Sparking Growth: A Data-Driven Strategy For Optimizing Sales, Inventory, and Logistics for Shri Hari Electric And Electronics

Final Submission for the Project on Business Data Management

Submitted by

Name: Sachin Maurya

Roll number: 24f2000305

IITM Online BS Degree Program, Indian

Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1.0 Executive Summary

This report presents a comprehensive strategic analysis for Shri Hari Electric and Electronics, a local retail firm facing significant operational challenges in inventory inefficiency, untargeted customer engagement, and sub-optimal logistics. Utilizing a primary dataset of 1000 transactions spanning from July 2024 to July 2025, this project transforms raw sales data into an actionable blueprint for sustainable growth.

The analysis employed a three-pronged methodology. First, an ABC Analysis was conducted to stratify the firm's product inventory. Second, an RFM (Recency, Frequency, Monetary) model was developed to segment the customer base. Third, a Geographical Profit Analysis was performed to optimize logistics and resource allocation.

The results of the analysis reveal critical operational insights. The ABC analysis demonstrates a severe misalignment between inventory and revenue; 39% of unique product SKUs (Category A) generate 80% of total revenue, while a staggering 30% of SKUs (Category C) contribute only 5%. The RFM model identified a highly diverse 139-customer base, containing a vital 'Best Customers' segment (11.5% of customers) and a large 'At-Risk' segment (32.4% of customers) requiring immediate intervention. Finally, the geographical analysis confirmed that profit is heavily concentrated in a few key postal codes, with the top 10 (out of 100) pincodes accounting for over 18% of total profit.

Based on these findings, this report provides a clear roadmap for action. Key recommendations include the implementation of a 3-tier inventory policy with distinct strategies for Category A, B, and C products; the launch of targeted, segment-specific marketing campaigns to improve customer retention and stabilize off-season revenue; and the adoption of a profit-focused logistics model to optimize delivery routes and reduce operational costs.

2.0 Detailed Explanation of Analysis Process and Method

This section provides the academic and business context for the analysis, establishing the credibility of the data and the appropriateness of the chosen methodologies.

2.1 Project Background and Objectives

Shri Hari Electric and Electronics is a B2C retail store established in 2022, serving the local community of Mariyahu, Jaunpur, with a diverse portfolio of electronics, furniture, and home wiring supplies. Despite its strategic location, the business faces three critical hurdles to sustainable growth, as identified in the initial project proposal:

Problem 1: Inefficient Inventory Management: A lack of systematic stock allocation results in capital being locked in slow-moving items and frequent stockouts of high-demand products.

Problem 2: Significant Revenue Volatility: Sales are heavily concentrated in festive periods, leading to unpredictable cash flow during off-season months, largely due to a lack of targeted customer engagement.

Problem 3: Sub-optimal Transportation Logistics: A reactive delivery process leads to inefficient routes, increasing fuel consumption and operational costs.

The overarching objective of this project is to leverage the firm's internal sales data to analyze these challenges and develop data-driven strategies that transform them into opportunities for growth and efficiency.

2.2 Metadata and Data Preparation

This section details the dataset used for the analysis and the preparation steps taken, addressing omissions from previous reports regarding data types and column descriptions.

Data Source :
The analysis is based on a master dataset provided by the business owner, containing 1000 transaction records from July 2024 to July 2025. This primary data provides a complete, one-year snapshot of the firm's sales, customer, and product operations.

Column Name	Data Type	Description	Business Purpose & Analytical Role
Transaction_ID	Numeric (Integer)	A unique identifier for each sale transaction.	Primary key for sales; used for counting Frequency in RFM.
Customer_ID	Numeric (Integer)	A unique identifier for each customer.	Links transactions to customers; essential for RFM analysis.
Transaction_Date	Datetime	The full timestamp of when the transaction occurred.	Used to calculate Recency in RFM and analyze sales trends.
Product_SKU	Alphanumeric (String)	The unique Stock Keeping Unit (SKU) for each product.	Primary key for products; essential for ABC analysis.
Product_Category	Categorical (String)	The general category of the product (e.g., 'TV', 'Fan').	Used for high-level sales visualization (Figure 2).
Quantity	Numeric (Integer)	The number of units of the product sold in the transaction.	Used to calculate Total Amount, Profit, and sales volume.
Unit_Price	Numeric (Currency)	The selling price of one unit of the product.	Used to calculate Total Amount and Profit.
Unit_Cost	Numeric (Currency)	The cost of one unit of the product to the business.	Essential for calculating Profit.
Total_Amount	Numeric (Currency)	The total revenue for the line item (Quantity *	Key metric for ABC and Monetary (RFM) analysis.
Customer_Pincode	Numeric (Integer)	The postal code for the customer's delivery location.	Essential for Geographical Profit Analysis.
Supplier_Lead_Time_ Days	Numeric (Integer)	The average time in days for a product to arrive from the	Critical data for developing inventory reorder point

Table 1 : Master Datasheet Metadata : This table provides a comprehensive dictionary for the 11 variables used in the analysis.

Data Preparation Prior to analysis, the dataset was processed to ensure accuracy and to derive necessary metrics:

1. **Data Cleaning:** The 1,000-row dataset was verified for completeness and integrity. No missing or null values were found in critical analytical columns such as Total_Amount, Customer_ID, Transaction_Date, Unit_Price, or Unit_Cost, indicating a high-quality source file.

2. **Feature Engineering:** A new and critical column, Profit, was engineered for each of the 1,000 transactions. This was calculated using the formula: Profit = (Unit_Price - Unit_Cost) × Quantity. This allows the analysis to move beyond simple revenue tracking to a more robust profit-based strategic assessment.

