# Data-Driven Problem Solving for a Growing Trading and Distribution Business

**A Mid Term Report for the BDM Capstone Project**

Submitted by

Name: Abhinav Saxena

Roll number: 23F3003763


IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

# Contents

# 1 Executive Summary

This project focuses on Greenaway Enterprises LLP, a micro-enterprise based in Kanpur, Uttar Pradesh, established on 7th August 2019 as a Limited Liability Partnership (LLP). The firm operates in the B2B trading and distribution sector, supplying plastic, chemical, and rubber-based raw materials such as low-density polymer, calcium carbonate, toluene, and methyl ethyl ketone. These materials are primarily delivered to small-scale manufacturers through a just-in-time delivery model. The business is supported by a team of five employees responsible for managing day-to-day inventory operations. This project aims to support the business in addressing key operational challenges using data-driven insights to enhance long-term growth and stability.

In the initial phase of the project, three core challenges were identified: a high dependency on a limited number of clients, rising inventory carrying costs, and the impact of raw material price fluctuations on profit margins. To begin understanding these issues, relevant sales and purchase data was collected from the business. Preliminary data cleaning and high-level exploratory analysis were conducted, focusing on sales patterns, inventory movement, and pricing trends over time. Early indicators suggest that a significant portion of revenue depends on a few key clients, and inventory holding costs may be rising due to overstocking practices and unpredictable purchase timing.

Although detailed analysis and modeling are reserved for the final phase, these early insights provide a foundational understanding of the problems and opportunities. The next steps will involve deeper investigation into client dependency, inventory optimization opportunities, and raw material cost management. The final phase will conclude with a comprehensive report, practical recommendations, and a presentation of findings to support real-world decision-making and operational improvements.

# 2 Proof of Originality

This project is based on primary data collected directly from Greenaway Enterprises LLP, a B2B trading and distribution firm located in Kanpur, Uttar Pradesh. The business owner has provided permission to access and use the company's internal data for academic analysis.
To establish authenticity and originality, the following evidence is provided:

- Business details, including owner's name and registered business address.

- Photograph of the inventory storage area (included below).
- Scanned invoices and purchase documents (available in the drive link).
- A signed permission letter from the business owner (available in the drive link).
- Video footage of the interaction/interview with the owner explaining business operations and challenges. (available in the drive link).

*Image 1: GoDown Storage Showing Calcium Carbonate Powder Inventory*

**Supporting documents and media can be accessed here:**

https://drive.google.com/drive/folders/1LtVm7I8AmXdoW5azT9GW8SacTNi-x6ey?usp=sharing

# 3 Meta Data

Dataset Details:

- Data Type: Transactional data (Sales & Purchase)
- Format: Excel
- Duration Covered: April 2023 - December 2024

The cleaned and structured sales and purchase datasets used for this analysis are available in google drive link.

## 3.1 Sales Metadata

The sales metadata consists of key variables that track outward flow of raw materials to manufacturing clients. Below are the core fields and their relevance to problem statements:

| Sno. | Field | Description | Relevance to Problem |
|---|---|---|---|
| 1. | Date | Transaction date | Time-based analysis of sales patterns. |
| 2. | Material Name | Type of raw material sold | Helps analyze demand trends and material-wise profitability. |
| 3. | Party Name | Name of the purchasing client | Used to detect high dependency on a few clients. |
| 4. | Quantity (kg) | Amount sold in kilograms | Helps analyze inventory movement and matching with purchases. |
| 5. | Rate (Rs.) | Selling price per unit | Used to track sales pricing trends. |
| 6. | Basic Amt. (Rs.) | Pre-GST total value | Helps in profit calculations. |
| 7. | Total Amt. (Rs.) | Final invoice amount (with GST) | Cash inflow analysis. |

*Table 1. Sales Meta Data with Description and Relevancy to problem statements*

### 3.2 Purchase Metadata

The purchase metadata captures details of inbound transactions from suppliers and forms the foundation for analyzing procurement efficiency, raw material cost changes.

