Credit Card Fraud Analysis

Identifying hidden fraud patterns in credit card transactions

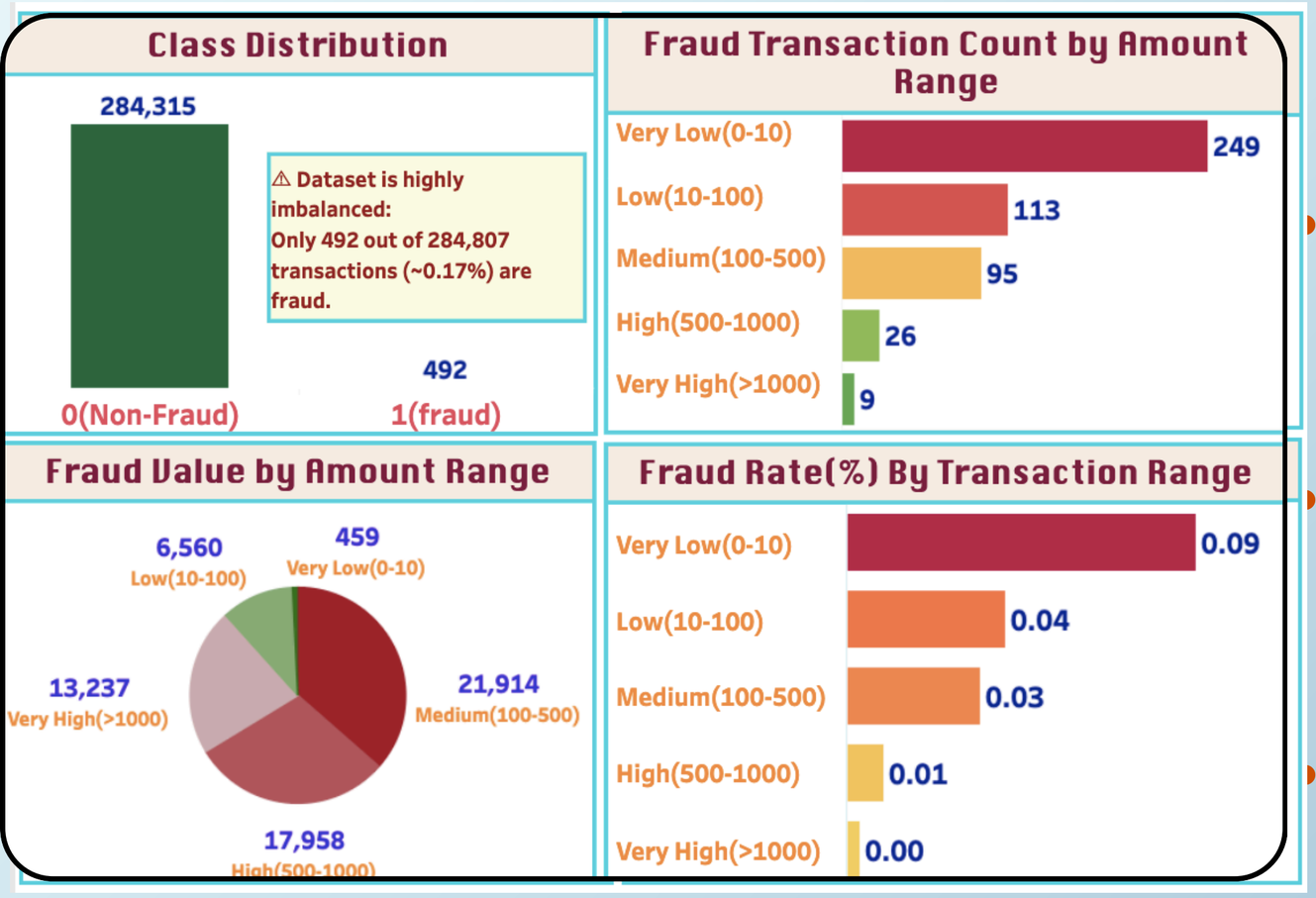
Project Objective

To detect fraudulent credit card transactions within a highly imbalanced dataset and uncover the hidden patterns behind them. By analyzing transaction amounts, time-of-day trends, and feature correlations, the goal is to identify behaviors that distinguish fraud from genuine activity.

Data And Methodology

- This project uses a Kaggle dataset of 284,807 credit card transactions made by European cardholders over two days in September 2013, which includes 492 fraud cases representing only 0.17% of the data.
- The dataset contains anonymised numerical features (V1–V28) obtained through PCA transformation, along with two original features, Time and Amount.
- After data cleaning and exploratory data analysis in Python, new fields such as Hour and AmountBin were engineered to reveal time- and value-based fraud patterns.
- The insights were then visualized through an interactive Tableau dashboard, combining KPI cards, distribution charts, time-series analysis,

Class Distribution and Transaction Amount Analysis



- Fraud is extremely rare: only 492 out of 284,807 transactions (0.17%).
- Small transactions (<€10) show the highest fraud frequency (249 cases, ~50%), but the financial loss is minimal.
- Medium (€100–500) and High (€500–1000) ranges have fewer frauds but cause the largest financial impact (~€40K combined).
- Fraud rate is highest in very low transactions (0.09%) and decreases as transaction amount increases.
- Fraudsters balance between many small, low-risk frauds and occasional high-value attempts.

