

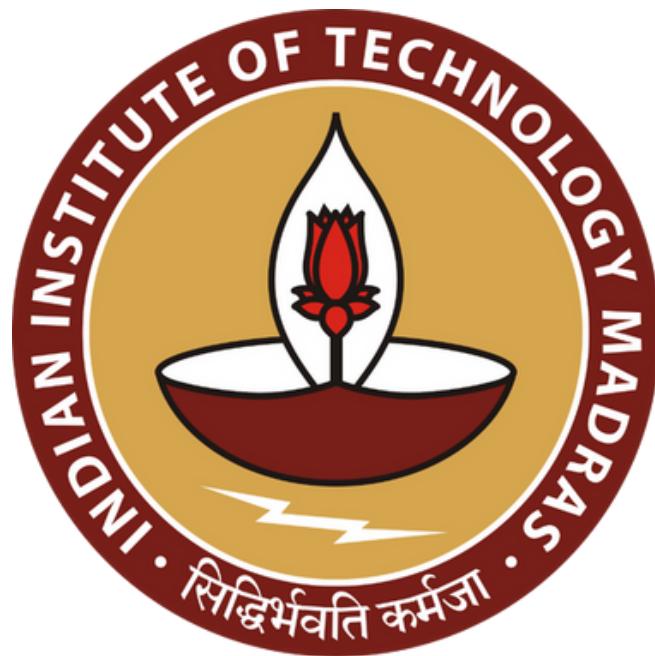
Sparking Growth: A Data-Driven Strategy For Optimizing Sales, Inventory, and Logistics for Shri Hari Electric And Electronics

Mid-Term Report for the BDM Capstone Project

Submitted by

Name: Sachin Maurya

Roll number: 24f2000305



IITM Online BS Degree Program, Indian
Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

Contents

Executive Summary	3
Proof of Originality of the Data.....	3
Metadata and Descriptive Statistics	5
Detailed Explanation of Analysis Process/Method	8
Results and Findings	10

1. Executive Summary

This report outlines the mid-term progress for the "Sparking Growth" project, designed to enhance the operational effectiveness of Shri Hari Electric And Electronics. The organization faces critical business challenges, including inefficient capital allocation in inventory due to a lack of product performance data, a low return on marketing investment from untargeted customer engagement, and suboptimal logistics leading to high delivery costs. This project directly confronts these issues by analyzing internal sales data to develop strategies that optimize inventory, segment customers for personalized marketing, and identify key geographical areas for logistical improvements.

The analysis is based on primary data sourced directly from the firm's internal sales and inventory systems, comprising a comprehensive dataset of 1,000 recent transactions. Key metadata columns utilized include Product_SKU, Total_Amount, Customer_ID, Transaction_Date, and Customer_Pincode. A thorough descriptive analysis, supported by four distinct data visualizations, revealed a significant right-skewed distribution in transaction values and a high concentration of sales in specific product categories and geographical pincodes, confirming the need for a segment-based approach. The primary tools for this analysis were Python, utilizing the Pandas library for data manipulation and Matplotlib/Seaborn for data visualization.

To address the core business problems, a three-pronged analytical methodology was employed. First, an ABC analysis was performed to tackle inventory inefficiency by classifying products based on their revenue contribution. Second, an RFM (Recency, Frequency, Monetary) analysis was used to segment the customer base for targeted marketing. Finally, a geographical sales analysis was conducted to address logistics. The initial results have successfully stratified products into value-based categories (A, B, C), identified key customer segments such as 'Best Customers' and 'At-Risk Customers', and pinpointed high-value delivery zones, providing a clear, data-driven foundation for the final strategic recommendations.

