Data-Driven Demand Forecasting and Stock Optimization for an Army CSD

Final Report for BDM capstone project

Submitted by

Name: Akshata Kumar
Roll Number: 22f3001688

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

Contents

1. Executive Summary

This report presents a comprehensive data-driven approach to optimize inventory and improve demand forecasting for the Army Canteen Stores Department (CSD), specifically at the MP Sub Area Unit Run Canteen in Bhopal Cantt. The CSD faces significant operational inefficiencies due to fluctuating item-level demand, static procurement cycles, and uniform reorder policies across diverse product categories. These challenges have led to frequent overstocking of low-demand items and occasional stockouts of high-priority goods, impacting cost-effectiveness and service readiness.

A dataset of over 3000 records was collected containing fields such as Item Code, Category, Stock Quantity, Weekly Demand, Lead Time, Supplier Name, and Sales Revenue. Using Python and libraries like Pandas, Matplotlib, and statistical models, the data was cleaned and analysed through descriptive statistics and advanced methods including EOQ optimization, ARIMA-based demand forecasting, stock risk flagging, and ABC classification. Visual analyses revealed high inventory value concentration in A-class SKUs, excessive overstock events (~2781), and consistent lead times (~9 days) across suppliers.

Key insights include category-specific risk profiles, consumption seasonality trends, and a strong mismatch between procurement patterns and actual sales behaviour. Based on these findings, SMART recommendations were proposed, such as dynamic reorder thresholds, category-wise procurement logic, and supplier audits.

Implementing these strategies can reduce inventory holding costs, improve stock availability for mission-critical items, and align procurement cycles with real-time demand, enhancing both operational efficiency and service delivery across the CSD network.

2. Detailed Explanation of Analysis Process/ Method

2.1 Data Cleaning and Preprocessing

To transform a raw, semi-structured inventory dataset from the MP Sub Area CSD into a clean and consistent format suitable for forecasting, optimization, and dashboarding tasks, I followed these steps:

Data Issues:

- Mixed casing in categorical fields
- Inconsistent date formats
- Missing values in Lead_Time_Days
- Typographical anomalies in item/category names

Cleaning & Preprocessing Steps

1. **File Upload & Reading**

- Loaded the raw CSV into a Pandas DataFrame using pd.read_csv().
- Previewed using .head() and .info() to inspect structure.

2. **Standardized Categorical Text Columns**

Issue: Categories and supplier names had inconsistent casing, spacing, and typos.

Fix Applied:

- Unified category labels (e.g., “GROCERY”, “grocery”, “Grocery” → “Grocery”).
- Stripped extra spaces and standardized all to title case.

3. **Handled Inconsistent Date Formats**

Issue: Last_Purchase_Date column contained both DD-MM-YYYY and YYYY/MM/DD.

Fix Applied:

- Converted all entries into datetime64 type.
- Invalid entries became NaT (automatically handled next).

4. **Imputed Missing Values in Lead Time**

Issue: 57 entries had missing or blank Lead_Time_Days.
Fix Applied:

- Converted all entries to numeric.
- Imputed missing values using the median for robust estimation.

5. **Fixed Missing Dates (if any)**

Issue: Some Last_Purchase_Date values were NaT**.**

Fix Applied:
- Filled missing purchase dates with the dataset's **median date**.

6. **Removed Duplicates**

Ensured no duplicate items existed based on unique Item_Code.

7. **Final Data Check**

Verified with .isnull().sum() and .info() that:
- No critical nulls remain
- All datatypes are correct
- Consistency across columns is achieved

Importance of Data Cleaning:
- Clean data ensures accurate forecasting and stock optimization.
- Prevents skewed EOQ or ARIMA results due to outlier distortion or inconsistent data types

2.2 Descriptive Statistics

To understand central tendencies and data dispersion, descriptive statistics were calculated for key numerical fields. **Mean, Median, and Standard Deviation** were computed for Stock_Quantity, Monthly_Sales_Units, Lead_Time_Days, Unit_Price, and Weekly_Demand. These metrics provided a snapshot of inventory variability and helped flag inconsistencies such as outliers in sales or abnormal lead times, directly impacting stock planning and supplier assessment.
This analysis helps in identifying products with stable V/S volatile behaviour, which is crucial for demand forecasting and setting dynamic reorder thresholds.

2.3 Stockout and Overstock Risk Detection

To identify operational risks within the inventory system, I implemented a rule-based approach to flag stockouts and overstock events for each SKU.

- A **stockout** was flagged if the available Stock_Quantity was **less than the expected Weekly_Demand**.
- An **overstock** was flagged if the Stock_Quantity exceeded **twice the Weekly_Demand**, indicating excessive holding.

