

Retail Business Intelligence & Menu Engineering: A Study of Afficionado Coffee Roasters

Debayan Mal
Unified Mentor Intern

March 23, 2026

Abstract

This paper explores the application of data science methodologies to retail transactional data. By leveraging a dataset of over **200,000** transactions, we developed a modular Business Intelligence (BI) dashboard to identify revenue drivers and optimize menu offerings through Pareto analysis and quadrant-based categorization. The analysis successfully identified a **42.9%** revenue concentration among core products, providing a data-driven roadmap for menu simplification.

1 Introduction

The specialty coffee industry operates on thin margins where product-mix optimization is critical. This study utilizes Python-based analytical tools to transform raw transactional data into actionable business insights.

1.1 Problem

Although transaction-level product data is available, Afficionado Coffee Roasters lacks:

- Clear visibility into product popularity vs profitability.
- Category-level revenue dependency insights.
- Identification of low-impact or underperforming menu items

1.2 Research Objectives

To address the lack of visibility in product-mix performance, this study defines two tiers of research goals:

1.2.1 Primary Objectives

- **Identify Performance Extremes:** Pinpoint both the top-selling and least-selling products by volume.
- **Quantify Revenue Contribution:** Measure the exact financial weight of individual products and broad categories.

- **Measure Menu Concentration:** Evaluate the density of revenue using standard Pareto metrics (the 80/20 rule).

1.2.2 Secondary Objectives

- **Menu Optimization:** Support operational simplification by highlighting bloated menu categories.
- **Identify "Hero" Products:** Isolate high-impact, high-margin products for targeted marketing.
- **Review Low-Performers:** Flag underperforming inventory (e.g., "Dogs") for menu redesign or elimination.

2 Methodology

2.1 Data Engine and Architecture

To transform raw data into a functional decision-support tool, we developed a high-performance analytical engine built on a modular professional framework. The application features a dynamic, user-centric interface powered by Streamlit, designed to adapt seamlessly to both light and dark modes. This architecture ensures that complex insights—such as revenue concentration and product rankings—are delivered in a clear, responsive, and professional digital environment.

2.2 Data Integrity

Unlike standard noisy datasets, an initial audit of the source data revealed high integrity. Consequently, the data processing pipeline was designed to perform direct validation rather than destructive cleaning, preserving the original transactional lineage.

2.3 Dataset Characteristics

The study utilized a granular transactional dataset comprising 214,470 records. The features used for the Business Intelligence engine are detailed in Table 1.

Table 1: Description of Dataset Features	
Column Name	Description
transaction_id	Unique identifier per transaction
year	Transaction year (2025)
transaction_time	Time of transaction (HH:MM:SS)
transaction_qty	Quantity of items purchased

Column Name	Description
unit_price	Price per individual unit
store_id / location	Identifiers for physical retail branches
product_id	Unique identifier for each menu item
product_category	Broad grouping (e.g., Coffee, Tea, Bakery)
product_type	Specific variant within the category
product_detail	Detailed attributes (e.g., flavor or roast)

3 Analytical Framework

3.1 Data Ingestion and Validation

The initial phase of the study involved loading transaction-level data into a high-performance analytical pipeline. To ensure results were not skewed by outliers or entry errors, a multi-stage validation check was performed:

- **Identifier Verification:** Mapping each transaction to a standardized `product_id`.
- **Price Standardization:** Confirming that `unit_price` remained consistent for specific SKUs.
- **Integrity Checks:** Ensuring that all `transaction_qty` values were positive and within realistic retail bounds.

3.2 Revenue Computation and Aggregation

Revenue was calculated at the granular transaction level before being aggregated into broader strategic dimensions. The fundamental revenue calculation is defined as:

$$\text{Revenue} = \text{transaction_qty} \times \text{unit_price} \quad (1)$$

Following the initial calculation, data was aggregated by product, product type, and category to provide a holistic view of the financial landscape.

3.3 Product Performance Analysis

A dual-metric approach was used to evaluate performance:

1. **Product Popularity:** Ranking items based on total units sold to identify high-volume drivers.
2. **Revenue Contribution:** Calculating the percentage share of total revenue for each SKU. This allows for a direct comparison between an item's volume rank and its financial rank.