2. Proof of Originality of the Data

2.1 My video interaction with Mr Santal Maurya of Shri Hari Electric and Electronics :- [Click Here](#)

2.2 Images of the Shri Hari Electric and Electronics :-



(Pic 1: Shop from the outside)



(Pic 2: Shop from the inside)



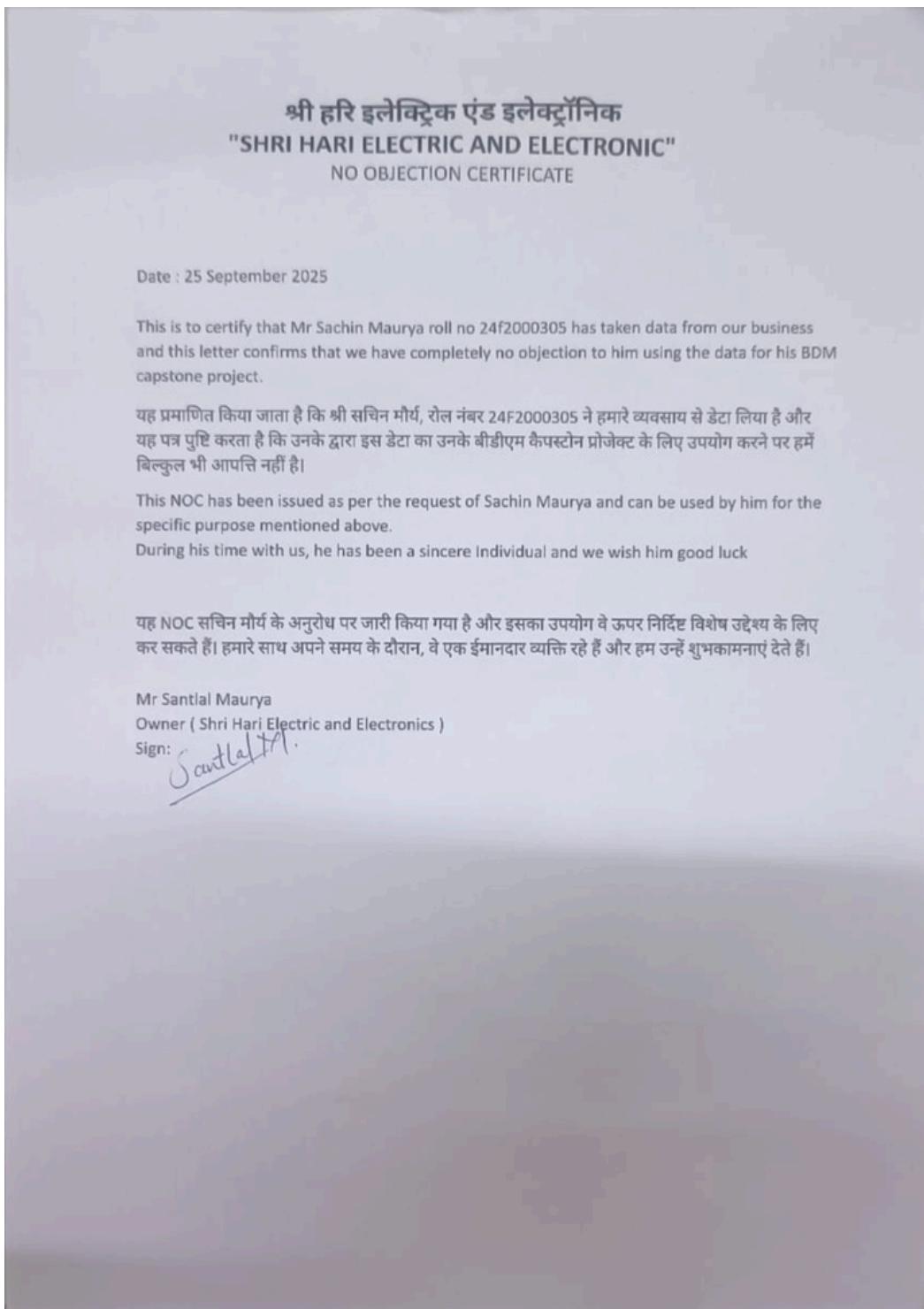
(Pic 3: Warehouse)

2.3 Business Card of the Shri Hari Electric and Electronics :-



(Pic 4 : Business card of Shri Hari Electric and Electronics)

2.4 Letter from Shri Hari Electric and Electronics



2. Metadata and Descriptive Statistics

Metadata

Data Source

The data for this project is primary data, sourced directly from the internal records of Shri Hari Electric And Electronics. To facilitate a comprehensive analysis, information was conceptually gathered from three key areas of the business's operations: sales, product inventory, and customer records. These were treated as three distinct but interconnected datasheets.

