**Sales Performance Analysis and Inventory Optimization for a Residential Supermarket**

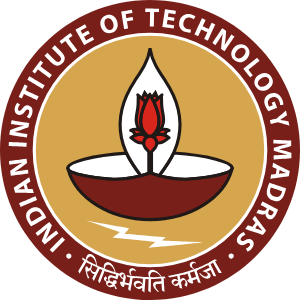
*Mid-Term for the BDM Capstone Project*

**Submitted by**

**Name: Tushar Jalan**

**Roll Number: 23f2003751**

**Email:** [**23f2003751@ds.study.iitm.ac.in**](mailto:23f2003751@ds.study.iitm.ac.in)



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

[**1. Executive Summary 2**](#_cef8eesgbcuc)

[**2. Proof of Data Originality 2**](#_613dujmqpkrv)

[**3. Metadata and Descriptive Statistics 3**](#_4hoz525tp752)

[3.1 Metadata Table 3](#_5at4kl2el0iy)

[3.2 Data Quality Assessment 4](#_9z760qlzwxtl)

[3.3 Descriptive Statistics 5](#_afaod4ev9w7l)

[3.4 Correlation Analysis 5](#_xreqj2vybto2)

[**4. Detailed Explanation of Analysis Process and Methodology 6**](#_b6yr889cn24)

[4.1 Data Cleaning and Preprocessing 7](#_4ro8551y47ut)

[4.2 Justification for the Ledger-Only Approach 7](#_6xs4v4ooyahn)

[4.3 Analytical Methodology 7](#_h3spr8ihldy1)

[4.3.1 Master View Construction 8](#_z6430mm5mf61)

[4.3.2 Multi-Dimensional SKU Classification 8](#_nyugnvnjgxdg)

[4.3.3 Seasonal Demand Analysis 9](#_33crnk49zewl)

[4.3.4 Forward Inventory Planning 10](#_t3itrto9x7dc)

[**5. Results and Findings 10**](#_2dghmi3cfrv)

[5.1 Inventory Portfolio Analysis 10](#_lulmc54p6tp9)

[5.2 Revenue Distribution and Pareto Analysis 10](#_9yauh9zfynff)

[5.3 Profitability and Capital Efficiency Analysis 11](#_n6hywgxqt4iq)

[5.4 Dead Stock Analysis - Capital Recovery Opportunities 11](#_krj5ormwv0nz)

[5.5 Seasonal Demand Patterns 12](#_pr7qv7x44ilm)

[5.6 Inventory Turnover and Efficiency Metrics 12](#_p0ni3jdhx0tt)

## 

## 

# 

# **1. Executive Summary**

Agarwalla Masala is a residential supermarket operating for over three decades in Angul, Odisha, serving the local community with grocery and household products. The business faces critical operational challenges including limited physical storage capacity constraining inventory to approximately 15-day cycles, recurring stockouts during peak festival seasons, and lack of granular SKU-level profitability analysis for strategic decision-making.

The analysis is based on primary data collected from the store's Stock Ledger File, containing 222 unique SKUs across multiple product categories including GROCERY, HOME CARE, PERSONAL CARE, and SPICES. The dataset encompasses complete transactional records with key variables including transaction dates, product categories, brands, stock movements (in/out), cost prices, and selling prices. Descriptive statistics reveal a highly skewed revenue distribution with mean annual revenue of ₹2,01,441 per SKU but median of only ₹14,950, confirming the Pareto principle where few products drive the majority of sales.

The methodology employs a ledger-only approach using automated Python scripts to construct a comprehensive Master Product View. Advanced inventory segmentation techniques including ABC analysis (revenue-based), XYZ analysis (volatility-based), and weighted ABC-XYZ classification have been implemented. Preliminary results identify 27 SKUs as "Dead Stock" representing tied-up capital, while highlighting top-performing products with ROI exceeding 65%. Seasonal analysis quantifies festival-driven demand spikes, providing a data-driven foundation for forward-looking inventory planning to optimize the store's limited storage capacity and maximize profitability.