3. **Snapshot Date Definition:** For the RFM analysis, a "snapshot date" was defined as 2025-08-01. This date is one day after the most recent transaction (2025-07-31) found in the dataset, ensuring that all customers have a valid Recency value.

2.3 Analytical Methodologies

The core of this project rests on a three-part analytical framework, with each method selected to address a specific business problem identified in Section 2.1.

2.3.1 ABC Analysis (Inventory Optimization)

Justification: This method was selected to directly address the "Inefficient Inventory Management" problem. The analysis is rooted in the Pareto Principle, which posits that approximately 80% of effects (revenue) come from 20% of causes (products). This method allows the business to stratify its inventory and focus its limited capital, warehouse space, and management attention on the products that generate the greatest return, rather than treating all items with equal importance.

- **Process:**
  a. The Total_Amount (revenue) was aggregated for every unique Product_SKU across all 1,000 transactions.
  b. All unique SKUs were then ranked in descending order based on their total revenue contribution.
  c. A cumulative revenue percentage was calculated for the ranked list.
  d. Products were classified into three distinct categories:
     - **Category A :** The "vital few" products that collectively account for the top 80% of total revenue.
     - **Category B :** The "intermediate" products that account for the next 15% of total revenue (i.e., from 80.1% to 95%).
     - **Category C :** The "trivial many" products that constitute the final 5% of total revenue (i.e., from 95.1% to 100%).

2.3.2 RFM Analysis (Customer Segmentation)

Justification: This method directly addresses the "Significant Revenue Volatility" problem. By segmenting customers based on their actual transactional behavior - Recency (how recently they bought), Frequency (how often they buy), and Monetary value (how much they spend), the business can move from an inefficient "one-size-fits-all" marketing approach to a highly targeted strategy. This allows for precise campaigns to stimulate demand during off-peak seasons, such as re-engaging at-risk customers.

- **Process:**
  a. **Recency (R) :** Calculated for all 139 unique customers in the dataset. This was computed as the number of days between the snapshot date (2025-08-01) and the customer's most recent Transaction_Date. A lower number of days results in a better score.
  b. **Frequency (F) :** Calculated as the total count of unique Transaction_IDs associated with each Customer_ID.
  c. **Monetary (M) :** Calculated as the total sum of Total_Amount from all transactions for each Customer_ID.
  d. **Scoring:** Each customer was assigned a score from 1 (worst) to 4 (best) for each of the R, F, and M metrics. These scores were based on quartiles, where the top 25% of customers in a metric received a score of 4, the next 25% a 3, and so on.

e.**Segmentation:** The individual R, F, and M scores were concatenated to create a three-digit RFM score (e.g., '444', '123'). These scores were then mapped to descriptive segments, such as 'Best Customers' or 'At-Risk Customers', to create actionable cohorts.

2.3.3 Geographical Profit Analysis (Logistics Optimization)

Justification: This analysis was chosen to address the "Sub-optimal Transportation Logistics" problem. Rather than simply mapping sales, this method maps *profitability* by Customer_Pincode. This provides a more powerful strategic insight, answering the critical question, "Where is our business *actually* making the most money?" This data-driven foundation is essential for optimizing delivery routes, allocating local marketing budgets, and improving overall logistical efficiency.

- **Process:**
  a. The entire 1,000-transaction dataset was grouped by the Customer_Pincode variable.
  b. For each of the 100 unique pincodes, four key performance indicators were aggregated:
     - Total Profit: The sum of the engineered Profit column.
     - Total Sales: The sum of the Total_Amount column.
     - Transaction Count: The count of all Transaction_IDs.
     - Unique Customer Count: The count of distinct Customer_IDs.
  c. The resulting summary table was then ranked by Total_Profit to identify the most and least lucrative geographical areas for the business.

This three-part analytical framework provides a cohesive, end-to-end strategic view. The ABC Analysis addresses *capital efficiency* by identifying which products to stock. The RFM Analysis addresses *customer efficiency* by identifying which customers to target. Finally, the Geographical Profit Analysis addresses *operational efficiency* by identifying where to focus logistics and marketing resources. This integrated approach ensures that decisions in one area (like inventory) support the goals of another (like customer retention and profit protection).

3.0 Results and Findings

This section presents the direct quantitative and visual outputs of the analytical methodologies detailed in Section 2. All results are derived from the complete 1000-transaction dataset.

