

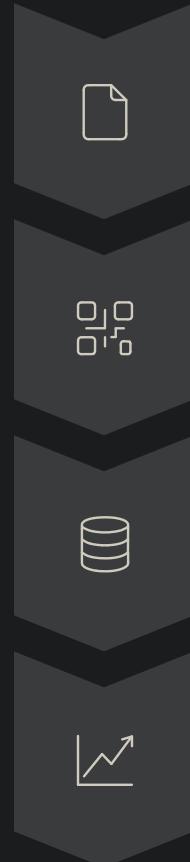
End-to-End Amazon Sales & Discount Analytics System

A Full-Stack Data Pipeline Integration — Python, PostgreSQL, Power BI

Raviteja Kudipudi

The Problem: Data Silos & Dirty Data

Raw e-commerce data → actionable business intelligence



Raw CSV

Unstructured strings, currency symbols, nulls

Python ETL

RegEx cleaning, normalization, feature engineering

PostgreSQL

Persistence, security, scalability

Power BI

Real-time interactive dashboards



Scope & Objectives



01

Automated Cleaning

Python RegEx for currency symbols & malformed strings

02

Database Integration

File-based → RDBMS for structured queries

03

Metric Engineering

Custom KPIs like "Total Amount Saved"

04

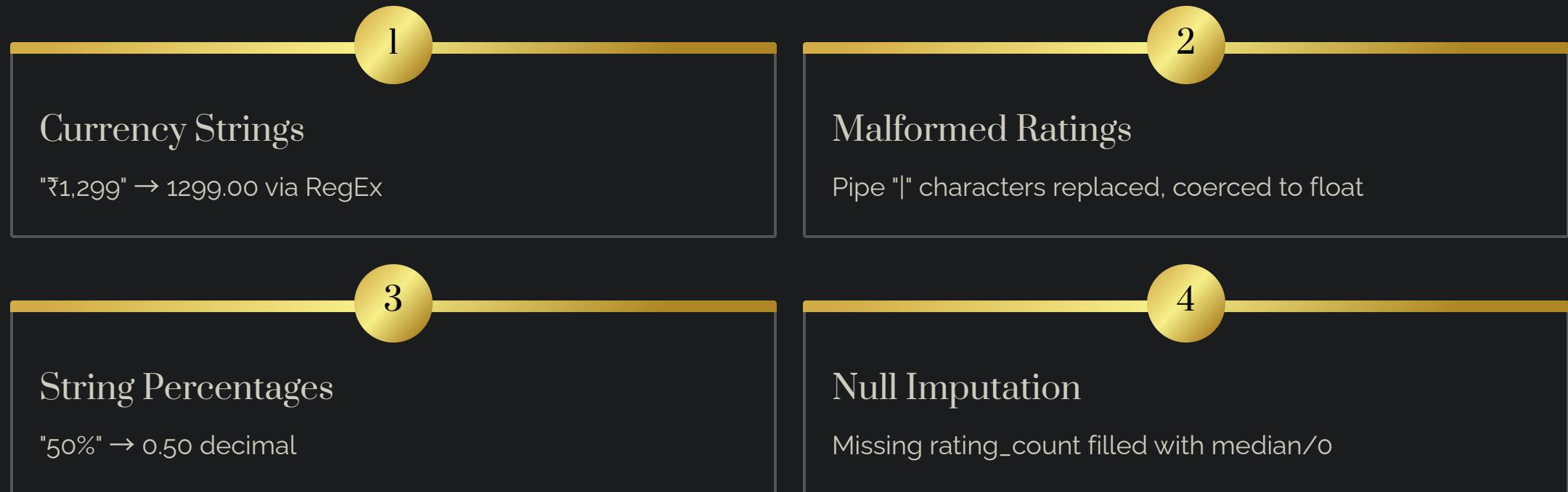
Interactive Visuals

Filter by category & performance instantly

- ❑ **Dataset:** 1,400+ Amazon products across Electronics, Home & Kitchen, Office Products — analyzing pricing, discounts, and ratings.