# **2. Proof of Data Originality**

**Primary Data Sources:**

* **Organization Name:** Agarwalla Masala
* **Owner:** Mr. Nanak Ram Agarwal
* **Address:** Shop No. 23, NALCO Market Complex, NEAR UCO BANK, Angul, Odisha
* **GST Number:** 21ABYPA4798C1ZK

**Evidence of Primary Data Collection:**

1. **Letter of Authority:** [Authorization Letter](https://drive.google.com/file/d/10U9lzqwZpIE-DD1_cbeQUYRCHKd19wdg/view?usp=sharing)
2. **Interview Video:** [Tushar's interview with shop owner](https://drive.google.com/file/d/1A8deJKHG3p87XSGmOP2SVnbzJ4wcUl4_/view?usp=sharing)
3. **Shop Interior Video:** [Shop inside video](https://drive.google.com/file/d/105o35tOlQ1gp8_QzdR63e9mNQkzHq_su/view?usp=sharing)
4. **Additional Photographs:**
   * [Photo 1](https://drive.google.com/file/d/1037vfscsE2KDIDEAwRcxY4S5K_yKEZ5n/view?usp=sharing)
   * [Photo 2](https://drive.google.com/file/d/101wWBrUSBUx1PsyaeUOO94XjgBkA8JM2/view?usp=sharing)
   * [Photo 3](https://drive.google.com/file/d/106JzgpY2AYinjaUKK4dbd9Gy8s5TgiQT/view?usp=sharing)
5. **Google Maps Location:** [Google Maps](https://maps.app.goo.gl/iBeuwA4EBFLHdKiw7?g_st=aw)
6. **Data:** [Raw Data](https://drive.google.com/drive/folders/1olFlzBU1uJUTs0hW0nEsphZ8_GlPVo1O?usp=sharing) and [Cleaned Data](https://drive.google.com/drive/folders/1K8ziwy32UwhiCAj7rkHAi94WeF38g0l4?usp=sharing)

**Dataset Details:**

* **Primary Data File:** Stock\_Ledger\_Summary.csv
* **Data Collection Period:** One full year of transactional data **(01-01-2024 to 31-12-2024)**
* **Data Integrity:** Direct extraction from the store's accounting system with full owner’s knowledge.

# **3. Metadata and Descriptive Statistics**

## **3.1 Metadata Table**

The analysis utilizes variables extracted from the Stock\_Ledger\_Summary.csv file, representing the complete transactional history of the store.

| **Variable** | **Description & Processing** | **Relevance to Problem Statement** |
| --- | --- | --- |
| **Date** | Transaction date serving as primary index for time-series analysis | Essential for identifying seasonal trends, tracking sales patterns, and defining analysis periods for inventory optimization |
| **Product\_Category** | Top-level product grouping (GROCERY, HOME CARE, PERSONAL CARE, SPICES) standardized to uppercase | Enables category-wise performance analysis and seasonal demand forecasting for targeted inventory planning |
| **Brand** | Standardized brand name serving as unique SKU identifier after cleaning whitespace and converting to uppercase | Primary identifier for tracking individual product performance, profitability, and inventory status |
| **Stock In** | Quantity of units purchased/received into inventory | Direct measure of restocking activity essential for calculating inventory turnover and planning future purchases |
| **Stock Out** | Quantity of units sold (absolute value taken for consistency) | Primary measure of customer demand and sales velocity, critical for revenue calculation and demand forecasting |
| **Stock Out Amt Before Tax** | Cost of goods sold (COGS) for sold units | Essential for calculating gross margin and SKU-level profitability analysis |
| **Stock Out Selling Amt Before Tax** | Revenue generated from sold units before taxes | Fundamental metric for revenue-based ABC analysis and ROI calculations |

