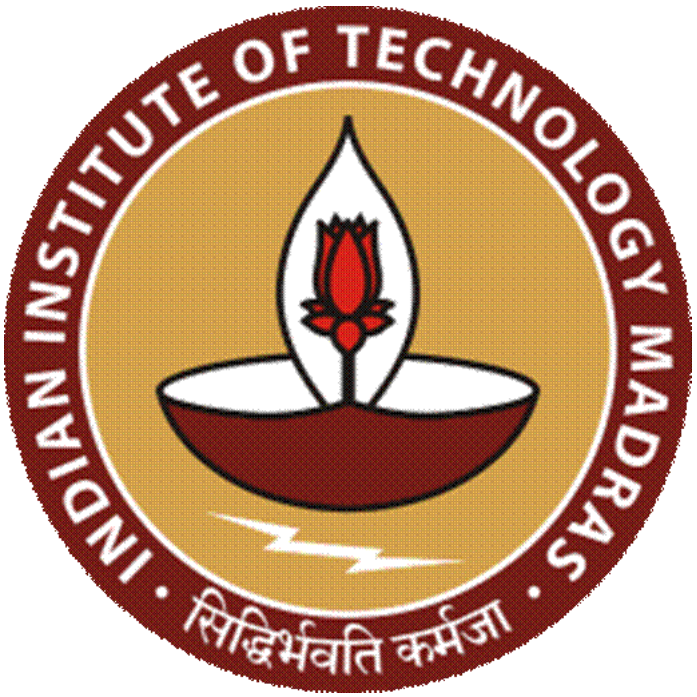
Optimizing sales and services of a leading Lab equipment firm

A Midterm Report for the BDM Capstone Project

Submitted by

Name: **V SANJANA**

Roll number: **23f2000029**



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1 Executive Summary

Sigma Scientific Products, established in 2015 and headquartered in Chennai, is a leading partnership firm specialized in manufacturing and selling scientific laboratory equipment to engineering colleges and schools for their laboratory needs.Despite its innovative offerings and growing presence, the company faces two major challenges: lack of insights on the product portfolio which is hindering them to rationalize, prioritize and market them effectively; lack of comprehensive understanding of the customer and corresponding sales which leads to ineffective marketing and sales effort.

This midterm report is based on primary data collected directly from Sigma Scientific Products, ensuring authenticity and originality. The data includes 12 month sales data from April 2024 to March 2025 and product data, provided by the business owner. Supporting evidence such as a signed letterhead, images of the organization and a video interaction with the owner is attached ensuring transparency and credibility of this research.

After cleaning and preprocessing the dataset, the preliminary analysis reveals high variability, with extreme outliers influencing overall revenue. Also, I have made various categorical analyses of data using excel and python libraries and charts to derive insights. The derived insights include information about ‘high ticket low frequency’ products, ‘low ticket high frequency’ products and redundant product details are derived. As part of customer segmentation and sales intelligence, sales and product data has been linked and analysed.

These insights emphasize the need for strategic demand planning and targeted marketing to stabilize sales and become more profitable. More understanding of the sales data linked with product data will help SIGMA Scientific Products to understand customer interests on potential products.

Extensive analysis of product portfolio data and customer data was done. The expected outcome is to rationalize the product portfolio and enhance targeted marketing effort which will ultimately increase companies’ sales and market presence which is presented in the ‘Results and Findings’ section.

2 Proof of Originality

This section validates the primary data collection, ensuring research authenticity. Supporting materials are available via the following Google Drive link.
The provided evidence includes:
• Datasets (Both Original and Cleaned)
• Official signed and stamped Letterhead
• Images of the organisation.
• Short Video of interaction with the owner.
Link: https://drive.google.com/drive/u/1/folders/1LlHYGCvC46Rgs8bzlONmOeedJ21Fyj3B

3 Metadata & Descriptive Statistics

3.1 Metadata

The product and sales data is recorded and maintained by Sigma in Tally software which is their internal accounting platform. The business owner provided the below records containing two datasets:

- Product Data – This is a master list of products sold by Sigma.
- Sales Data – This has all the sales made by Sigma from 1st April 2024 to 31st March 2025.

