

# **Indian Institute of Technology Madras**



**A Mid-Term report**

## **Revenue Optimization and Demand Forecasting Analysis of a Dairy Firm**

---

**BUSINESS DATA MANAGEMENT – CAPSTONE PROJECT**

---

**Submitted by**  
Name : Abhay Sahu  
Roll number : 23f1001119

Indian Institute of Technology, Madras, Chennai  
Tamil Nadu, India, 600036

---

## Table of Contents

<b>CONTENT</b>	<b>PAGE NO</b>
○ Executive Summary	2
○ Proof of Originality	3
○ Metadata and Descriptive Statistics	4-5
○ Detailed Explanation of Ananlysis Process/Method	5-9
○ Results and Findings	9-11

## 1. Executive Summary

Shree maa dairy is a regional dairy supplier dealing in fresh milk, curd, ghee, and other dairy products. The business problem that could be identified from the discussion with the owner are as follows:

- o The problem of reaching a large customer base
- o Spoilage due to overstocking
- o Seasonal demand fluctuations

**Goal:** Optimize dairy sales and distribution using 6-month sales data (Nov 2024 to Apr 2025).

The approach followed to do this capstone project is as follows:

Step 1: Define objectives

Step 2: Collect & clean data (Excel sheets from invoices)

Step 3: Conduct descriptive and exploratory analysis(Using Python)

Step 4: Visualize results with charts (bar, pie, histogram, heatmap, Pareto)

Step 5: Interpret insights for decision-making

In the proposal stage, the objective was defined as optimizing revenue generation and demand forecasting for Shree Maa Dairy. The firm has been facing challenges of seasonal demand fluctuations, overproduction of perishable items, and inconsistent revenue patterns. The chosen goal of this study is to leverage data-driven methods to improve product mix decisions, reduce wastage, and ensure stable cash flows.

The data collected spans three months of sales (Nov 2024, Jan 2025, Apr 2025) across more than 25 dairy products such as Fresh Milk, Neno Milk Pouches, Curd, Ghee, and Lassi. But after taking the bootcamp Sir said take more month data three month data is not efficient to do demand forecasting as of which is my problem statement so I have collected total six month data from (Nov-24 to April 2025). Each dataset captures product name, quantity sold, rate per unit, and revenue generated, which allows a detailed breakdown of product-level and category-level performance.

Initial descriptive analysis revealed that monthly revenues ranged from ₹83.6 lakh to ₹1.05 crore, with Fresh Milk and Neno Milk Pouches emerging as the most consistent contributors. Seasonal products like Ghee tins showed sharp spikes in winter and festival months, whereas Curd and Lassi displayed high variability and perishability risk. Together, the top five products contribute nearly 40% of total revenue, demonstrating the classic Pareto principle (80/20 rule) in action.

Pictorial representations such as line charts, bar graphs, pie charts, heatmaps, and Pareto curves have been created for better visualization and understanding. These analyses help uncover the story behind the numbers: stable core products drive regular income, while seasonal and perishable items require smarter stocking and forecasting strategies.

Using forecasting tools and optimization methods taught in the Business Data Management course. The project aims to help Shree maa dairy achieve stable revenue growth by aligning production with seasonal demand patterns.

## 2. Proof of Originality



**Shop Location**



**Product Banner**



**Owner with Staff**

Video link : [Interaction with owner](#)



**Photo with owner(office)**

Letter : [Signed Letter From Organization](#)

### 3. Metadata & Descriptive Statistics

The dataset used in this study represents detailed sales records for various dairy products of Apra Milk Company across multiple months. Each record captures both qualitative and quantitative attributes relevant to understanding product-level performance and overall business trends.

The metadata of the above mentioned column headers are explained as follows:

Particulars	Outwards Qty	Rate	Amount
-------------	--------------	------	--------

The **metadata** provides structural information about the dataset. It includes five key attributes:

Attribute	Description	Type	Example Value
Particulars	Product name	object	Full Fat Milk Pouch 500 ML
Outwards Qty	Quantity sold	Numeric (Float)	44773
Rate	Selling rate per unit (₹)	Numeric (Float)	30.47
Amount	Total revenue per product	Numeric (Float)	13,64,431.60
Month	Month of record (added column)	object	January

- Particulars** – the product name or description (e.g., *Full Fat Milk Pouch 500 ML*), represented as text data.
- Outwards Qty** – the total quantity of each product sold during the month, stored as a numeric (float) value.
- Rate** – the selling price per unit of the product (in ₹), also a numeric attribute.
- Amount** – the total revenue generated by the product, calculated as quantity multiplied by rate.
- Month** – an added column using python representing the month of record (e.g., January, February, etc.), allowing for temporal comparison and trend analysis.

