

# BUSINESS DATA MANAGEMENT

Capstone Project Mid Term



## **Optimizing Inventory and Operations for Sustainable Growth in a Family-Run Grocery Business**

Submitted By :-

Name : Sumit Kumar

Roll No. : 22f2000848

Mail : 22f2000848@ds.study.iitm.ac.in

## Executive Summary :

This study is rooted in the original data collected from Jai Maa Jhandewali Store, a grocery business in Delhi owned by Mr. Satyam Prakash. The sales data, meticulously recorded over months, was digitized and cleaned to ensure accuracy and readiness for analysis. The cleaning process addressed missing values, standardization of formats, and categorization of SKUs into three primary groups—Dairy, Groceries, and FMCG—offering a structured dataset. The final dataset included daily sales figures for 13 SKUs and meta-data such as date and category classifications.

Descriptive statistics revealed sales patterns and variability across SKUs, with mean, range, standard deviation, and skewness providing valuable insights. For instance, FMCG showed the highest average daily sales and variability, while Dairy exhibited steady demand with occasional spikes. Advanced analytical methods were employed to address specific business challenges. Monthly sales trends were plotted to identify seasonal demand variations, segmentation analysis categorized products for detailed insights, and a correlation matrix highlighted interdependencies among categories. These methods were chosen to align with the store's challenges, such as understanding seasonal trends, optimizing inventory, and addressing the sales variability of slow-moving SKUs.

The analysis revealed that FMCG consistently outperformed other categories, with peaks from May to July driven by soft drink sales. Dairy sales remained stable but showed a slight increase during periods of low FMCG performance, suggesting complementary consumption patterns. Groceries exhibited the lowest sales but maintained a consistent correlation with FMCG. These findings underline the importance of category-specific strategies, such as targeted promotions for FMCG, improved stock management for Dairy, and leveraging Grocery items for revenue.

By integrating these insights, Mr. Satyam can enhance inventory efficiency, drive sales through targeted campaigns, and better align the store's offerings with customer demand. This actionable analysis supports the store's growth while addressing its operational challenges.

## Proof of Originality :

### Details :

- Shop Name : Jai Maa Jhandewali Store
- Type of Organisation : B2C (Business-to-Consumer)
- Shop's Address : VP Block, Pitampura, North West Delhi, 110034
- Shop Owner's Name : Mr. Satyam Prakash
- Owner's Phone Number : 9868424333
- Number of Employee : 1
- Working Time : 09:00 AM to 9:00 PM (12 hours)