| Sno. | Field | Description | Relevance to Problem |
|---|---|---|---|
| 1. | Date | Purchase date | Enables monthly procurement analysis. |
| 2. | Material Name | Raw material purchased | To track cost per material over time. |
| 3. | Supplier Name | Vendor from whom material was procured | Can highlight over-reliance on specific suppliers. |
| 4. | Quantity (kg) | Inbound material in kilograms | Key for inventory flow and matching with sales. |
| 5. | Rate (Rs.) | Price per unit | Helps monitor input cost fluctuations. |
| 6. | Basic Amt. (Rs.) | Total pre-tax amount | Cash flow tracking. |
| 7. | Total Amt. (Rs.) | Total payment made including GST | Full cost calculation for procurement. |

*Table 2. Sales Meta Data with Description and Relevancy to problem statements*

# 4 Descriptive Statistics

## 4.1 Sales Data:

| Metric | Sales Quantity (kg) | Basic Amount (Rs) | Total Revenue (incl. GST) (Rs) |
|---|---|---|---|
| Total | 12,51,790.00 | ₹ 2,48,48,520.00 | ₹ 2,93,21,253.60 |
| Mean | 2,009.29 | ₹ 39,885.26 | ₹ 47,064.61 |
| Median | 2,000.00 | ₹ 37,500.00 | ₹ 44,250.00 |
| Min | 180.00 | ₹ 6,000.00 | ₹ 7,080.00 |
| Max | 6,000.00 | ₹ 3,25,000.00 | ₹ 3,83,500.00 |
| Standard Deviation | 1,153.12 | ₹ 43,811.48 | ₹ 51,697.55 |
| Range | 5,820.00 | ₹ 3,19,000.00 | ₹ 3,76,420.00 |

*Fig 1. Summary Statistics for Combined Sales Data (April 2023 - Dec 2024)*

| Metric | Sales Quantity (kg) | Basic Amount (Rs) | Total Revenue (incl. GST) (Rs) |
|---|---|---|---|
| Total | 5,59,870.00 | ₹ 1,06,68,160.00 | ₹ 1,25,88,428.80 |
| Mean | 2,073.59 | ₹ 39,511.70 | ₹ 46,623.81 |
| Median | 3,000.00 | ₹ 37,500.00 | ₹ 44,250.00 |
| Min | 180.00 | ₹ 6,000.00 | ₹ 7,080.00 |
| Max | 6,000.00 | ₹ 3,25,000.00 | ₹ 3,83,500.00 |
| Standard Deviation | 1,169.22 | ₹ 43,821.13 | ₹ 51,708.94 |
| Range | 5,820.00 | ₹ 3,19,000.00 | ₹ 3,76,420.00 |

*Fig 2. Stats for the Year 2023*

| Metric | Sales Quantity (kg) | Basic Amount (Rs) | Total Revenue (incl. GST) (Rs) |
|---|---|---|---|
| Total | 6,91,920.00 | ₹ 1,41,80,360.00 | ₹ 1,67,32,824.80 |
| Mean | 1,960.11 | ₹ 40,170.99 | ₹ 47,401.77 |
| Median | 2,000.00 | ₹ 37,500.00 | ₹ 44,250.00 |
| Min | 180.00 | ₹ 6,000.00 | ₹ 7,080.00 |
| Max | 5,000.00 | ₹ 3,25,000.00 | ₹ 3,83,500.00 |
| Standard Deviation | 1,138.21 | ₹ 43,801.95 | ₹ 51,686.30 |
| Range | 4,820.00 | ₹ 3,19,000.00 | ₹ 3,76,420.00 |

*Fig 3. Stats for the Year 2024*

The sales data from Greenaway Enterprises LLP was analyzed for the combined period of FY 2023 and 2024, as well as separately for each year. The combined sales quantity over this duration stood at 12,51,790 kg, generating a basic amount of Rs. 2.48 crore and total revenue (including GST) of Rs. 2.93 crore.

In FY 2023, the sales volume was 5,59,870 kg with a total revenue of Rs. 1.25 crore, while FY 2024 witnessed a higher volume of 6,91,920 kg and revenue of Rs. 1.67 crore, indicating significant business growth.

Both the average sales volume and revenue value remained consistent over the years, although 2024 saw a slight decrease in the average volume. The range and standard deviation in quantity and revenue shows moderate variability in order sizes and client demand.