Dataset: Product_Data.xlsx

Table 1: Structure of Product data

Category	Product Name	Approx. Price (₹)	Usage	Sold By	Key Attributes	Model No.
Weighing Balance	WENSAR PGB 200	23100	Precision weighing in labs	WENSAR	200g, LCD, external cal	PGB-200
Bunsen Burner	Electric Bunsen Burner (with regulator)	3600	Flame heating in labs	SIGMA	Stainless steel, 800°C	BB-E01
Oven	High Temperature Oven	38000	High-temp heating	SIGMA	450°C, 343L, thermostat	HTO-450
Oven	Vacuum Oven (300x300 mm)	74000	Vacuum drying	SIGMA	200°C, 230V	VO-300
Autoclave	Vertical Autoclave	42000	Sterilization	SIGMA	20L, SS body	VA-20SS
Autoclave	Double Wall Autoclave 100L	110000	Sterilization	SIGMA	100L, 450x600 mm	DW-100

Overview:

The dataset “Product_Data.xlsx” contains the master list of products sold by Sigma. It includes 17 entries and provides insights into products like product category, product name, price, usage etc.

Data Composition:

Rows: 17
Columns: The dataset comprises 7 columns:

- CATEGORY: Broad grouping of the products like Oven, Autoclave, Weighing Balance. Helps analyze product diversity.
- PRODUCT NAME: Specific name of the product like Vertical Autoclave. Useful for cataloging and mapping to sales.

- APPROX. PRICE (₹) :Indicates approximate unit price. Aids in comparing against sales values and analyzing pricing strategy.
- USAGE: Describes primary applications like Sterilization,Precision weighing in labs. Useful for market segmentation.
- SOLD BY :Supplier/brand associated with the product such as *SIGMA*, *WENSAR*. This helps in evaluating vendor reliance.
- KEY ATTRIBUTES: Key features/specifications enable product differentiation analysis.
- MODEL NO.:Unique identifier provides linkage between sales data and product master.

Data Integrity & Anomalies:

- No missing values.
- Pricing appears consistent and within expected lab equipment ranges (₹3,600 – ₹74,000).

Dataset: Sales_Data.xlsx

Table 2: Structure of Sales data

Date	Particulars	Credit	Vch No	Product code	Vch Type
4-Apr-24	Jayam Scientific Company - Coimbatore	5400	01/24-25	BB-E01	Sales
6-Apr-24	SRF Limited (Viralimalai)	6000	06/24-25	BB-E01	Sales
11-Apr-24	Geolag International	78490	07/24-25	C24BL	Sales
13-Apr-24	Karunya University - Coimbatore	32000	09/24-25	SVP-01	Sales
18-Apr-24	Accuracy Analabs Pvt Ltd	55000	011/24-25	VO-300	Sales
22-Apr-24	Anabond Limited	7000	012/24-25	BB-E01	Sales

Overview:

The dataset “Sales Data” captures the sales transactions for the entire financial year (from April 2024 till March 2025) covering clients, product codes, sales values and voucher #. It enables tracking of revenue, customer segments and product level sales performance.

Data Composition:

Rows: 139
Columns: The dataset comprises 6 columns:

- DATE: Transaction date helps to identify sales patterns and seasonality.
- PARTICULARS: Buyer name for customer segmentation and identifying top clients.
- CREDIT: Transaction value in rupees serves as the main revenue indicator.
- VCH NO.: Voucher number acts as a unique transaction reference.

- PRODUCT CODE: Code linking to Product data that is MODEL NO. Enables product wise sales analysis.
- VCH TYPE: Type of entry.

Data Integrity & Anomalies:

- No missing values
- Credit values range widely (₹222 – ₹3,00,000), suggesting both small and large orders.
- Product codes are consistent with the product data ensuring clean linkage between datasets.

