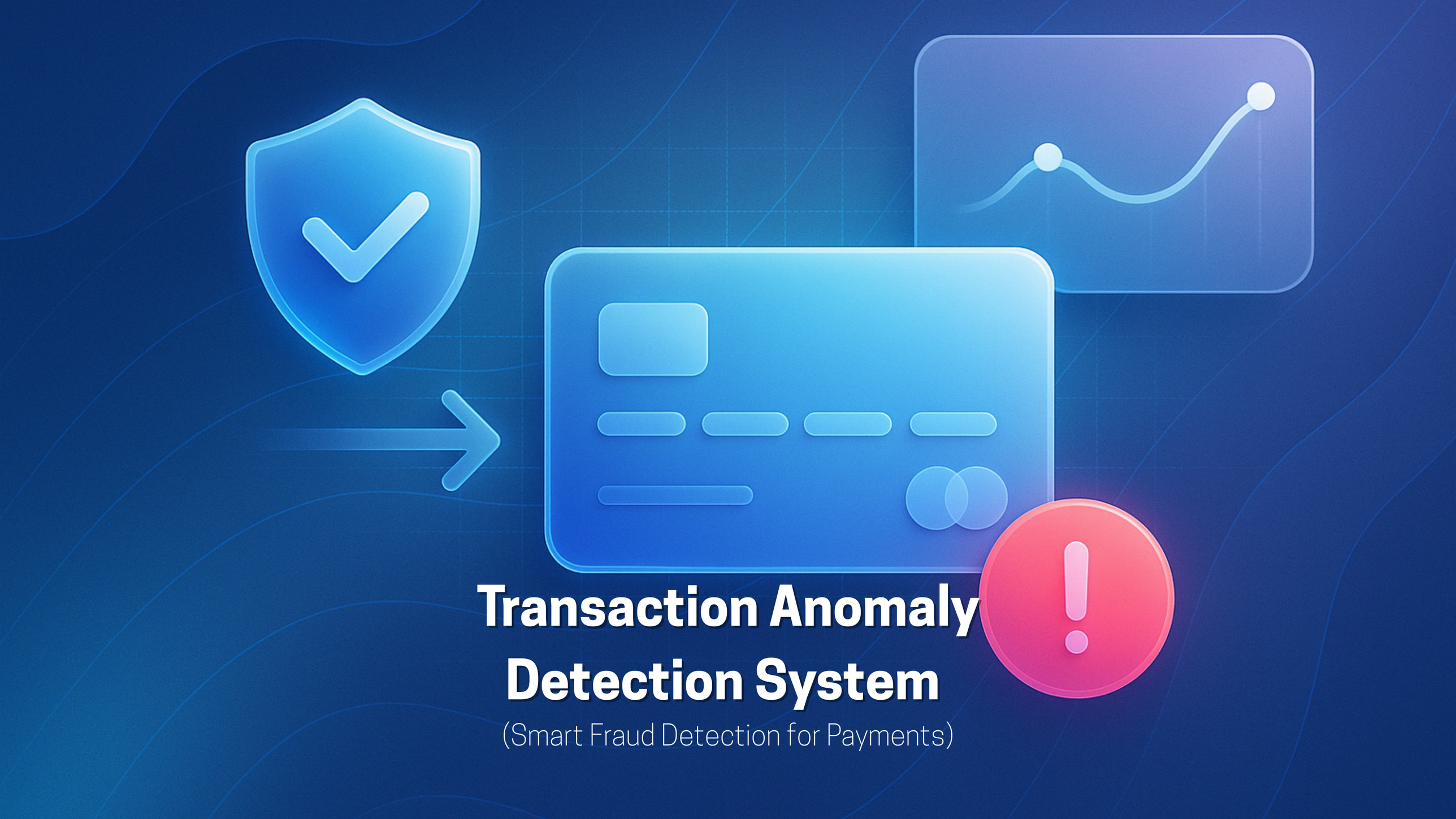


Transaction Anomaly Detection System

(Smart Fraud Detection for Payments)

Project Summary:

In today’s fast-paced digital economy, ensuring the security of payment transactions is of paramount importance. With the growing volume of online transactions, fraud prevention has become a critical challenge for businesses and financial institutions. This report provides a detailed analysis of a dataset of 200,000 payment transactions, focusing on transaction trends, fraud patterns, and potential areas for enhancing fraud detection systems. By analyzing the data, we gain key insights into payment behaviors, fraud occurrences, and operational patterns, which will be instrumental in developing an effective fraud detection framework.