3.1 Inventory Optimization: ABC Analysis Results

The ABC analysis of 502 unique product SKUs reveals a critical misalignment between the volume of inventory SKUs and their actual contribution to revenue. This confirms the hypothesis that the business's capital is disproportionately locked in a large number of low-value items.
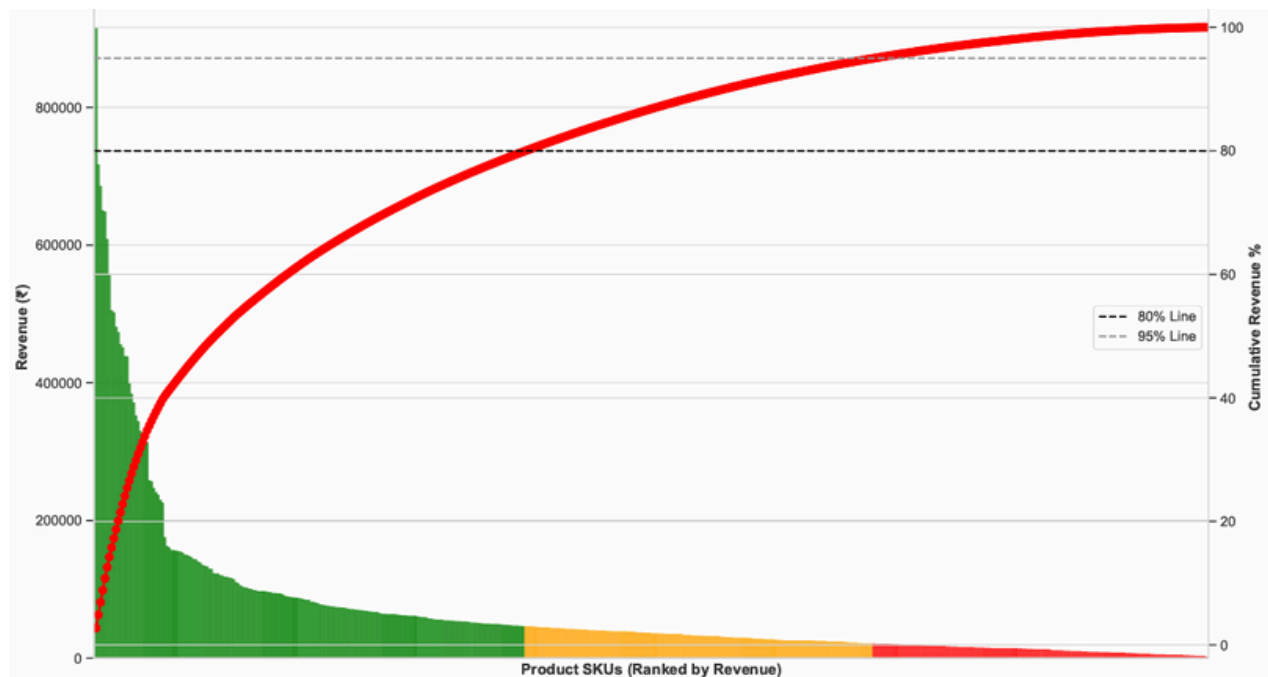

Figure 1 : ABC Analysis - Revenue vs. SKU Contribution (Pareto Chart)

Finding 1: The Pareto chart (Figure 1) visually confirms that a small subset of products, the "vital few," drives the vast majority of revenue. The cumulative revenue line rises sharply, indicating that the first 39% of SKUs (Category A) achieve 80% of the total revenue.

Finding 2: This imbalance is quantified in Table 2. Category C, the "trivial many," represents 30% of all unique products the business stocks, but this entire category is responsible for only 5% of total revenue. Conversely, just 194 unique SKUs (Category A) generate 80% of the income.

Category	Revenue Contribution	% of Total Revenue	No. of Unique % of SKUs	Total SKUs
Category A	Top 80%	79.9%	194	38.6%
Category B	Next 15%	15.0%	157	31.3%
Category C	Bottom 5%	5.1%	151	30.1%
Total		100.0%	502	100.0%

Table 2 : ABC Category Summary (N = 502 Total SKUs)

