Optimizing Profits: A Data-Driven Approach for Har-Hith Grocery Store's Inventory Management

Mid-term report for the BDM capstone Project

Submitted by

Name: Lakshay

Roll number: 21f3001076

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1) Executive Summary

The project focuses on addressing challenges encountered by the B2C business, specifically the grocery store known as "Har-Hith Grocery Store" .This establishment offers a diverse range of essential and everyday products to its customers. Several challenges have been identified, including issues related to inventory management, customer retention, and cyclic variations, all of which have a direct impact on the store's overall revenue and net profit. This project is dedicated to providing solutions to these unique business challenges.

The primary objectives of the project include evaluating the declining sales rate, managing cash flow effectively, and implementing strategies to control inventory. To achieve this, data collected from the Point of Sale (POS) machine over the past year is analyzed in this project. For the proof of originality of data images, letter from organization and other related document provided.

After the data collection process, the data was meticulously cleaned and thoroughly examined to ensure its proper format and structure. Essential context and details were provided to facilitate a comprehensive understanding, effective management, and efficient utilization of the data. Following the examination of the data, various techniques were employed for its analysis, commencing with descriptive analysis.

The analysis has revealed some interesting insights about the business, such as the highest revenue generating category, month in which highest sale occur, average day of inventory, and FSN percentage of categories, which contribute to effective inventory management, and identification of seasonal demands. In conclusion, by adopting a data-driven approach, this initiative aims to optimize profits for Har-Hith Grocery Store and enhance its overall business performance.

2) Proof of originality of the Data

The collected data spans a period of one year, from January 01, 2023 to December 31, 2023. For the proof of originality of data all required things mentioned as below:

Link of Dataset excel file – Click Here

- **Photographs:** Images of the POS machine and images of store front as well as from inside, are attached herewith to serve as proof of the originality of the data.

Figure 1: Mr. Sanjay and his brother in store

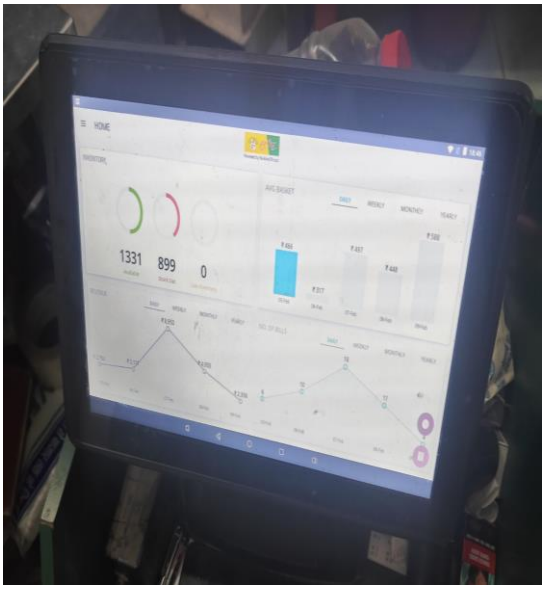

Figure 2: Point of Sale (POS) machine

Figure 3: Picture of store front

Figure 4: Inside area of Store

- **Letter from the Organization:** The attached letter, authorized and signed by the store owner, Mr. Sanjay Karwasra, confirms the store's permission for the collection of data and its implementation in the project.
  Click here to access the Letter from the organization
- **Short Video with Owner:** A short video of interacting with owner uploaded,
  Click here to access the video

3) Metadata and Descriptive Statistics

3.1 Meta Data

- **Data Collection Process:** Data collected from the store's point-of-sale (POS) machine includes information on daily sales, purchase orders, stock movement details, etc. Data collection took 4 days as per availability of POS machine. The collected data spans a period of one year, from January 01, 2023 to December 31, 2023.There are a total of 1,057 types of products categorized across 21 categories.