## 4.2 Purchase Data:

| Metric | Purchase Qty (kg) | Purchase Amount (Rs) |
|---|---|---|
| Total | 14,03,660.00 | ₹ 2,17,59,330.00 |
| Mean | 13,496.73 | ₹ 2,09,224.33 |
| Median | 10,000.00 | ₹ 2,51,250.00 |
| Min | 180.00 | ₹ 5,300.00 |
| Max | 25,000.00 | ₹ 9,06,750.00 |
| Standard Dev. | 10,939.93 | ₹ 1,87,821.94 |
| Range | 24,820.00 | ₹ 9,01,450.00 |

*Fig 4. Combined Purchase Data Stats (2023 - 2024)*

| Metric | Purchase Qty (kg) | Purchase Amount (Rs) |
|---|---|---|
| Total | 6,58,240.00 | ₹ 95,78,470.00 |
| Mean | 13,164.80 | ₹ 1,91,569.40 |
| Median | 9,000.00 | ₹ 2,51,250.00 |
| Min | 180.00 | ₹ 16,740.00 |
| Max | 25,000.00 | ₹ 8,70,000.00 |
| Standard Dev. | 11,000.58 | ₹ 1,69,894.43 |
| Range | 24,820.00 | ₹ 8,53,260.00 |

*Fig 5. Stats for Year 2023*

| Metric | Purchase Qty (kg) | Purchase Amount (Rs) |
|---|---|---|
| Total | 7,45,420.00 | ₹ 1,21,80,860.00 |
| Mean | 13,804.07 | ₹ 2,25,571.48 |
| Median | 12,500.00 | ₹ 2,52,500.00 |
| Min | 180.00 | ₹ 5,300.00 |
| Max | 25,000.00 | ₹ 9,06,750.00 |
| Standard Dev. | 10,874.44 | ₹ 2,01,641.19 |
| Range | 24,820.00 | ₹ 9,01,450.00 |

*Fig 6. Stats for Year 2024*

Descriptive figures for purchasing data in 2023 and 2024 provide valuable insight into the raw material purchase trends of Greenway Enterprises LLP. The combined purchase volume totaled 14,03,660 kg, with a corresponding spend of Rs. 2.17 crore, showing the firm's scale of operations. The mean purchase quantity per transaction was around 13,496.73 kg, with noticeable variance as indicated by a high standard deviation of 10,939.93 kg.

When broken down by year, 2024 showed a marginally higher mean purchase quantity (13,804.07 kg) and spend, reflecting a scale-up in operations compared to 2023. Consistency is observed in median values (~10,000 - 12,500 kg), suggesting regularity in ordering patterns, despite the wide range in volumes and amounts. These statistics help identify cost instability and can guide the inventory plan and sellers negotiations to customize the purchase strategies.

# 5 Detailed Explanation of Analysis Process/Method

### Overview

This section outlines the structured approach adopted to begin addressing the three identified business challenges using the available sales and purchase datasets. The aim at this mid-term stage is to clean the data, explore initial trends, and set up a roadmap for advanced analysis in the final report.

## 5.1 Data Cleaning and Preprocessing

To ensure consistency and accuracy across datasets, both sales and purchase data for 2023 and 2024 were first merged using Power Query in Excel. Key columns such as invoice numbers, which contained missing values, were removed. Data types for monetary and quantity fields like Rate, Value, GST, and Sale Value were standardized to decimal format.

In the sales data, inconsistent and incorrect date formats were identified and corrected, including resolving mixed MM/DD and DD/MM entries and updating miswritten years (e.g., 2023 marked as 2024). Material and party names were cleaned for uniformity.

For purchase data, similar issues were addressed, including date format corrections (e.g., '3034' instead of '2024') and standardization of supplier names with inconsistent spellings. Additional derived columns like Year and Month Name were created to facilitate temporal analysis. These preprocessing steps were critical in establishing a reliable foundation for meaningful and error-free analysis in subsequent stages.