3.2 Descriptive Statistics

Numerical Summary

	count	mean	std	min	25%
pprox. Price (₹)	17.0	63111.764706	73191.383393	3600.0	18900.0

	50%	75%	max	IQR	skewness
pprox. Price (₹)	32500.0	85000.0	300000.0	66100.0	2.347602

Figure 3.2.1.1: Value analysis of product data

The product dataset contains 17 products with their approximate prices ranging from ₹3,600 to ₹3,00,000 indicating a wide variation in product pricing. The mean is around ₹63000 is heavily influenced by a few high-priced items that is the max value meaning it does not represent the typical product in the catalog.The median is substantially lower than the mean, at ₹32,500 which confirms that most products fall in the moderate price segment, while only a few expensive products raise the overall average. The large standard deviation of ₹73191 compared to the mean reflects significant variability in product prices. The interquartile range indicates that the middle 50% of products are priced between approximately ₹18,900 and ₹85,000. This suggests that the core product range is mid-priced.
The positive skew shows that the distribution has a long right tail. A few very high-priced outliers disproportionately affect the average.The full pricing range spans from low-cost lab equipment ₹3,600 to high-end machines ₹3,00,000. This demonstrates the company's strategic coverage of multiple market segments, catering to both cost-sensitive and high-value customers.

Categorical Analysis

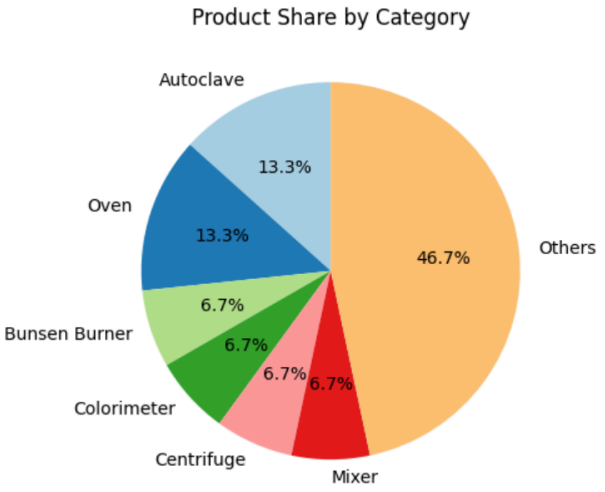


Figure 3.2.1.2:Category distribution analysis using pie chart

The product catalog is extremely **broad but shallow** with 13 categories (7 categories are grouped under others) for 17 products.Only two categories have 2 products that are Oven and Autoclave. The rest are single-product categories (e.g., Refrigerator, Dryer, Spectrophotometer, Bunsen Burner).

4 Detailed Explanation of Analysis Process/Method

4.1 Data Cleaning and Preprocessing

The original dataset required thorough cleaning and preprocessing to ensure accurate analysis.
Key preprocessing steps included:

- Normalized column names and trimmed whitespace.
- Normalized into a datetime column and derived Month for monthly analysis.
- Normalized product key by joining Sales's Product code with Products's Model No.to attach Category, Product Name, Usage, Approx. Price, etc.
- Cleaned Approx. Price (₹) into numeric where present by removing symbols like comma and rupees.
- Standardized customer names. For example some entered IITM and some as Indian Institute of Technology, Madras for accurate customer segmentation.
- Extracted the city names for location mapping from the Particulars.
- The product code in Product_data.xlsx were mapped to the Sales_data.xlsx with the Approx price for creating a merged dataset to work with.
- The dataset was also checked for missing values using python commands.

4.2 Methods/Analysis

After cleaning the dataset, I first analyzed Product_data.The column category and price was analyzed well for deriving insights about the product portfolio.

4.2.1 Product Rationalization

A pie chart was drawn, *Figure 3.2.1.2* to see which category is mostly available in the company.This pie chart was plotted by grouping the Category column in the product data after standardizing the names.

