Smart Analytics for a Vegetable Seller

The Mid-Term Report for the BDM Capstone Project

Submitted by

Name: Anuj Krishna S

Roll number: 21F2000852

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1 Executive Summary

Mr. Muthuraj A, a longtime resident of Chennai, has been running a roadside vegetable shop in Ashok Nagar for over 35 years. Despite his strong local presence and established customer relationships, his B2C business faces significant challenges, including stagnant sales that hinder revenue growth and excessive vegetable wastage that impacts profitability.

The root cause of these issues is the lack of structured data on sales, stock levels, and procurement. To address this, I am systematically collecting data by tracking daily customer purchases, market procurement, existing stock, and excess inventory. This data is then analysed to identify sales patterns, optimize stock management, and implement waste reduction strategies.

By leveraging data-driven insights, the project seeks to enhance inventory control, reduce wastage, and introduce efficient sales strategies. The expected outcome is a more profitable and sustainable business, with improved revenue and savings. This initiative not only benefits Mr. Muthuraj but also sets a framework for similar small vendors to adopt smarter business practices.

2 Proof of Originality of the Data

Business Name: Vegetable shop

Owner: Mr. Muthuraj A

Address: 7th Avenue, Ashok Nagar, Chennai, India

Proof of Originality: Letter

Video of the interaction with the Owner: Video

Fig 1. Myself with the shop owner

Fig 2. The Owner and his shop

Fig 3. Myself with the Shop Owner

3 Metadata

- Data Format: The data is stored in Excel format (XLSX)
- Data range: The data is available from January 01, 2025 to March 02, 2025
- Unit of Measurement: Quantity of the products in “Kilogram” (Kg), Price in Rupees (Rs)

The dataset contains 4 datasheets. Link

3.1 Vegetable_Price_Data
3.2 Sales_Data
3.3 Customer_Purchase_Data
3.4 Product_Id

3.1 Vegetable_Price_Data:

This dataset contains 1159 rows and 5 columns. It explains about the products being sold in the Wholesale market. The important features and its description for Vegetable Price Data datasheet are listed below.

- **Date:** This column indicates the specific date on which the price information for the vegetables were recorded.

- **Product_Id:** This column contains a unique identifier code assigned to each vegetable product for easy tracking and reference.

- **Product_Name:** This column specifies the name of the vegetable (e.g., "Tomato," "Carrot," "Onion").

- **Wholesale_Price(in Rs):** This column shows the price in Indian Rupees at which the vegetables are sold in bulk to retailers or other businesses.

- **Qty(in KG):** This column represents the quantity of the vegetable being considered, measured in kilograms.

3.2 Sales_Data:

This dataset contains 1159 rows and 8 columns. It explains about the products bought and sold by the vegetable seller. The important features and its description for Sales Data datasheet are listed below.

- **Date:** This column records the specific date of the transaction or stock movement.

- **Product_Id:** This column contains the unique identification number assigned to each product.

- **Product_Name:** This column specifies the name of the product being tracked.

- **Stock_Purchased(in KG):** This column indicates the amount of the product bought, measured in kilograms.

- **Stock_Sold(in KG):** This column shows the amount of the product sold, also measured in kilograms.

- **Selling_Price_per_Kg(in Rs):** This column represents the price at which one kilogram of the product was sold, in Indian Rupees.

- **Revenue_Generated(in Rs):** This column displays the total revenue earned from the sale of the product, in Indian Rupees.

- **Stock_remained(in KG):** This column shows the remaining quantity of the product after sales, measured in kilograms.

3.3 Customer_Purchase_Data:

This dataset contains 4490 rows and 8 columns. It explains about the customer behaviour in buying and selling of the products. The important features and its description for Customer_Purchase_Data datasheet are listed below.

- **Date:** This column records the specific date on which the customer made the purchase.

- **Customer_Id:** This column contains a unique identifier assigned to each customer for tracking their purchases.

- **Customer_Name:** This column displays the name of the customer who made the purchase.

- **Product_Id:** This column contains the unique identification number assigned to the product that was purchased.

- **Product_Name:** This column specifies the name of the product purchased by the customer.

- **Quantity_Purchased(KG):** This column indicates the amount of the product the customer bought, measured in kilograms.

- **Price_per_Kg(Rs):** This column represents the price the customer paid for one kilogram of the product, in Indian Rupees.