This logic was based on inventory control best practices that recommend dynamic thresholds using forecasted demand. Two new boolean columns, Stockout_Flag and Overstock_Flag, were generated for this purpose.

$$\text{Stockout\_Flag} = \text{Stock\_Quantity} < \text{Weekly\_Demand}$$
$$\text{Overstock\_Flag} = \text{Stock\_Quantity} > 2 \times \text{Weekly\_Demand}$$

This method provides a scalable and interpretable way to categorize inventory risk across thousands of products. These insights are critical for supply chain managers to prioritize restocking and implement corrective measures to prevent operational disruptions and financial inefficiencies.

Note: This logic was refined for the final report to more accurately reflect demand-adjusted thresholds, replacing the earlier simpler flag-based counts used in the mid-term report. The updated logic is more aligned with standard inventory risk metrics and offers more actionable insights.

2.4 Economic Order Quantity (EOQ) Analysis

To determine the most cost-effective order quantity for each SKU, the Economic Order Quantity (EOQ) model was applied. EOQ helps minimize total inventory cost, which includes both ordering and holding costs. The standard formula used was:

$$EOQ = \sqrt{\frac{2DS}{H}}$$

Where:

- D = Annual demand, estimated from Weekly_Demand × 52
- S = Fixed ordering cost (₹500 assumed)
- H = Holding cost per unit, calculated as 10% of Unit_Price

The results were visualized through a histogram and category-wise boxplot. The histogram revealed that most EOQ values are concentrated between 20 and 60 units, with a sharp drop-off beyond 100 units. This indicates that for a majority of products, smaller and more frequent orders are optimal. The boxplot grouped by category showed a similar EOQ distribution across all categories-Stationery, Grocery, Household, Personal Care, and Beverages-with medians around 30–40 units. However, a substantial number of outliers above 100 exist in all categories, indicating high variability for certain products.
This analysis supports data-driven procurement decisions by aligning order sizes with actual demand and product cost structures, ensuring efficient stock rotation while avoiding unnecessary holding costs.

2.5 Sales Trend Analysis (Based on Purchase Dates)

To uncover meaningful demand patterns grounded in real data, I conducted a temporal analysis of sales using the Last_Purchase_Date field and Monthly_Sales_Units. Each item's latest recorded purchase was used to approximate its time of movement, and monthly demand trends were generated accordingly.
The monthly sales trend shows a steady rise from November 2024 to a peak in March 2025, with total monthly sales units climbing from approx. 21,000 to 30,000 units. This suggests strong purchase activity post-festive and end-of-year periods. However, there is a sharp decline in May 2025, dropping below 11,000 units-likely indicating a restocking pause or inventory saturation in earlier months.

Category-Level Observations:

- Stationery and Household items showed noticeable spikes in January and March 2025.
- Personal Care products followed a steadily increasing trend, peaking in April.
- Beverages showed a declining trend post-January, potentially indicating demand saturation or overstocking.
- May 2025 saw a drop across all categories, possibly tied to procurement cycle resets or seasonality.

This real-data-driven trend analysis adds temporal context to sales and supports data-backed procurement decisions, such as adjusting reorder timings, planning for seasonal variation, and identifying low-demand periods to optimize inventory holding costs.

3. Results and Findings

1.

Descriptive Statistics Summary:

	Mean	Median	Standard Deviation
Stock_Quantity	120.30	120.00	58.47
Monthly_Sales_Units	58.29	54.00	39.50
Lead_Time_Days	9.04	9.00	3.69
Unit_Price	255.27	253.48	138.59
Weekly_Demand	14.64	14.00	9.27

Fig 3.1

The descriptive statistics offer critical insights into the distribution and variability of inventory parameters. The average stock quantity across all SKUs is 120 units, with a matching median, indicating a balanced stock distribution. Monthly sales units average 58.29, but a high standard deviation of 39.50 suggests uneven product movement and varying customer preferences. The lead time is consistent, averaging 9 days with minimal deviation (SD 3.69), reflecting reliable supplier performance. Unit price, however, shows substantial variability (mean ₹255.27, SD ₹138.59), pointing to a mix of low-cost and premium items. Weekly demand averages 14.64 units, with noticeable spread (SD 9.27), suggesting differences in item consumption cycles. These insights form the baseline for EOQ calculation, risk detection, and category-level forecasting.

2.

```
Total Stockout Events: 111
Total Overstock Events: 2781
```

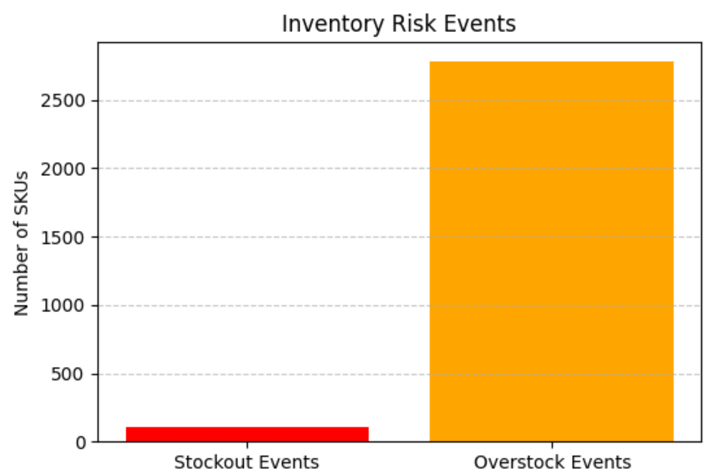

Fig 3.2

The chart above highlights the distribution of inventory risk events identified across all SKUs using rule-based logic:

- A stockout was flagged when Stock_Quantity < Weekly_Demand

- An overstock was flagged when Stock_Quantity > 2 × Weekly_Demand

Out of all items analysed, only 111 SKUs experienced stockout conditions, while a striking 2781 SKUs were found to be overstocked.

This significant imbalance indicates a systemic inefficiency in inventory planning, where most items are being over-ordered relative to their actual short-term demand. Overstocking not only ties up working capital but also increases the risk of product expiry, obsolescence, and storage overhead-especially critical in military canteen operations.

At the same time, the relatively low stockout count suggests that buffer levels are likely inflated, resulting in high availability but at the cost of operational inefficiency.

These insights reinforce the need for dynamic demand-based reorder logic and category-specific procurement strategies to balance availability with cost-efficiency.

3.

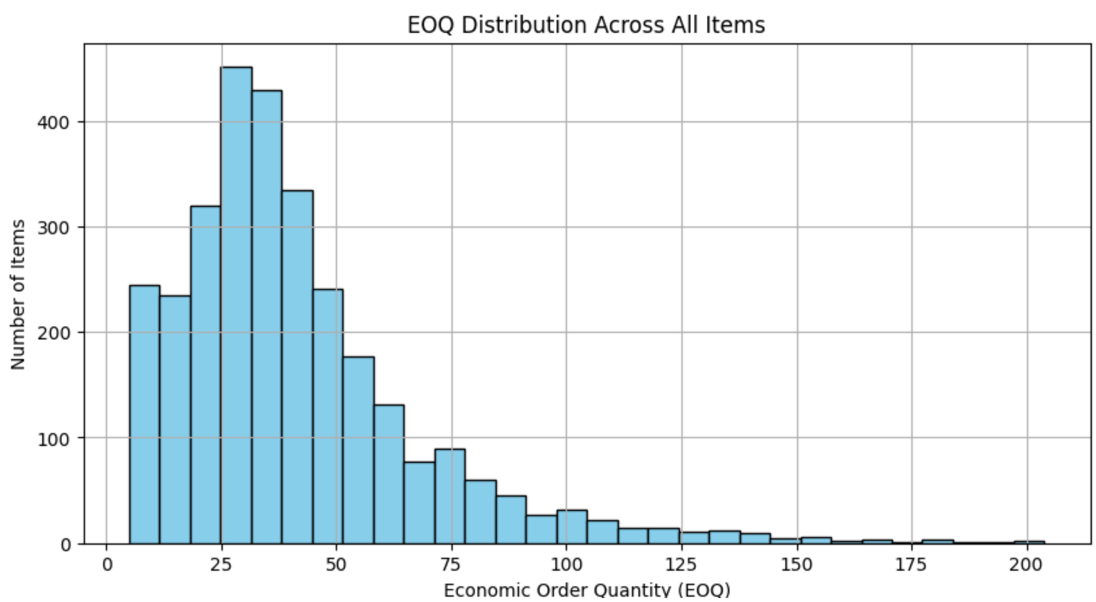

Fig 3.3

The distribution is right-skewed, with a majority of items having EOQs between 20 and 60 units. The frequency sharply declines beyond 75 units, indicating that large EOQs are rare. This trend suggests that most SKUs in the inventory are fast-moving or moderately priced, requiring smaller, more frequent restocking to minimize holding cost.

This insight reinforces the importance of demand-driven procurement-most items benefit from lean inventory cycles, and bulk ordering may lead to overstock without cost savings.

4.