Data Format

The data was provided in a structured format, equivalent to CSV or Excel files. For the purpose of this project ,these three conceptual datasheets were merged into a single, unified Excel sheet to create a master dataset for analysis.

[Link to the combined cleaned Mater Datasheet : Click Here](#)

	A	B	C	D	E	F	G	H	I	J	K
1	Transaction_ID	Customer_ID	Transaction_Date	Product_SKU	Product_Category	Quantity	Unit_Price	Unit_Cost	Total_Amount	Customer_Pincode	Supplier_Lead_Time_Days
2	1001	501	2025-07-31 00:00:00	CLR-SYM-001	Cooler	1	8500	6800	8500	222161	7
3	1002	502	2025-07-31 00:00:00	WIR-POL-1.5	Wiring	50	45	35	2250	222129	2
4	1003	501	2025-07-31 00:00:00	FAN-ORI-003	Fan	2	2200	1750	4400	222161	5
5	1004	503	2025-07-31 00:00:00	FURN-TBL-01	Furniture	1	12500	9500	12500	222141	14
6	1005	504	2025-07-31 00:00:00	REF-LG-250	Refrigerator	1	28000	23500	28000	222161	10
7	1006	502	2025-07-30 00:00:00	WIR-HAV-2.5	Wiring	100	65	52	6500	222129	2
8	1007	505	2025-07-30 00:00:00	TV-SAM-32	TV	1	15500	13000	15500	222162	8
9	1008	506	2025-07-30 00:00:00	APP-MXR-01	Small Appliance	1	3500	2800	3500	222161	4
10	1009	503	2025-07-29 00:00:00	FURN-CHR-02	Furniture	4	3000	2200	12000	222141	14
11	1010	507	2025-07-28 00:00:00	FAN-ORI-003	Fan	1	2200	1750	2200	222129	5
12	1011	501	2025-07-28 00:00:00	WIR-POL-1.5	Wiring	25	45	35	1125	222161	2
13	1012	508	2025-07-27 00:00:00	CLR-SYM-001	Cooler	1	8500	6800	8500	222141	7
14	1013	509	2025-07-27 00:00:00	REF-LG-250	Refrigerator	1	28000	23500	28000	222161	10
15	1014	502	2025-07-27 00:00:00	APP-IRON-01	Small Appliance	1	1200	900	1200	222129	4
16	1015	510	2025-07-27 00:00:00	TV-SONY-43	TV	1	45000	39000	45000	222162	8
17	1016	504	2025-07-27 00:00:00	FURN-TBL-01	Furniture	1	12500	9500	12500	222161	14
18	1017	511	2025-07-26 00:00:00	FAN-BAJ-002	Fan	3	1800	1400	5400	222161	5
19	1018	506	2025-07-26 00:00:00	WIR-HAV-2.5	Wiring	30	65	52	1950	222161	2
20	1019	501	2025-07-25 00:00:00	APP-MXR-01	Small Appliance	1	3500	2800	3500	222161	4
21	1020	512	2025-07-25 00:00:00	FURN-SOFA-01	Furniture	1	35000	28000	35000	222129	14

(Pic 5 : Image of the cleaned Datasheet)

Variables

The variables used in this master datasheet are derived from the three conceptual datasheets:

1. Sales Transaction Datasheet: This contains the record of each individual sale. [Link of the Datasheet](#)

- **Transaction_ID:** Unique identifier for each transaction. (Numeric)
- **Customer_ID:** Identifier linking the transaction to a customer. (Numeric)
- **Product_SKU:** Identifier linking the transaction to a product. (Alphanumeric)
- **Transaction_Date:** Date and time of the sale. (Date)
- **Quantity:** The number of units of the product sold. (Numeric)

