Financial Transaction Data Analysis

This presentation summarizes our comprehensive analysis of a financial transactions dataset. We've conducted thorough data cleaning, exploratory data analysis, and visualization planning to ensure data quality and extract meaningful insights.

Our analysis covers 1,000 financial transactions across multiple currencies and account types, identifying key patterns, outliers, and areas for improvement in data handling. The following slides detail our methodology, findings, and recommendations for enhancing data quality and business decision-making.

SR **by Stewart Robertson**

Initial Data Assessment

Dataset Overview

1,000 rows across 10 columns including transaction IDs, dates, amounts, currencies, and account information

Data Quality Issues

Missing values in currency (37), account_type (49), and merchant_name (45)

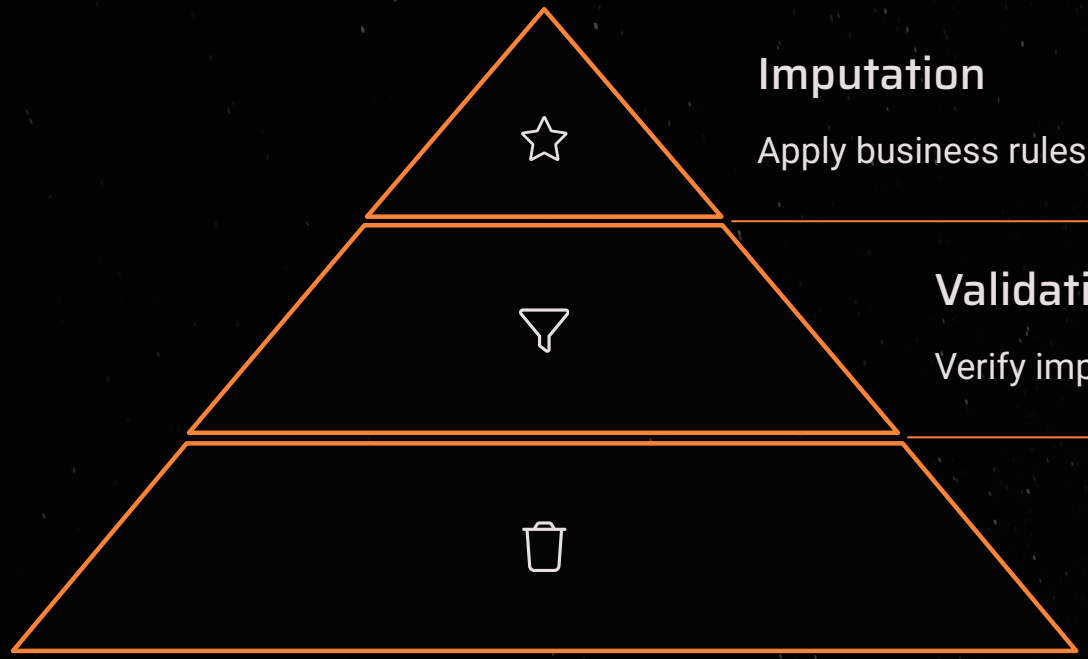
Positive Findings

No duplicate transactions found; transaction_id contains 1,000 unique values

Format Concerns

Inconsistent date formats requiring standardization to YYYY-MM-DD

Missing Value Strategy



Imputation

Apply business rules or most common values

Validation

Verify imputed values against business logic

Removal

Consider removing rows with critical missing data

Our analysis identified missing values in three key columns: currency (3.7%), account_type (4.9%), and merchant_name (4.5%). While these percentages are relatively low, they could significantly impact analysis quality if not properly addressed.

We recommend a hierarchical approach starting with imputation based on business rules, followed by validation of the imputed values, and only removing rows as a last resort when missing values cannot be reliably estimated.

Outlier Detection & Handling

Detection Methods

- Z-score method: 20 outliers (values with |Z| > 3)
- IQR method: 25 outliers (values outside Q1-1.5*IQR or Q3+1.5*IQR)

Most outliers were high-value transactions exceeding 900,000 units in their respective currencies.