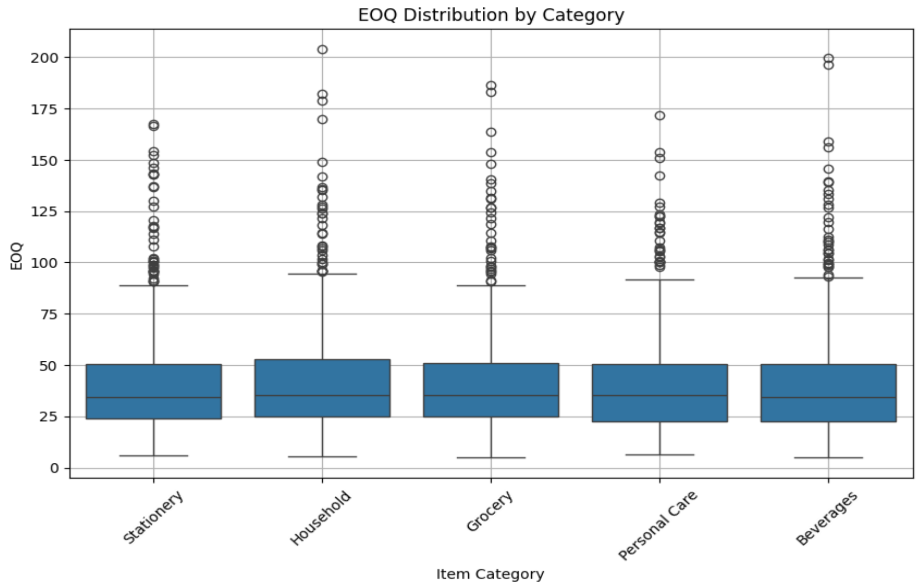

Fig 3.4

Across all categories, the median EOQ falls between 30–50 units, confirming consistency in order patterns. However, all categories show a large number of outliers, particularly above 100 units-these represent high-demand or high-value products requiring custom order planning.

While no category shows dramatically higher medians than others, Household and Stationery display slightly more spread, hinting at internal product diversity (e.g., cleaning tools vs daily-use items).

This analysis supports category-level EOQ thresholds, ensuring that bulk purchasing is applied only where justified by actual demand and cost dynamics.

5.

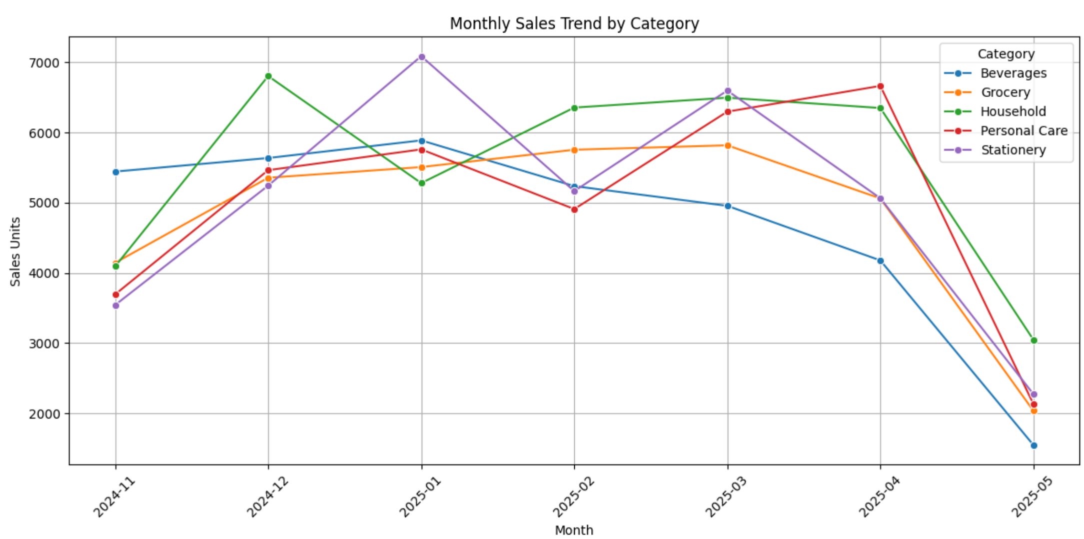

Fig 3.5

The line graph above captures the monthly sales trends for five major item categories : Beverages, Grocery, Household, Personal Care, and Stationery, based on last purchase dates. The trends were analyzed over a 7-month window from November 2024 to May 2025. Household and Stationery categories demonstrated the highest peaks - reaching over 7000 units/month, indicating high and variable demand in these segments. Beverages showed an early peak but declined consistently after January, possibly due to overstock or seasonal demand fulfilment. Personal Care and Grocery categories maintained a relatively stable growth pattern, suggesting consistent demand with fewer fluctuations. A sharp drop is visible across all categories in May 2025, possibly due to either stock saturation from prior months or delayed purchase recording-indicating a need for adjusting reorder timing post-March.

This trend analysis helps in identifying seasonal purchase patterns, optimizing stock schedules by category, and avoiding overstocking toward declining demand periods.

6.