Introduction:

The growing digital payments landscape, particularly through mobile platforms and online wallets, presents both opportunities and challenges. The sheer volume of transactions coupled with the increasing sophistication of fraud requires advanced detection techniques. This report presents an analysis of transaction data from multiple dimensions, aiming to identify anomalies, fraud patterns, and insights that will enhance the ability to detect and prevent fraudulent activities within the payment ecosystem.

The analysis was conducted using Python and various analytical libraries (e.g., Pandas, Matplotlib, Seaborn) to extract meaningful insights. These insights will guide the development of fraud detection models and contribute to building a more secure payment environment.

Problem Statement

The main objectives of this report are:

- To determine the total number of transactions and the overall money processed.
- To identify the most commonly used payment methods and devices.
- To understand the geographic distribution of transactions.
- To analyze the incidence of fraud within the dataset and identify any patterns related to transaction amounts, payment modes, and device types.
- To uncover time-based trends that may correlate with transaction volumes and fraud occurrences.

Methodology:

The dataset used for this analysis contains **200,000** transactions, which includes both fraudulent and non-fraudulent records. The analysis was carried out using the following approach:

- Descriptive Statistics were employed to summarize the dataset, including transaction counts and amounts.
- Data Visualization was used to identify trends and patterns in payment methods, geographic distribution, and device usage.
- Comparative Analysis was used to examine fraud rates across different categories, including transaction size, payment method, and device type.
- Time-Series Analysis allowed us to detect transaction and fraud patterns based on the time of day and day of the week.

Key Findings:

- **Total Transactions:**
  The dataset includes a total of **200,000** transactions, providing a broad representation of payment behavior.

```
total_transactions = df.shape[0]
print(f"Total transactions: {total_transactions}")
```

- **Total Money Processed:**
  The total amount of money processed across all transactions is **₹182,632,874.03**. This large financial volume highlights the scale of the payment ecosystem and the need for robust fraud detection mechanisms.

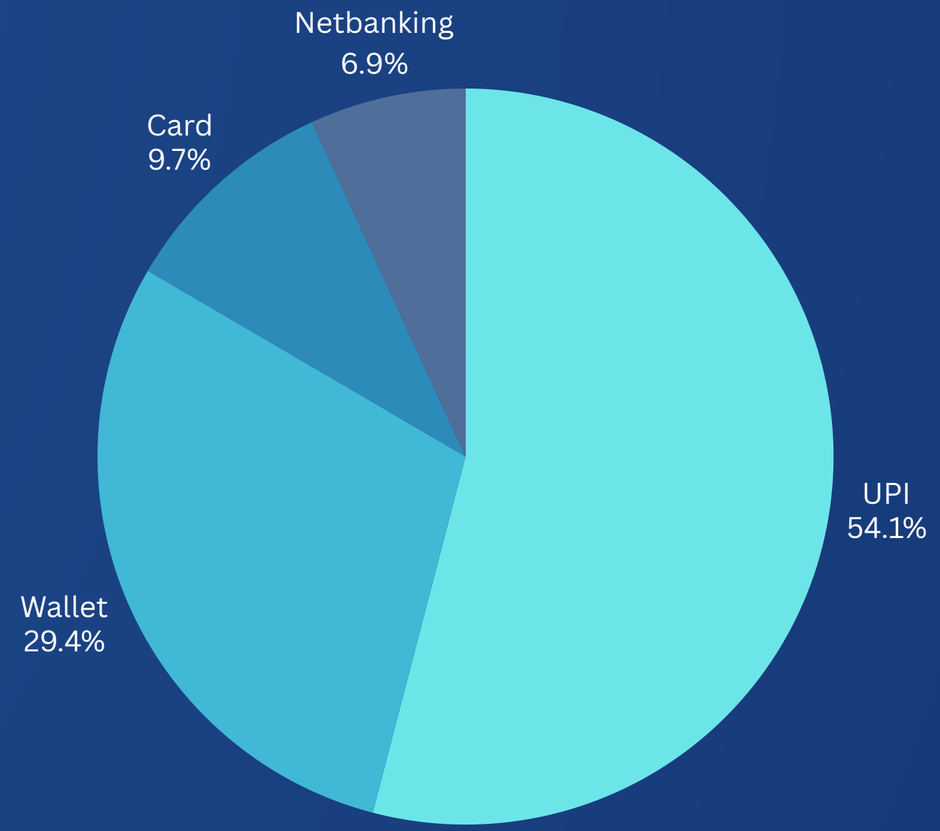
```
total_money_processed = df['amount'].sum()
print(f"Total money processed: {total_money_processed}")
```