Finding 3: At a higher level, the revenue is concentrated in high-value categories. The Treemap in Figure 2 illustrates that 'Cooler', 'Refrigerator', 'TV' are the dominant product categories, which aligns with the SKU-level finding that high-ticket items form the core of the business's revenue.
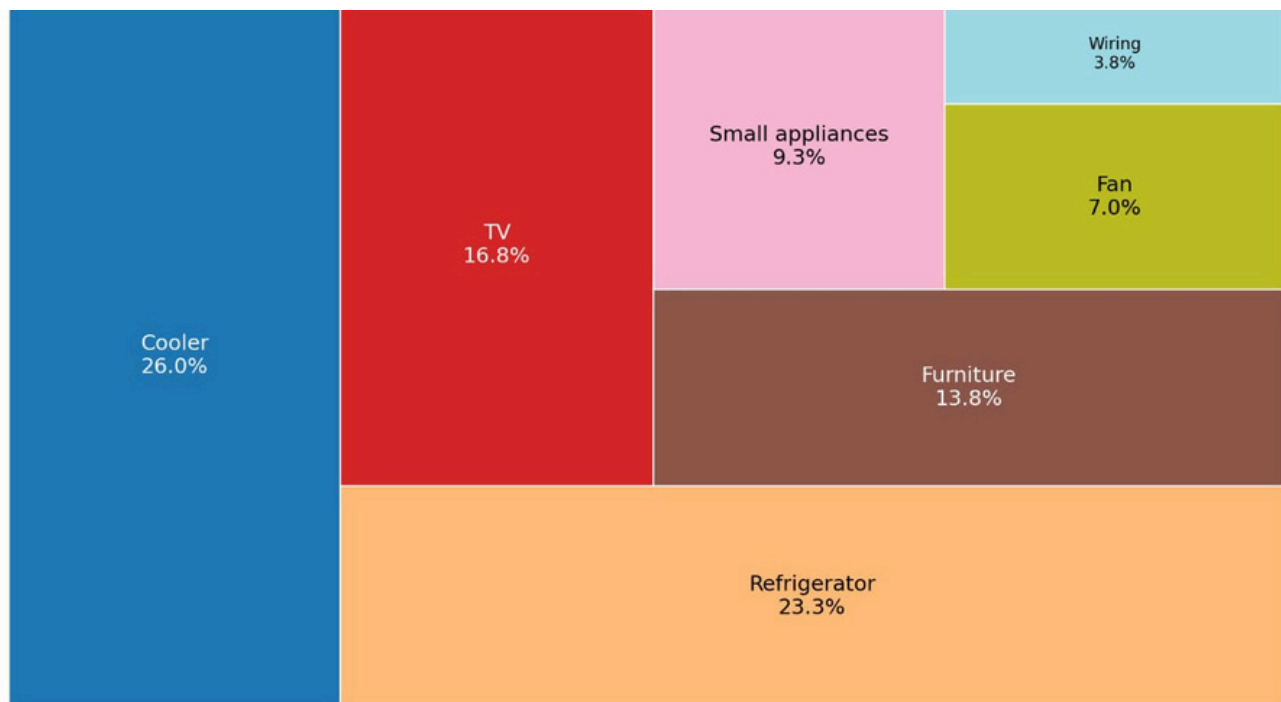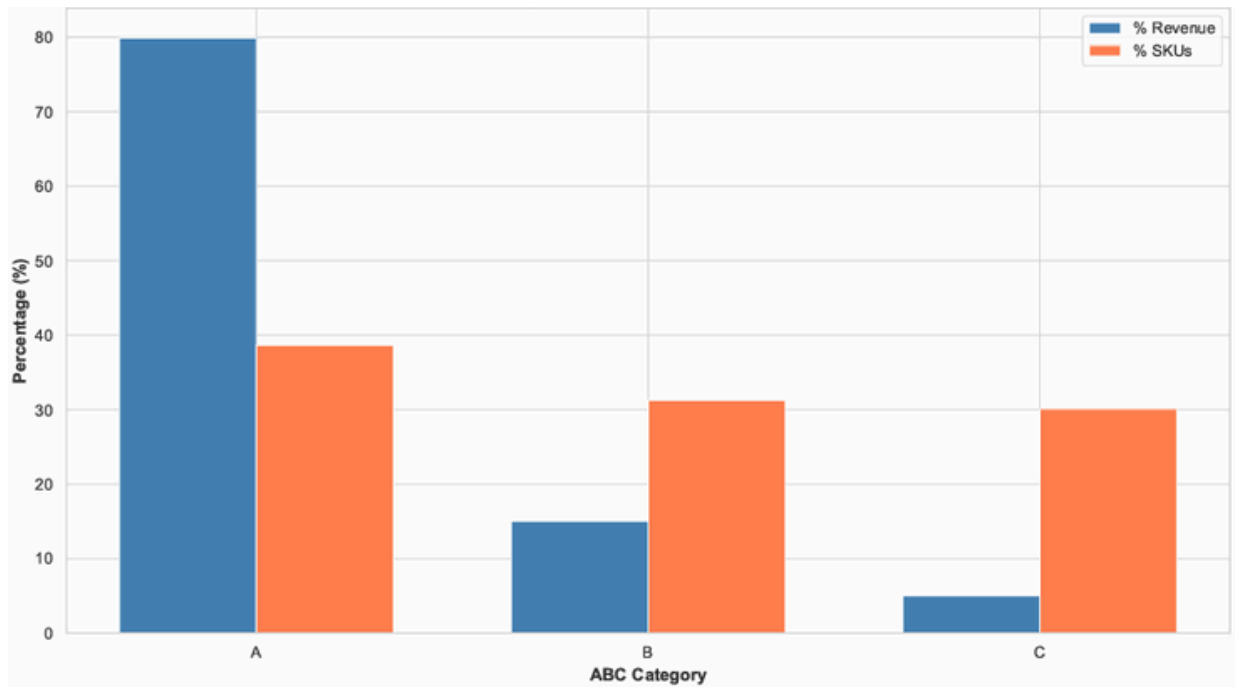

Figure 2 : Revenue by Product Category

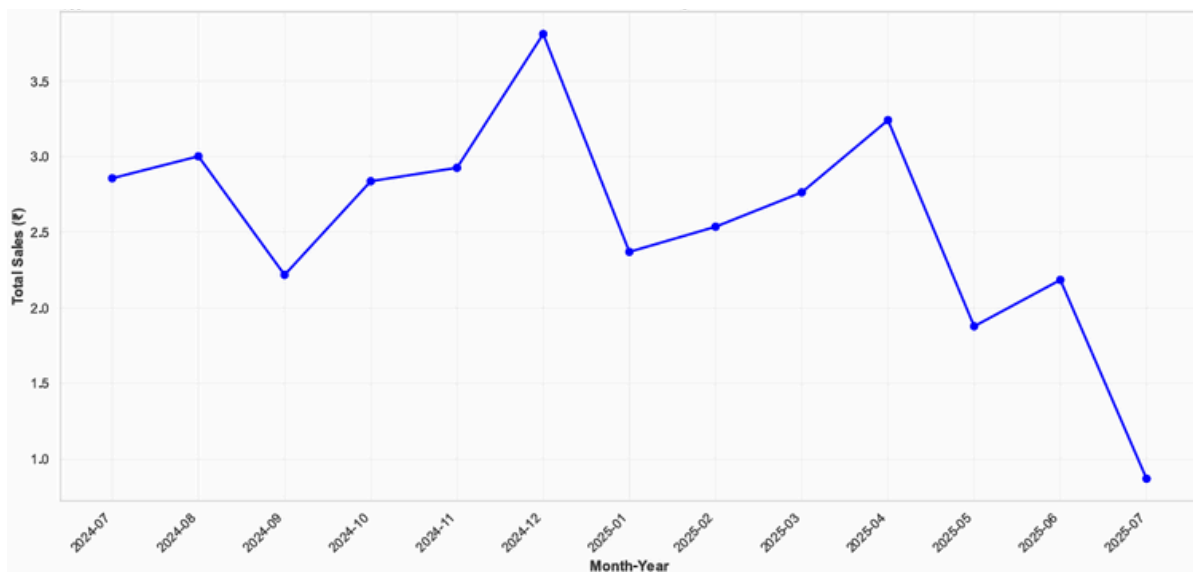**Figure 3 : ABC Analysis: Revenue vs SKU Distribution**

Figure 4 : Total Sales Amount by Month

3.2 Customer Engagement: RFM Segmentation Results

The RFM analysis segmented the entire 139-customer base into distinct, behavior-based cohorts. This result confirms that the customer base is not a monolith but a diverse portfolio of assets, each requiring a different management strategy. This section directly addresses the mid-term feedback by providing the size and percentage of each customer segment.