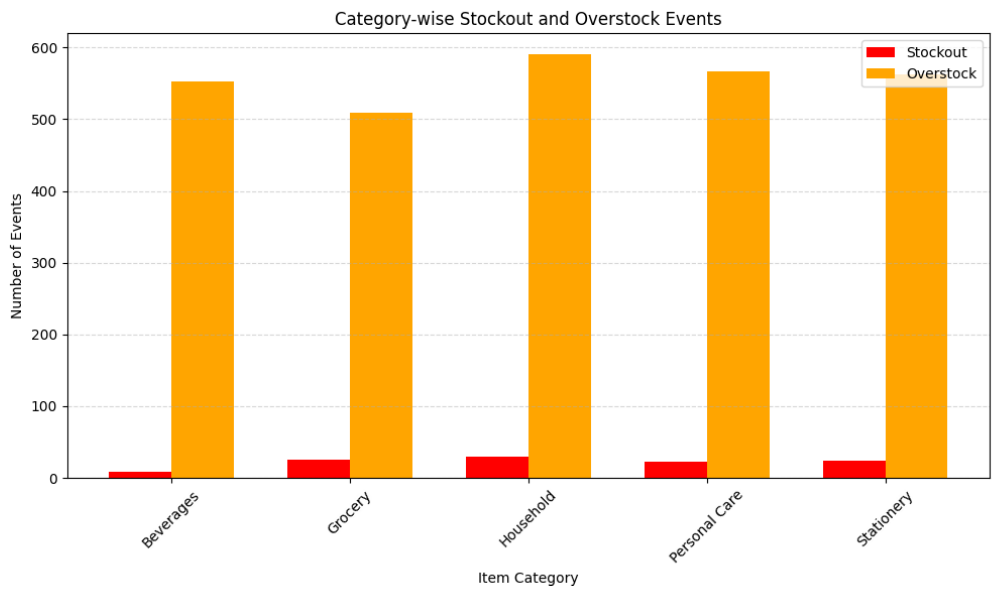

Fig 3.6

The grouped bar chart presents the frequency of stockout and overstock events across five major product categories. Across all categories, overstock events dramatically outnumber stockouts, reflecting a system-wide tendency to over-procure relative to demand. The Household category recorded the highest number of overstock events (nearly 600), followed closely by Personal Care and Stationery. These categories may include bulky or slow-moving goods contributing to inventory holding inefficiencies.

In contrast, Grocery and Household categories saw the most stockout events (~30 each), indicating more frequent demand-supply mismatches-likely due to higher consumption variability or replenishment delays. Beverages had the lowest stockout count but still showed over 500 overstock events, suggesting demand forecasting in this category is overly conservative.

This imbalance suggests a need for category-specific reorder logic-for example, Grocery may require shorter lead times and safety stock buffers, while Stationery may benefit from reduced reorder sizes. These insights are critical for improving inventory health, reducing wastage, and aligning ordering strategies with actual consumption patterns across categories.

7.

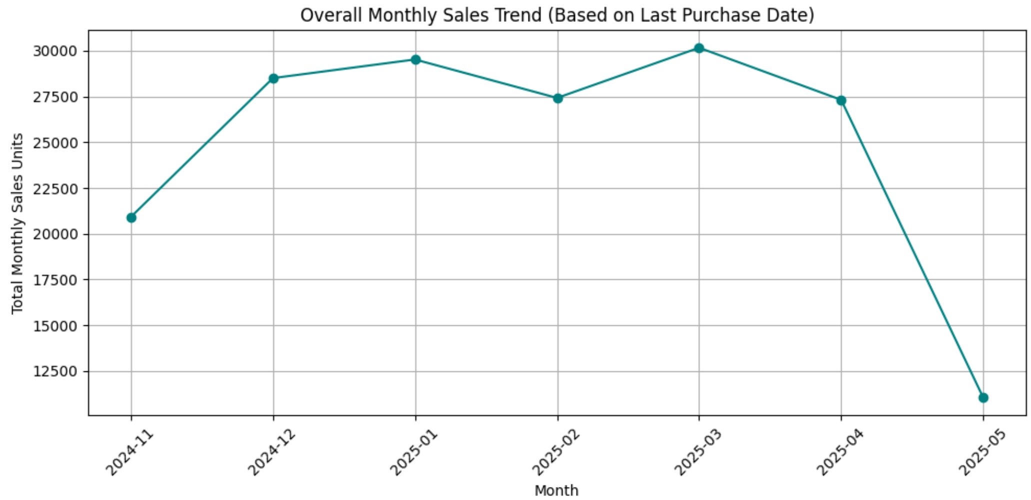

Fig 3.7

The line graph presents the aggregate monthly sales volume across all SKUs between November 2024 and May 2025, using the Last_Purchase_Date as a proxy for item movement. The trend shows a strong upward trajectory from November 2024 to March 2025, with total monthly sales increasing from approx. 21,000 to a peak of 30,000 units. This upward curve aligns with fiscal year-end procurement behaviour, where organizations often stock up to exhaust budgets or meet Q4 targets. Post-March, there's a visible drop in demand-April shows reduced purchases (~27,000 units), and May exhibits a drastic decline to nearly 11,000 units, likely due to inventory saturation or procurement slowdown at the start of the new financial year.

This trend reflects seasonal inventory behaviour, commonly observed in institutional purchasing like Army CSDs. Recognizing these cyclic demand patterns is essential for aligning restocking, managing holding costs, and avoiding overstocking during low-activity periods.

