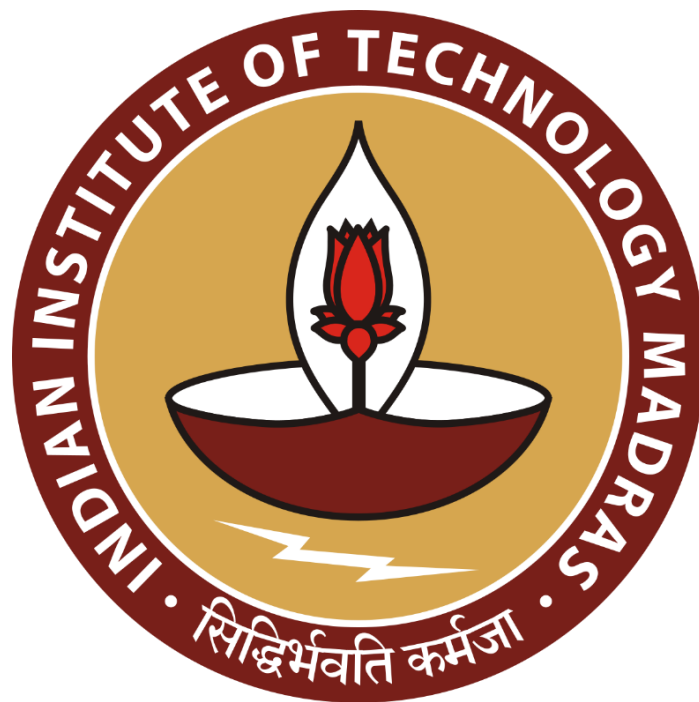# Business Data Management

Capstone Project Mid-term

Data-Driven Demand and Inventory Planning for a Rural Grocery Shop

Submitted by

Name : Sk Noor Alam
Roll No. : 23f1000205

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. Executive Summary:

Mallick Bhander is a small, family-run grocery store located in the village of Khoriop, Howrah, West Bengal. The store serves a small number of locals, most of whom earn through daily wage labour, leading to inconsistent shopping patterns. The business lacks tools to track seasonal fluctuations and festival-driven demand, resulting in frequent understocking of key items and ad-hoc inventory decisions. These inefficiencies impact both customer satisfaction and profitability.

To address these challenges, primary data was collected manually over a period of 12 months from June 2024 to May 2025. The shop maintains daily handwritten bills for customers, which were later digitized into Excel. The dataset includes 18 essential product categories such as Rice (Local and Premium), Wheat Flour, Toiletries, Pulses, Soft Drinks, and Eggs. Metadata was organized in a daily-record format with quantities sold per product. Descriptive statistics (mean, median, standard deviation) were computed to uncover trends, variability, and seasonal changes across products.

A hybrid analysis approach was applied: Excel was used for basic summarization and Python (Pandas, NumPy, Matplotlib) enabled detailed analysis. Early insights revealed strong seasonal demand shifts (e.g., winter spikes in tea and dairy), and festival-induced peaks in premium rice, eggs, and oils. The study has laid the foundation for a predictive inventory model that aligns stock levels with local consumption behaviour, reducing understocking risks and improving operational efficiency.

2. Proof of Originality of the Data:

Store Details:

- Name: Mallick Bhandar
- Type of Business: Business to consumer (B2C)
- Owner: Mr. Ali Mallick (F/O Mr. Mukhtar Ahmed Mallick)
- Contact: +918967354277 (Muktar Ahmed Mallick)
- Address: Khoriop, Amta, Howarah, West Bengal, 711401
- Work hours: 6:30 to 13:00 and 16:30 to 21:00
- Area: 60 sq. ft (approx.)
- Location: Click the **link** to view in map.

Image 1&2: Rear view of the shop & with the shop representative

Video Interaction & Shop Images Links:

- [Video Interaction](#)
- [Shop Images](#)

Letter From Organization:

- [Link of the letter](#) (pdf)
- [Snapshot of the letter](#) (jpg)

3. Metadata & Descriptive Statistics:

Cleaned Dataset Link: *[BDM Project Data](#)* (365 rows & 19 columns)

• Metadata

The dataset collected from Mallick Bhandar consists of 365 rows, representing daily sales data over one year. Each record includes quantities sold for the following 18 product categories:

- o **Grains & Staples**: Rice (Local), Rice (Premium), Wheat Flour, Flattened Rice, Puffed Rice
- o **Cooking Essentials**: Pulses, Cooking Oils, Coconut Oil, Sugar, Salt, Spices
- o **Snacks & Beverages**: Snacks, Tea, Soft Drinks
- o **Perishable & Others**: Dairy Product, Potatoes & Onions, Egg and Toiletries

• Descriptive Statistic

	Rice (Local)	Rice (Premium)	Wheat Flour	Flattened Rice	Puffed Rice	Pulses	Cooking Oils	Coconut Oil	Sugar	Salt	Spices	Snacks	Tea	Dairy Product	Potatoes & Onions	Toiletries	Soft Drinks	Egg
Mean	14.52	9.27	8.26	4.09	5.33	13.24	5.16	2.04	5.80	3.89	4.08	8.17	7.10	5.18	13.14	2.59	8.48	11.73
Standard Error	0.23	0.39	0.13	0.07	0.10	0.24	0.15	0.09	0.09	0.09	0.13	0.17	0.14	0.10	0.19	0.10	0.15	0.21
Median	16.00	8.00	9.00	4.00	6.00	14.00	5.00	2.00	6.00	4.00	4.00	8.00	7.00	5.00	14.00	2.00	8.00	12.00
Mode	17.00	0.00	9.00	4.00	6.00	15.00	5.00	1.00	6.00	5.00	3.00	8.00	7.00	5.00	16.00	1.00	8.00	12.00
Standard Deviation	4.47	7.45	2.42	1.37	1.97	4.49	2.82	1.70	1.70	1.64	2.52	3.23	2.64	1.89	3.68	1.98	2.91	3.95
Sample Variance	19.94	55.44	5.84	1.88	3.90	20.16	7.94	2.90	2.88	2.70	6.34	10.46	6.96	3.58	13.57	3.91	8.49	15.62
Kurtosis	-0.09	0.82	0.36	-0.05	-0.03	3.28	17.35	-0.03	-0.16	-0.21	7.90	12.70	-0.33	0.29	-0.57	0.25	1.24	10.45
Skewness	-0.68	0.92	-0.64	-0.11	-0.40	0.74	3.00	0.77	-0.29	-0.24	1.75	2.31	0.16	0.34	-0.46	0.70	0.65	2.08
Range	21	39	13	7	10	33	27	8	9	9	22	28	13	12	17	10	21	35
Minimum	2	0	1	1	0	1	0	0	1	0	0	1	1	0	3	0	0	1
Maximum	23	39	14	8	10	34	27	8	10	9	22	29	14	12	20	10	21	36
Sum	5299	3383	3015	1493	1945	4831	1884	743	2116	1419	1490	2981	2590	1889	4797	946	3094	4280
Count	365	365	365	365	365	365	365	365	365	365	365	365	365	365	365	365	365	365

Figure 1: Descriptive Statistics of 18 SKUs

1. **High Demand and Stable Products: -**

   - Rice (Local), Pulses, Potatoes & Onions have high means with 14.52, 13.24 and 13.14 respectively, indicating they are essential household items with consistent daily demand.
   - Their **median** and **mode** are also close to or above the mean, confirming their stable and frequent purchase behaviour. These should be consistently stocked.

2. **Festival Sensitive Products: -**

   - Products like Rice (Premium), Spices & Snacks have high standard deviation (7.45, 2.52, and 3.23) and large range values (39, 22, and 28), indicating variability in purchase quantities, likely due to festivals like Durga Puja, Eid etc.
   - Positive skewness and high kurtosis for these items (Spices: Skewness = 1.75, Kurtosis = 7.90; Snacks: Skewness = 2.31, Kurtosis = 12.70) confirm that demand spikes sharply during certain periods—suggesting need for predictive stocking.

3. **Volatile but Essential Perishable Products: -**

   - Egg & Dairy Product have high means (11.73 and 5.18) and noticeable standard deviation (3.95 and 1.89), giving hints of inconsistent but strong demand. Their restocking needs regular, monitored attention because of their perishability.

4. **Occasionally Bought Items: -**

   - Coconut Oil, Toiletries, & Soft Drinks have low mean values (2.04, 2.59, and 8.48), and wide ranges (8, 10, and 21), indicating that purchases are less frequent but may be done in bulk or during specific needs/events.

5. **Balanced Demand Items: -**

- Wheat Flour, Puffed Rice & Tea have moderate means (between 5 and 9) and manageable standard deviations, suggesting they are reliable and moderate demand items. Although they don't need frequent restocking but weekly stock replenishment.