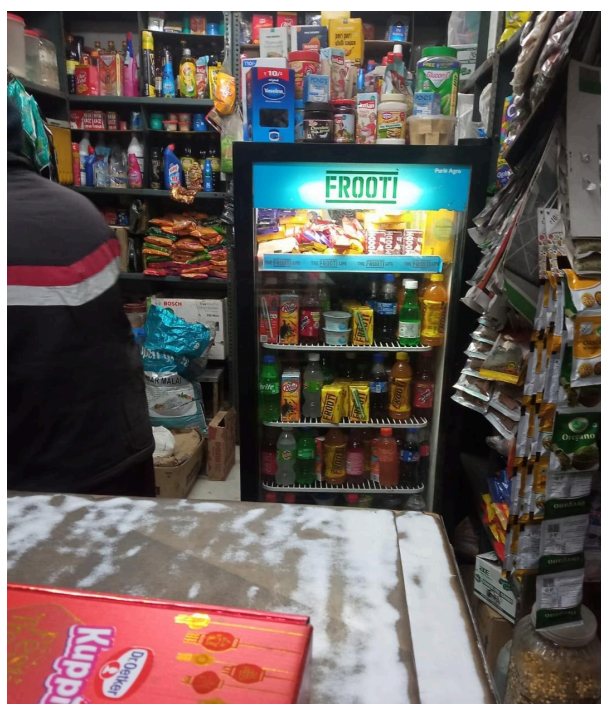
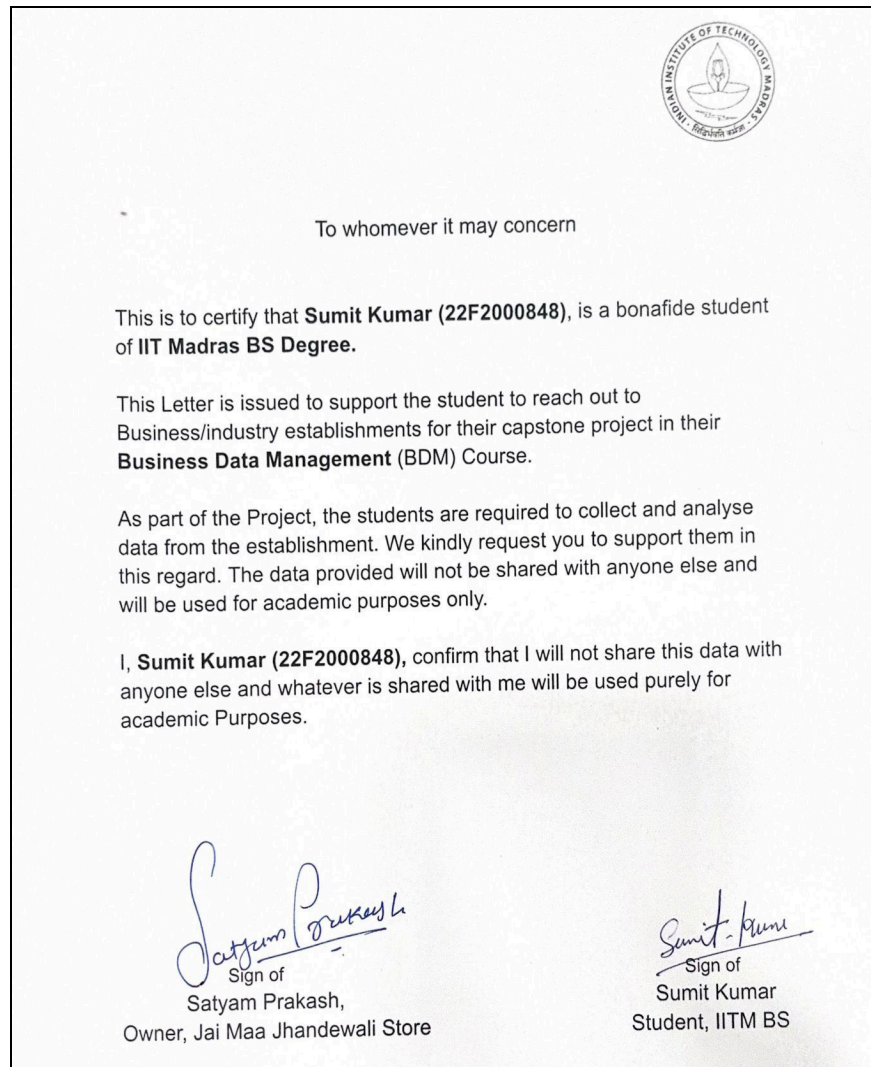


Image 1 & 2 : Jai Maa Jhandewali Store, Delhi

### Proof Links :

- Interview Video : [Sumit interview with shop owner.mp4](#)
- Video of Shop : [shop inside video.mp4](#)
- Some more photographs : [Photo\\_1.jpg](#)
- Some more photographs : [Photo\\_2.jpg](#)
- Some more photographs : [Photo\\_3.jpg](#)

- Authorization Letter :  Authorization letter.pdf



**Image 3 : Authorization from Mr. Satyam to use data for BDM Capstone project**

## Meta Data :

The sales data was initially received in raw format, either handwritten or system-generated, which required digitization. Through this process, a total of 8,563 individual bills spanning a 13-month period were compiled. To streamline the analysis, the digitized data was further consolidated, with daily sales aggregated into a single row for each day. This consolidation provided a clear summary of total sales for each SKU, making the dataset more manageable and suitable for further analysis.

