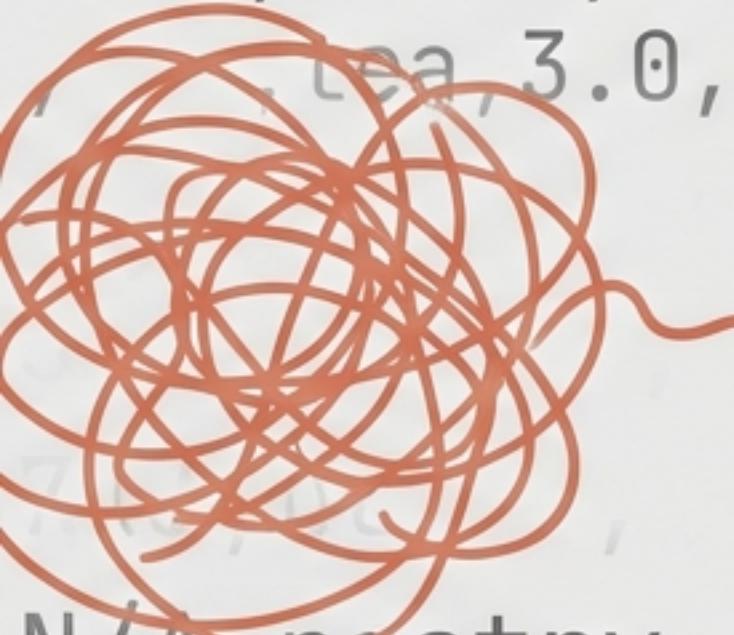


Transaction_ID,Item,Price,,
ERROR,coffee,2.0,,
2023-Unknown,,,NaN
3456,sandwich,5.50,,
N/A,pastry,,,
5678,latte,4.00,,
,tea,3.0,,
N/A,pastry,,,
5678,latte,4.00,,
,tea,3.0,,
7890,Unknown,NaN,,
ERROR,muffin,,,



Cafe Sales Analytics

Transforming Raw Data into Business Intelligence

An end-to-end analysis from Python-based data cleaning to Tableau dashboarding and strategic roadmap.

Executive Summary: Converting Data Chaos into Revenue Strategy

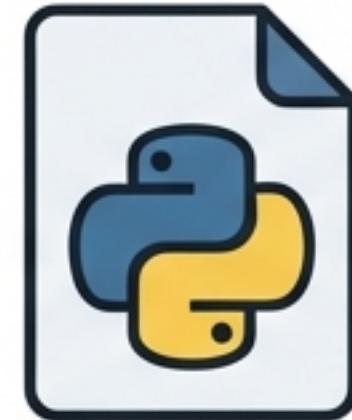
The Goal

Transform unstructured, dirty transaction logs into a navigational dashboard for business owners.



The Methodology

Utilized Python (Pandas/NumPy) for a 5-phase cleaning pipeline. Key Innovation: 'Deductive Imputation' recovered 45% of missing item data using price logic.



Key Findings

Total Revenue: \$73,000

Transactions: 8,159

AOV: \$8.9



Top Recommendation

Urgent operational audit required. 31% of transactions have "Unknown" payment methods, posing a severe reconciliation risk.