- **Total_Price(Rs):** This column shows the total amount the customer paid for the purchased quantity of the product, in Indian Rupees.

3.4 Product_Id:

This dataset contains 19 rows and 2 columns. The important features and its description for Project_Id datasheet are listed below,

- **Product_Id:** This column indicates the Id's for the products.

- **Product_Name:** This column indicates the name of each product.

4 Descriptive Statistics

a). Vegetable_Price_Data:

	Wholesale_Price(Rs)
count	1159.000000
mean	29.682183
std	11.731687
min	10.050000
25%	18.985000
50%	29.400000
75%	39.650000
max	49.970000

Fig 4. Descriptive Statistics for the "Vegetable_Price_Data" dataset.

The wholesale vegetable prices exhibit a relatively narrow range (Rs. 10 to Rs. 50), with the majority of prices clustered around the mean of Rs. 29.68 and a standard deviation of Rs. 11.73, indicating moderate variability.

b). Sales_Data:

	Stock_Purchased(KG)	Stock_Sold(KG)	Selling_Price_per_Kg(Rs)	Revenue_Generated(Rs)	Stock_remained_(KG)
count	1159.000000	1159.000000	1159.000000	1159.000000	1159.000000
mean	7.522002	5.240725	40.095298	209.472899	2.281277
std	1.722850	1.366967	15.878443	99.076728	1.392152
min	5.000000	4.000000	13.070000	54.560000	1.000000
25%	6.000000	4.000000	25.740000	130.185000	1.000000
50%	8.000000	5.000000	39.650000	197.920000	2.000000
75%	9.000000	6.000000	53.625000	266.280000	3.000000
max	10.000000	9.000000	69.430000	562.590000	6.000000

Fig 5. Descriptive Statistics for "Sales_Data" dataset

This dataset contains 1,159 records with stock, sales, pricing, and revenue details, showing a mean stock purchase of 7.52 KG, average sales of 5.24KG, a mean selling price of ₹40.00 per KG, and an average revenue of ₹209.47.

c). Customer_Purchase_Data:

	Quantity_Purchased(KG)	Price_per_Kg(Rs)	Total_Price(Rs)
count	4490.000000	4490.000000	4490.000000
mean	0.748831	40.183639	29.993726
std	0.354279	16.039843	19.321670
min	0.250000	13.070000	3.370000
25%	0.500000	25.570000	14.230000
50%	0.750000	39.650000	25.605000
75%	1.000000	54.000000	42.335000
max	1.250000	69.430000	86.480000

Fig. 6 Descriptive Statistics for "Customer_Purchase_Data" dataset

The descriptive statistics for the "Customer_Purchase_Data" reveal that the customer purchases are typically small (averaging under 1kg), with a wide price range (Rs. 13 to Rs. 69 per kg) and a correspondingly varied total purchase cost (Rs. 3.37 to Rs. 86.48).

5 Detailed Explanation of Analysis Process and Methods

To address the client's challenges of stagnant sales and vegetable wastage, a structured data-driven approach is necessary. The most appropriate method for this problem is Time-Series Analysis combined with Exploratory Data Analysis (EDA) and Inventory Optimization.

Time-series analysis is essential because vegetable sales exhibit daily and seasonal variations influenced by factors such as customer demand, pricing, and local market trends. By leveraging historical sales data, it is possible to identify demand patterns, predict future sales trends, and optimize procurement strategies accordingly. This method is superior to traditional manual tracking because it provides quantitative insights rather than relying on intuition. Furthermore, time-series forecasting enables dynamic adjustments in pricing and stock levels, ensuring a balance between supply and demand.

Exploratory Data Analysis (EDA) plays a crucial role in uncovering key business insights from the collected data. Using descriptive statistical techniques, we can analyze the variations in Stock_Purchased(KG), Stock_Sold(KG), Stock_remained(KG), Revenue_Generated(Rs), and Selling_Price_per_Kg(Rs) columns. A correlation analysis between selling price and sales volume can reveal pricing sensitivity, helping the client to determine optimal price points that maximize revenue without reducing customer demand. Additionally, outlier detection can

highlight abnormal sales patterns, such as unexpectedly high wastage on certain days, which may indicate inefficient stock management or external market fluctuations. These analytical insights support strategic decision-making, making EDA an indispensable method for improving sales performance.