Payment Methods:

The most frequently used payment methods are:

- UPI: 108,114 transactions
- Wallet: 58,727 transactions
- Card: 19,455 transactions
- Net banking: 13,704 transactions

```
payment_mode_counts = df['payment_mode'].value_counts()
print("Top payment modes used:")
print(payment_mode_counts)
```



These figures indicate that **UPI** is the dominant payment mode, which should be a key focus area in fraud detection efforts.

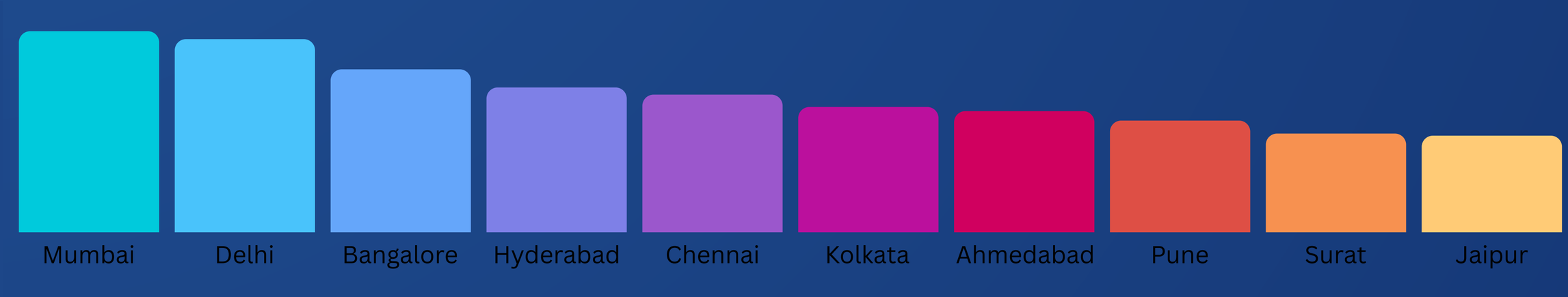
Geographic Distribution:

The top cities and states with the highest transaction volumes are:

- Mumbai, Maharashtra: 11,885 transactions
- Delhi, Delhi: 11,417 transactions
- Bangalore, Karnataka: 9,629 transactions
- Hyderabad, Telangana: 8,555 transactions

```
city_transaction_counts = df['location'].value_counts().reset_index()
city_transaction_counts.columns = ['City/State', 'Transaction Count']
print("TRANSACTIONS BY CITY/STATE - COMPLETE LIST")
print(city_transaction_counts)
print(f"Total locations: {len(city_transaction_counts)}")
print(f"Total transactions: {city_transaction_counts['Transaction Count'].sum()}")
```

This geographical spread offers insights into regions with higher payment activity, which may require additional monitoring for fraud prevention.