- **Data Format:** The collected data is in the format of csv (comma-separated values) files, which are organized into different files for product-wise monthly sales, daily sales, and stock movement. One spreadsheet encompasses the monthly sales data for each product, the daily sales CSV file, and the stock movement CSV file.
- **Data Structure:** There is three types of sheets and their data structure is as following:
  - Daily Sales sheet – It consist two columns Date of Sale and Revenue.
  - Product wise Sales – This sheet contains multiple columns Month of Sale, Product, Unit of Measurement, Category, Sub-Category, Quantity, Revenue, MRP, Sell price (S.P.), Cost Price (C.P.).
  - Stock Movement – This sheet contains Product, Category, Sub-Category, Opening Stock, In-ward Stock, Out-ward Stock and Closing Stock
- **Data Formation:** Data types of each column described in below table –

Data Formation	
Data Types Of Column	**Columns**
Categorical Data	Product, Category, Sub-Category, Months, Date
Numerical Data	Revenue, MRP, CP,SP, Opening Stock, Closing Stock, In-ward Stock, Out-ward Stock

- **Unit of Measurement (UOM):** Two types of units of measurement, namely pieces (pcs) and kilograms, are utilized. For packed products, pieces are used, while kilograms are employed as the unit of measurement for loose products.
- **Dependencies and Relationships:** Product wise Sales linked to Stock Movement sheet by product name, category and issued quantity. And Daily Sales linked by Month and revenue over month.
- **Data Limitation:** The data is collected through an automated Point of Sale (POS) system, and limitations may arise from discrepancies in manual entry or system malfunctions. Regular reconciliation with physical inventory counts is conducted, but variations may occur due to factors such as human error or discrepancies in recording.

3.2 Descriptive Statistics

Descriptive Statistics Based on Monthly Revenue:

- Average monthly revenue was **₹171,639**. Average daily revenue was **₹5,643**. The highest revenue was observed in the month of June, amounting to **₹256,188**, whereas the lowest revenue, amounting to **₹122,039**, was recorded in the month of February.
- With a standard deviation of **₹38,153**, a notable variability observed in monthly revenues. This suggests that the actual revenues from month to month deviate from the average by a considerable amount.
- The median revenue is **₹163,721**. Given that the median is close to the mean, it indicates a relatively symmetrical distribution, further insights can be gained by considering skewness. A positive skewness of **0.93** suggests that the distribution of revenues is skewed to the right. This means there may be a few months with exceptionally high revenues, contributing to the positive skew.
- The range of **₹134,150** shows the spread between the lowest and highest monthly revenues. A larger range indicates more variability in revenue.
- A kurtosis value of **0.75**, close to 0, suggests a distribution with moderately heavy tails. This means there might be a moderate number of outliers, but the distribution is not extremely peaked or flat.

Sales Trend: The total revenue for the year 2023 was recorded as **₹2,059,672**. From revenue trend graph (Figure 5) it is observed that sale increases april onwards reach at peak in june.