Inventory optimization is another critical aspect of this project. Since vegetables are perishable, excess stock leads to financial losses due to spoilage. By analyzing historical sales and remaining stock data, we can calculate wastage rates and implement a procurement strategy that minimizes over-purchasing. One approach is demand-based stock purchasing, where stock levels are adjusted based on sales predictions rather than arbitrary ordering. This method is more effective than simple trend analysis because it accounts for daily fluctuations, seasonal demand, and external factors like festivals or weather conditions that impact purchasing behavior. As a result, this optimization approach reduces wastage while ensuring enough stock is available to meet customer needs.

In conclusion, Time-Series Analysis, EDA, and Inventory Optimization together form a robust methodology for solving the identified challenges. This structured approach allows Mr. Muthuraj to make data-backed decisions, improve revenue, and reduce wastage systematically. Compared to manual tracking and reactive decision-making, this method offers a proactive, predictive, and scalable solution, ensuring long-term business sustainability.

6 Results and Findings

6.1 Quantity of Stock remained for the months:

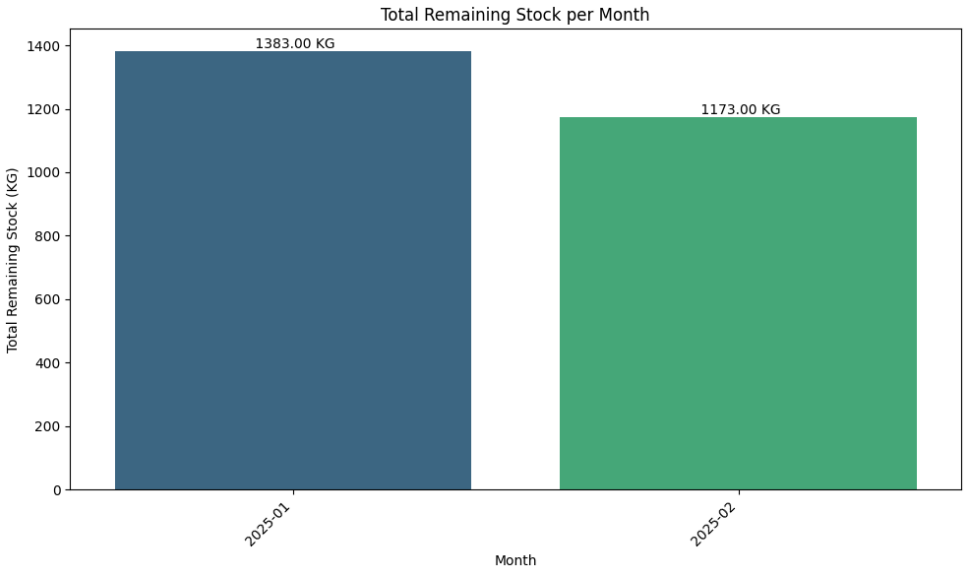

Fig. 7 Quantity of Stock remained for the months

This bar graph illustrates a rapid decline in stock quantity (in kilograms) for the months, January and February.

Key Observation:

- Significant Initial Stock: In January 2025, the stock quantity is substantial (1383 KG).
- Slight decrease in February: There's a drop in stock levels by February, with the quantity falling to approximately 1173 KG.
- High Demand or Sales: The rapid decrease could indicate high sales or demand for the product during this period.
- Limited Supply: There might have been limited replenishment of stock, leading to the depletion.