Fraud vs Non-Fraud Transactions:

Out of the 200,000 transactions, 197,917 were non-fraudulent, while 2,083 were flagged as fraudulent. This implies a fraud rate of 1.04% in the dataset. Although fraud accounts for a small proportion, it remains significant enough to warrant attention.

```
fraud_counts = df['is_fraud'].value_counts()
print("Fraud vs non-fraud transactions:")
print(fraud_counts)
```

Average Transaction Amount:

The average transaction amount across all transactions was calculated at **₹913.16**. This value serves as a benchmark for detecting anomalies, especially in higher-value transactions that might pose greater risks for fraud.

```
average_transaction_amount = df['amount'].mean()
print(f"Average transaction amount: {average_transaction_amount}")
```

Distribution of Transaction Amounts:

Transactions were categorised into three groups based on the transaction value:

- Small: < ₹1,000
- Medium: ₹1,000 - ₹5,000
- Large: > ₹5,000
-

The distribution is as follows:

- Small: 143,761 transactions (71.88%)
- Medium: 53,845 transactions (26.92%)
- Large: 2,066 transactions (1.03%)

```
bins = [0, 1000, 5000, np.inf]
labels = ['Small', 'Medium', 'Large']
df['transaction_size'] = pd.cut(df['amount'], bins=bins, labels=labels)
transaction_size_distribution = df['transaction_size'].value_counts()
print("Transaction amount distribution:")
print(transaction_size_distribution)
```

The majority of transactions are small, while large transactions, though fewer, carry higher potential risk.

Fraud Patterns Based on Transaction Amount:

Fraudulent transactions are more likely to involve higher amounts, though they still vary widely. Key findings include:

- Average fraud transaction amount: **₹2,034.80**
- Standard deviation: **₹2,452.92**
- Range: From **₹162.44** to **₹30,423.71**

```
fraud_transactions = df[df['is_fraud'] == 1]
fraud_amount_stats = fraud_transactions['amount'].describe()
print("Fraud transactions amount stats:")
print(fraud_amount_stats)
```

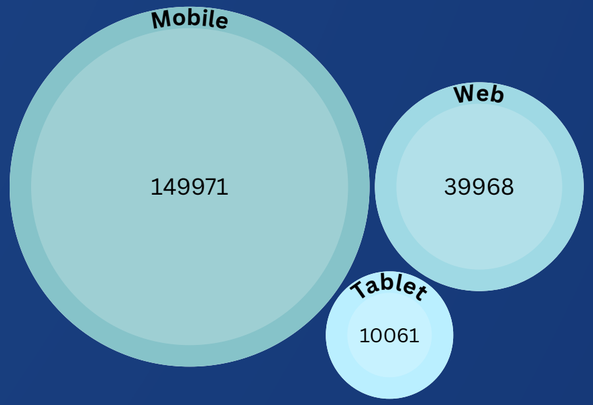
Most fraudulent transactions fall between **₹687.25** and **₹2,412.96**, with a noticeable increase in fraud for higher-value transactions. This suggests that fraudsters tend to target larger transactions, highlighting the need for focused fraud detection on higher-value transactions.

Device Usage:

The distribution of transactions across device types is as follows:

- Mobile: **149,971** transactions
- Web: **39,968** transactions
- Tablet: **10,061** transactions