Clean Dataset Link : [📄 BDM Project Data](#)

Data collection duration :13 months

Data collection dates : 01-01-2024 to 31-01-2025

Dataset Dimensions : The dataset contains **14 columns** and **398 rows**.

Date	Milk	Bread	Butter	Soft Drinks	Chips	Biscuits	Rice	Wheat	Pulse	Sugar	Salt	Toiletries	Spices
2024-01-01	24	15	5	6	11	12	2	3	4	5	1	10	1
2024-01-02	13	11	6	7	10	7	2	4	5	4	2	6	1
2024-01-03	19	17	6	10	13	8	2	4	4	6	1	6	1
2024-01-04	30	16	6	6	12	9	3	3	7	4	3	7	0
2024-01-05	20	11	4	9	10	12	1	1	4	6	3	11	1
2024-01-06	51	21	5	7	17	11	2	3	4	4	3	9	1
2024-01-07	64	21	6	6	19	10	2	2	4	5	1	11	1
2024-01-08	30	22	4	6	16	16	3	3	4	6	2	12	1
2024-01-09	36	14	6	4	14	4	3	3	5	4	1	7	1
2024-01-10	30	7	5	7	15	7	2	3	5	6	2	9	7
2024-01-11	21	21	5	2	19	7	3	3	4	6	2	9	0
2024-01-12	27	14	2	4	15	5	0	3	5	5	2	8	0
2024-01-13	60	13	3	6	10	8	3	3	4	7	2	8	1
2024-01-14	32	11	6	8	13	12	2	2	6	5	2	12	1
2024-01-15	17	17	5	9	14	8	2	2	8	7	1	9	1
2024-01-16	16	27	4	6	8	9	2	4	4	8	2	9	1
2024-01-17	23	15	6	6	11	5	2	2	6	5	2	7	1

.Image 4 : A snapshot of how the data looks like (top 18 entries)

Columns :

- Date** : The date for which the entry is maintained
- Dairy Products**:
  - Milk, Butter
- Groceries**:
  - Rice, Wheat, Pulse, Sugar, Salt, Spices
- FMCG Products**:
  - Bread, Soft Drinks, Chips, Biscuits, Toiletries

## Descriptive Statistics :

	<i>Milk</i>	<i>Bread</i>	<i>Butter</i>	<i>Soft Drinks</i>	<i>Chips</i>	<i>Biscuits</i>	<i>Rice</i>	<i>Wheat</i>	<i>Pulse</i>	<i>Sugar</i>	<i>Salt</i>	<i>Toiletries</i>	<i>Spices</i>
<b>Mean</b>	38.16	24.69	6.15	8.35	15.63	13.49	3.27	5.04	8.23	6.83	2.65	12.21	2.89
<b>Standard Error</b>	0.89	0.53	0.15	0.18	0.33	0.31	0.09	0.14	0.24	0.14	0.06	0.26	0.27
<b>Median</b>	35	23	6	8	14	12	3	4	7	6	2	11	1
<b>Mode</b>	23	25	4	6	13	8	2	3	5	6	2	9	1
<b>Standard Deviation</b>	17.65	10.50	2.93	3.50	6.59	6.21	1.89	2.89	4.85	2.83	1.15	5.17	5.46
<b>Sample Variance</b>	311.61	110.28	8.56	12.28	43.40	38.55	3.56	8.32	23.48	7.98	1.33	26.76	29.77
<b>Kurtosis</b>	0.68	-0.19	1.71	-0.19	-0.15	0.10	2.95	3.65	4.30	-0.58	-0.11	-0.08	30.89
<b>Skewness</b>	0.83	0.50	1.14	0.69	0.65	0.73	1.52	1.62	1.83	0.53	0.58	0.79	4.95
<b>Range</b>	104	54	19	16	35	34	11	18	28	12	6	26	53
<b>Minimum</b>	4	3	0	2	0	0	0	1	2	2	0	2	0
<b>Maximum</b>	108	57	19	18	35	34	11	19	30	14	6	28	53
<b>Sum</b>	15148	9801	2443	3313	6206	5356	1298	2000	3266	2713	1052	4849	1147
<b>Count</b>	397	397	397	397	397	397	397	397	397	397	397	397	397