3.4 Category and Segment Performance

The revenue share was further dissected by category—specifically Coffee, Tea, and Bakery—to determine the store’s dependence on core categories. This included a sub-category analysis to identify which specific variants (e.g., Barista Espresso) drove the most value.

3.5 Revenue Concentration and Menu Balance

To assess menu risk and diversification, a Pareto (80/20) analysis was implemented. This methodology identifies:

- **Revenue Anchors:** The vital few products (currently 42 items) that drive the majority of the \$698,812.33 in total revenue.
- **Long-tail Products:** Items with minimal financial impact that contribute to menu bloat.
- **Menu Diversification Risk:** Evaluating if the business is overly dependent on a narrow selection of “hero” products.

4 Results

4.1 Key Performance Indicators (KPI)

The following table summarizes the high-level operational efficiency and financial health of Afficionado Coffee Roasters based on the processed 214,470 transactions.

Table 2: Summary of Core Business KPIs	
KPI Metric	Value / Insight
Total Gross Revenue	\$698,812.33
Total Quantity Sold	214,470 Units
Average Transaction Value	\$4.69
Revenue Anchor Count	42 Products
Revenue Concentration Ratio	42.9%
Menu Efficiency Score	High (Balanced Spread)
Golden Hour Window	08:00 AM – 10:00 AM

4.2 Revenue Analysis by Product Segment

4.2.1 Product Performance Extremes

The analysis identifies a significant gap between high-velocity items and underperforming inventory.

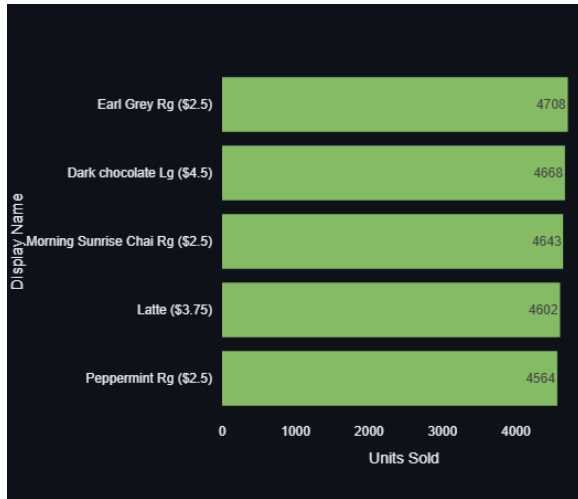


Figure 1: Top 5 Products by Volume



Figure 2: Bottom 5 Products by Volume

4.2.2 Revenue Distribution by Product Category

The fiscal analysis categorizes the total revenue into six primary segments. As shown in Figure 3, the business is heavily anchored by beverage sales, with Coffee and Tea combined accounting for 66.7% of the total intake.