### 5.2 Initial Analytical Approach with Justification

## A. Foundational Analysis

**1. Descriptive Statistics**: Core statistical metrics such as Total, Mean, Median, Minimum, Maximum, Standard Deviation, and Range were calculated for important numerical fields, including Quantity (kg), Basic Amt (Rs), and Total Sale Value (Rs) in the sales data, and equivalent fields in the purchase data.

**Justification**: These metrics help in understanding the central tendency and spread of the data. For example, high standard deviation or range indicates significant variability, reflecting uncertain purchases or sales patterns, value fluctuations, or customer-specific order behavior.

**2. Time Series Aggregation:** Monthly and quarterly aggregations of both sales and purchase data were created using Pivot Tables grouping features. Key metrics such as total monthly Quantity, Sales Value, and Purchase Spend were computed to visualize and compare performance over time.

**Justification**: Aggregating data by time allows for early detection of seasonal patterns, demand fluctuations, and purchasing cycles. For instance, a consistent rise in purchase volume without a corresponding increase in sales may signal inventory build-up.

## B. Problem-wise Preliminary Analysis Plan

**1. High Dependency on a Few Clients**

**Method Used:** Pareto Analysis (80/20 Rule)

For this purpose, relevant sales data was collected and cleaned, containing essential fields such as Date, Material Name, Party's/Customer Name, Sale Value (Rs) etc. These columns allow the calculation of total revenue per client, enabling the ranking of clients based on contribution.

This will help identify whether the business depends too heavily on a few clients, thereby guiding strategies for diversification in the final analysis.

**Justification**: Pareto analysis is ideal in identifying whether a small group of clients is contributing to most of the sales revenue. This is more appropriate than cluster analysis or regression at this stage because it gives clear, immediate insight into client contribution, without requiring predictive modeling.

**2. Rising Inventory Carrying Costs**

**Method Used:** Monthly Comparison of Purchase & Sales Data

To evaluate potential inventory build-up, monthly totals of Purchase Quantity and Sales Quantity were calculated separately using Pivot Tables. An approximate ending inventory for each month was also derived using the following running formula in Excel:

Ending Inventory = Previous Month Inventory + Current Month Purchase - Current Month Sales

**Justification**: Monthly aggregation is effective for spotting mismatches between purchased and sold quantities. It provides early indications of over-purchasing or slow-moving materials. Other techniques like EOQ (Economic Order Quantity) or inventory turnover ratios will be considered in the final report.

**3. Fluctuating Raw Material Prices Impacting Profit Margins**

**Method Used:** Trend Analysis of Purchase and Sale Rates

To assess how raw material cost fluctuations impact profitability, I computed the monthly average purchase rate and monthly average sale rate per kg in Excel. This was done using Pivot Tables by aggregating “Rate (Rs)” from both purchase and sales data across each month for both 2023 and 2024 separately.

**Justification:** This method is appropriate for the mid-term phase as it provides a clear visual and numerical comparison of input and output pricing trends over time. Unlike regression or margin modeling, this trend-based approach is straightforward and immediately highlights potential problem periods. More advanced analysis like profit margin variance, cost-volume-profit (CVP) analysis, and material-specific pricing trends will be explored in the final report.

# 6 Results and Findings

The following section highlights initial findings derived from the foundational and problem-specific analyses.

