

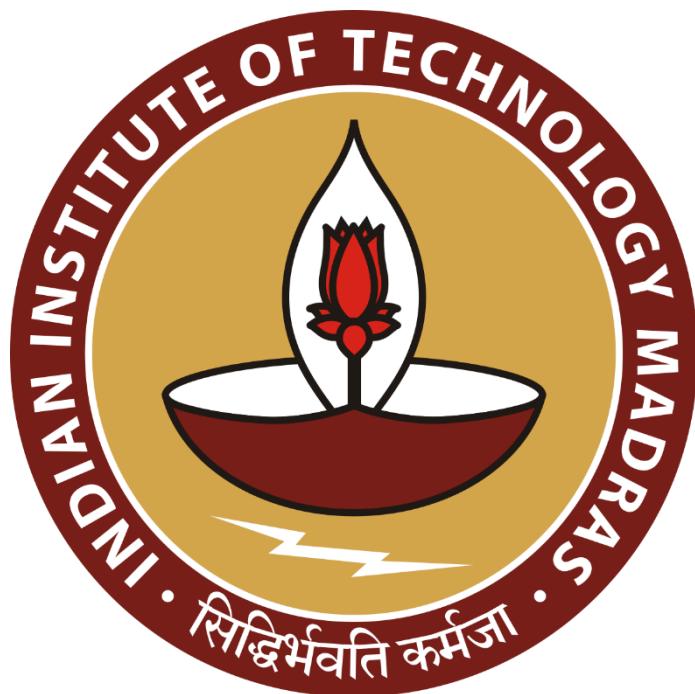
# **Optimized Business Data Management for Strategic Growth at Arihant Enterprises**

**A Mid-Term report for the BDM capstone Project**

Submitted by

Name: Deepesh Kumar Dawar

Roll number: 22f2001290



IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

## **Contents**

1. Executive Summary and Title	2
2. Proof of Originality of Data	2
3. Metadata	5
4. Descriptive Statistics	6
5. Detailed Explanation and Analysis	8
6. Results and Findings	10

## 1. Executive Summary and Title

Safai Ghar, operating in the healthcare and hygiene sector, recorded a total turnover of ₹4,29,63,849.40 between January 1, 2024, and February 15, 2025. Despite steady business activity, the company faces several operational challenges, including inventory inefficiencies, unbalanced product pricing, supplier dependency, and delayed receivables from institutional clients. While high-volume products such as 3 PLY masks dominate sales, they contribute relatively less to overall revenue due to suboptimal pricing. Simultaneously, a few low-volume, high-value products drive substantial revenue but are not prioritized in inventory planning.

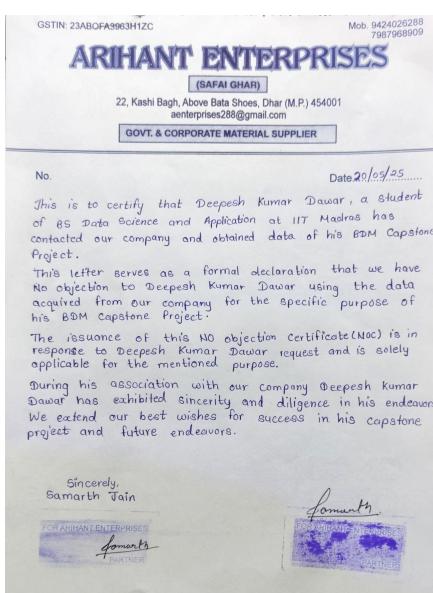
To tackle these inefficiencies, a data-driven methodology was adopted. Initial data cleaning and tabular analysis were carried out using Microsoft Excel, while Python (with libraries such as pandas, matplotlib, and seaborn) enabled deeper statistical analysis and visualizations. The integration of tools like Power BI or Tableau was proposed to create dynamic dashboards for real-time decision-making. These methods ensured that all insights were both reliable and actionable.