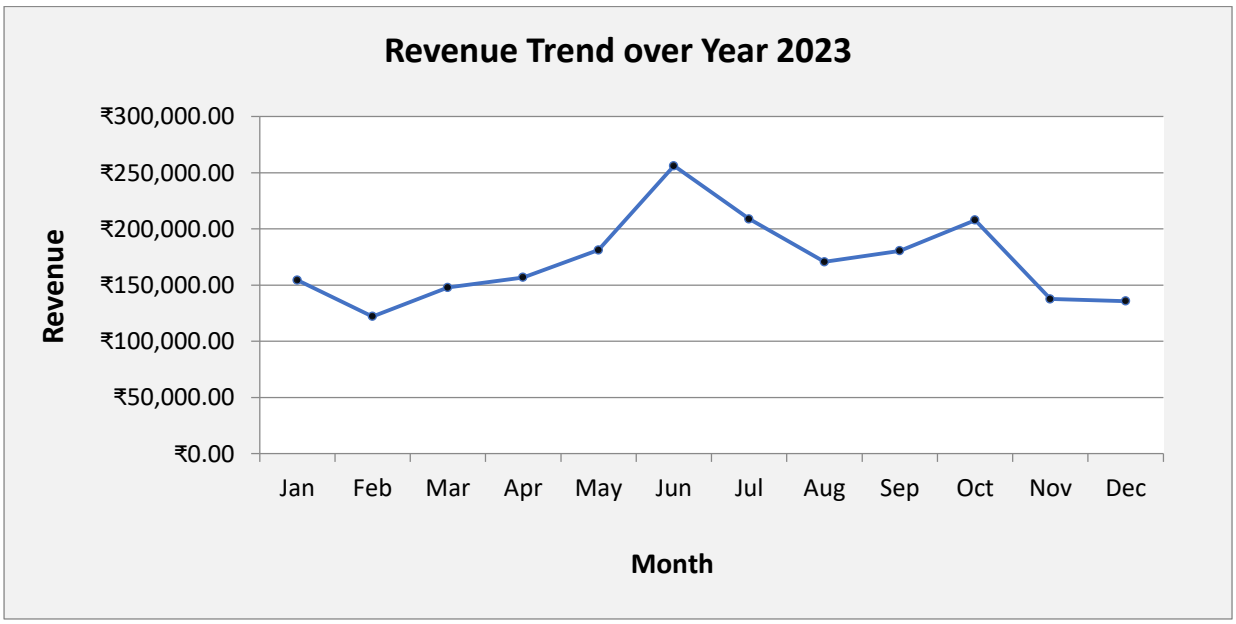

Figure 5: Line chart showing month wise variation of revenue in year

Product Performance: By comparing sales of different categories as shown in category wise revenue (Figure 6), Beverages & Soft drinks category emerges as the dominant one. The revenue generated by this category was **₹398,117.**