These attributes form the foundation for all subsequent analyses, including monthly aggregation, product performance comparison, and seasonal trend evaluation.

### Descriptive Statistics

The **descriptive statistics** provide a quantitative summary of the dataset. The dataset contains **28 product entries**, showing a wide range of sales and pricing behaviors across product types.

index	Outwards Qty	Rate	Amount
count	29.0	29.0	29.0
mean	11993.975862068964	397.4106896551724	327560.5024137931
std	34088.889245216116	1142.3154521438685	525399.5411650811
min	14.0	6.67	5508.0
25%	333.0	17.2	20741.64
50%	680.0	55.28	76346.43
75%	2949.0	334.93	302874.68
max	169355.0	6191.39	1762273.0

Figure 1

- The **average quantity sold** (mean Outwards Qty) is approximately **11,994 units**, but the high standard deviation (~34,089) indicates large variation — some products like *Fresh Milk* and *Neno Milk Pouches* sell in bulk, while others like *Ghee* or *Shri Khand* have lower, premium-volume sales.
- The **average selling rate** per product is around **₹397**, with prices ranging from ₹6.67 (low-cost items like neno milk) to ₹6,191 (high-value ghee tins).
- The **average revenue (Amount)** per product is about **₹3.28 lakh**, with a substantial standard deviation (~₹5.25 lakh), reflecting a strong skew toward a few high-selling items.
- The **minimum total sales** recorded for any product is ₹5,508, while the **maximum** reaches ₹17.6 lakh — reaffirming the 80/20 principle where a few products dominate overall revenue.

In summary, the metadata and descriptive analysis reveal that Apra Milk Company's dataset is well-structured for exploratory data analysis, with strong variability across products and months. The large spread in both sales volume and rate demonstrates the diversity of the product portfolio — from low-cost, high-volume items like milk pouches to premium, low-volume products such as ghee and paneer. These insights set the stage for deeper analyses such as month-wise trend evaluation, category performance, and forecasting.

## 4. Detailed Explanation of Analysis Process/Method

### Introduction

After cleaning and structuring the monthly datasets represents six months of product-wise sales across a diverse SKU portfolio — ranging from **milk, curd, ghee, beverages, paneer, and sweets**. Since this dataset records information across consecutive months (November 2024 to April 2025), a **time-series analysis technique** was adopted.

Time-series analysis allows us to identify **seasonal sales trends, monthly fluctuations, and demand patterns** influenced by weather, festivals, and consumer habits. To achieve this, the study employed the four fundamental stages of data analysis:

- **Descriptive Analysis** – Summarized total sales, product performance, and category shares.
- **Diagnostic Analysis** – Compared categories to detect causes of sales increase or decrease (e.g., seasonal demand).
- **Predictive Analysis** – Used regression and trend lines to forecast future sales.
- **Prescriptive Analysis** – Suggested actions for inventory optimization, reducing spoilage, and increasing profitability.

The analysis was done using a combination of **Microsoft Excel 2007 (Pivot Tables)** and **Python (pandas, matplotlib, seaborn)** for visualization and validation.

## Objective

The main objective of this project is to identify **monthly and category-wise sales performance** and analyze **demand seasonality** in order to optimize stocking, pricing, and distribution.

Secondary objectives include:

- Understanding **which SKUs (Stock Keeping Units)** contribute most to overall revenue.
- Comparing **product categories (Milk, Ghee, Curd, Beverages, Paneer, Sweets)** to determine their sales dominance.
- Assessing **consumer demand behavior** during specific months such as winter (Dec–Jan) vs. summer onset (Mar–Apr).
- Generating insights that can guide future **inventory and marketing decisions**.