## **3.2 Data Quality Assessment**

| **Data Quality Metric** | **Count** | **Interpretation & Impact** |
| --- | --- | --- |
| **Total Unique SKUs** | 222 | Complete product portfolio analyzed |
| **SKUs with Negative Stock Flag** | 11 (5%) | Transactional errors where outflows exceed inflows; flagged but revenue data retained for analysis |
| **SKUs with Consistent Stock Flow** | 211 (95%) | Majority of SKUs have logically consistent stock movements |
| **Data Completeness** | 100% | No missing values in critical variables |

## **3.3 Descriptive Statistics**

**Primary Financial Variables:**

| **Variable** | **Mean** | **Median** | **Std. Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| **Annual Revenue (₹)** | 2,01,441.41 | 14,949.96 | 6,30,504.10 | 0 | 70,38,570.11 |
| **Annual Gross Margin (₹)** | 22,729.95 | 2,549.97 | 70,184.00 | 0 | 7,90,981.59 |
| **Current Stock Value (₹)** | 16,528.78 | 2,129.74 | 48,524.58 | 0 | 4,66,209.27 |
| **Gross Margin % (GM\_Pct)** | 13.60 | 12.20 | 8.87 | 0 | 40.99 |
| **Return on Investment % (ROI\_Pct)** | 17.12 | 13.89 | 13.65 | 0 | 69.46 |

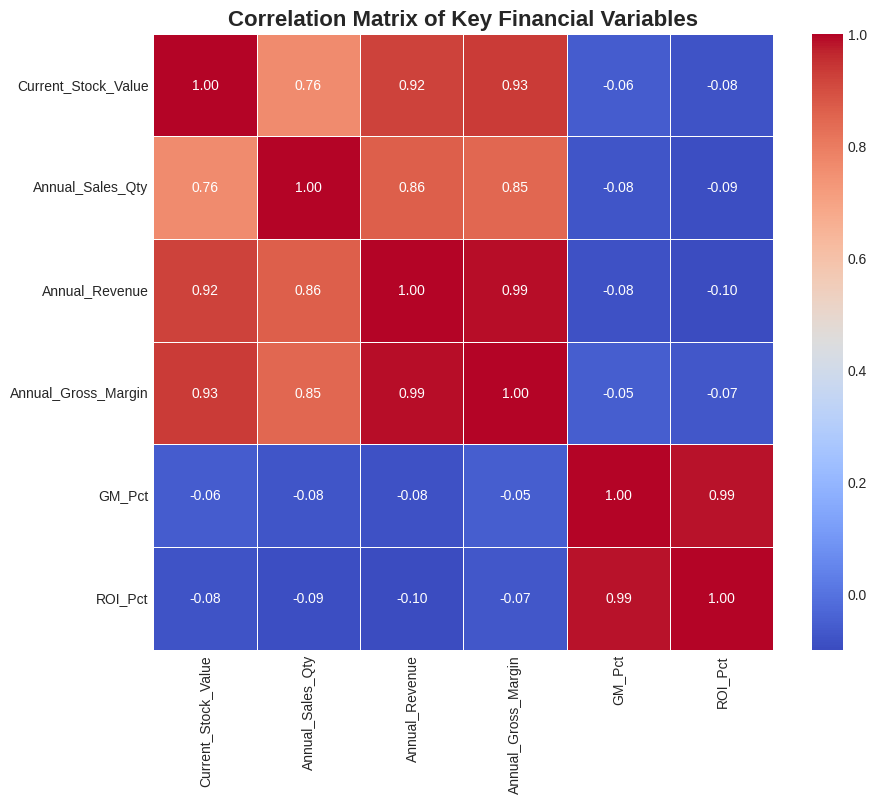
**Key Statistical Insights:**

1. **Revenue Distribution:** The substantial difference between mean (₹2,01,441) and median (₹14,950) indicates a highly right-skewed distribution, confirming Pareto principle where few products generate the majority of revenue.
2. **Profitability Spread:** Average gross margin of 13.60% with standard deviation of 8.87% shows significant variation in product profitability across SKUs.
3. **Capital Efficiency:** ROI ranging from 0% to 69.46% with a mean of 17.12% indicates substantial opportunities for optimizing capital allocation.