.Image 5 : Descriptive Statistics Summary of Sales Data across the categories

### 1. Dairy Products

- **Milk** : The average daily sales for Milk are 38.16 units, with a notable range of 4 to 108 units, reflecting fluctuating demand. The standard deviation of 17.65 indicates significant variability, often with spikes during weekends and festivals. The positive skewness of 0.83 suggests occasional days of exceptionally high sales, while the 25th and 75th percentiles of 25 and 50 highlight that most days experience moderate sales volumes.
- **Butter** : The average daily sales for Butter are 6.15 units, with a range spanning from 0 to 19 units and a standard deviation of 2.93, indicating moderate variation in demand. The positive skewness of 1.14 highlights the presence of a few high-demand days, likely corresponding to holidays or special occasions. The median of 6 units suggests that sales remain regular and steady on most days.

### 2. Groceries

- **Rice** : Daily sales for Rice average 3.27 kg, with a range from 0 to 11 kg and a standard deviation of 1.89, indicating moderate variation in demand. A skewness of 1.52 points to occasional bulk purchases, often observed during month-ends or festive periods. The median sales of 3 kg closely align with the mean, reflecting consistent purchasing behavior on most days.



- **Wheat** : Daily sales for Wheat average 5.04 kg, with a range of 1 to 19 kg and a standard deviation of 2.89, reflecting trends of bulk purchases on certain days. A skewness of 1.62 indicates occasional high-demand periods, likely during month-ends or festive seasons. The median sales of 4 kg and a 75th percentile of 6 kg represent typical buying patterns for most days.
- **Pulse** : Daily sales for Pulses average 8.23 kg, with a range of 2 to 30 kg and a standard deviation of 4.85, reflecting considerable variability in purchasing patterns. A skewness of 1.83 suggests significant outliers, likely due to bulk purchases during festivals or monthly restocking. The median sales of 7 kg indicate a consistent and steady demand on most days.
- **Sugar** : Daily sales for sugar average 6.83 kg, with a range spanning from 2 to 14 kg and a standard deviation of 2.83, reflecting a consistent and reliable demand. The skewness of 0.53 suggests occasional demand surges, possibly due to festivals or seasonal needs. The median sales of 6 kg and the 75th percentile at 8 kg indicate strong, steady consumption patterns, making sugar a staple product with predictable trends.
- **Salt** : Salt sales average 2.65 kg daily, with a narrow range between 0 and 6 kg and a standard deviation of 1.15, highlighting its stable demand. The slight positive skewness of 0.58 suggests infrequent instances of higher sales, likely tied to larger purchases for bulk needs. Median sales of 2 kg and a 75th percentile of 3 kg underline its status as an essential commodity with consistent buying habits.
- **Spices** : Average daily sales are 2.89 units, with a broad range spanning from 0 to 53 units and a high standard deviation of 5.46, indicating sporadic bulk purchases. The high positive skewness of 4.95 reflects rare but significant demand spikes, typically aligned with festivals or special occasions. Median sales of 1 unit highlight the regular low consumption of spices, with occasional bulk buys catering to specific needs or events.

### 3. FMCG Products

- **Bread** : Average daily sales are 24.69 units, with a range of 3 to 57 units and a standard deviation of 10.50, reflecting moderate variability. A slightly positive skewness of 0.50 indicates occasional higher sales days, likely on weekends or during holidays. The

median sales of 23 units and a 75th percentile of 31 units point to consistent and steady demand for bread.

