ENHANCING OPERATIONAL EFFICIENCY AT SFP SONS

A MID-TERM REPORT FOR THE BDM CAPSTONE PROJECT

Submitted by

Name: DARSHAN GANATRA

Roll number: 22f2000931

IITM BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary

SFP Sons (India) Pvt. Ltd., founded in 2004, is a B2B fragrance manufacturer producing perfumes, attars, and deodorants for both domestic and export markets. While the company has grown with a wide product range, it struggles with key operational issues such as delays in fulfilling high-demand items, underused capacity for fast-moving products, and inefficient focus on low-performing ones.

The company provided primary data, including production, cost, and sales records, along with supporting proof like NDA. Proof of originality included factory images, dataset links and interview videos. The Meta-Data section outlines details of the datasets: Production , Cost and Sales. Descriptive statistics revealed a consistent production shortfall of an average 11,662 units/day and a skewed sales distribution.Notably, negative daily sales values (as low as ₹-4,73,905) suggest significant order fulfillment issues, likely caused by mismatches between production output and actual market demand.

Three key analysis techniques were applied to tackle these inefficiencies: Pareto Analysis, K-Means Clustering, and Monthly Trend Analysis. Pareto Analysis showed that just 10 products, such as JASS CLASSIC EDP and AHSAN ATTAR FULL EDP LIVE FRESH, contribute over 80% of the revenue. K-Means Clustering grouped products based on demand to support improved production prioritization. Monthly Trend Analysis revealed strong sales peaks in November, pointing to seasonal or promotional impacts. The objective of this project is to enhance production efficiency by identifying top-performing products and aligning production resources with actual market demand.

2. Proof Of Originality

The data used in this project has been provided by Mrs Gayathri Patel who is Marketing Executive at SFP Sons Pvt Ltd. The company was incorporated on 17th December 2004 (CIN: U24246TN2004PTC053500). The business's GSTIN is 24AAICS9235N1ZS.

2.1 Letter from the Organization:

The Non-Disclosure Agreement on the company's official letterhead, confirming access and data sharing for academic purposes, has been included

Link to the Document(NDA) - **NDA DARSHAN GANATRA.pdf**

link to the all the proofs including images, documents , Raw Dataset Links and video proofs are available here **:** Proof of originality

2.2 Conversational videos of Interaction:

Video recordings of the conversation with Mr. Dinesh S. Patel (Managing Director) and Mrs. Gayathri Patel (Marketing Executive) at SFP Sons (India) Pvt. Ltd. are attached as part of the proof of originality.

Interaction with the Founder and MD, Mr. Dinesh S. Patel: **dinesh sir sfp.mp4**

Interaction with the Marketing Executive, Mrs. Gayathri Patel : **gayathri maam sfp.mp4**

3. Meta Data and Descriptive Statistics

This section briefly outlines the Production, Cost, and Sales datasets used in the analysis, shared via email by the company's Marketing Executive in Excel Format. Since the raw data included columns that are only relevant to the business , only the relevant columns required for analysis were retained. Descriptive statistics were then used to provide a clearer view of the dataset's structure.

3.1 Metadata

3.1.1 Production Data (Link : Production (1))

The Production dataset, covering November 2024 to January 2025, provides insight into manufacturing performance by comparing planned output with actual completion, helping identify production efficiency and operational gaps across products. It consists of 6 columns and 261 rows.

Column Name	Type	Description
Order Date	Date	Date of production order placed
ItemName	String	Name of the product manufactured
Planned Quantity	Integer	Number of units scheduled for production
Completed Quantity	Integer	Actual number of units produced
DiffQty	Integer	Difference between planned and completed quantity
ML	String	Volume of the product in ML (e.g., 100ML, 8ML)

Table 3.1.1 Fields of Production Data

This dataset highlights operational inefficiencies in execution, as shortfalls between planned and actual quantities directly contribute to delayed order fulfillment. This enables better production planning and resource allocation.

3.1.2 Cost Data (Link : Cost (1))

The Cost dataset comprises cost-related information and the opening stock for each product as of the production data's start date, structured across 5 columns and 196 rows. Costs were mapped to their respective products based on a provided ML-driven cost mapping.