## Dataset and Variables

The dataset used for this analysis contains detailed monthly sales records of more than 28 dairy products from **November 2024 to April 2025**. Each entry corresponds to a unique product with its respective sales quantity, selling rate, and total revenue.

The data fields include the **product name (Particulars)**, **outwards quantity, unit rate**, and **amount**, along with the **month** in which the sale occurred. These attributes (already defined in the Metadata section) form the foundation for performing descriptive, comparative, and time-series analyses.

Additionally, each product has been categorized using the **SKU table shared by the company** (refer to Figure 2) into broader groups such as *Milk Products*, *Beverages*, *Ghee & Fat Products*, *Curd & Fermented Products*, and *Paneer, Sweets & Others*.

This structured classification allowed for more meaningful **category-wise aggregation**, trend detection, and revenue contribution analysis during the later stages of the project.

SKU OF THE PRODUCTS				
Milk Products				
SKU Code	Product Name	Packaging Size	Unit Type	Remarks
M001	Fresh Milk	1 Liter Pouch	Litre	Stable household demand
M002	Neno Milk Pouch 200ml	200 ml	Pouch	High-volume daily product
M003	Full Fat Milk Pouch 500ml	500 ml	Pouch	Strong performer
M004	Toned Milk Pouch 500ml	500 ml	Pouch	Moderate demand
M005	Loose Milk	Variable	Litre	Used by hotels & canteens
Beverages				
SKU Code	Product Name	Packaging Size	Unit Type	Remarks
B001	Lassi Bottle	250 ml	Bottle	Seasonal summer drink
B002	Mattha Bottle	250 ml	Bottle	Popular during hot months
B003	Butter Milk	200 ml	Bottle	Low margin, high turnover
Ghee & Fat Products				
SKU Code	Product Name	Packaging Size	Unit Type	Remarks
G001	Ghee Tin 15Kg	15 Kg	Tin	High-margin, festive product
G002	Ghee Jar 1L	1 Litre	Jar	Premium retail SKU
G003	Butter	500 gm	Pack	Consistent sales
Curd & Fermented Products				
SKU Code	Product Name	Packaging Size	Unit Type	Remarks
C001	Curd Pouch	500 gm	Pouch	Daily use, high spoilage risk
C002	Dahi Cup	200 gm	Cup	Used in small households
C003	Lassi Cup	200 ml	Cup	Summer seasonal item
Paneer, Sweets & Others				
SKU Code	Product Name	Packaging Size	Unit Type	Remarks
P001	Paneer Poosie	500 gm	Kg	Niche, festive product
P002	Khoya	1 Kg	Kg	Used in sweets
P003	Shri Khand Cup	100 gm	Cup	Very low sales, disconti

Figure 2

The SKU mapping also helps in comparing similar product types—such as milk variants, ghee packs, or fermented products—based on their sales volume, packaging, and seasonal behavior. The following table represents the SKU categorization used for further analysis.

## Pre-Processing of Data

Data preprocessing was a crucial step before analysis. Each month's sales were recorded in separate Excel sheets named *Nov*, *Dec*, *Jan*, *Feb*, *Mar*, and *Apr*. These were merged into a single dataset using Python and cross-checked in Excel.

### Steps followed:

1. **Data Import and Merging:**
  - o Each month's sheet was imported using `pandas.read_excel(sheet_name=None)` and concatenated into one master DataFrame.
  - o A new column “Month” was added to track the source of each record.
2. **Data Cleaning:**
  - o Missing or blank Rate and Amount values were corrected using:  
→ `Amount = Rate × Outwards Qty` where applicable.
  - o Duplicate rows were removed using `df.drop_duplicates()`.
  - o Data types were standardized (`float` for numeric, `object` for categorical fields).
3. **Category Mapping:**
  - o Using the SKU classification table (Figure 2), each product was mapped to a category — *Milk*, *Beverages*, *Ghee*, *Curd*, or *Panner & Sweets*.
  - o Example: “Lassi Cup” → Beverages; “Ghee Jar” → Ghee & Fat Products.
4. **Validation:**
  - o Pivot tables in Excel were used to confirm monthly total sales matched with the computed totals from Python.
  - o Totals were compared manually to ensure accuracy across sources.