Category % = (Count of products in the category / Total products) × 100

A box plot, *Figure 5.1.1* is the best way to find how the price of the products are spread.Here, the minimum product price, maximum price, average and also the outliers are specified.Since I had to understand how the price alone is spread, I plotted a box plot for the continuous values in the dataset. Also here, since the prices of the products belonging to same category varies, to find a uniform price for one product, the mean has been calculated using this formula

Approx. Price (₹) =(Sum all products grouped by category)/(No of products in that category)

Q1 = 25th percentile (lower quartile) = ₹16,950
Q2 = 50th percentile (median) =₹ 38,000
Q3 = 75th percentile (upper quartile) = ₹97,500
IQR = Q3 - Q1 => ₹ 80,050

While the pie chart gave the category distribution, A bar chart, *Figure 5.1.2* is the best way to understand the price vs category. Category is categorical data, while Price is continuous data. The bar cart here, tells us how price varies between categories. The category of Bunsen Burner seems to be redundant.A bar chart between price and category clearly tells which category is highly priced, average and low.

4.2.2 Comparative Pricing and Sales Intelligence

A bubble chart,*Figure 5.2.1,* is plotted to understand the average price and the total sales value of that particular product. Here, using the product code the sales and product data is merged.The sales column is first detected by the column name credit. A new column called quantity was introduced by counting the product_code. column. A column named “avg_price_prod” was created by mapping the product_code, and its corresponding price’s mean. (both grouped by code)

avg_price_prod = total_price_sum / number_of_records

total_sales column is added by again summing up the sale value by mapping the product_code.

$$\textbf{total\_sales} = \Sigma(\textbf{quantity} * \textbf{price})$$

The bubble here represents sold/product code and the size is the quantity of a particular product .

$$\textbf{quantity} = \sum \textbf{units sold for that product}$$

$$\textbf{bubble size} = \min(\max((\text{Quantity} / \text{Quantity\_max}) \times 1200 + 30 , 30), 4000)$$

This ensures that the smallest bubble has a minimum size of 30 and largest bubble has a maximum size of 4000 preventing very small or very large bubbles from dominating the visualization.Multiplying by 1200 scales that normalized number to a reasonable display size range.All other bubbles are scaled proportionally based on product sales quantity.

4.2.3 Customer Segmentation

The products were grouped by total sales and transaction count. The results were sorted and finally displayed in a horizontal bar chart for,*Figure 5.3.1* the top ten companies.Here, the data was processed by adding a new column stating Customer Name by standardizing all the duplicates such as VIT and Vit University. Also, the location details were trimmed from the Particulars for further analysis.

To analyze the monthly income, a line chart,*Figure 5.3.2* is plotted by extracting the month and year in one axis and the total sales made in that month in another axis.Here the peak and low sales cycles are identified for further forecasting.