Strategically, the business must re-evaluate its pricing models, align stock levels with product performance, and introduce predictive inventory management. On the procurement side, the data showed a heavy reliance on a few suppliers, notably Satol Chemicals and Saify Healthcare. This dependency reduces bargaining power and increases risk, highlighting the need for vendor diversification and volume-based negotiations.

By implementing these data-backed strategies, Safai Ghar can optimize inventory, reduce procurement costs, improve liquidity through tighter credit control, and ultimately foster long-term growth and financial stability.

## 2. Proof of Originality of Data

### 1. Letterhead:



2. Business Location:

[https://www.google.com/maps/place/Safai+ghar/@22.6024217,75.3105213,63m/data=!3m1!1e3!4m6!3m5!1s0x396237b445330d75:0xcb8fbb5d2b897401!8m2!3d22.60244!4d75.310532!16s%2Fg%2F11fwswymz?entry=ttu&g\\_ep=EgoyMDI1MDUyMS4wIKXMDSoASAFQAw%3D%3D](https://www.google.com/maps/place/Safai+ghar/@22.6024217,75.3105213,63m/data=!3m1!1e3!4m6!3m5!1s0x396237b445330d75:0xcb8fbb5d2b897401!8m2!3d22.60244!4d75.310532!16s%2Fg%2F11fwswymz?entry=ttu&g_ep=EgoyMDI1MDUyMS4wIKXMDSoASAFQAw%3D%3D)

3. Images:



4. Video:

<https://drive.google.com/drive/folders/1kAbQakDLSXywITIPQs3kKoy8c6W-Xbpe?usp=sharing>

5. Original Data Provided by Company:

[https://drive.google.com/drive/folders/1OPbX1b\\_ZFheEzINDCz5AQ80EeO2Pn31g?usp=sharing](https://drive.google.com/drive/folders/1OPbX1b_ZFheEzINDCz5AQ80EeO2Pn31g?usp=sharing)

6. Data made by me for further analysis:

[https://docs.google.com/spreadsheets/d/1Jg\\_dE-qwCxXNHKev7bLKN\\_eYq5r08sMV8Z\\_f7SqE-3Y/edit?usp=sharing](https://docs.google.com/spreadsheets/d/1Jg_dE-qwCxXNHKev7bLKN_eYq5r08sMV8Z_f7SqE-3Y/edit?usp=sharing)

### 3. Metadata

The business uses a software to maintain all its transactions, so from it the business owner provided me with the one year data, listed as below:

- Purchase Register:** The owner provided me with the One year Purchase Transaction held within the company, for the Time period of January 2024 to February, 2025. This information was provided by an Excel sheet containing below details as Column Headers.

Column Name	Description
Invoice No.	A unique number assigned to each invoice for identification and tracking.
Invoice Date	The date on which the invoice was issued.

Supplier	The business entity providing the goods.
Product	The name or description of the product sold.
Quantity	The number of units of the product purchased.
Rate	The cost of one unit of the product.
Amount	The total cost before applying any discount or tax (Quantity × Rate).
Discount	The reduction in price offered by the supplier.
GST	The Goods and Services Tax applied to the transaction based on applicable tax rates.
Net	The final payable amount after applying discounts and adding GST (Amount - Discount + GST).

and other relevant columns.

- 2. Sales Summary(Bill Wise):** The owner provided me with the One year Purchase Transaction held within the company, for the Time period of January 2024 to February, 2025. This information was provided by an Excel sheet containing below details as Column Headers.

Column Name	Description
Party	Name of the customer involved in the transaction
Bill No.	A unique identifier assigned to the bill
Bill Date	The date on which the bill was issued
Cash/Credit	Indicates whether the transaction was paid in cash or on credit
Bill/Est	Specifies if the document is a final bill or just an estimate
Gross	Total amount before applying any discounts
Discount	Reduction applied to the gross amount
Amount	Final amount payable after discount

and other relevant columns.