6. **Risk of Overstocking or Waste: -**

- Cooking Oils & Snacks have high kurtosis values (17.35 & 12.70) implying their large purchase value. Stocking these items lead to overstocking and wastage. They need to be stocked with proper planning and analysis.

7. **Skewness Pattern and Strategic Planning: -**

- Some products have positive skewness (like Egg, Snacks), showing their smaller purchase but relatively high means (like Egg:11.73, Snacks:8.17), confirm sudden spikes, good for festival promotion.
- Some items have negative skewness (like Rice (Local), Wheat Flour), showing their larger purchases are common.

4. Detailed Explanation of Analysis Process/Method:

- **Data Collection & Preprocessing:**
  - The data was collected from handwritten bill of Mallick Bhandar and later digitized in excel. The data has 365 days of records from June 2024 to March 2025 over 18 different essential products. The final dataset, has 365 rows and 19 columns, contains daily sales of 18 products along with date.
  - During manual data entry into Excel, preprocessing was conducted concurrently. Zero sales were used to denote valid days with no purchase, helping retain the continuity of time series analysis. A box-plot was used to detect extreme outliers those might be input errors and then each suspected outlier was cross-verified against original handwritten bills or discussed with the shopkeeper. Erroneous entries were corrected using the median value of surrounding days. All columns, except for the Date column, were formatted as numerical to facilitate quantitative analysis.
  - **Justification:** Preprocessing ensured that data are reliable, accurate and consistent which is crucial for future analysis. Any errors in the data may skew the analysis, lead

to wrong interpretation specially in seasonal trend analysis.

- **Descriptive Statistics:**
  - Key statistical measures were calculated across 18 SKUs. Metrics of central tendency included mean, median, and mode; measures of dispersion comprised variance, standard deviation, and range; and distributional characteristics were assessed using skewness and kurtosis.
  - **Justification:** The computed descriptive statistics provided valuable insights into product demand patterns. The mean revealed the average daily demand for each product, offering a baseline for inventory planning. The mode highlighted the most frequently occurring demand levels, indicating common purchasing behaviour. The standard deviation captured the variability in demand, helping to assess the stability or volatility of each product's sales. These insights support more informed decisions in demand forecasting, stock management, and resource allocation.

- **Time Series Plot:**
  - Time series plot shows how change of sales occurs over time. It visually tracks how demand fluctuate over a fixed time frame. Monthly time series plot is used to reveal seasonal pattern, such as monsoon dip or winter spike of demand. This method provides a continuous view of the data across time, making it easier to observe cyclical or repeating demand trends.
  - The dataset comprises daily sales records of 18 SKUs over a one-year period. To ensure clarity and avoid visual clutter in the time series analysis, individual SKUs were not plotted separately. Instead, the products were grouped into four major categories based on their nature and usage patterns: **Grains & Staples**, **Cooking Essentials**, **Snacks & Beverages**, and **Perishables & Others**. This categorization enables a more interpretable visualization of monthly sales trends while still preserving meaningful insights across product types.
  - **Justification:** It directly helps address Problem Statement 1 – "Lack of Visibility into Seasonal Patterns". By clearly revealing high-demand windows, this plot enables the business to prepare ahead of time, pre-stock inventory, and avoid stockouts.

- **Bar Plot:**
  - Bar plots aggregate and visualize the total monthly sales per product or category, summarizing daily data into 12 monthly totals. These types of plots are effective for comparing overall sales volumes across months. They reveal which products are

highest-selling and when, helping to pinpoint peak demand periods.

o To simplify the analysis and enhance interpretability, the 18 SKUs were grouped into four broad categories: **Grains & Staples**, **Cooking Essentials**, **Snacks & Beverages**, and **Perishables & Others**. Monthly sales were aggregated within each category and visualized using bar plots to highlight overall sales volume and seasonal demand patterns.

o **Justification:** This directly connects to both Problem Statements. It will help to understand seasonality (Problem 1) and assists in data-driven inventory planning by identifying months that require higher inventory levels (Problem 2).

