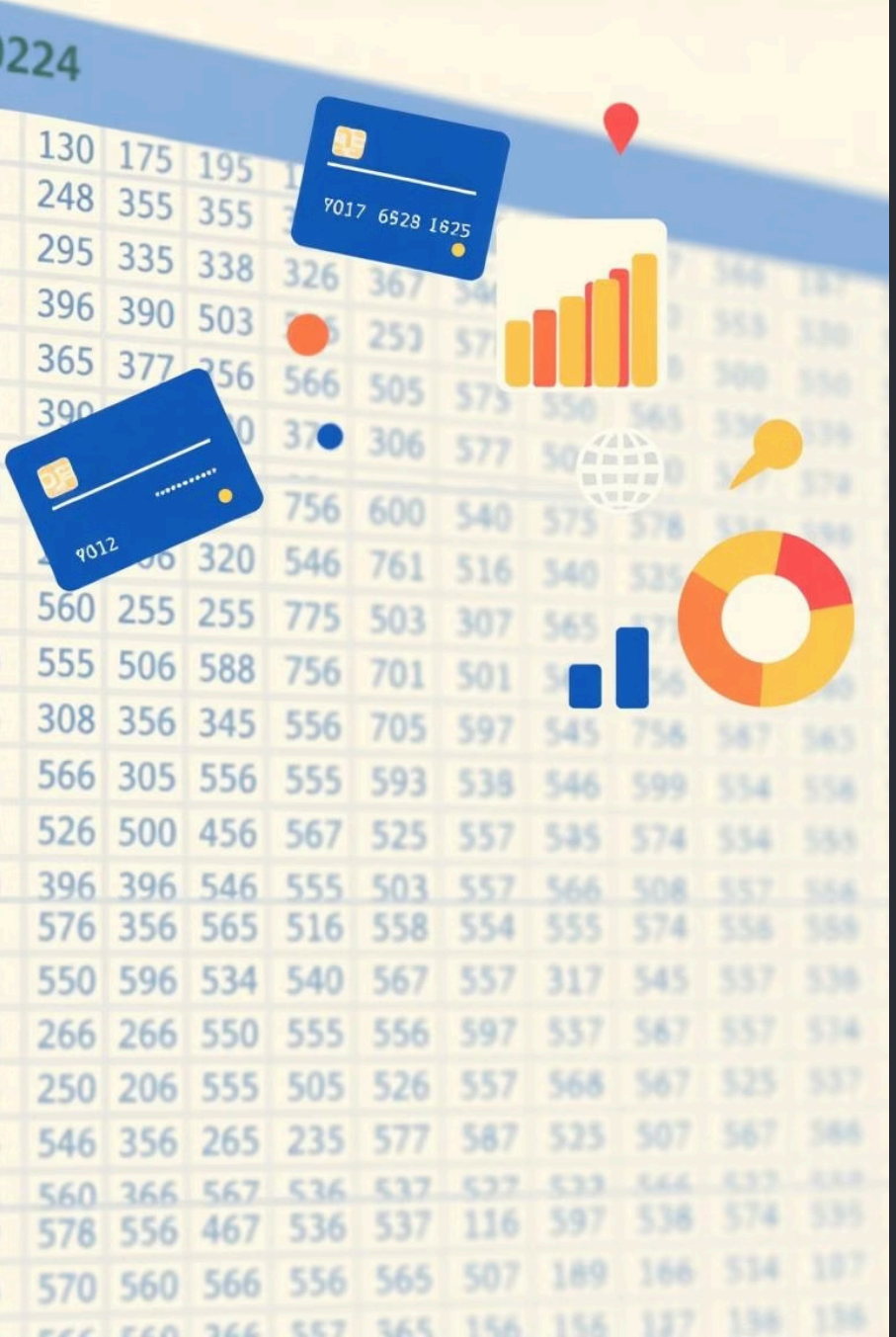
Bank Transaction Fraud Detector

A Machine Learning Approach to Identify Suspicious Transactions

The Problem: Battling Financial Fraud

Financial fraud represents a substantial and growing threat to banks and their customers alike. It leads to significant monetary losses, erodes customer trust, and can incur substantial operational costs for investigation and remediation. Our objective with this project is to develop an automated, intelligent system capable of accurately detecting fraudulent transactions in real-time, thereby mitigating these risks effectively.