Recommendations:

Low-Value Transactions (<€10)

- Implement velocity checks (multiple small transactions in short time windows).
- Flag accounts with suspicious small-amount testing.

Medium & High-Value Transactions (€100–1000)

- Apply stricter authentication (e.g., 2FA, OTP confirmation).
- Use dynamic risk scoring → raise alerts for unusual spending compared to cardholder history.

Very High Transactions (>€1000)

- Trigger real-time alerts to customers for manual verification.
- Require secondary validation (e.g., contact bank, biometric verification).

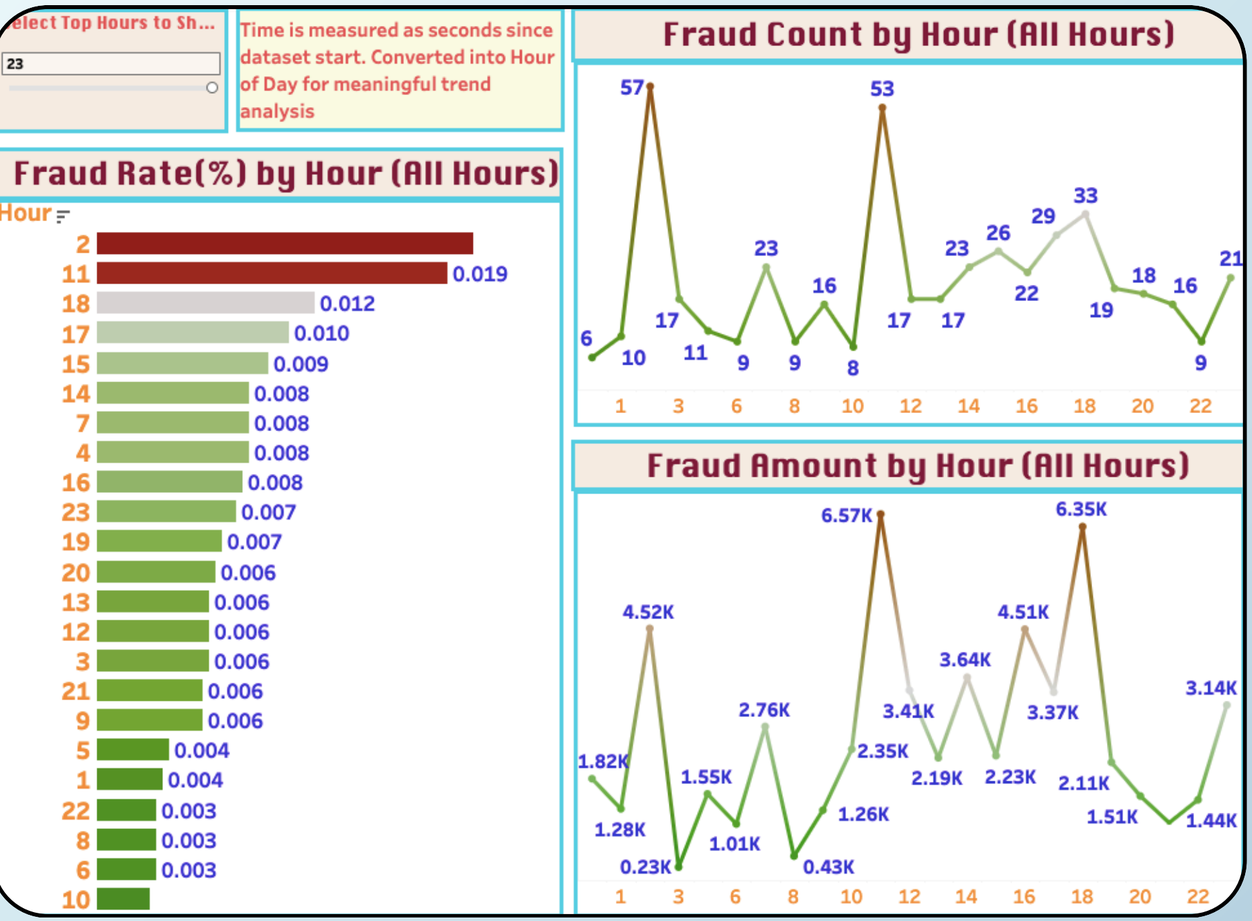
General Fraud Monitoring

- Develop hybrid fraud detection models
- Continuously retrain models to adapt to evolving fraud strategies.