## Analysis Approach

The analytical process was divided into **five stages**:

### *Stage 1: Descriptive Analysis*

Initial descriptive statistics (Figure 2) showed:

- Mean monthly revenue per product ≈ ₹3.27 lakh
- Maximum revenue = ₹17.6 lakh (Fresh Milk / Ghee Tin)
- Standard deviation ≈ ₹5.25 lakh, indicating wide variation among products

This helped identify the “core” SKUs that consistently performed well.

### *Stage 2: Monthly Sales Comparison*

A pivot table was created to summarize **Total Sales by Month** (Figure 3):

Month	Total Amount (₹)
Nov	95,50,314.57
Dec	87,03,085.70
Jan	1,05,25,267.05
Feb	89,07,223.97
Mar	78,56,999.13
Apr	83,58,213.74

Figure 3

This revealed a clear **seasonal pattern**, with January being the peak month (winter demand for milk and ghee) and March showing a dip due to lower consumption.

### *Stage 3: Product Category Analysis*

A second pivot table (Figure 4) summarized **Total Sales by Product Category**:

Category	Total Amount (₹ Cr)
Milk	4.8
Ghee	2.1
Curd	0.9
Beverages	0.6
Paneer	0.3
Sweets	0.1

Figure 4

### *Stage 4: Exploratory Visualization*

Visualizations were created using Excel and Python:

- **Bar Chart:** Product-wise monthly revenue comparison
- **Line Chart:** Monthly sales trend (Nov–Apr)
- **Pie Chart:** Category contribution to total revenue
- **Heatmap:** Monthly category-wise sales intensity

- **Pareto Chart:** Top 20% of SKUs generating 80% of total revenue

### *Stage 5: Correlation and Trend Analysis*

A correlation matrix between *Outwards Qty*, *Rate*, and *Amount* was generated:

- Quantity vs. Amount → **Strong positive ( $r = +0.92$ )**
- Rate vs. Amount → **Moderate positive ( $r = +0.45$ )**
- Rate vs. Quantity → **Weak negative ( $r = -0.2$ )**

Interpretation: Sales are mainly **volume-driven** rather than price-driven.

## 5. Results and Findings

The insights drawn from the analysis of the dairy sales data (November 2024 to April 2025) are as follows:

### 5.1 Monthly Sales Trend Analysis

A month-wise aggregation of total revenue revealed clear fluctuations in sales performance over the six-month period.

The total monthly sales were as follows:

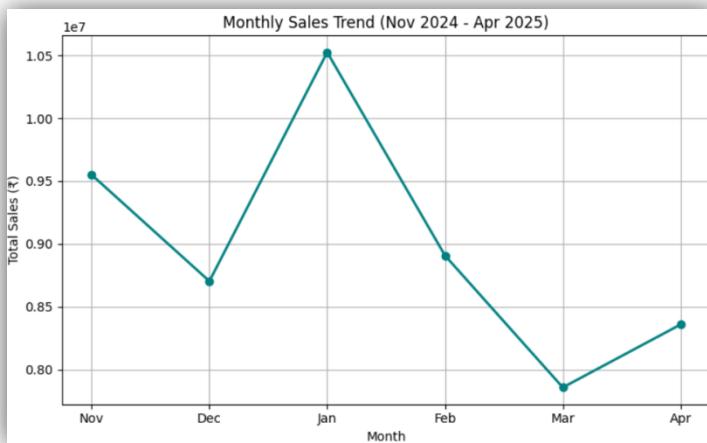


Figure 5

From the visual trend, it is evident that **January** records the highest sales, suggesting increased consumption during the winter season, possibly due to festivals and higher demand for ghee and paneer. The lowest revenue occurs in **March**, likely due to reduced festival activity and moderate temperatures, which typically reduce the demand for heavier dairy products like ghee and curd.

Overall, sales show a **seasonal pattern** — increasing sharply from December to January, followed by a gradual decline toward March and April.