**3. Sales Summary(Customer Wise):** The owner provided me with the One year Purchase Transaction held within the company, for the Time period of January 2024 to February, 2025. This information was provided by an Excel sheet containing below details as Column Headers.

Column Name	Description
Party	Name of the customer involved in the transaction
Bill No.	A unique identifier assigned to the bill
Bill Date	The date on which the bill was issued
Cash/Credit	Indicates whether the transaction was paid in cash or on credit
Bill/Est	Specifies if the document is a final bill or just an estimate
Gross	Total amount before applying any discounts
Discount	Reduction applied to the gross amount
Amount	Final amount payable after discount

and other relevant columns.

**4. Sales Summary(Product Wise):** The owner provided me with the One year Purchase Transaction held within the company, for the Time period of January 2024 to February, 2025. This information was provided by an Excel sheet containing below details as Column Headers.

Column Name	Description
Product Group	Category under which the product falls(e.g., electronics, stationery).
Product	Specific item or product name being sold.
Unit	The measurement unit used for the product (e.g., pieces, kg, liters).
Total Quantity	The total number of units sold across all entries.
Total Amount	The overall value of the total quantity of the product.

## 4. Descriptive Statistics

Description Statistic Measure	Description Statistic Definition
Count	The total number of observations
Sum	The total of all the values
Mean	The average of all the values
Standard Deviation	A measure of how spread out the data points are from the mean
Median	The middle value when the data is ordered
Minimum	The smallest value in the dataset
Maximum	The largest value in the dataset.

### Overall Sales Data - Product Wise(January 2024 - February 2025) :

Description Statistic Measure	Quantity	Total Revenue
Count	6760	6760
Sum	837087	42964299.40
Mean	123.82956065088759	6355.665591715977
Standard Deviation	2732.899074142871	23219.759550245584
Median	8	1112.25
Minimum	1	70
Maximum	176349	614142

The company collaborates with multiple suppliers for procurement of essential goods and chemicals. The suppliers listed below are selected based on their **purchase volume and value**, and form the **core group** of vendors used for strategic analysis.

A short **descriptive statistical summary** of these top suppliers is as follows:

<b>Supplier: SATOL CHEMICALS</b>		
<b>Description Statistic Measure</b>	Quantity	Amount
Total Transactions	146	146
Total Purchased	3655.0	₹634,040.01
Mean	25.03	₹4342.74
Standard Deviation	43.52	₹4758.87
Minimum	2.0	₹454.06
Maximum	300.0	₹29700.0

<b>Supplier: SAIFY HEALTHCARE AND MEDI-DEVICES (INDIA) PRIVATE LIMITED</b>		
<b>Description Statistic Measure</b>	Quantity	Amount
Total Transactions	47	47
Total Purchased	826	₹548,732.86
Mean	17.57	₹11675.17
Standard Deviation	23.54	₹19093.28
Minimum	1.0	₹113.16
Maximum	100.0	₹103775.1

<b>Supplier: PUSHKAR AGENCIES</b>		
<b>Description Statistic Measure</b>	Quantity	Amount
Total Transactions	332	332
Total Purchased	5191.0	₹541,271.59
Mean	15.64	₹1630.34

Standard Deviation	15.16	₹2962.65
Minimum	1.0	₹108.0
Maximum	120.0	₹34540.0

## 5. Detailed Explanation and Analysis

The analytical process undertaken for this project on Safai Ghar's one-year sales and purchase data was driven by a structured combination of quantitative methods, data visualization, and business-contextual interpretation. The primary aim was to uncover actionable insights regarding product performance, supplier dependencies, cash flow bottlenecks, and inventory management challenges. To achieve this, the analysis relied heavily on exploratory data analysis (EDA) techniques using Python (primarily pandas and matplotlib) and Excel pivoting and filtering functionalities to derive trends, correlations, and anomalies from the raw data.