Column Name	Type	Description
ItemName	String	The product’s name
Category	String	The product’s classification (e.g., Perfume, Deodorant)
Opening Stock	Integer	The quantity of product at the start of the production data
ML	String	Volume or size of the product (e.g., 100ML, 8ML)
Cost	Float	The cost per unit (in ₹)for manufacturing or procurement

Table 3.1.2 Fields of Cost Data

This helps assess opening inventory and align production decisions based on stock availability, ensuring resource planning is both demand- and cost-informed.

3.1.3 Sales Data (Link : Sales)

The Sales dataset provides a detailed record of sales transactions across different regions, helping track product performance and sales trends over time. Spanning from April 2024 to February 2025, it contains 13 columns and 32,714 rows.

Column Name	Type	Description
Posting Date	Date	The date when the sales transaction was recorded
Month	String	The month of the transaction
City	String	The city where the sale occurred
State	String	The state where the city is located
ItemName	String	The name of the product sold
Quantity	Integer	The number of units sold
MRP	Float	Maximum Retail Price(₹)

Distributor Rate	Float	Price for the Distributors(₹)
Final Sale	Float	Total sale from the transaction(₹)
Branch Name	String	Branch of the unit from where the sale happened
Brand Name	String	The brand under which the product is marketed
Category	String	The product's classification (e.g., Perfume, Deodorant)
ML	String	Volume of the product in ML (e.g., 100ML, 8ML)

Table 3.1.3 Fields of Sales Data

This shows sales trends across products over the time period of 10 months .It helps identify which products are generating consistent revenue and when demand peaks occur,both essential for aligning production schedules and avoiding stockouts.

3.2 Descriptive Statistics

This section outlines key statistical measures from the Production, Cost, and Sales datasets, offering an initial understanding of data distribution and variation before deeper analysis. For this summary, only the relevant columns from each dataset were selected to ensure clarity in interpreting the data.

	Production			Cost	Sales		
Metrics	**Comp. Quantity Per Day**	**Planned Quantity Per Day**	**DiffQty Per Day**	**Cost**	**Final Sale Per Day(₹)**	**Distributor Rate(₹)**	**MRP (₹)**
Average	81,357	93,019	11,662	9.86	7,01,532.29	94.5	135.89
Standard Deviation	88,587.26	89,264.79	24,535.87	4.12	6,60,482.02	72.86	119.73
Minimum	0	289	0	4.5	-4,73,905.68	0	0
Maximum	5,55,495	5,66,436	1,41,108	18	42,50,246.21	521.81	950

Table 3.2 Descriptive statistics of relevant column

The descriptive statistics presented in Table 3.2 provide a clear overview of key metrics across production, cost, and sales data.In the Production segment, the average Planned Quantity per Day was approximately 93,019 units, while the average Completed Quantity per Day stood lower at 81,357 units, highlighting recurring shortfalls in meeting planned targets. The production volumes showed substantial variability, ranging from a minimum of just 289 units to a maximum of 5,66,436 units per day. This wide range suggests that both small and large batch sizes were common. The Difference in Quantity (DiffQty), representing the gap between planned and completed production, reached up to 1,41,108 units, with a daily average shortfall of 11,662 units. This recurring gap highlights a need for more accurate production planning, as such shortfalls can delay delivery and contribute to backorders.In the Cost data, the average unit cost was ₹9.86, with a relatively narrow spread between ₹4.50 and ₹18.00. The standard deviation of ₹4.12 implies a stable and predictable cost structure (as the cost of the product depends on the volume)which can be beneficial for pricing and margin management.

The Sales data revealed substantial variation. The average Final Sale per Day was ₹7,01,532.29, but this figure was heavily influenced by high-value days, as reflected by a large standard deviation of ₹6,60,482.02. The minimum daily sale recorded was ₹-4,73,905.68, which corresponds to registered orders that were unfulfilled. This points to misalignment between production and real-time demand. On the other end, daily sales peaked at ₹42,50,246.20, indicating the occurrence of large transactions. The Distributor Rate ranged widely from ₹0 to ₹521.81, with an average of ₹94.50, suggesting different pricing models or tiers across products. Similarly, the MRP (Maximum Retail Price) ranged from ₹0 to ₹950, with an average of ₹135.89, reflecting a diverse product portfolio catering to both mass-market and premium segments. These insights from sales distribution, production gaps, and cost variability form the baseline for applying targeted analytics to improve fulfillment and reduce operational strain.