Finding 1 : Table 3 provides the definitive breakdown of the customer base. The analysis reveals two critical findings: a small, high-value cohort of 'Best Customers' (16 customers, 11.5%) and a dangerously large 'At-Risk' and 'Lost Customers' cohort, which combined makes up 52 customers (37.4% of the entire customer base). Additionally, an 'Other' segment of 10 customers (7.2%) represents customers with mixed RFM patterns that do not fit cleanly into the primary strategic categories.

RFM Segment	Segment Definition (R-F-M Scores)	Customer Count (Size)	% of Total Customer Base
Best Customers	4-4-4	16	11.51%
Loyal Customers	(3,4)-(3,4)-(3,4) (excl. 444)	38	27.34%
Big Spenders	(any)-(any)-4 (excl. 444)	0	0.00%
New Customers	4-1-(any)	0	0.00%
Promising	(3,4)-(2,3)-(2,3)	6	4.32%
Need Attention	(2,3)-(2,3)-(2,3)	17	12.23%
At-Risk Customers	(1,2)-(any)-(any) (excl. 111)	45	32.37%
Lost Customers	1-1-1	7	5.04%
Other	Mixed patterns (all other combination)	10	7.19%
Total		**139**	**100.00%**

Table 3 : RFM Segment Distribution and Size (N = 139 Customers)

Finding 2 : The pie chart in Figure 5 visualizes the segment sizes from Table 3. It provides an immediate sense of scale, showing that the 'At-Risk' segment alone (45 customers, 32.4%) is the largest single segment and is larger than the 'Best Customers' (16 customers, 11.5%) and 'Loyal Customers' (38 customers, 27.3%) segments combined. This highlights a significant "leaky bucket" problem, where customer churn is a major threat to revenue stability.
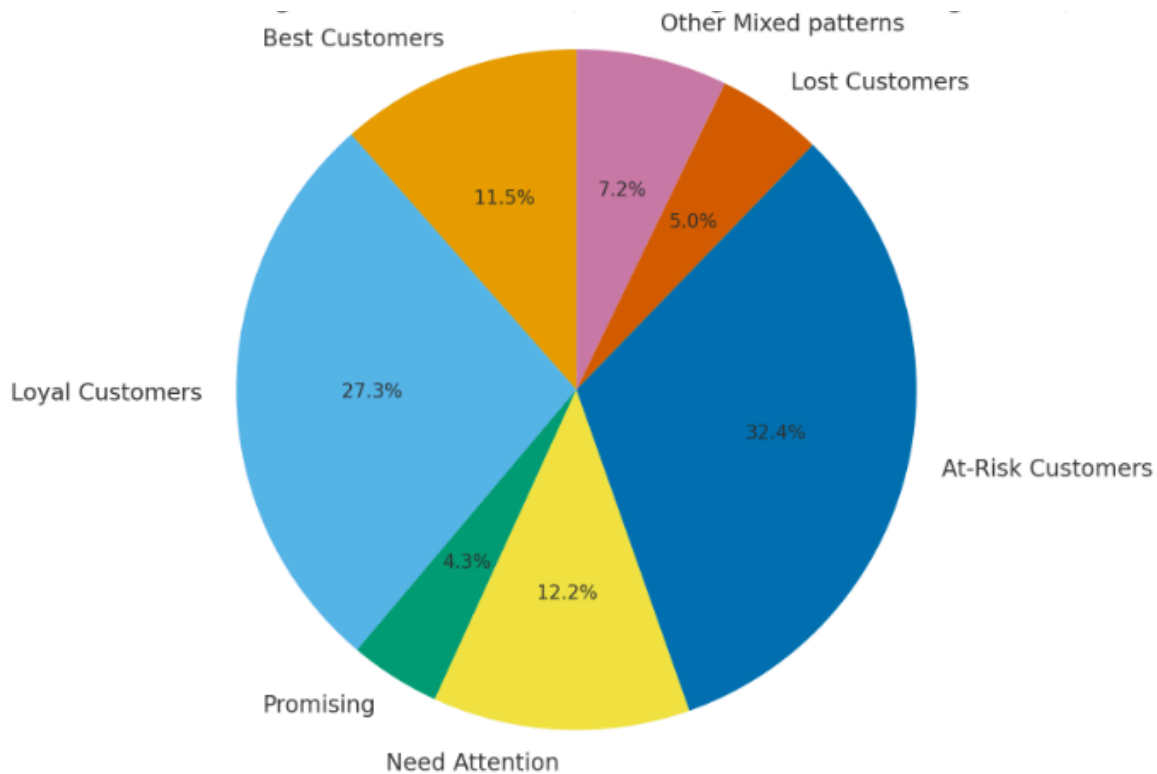

Figure 5 : Customer Segment Distribution (% of Total Customers)