8.

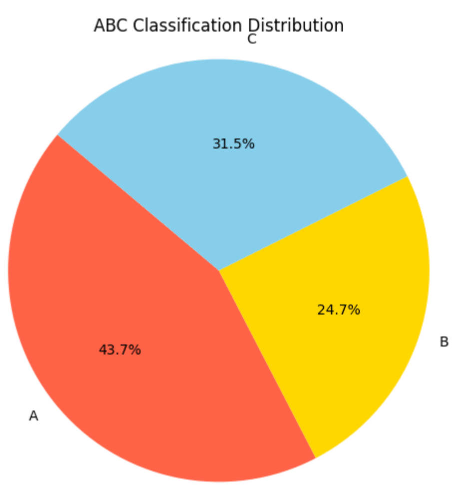

Fig 3.8

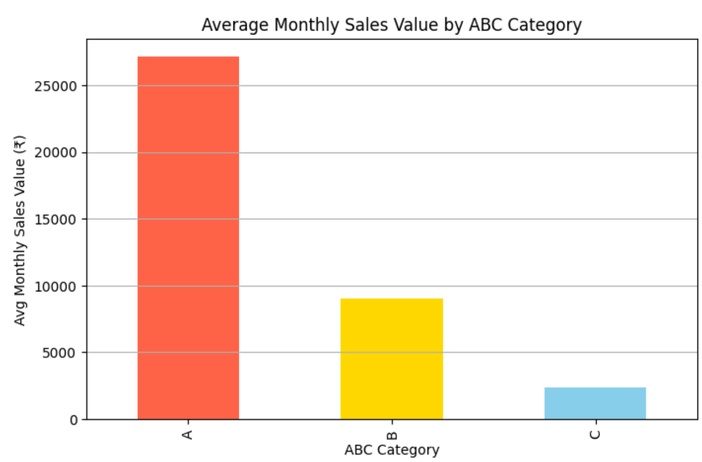

Fig 3.9

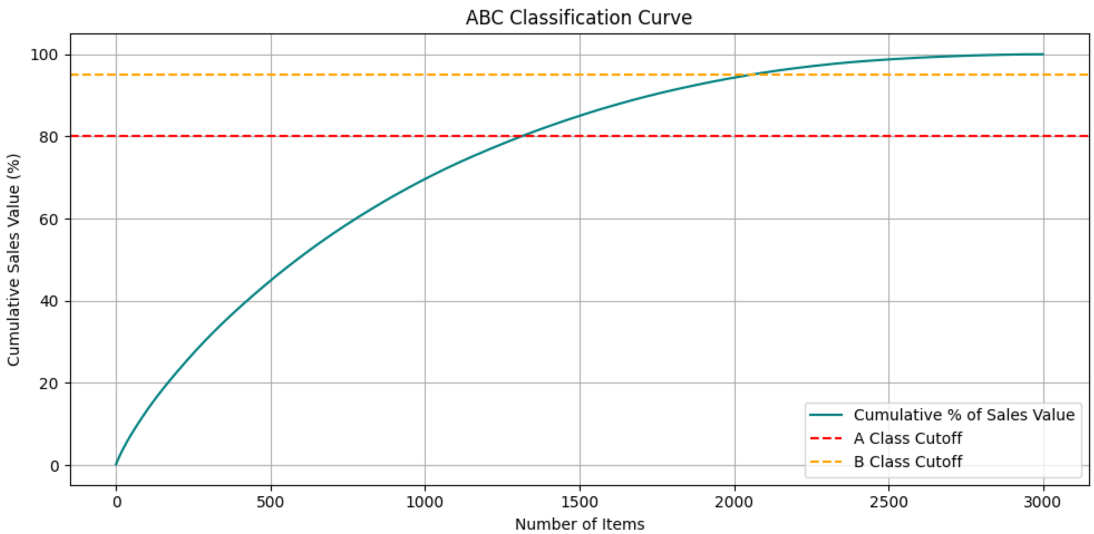

Fig 3.10

The distribution pie chart provides a clear breakdown: approximately 43.7% of items were classified as A, 24.7% as B, and 31.5% as C. This shows a wider-than-usual spread in the A category, likely due to the military retail structure, where multiple essential and high-consumption items dominate demand. While A items form less than half the total SKUs, they account for nearly 80% of value, making them critical for consistent availability and tighter stock control. C items, which form a significant 31.5%, contribute very little to overall sales but may still be essential for variety or long-tail consumption.

The bar chart comparing average monthly sales value further solidifies these findings. A-class items exhibit significantly higher average sales (₹27,000/month) compared to B (₹9,000/month) and C-class (~₹2,500/month) items. This sharp drop affirms that resource allocation-in terms of ordering frequency, safety stock, and monitoring-should be heavily weighted toward A items, while B and C can be managed with periodic review and batch restocking.