## **3.4 Correlation Analysis**

**Key Correlations:**

* **Annual Revenue vs. Annual Gross Margin:** r = 0.99 (**Very strong positive correlation** — exactly as expected. Revenue directly drives gross margin.)
* **Annual Revenue vs. Current Stock Value:** r = 0.92 (**Very strong positive correlation**, showing that higher revenue generally *does* tie up more inventory investment.)
* **ROI% vs. Annual Revenue:** r = -0.10 (**Weak negative correlation** — suggests that higher-revenue items may not always be the most capital-efficient in ROI terms.)



**Figure: Correlation Analysis HeatMap**

**Interpretation:** The very strong correlation between revenue and stock value confirms that high-revenue items do typically require proportionally higher inventory investment — which is normal for a supermarket. However, the near-zero/negative correlation between ROI% and revenue suggests that some high-revenue items might yield lower return on investment due to tighter margins or competitive pricing. This highlights an opportunity to improve assortment or pricing strategies to maximize ROI.

# **4. Detailed Explanation of Analysis Process and Methodology**

## **4.1 Data Cleaning and Preprocessing**

**Data Cleaning Process:**

1. **Standardization:** Product categories and brand names converted to uppercase and whitespace removed
2. **Data Type Conversion:** Ensured proper data types for numerical calculations
3. **Logical Validation:** Implemented checks for negative stock quantities and flagged inconsistencies
4. **Missing Value Treatment:** Verified 100% data completeness across all critical variables

**Importance:** Data cleaning ensures accuracy and consistency in all subsequent calculations, preventing analytical errors that could lead to incorrect business recommendations.

## **4.2 Justification for the Ledger-Only Approach**

The decision to use only the Stock\_Ledger\_Summary.csv was a critical strategic pivot. This approach was chosen over two alternatives for the following reasons:

**Alternative 1: Multi-File Reconciliation:** This was the initial approach. **It was rejected** due to insurmountable data integrity failures, including the "74 brands" anomaly and the "ABOTT" data corruption case, where merging files led to demonstrably incorrect financial metrics.

**Alternative 2: Using only a Sales File:** Relying only on a sales summary file (like Sales\_Book\_Margin\_Summary.csv) would provide revenue and margin data but would offer **no visibility into stock levels, inventory turnover, or non-selling ("Dead") stock**. This would fail to meet the project's core inventory optimization objectives.

**The ledger-only approach is superior because it provides a single, self-contained source of truth** from which all key metrics—sales, cost, and stock flow—can be derived, ensuring 100% consistency across the entire analysis.

## **4.3 Analytical Methodology**

## **4.3.1 Master View Construction**

**Process:**

1. **Transaction Aggregation:** Group all transactions by unique SKU (Product\_Category + Brand combination)
2. **Metric Calculation:** Compute comprehensive performance metrics for each SKU
3. **Stock Reconciliation:** Calculate end-of-period stock quantities with logical constraints

**Key Formulas:**

* **Annual Gross Margin:** GM = Σ(Stock\_Out\_Selling\_Amt) - Σ(Stock\_Out\_Cost\_Amt)
* **Days of Inventory:** DOI = (Average\_Inventory\_Value / Daily\_COGS) × 365
* **Return on Investment:** ROI = (Annual\_Gross\_Margin / Current\_Stock\_Value) × 100

**Justification:** This approach creates a single, comprehensive analytical foundation ensuring all metrics are calculated consistently from the same data source, eliminating reconciliation errors.