### Data Cleaning and Structuring

The analytical process began with the compilation of raw datasets from multiple sources, including product-wise sales summaries, customer-wise and bill-wise sales records, and a supplier-wise purchase register. These datasets were initially inconsistent and required extensive cleaning and restructuring. Using a combination of Python's Pandas library and Microsoft Excel, redundant headers and empty rows were systematically removed. Column names, which varied across files, were standardized into uniform formats such as "Qty", "Amount", "Supplier" and "Customer" to enable cross-sheet compatibility. Missing or null values were addressed through filtering techniques, while negative quantities and entries with obvious errors were eliminated to maintain the integrity of the dataset. The entire cleaning process was guided by a step-by-step validation of row-level entries using conditional filters and custom scripts to log and trace corrections. These preprocessing steps were crucial in ensuring that subsequent analyses were based on reliable and accurate data.

### Inventory and Product-Level Analysis

To evaluate product performance and uncover inefficiencies in the inventory strategy, product-level statistics were aggregated. Metrics such as total quantity sold, total revenue generated, mean sale price, standard deviation, minimum, and maximum values were computed. These calculations were performed using grouped aggregations in Pandas (groupby, agg, describe), allowing for easy segmentation by product ID and name. The analysis uncovered a notable trend: several products with high sales volumes contributed relatively little to overall revenue, signaling potential underpricing or thin profit margins. On the other hand, some low-volume products yielded substantial revenue, indicating premium pricing or higher profit margins. These trends were visualized using bar plots and ranking techniques to identify the top-selling products by quantity, the highest revenue-generating items, and those performing strongly in both areas. Visualization libraries

such as seaborn and Excel charts were employed to represent findings in an intuitive format. This disparity between volume and value emphasized the need for a data-driven pricing model and strategic inventory planning to enhance profitability.

## **Supplier Dependency Analysis**

The supplier-wise purchase register was analyzed to understand procurement patterns and vendor concentration risks. The data was grouped by supplier, and key statistics—including total quantity purchased, total expenditure, mean transaction size, standard deviation, minimum, and maximum values—were calculated. This was executed using aggregation functions in Pandas, with visual aids such as supplier-wise bar graphs and pie charts illustrating the distribution of purchase values. This analysis identified the top ten suppliers by purchase value, with Satol Chemicals and Saify Healthcare and Medi-Devices (India) Pvt. Ltd. emerging as dominant procurement partners. These suppliers accounted for a significant portion of the total purchase volume and value, which suggests a dependency that could pose a supply chain risk. Further, Z-score analysis was considered to identify outlier transactions and evaluate whether certain suppliers had unusually large or small order volumes. The business's reliance on a few key suppliers highlights the importance of diversifying vendor partnerships to reduce potential disruptions and negotiate better terms.

## **Customer and Cash Flow Insights, Temporal Trends, and Visualization**

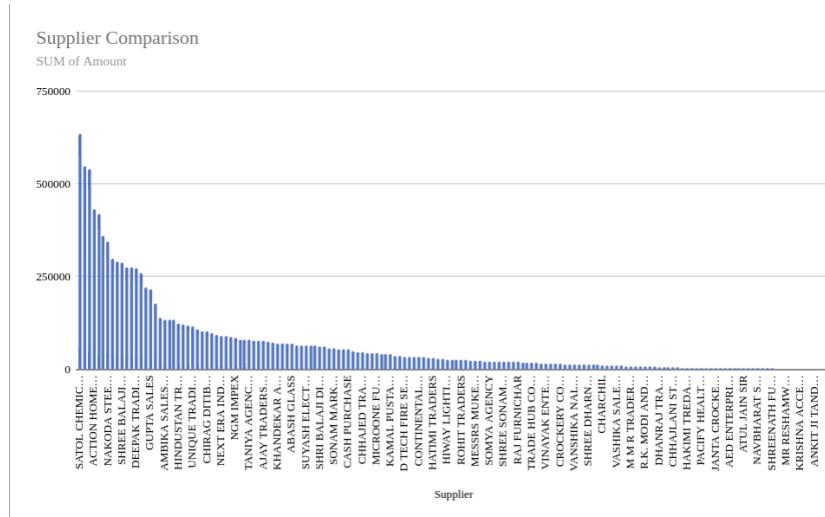
An analysis of customer-wise and bill-wise sales data revealed that institutional buyers often delay payments, leading to cash flow strain. The data was sorted and grouped by customer and bill date, followed by the computation of total due amounts and average payment delays. Although month-wise trends weren't directly available, the dataset supports monthly aggregation, enabling time-series analysis to identify seasonal demand patterns. This was achieved by converting transaction dates into datetime formats and using resample('M') in Pandas to evaluate monthly patterns. This can help align procurement with actual needs and avoid stock imbalances. Throughout the project, tools like Python and Excel were used to generate charts and pivot tables, simplifying insights and aiding quick decision-making. Heatmaps, line graphs, and temporal plots provided clarity on sales cycles, while cash flow gaps were visualized with cumulative payment and due trackers.