6.2 Revenue Generated over the period:

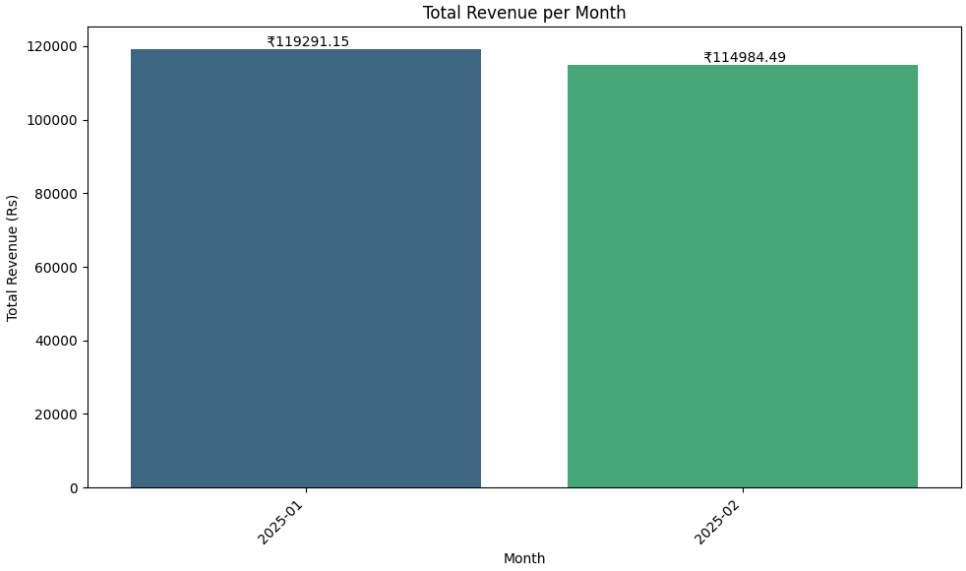

Fig. 8 Revenue Generated over the months

The bar chart displays the total revenue (in INR) generated in the months of January 2025 and February 2025**.**

Key Observations:

- In January 2025, the total revenue was ₹1,19,291**.**

- In February 2025, the revenue slightly increased to ₹1,14,984**.**
- This indicates a monthly revenue decrease of ₹4,307, from January to February.

6.3 Daily Profit Trend:

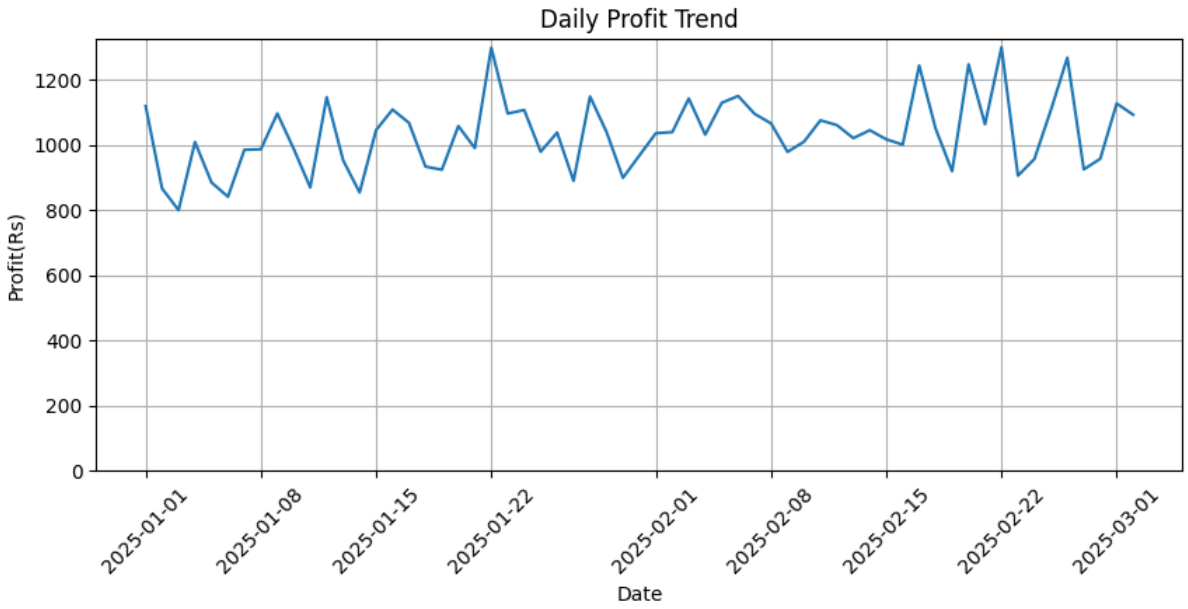

Fig 9. Daily Profit Trend

The daily profit exhibits regular fluctuations over the two-month period (Jan - Feb 2025).

Key Observation:

- Profits generally range between ₹900 and ₹1200, with a few days showing exceptional spikes above ₹1200.
- The overall pattern does not indicate a consistent upward or downward trend, suggesting stable yet variable daily profits.
- High-Profit Spikes: Occur notably around January 21 and February 21, reaching peak values.
- Low-Profit Dips: A few dips closer to ₹800 are seen in early January.
- Stable Period: A more stable profit pattern is observed during early to mid-February.