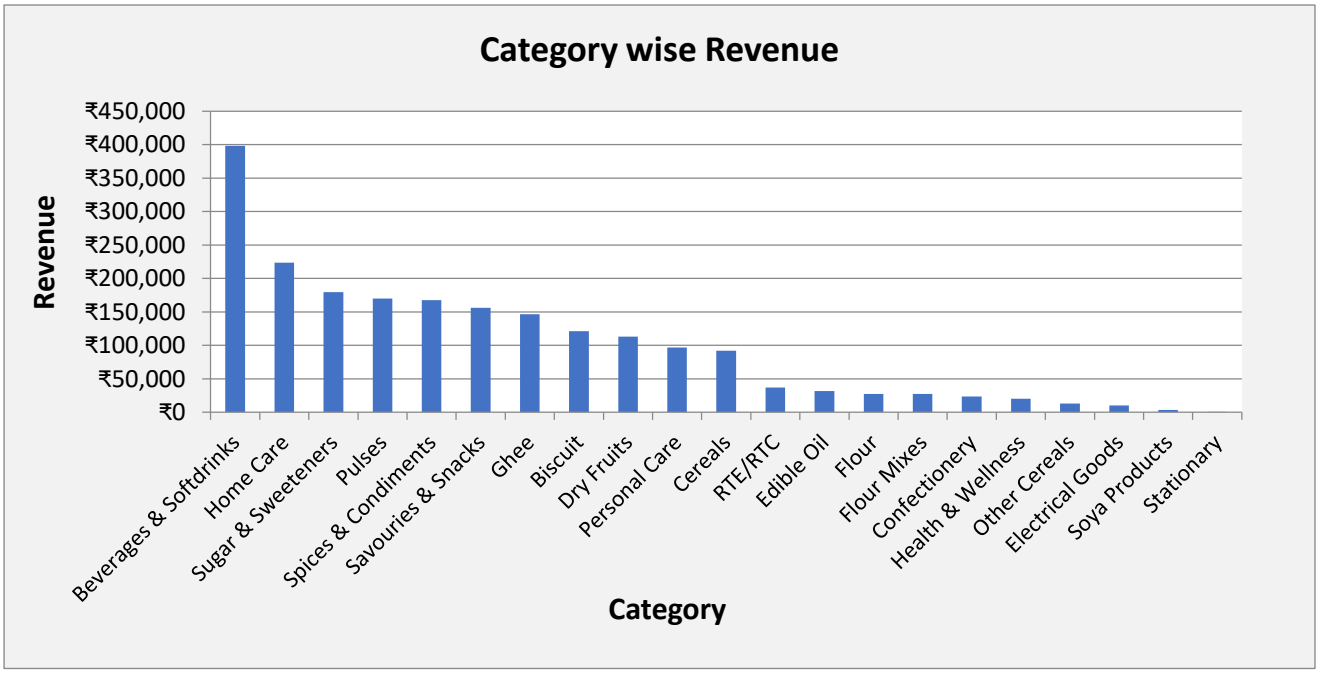

Figure 6: Column chart showing revenue category wise in year 2023

Stock Movement: Total inventory value on January 1, 2023 was **₹343,245** and total expenditure on received inventory was **₹1,799,497.** Closing Stock value at the end of year 2023 was **₹291,517.**

4) Detailed Explanation of Analysis Process/Method

Data was collected over a one-year period, and the gathered information was initially stored in distinct files in CSV format. To streamline the analysis process, all files were consolidated into a single spreadsheet. Subsequently, data was carefully examined to ensure its structure and format were clean and organized, facilitating effective analysis. This step involved identifying and rectifying inconsistencies, addressing missing values and identifying potential outliers that could influence the accuracy of the results. Additionally, as part of the data cleaning process, extraneous columns such as HSN code, barcode, and tax were removed to refine the dataset for further analysis.

For Statistical Analysis following techniques used in process:

- **Exploratory Data Analysis (EDA):** A variety of descriptive statistical techniques were applied to summarize and explore the data. This included calculating measures such as mean, median, max revenue, min revenue, standard deviation, kurtosis, skewness, range and inventory stock movements to gain insights into the overall dynamics of the store. By examining statistical measures, EDA helps identify outliers that might impact the overall analysis. Outliers can be further investigated to understand their impact on the data.
- **Sales Analysis:** Charts (column, bar, line, and pie) visually representation of sales trends, revenue proportion, Average daily sales, making it easier to interpret and communicate insights. Evaluated sales performance by analyzing sales trends over time, identifying peak sales periods. Identifying key drivers contributing to sales variations such as specific product categories using revenue chart enable to increase profit.
- **Pareto analysis:** Pareto analysis performed on revenue. This analysis allows for a clear and effective representation of the cumulative contribution of each category to the overall revenue. Pareto analysis highlights the vital few categories that contribute the most to overall revenue. This insight enables strategic decision-making and ensures that high-impact categories are consistently prioritized in stock management.
- **Inventory Analysis:** For inventory analysis stock movement data used. Using this data, key components such as Average Inventory, Cost of Goods Sold (COGS), Average Day of Inventory and Inventory Turnover Ratio (ITR) calculated. These components help to determine that the store meets demand efficiently while minimizing overstock or stockouts. Further insights can be gained by FSN analysis.
- **FSN analysis:** Average day of Inventory and Inventory turnover rates, identify slow-moving or fast-moving items, and this analysis helps to optimize stock levels. This ensures that resources are allocated efficiently, preventing overstocking or shortages.

5) Results and Findings

Based on the analysis conducted using the graphs and charts, several key findings:

- **Top-Selling Categories:** From the sales trend, it is observed that the highest revenue is generated in the second quarter. The reason behind this increase in sales is the demand for beverages and soft drinks, which collectively contribute approximately 20% to the overall revenue, as shown in Figure 7 revenue proportion by category. The

top five revenue-generating categories include Home Care, Sugar & Sweeteners, Pulses, and Spices & Condiments.