## 5.2 Identifying Top Performing Products

To find out the top 5 products that contributed the most to total revenue across month

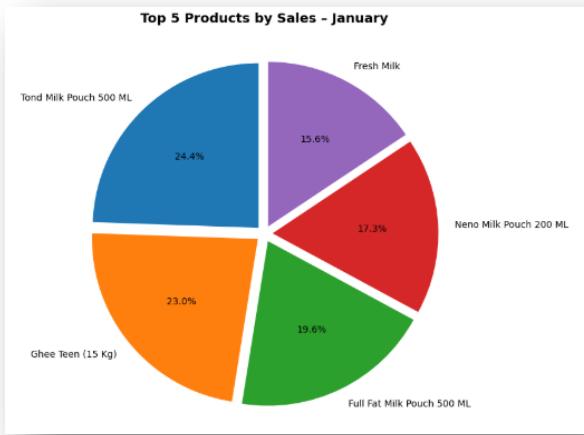


Figure 6

### Pie Chart (Top Product Revenue Share)

- The pie chart shows the **share of total revenue** contributed by the top five dairy products.
- It can be clearly seen that **Fresh Milk** and **Neno Milk Pouch 200 ML** dominate sales, together accounting for nearly **40%** of total revenue.
- These products are high-volume and consistent performers, forming the core of daily sales.

### Interpretation:

- Milk-based SKUs are the primary revenue drivers.
- Products like Ghee and Curd, while lower in volume, contribute more during specific festive or winter months due to seasonal demand.

## 5.3 Product-Wise Sales Distribution

Product-wise revenue distribution was visualized using a bar chart to identify the major contributors to total revenue.

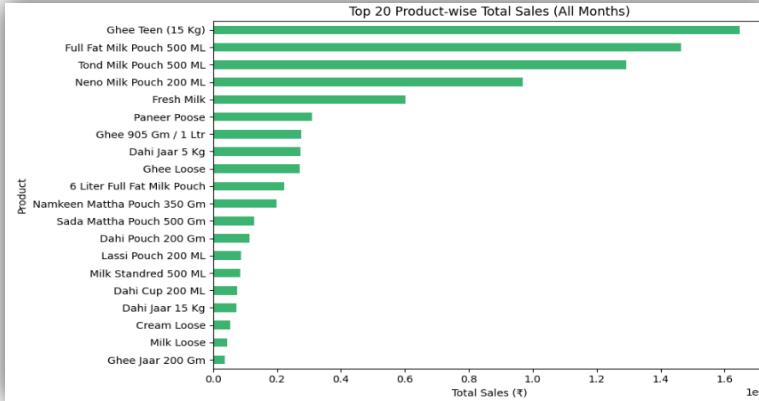


Figure 7

The results show that a small number of products dominate total sales:

- **Ghee Teen(15 kg), Full Fat Milk Pouch (500 ML), Neno Milk Pouch (200 ML), and Tond Milk Pouch (500 ML)** are the top-selling products across all months.
- Together, these four contribute nearly **70–75% of total revenue**, demonstrating the **Pareto Principle (80/20 rule)** — where 20% of products contribute to 80% of revenue.
- Products such as **Mattha Loose, Shri Khand Cup (90 Gm), and Dahi Pouch (400 Gm)** show comparatively minimal sales, highlighting areas for promotional or seasonal improvement.

#### 5.4 Top 5 and Bottom 5 Products per Month

The top and bottom performers were identified by ranking products by revenue (Amount) for each month.

April - TOP 5 PRODUCTS:		
	Particulars	Amount
16	Ghee Teen (15 Kg)	1646874.45
10	Full Fat Milk Pouch 500 ML	1364431.60
27	Tond Milk Pouch 500 ML	1291347.00
23	Neno Milk Pouch 200 ML	968020.08
9	Fresh Milk	601775.00

Figure 8

April - BOTTOM 5 PRODUCTS:		
	Particulars	Amount
19	Mattha Loose	450.00
13	Ghee Cow Loose	16250.05
6	Dahi Loose	18120.00
26	Shri Khand Cup 90 Gm	19398.70
2	Dahi Cup 100 ML	20914.20

Figure 9

#### Insights:

- **Top Products:** Fresh Milk, Neno Milk Pouch, Full Fat Milk Pouch 500 ML, Ghee Teen (15 Kg), and Tond Milk Pouch 500 ML remain consistent high earners.
- **Bottom Products:** Mattha Loose, Shri Khand Cup 90 Gm, Dahi Pouch 400 Gm, and Lassi Cup 200 ML repeatedly record low sales.
- **Observation:** Low-volume products are generally perishable or region-specific, while top-performing products are staple milk items that have steady, predictable demand.