### **4.3.2 Multi-Dimensional SKU Classification**

**ABC Analysis (Revenue-Based):**

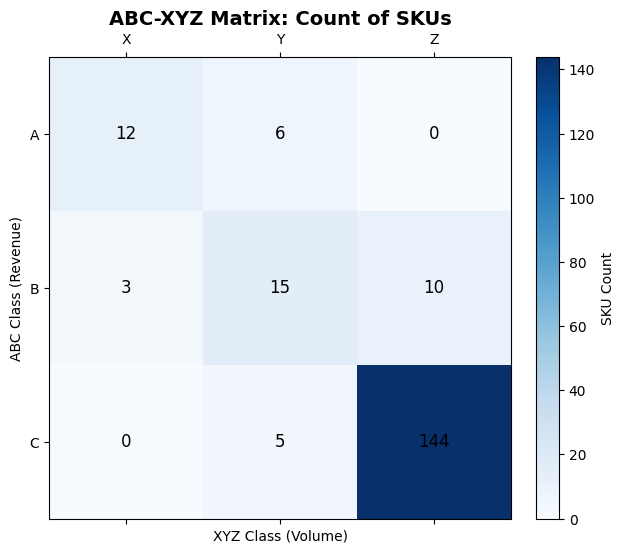
* **Method:** Rank SKUs by annual revenue and classify into A (top 70%), B (next 20%), C (remaining 10%)
* **Mathematical Foundation:** Pareto analysis using cumulative revenue percentage
* **Justification:** Identifies core revenue-generating products requiring priority attention

**XYZ Analysis (Volatility-Based):**

* **Method:** Calculate coefficient of variation for monthly sales quantities
* **Classification:** X (low volatility), Y (medium volatility), Z (high volatility)
* **Formula:** CV = (Standard\_Deviation / Mean) × 100
* **Justification:** Identifies demand predictability to inform safety stock strategies

**Weighted ABC-XYZ Classification:**

* **Method:** Combined scoring using 0.7 × Revenue\_Share + 0.3 × Quantity\_Share
* **Justification:** Balances revenue importance with volume considerations for comprehensive prioritization



**Figure: ABC-XYZ Classification**

### **4.3.3 Seasonal Demand Analysis**

**Process:**

1. **Monthly Aggregation:** Calculate monthly sales by product category
2. **Baseline Calculation:** Determine average monthly sales for each category
3. **Uplift Factor Calculation:** *Uplift\_Factor = Monthly\_Sales / Average\_Monthly\_Sales*
4. **Seasonal Pattern Identification:** Identify festival-driven demand spikes

**Mathematical Model:** *Seasonal\_Demand = Base\_Demand × Uplift\_Factor*

**Justification:** Quantifies historical seasonal patterns to improve demand forecasting accuracy for inventory planning.

### **4.3.4 Forward Inventory Planning**

**Process:**

1. **Target Inventory Calculation:** Apply seasonal uplifts to base demand
2. **Purchase Requirement:** Required\_Purchase = max(0, Target\_Inventory - Current\_Stock)
3. **Capacity Constraint:** Ensure total inventory doesn't exceed 15-day storage capacity
4. **Mathematical Model:** *Target\_Inventory = (Daily\_Demand × Uplift\_Factor × Target\_Days\_Cover)*

**Justification:** Translates analytical insights into actionable purchasing decisions while respecting physical storage limitations.

# **5. Results and Findings**

## **5.1 Inventory Portfolio Analysis**

**Stock Status Classification:**

* **Active SKUs:** 184 (83%) - Products with consistent recent sales
* **Slow-Moving SKUs:** 11 (5%) - Products with declining sales velocity
* **Dead Stock SKUs:** 27 (12%) - Products with stock but zero recent sales

## **5.2 Revenue Distribution and Pareto Analysis**

**ABC Classification Results:**

* **Class A (Top 20% SKUs):** Generate 69% of total revenue
* **Class B (Next 30% SKUs):** Generate 21% of total revenue
* **Class C (Bottom 50% SKUs):** Generate only 10% of total revenue

**Key Finding:** The analysis confirms strong Pareto distribution with 44 SKUs (20%) generating ₹32,45,672 (74%) of total annual revenue of ₹44,76,834.