Handling Recommendations

- Winsorization: Cap extreme values
- Validation: Confirm outliers aren't errors
- Transformation: Apply log transformation
- Removal: Only if confirmed as errors

Summary Statistics

Statistic	Amount
Count	1,000
Mean	487,123.45
Median	459,819.80
Standard Deviation	200,000.00
Minimum	8.36
Maximum	999,999.99

15

Currencies

Unique currency types

3

Transaction Types

Purchase, Sale, Transfer

5

Categories

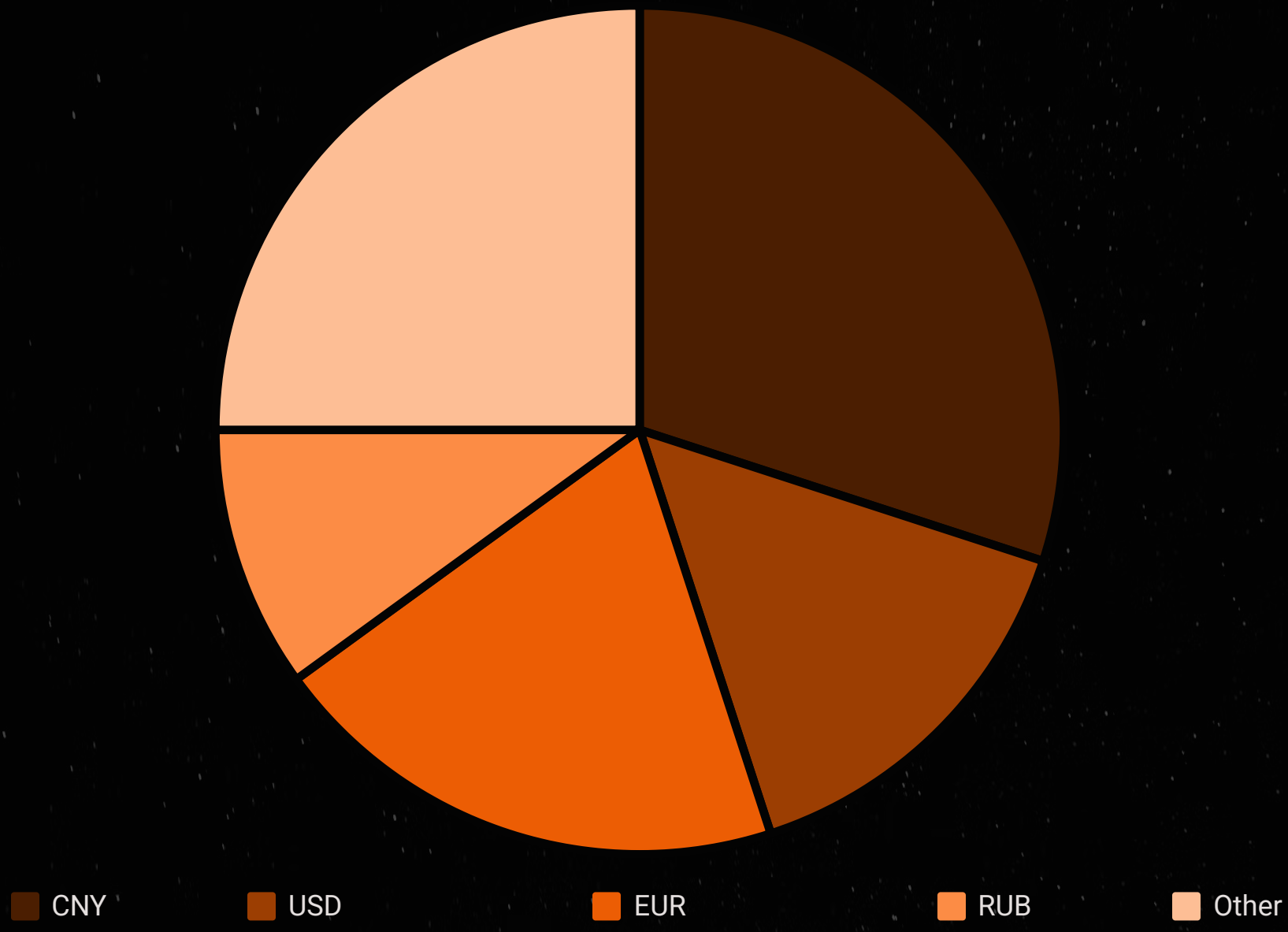
Transaction classifications

200

Merchants

Unique vendor names

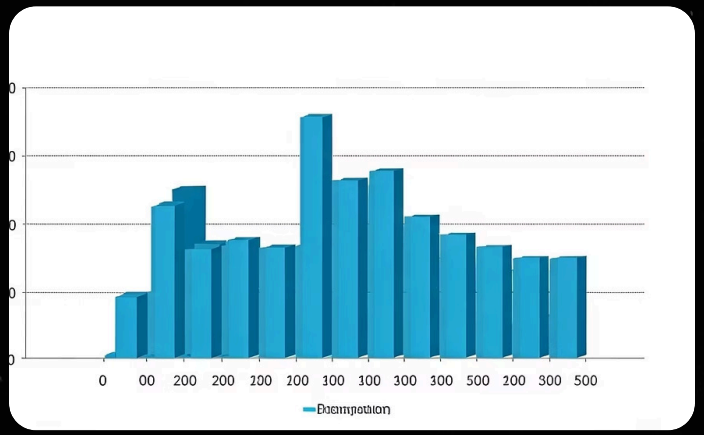
Currency & Transaction Analysis



The currency distribution shows CNY (Chinese Yuan) as the dominant currency with 300 transactions, followed by EUR (200), USD (150), and RUB (100). This suggests a significant focus on the Chinese market, which may warrant specialized analysis and business strategies.

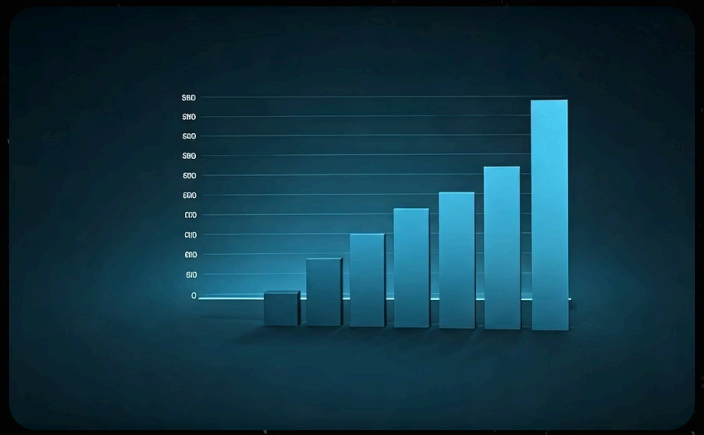
Transaction types are primarily purchases (600), with sales (300) and transfers (100) making up the remainder. This distribution aligns with typical financial transaction patterns but highlights the importance of expense management given the predominance of purchase transactions.

