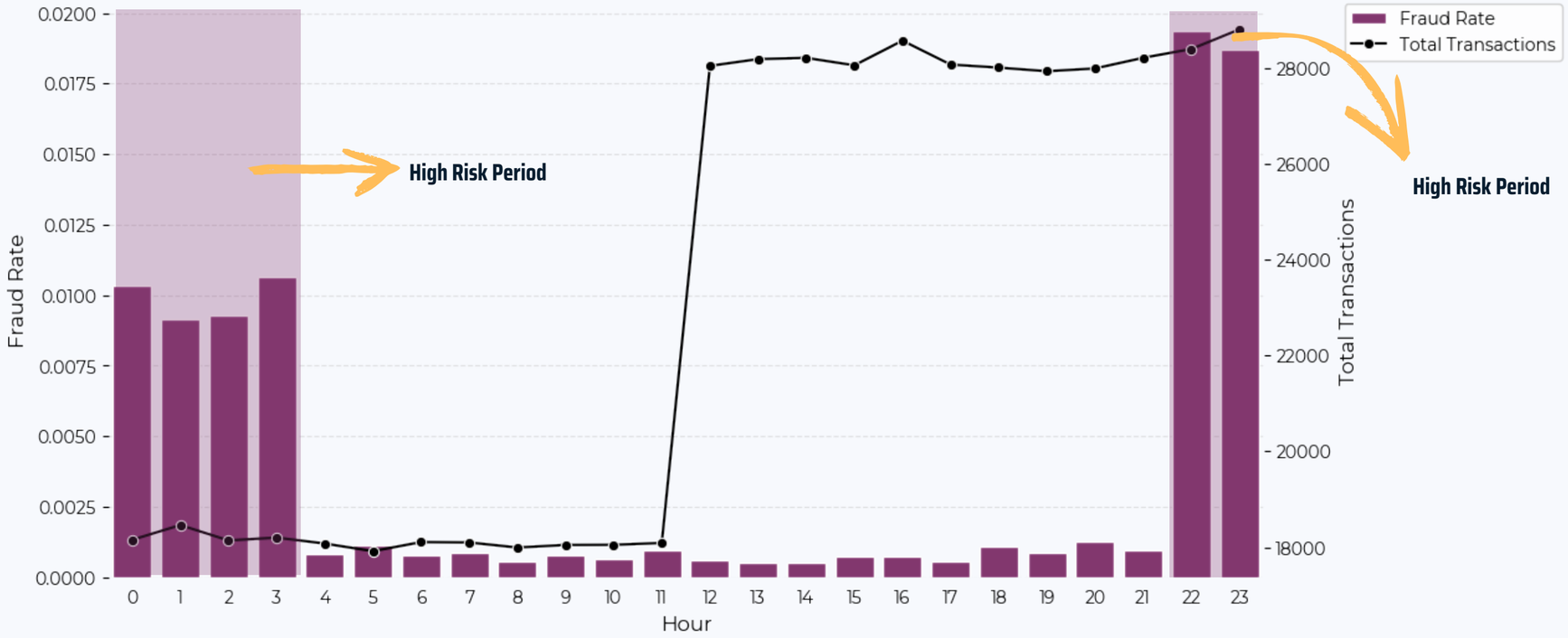
Non-fraudulent transactions spike from 11 AM due to several factors:

- People are more likely to make purchases during their lunch break, whether it's for shopping, paying bills, or transferring money.
- Businesses may also have higher transaction volumes during the day.

FRAUD TRANSACTIONS OVER TIME



FRAUD TRANSACTIONS OVER TIME



Fraudulent transactions spike during the night, this may be due to different reasons like

- Reduced Oversight: Fraudsters may prefer operating during late-night hours, where fewer people are around to detect suspicious activity. This means they might target individuals while they are asleep, with the hope that the fraud is less likely to be detected immediately.
- Global Time Zone Differences: Fraudsters can be located anywhere in the world, they may target accounts when legitime customers are less likely to notice or report suspicious activity.
- **Actionable Insights**: Increased user verification during late-night transactions.

FRAUD TRANSACTIONS OVER TIME

Non-Fraud Transactions:

- The distribution of non-fraudulent transactions does not show drastic differences between daytime and nighttime.
- This suggests that legitimate transactions follow a relatively stable pattern, with people conducting transactions throughout the day.
- Typically smaller during the day, likely due to common daily transactions such as coffee, lunch, online purchases, and small transfers.