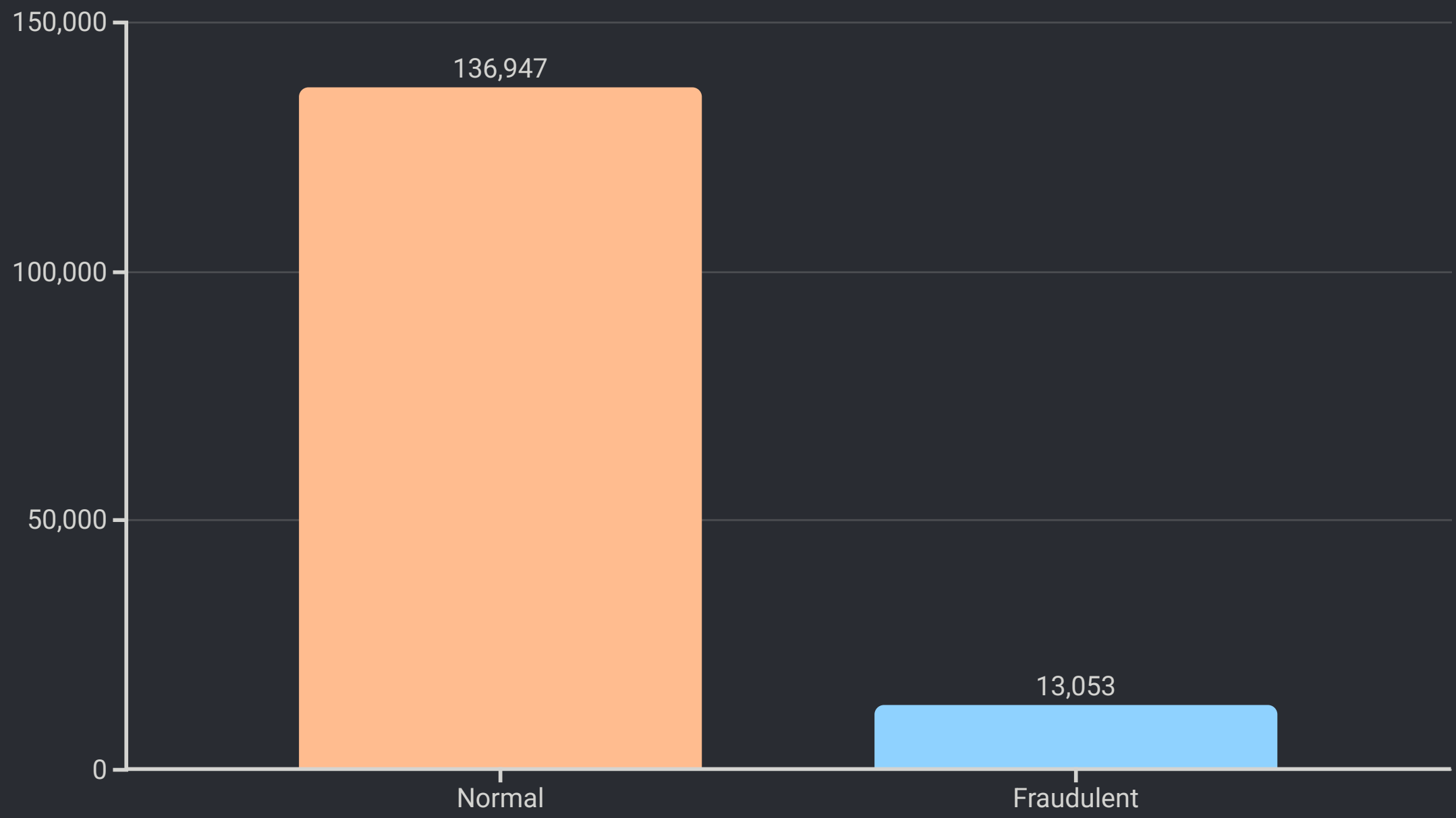
The Dataset: Unveiling Transaction Patterns

Our analysis leverages the "**creditcard_sample.csv**" dataset, a comprehensive collection of anonymised transaction records. This dataset provides crucial insights into transactional behaviour, with key features including:

- **ratio_to_median_purchase_price:** Indicating the transaction amount relative to the cardholder's median purchase.
- **online_order:** A binary flag differentiating between online and in-person transactions.
- **fraud:** Our target variable, identifying whether a transaction is legitimate or fraudulent.