Operational Strategies

- Focus fraud analyst teams more during high-risk transaction ranges.
- Share findings with merchant partners (e.g., e-commerce platforms) to implement amount-specific fraud checks.
- Educate customers → highlight risks of small “testing” transactions as potential fraud indicators.

Hourly Fraud Patterns

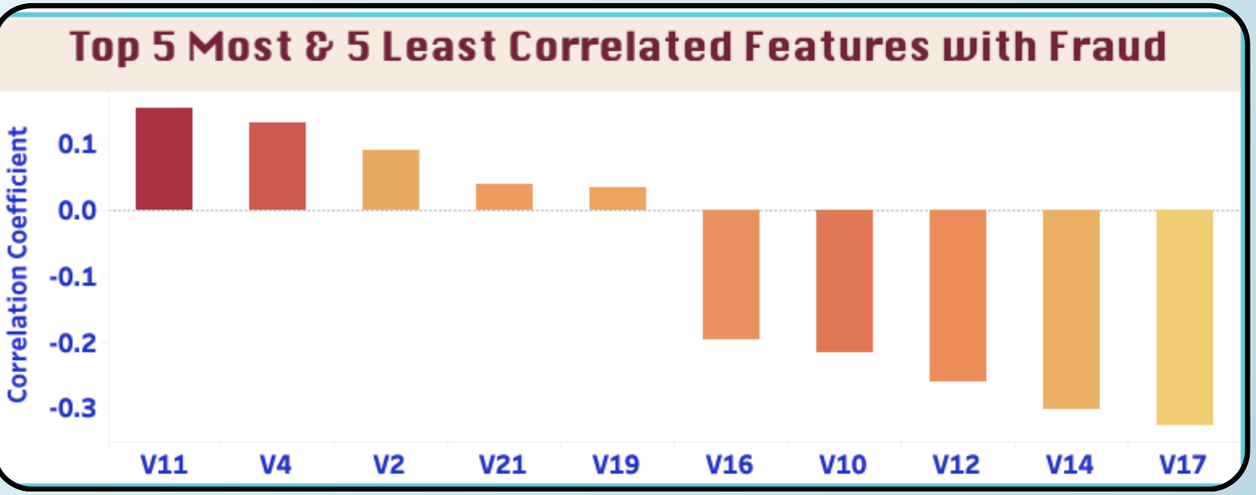


- Fraud peaks at 2 AM and 11 AM, with additional spikes in the evening (16–18 hrs).
- Overall fraud rate is low per hour, but concentrated during specific time windows.
- High financial impact observed at peak hours: €6.5K–6.7K in a single hour.
- Fraud count and fraud rate often rise together (e.g., 2 AM, 11 AM), but fraud amount does not always align medium-count hours like 5 PM and 12 PM show disproportionately high losses, indicating targeted high-value frauds.

Recommendations:

- Prioritise fraud detection models and real-time alerts between 2–3 AM and 10–12 AM, as both fraud rate and amount are highest.
- Flag unusually high-value transactions during 2 AM, 11 AM, and 6 PM since these times show fraud attempts with higher monetary losses.
- Schedule more fraud analysts or automated system checks during late night (2 AM) and midday (10 AM–12 PM) when fraud spikes.
- Educate customers to be more cautious during late night/early morning hours, encouraging use of extra authentication for high-value payments.
- Build hour-of-day features into fraud detection models to capture time-based fraud trends.

Feature Correlation



- Top Correlated Features: Features like V11, V4, and V2 show the strongest positive correlation, suggesting that higher values in these features are more indicative of fraud.
- Least Correlated Features: Features like V17, V14, and V12 show the strongest negative correlation, meaning lower values in these features are more common in fraudulent transactions.
- Correlations are relatively weak in magnitude (close to zero), meaning no single feature is strongly predictive on its own.
- Fraud detection likely requires combining multiple features, not relying on one.

Recommendations:

- Focus on V11, V4, V2 for building fraud prediction models (they may capture fraud-driving patterns).
- Incorporate both positively and negatively correlated features into fraud detection models negative correlations are equally useful for separating fraud vs. non-fraud.
- Since correlations are weak, avoid using correlation alone as a filter. Instead, use feature importance techniques (e.g., tree-based models, permutation importance) to validate.
- Use positively correlated features (V11, V4, V2) for real-time fraud flags.
- Use negatively correlated features (V16, V10, V12, V14, V17) to reduce false positives (helping balance detection).
- Perform multivariate analysis (logistic regression, random forest feature importance) to confirm which of these features are truly significant.

Dashboard Link

https://public.tableau.com/views/CreditCardFraudAnalysis_17583612937270/Dashboard1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Thank you