- **Soft Drinks** : Daily sales average 8.35 units, ranging from 2 to 18 units, with a standard deviation of 3.50, indicating moderate variation. A skewness of 0.69 suggests periodic surges, especially during the summer months or festive seasons. The median sales of 8 units and a 75th percentile of 11 units reflect a steady consumption pattern with occasional peaks.
- **Chips** : Average daily sales are 15.63 units, with a range of 0 to 35 units and a standard deviation of 6.59, showing significant variability. A skewness of 0.65 highlights occasional higher sales days, likely during celebrations or social gatherings. Median sales of 14 units and a 75th percentile of 20 units suggest strong regular demand with occasional surges.
- **Biscuits** : Daily sales average 13.49 units, ranging from 0 to 34 units, with a standard deviation of 6.21, reflecting moderate variability. A skewness of 0.73 indicates occasional spikes, possibly linked to promotional events or seasonal trends. The median sales of 12 units and a 75th percentile of 17 units indicate consistent daily demand with room for periodic increases.
- **Toiletries** : Sales average 12.21 units per day, with a range of 2 to 28 units and a standard deviation of 5.17, showcasing moderate fluctuations. A slight skewness of 0.79 points to occasional higher sales, potentially due to stock-ups or promotions. The median sales of 11 units and a 75th percentile of 16 units reflect a steady demand for essential toiletries.

## Methods of Analysis with Justification :

### 1. Data Cleaning and Consolidation

- **Process:**

The store's sales data was maintained in handwritten records. Over multiple visits, this data was digitized into Excel format to facilitate analysis. During this process, issues like missing values (name of product), duplicate entries, and discrepancies in SKU names were identified and resolved. After cleaning, the data was consolidated to summarize



daily sales for each product category. For instance, instead of multiple entries for biscuits or milk on a single day, these were aggregated into a single-row summary for each day.

- **Justification:**

Cleaning ensures accuracy and consistency, while consolidation provides a holistic view of sales trends. For a small store like Jai Maa Jhandewali, where resources are limited, this streamlined dataset is essential to understand daily sales performance without being bogged down by raw, unorganized data. Consolidation also makes it easier to analyze long-term trends and make data-driven decisions.

## 2. Descriptive Statistics Analysis

- **Process:**

Key statistical metrics were calculated for each SKU, such as mean, median, range, standard deviation, skewness, and percentiles. These metrics were used to summarize the performance of each product. For example, sugar showed consistent daily demand with a mean of 6.83 kg and low variability, while soft drinks exhibited high variation due to spikes during festivals or hot weather.

- **Justification:**

This step provides a detailed understanding of each SKU's behavior. For staples like rice, wheat, and sugar, steady demand patterns were identified, ensuring optimal stock levels without overstocking. For discretionary items like chips and soft drinks, high variability highlighted the need for dynamic stocking strategies to meet peak demand during festivals or holidays. Descriptive statistics allow Mr. Satyam to prioritize essential items and manage shelf space effectively.

## 3. Trend Analysis

- **Process:**

Sales data was analyzed over time to identify seasonal patterns, weekday-weekend variations, and festival-related spikes. For instance, soft drink sales surged during summer, while dairy products saw increased demand on weekends. A monthly sales trend was also examined to predict high-demand periods, such as Diwali and New Year.

- **Justification:**

Understanding trends allows Mr. Satyam to prepare for high-demand periods in advance. For example, stocking more dairy products and soft drinks during summer ensures customer satisfaction and prevents lost sales. Similarly, identifying seasonal declines helps avoid overstocking, reducing waste and costs. Trend analysis equips the store to adapt to fluctuations in demand efficiently.

#### 4. ABC Segmentation Analysis

- **Process:**

Sales data was segmented using the ABC analysis method based on the contribution of each SKU to the total revenue.

- Category A: High-priority items (e.g., Dairy products and staples like rice and sugar) that contribute to 70-80% of revenue but constitute a small percentage of SKUs.
  - Category B: Moderate-priority items (e.g., FMCG products like chips and soft drinks) that contribute around 15-20% of revenue.
  - Category C: Low-priority items (e.g., slow-moving products like specific spices or toiletries) that contribute the remaining 5-10% of revenue.
- Further segmentation within each category was conducted based on sales frequency during weekdays, weekends, and festivals to identify distinct patterns. For example, Category A items like dairy saw peak demand on weekends, while Category B items like soft drinks spiked during festivals.