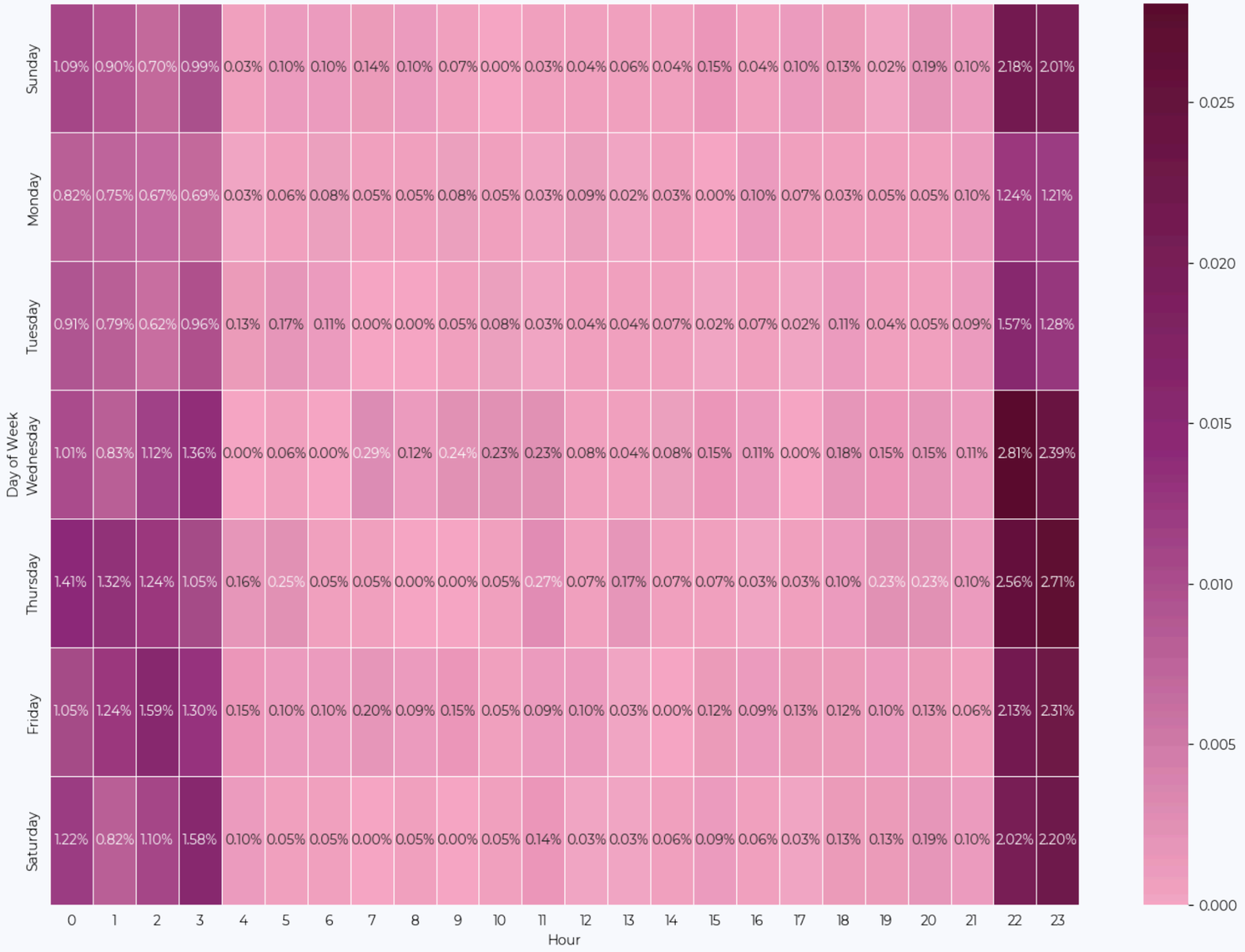
Fraud Transactions:

- Higher median and a larger IQR in both time periods suggests that fraudulent transactions tend to involve larger amounts with more variability compared to non-fraudulent ones.
- The large median during the day is surprising but could indicate:
  - Scams might be more common during the day when people are active and may respond to phishing attempts.
  - Larger sums of money are being moved during the day, making fraudulent transactions blend in more easily or in hours with less user verification protocols.

Actionable Insights:

- Inform customer about typical phishing attempts and scams during working hours
- Implement transaction limits and alerts functionalities

FRAUD TRANSACTIONS OVER TIME



Key Observations:

- This aligns with previous findings that fraud spikes after 10 PM.
- Monday & Tuesday Have the Lowest Fraud Rates:
  - People are more focused on work and responsibilities at the beginning of the week.
  - Fewer non-essential transactions, leading to fewer fraud opportunities.
- Midweek Fraud Peak:
  - Banks may have stricter fraud prevention measures over the weekend, such as extra verification steps or delayed transaction approvals. As a result, fraudsters might shift their activities to weekdays

Actionable Insights:

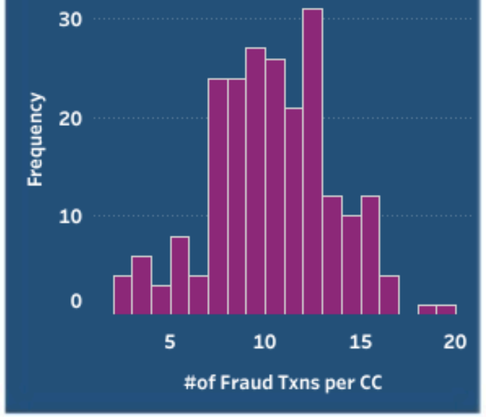
- Flag high-value transactions for manual review on midweek afternoons.
- Apply higher risk scores to first-time transactions from new devices on Wednesdays & Thursdays.