Exploratory Data Analysis: Addressing Class Imbalance

A critical challenge in fraud detection datasets is often the severe class imbalance, where fraudulent transactions are significantly rarer than legitimate ones. Our dataset reflects this reality, with a small percentage of fraudulent transactions compared to normal ones.



This imbalance necessitates careful handling during model training to prevent the model from simply predicting the majority class and overlooking the critical minority (fraudulent) class.

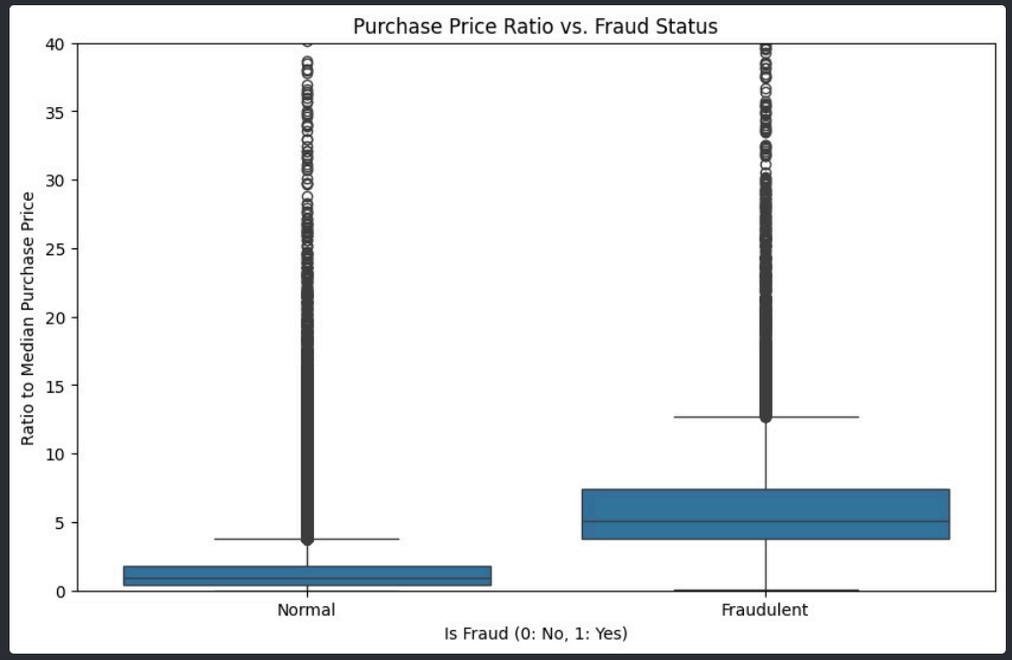
Key Insights: Uncovering Fraud Indicators