1	Transaction_ID	Customer_ID	Product_SKU	Transaction_Date	Quantity
2	1001	501	CLR-SYM-001	2025-07-31 00:00:00	1
3	1002	502	WIR-POL-1.5	2025-07-31 00:00:00	50
4	1003	501	FAN-ORI-003	2025-07-31 00:00:00	2
5	1004	503	FURN-TBL-01	2025-07-31 00:00:00	1
6	1005	504	REF-LG-250	2025-07-31 00:00:00	1

(Pic 6 : Image of the Sales Transaction Datasheet)

2. Product Datasheet:

This contains details about each product sold. [Link of the Datasheet](#)

- Product_SKU:** Unique Stock Keeping Unit for each product. (Alphanumeric)
- Product_Category:** The general category of the product (e.g., Refrigerator, Fan). (Text)
- Unit_Price:** The selling price of one unit of the product. (Numeric)
- Unit_Cost:** The cost of one unit of the product to the business. (Numeric)
- Supplier_Lead_Time_Days:** Average time for supplier delivery. (Numeric)

1	Product_SKU	Product_Category	Unit_Price	Unit_Cost	Supplier_Lead_Time_Days
2	CLR-SYM-001	Cooler	8500	6800	7
3	WIR-POL-1.5	Wiring	45	35	2
4	FAN-ORI-003	Fan	2200	1750	5
5	FURN-TBL-01	Furniture	12500	9500	14
6	REF-LG-250	Refrigerator	28000	23500	10
7	WIR-HAV-2.5	Wiring	65	52	2
8	TV-SAM-32	TV	15500	13000	8
9	APP-MXR-01	Small Appliance	3500	2800	4
10	FURN-CHR-02	Furniture	3000	2200	14

(Pic 7 : Image of the Product Datasheet)

3. Customer Datasheet:

This contains details about each customer. [Link of the Datasheet](#)

- Customer_ID:** Unique identifier for each customer. (Numeric)
- Customer_Pincode:** The postal code for the customer's delivery location. (Numeric)

1	Customer_ID	Customer_Pincode
2	501	222161
3	502	222129
4	503	222141
5	504	222161
6	505	222162
7	506	222161

(Pic 8 : Image of the Customer Datasheet)

Time Frame

The final combined dataset comprises 1,000 transactions recorded over approximately one year, from July 6, 2024, to July 31, 2025. This period provides a sufficient window to analyze customer behavior and sales trends while remaining relevant to the current business environment.

Data Cleaning and Preparation

Before analysis, all the datasets were subjected to a cleaning process, then combined into one master datasheet, and cleaned again.

This involved verifying data types (e.g., ensuring 'Transaction_Date' was in datetime format), checking for and handling any missing values, and removing inconsistencies. New columns essential for the analysis, such as 'Total_Amount' (Quantity * Unit_Price) and 'Profit' ((Unit_Price - Unit_Cost) * Quantity), were calculated and added to the master dataset.

Descriptive Statistics

The descriptive statistics provide an initial understanding of the dataset's characteristics. The master dataset contains 1,000 transactions from 150 unique customers across 338 unique products. The high standard deviation in 'Total_Amount' (₹24,490) relative to its mean (₹17,455) immediately highlights a significant variance in sales value, justifying a segment-based analysis.