4. Explanation of Analysis Process/Methods

4.1 Data Cleaning

The datasets collected Production, Cost, and Sales required varying levels of preprocessing before analysis. While Production and Cost data were largely clean, the Sales dataset required more extensive refinement to ensure consistency and relevance.

To support better analysis across all datasets, the product volume (ML), which was originally embedded within the product name, was extracted into a separate column. This allowed for standardized cost mapping and comparison. In the Cost dataset, unit costs were assigned using a provided ML-to-cost mapping file, ensuring accurate cost allocation for each product.

In the Sales dataset, non-relevant columns such as internal document numbers and HSN codes were removed to simplify the structure. Additionally, entries tagged as gifts or unrelated products (e.g., shampoos) were excluded to focus the analysis on fragrance-related items. Column names were also standardized for clarity—for example, “Description” was renamed to “Category” to better reflect its content.These steps were essential to maintain data quality and ensure the reliability of the results.

Here is the Google Colab Link for reference of processing and cleaning done on the data : BDM_MIDTERM.ipynb

4.2 Data Analysis

This section presents the analysis techniques used to explore operational inefficiencies at SFP Sons, particularly the challenges of managing a wide product range with limited production capacity. To address this, three targeted methods were applied.Pareto Analysis, to identify the small subset of SKUs driving the majority of sales; K-Means Clustering**,** to group products based on sales performance and production impact; and Monthly Trend Analysis, to uncover seasonal demand patterns and misalignments in planning. Each method was selected based on its relevance to the production and planning challenges observed in the data.

4.2.1 Pareto Analysis (80/20 Rule)

Pareto Analysis is a method based on the 80/20 principle, which states that roughly 80% of outcomes come from 20% of causes. In a business context, this often means that a small number of products generate the majority of revenue. This technique helps identify those key products, so that resources can be focused where they make the most impact.

- For this method , total sales were calculated for each product using the sales data.
- The products were then sorted from highest to lowest based on their contribution to revenue.
- A cumulative percentage of sales was plotted to show how many products account for 80% of total revenue.
- This created a clear visual of which SKUs are driving the business and which have minimal impact.

How it helps the problem:

SFP Sons has a wide range of products but limited production capacity.Pareto Analysis helps simplify this complexity by identifying the small group of high-impact products that deserve production priority. By focusing on these, the company can reduce overload, improve fulfillment of popular items, and use resources more efficiently which directly supports the goal of enhancing operational efficiency.

4.2.2 K-Means Clustering

K-Means Clustering is an unsupervised machine learning algorithm used to group data points into K-distinct clusters based on similarity. The goal is to minimize the distance between data points and the center of their assigned cluster. It was used to segment products based on their sales performance, allowing us to group similar-performing items together. This helped simplify the complexity of analyzing a large and diverse product portfolio by clearly highlighting high-demand, moderate-demand, and low-demand segments.

Formally, K-Means aims to minimize the **within-cluster sum of squares (WCSS)**, also known as inertia. The objective function is:

$$J = \sum_{i=1}^{K} \sum_{x \in C_i} \|x - \mu_i\|^2$$

Where:

- K = number of clusters
- C_i = set of points in cluster i
- μ_i = centroid (mean) of cluster i
- $\|x - \mu_i\|^2$ = squared Euclidean distance between a point x and its cluster centroid

The sales data was grouped by product, and total revenue and quantity sold were calculated for each item. These two features served as inputs for clustering.

The algorithm works iteratively by:

1. Initializing K centroids.
2. Assigning each data point to the nearest centroid using Euclidean distance
3. Recalculating the centroids as the mean of all points assigned to each cluster
4. Repeating steps 2 and 3 until convergence (i.e., cluster assignments no longer change significantly).