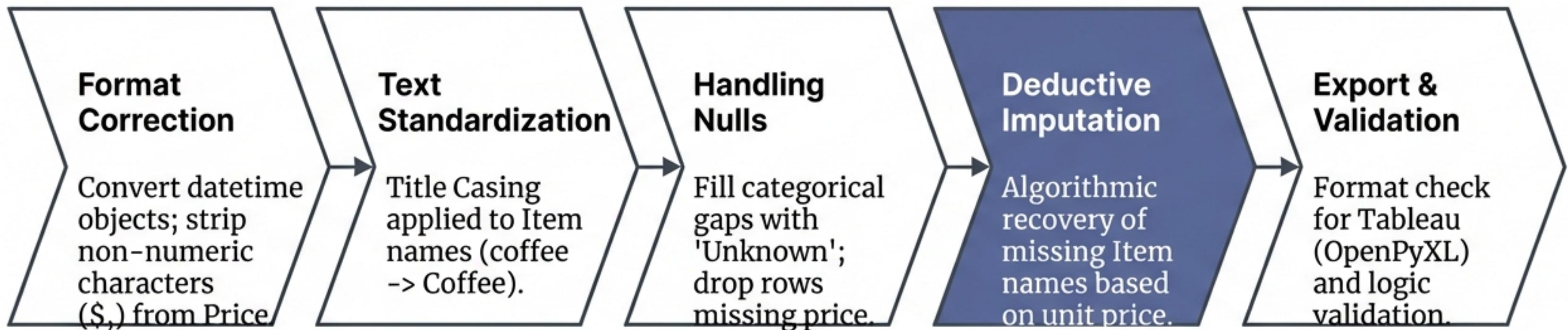


Business insights were trapped behind significant data quality barriers

1. Wrong Data Types: Transaction dates in string format.	2023/01/15	coffee	2.0	Cash	3. Missing Values: Critical gaps in Location and Payment Method.
2. Inconsistent Text: Case sensitivity issues create duplicate items.	Unknown Date	Coffee	\$2.00	Unknown	
	2023-01-16	NaN	2.0	Credit	4. Structural Errors: Rows missing essential transaction info.
	ERROR				

Without intervention, these errors prevent accurate revenue tracking and inventory planning.

A 5-Phase Python Pipeline ensured data integrity



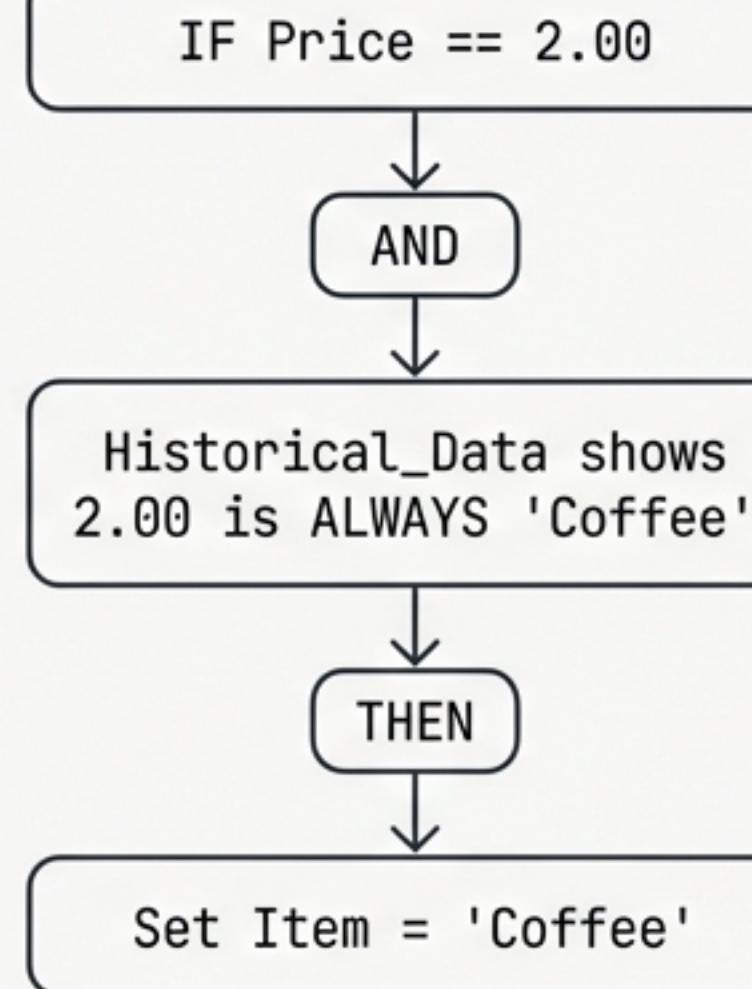
Deductive Imputation recovered 45% of missing item data

Raw Data

Item:	Unknown
Price:	\$2.00
Status:	Valid Transaction

JetBrains Mono
Transactions with valid prices
but missing names would
normally be deleted.

The Algorithm



Recovered Data

Item:	Coffee
Price:	\$2.00
Status:	Analytics Ready

JetBrains Mono
45% of 'Unknown' items were
salvaged, improving product
mix accuracy.