## 6. Results and Findings

### 6.1 Insights from Purchase Data : Supplier Dependence

Some insights gained from the Purchase data:

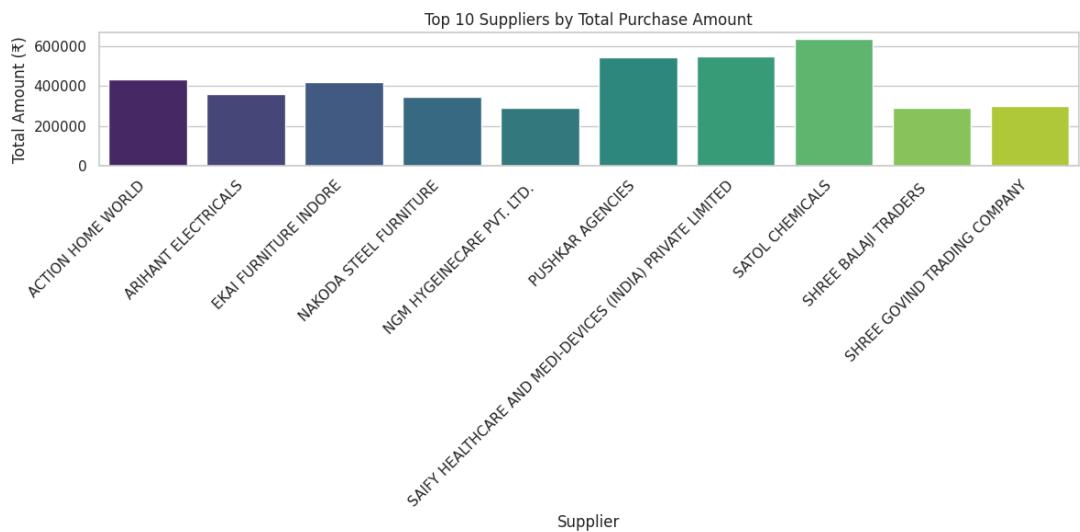
The total purchase amounts distributed across a wide range of suppliers. The data is highly **right-skewed**, with a small number of suppliers accounting for a significant portion of the total purchases.