5 Results and Findings

5.1 Product Portfolio Rationalization

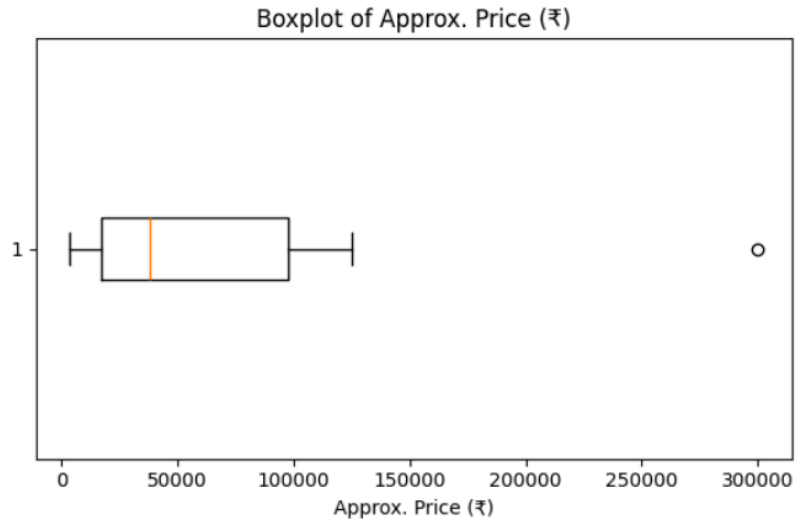


Figure 5.1.1: Product pricing spread using box plot

The box plot shows a strong right skew which means a few very expensive products drive the mean, which includes the ₹3,00,000 refrigerator and the ₹1,25,000 dryer/spectrophotometer.Boxplot highlights these as outliers or **high-ticket items,** the presence of specialized, high-value items that should be monitored for profitability.The high interquartile range (₹80,050) shows a **broad spread of prices** meaning the company offers products which supports different market segments but may also lead to redundancy. This can attract diverse customer segments, but it also makes pricing strategy and inventory management more complex.

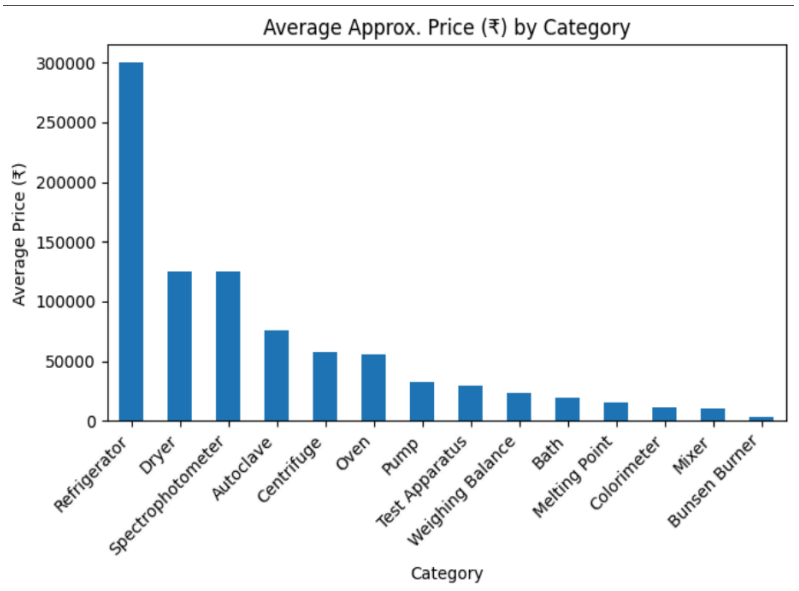


Figure 5.1.2: Category vs average price using bar chart

From the above analysis it is clearly evident that the major effect on the company's sales might be due to LR-1355 as the average price of the product is very high and has high influence on other factors such as mean etc., Thus, losing LR-1355 might be riskier for the company.

After comparing the key findings from the above three graphs and plots we can conclude two distinct selling modes

- **High-ticket, low-frequency items** that is the refrigerator for ₹3,00,000.
- **Low-ticket, high-frequency items** for example, Autoclave and Ovens (derived from the pie chart) which are from ₹19,000 to ₹85,000 which in the IQR.Each needs a different GTM, inventory and margin treatment.
- Products with similar price/features in the same category may be **redundant.**

5.2 Comparative Pricing and Sales Intelligence

Comparative Pricing & Sales Intelligence by linking Product and Sales Data.