Finding 3 : Figure 6 contrasts the customer count with the revenue they generate. It shows that the 'Best Customers' (16 customers) and 'Loyal Customers' (38 customers) together form the core of the business (38.8% of the customer base), contributing a disproportionately massive share of the total revenue. This illustrates their outsized importance and the high financial risk associated with losing them. The 'Other' segment (10 customers, 7.2%), while relatively small, represents customers with inconsistent purchasing patterns who may benefit from re-engagement campaigns to move them into more valuable segments.
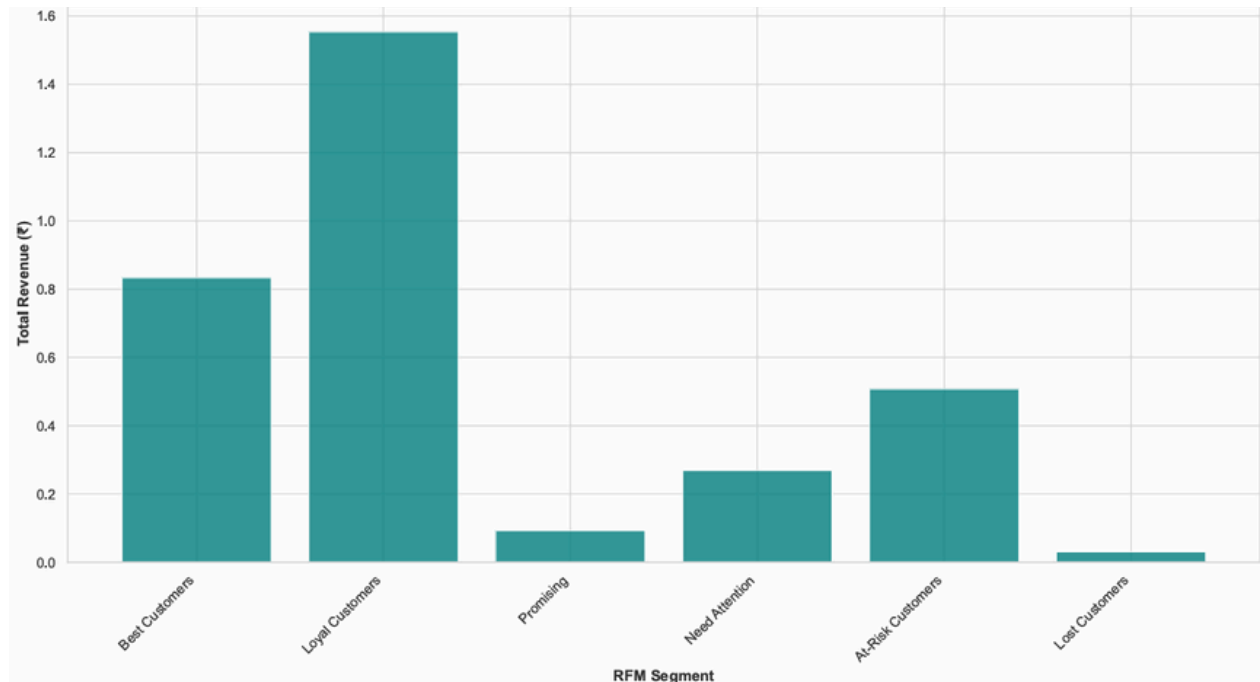
Figure 6 : Revenue Contribution by RFM Segment

3.3 Logistical Optimization: Geographical Analysis Results

The geographical analysis reveals that the business's success and profitability are highly concentrated in a few key pin codes. This finding suggests that a uniform, "one- size-fits-all" logistics and delivery policy is highly inefficient, with profitable "hotspots" likely subsidizing costly deliveries to "cold spots."

Finding 1: Table 4 lists the top 10 most profitable pin codes for the business, ranked by the engineered Total_Profit metric. These10pincodes (out of 100 total) account for over 18% of the firm's total profit. The data also provides context, showing that some areas are valuable due to a high volume of transactions and customers, while others may be valuable due to a few, high-profit sales.

Finding 2: Figure 7 provides a clear visual representation of this profit concentration. The dominance of the top pin codes is immediately apparent, making it a powerful tool for communicating where strategic focus should be applied.

Rank	Customer Pincode	Total Profit (₹)	Total Sales (₹)	Transaction Count	Unique Customer Count
1	222141	126,270	712,564	34	21
2	222171	112,201	644,796	15	14
3	222139	105,255	648,001	13	13
4	222161	103,131	600,453	44	25
5	222160	101,381	620,743	16	15
6	222176	96,725	548,641	9	9
7	222155	92,981	482,275	13	12
8	222162	89,710	508,896	26	18
9	222107	89,191	549,361	10	9
10	222164	88,839	594,160	10	10