To determine the optimal number of clusters K, we used:

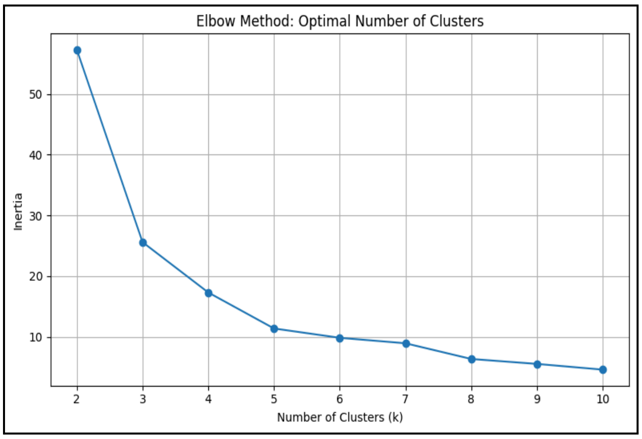
- Elbow Method (Plots the WCSS against different values of K. The 'elbow point' is where the decrease in WCSS slows down, suggesting the best K.
- Silhouette Score: Measures how similar a point is to its own cluster versus other clusters. The formula is:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Where:

- $a(i)$ = average distance between point i and all other points in the **same cluster**
- $b(i)$ = average distance between point i and points in the **nearest neighboring cluster**

Both methods suggested that three clusters (K=3) provided a meaningful balance of separation and similarity(Figure 4.2.2.1 and Figure 4.2.2.2).



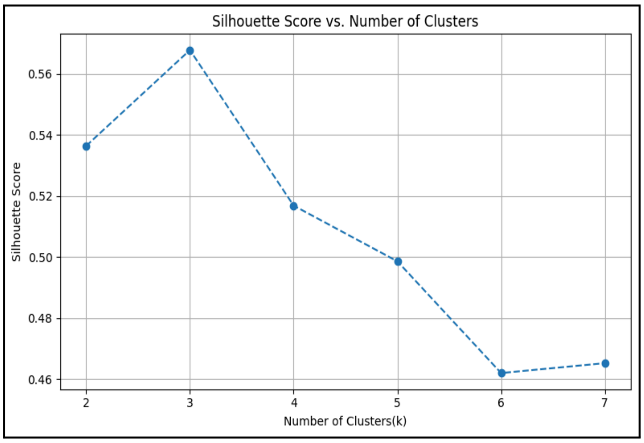


Figure 4.2.2.1 Elbow method

Figure 4.2.2.2 Silhouette Score

K-Means was then applied to form three groups of products typically representing high, medium, and low performers. The top products in each cluster were reviewed to understand the characteristics of each group and to explore how product performance is distributed across the catalog.

How it helps the problem:

SFP Sons faces challenges due to the size and complexity of its product range. K-Means Clustering makes it easier to segment products into clear priority levels. This helps in making focused production decisions by identifying which products to scale, monitor, or possibly reduce. It supports better use of limited resources and directly contributes to streamlining operations and reducing inefficiencies.

4.2.3 Monthly Sales Trend Analysis

Monthly Sales Trend Analysis is used to understand how sales change over time. It helps identify which months have high demand and which have low, allowing the company to plan production more effectively.

- In this project, the sales data was grouped by month using the Posting Date and the total revenue (Final Sale) for each month was calculated.
- A line chart was then used to plot monthly sales, revealing clear patterns in how demand shifts throughout the year. This helped highlight seasonal peaks, dips, and overall sales fluctuations over the 10-month period (February was excluded due to partial data).

How it helps the problem:

One of the challenges at SFP Sons is that production doesn't always align with real customer demand. High demand and low stock leads to missed sales opportunities. When demand is low and production continues as usual, resources are wasted. This time-based analysis gives the company a clear view of when to scale up production and when to slow it down,making operations more responsive, reducing overproduction, and improving fulfillment during peak sales periods.