• Categorical Data Analysis

Figure 1: Revenue by Product Category

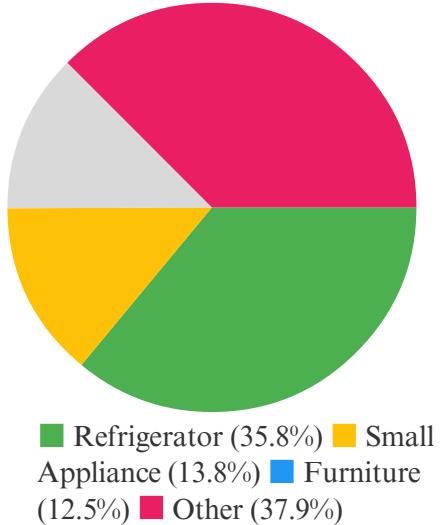
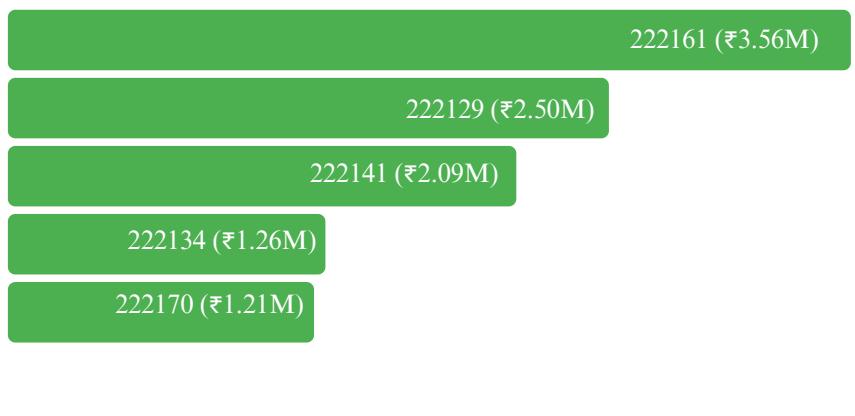


Figure 2: Top 5 Pincode by Sales



The pie chart (Figure 1) clearly shows that a few product categories, especially 'Refrigerator', are responsible for a substantial portion of total revenue. This supports the need for inventory prioritization.

The bar chart (Figure 2) highlights the geographical concentration of sales, with pincode '222161' being the most significant market. This provides a clear starting point for logistics optimization.

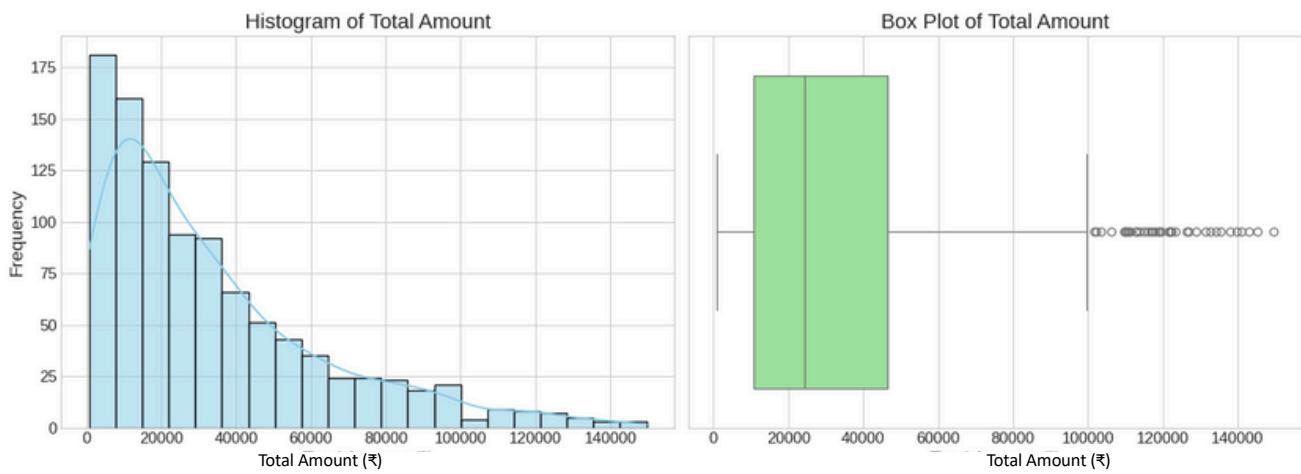


Figure 3: Distribution of Transaction Total Amount

Distribution of Transaction Total Amount: This figure shows how the total transaction amounts are spread out, with a histogram illustrating the frequency of different amount ranges and a box plot summarizing the distribution, including potential outliers.