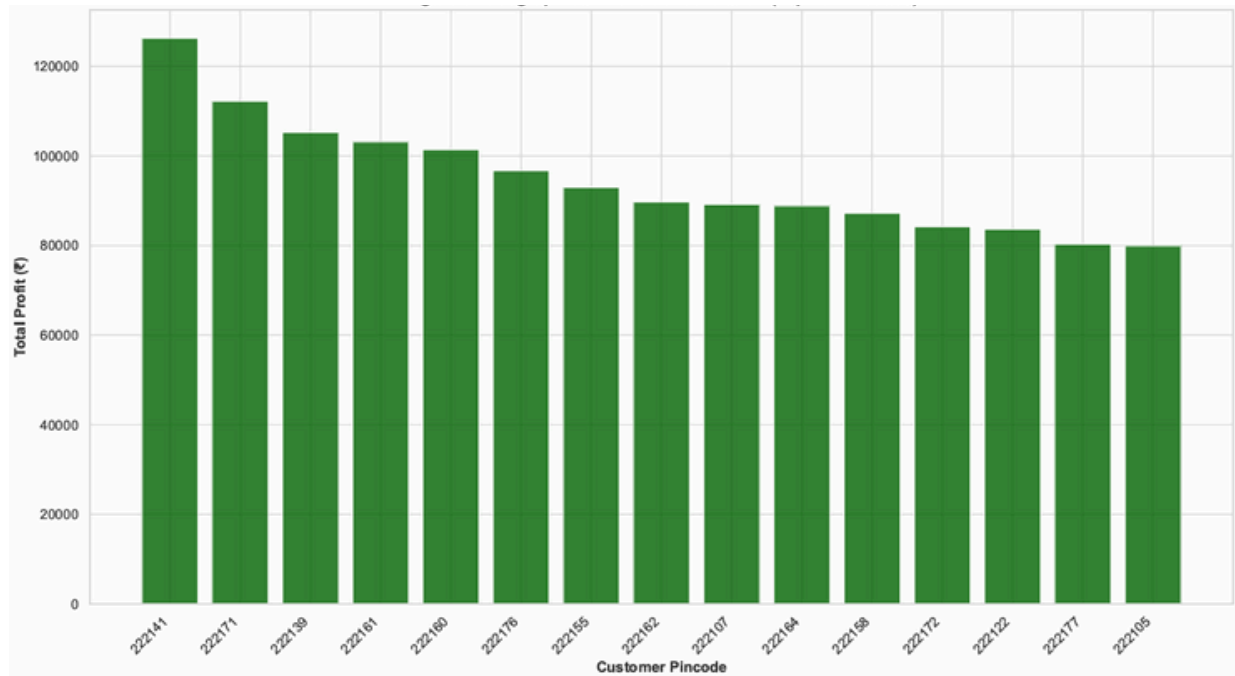**Table 4 : Top 10 Pin codes by Total Profit**

Figure 7 : Geographical Profit Concentration (Top 15 Pin codes)

4.0 Interpretation of Results and Recommendation

This section provides the strategic interpretation of the findings from Section 3.0, translating the "what" (the data) into the "so what" (the business implications). It delivers actionable recommendations for each of the three core business problems.

4.1 Recommendation 1: Implement a 3-Tier Inventory Management Strategy (from ABC Analysis)

Interpretation : The ABC analysis results in Table 2 are stark: 39% of product SKUs generate 80% of revenue. This is a classic case of capital inefficiency, where valuable warehouse space, management time, and funds are tied up in products that do not contribute to the bottom line (Category C). A stockout in Category A is a significant lost-sale event, while overstocking Category C is a silent, persistent financial drain.

Recommendation (Category A: The "Crown Jewels")
Action : These 194 SKUs are the business's lifeblood. A perpetual inventory system (daily or weekly counts) must be implemented for this category. Maintain high safety stock levels to never have a stockout. Prioritize and strengthen supplier relationships for these items, using the Supplier_Lead_Time_Days data to build a predictive reordering model.

Justification : The cost of a stockout (a lost sale of a high-value refrigerator or TV) is far higher than the cost of holding a few extra units.

Recommendation (Category B: The "Engine Room")
Action : This section directly addresses the feedback regarding the omission of Category B. This category of 157 SKUs (31% of all products) is the firm's "engine room."

Interpretation : These items (e.g., Fans, Wiring, Small Appliances) are not high-revenue stars, but they are the crucial stabilizers that customers expect a full-service electronics store to carry. They are too large a group to ignore (like C) and not valuable enough for the high-touch "A" strategy.

Strategic Recommendation : These 157 SKUs are the key to increasing the average order value and customer frequency. They should be strategically bundled with Category A items (e.g., "Purchase a Category A Refrigerator, receive a Category B Small Appliance at 20% off"). They are also the perfect products for re-engagement campaigns (see RFM Recommendation 4.2), as they are lower-cost, lower-risk purchases to entice an at-risk customer back into the store. A periodic review system (e.g., checking stock levels bi-weekly) is sufficient.

Recommendation (Category C: The "Long Tail")
Action : Aggressively rationalize this category. The bottom 50% of these 151 SKUs should be delisted immediately upon selling through current stock.

Justification : This category is locking up capital and physical space. The funds freed from liquidating this overstock can be reinvested into ensuring Category A items are always in stock, directly increasing revenue and profitability.

4.2 Recommendation 2: Launch Targeted, Segment-Specific Marketing Campaigns (from RFM Analysis)

Interpretation : The RFM results (Table 3) prove that a "one-size-fits-all" marketing approach is deeply inefficient and is the primary driver of the "Revenue Volatility" problem. The business is demonstrably failing to retain customers, with 37.4% of its base (52 customers) being 'At-Risk' or 'Lost'. This represents a critical and actionable revenue leak that must be addressed immediately.

Recommendation (High-Value Segments: 'Best Customers', 'Loyal Customers')
Action: Nurture and reward. Do not discount.

Strategy : This group (54 customers, or 38.8% of the base) is the firm's most valuable asset. Create a "Shri Hari VIP Club" for them. Offer non-monetary perks: first access to new product arrivals, free delivery (their pin codes are known from Table 4), or a small, unexpected "thank you" gift. Their value is in retention and high-quality word of-mouth. The 'Best Customers' (16 customers, 11.5%) are the absolute top priority losing even a single customer from this segment would be a significant revenue loss.