Technical Implementation & Validation

The Toolkit



Python 3.10
(Core Logic)



Pandas & NumPy
(Vectorization)



OpenPyXL
(Formatting)



Git
(Version Control)

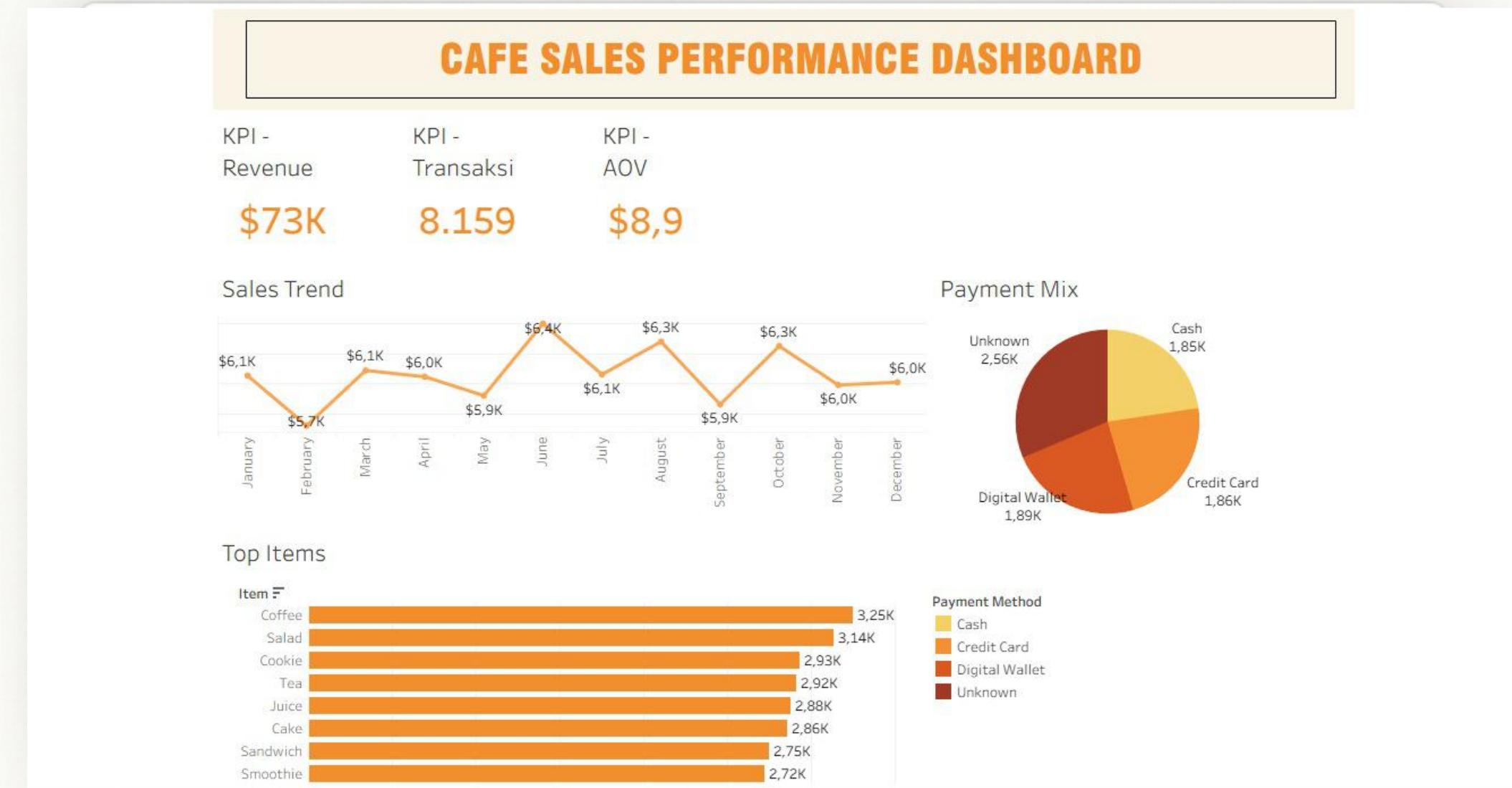
The Integrity Check

Before Tableau ingestion, data is validated against the master equation:

$$\text{Quantity} \times \text{Price} = \text{Total Spent}$$

✓ 0 Errors Found in Final Dataset.