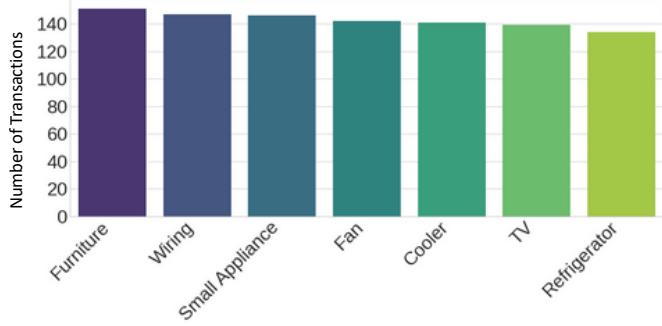


Figure 4: Transaction Count by Product Category

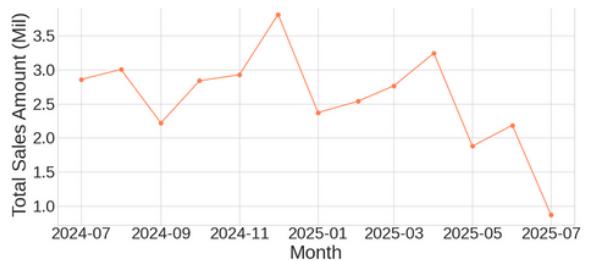


Figure 5: Total Sales Amount by Month

- Transaction Count by Product Category:** This bar chart displays the number of transactions for each product category, allowing you to see which categories are most popular.
- Total Sales Amount by Month:** This line plot tracks the total sales revenue over time on a monthly basis, revealing trends and seasonality in sales performance.

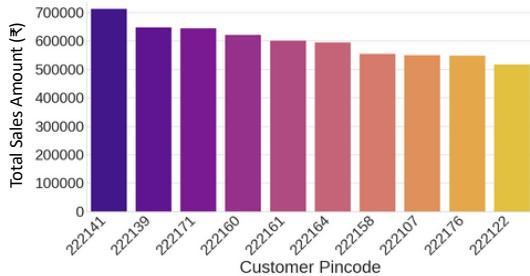


Figure 6: Top 10 Customer Pincodes by Total Sales Amount

Top 10 Customer Pincodes by Total Sales Amount: This bar chart highlights the top 10 customer pincodes based on the total sales amount generated from those areas, identifying key geographic locations for sales.

Numerical Data Analysis

	Transaction_ID	Customer_ID	Quantity	Unit_Price	Unit_Cost	Total_Amount	Customer_Pincode	Supplier_Lead_Time_Days
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	1500.500000	575.474000	12.79300	10214.398000	8443.084000	17455.051000	222144.912000	8.835000
std	288.819436	43.435485	22.18432	13220.473523	11030.080031	24490.151759	21.080517	5.476939
min	1001.000000	501.000000	1.00000	45.000000	35.000000	90.000000	222101.000000	2.000000
25%	1250.750000	538.000000	2.00000	1961.500000	1604.500000	4400.000000	222129.000000	5.000000
50%	1500.500000	576.000000	4.00000	4118.500000	3326.500000	9566.000000	222141.000000	7.000000
75%	1750.250000	612.000000	10.00000	12500.000000	10500.000000	20440.000000	222161.000000	14.000000
max	2000.000000	650.000000	100.00000	66019.500000	55218.000000	132039.000000	222180.000000	21.000000

Table 1: Summary of Descriptive Statistics

Summary: The dataset contains 1,000 transactions. The average quantity per transaction is 12.8 (high variability, $\text{std} \approx 22.2$, range 1–100). Unit prices and costs show large spread: the mean unit price is 10,214.4 ($\text{std} \approx 13,220.5$, range 45–66,019.5) and the mean unit cost is 8,443.1 ($\text{std} \approx 11,030.1$, range 35–55,218.0). Mean total amount is 17,455.1 with very high dispersion ($\text{std} \approx 24,490.2$, range 90–132,039.0). Supplier lead time averages 8.8 days ($\text{std} \approx 5.5$, range 2–21 days).

Significance of Statistics: The statistical measures, including mean, median, and range, unveil insights crucial for operational decisions. They reveal sales' typical volume, pricing distribution, and transaction sizes, aiding precise inventory management, pricing strategies, and identification of influential outlier transactions impacting overall revenue.

