Customer Analytics & Pricing Logic Deep-Dive

This project represents a technically rigorous and business-driven analytical study of an e-commerce dataset. Rather than focusing on surface-level metrics or visualization-heavy exploration, the analysis prioritizes data validation, pricing logic verification, and decision-making grounded in real commercial behavior. All insights, assumptions, and exclusions are explicitly justified through SQL-based reasoning.

Business Problem Definition

E-commerce pricing is not determined by unit price alone. Delivery charges, discounts, tax policies, and transaction structure collectively shape the final invoice value and customer perception. This project addresses the following core questions:

- Are pricing components applied consistently and logically across transactions?
- Which extreme values represent data quality issues versus valid commercial scenarios?
- How do delivery, tax, and discount mechanisms interact to influence final pricing?
- Are customer and product behaviors aligned with expected business rules?

Data Source

The dataset was sourced from Kaggle and simulates a real-world e-commerce environment. It includes transactional, customer, pricing, discount, tax, and marketing spend data.

Dataset Structure & Scope

The analysis is based on five core tables. Only analytically relevant features are highlighted below.

Table	Rows	Key Features	Analytical Role
customers	1,468	Gender, Location, Tenure_Months	Customer segmentation & tenure analysis
discount_coupon	204	Product_Category, Coupon_Code, Discount_pct	Discount logic & campaign validation
marketing_spend	365	Offline_Spend, Online_Spend	Contextual marketing reference
online_sales	52,924	Transaction_ID, SKU, Quantity, Unit_Price, Delivery_Charges	Core transactional analysis
tax_amount	20	Product_Category,	Tax application

GST verification

Analytical Methodology

1. Data Quality Assessment: Missing values, duplicate transactions, and structural inconsistencies were systematically identified across all tables.
2. Validation of Commercial Logic: Pricing components were evaluated at the correct business granularity (transaction-level vs. row-level).
3. Layered SQL Modeling: CTEs and Views were used to isolate pricing effects and prevent logic leakage.
4. Decision-Oriented Analysis: Each analytical step resulted in an explicit decision—retain, exclude, or reinterpret the data.
5. Business Translation: Technical findings were translated into pricing, tax, and delivery policy recommendations.

Analytical Context & Interpretation Boundaries

Several analyses revealed limited variance across customer segments and product categories.
This outcome is not attributed to analytical limitations, but rather reflects the structure of the underlying business model, which enforces largely uniform pricing, tax, and delivery policies.

Recognizing and validating the absence of strong segmentation is a meaningful analytical outcome, as it highlights structural constraints within the commercial system and prevents the introduction of artificial or misleading insights.

Key Analytical Decisions

- Delivery charges were calculated per transaction rather than per row, reflecting real-world invoicing behavior.
- High delivery values were retained after validation, as they correspond to bulk or low-priced orders, not data errors.
- Delivery-to-price ratios exceeding 100% were analyzed contextually instead of being removed as outliers.
- Coupon status was deliberately excluded from unit price logic and evaluated at the invoice level.
- Tax impact was isolated from discounts to avoid misattributing price inflation to promotional activity.
- Identical SKUs with inconsistent pricing were flagged as pricing logic gaps rather than noise.

Core Insights

- Approximately 1.7% of transactions exhibit delivery costs equal to or exceeding product value, primarily due to fixed delivery fees on symbolic or low-priced items.
- Smart device categories experience stronger tax impact despite lower tax rates, driven by

higher base prices.
- Unit price inconsistencies for identical SKUs suggest undocumented pricing rules or data entry issues.
- Customer purchasing behavior is stable across gender and tenure, reinforcing category-level rather than demographic-driven strategies.

Business Recommendations

- Reevaluate delivery pricing for low-priced, single-unit products to prevent customer friction.
- Document and standardize SKU pricing logic to reduce unexplained variance.
- Treat tax impact as a primary pricing driver for high-value categories, not as a secondary adjustment.
- Introduce compensatory discounts for high-tax, low-demand categories to balance final invoice perception.
- Maintain transaction-level pricing validation as part of routine data quality monitoring.

Technical Depth & SQL Design

The project demonstrates advanced SQL usage beyond exploratory querying：
- Extensive use of (CTEs, Nested Subqueries and Views) for stepwise validation and analytical clarity
- Window functions for ranking, anomaly detection, and behavioral comparison
- Transaction-level aggregation logic to align analysis with business reality
- Defensive handling of NULLs using COALESCE to prevent silent calculation bias
- Reusable SQL views to abstract complex pricing logic into a single analytical layer

Note on Visualization Tools

Power BI was intentionally excluded from this project. The analytical objectives required transaction-level validation, pricing logic verification, and anomaly interpretation—tasks better served through SQL-based analysis. Visualization was deprioritized in favor of analytical depth. Future projects will incorporate BI tools where visual storytelling adds incremental decision value.