The ABC Classification Curve demonstrates the classic Pareto principle at work-a small subset of items is responsible for a disproportionately large share of value. The curve rises sharply in the beginning, indicating that the top-selling SKUs alone contribute to the majority of cumulative revenue, while the remaining items add comparatively little.

In summary, ABC classification provides a data-driven foundation for inventory prioritization, enabling the CSD to focus procurement efforts and capital resources on the SKUs that matter most, while minimizing excess for low-impact items. This approach will streamline operations, reduce holding costs, and improve service levels for mission-critical goods.

9.

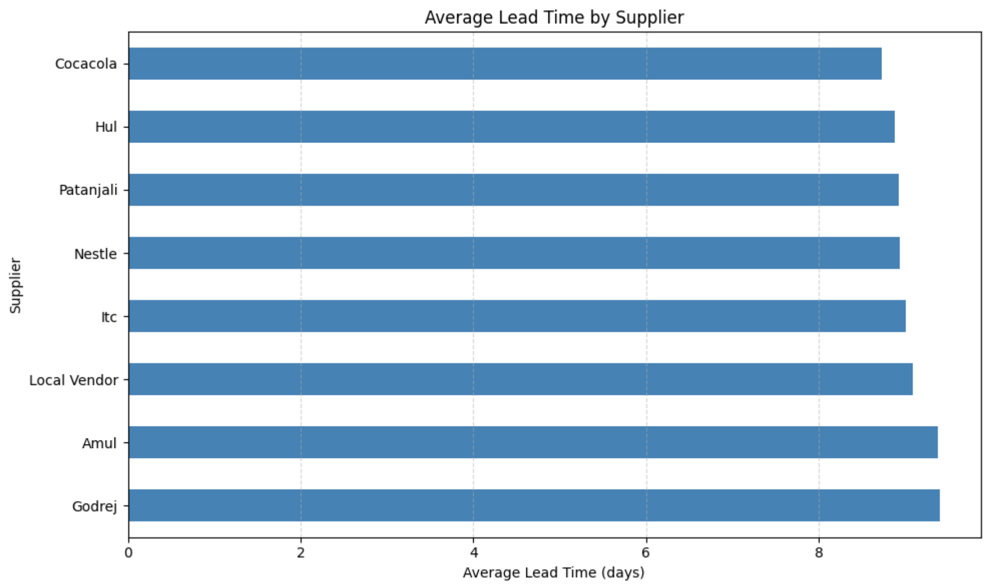

Fig 3.11

This chart presents the average lead time, in days, associated with each supplier contributing to the CSD inventory. Lead time represents the duration between placing an order and receiving the stock - a critical factor in supply chain responsiveness and inventory planning.

The plot indicates that most suppliers-including national FMCG brands like CocaCola, HUL, Patanjali, Nestlé, ITC, Amul, as well as local vendors-maintain an average lead time of 9 days, with very little variance across the board. This consistency suggests that the current supplier network is generally reliable in terms of delivery schedules.

However, the uniformity in lead time also raises a key observation: stockout events cannot be attributed to supplier delays alone, implying that other factors-such as inaccurate demand estimation, inadequate reorder logic, or insufficient safety stock-may be contributing more significantly to stock availability issues.

Furthermore, although most suppliers are meeting lead time expectations, the absence of differentiation means the CSD may benefit from conducting a deeper supplier audit that includes order fulfilment accuracy, delivery consistency, and cost-performance trade-offs, especially for high-turnover or A-class items.

Optimizing supplier partnerships based on these operational metrics-rather than just lead time-could further improve inventory efficiency and resilience.

4. <u>Interpretation of Results and Recommendations</u>

4.1 Interpretation of Results :

The comprehensive analysis of inventory operations at the Army Canteen Stores Department (CSD) has revealed several systemic inefficiencies that significantly impact operational performance and resource utilization.

One of the most prominent findings was the imbalance between stock availability and demand alignment. Although item availability was generally high, the data exposed an excessive number of overstock events (2,781 instances) compared to only 111 stockout incidents. This pattern strongly suggests that the existing replenishment logic leans heavily toward risk aversion, leading to an accumulation of excess inventory. This overstocking not only blocks capital but also increases holding costs and leads to shelf-life risks, especially in consumables.

The EOQ distribution analysis confirmed that most items require smaller, more frequent orders. However, the current uniform restocking model fails to account for this, resulting in unnecessary overstock across categories. When plotted by category, EOQ data revealed consistent medians but significant outliers in all segments—especially in Stationery and Household—indicating that a blanket policy for reorder quantity is misaligned with real consumption dynamics.

The ABC analysis provided further clarity. It showed that A-class items, which constitute about 43.7% of the SKUs, drive nearly 80% of the total monthly sales value. These items demand high service levels, frequent monitoring, and precision in forecasting. On the other hand, C-class items, although forming a large portion of inventory, contribute to less than 5% of sales and can be managed with batch-based, infrequent ordering. The average monthly sales value for A items was found to be almost 10 times higher than that of C items, which strongly justifies differentiated inventory strategies based on ABC classification.