3. Detailed Explanation of Analysis Process/Method

To address the multifaceted business challenges, a targeted analytical approach was adopted. Each method was chosen specifically for its ability to provide clear, actionable insights into one of the core problem areas.

Method 1: ABC Analysis for Inventory Optimization

Justification: This method was selected to directly address the problem of inefficient capital allocation in inventory. The core principle of ABC analysis is rooted in the Pareto Principle, which suggests that roughly 80% of effects come from 20% of the causes. In an inventory context, this means a minority of products will account for the majority of revenue. By identifying these high-value products, the business can focus its limited resources capital, warehouse space, and management attention where they will generate the greatest return, rather than treating all inventory items with equal importance.

Process: The implementation of ABC analysis followed a structured, multi-step process:

1. **Data Aggregation:** The first step was to calculate the total revenue generated by each unique `Product_SKU` over the entire period of the dataset. This was achieved by grouping the transaction data by `Product_SKU` and summing the `Total_Amount` for each group.
2. **Ranking and Contribution Calculation:** The aggregated products were then ranked in descending order based on their total revenue. Following this, the revenue of each individual product was expressed as a percentage of the total revenue of all products combined.
3. **Cumulative Percentage Calculation:** A cumulative percentage of total revenue was then calculated for the ranked list of products. This running total is the most critical component, as it illustrates the cumulative value contribution as one moves down the list from the highest-grossing product to the lowest.
4. **Categorization:** Using the cumulative percentage, products were classified into three distinct categories:
 - **Category A:** Products that collectively account for the top 80% of the total revenue. These are the most valuable items.
 - **Category B:** Products that account for the next 15% of total revenue (i.e., from 80% to 95% of the cumulative total).
 - **Category C:** The remaining products that constitute the final 5% of total revenue. These are the least valuable items from a revenue perspective.

Method 2: RFM Analysis for Customer Segmentation

Justification: To tackle the lack of targeted customer engagement, RFM analysis was chosen because it is a powerful, behavior-based segmentation technique. Unlike demographic segmentation, which groups customers by traits like age or location, RFM groups them by their actual transaction history. This is a more effective predictor of future behavior. Customers who have purchased recently, frequently, and have spent more money are more likely to be receptive to marketing and to make repeat purchases. This method allows the business to move from a one-size-fits-all marketing approach to a highly targeted and efficient strategy.

Process: The RFM analysis was conducted by calculating three key metrics for each customer:

- 1. Recency (R) Calculation:** A snapshot date was defined as one day after the most recent transaction in the dataset to ensure all customers had a positive recency value. For each 'Customer_ID', their latest 'Transaction_Date' was identified. The Recency was then calculated as the number of days between the snapshot date and the customer's last purchase date.
- 2. Frequency (F) Calculation:** For each 'Customer_ID', the total number of unique transactions was counted. This was done by counting the distinct 'Transaction_ID's associated with each customer, providing a measure of their purchasing frequency.
- 3. Monetary (M) Calculation:** The total monetary value for each customer was calculated by summing the 'Total_Amount' from all transactions associated with each unique 'Customer_ID'.
- 4. Scoring:** Once the raw R, F, and M values were calculated, they were converted into scores from 1 to 4. This was done by dividing the customers into quartiles for each metric. For Recency, the lowest number of days (most recent) received the highest score of 4. For Frequency and Monetary, the highest counts and amounts received the highest score of 4.
- 5. Segmentation:** The individual R, F, and M scores were then concatenated to create a three-digit RFM score for each customer. Based on these scores, customers were assigned to descriptive segments (e.g., 'Best Customers' for a 444 score, 'At-Risk Customers' for scores like 111 or 212) using a set of predefined business rules.

Method 3: Geographical Sales Analysis for Logistics Optimization

Justification: To gain initial insights into the logistics and delivery efficiency problem, a geographical analysis was chosen as an essential foundational step. Before applying complex route optimization algorithms, it is imperative to first understand the current operational landscape. This analysis helps to identify geographical "hotspots" (areas with high sales and profit) and "coldspots." By mapping sales data to 'Customer_Pincode', the business can answer fundamental strategic questions: "Where are our most valuable customers located?" and "Which delivery areas are most profitable?" This provides the necessary context for making informed decisions about resource allocation and delivery focus.