5. Results and Insights

5.1 Pareto Chart Analysis

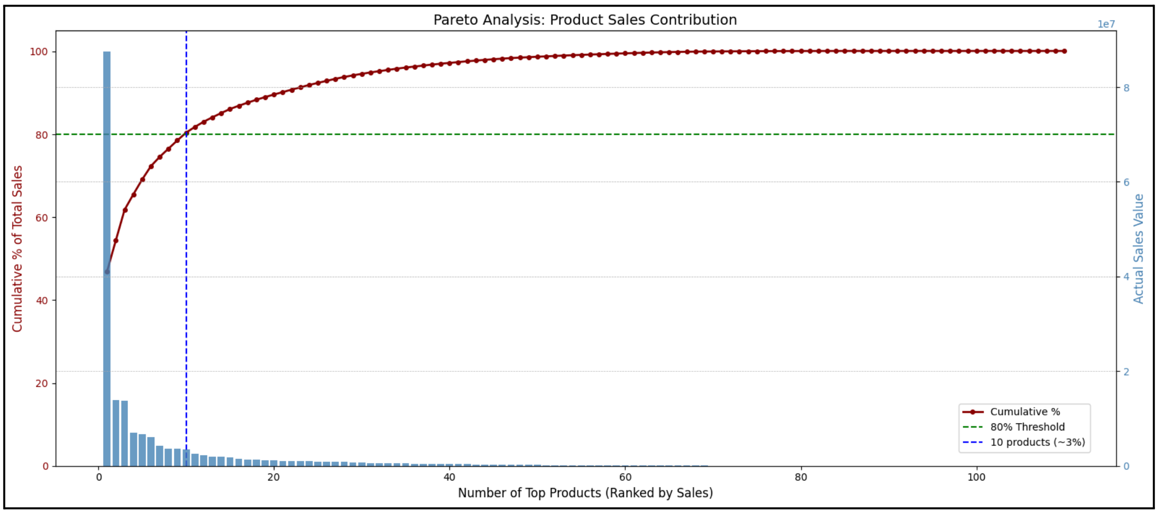


Figure 5.1 Pareto Analysis

As presented in analysis Section 4.2.1, a Pareto analysis revealed that out of more than 400 products sold by SFP Sons, just 10 products, around 3% of the entire range, account for nearly 80% of total sales revenue. This indicates that sales are highly concentrated in a small portion of the catalog. The top-selling product, JASS CLASSIC EDP, alone contributes around ₹87 million, while AHSAN ATTAR FULL EDP LIVE FRESH and JASS CLASSIC DEODORANT each generate around ₹13.9 million and ₹13.8 million respectively. Other products like CRAZY MOMENTS PINK EDP, AHSAN BLACK MAGNET EDP, and JASS NUMBER ONE DEODORANT also show strong performance, earning between ₹6 to ₹7 million each. These products lie in the steep section of the cumulative sales curve, showing their high contribution to overall revenue. In contrast, the majority of the remaining products generate less than ₹5 million each, with many contributing very little. This pattern suggests that equal focus on all SKUs may lead to wasted effort and resources.

5.2 K-Means Clustering

As presented in section 4.2.2, with K-Means clustering (Figure 5.2) products are grouped into three distinct clusters:

- Low Demand (Green): Products in this cluster have low sales volumes and low revenue. An example is AHSAN ATTAR FULL EDP SMALL BOTTLE.
- Moderate Demand (Orange): These products fall in the mid-range of both sales and revenue. CRAZY MOMENT'S BLACK NEW EDP (LOCAL) is an example from this group.
- High Demand (Blue): Items in this cluster show high values for both sales quantity and revenue. One such product is AHSAN ATTAR FULL EDP LIVE FRESH.

- Top Performer (Yellow Cross Marker): JASS CLASSIC EDP lies outside the three main clusters, exhibiting exceptionally high values in both quantity sold and total revenue. Its separation from the clusters marks it as a clear performance outlier within the product range.