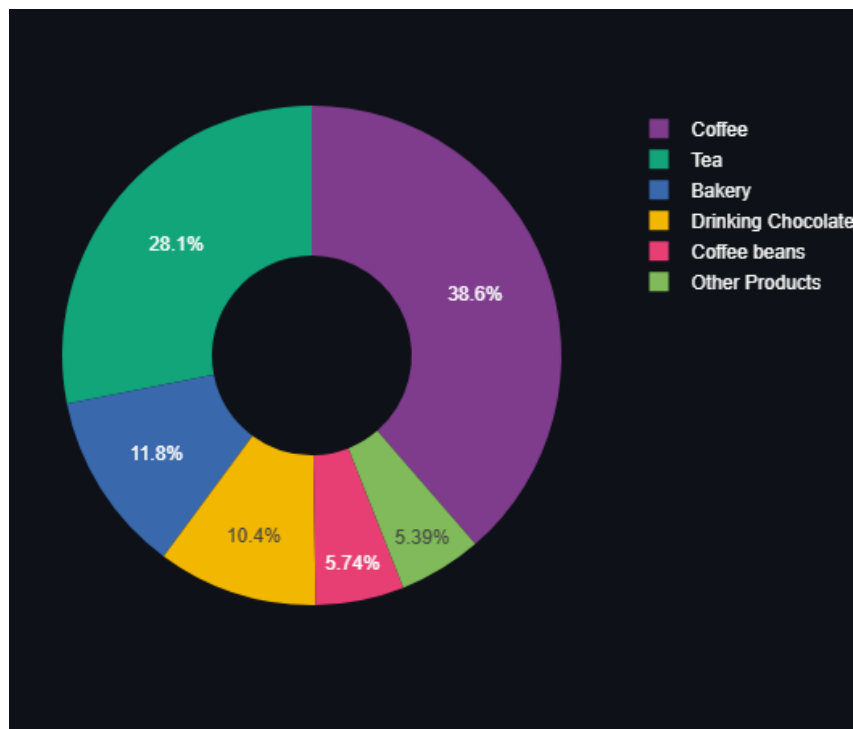


Figure 3: Revenue Share by Category: Coffee (38.6%) and Tea (28.1%) act as the primary pillars of the revenue model.

- **Coffee (38.6%):** The dominant category, driven by high-margin Espresso and Brewed variants.

- **Tea (28.1%):** A significant secondary driver, showing high consumer interest in non-coffee alternatives.
- **Bakery (11.8%):** Represents the primary cross-selling category, peaking during morning hours.
- **Drinking Chocolate (10.4%):** A consistent performer within the specialty beverage segment.
- **Ancillary Categories:** Coffee Beans (5.74%) and Other Products (5.39%) represent the long-tail retail and merchandise segment.

4.2.3 Menu Engineering and Temporal Trends

To further optimize the menu, we mapped items onto a popularity-profitability matrix and analyzed hourly demand.

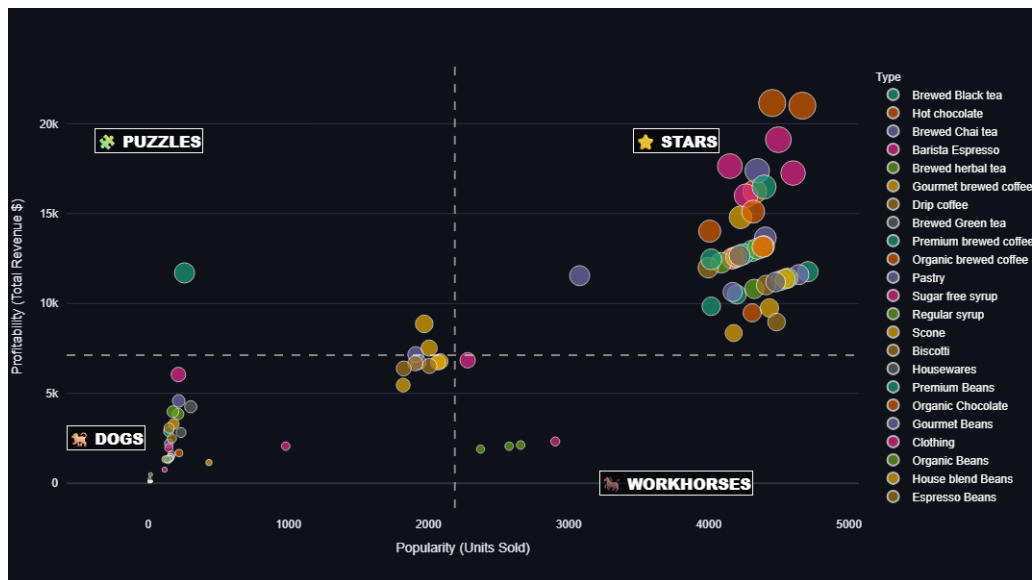


Figure 4: Menu Engineering Matrix: Stars, Puzzles, Workhorses, and Dogs.