The Dashboard reveals a \$73k business driven by high-volume transactions



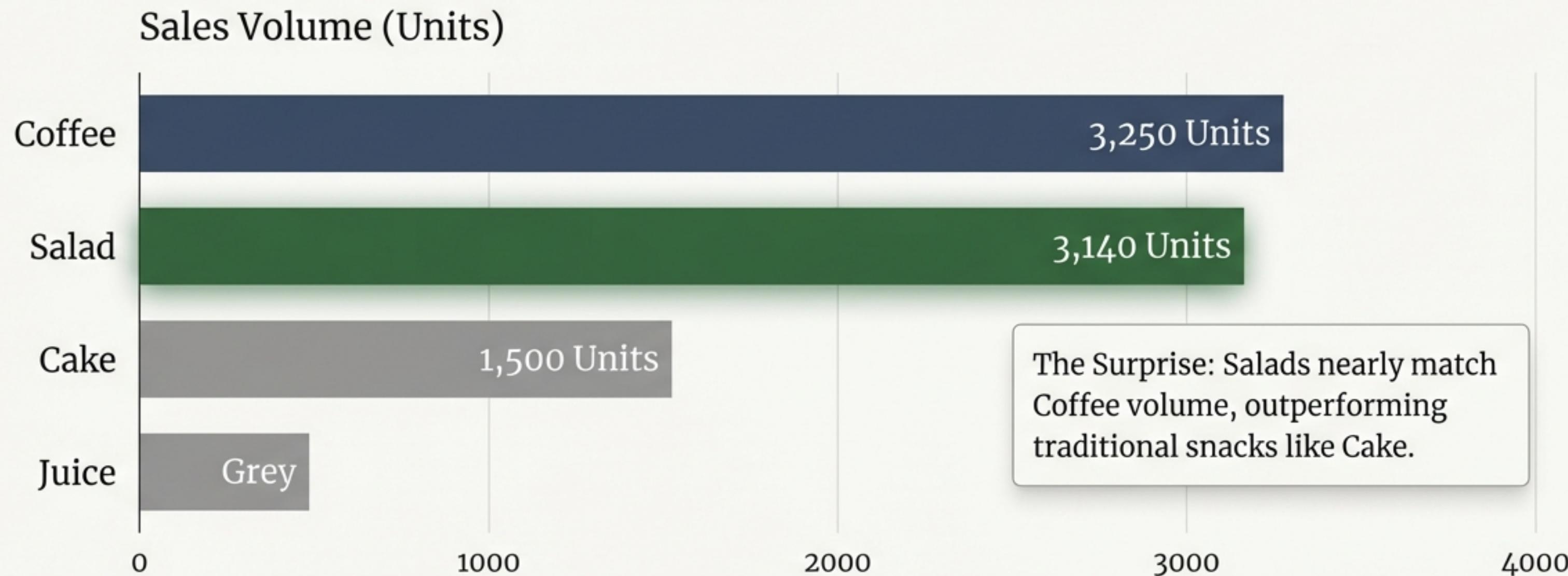
The dashboard consolidates 12 months of cleaned transaction data into a single interactive view for the business owner.