### 1. Time Series Trend - Sales and Purchase Patterns:

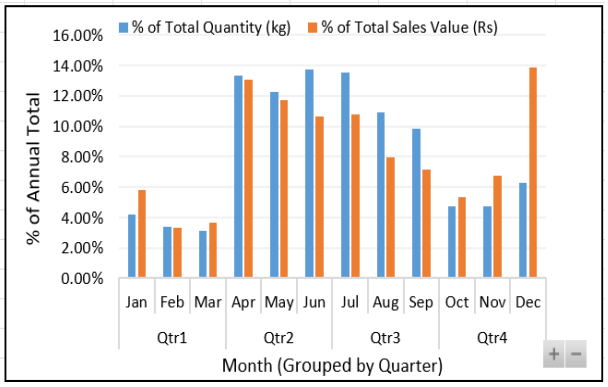


*Fig7. Monthly Contribution to Sales Quantity and Value (%)*

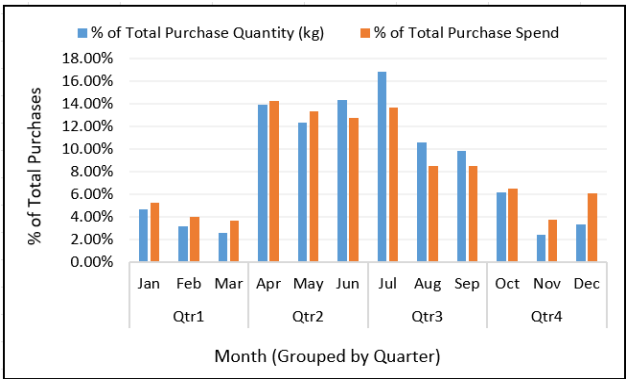


*Fig8. Monthly Contribution to Purchase Quantity and Value (%)*

**Sales Data Insight**: April to July (Q2-early Q3) shows peak sales volume and value (~13-14% per month), indicating a seasonal demand surge, November to February (Q4-Q1) records lower sales, possibly due to off-season demand or supply slowdowns.

**Purchase Data Insight:** Purchases also peak in April to July, aligning with high sales. July sees the highest purchase spike (16.5%), likely for pre-stocking, Q4 shows minimal purchases, possibly to reduce inventory costs.

### 2. High Dependency on a Few Clients (Pareto Analysis):

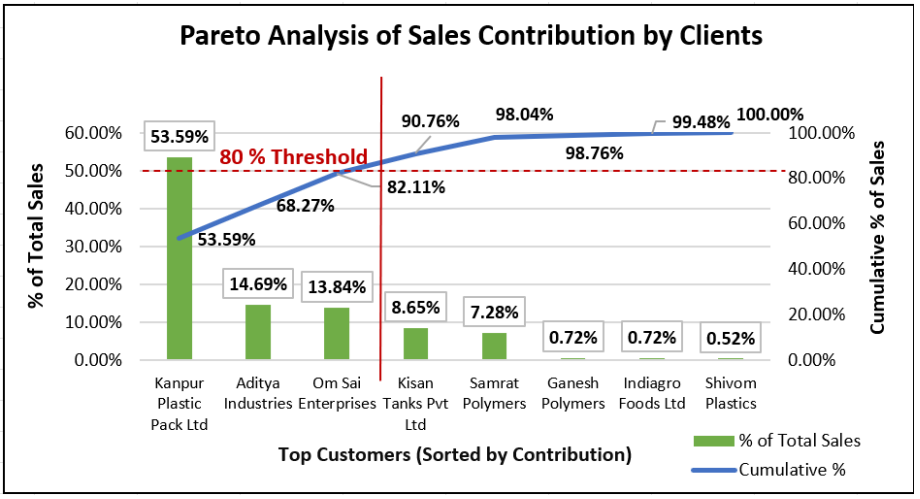


*Fig 9. Pareto Chart of Client-wise Sales Contribution (% of Total Sales and Cumulative %)*

- The Pareto analysis reveals that **approximately 80% of the total sales revenue is generated by just 3 out of 9 key clients**. This concentration poses a business risk if any major client disengages.
- **Remaining 6 clients** contribute only **~18%** collectively - a long tail with minimal individual impact.
- Heavy reliance on a single client (Kanpur Plastic Pack Ltd) - accounts for **over half (53.59%)** of total sales.