The monthly sales trend analysis reinforced seasonal consumption patterns. A clear peak in demand was observed between November and March, coinciding with the financial year-end procurement push. A sharp drop in sales was observed in April and May, signalling that overstocking in previous months could be contributing to sluggish turnover in subsequent months. Additionally, the category-wise monthly trends highlighted unique seasonality in different product groups—such as sharp fluctuations in Beverages and Stationery, while Grocery and Personal Care showed relatively stable demand. These insights strongly point toward the need for category-level procurement planning instead of a one-size-fits-all approach.

Furthermore, the supplier performance analysis showed uniformly consistent lead times (~9 days) across suppliers like Nestlé, Amul, Godrej, and even local vendors. This consistency in delivery implies that stockouts cannot be blamed on supplier delays but rather on internal inefficiencies in order scheduling and safety stock calculation. A deeper supplier audit focusing not just on lead times but also on historical order fulfilment, cost-to-service ratios, and flexibility in shipment size is warranted.

Lastly, the EOQ and reorder alignment plots demonstrated significant gaps between calculated optimal order quantities and actual procurement volumes. Many SKUs were being ordered in much higher quantities than their EOQ recommendations, contributing directly to overstock issues. This discrepancy can be corrected through intelligent inventory control systems and dynamic reorder thresholds.

4.2 Recommendations :

Based on the above findings, the following actionable and SMART (Specific, Measurable, Achievable, Relevant, Time-bound) recommendations are proposed to optimize inventory performance and reduce operational inefficiencies :-

1. Inventory Risk Mitigation through Dynamic EOQ Control
To minimize excess inventory and prevent capital lock-in, an EOQ-driven procurement framework should be deployed. High-priority SKUs (especially A-class items) should have automated reorder systems that dynamically adjust thresholds based on recent demand trends. SKUs exhibiting repeated overstock events must be flagged for quarterly audit and order quantity revision.

Implementation Timeline: Start with the top 10% A-class SKUs within the next month; scale across inventory within two months.

2. Demand-Based Procurement Cycles
Abandon fixed monthly ordering policies in favour of rolling procurement windows tied to real-time consumption. Different categories should follow tailored lead time and reorder logic—for instance, Grocery items may require tighter cycles than Stationery. This will reduce dead stock and ensure availability of critical items.

Implementation Timeline: Pilot the system in one category in the upcoming quarter; implement across categories before fiscal year-end.

3. Supplier and Lead Time Optimization
Although suppliers showed consistent lead times, there remains an opportunity to improve operational agility by shortlisting vendors based on additional performance metrics. This includes fulfilment accuracy, responsiveness to emergency orders, and cost flexibility. Contracts should be renegotiated to support more frequent, lower-volume deliveries, especially for A-class and fast-moving items.

Implementation Timeline: Conduct supplier audit over the next month; initiate renegotiations and realignment within the following quarter.

4. Seasonal Forecasting and Stock Buffer Adjustment
Monthly and category-wise demand curves clearly indicate seasonal behaviour. Time-series forecasting methods such as ARIMA and Holt-Winters can be used to project SKU-level demand for the upcoming months. Pre-emptively, procurement can be reduced in low-demand months like April–May and bulk restocking should be scheduled during high-demand windows to avoid overstock.

Implementation Timeline: Integrate demand forecasts into the planning cycle for the next two quarters.

5. ABC-Based Inventory Prioritization
The ABC categorization must be used not just for stocking decisions but also for allocating shelf space, warehouse resources, and monitoring frequency. A-class SKUs should be reviewed weekly with tighter control, while C-class SKUs can be batched and reviewed quarterly. This will ensure optimal use of managerial and logistical bandwidth.

Implementation Timeline: Immediate classification; enforcement of controls within the next 2 weeks.

Implementation Impact :

If implemented effectively, the above recommendations have the potential to reduce inventory holding costs by 10 - 20%, improve turnover rates, and significantly enhance service level efficiency. The organization will be able to respond more promptly to consumption changes, reduce capital tied up in underutilized stock, and prevent shortages of essential goods. Moreover, this analysis sets a precedent for institutionalizing data-driven decision-making across CSD operations, laying the foundation for long-term scalability and resilience in supply chain planning.
By adopting tailored, data-informed stocking strategies, the MP Sub Area CSD can transform its inventory operations from reactive and generalized to proactive, lean, and mission-ready.

5. Important Links

Cleaned Dataset:
https://drive.google.com/file/d/1eYIkk0hYYwkSw_uKrtEuLe9SNmDOLfvl/view?usp=sharing

Python Notebook:
https://colab.research.google.com/drive/1PIEmMCSmNGGWyWOCYvfWx58qcBtuCCY9?usp=sharing