Recommendation (High-Potential Segments: 'Promising', 'Need Attention')

Action : Convert and grow before they slip away.

Strategy : The 'Promising' segment (6 customers, 4.3%) and 'Need Attention' segment (17 customers, 12.2%) represent untapped potential. These customers have shown some engagement but are at risk of slipping into the 'At-Risk' category. They should be targeted with personalized product recommendations based on their purchase history and limited-time offers (e.g., "Complete your setup with...") to increase their engagement frequency and monetary value. This is preventive medicine it is far cheaper to retain these customers now than to try to win them back later.

Recommendation (Low-Value/At-Risk Segments: 'At-Risk', 'Lost Customers')

Action : Re-activate to stabilize off-season revenue.

Strategy: This group of 52 customers (37.4% of the base) is the key to solving the revenue volatility problem. The 'At-Risk' segment (45 customers, 32.4%) should receive a high-incentive "We Miss You!" discount campaign. This should be deployed during the off-season with offers for Category B products (e.g., Fan, Wiring, Small Appliances), which are lower-risk purchases that get them back into the ecosystem without eroding margins on Category A products. The 'Lost Customers' (7 customers, 5.0%) may require deeper discounts or a "Win-Back Guarantee" (e.g., "Try us again with 30% off your next purchase"). While these customers have the lowest recovery probability, even reclaiming a small percentage would generate incremental revenue with minimal additional cost.

Recommendation ('Other' Segment: Mixed-Pattern Customers)

Action : Analyze and redirect into defined segments.

Strategy : The 'Other' segment (10 customers, 7.2%) consists of customers with inconsistent purchasing patterns that do not fit cleanly into the primary strategic segments. These customers should be individually analyzed to understand their behavior patterns. Some may be seasonal buyers who only purchase during specific times of the year, while others may be bulk buyers with irregular patterns. Based on this analysis, develop micro-campaigns to move them into more predictable and valuable segments. For example, if analysis reveals they are festival-season buyers, send them pre-festival reminders and early-bird offers to increase their purchase frequency and move them toward the 'Promising' or 'Loyal' segments.

4.3 Recommendation 3: Optimize Logistics and Marketing Based on Profit Hotspots (from Geographical Analysis)

Interpretation : The geographical analysis (Table 4) clearly indicates that the business's profit is geographically concentrated. The top 10 pin codes are the profit engine. This means the firm is likely wasting significant resources (fuel, time) on inefficient, one-off deliveries to low-profit "cold spot" areas.

Recommendation (High-Profit Pin codes: e.g., 222141, 222171, 222139)
Action : Consolidate and dominate.

Strategy (Logistics) : Consolidate all deliveries for these top pincodes into dedicated "delivery days" (e.g., "All deliveries to 222141 and 222139 occur on Tuesdays and Fridays"). This maximizes route density and immediately slashes the cost-per-delivery.

Strategy (Marketing) : Focus 80% of the local marketing budget (flyers, local advertisements) only in these 10 pin codes. The customer density and profitability are proven; this is the most efficient use of marketing capital.

Recommendation (Low-Profit Pin codes)
Action : Batch or charge.

Strategy : For all other 90 pincodes, implement a new delivery policy. Either (a) implement a delivery surcharge to ensure profitability, or (b) offer free delivery with a longer window (e.g., "delivery within 7-10 days"). This allows the business to batch multiple low-profit orders into a single, efficient trip, turning them from a net loss to a net gain.

5.0 Conclusion

This capstone project successfully translated a raw dataset of 1000 transactions into a comprehensive strategic blueprint for Shri Hari Electric and Electronics. The analysis directly addressed the three core operational challenges identified in the initial proposal: debilitating revenue volatility, inefficient inventory management, and sub-optimal logistics. By systematically applying a tripartite analytical framework, consisting of ABC Analysis, RFM Customer Segmentation, and Geographical Profit Analysis, this report moves beyond surface-level metrics to uncover the underlying drivers of profitability and operational friction.

The findings from this analysis were definitive and exposed significant operational misalignments. The ABC analysis revealed a stark 39/80 Pareto imbalance, with 39% of product SKUs generating 80% of revenue, while 30% of SKUs (Category C) contributed only 5%. The RFM model quantified a critical customer retention problem, identifying a vital 'Best Customers' segment (11.5% of the base) but also a dangerously large 'At-Risk' and 'Lost' cohort (37% combined). Finally, the geographical analysis confirmed that profitability is not evenly distributed, with over 18% of total profit originating from just 10 of the 100 service pincodes.

These findings form the basis of an integrated set of actionable recommendations. The 3-tier inventory strategy (including specific, targeted actions for Category B products) directly realigns capital with high-value Category A items. The segment-specific marketing campaigns, derived from the RFM analysis, provide a precise toolkit for nurturing high-value customers and re-engaging the 'At-Risk' segment to stabilize off-season revenue. The profit-focused logistics model, in turn, uses the geographical findings to reduce operational costs by concentrating resources on proven profit hotspots.

This project provides the necessary blueprint for Shri Hari Electric and Electronics to transition from an intuition-based operation to a data-driven, profitable, and sustainable enterprise. The recommendations are not isolated fixes but a unified strategy: freeing capital from inefficient inventory (ABC) allows for investment in targeted marketing (RFM), which is made more profitable by reducing delivery costs (Geographical).

While this analysis provides a robust one-year baseline, future work should focus on implementing these systems and continuously monitoring these new metrics (e.g., Category C SKU count, customer churn rate, and cost-per-delivery) to ensure long-term, adaptive growth in a competitive market.

--------------------X--------------------