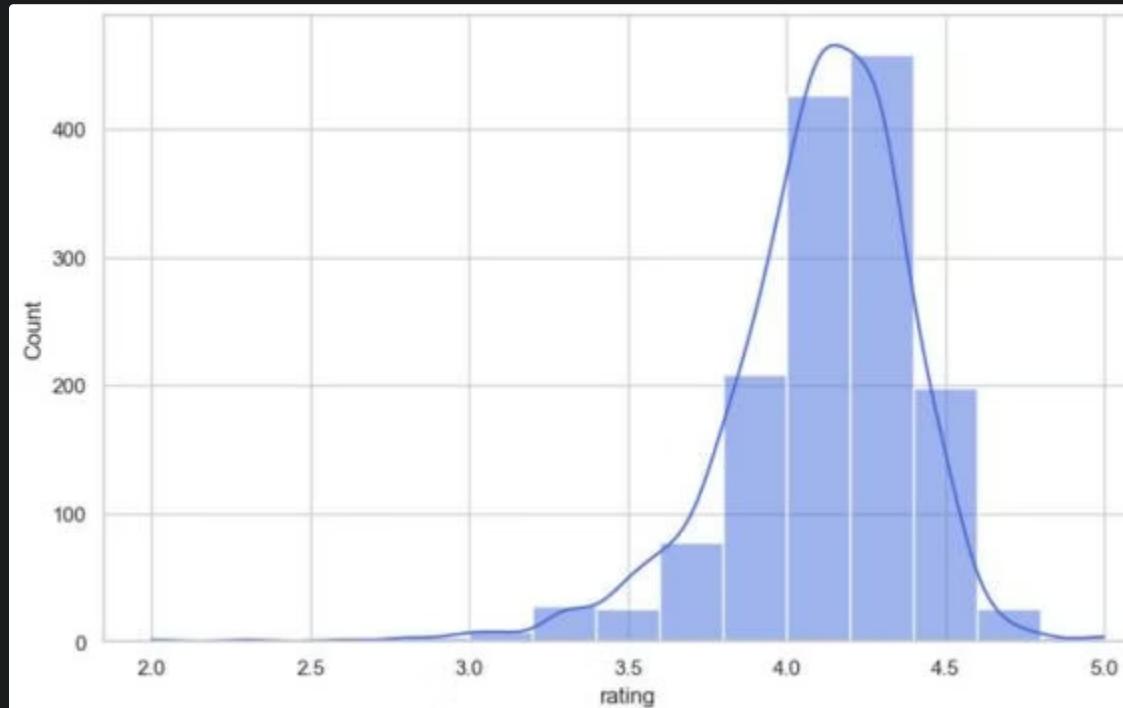
Python ETL: Cleaning the Mess

Raw data "smells" identified and resolved:



Feature Engineered: `Amount_Saved = Actual_Price - Discounted_Price`

What the Data Reveals

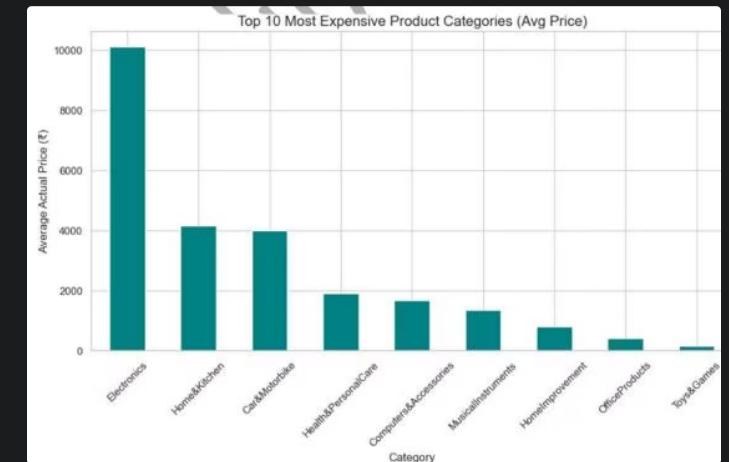
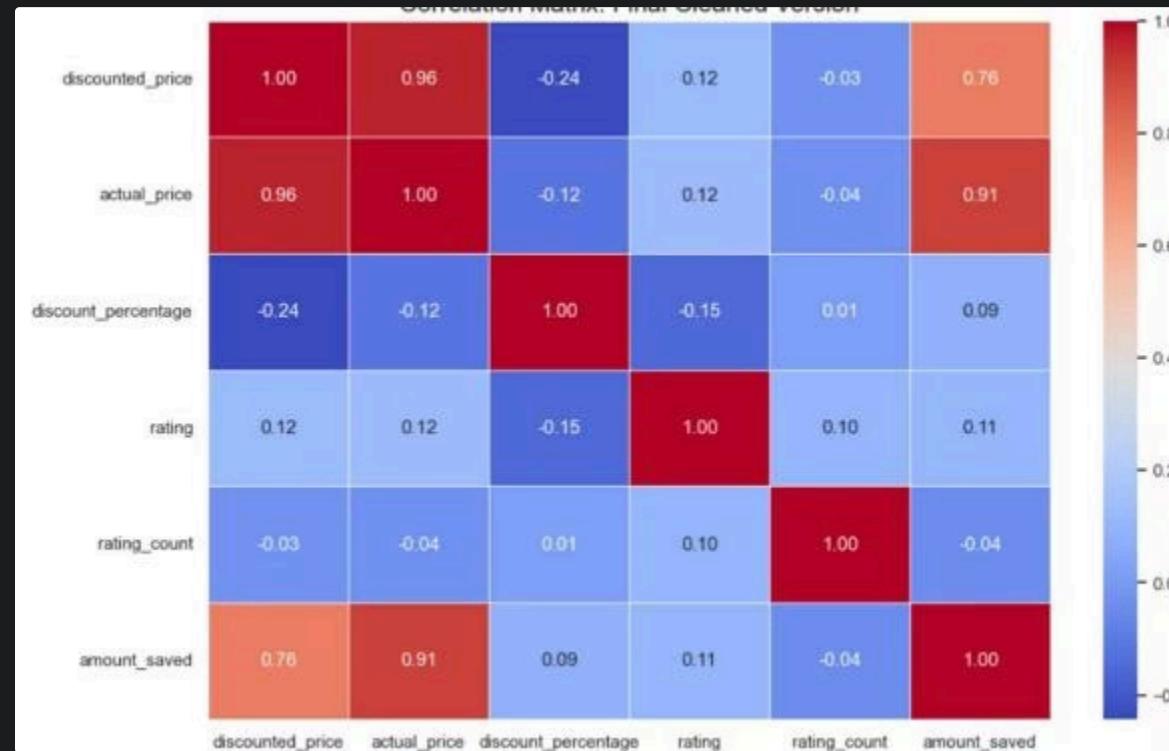


Ratings cluster around **4.0–4.3 stars** — right-skewed distribution



Slight positive trend — but **mid-range products outperform** premium items on ratings

Correlation Matrix & Category Pricing



0.96

Actual vs. Discounted Price
correlation

-0.15

Discount % vs. Rating — deep
discounts hurt trust

#1 Electronics

Highest avg price at ~₹10,000

PostgreSQL: Single Source of Truth

Schema Highlights

- `product_id` → VARCHAR(50), PK
- `actual_price` → NUMERIC(12,2)
- `rating` → DECIMAL(3,2)
- `category` → TEXT

Injection & Validation

SQLAlchemy bridge with `to_sql(if_exists='replace')` — idempotent pipeline

1,465 rows

Migrated & confirmed via `SELECT COUNT(*)`

Zero nulls

On `product_id` — validated in pgAdmin

CHAPTER 5 · PHASE III

Power BI Dashboard

₹7.98.. 26.77.. 48%

Total Revenue

Total Ratings

Avg Discount

Dark UI design with F-pattern KPI layout · Interactive category slicer for cross-filtering · Bar chart + donut chart for volume & savings



Key Findings & Business Insights

Quality-Price Gap

Mid-range products (₹500–₹2,000) rated **4.3+** — outperforming premium items. Customers prioritize value over brand.

Discount Sweet Spot

35%–50% discounts drive highest engagement. Discounts >70% plateau — "too good to be true" deters shoppers.

Category Competition

Electronics & Computers lead revenue but face highest competition — new entries need heavy discount strategy.



CONCLUSION

Summary & Future Scope

Data Resilience

Messy real-world strings handled via Python ETL

Structural Integrity

RDBMS schema ensures financial precision

Visualization Excellence

High-contrast, user-centric Power BI interface