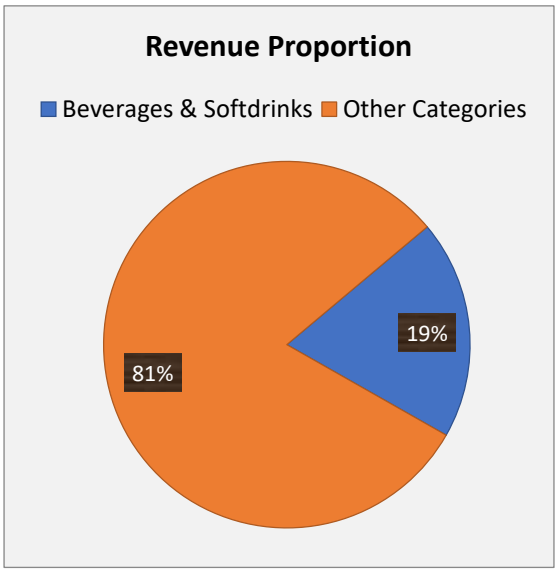

Figure 7: Pie chart showing revenue proportion

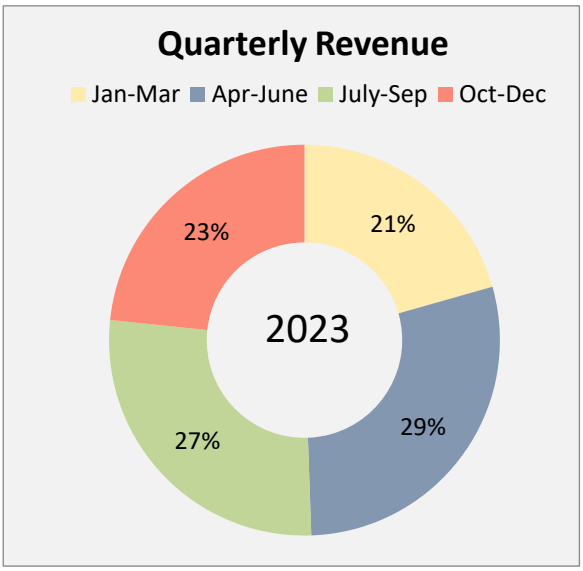

Figure 8: Doughnut chart showing Revenue Proportion Quarterly

- **Seasonal Revenue Patterns:** Analysis of the Quarterly Revenue Chart (Figure 8) indicates the lowest sales occurring in the first quarter of the year. This signifies a decline in sales during the winter season. The contributing factor was drop in sales of high-performing categories.