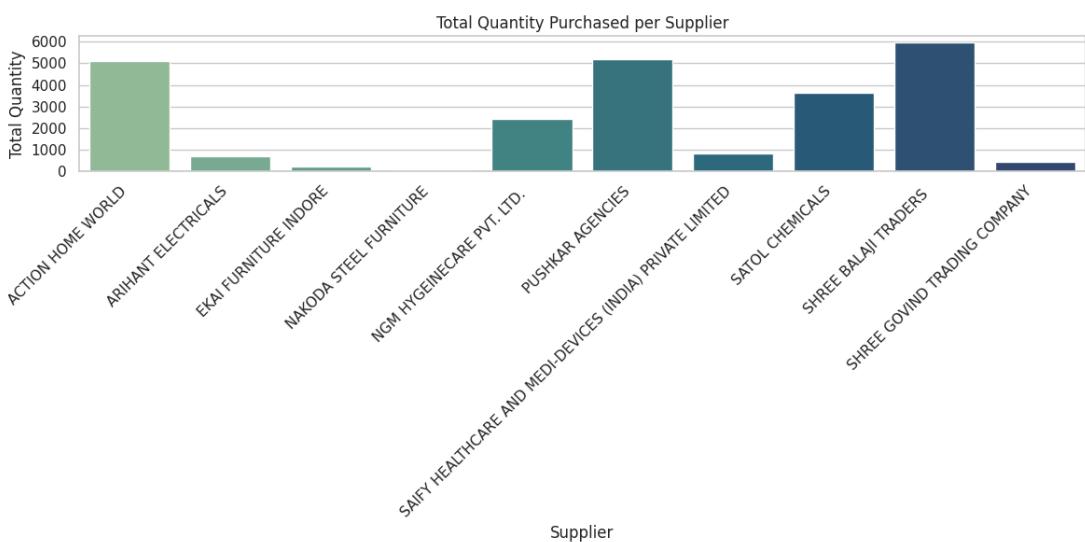
Graph 6.1.1

Top 5 suppliers, including "SATOL CHEMICALS", "ACTION HOME WORLD", "SAIFY HEALTHCARE AND MEDI-DEVICES (INDIA) PRIVATE LIMITED", and "PUSHKAR AGENCIES", contribute disproportionately to the overall procurement value, each with purchase amounts above ₹500,000. The **top 20 suppliers** dominate the expenditure landscape, likely contributing to **over 70% of total purchase spend**, which suggests a **Pareto-like (80/20) distribution** in supplier spending.

A long tail of suppliers with very low purchase amounts indicates low-frequency or low-volume procurement, which could be due to specialized items or one-off purchases. This concentration suggests that vendor management strategies should be focused on a few high-value suppliers, possibly through better negotiation, contracts, or partnerships to optimize procurement costs.



Graph 6.1.2



Graph 6.1.3

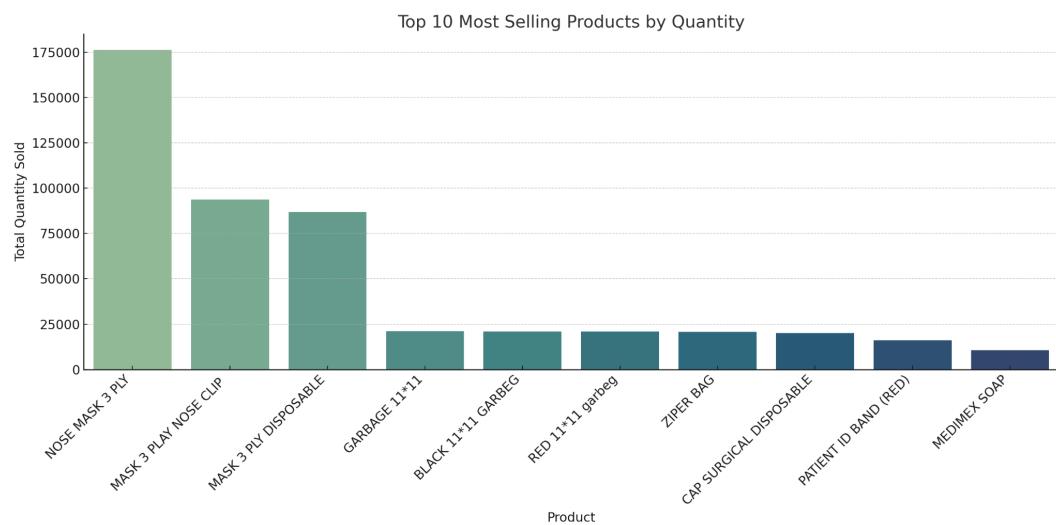
Volume vs. Value Balance: Some vendors provide fewer but more expensive items (like medical equipment or chemicals), while others supply low-cost, high-frequency goods. Both are crucial in sustaining daily operations and long-term infrastructure.