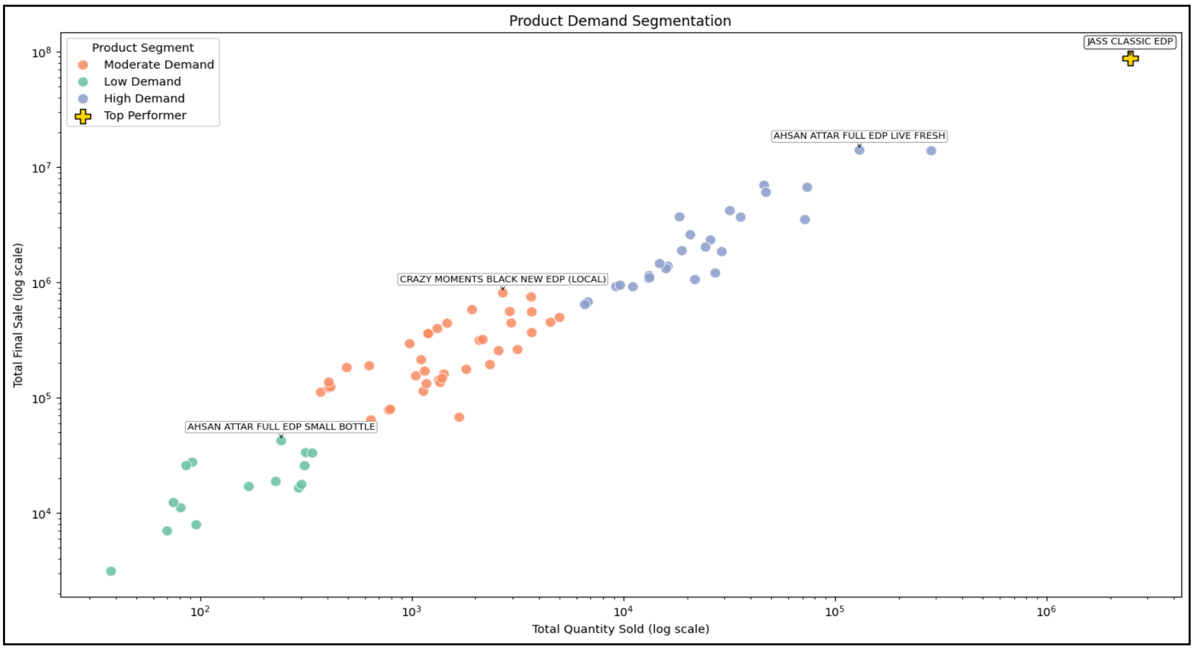


Figure 5.2 K-means clustering

5.3 Monthly Trend Analysis

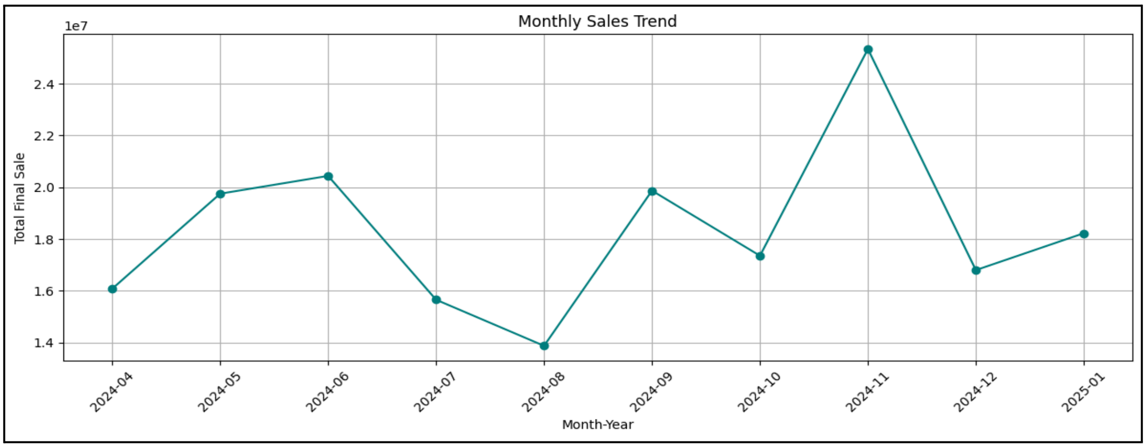


Figure 5.3 Monthly Trend Analysis

As presented in analysis section 4.2.3, the monthly trend chart in Figure 5.3 shows how total sales varied over time during the observed period (February was due to partial data).The graph highlights clear fluctuations, with noticeable peaks in June, September, and November, and a drop in August and December.The spike in November marks the highest sales month, suggesting strong seasonal or promotional influence.