Fraud Detection Insights

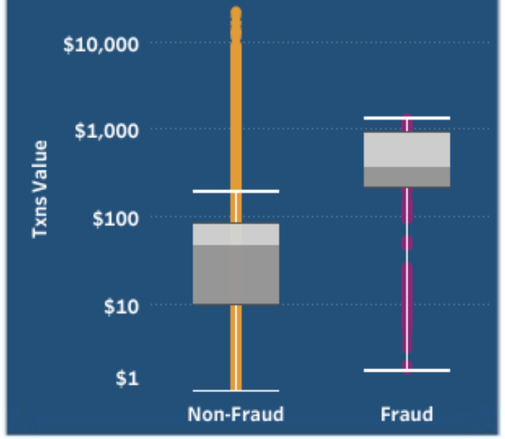
Gender: (All) | Age group: (All) | Population Group: (All)

# of Txns	# of Fraud Txns	Fraud Percentage	Avg. Non-Fraud Txns	Avg. Fraud Txns
555,719	2,145	0.39%	$67.61	$528.36

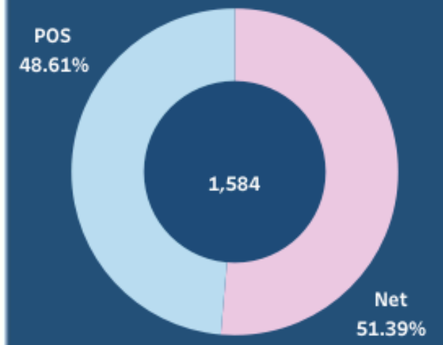
Fraud Frequency per CC

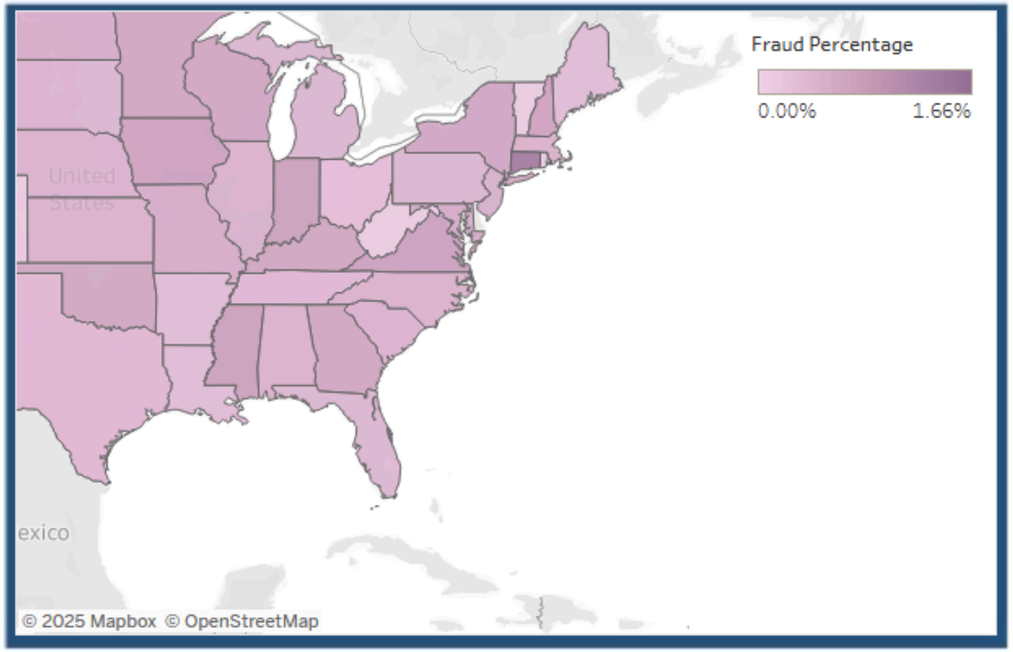


Txns Value Range

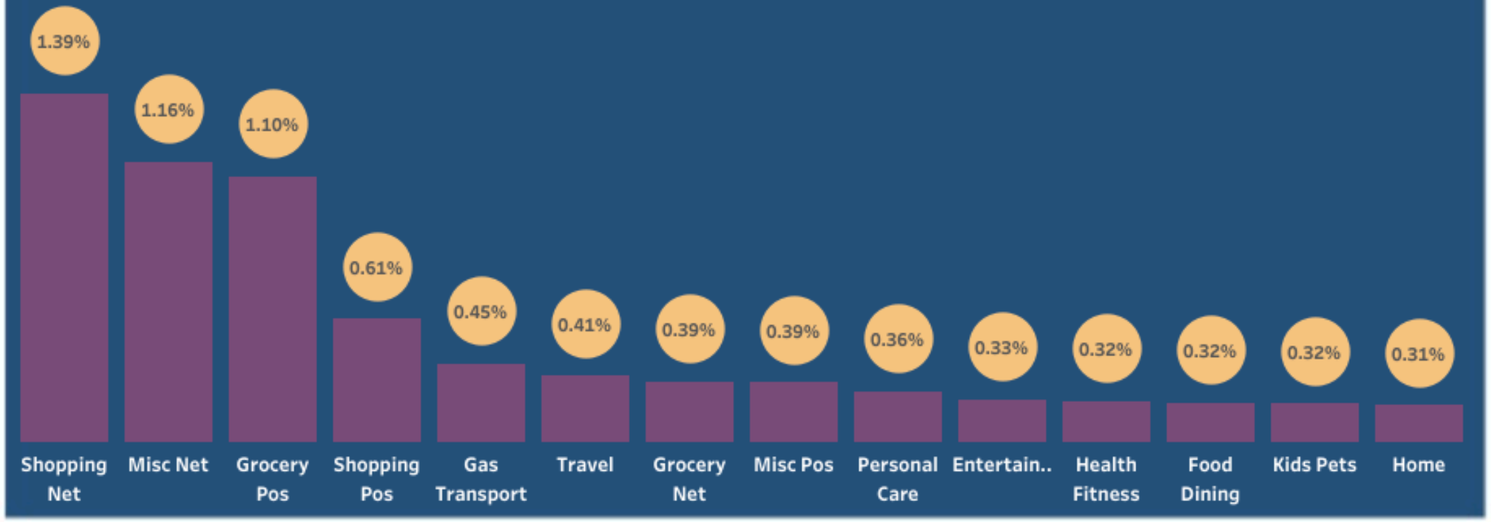


Online vs. Point of Sale

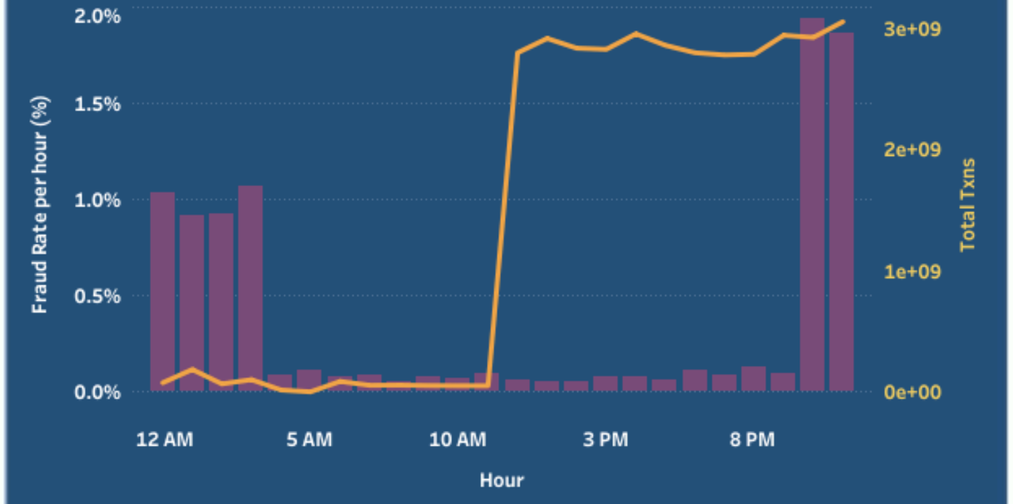




Fraud rate per merchant category



Fraud Txns per Hour



KEY TAKEAWAYS

TIME-OF-DAY PATTERNS SHOW FRAUD PEAKING AT NIGHT

MERCHANT CATEGORIES AS SHOPING AND GROCERY ARE AT HIGHER RISK

AGE AND GENDER HAVE AN IMPACT FRAUD VULNERABILITY

RECOMMENDATIONS FOR FRAUD PREVENTION

- Real-time fraud monitoring during high-risk periods.
  - Enhanced fraud detection between 10 PM – 3 AM.
- Geographic-based fraud alerts.
  - Custom fraud detection models for high-risk cities.
- Educational campaigns for high-risk demographics.
- Strengthening POS security with advanced authentication methods can help reduce fraud in POS systems.


THANK YOU