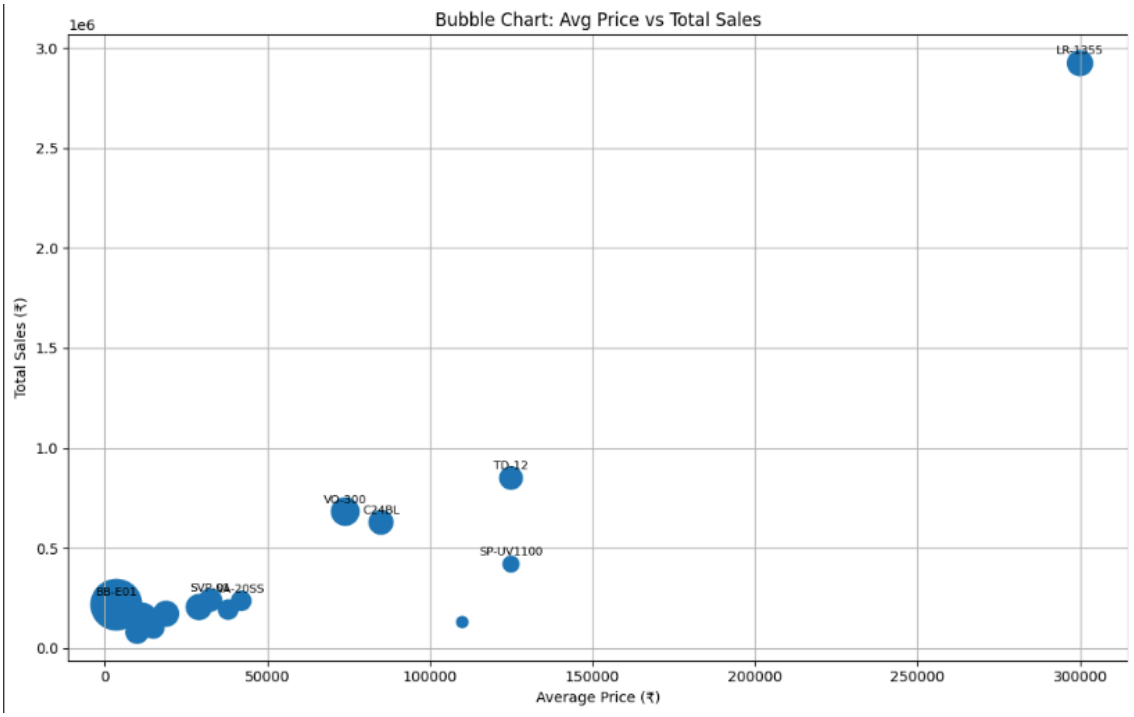


Figure 5.2.1: Average price vs total sales using bubble chart

The bubble plot derives the following categorization of products.

Top-right quadrant represents a high price and makes high sales for the company:They bring big revenue and are high-ticket .E.g., LR-1355.

Top-left represents low price, high sales: Volume movers or low-price fast sellers in the above data. For example, BB-E01.

Bottom-left represents low price and low sales:A lot of products are low-moderately priced, i.e., between ₹19,000 to ₹85,000. Therefore, they have made better sales. Also, the size of the bubbles are bigger here which denotes more quantity has been sold but made less sales value compared to others . For example, BB-E01, Bunsen burner.

Bottom-right high price and low sales: Niche products with low demand.