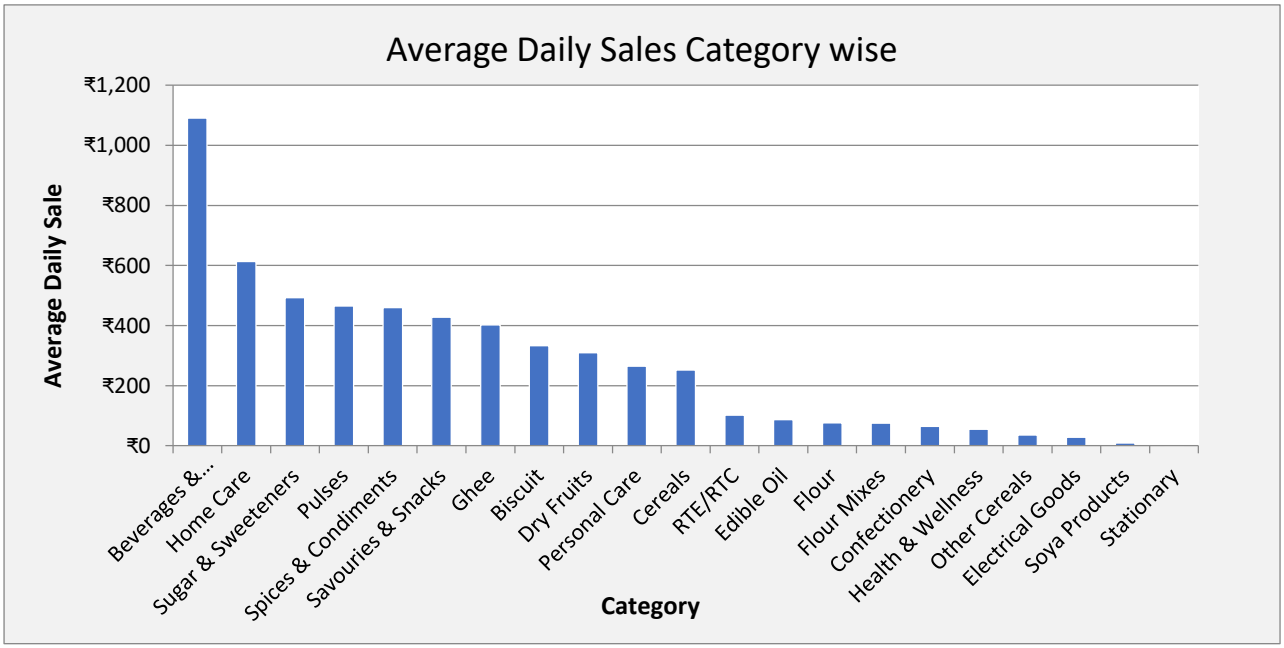

Figure 9: Column chart showing sales trend category wise over the year 2023

- **Daily sales patterns:** Average daily sales category wise (Figure 9), reveal key insights about some categories. Products under the Cereals, Other Cereals, Ready to Eat or Cook (RTE/RTC) categories show low sales, even though these products are consumed daily or on alternate days in general households

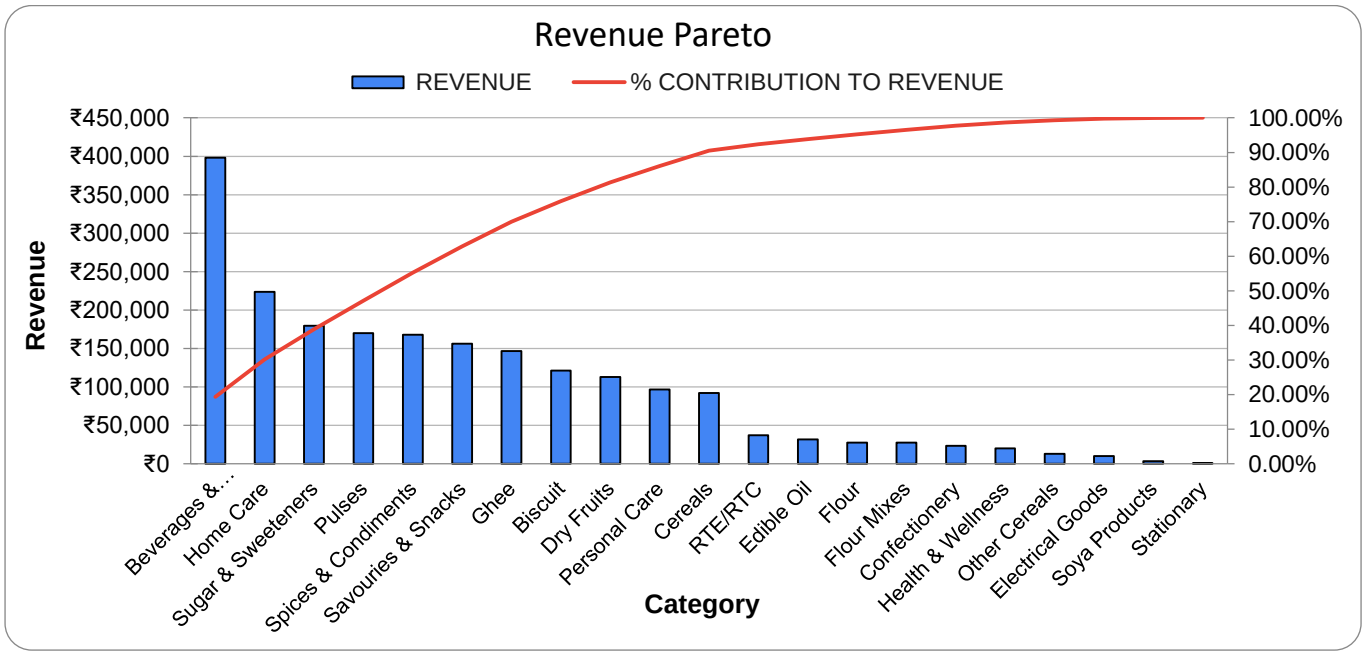

Figure 10: Revenue Pareto Chart

- **Pareto Analysis:** The Revenue Pareto chart (Figure 10), reveals that 80% of the revenue is generated by 40% of the categories. This insight is invaluable for strategic decision-making, such as ensuring that high-impact categories are consistently maintained in stock.