- **Box Plot:**

o Box plots provide a concise summary of the distribution of daily sales by displaying key statistical measures, including the median (representing the central tendency), interquartile range (IQR) (indicating the spread of the middle 50% of data), and outliers (highlighting unusual spikes or drops in sales).

o This visualization is particularly useful for comparing sales variability across products. For instance, items like *Rice (Premium)*, *Snacks*, and *Eggs* may exhibit high variability in daily sales, whereas products such as *Salt* and *Toiletries* tend to show more stable patterns.

o **Justification:** Box plots play a critical role in informing restocking strategies:

▪ Products with high variability in demand may require buffer stock to accommodate fluctuations and avoid stockouts.

▪ Products with consistent, stable demand are well-suited for a lean stocking strategy, minimizing excess inventory and reducing holding costs.

By visually identifying these patterns, Mallick Bhandar can tailor their replenishment plans to each product's behaviour, ultimately supporting more efficient and data-driven inventory management.

5. <u>Results and Findings:</u>

- **Monthly Trend Analysis:**

After analysing monthly trend analysis across these four categories here are key observations by category –

➤ **Grains & Staples (Green Line in Figure 2):**

▪ Consistently the highest-selling category throughout the year.

- Significant spike in September–December 2024, peaking in December, likely due to festivals like Durga Puja, Eid-e-Milad and Kali Puja.
- A minor dip in February, possibly a post-festival cooling period.

➢ **Cooking Essentials (Red Line in Figure 2):**
- Follows a similar trend to Grains & Staples with:
  o Noticeable surge during October–December (peak demand during festival season).
  o Low sales during monsoon months (June–August), then a steep rise starting from September.
  o Dips again in January–February, possibly due to off-peak season demand.

➢ **Perishables & Others (Orange Line in Figure 2):**
- Steady growth from August, peaks in December, and fluctuates thereafter.
- Highly seasonal: spikes around festivals and drops during off-seasons.
- Noticeable drop in February and April 2025, possibly due to lower footfall or price sensitivity.

➢ **Snacks & Beverages (Blue Line in Figure 2):**
- This is the lowest-selling category, yet it shows strong seasonal behaviour:
  o Gradual increase from August, peaks in December, then falls again.
  o This mirrors festive and winter consumption habits (more snacks and beverages like tea during winter gatherings).
  o Growth again in March–April, possibly due to warmer weather and beverage (perhaps Soft Cold Drinks) demand.
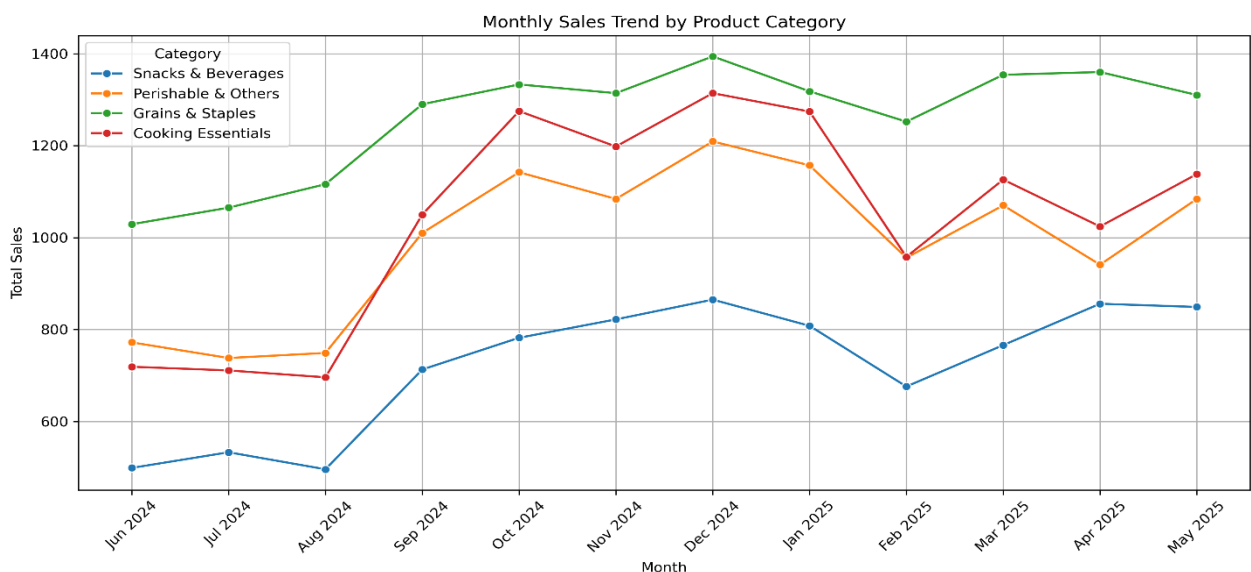

Figure 2: Monthly Trend Analysis

- **Bar Plot (Monthly Aggregated):**

    Monthly aggregation of bar plot depicts that sales peak between September to December, especially in Grains & Staples and Cooking Essentials. This aligns with major local festivals (Durga Puja, Kali Puja, Eid, Diwali), confirming seasonal demand spikes.