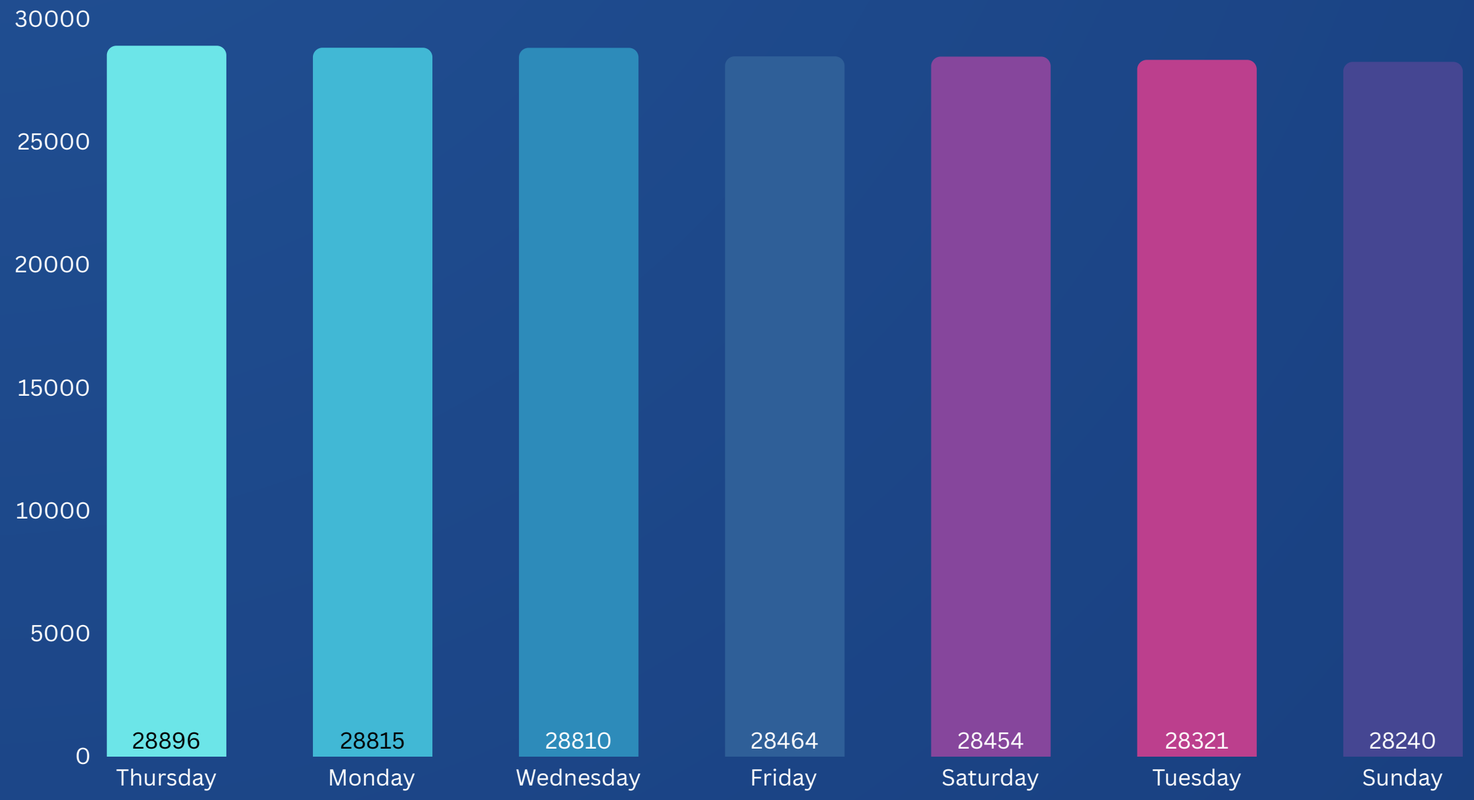
```
device_type_counts = df['device_type'].value_counts()
print("Transactions by device type:")
print(device_type_counts)

fraud_device_type = fraud_transactions['device_type'].value_counts()
print("Fraud transactions by device type:")
print(fraud_device_type)
```