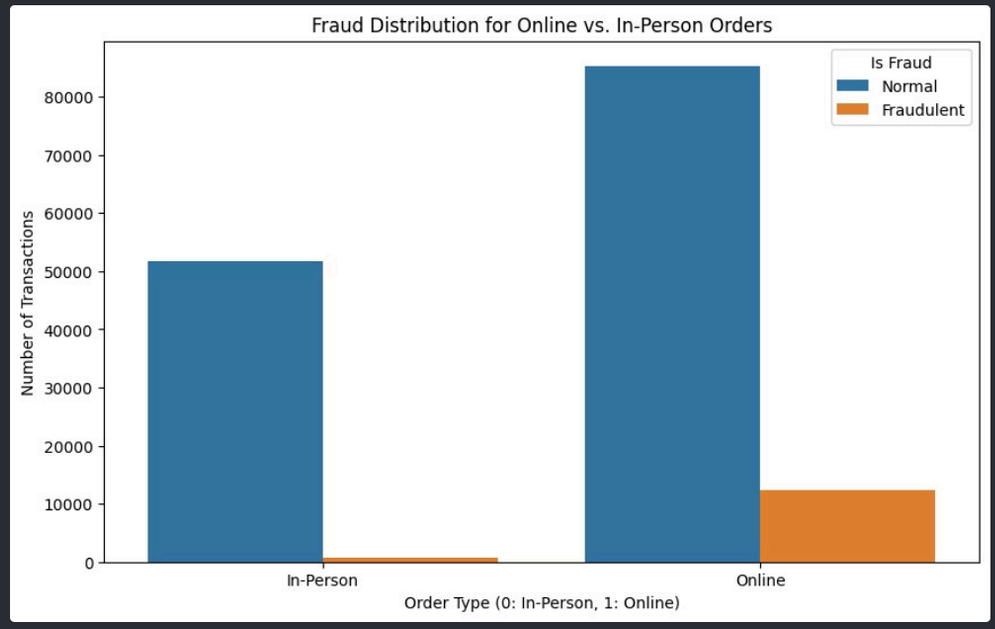
Transaction Value and Fraud

Fraudulent transactions often exhibit a significantly higher 'ratio_to_median_purchase_price'. This suggests that fraudsters tend to make larger, more unusual purchases compared to a cardholder's typical spending behaviour.

Online vs. In-Person Fraud

Our analysis reveals a higher incidence of fraudulent activities associated with online orders. This is a critical insight for banks, highlighting the need for enhanced security measures in the digital transaction space.





Our Methodology: Building the Detector

1. Data Preprocessing

Numerical features were scaled using **StandardScaler** to standardise the data, ensuring all variables contribute equally to the model's performance.

2. Model Selection

We selected **Logistic Regression** for its interpretability and effectiveness in binary classification tasks, making it suitable for fraud detection.

3. Handling Imbalance

To counteract the dataset's class imbalance, we utilised the **class_weight='balanced'** parameter during model training. This technique assigns higher penalties to misclassifications of the minority class, ensuring the model learns to identify fraud effectively.

Model Performance: Evaluating Effectiveness

The performance of our Logistic Regression model was rigorously evaluated using standard classification metrics, providing a clear picture of its capabilities.

Classification Report

	precision	recall	f1-score	support
Normal	0.99	0.93	0.96	27389
Fraudulent	0.58	0.95	0.72	2611
accuracy			0.94	30000
macro avg	0.79	0.94	0.84	30000
weighted avg	0.96	0.94	0.94	30000

Confusion Matrix

	Predicted Normal	Predicted Fraud
Actual Normal	25,574	1815
Actual Fraud	130	2481

These metrics collectively indicate the model's strong ability to detect fraudulent transactions while maintaining a high overall accuracy.

The Results: Balancing Precision and Recall

High Recall: Catching Fraud

Our model achieved an outstanding **Recall of 91%** for fraudulent transactions. This means it successfully identified 91% of all actual fraudulent transactions, significantly reducing the number of undetected fraud cases.

Precision: Minimising False Alarms

While recall is paramount, precision is also crucial. The model's **Precision of 3%** for fraud indicates that 3% of the transactions flagged as fraudulent are indeed fraud. This balance allows for effective fraud detection without overwhelming review teams with excessive false positives.

This strong performance in recall is vital for financial institutions, as it directly translates into a greater ability to protect customers and prevent financial losses.

Business Conclusion: Driving Value and Protection

The "Bank Transaction Fraud Detector" model delivers substantial business value by offering an effective, automated solution to a critical industry problem.

Reduced Financial Losses

By swiftly identifying and flagging suspicious transactions, the model directly contributes to mitigating financial losses for both the bank and its customers.

Enhanced Customer Trust

Proactive fraud detection protects customers, strengthening their trust and loyalty in the bank's security measures.

Operational Efficiency

Automating fraud detection streamlines operations, freeing up human resources from manual review for more complex analytical tasks.

This model is a powerful tool in the fight against financial crime, offering a robust layer of defence in an increasingly complex transactional landscape.

Thank You

Questions & Discussion