### 3. Rising Inventory Carrying Costs - Trend-Based Observation:

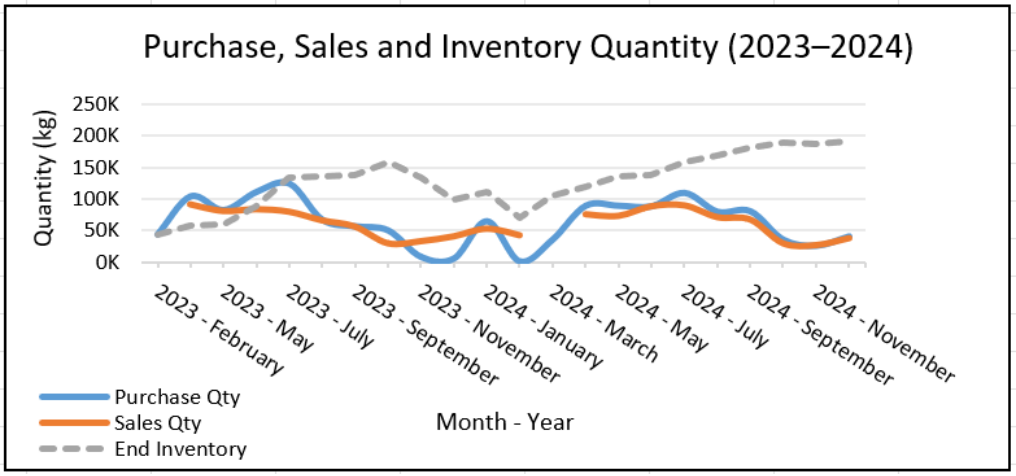


*Fig 10. Inventory Accumulation Due to Purchase-Sales Mismatch (2023-2024 Monthly Trend)*

Note: Sales data is available from April 2023 onwards; purchase data starts from February 2023

- Consistent Excess Inventory: **Rising Carrying Costs** - there's a visible trend of continuous Excess Inventory, with Purchase Qty exceeding Sales Qty.
- The grey **End Inventory** line shows a **steady upward trend**, especially from mid-2023 to late 2024, despite fluctuations in purchase and sales.
- Inventory decisions are likely **not aligned with demand**, leading to excess stock and potential wastage or storage burden.

**4. Fluctuating Raw Material Prices Impacting Profit Margins:**

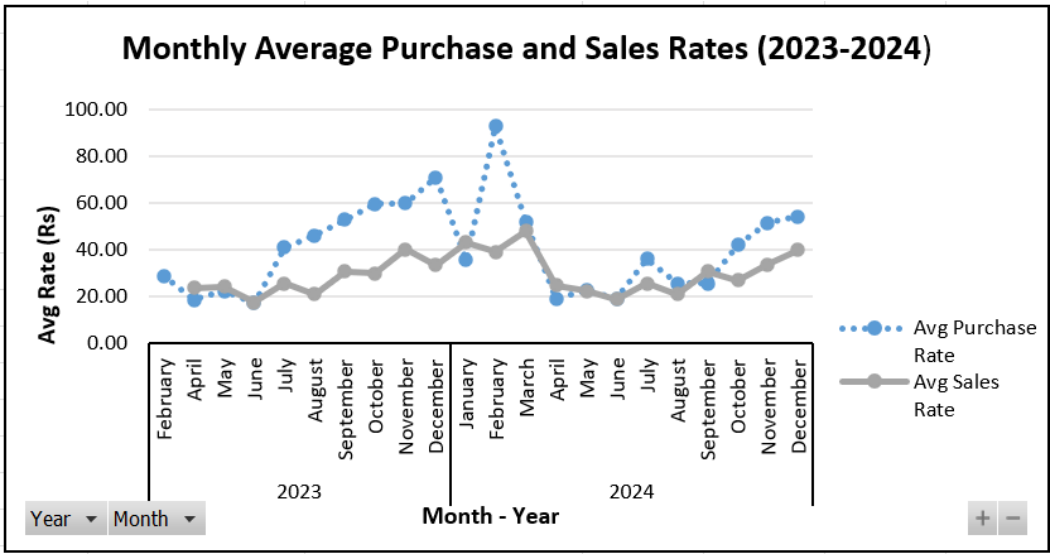


*Fig. 11 Monthly Average Purchase and Sales Rates (2023-2024)*

- Across most months, **sales rate (grey)** is **lower than purchase rate**, suggesting potential **profit margin pressure** or inefficient pricing strategy.
- The **Avg Purchase Rate** (blue) shows sharp peaks (notably **Jan-Feb 2024**), while sales rate is more stable.
- The cost of raw materials keeps changing a lot, but the selling prices are not adjusted in time - this is reducing the profit the business can make.
- This needs further deep-dive into material-wise margin trends to identify specific high-risk materials in the final analysis.

These insights represent preliminary trends and problem diagnostics. Deeper modeling and recommendation-based conclusions will be presented in the final report.