- **Justification:**

ABC segmentation allows Mr. Satyam to prioritize resources and focus on high-value items. For example, ensuring sufficient stock of Category A items like staples and dairy prevents revenue loss due to stockouts. Category B items can be strategically promoted during peak times, such as festival seasons, to boost sales. Low-priority Category C items can be analyzed for potential removal or optimized stocking to save shelf space.

#### 5. Correlation Analysis

- **Process:**

Relationships between SKUs were analyzed using correlation coefficients. For example, strong correlations were observed between bread and butter, chips and soft drinks, and

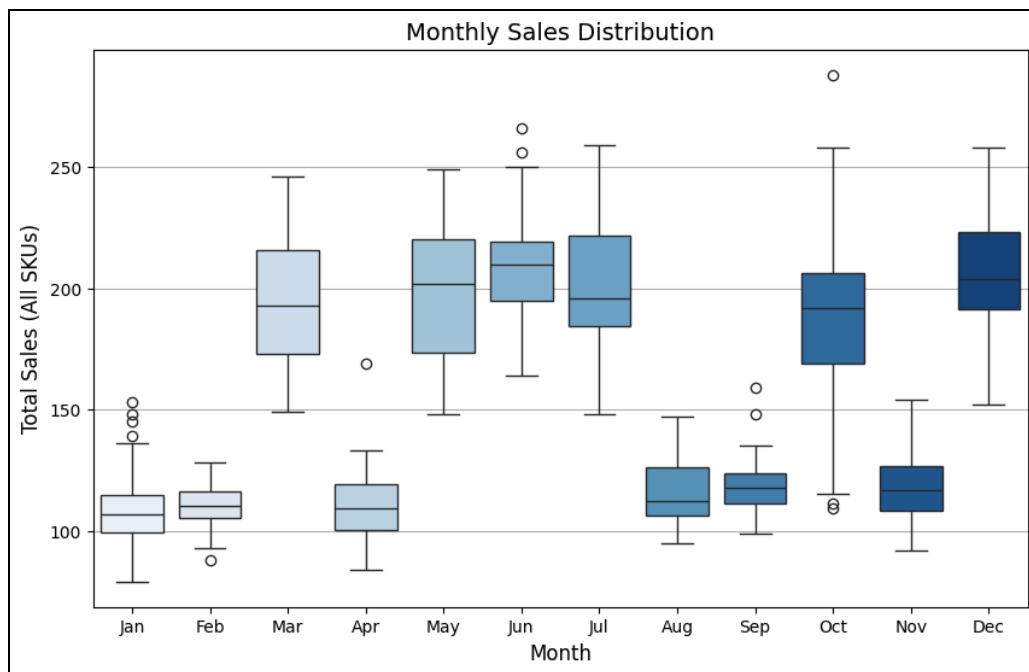
spices and pulses. Correlation with external factors like festivals, temperature, or holidays was also examined to understand demand influencers.

- **Justification:**

Correlation analysis identifies complementary products, enabling the store to create bundled offers. For example, bread and butter or chips and soft drinks could be promoted together, increasing overall sales. Identifying external influencers, such as higher soft drink sales during summer, allows for better inventory planning. This ensures the store remains agile in meeting customer demands while maximizing profitability.

## Results and Findings :

The box plot analysis of monthly sales reveals significant patterns across different months. October emerges as the month with the highest sales volume and spread, with sales ranging from 200 to 250 units. This trend is likely driven by major festivals such as Navratri and Diwali, which stimulate consumer demand. December follows closely, fueled by the festive spirit of Christmas and New Year celebrations. July also shows relatively higher sales, possibly linked to seasonal factors like monsoons that influence consumer behavior.



**Image 6 : Box plot showing distribution of sales over the months**

On the lower end, January and February exhibit the least sales. This could be attributed to Delhi's cold and foggy winter weather, which often reduces customer footfall and purchasing activity. However, to gain a deeper understanding of these sales patterns, it's essential to explore trends within individual product categories.