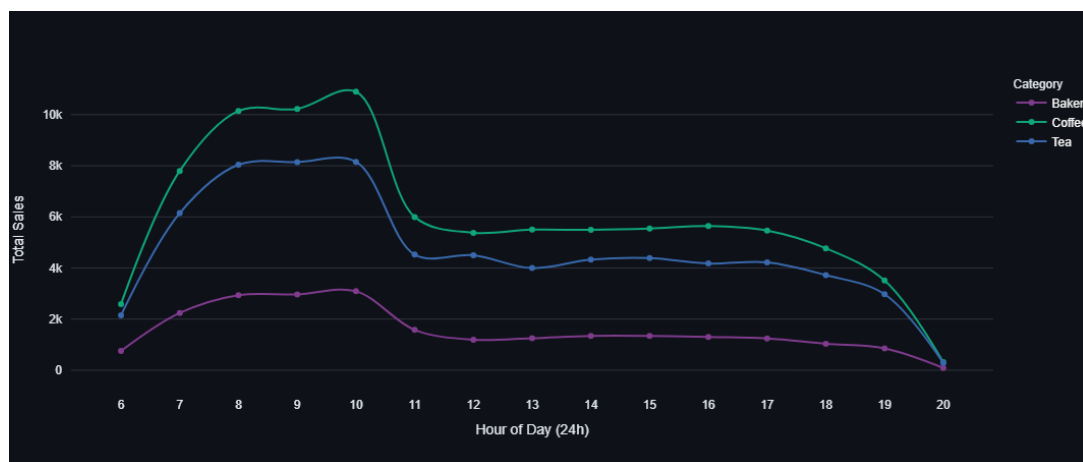


Figure 5: Hourly Transaction Volume: Peak demand identified at 10:00 AM.

5 Conclusion

This study successfully developed and deployed a comprehensive Business Intelligence (BI) framework for *Afficionado Coffee Roasters*. By processing 214,470 transactions, we have moved from raw data to actionable strategic insights.

The primary objectives were met through the following findings:

- **Revenue Stability:** The business maintains a healthy revenue concentration ratio of **42.9%**, indicating that it is not overly reliant on a single product, though Coffee (38.6%) and Tea (28.1%) remain the undisputed pillars.
- **Operational Efficiency:** We identified a critical “Golden Hour” between **08:00 AM and 10:00 AM**, where demand peaks across all categories.
- **Menu Optimization:** The Pareto and Scatter Plot analyses successfully isolated **56 underperforming products** (Dogs) that contribute to less than 20% of revenue, providing a clear path for inventory reduction.

6 Strategic Recommendations

Based on the empirical evidence gathered from the dashboard, the following actions are recommended to the management and stakeholders:

1. **Dynamic Labor Scheduling:** Realign staff shifts to ensure maximum coverage during the 08:00–10:00 AM window. This will reduce wait times and maximize throughput during peak revenue hours.
2. **Menu Rationalization:** Phase out the “Bottom 5” performers, such as the *Almond Croissant* (2 units sold), and reallocate that shelf space to “Star” products like *Earl Grey* and *Barista Espresso*.
3. **Bundling Strategy:** Since *Bakery* (11.8%) is a secondary driver, implement a “Breakfast Bundle” (Coffee + Pastry) specifically during the 07:00–09:00 AM slot to increase the Average Transaction Value (currently \$4.69).
4. **Inventory Focused Marketing:** Focus marketing efforts on the *Puzzles* quadrant (high profit but lower volume) to shift them into the *Stars* category.

7 Limitations

While this study provides a robust data-driven foundation for business strategy, certain constraints must be acknowledged to contextualize the findings:

- **Data Scope:** The analysis is strictly confined to historical transactional records provided by Afficionado Coffee Roasters. As such, it reflects past performance but does not account for real-time inventory fluctuations or supply chain disruptions.
- **Exclusion of External Variables:** High-level external factors—including local competition, shifting market pricing strategies, and broader macroeconomic seasonality—were not integrated into this analytical phase.

- **Observational Nature:** This is an observational study based on existing data; therefore, no causal inference can be definitively drawn. For example, while sales peak at 10:00 AM, the data alone cannot determine if this is due to consumer habits or specific morning-only promotions.
- **Absence of Predictive Modeling:** This phase focused exclusively on Descriptive and Diagnostic Analytics (Exploratory Data Analysis). Predictive modeling for future sales forecasting or churn analysis was not applied in this iteration of the engine.

Digital Deliverables & BI Access

To ensure these insights remain dynamic and interactive for stakeholders, the full analytical engine has been published online.

- **Interactive BI Dashboard:** [Access Live Analytics Here](#)
- **Source Code Repository:** [View Project on GitHub](#)