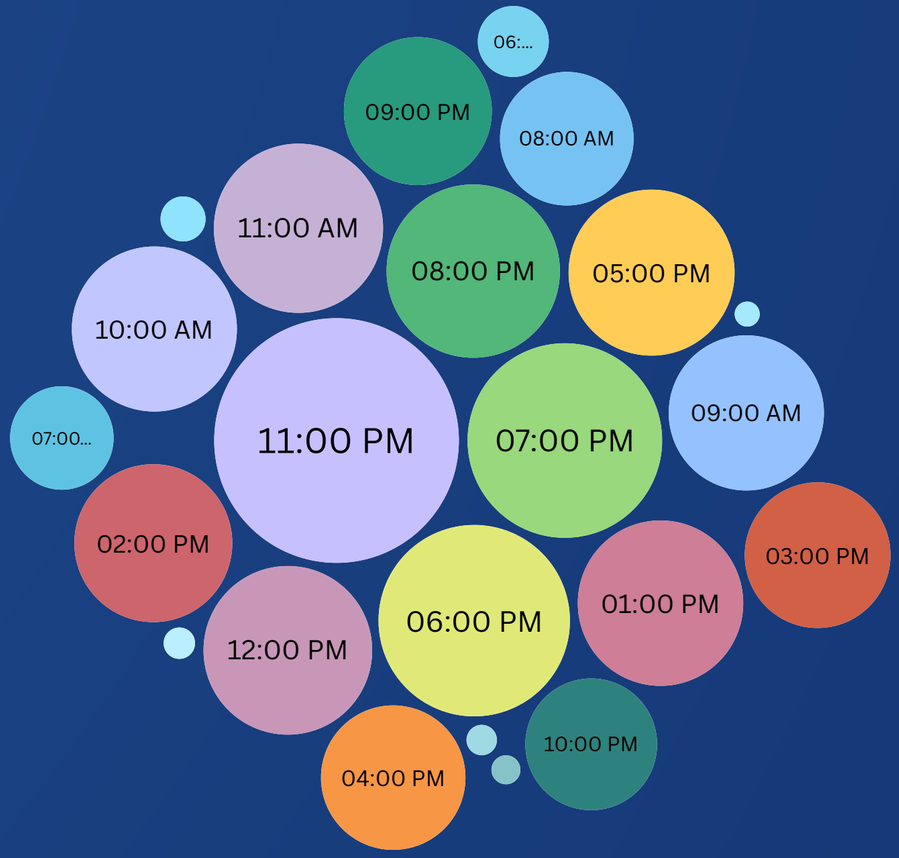


Fraudulent transactions were more prevalent on mobile devices, with **1,549** frauds compared to **424** on web and **110** on tablets. This suggests that mobile platforms may be a target for increased fraud activity.

Time Trends:

Transaction volumes fluctuated throughout the week, with the highest volume occurring on Thursday **(28,896 transactions)**. The peak hours for transactions were observed between **11 AM and 1 PM**, with over **12,000 transactions** recorded during these hours. Fraud detection systems should be optimized for high-volume periods, particularly on Thursdays and during peak hours.





🔶 Discussion:

The analysis uncovers several key insights for the development of a fraud detection system:

- Payment Modes: Focus on UPI and Wallet transactions as these represent the highest volume of payments.
- Fraud Patterns: Fraudulent transactions tend to have a higher value, particularly on mobile devices. These trends suggest that a focused detection mechanism for high-value mobile transactions could significantly improve fraud detection accuracy.
- Time-based Patterns: High transaction volumes and fraud rates occur during specific times of the day and week, particularly Thursday afternoons. Detecting anomalies during these periods could improve fraud prevention.
- Geographic Focus: Cities like Mumbai and Delhi have high transaction volumes and may require more attention to detect potential fraudulent activities.

Conclusion:

The analysis has provided comprehensive insights into transaction patterns and fraud occurrences. The findings reveal that fraudulent transactions tend to be higher-value and more frequent on mobile devices. Additionally, the geographic distribution of transactions highlights the need for tailored fraud detection models based on regional transaction trends. Time-based analysis indicates peak hours for transaction activity, where fraud prevention systems can be optimized.

This analysis lays the foundation for a more focused and effective Transaction Anomaly Detection System, capable of identifying potential fraud through targeted insights based on transaction amounts, device types, and time trends.