The quantity-wise chart shifts the focus. Shree Balaji Traders and Action Home World emerge as leaders in terms of total units purchased. This suggests that although their individual product prices may be lower, they provide essential consumables or frequently used items. Pushkar Agencies also stands out with both high quantity and amount, indicating both value and volume procurement—possibly making it a strategic supplier.

Procurement Strategy Diversification: The data reflects a diversified vendor strategy. Suppliers are not just chosen for cost but also for value and volume, optimizing operational needs.

## 6.2 Insights from Sales Data: Inventory & Pricing Strategy

A detailed analysis of the sales data reveals significant patterns in product performance, pricing, and inventory distribution. One of the most striking observations is the disproportionate contribution of certain products to total sales volume.

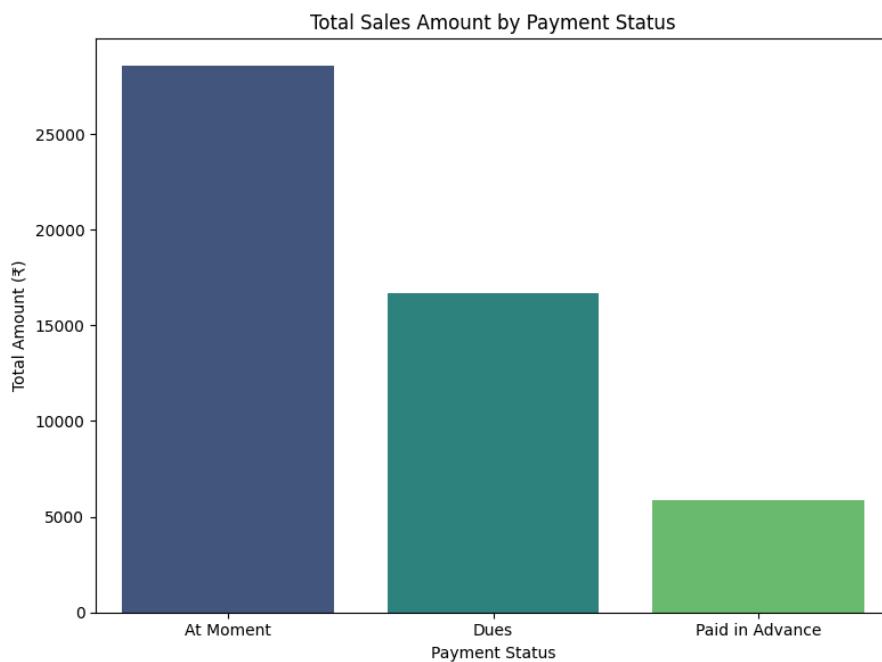


Graph 6.2.1

The chart shows that **masks dominate sales by quantity**, with the rest of the items trailing significantly. This indicates a **heavy reliance on a few key products**, which can be useful for inventory planning and sales focus. **NOSE MASK 3 PLY** is the clear leader, with over **175,000 units sold**, contributing a significantly higher volume than any other product. The top three products — all variations of **3 PLY MASK** — dominate the chart, indicating **strong and consistent demand for protective masks**.

## 6.3 Insights from Sales Data: Cash Flow Challenges

Visual data analysis further confirms the risks associated with current credit practices. While a majority of the transactions are processed as cash sales, it is the credit-based transactions, though fewer in volume, that contribute the most to the business's revenue. These credit-based invoices are predominantly associated with institutional clients and carry higher individual values. Unfortunately, a significant portion of these remain unpaid for extended periods. Meanwhile, only a small number of customers pay in advance, revealing an opportunity for the business to renegotiate terms that encourage upfront payments or at least partial advance settlements.



Graph 6.3.1

The financial strain from delayed receivables is beginning to show tangible operational consequences. The business has been unable to invest in time-sensitive inventory restocking or capitalize on early payment discounts offered by key suppliers. This liquidity crisis, caused not by poor sales but by delayed collections, is an urgent issue that needs to be addressed.