    ➢ **Grains & Staples (Green Bars in Figure 3) Dominate:**
        - This category consistently leads in monthly sales, indicating its critical role in household consumption and planning.
        - Suggests the need to prioritize buffer stock for items like rice and wheat during peak months.

    ➢ **Cooking Essentials (Red Bars in Figure 3) Show Volatility:**
        - Noticeable increase in October to December, followed by a dip and a secondary rise in March–May (possibly linked to pre-summer purchases).

    ➢ **Snacks & Beverages (Blue Bars in Figure 3):**
        - Though lower in volume, this category shows a gradual upward trend—especially in December, April, and May, hinting at summer preparation and festivities.

    ➢ **Perishables & Others (Orange Bars in Figure 3) Are Mid-Level but Stable:**
        - This category stays relatively steady throughout the year but sees noticeable spikes around September–January.
        - Suggests a moderate but essential inventory strategy for perishable products like eggs, potatoes, dairy, etc.
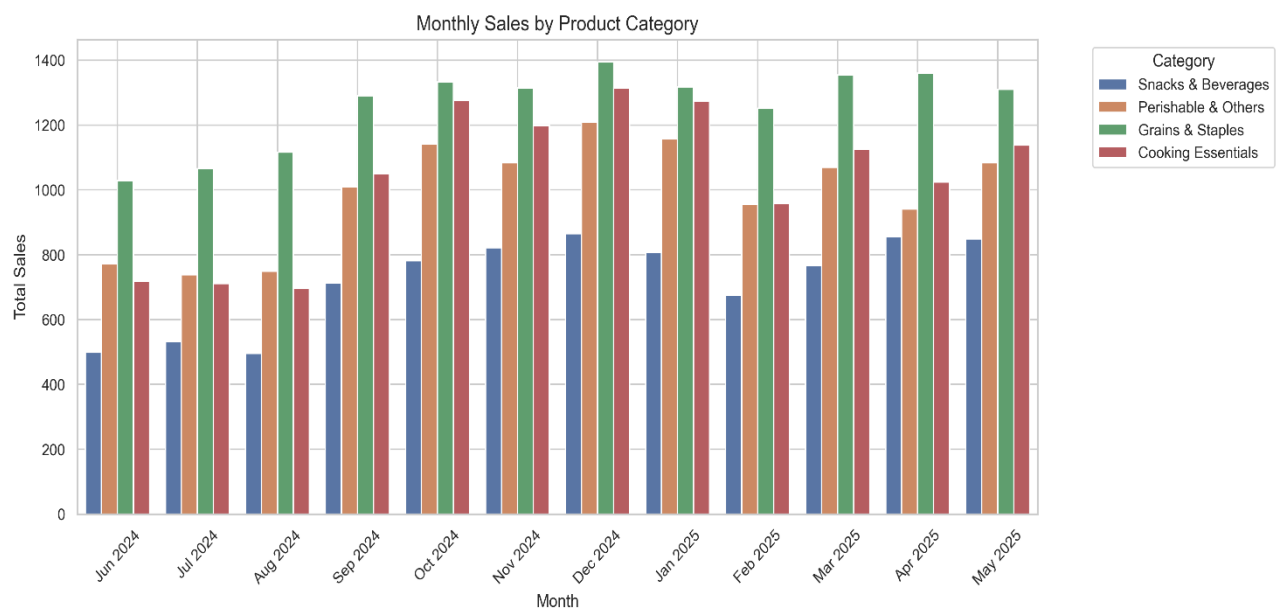

Figure 3: Bar Plot (Monthly Sales Aggregated)

- **Box Plot:**

To understand the spread, central tendency, and outliers of daily unit sales across 18 SKUs. This aids in differentiating between stable-demand products and those with volatile or festival-driven consumption patterns.