To better interpret these trends, we will group the SKUs into the following categories:

- Dairy: Milk and Butter
- Groceries: Wheat, Rice, Pulses, Sugar, Salt, and Spices
- FMCG: Biscuits, Soft Drinks, Bread, Chips, and Toiletries

The line chart, visualizing sales trends across the three categories (FMCG, Dairy, and Groceries), reveals some insightful patterns. Every month, the red line, representing FMCG sales, consistently occupies the top position, indicating the dominant demand for FMCG items. Dairy, despite containing only two SKUs (Milk and Butter), is in second place with the blue line, outperforming the more extensive Grocery category. The green line, representing Grocery sales, remains at the bottom throughout the year.

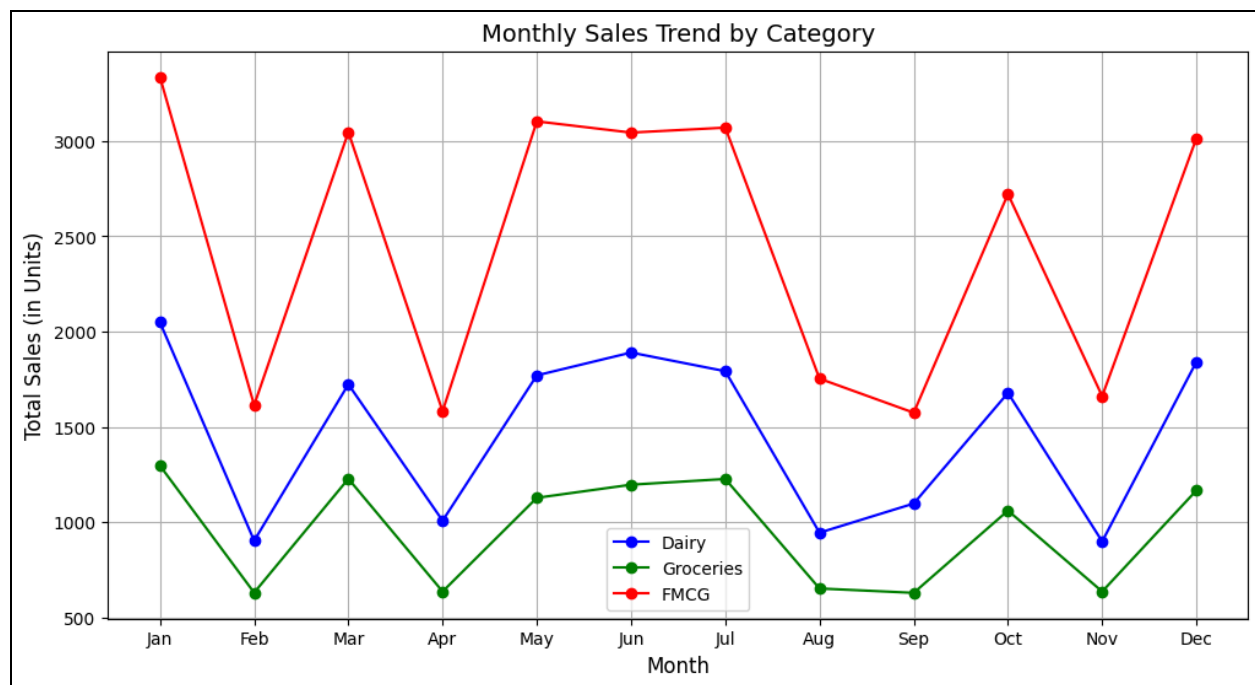



Image 7 : Line Chart shows sales trend of 3 categories (Dairy, FMCG and Groceries)



**Seasonal Peaks:** There is a noticeable peak in FMCG sales between May and July, likely driven by the demand for soft drinks during the summer season. This highlights the influence of seasonality on sales trends.