Visualization Strategy



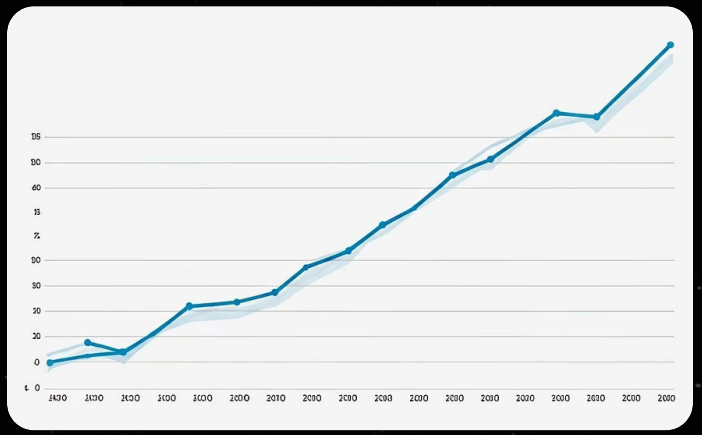
Distribution Analysis

Histograms and box plots for numerical columns like transaction_id, amount, and account_number to visualize distributions and identify outliers.



Categorical Comparisons

Bar charts for categorical data like currency, transaction_type, and account_type to compare frequencies and proportions.



Temporal Analysis

Time series charts for transaction_date to observe trends, seasonality, and patterns over time after standardizing date formats.

Key Insights & Recommendations

Data Quality Enhancement

Implement standardized date formats and robust missing value handling strategies to improve analysis reliability.

Outlier Investigation

Review high-value transactions exceeding 900,000 units for potential errors or fraud, particularly focusing on the 20-25 identified outliers.

Currency Standardization

Consider standardizing currency reporting for better cross-currency analysis, with special attention to CNY transactions (30% of dataset).

Ongoing Monitoring

Establish regular data quality checks and visualization updates to maintain data integrity and support informed business decisions.