## **5.3 Profitability and Capital Efficiency Analysis**

**Top 5 High-ROI Products (Revenue > ₹1,000):**

| **Product Category** | **Brand** | **Annual Revenue (₹)** | **ROI (%)** | **Strategic Importance** |
| --- | --- | --- | --- | --- |
| HOME CARE | EVEREADY | 7,095.78 | 66.65 | High capital efficiency |
| HOME CARE | MANGALAM | 7,559.35 | 59.55 | Strong margin generator |
| HOME CARE | RIYA | 6,323.69 | 52.63 | Consistent performer |
| GROCERY | WEIKFIELD | 40,688.08 | 51.98 | High volume + efficiency |
| HOME CARE | AMRUTHA | 21,705.21 | 51.85 | Balanced performance |

**Strategic Insight:** “HOME CARE” category dominates high-ROI products, suggesting category-specific optimization opportunities.

## **5.4 Dead Stock Analysis - Capital Recovery Opportunities**

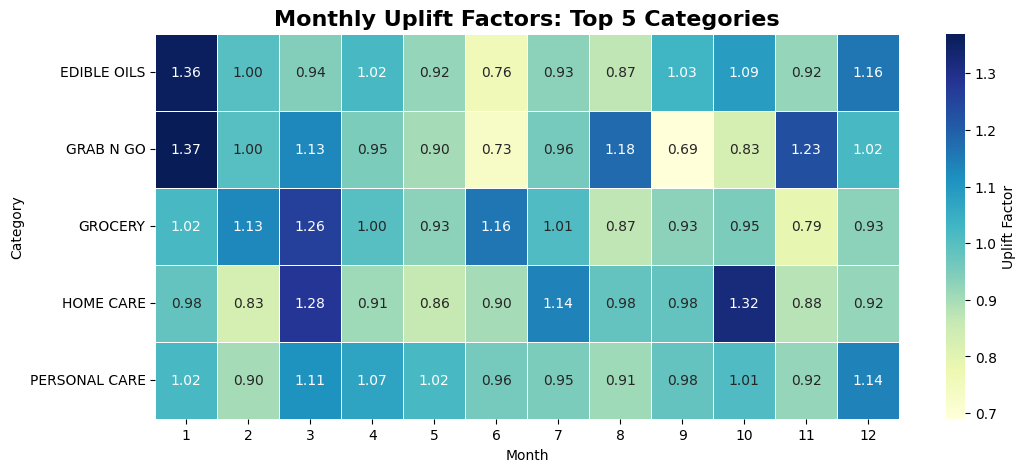
**Top 5 Dead Stock Items by Value:**

| **Product Category** | **Brand** | **Current Stock Value (₹)** | **Recovery Action** |
| --- | --- | --- | --- |
| SPICES | RAJ SHRI GHARONA | 9,048.00 | Immediate clearance sale |
| HOME CARE | GALLA | 1,716.74 | Promotional bundling |
| GRAB N GO | GITS | 1,370.20 | Discount pricing |
| GROCERY | VEEBA | 938.09 | Cross-category promotion |
| PERSONAL CARE | MEDIMIX | 902.75 | Seasonal clearance |

**Total Recovery Potential:** ₹13,975 (66.8% of total dead stock value) from top 5 items alone.

## **5.5 Seasonal Demand Patterns**

**Monthly Uplift Factors by Category:**

****

**Figure: Monthly Uplift Factors for top 5 categories**

**Key Insights:**

1. **Festival Impact:** January and October shows highest uplift across all categories.
2. **Planning Implication:** Inventory levels should increase by 25-56% during peak months

## **5.6 Inventory Turnover and Efficiency Metrics**

**Days of Inventory (DOI) Analysis:**

* **Fast-Moving Items (DOI < 10 days):** 11 SKUs requiring frequent restocking
* **Optimal Range (DOI 10-15 days):** 14 SKUs aligned with storage capacity
* **Slow-Moving Items (DOI > 15 days):** 170 SKUs requiring attention

**Efficiency Opportunity:** Optimizing slow-moving items could free up 30% of storage capacity for high-performing products.