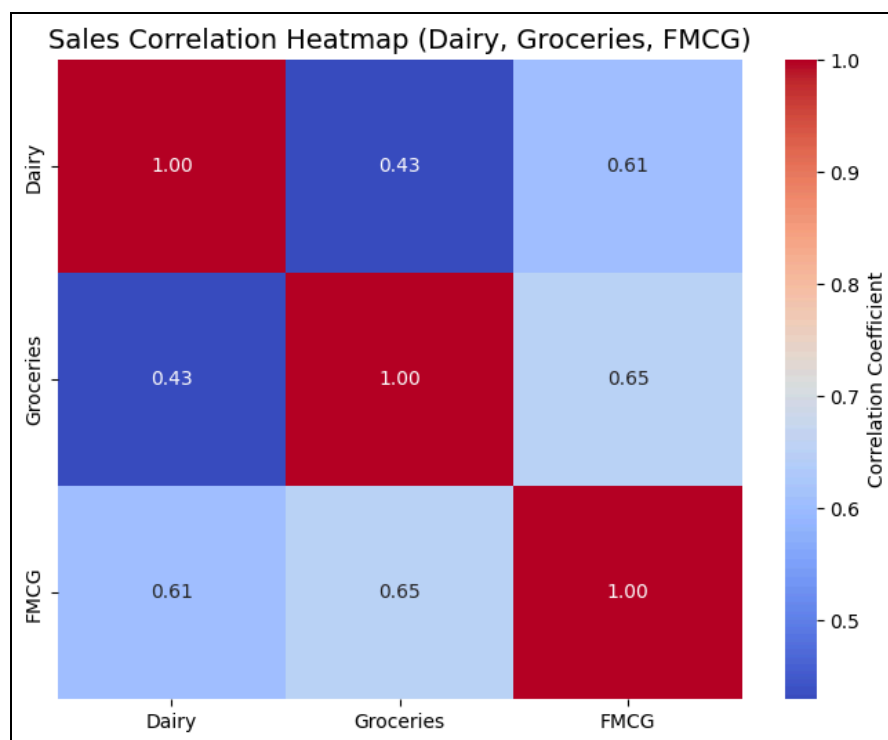
**Consistent Patterns Across Categories:** In most month-to-month transitions, the sales trends of all three categories follow a similar direction—either increasing or decreasing together. This suggests shared external factors such as festivals, weather, or consumer spending cycles driving overall demand.

**Exception in August to September Transition:** While moving from August to September, a dip in sales is observed in both FMCG and Grocery. However, Dairy shows a slight increase during this period. This deviation calls for further exploration to identify what factors, such as changes in consumer preferences or promotions, might have contributed to this anomaly.

**Next Steps:**

- **SKU-Level Analysis:** To better understand these patterns, we will analyze the trends at the SKU level. This will help identify specific products driving the peaks and dips in each category.
- **Revenue Contribution Analysis:** After discussing with Mr. Satyam, we recognize that Grocery SKUs, while slower-moving, may significantly contribute to revenue due to their higher selling prices. To account for this, we will incorporate revenue analysis in addition to sales volume trends, providing a more comprehensive view of category performance.

The sales correlation matrix reveals significant insights into the relationships between the three categories—Dairy, Groceries, and FMCG. The strongest correlation is observed between FMCG and Groceries (0.65), indicating that sales in these categories are often influenced by similar factors such as customer footfall, seasonal demand, or promotional events. For instance, during festivals or peak shopping periods, customers tend to purchase both FMCG products like snacks and beverages alongside essential groceries, demonstrating complementary buying patterns.



**Image 8 : Correlation heatmap between categories**

A moderate correlation exists between FMCG and Dairy (0.61), suggesting that customers buying everyday FMCG items also frequently purchase Dairy products, given their essential and recurring nature. However, the slightly weaker correlation compared to FMCG and Groceries might be due to Dairy's consistent demand driven by daily consumption rather than seasonal trends. In contrast, Dairy and Groceries show the weakest correlation (0.43), reflecting their distinct consumption patterns—Dairy being a daily necessity, while Groceries like rice, wheat, and pulses are typically purchased in bulk or less frequently.