June drives annual revenue peaks, while February lags behind



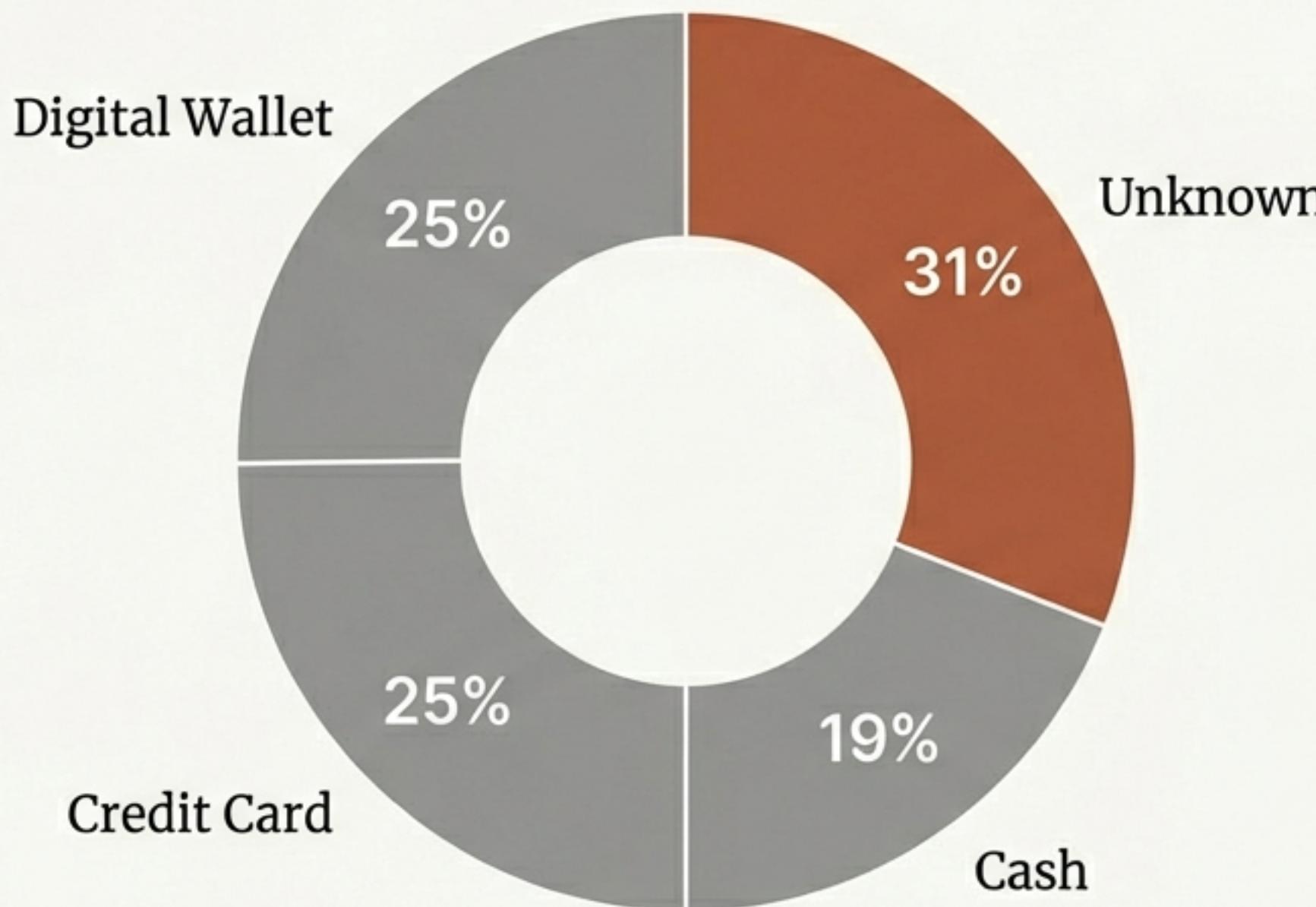
Volatility fluctuates between a floor of \$5.7k and a ceiling of \$6.4k, indicating stable but seasonal demand.

Customers are health-conscious, favoring Salads over Sweets



Insight: The customer base prioritizes lunch/meals over sweets.

31% of payments are unrecorded, creating a critical reconciliation risk



Impact Analysis

The Risk: 2,560 transactions cannot be reconciled against bank deposits.

Root Cause: Likely a POS (Point of Sale) system failure or a gap in staff training regarding payment entry.

Strategic Roadmap: Marketing, Menu, and Operations



Marketing

Prevent the Slump

Launch targeted campaigns in February and May. Use 'Valentine's Bundles' to smooth out revenue dips.



Menu

Bundling Strategy

Lift low-performing Smoothies (2.7k units) by pairing them with best-selling Salads in a 'Healthy Lunch Set'.



Operations

POS Audit

Immediate technical audit required to fix the 31% 'Unknown' payment anomaly. Retrain staff on transaction entry.

Project Resources & Credits

Analyst: DjuandaF13

Repository: GitHub / Project-Cafe-Sales

Tools: Python 3.10, Tableau Public, OpenPyXL

*Data cleaning is not just about error removal;
it is the process of recovering lost value.*