- Strong Core Products:
  - Rice (Local), Pulses, Potatoes & Onions exhibit high medians with tight interquartile ranges (IQRs), confirming regular daily purchases.
  - These staples remain above the overall average line, justifying daily or alternate-day restocking strategies with low buffer risk.
- Volatile Products with Demand Surges:
  - Egg, Rice (Premium), Soft Drinks, and Spices show long whiskers and many outliers, implying erratic demand.
  - High variance is typical for items consumed during festivals, gatherings, or weekends, so stock flexibility and event-based forecasting are crucial.
- Outlier-Heavy SKUs (Bulk Purchases or Promotions):
  - Spices, Snacks, and Cooking Oils display extreme outliers far above the upper quartile—likely reflecting bulk purchases during festive preparations or stockpiling.
  - These products should be analysed for batch-wise movement, and order quantity limits may be useful to avoid overstocking.
- Low & Predictable Items (Stable but Infrequent Sales):
  - Coconut Oil, Salt, Toiletries, and Flattened Rice lie consistently below the red mean line, with narrow boxes—suggesting low volatility and low daily movement.
  - These can be managed using longer replenishment cycles and lower holding costs.
- Balanced Performers:
  - Wheat Flour, Tea, and Sugar fall around the overall mean, with moderate spread—indicating steady yet moderate consumption.
  - Weekly restocking and maintaining minimum safety levels are recommended for these SKUs.
- Inventory Risk Flagged:

- Products like Cooking Oils and Snacks not only have high kurtosis (as seen in stats) but also visually wide spread and extreme outliers, signaling risk of dead stock or spoilage if overstocked. Just-in-Time (JIT) strategies could be explored here.
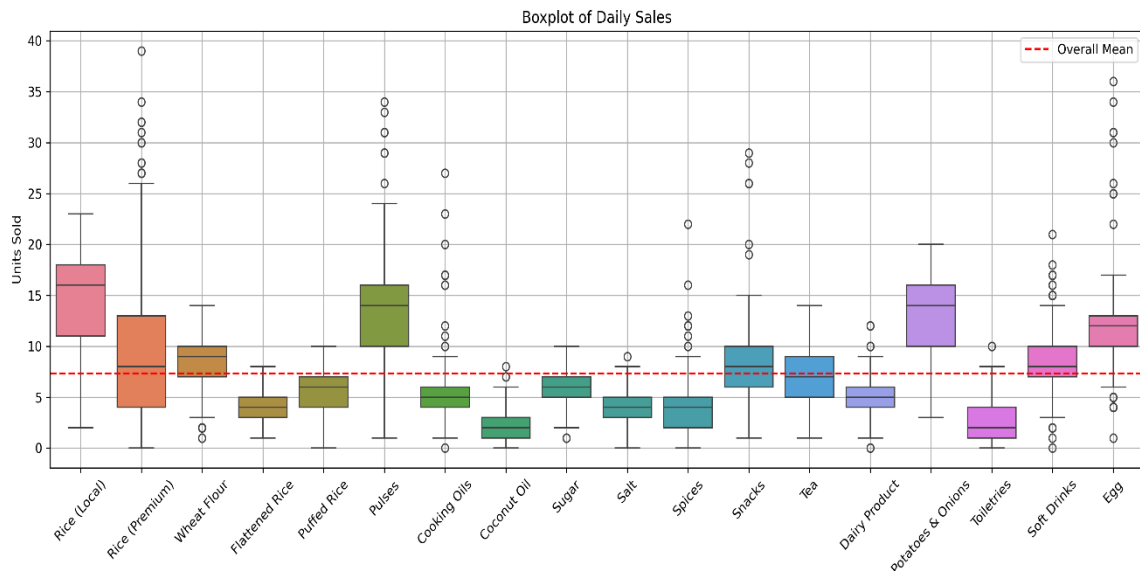

Figure 4: Box Plot across all SKUs

• **Preliminary Insights:**

- **Seasonal Peaks (Oct–Dec 2024):**
  Sales across all categories surge during the festive season (Durga Puja, Diwali, Eid-e-Milad), highlighting the need for proactive stock buildup at least one month in advance to avoid stockouts.

- **February Dip:**
  February consistently underperforms across categories, offering an opportunity for discounting or inventory liquidation to clear unsold stock.

- **Staple Consistency:**
  Grains & Staples dominate consistently throughout the year, reaffirming their priority for space allocation and regular restocking.

- **Snacks & Beverages – Growth Potential:**
  While lower in overall volume, this category shows sharp festive spikes, suggesting potential for targeted promotions and seasonal bundling.