5.3 Customer Segmentation

Customer Segmentation by Sales Volume and Transactions

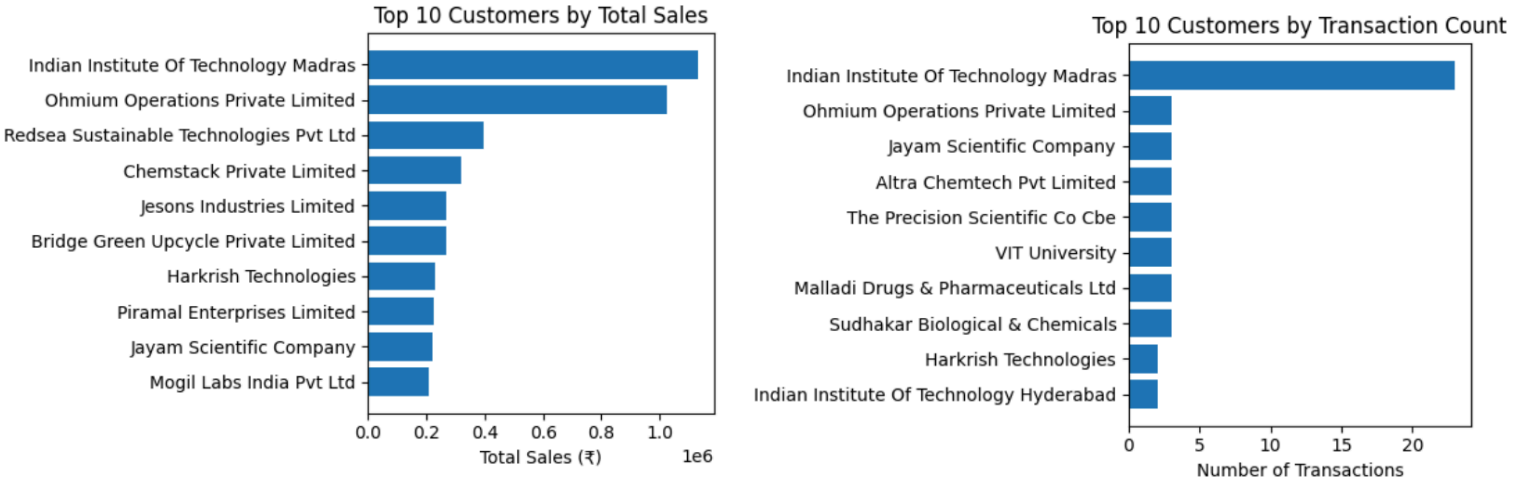


Figure 5.3.1: Top 10 customers by credit value and transaction count using bar graph

The bar chart shows us that the Indian Institute of Technology Madras has made the highest sales as well as made the maximum number of transactions compared to the rest of the companies. The graph also tells us that some companies who buy more products do not buy products frequently. Hence, Indian Institute of Technology Madras is a high-value customer buying products frequently.A small number of customers drive a large share of revenue. This is an account management opportunity and a concentration risk.

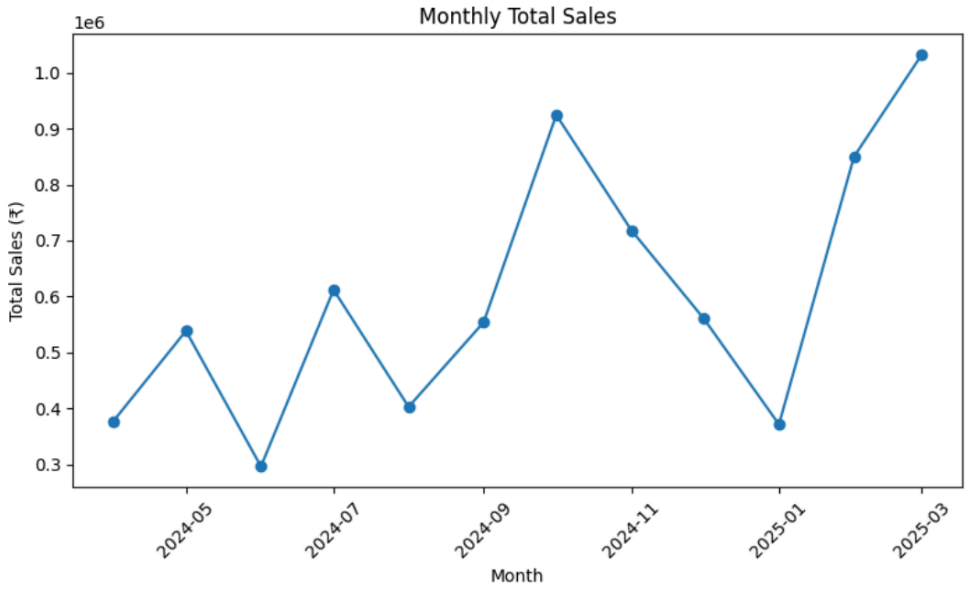


Figure 5.3.2: Monthly sales using a line chart

The monthly sales vary month-to-month peaks and troughs for categories spike at different months (project-driven purchases for high-ticket items).Knowing peaks helps inventory planning and sales campaigns.From the above line chart it is clearly evident that highest sales are in the month of March. As discussed with the owner, the main reason behind this is that colleges and industries try to exhaust their funds by the end of March.

In the Final-Term Submission, I would discuss the numerical aspect of the data and find the solutions to the problem statements by the interpretation of the results that we will get by some more analysis of the data.