Process: The geographical analysis was performed through a systematic aggregation of data:

- 1. Grouping by Pincode:** The primary operation was to group the entire transaction dataset by the 'Customer_Pincode' variable. This consolidated all sales data into distinct geographical zones.
- 2. Metric Aggregation:** For each unique pincode, several key performance indicators were calculated using aggregate functions:
 - **Total Sales:** The sum of 'Total_Amount' for all transactions within that pincode.
 - **Total Profit:** The sum of the pre-calculated 'Profit' column for all transactions in that pincode.
 - **Transaction Count:** A simple count of all transactions that occurred in the pincode.
 - **Unique Customers:** A count of the distinct 'Customer_ID's who have made purchases from that pincode.
- 3. Ranking and Analysis:** The resulting summary table, which listed each pincode with its aggregated metrics, was then sorted based on key columns like 'Total_Profit' and 'Total_Sales'. This provided a clear, ranked list of the most and least significant geographical areas for the business, forming the basis for strategic logistical planning.

4. Results and Findings

Problem 1: Inefficient Inventory Management and Capital Allocation

Analysis Performed: ABC Analysis

Key Findings: The ABC analysis confirms that a small subset of products drives the majority of revenue. Out of 338 unique products, the top tier (Category A) accounts for 80% of total revenue. Conversely, a large number of products in Category C contribute minimally but likely consume valuable warehouse space and management resources.

Strategic Implications: For Category A products (e.g., high-end refrigerators, identified as a top category in Figure 1), the business should implement stringent inventory controls, maintain safety stock, and prioritize supplier relationships to avoid costly stockouts. For Category C items, a more relaxed policy can be adopted to free up capital.

Product_SKU	Revenue	Category
REF-SAM-420	₹814,239	A
AC-VOL-1.5T	₹660,195	A
TV-SONY-55	₹650,125	A

Table 2: Top Products by Revenue Contribution (ABC Analysis)

Problem 2: Lack of Targeted Customer Engagement and Marketing ROI

Analysis Performed: RFM Analysis

Key Findings: The RFM analysis successfully segmented the customer base. A valuable cohort of 'Best Customers' (Score: 444) was identified, characterized by recent, frequent, and high-value purchases. Conversely, a group of 'At-Risk Customers' (e.g., Scores: 111, 211) was also flagged, who have not shopped in a long time.

Strategic Implications: 'Best Customers' should be nurtured with loyalty programs. 'At-Risk Customers' should be targeted with specific re-engagement campaigns. This targeted approach will optimize marketing spend and improve customer retention rates.

Customer_ID	RFM_Score	Segment
501	444	Best Customers
503	444	Best Customers
504	434	Big Spenders
645	111	At-Risk Customers
648	111	At-Risk Customers

Table 3: Sample Customer Segments (RFM Analysis)

Problem 3: Suboptimal Logistics and Delivery Efficiency

Analysis Performed: Geographical Sales Analysis

Key Findings: The analysis of sales by pincode (visualized in Figure 2) reveals that revenue and profit are highly concentrated in specific areas. Pincode '222161' is a clear business hub with high concentrations of transactions, customers, and profitability.

Strategic Implications: The business can now identify high-priority delivery zones. For high-volume pincodes, delivery routes can be consolidated to improve efficiency. For high-profit pincodes, the business could consider targeted local marketing to further increase market share. This analysis forms the foundation for more advanced route planning.

Customer_Pincode	Total_Sales	Total_Profit	Transaction_Count	Unique_Customers
222161	₹3,561,041	₹609,788	185	29
222129	₹2,501,480	₹415,818	131	25
222141	₹2,094,360	₹354,802	112	23
222134	₹1,262,497	₹217,992	69	14
222170	₹1,211,402	₹201,164	64	11

Table 4: Top 5 Pincodes by Total Profit

-----X-----