Demand Forecasting for a Screen-Printing Company

A Mid-Term report for the BDM capstone Project

Submitted by

Name: RITEESH TM

Roll number: 23f3002667

IITM

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1. EXECUTIVE SUMMARY

This mid-term report presents the progress of the Business Data Management (BDM) project conducted for Sri Meenakshi Screen Printing, a Thamboolam bag manufacturer based in Madurai. The analysis is based on 500 historical order records (2022–2025), as verified in the Proof of Originality section. The project aims to analyze customer demand trends, identify key clients, and understand regional performance using structured business data.

The Metadata section outlines the dataset's structure, features, and sources, detailing key fields such as order dates, client type, bag specifications, and quantities. Data cleaning and preprocessing were performed to standardize dates, normalize text fields, remove duplicates, and ensure numerical validity, producing a reliable dataset for analysis.

The Descriptive Statistics section explores order behaviour through measures of central tendency and dispersion (mean, median, mode, standard deviation, skewness) across both client categories and regional segments. This analysis performed using Excel and Python (pandas, matplotlib, seaborn), revealed balanced order distributions, moderate variability, and consistent demand patterns across business and individual clients.

The Analysis Process, planned for the final stage, will implement Linear Regression for demand forecasting, statistical segmentation (Mean ± SD) for identifying high-value customers, and descriptive comparative analysis for regional demand. The Results and Findings indicate seasonal demand fluctuations, client concentration, and regional dominance in Madurai, supporting data-driven planning and market strategy.

2. PROOF OF ORIGINALITY OF DATA

Video of Interaction with Business Owner: Video

Letter Head with approval: Letter-Head

Images: Images

3. METADATA

The raw dataset was originally provided by Sri Meenakshi Screen Printing, a Thamboolam bag manufacturer based in Madurai. It represents business order records collected between 2022 and 2025, containing transactional details of customer orders such as dates, client information, product specifications, and quantities. The data was shared in Excel (.xlsx) format as part of this Business Data Management project.

The raw dataset contained several inconsistencies including mixed date formats, irregular text capitalization, missing entries, and duplicate records. A systematic data cleaning process was performed by me to ensure consistency and reliability before analysis.

The following pre-processing steps were applied:

1) Standardized date formats to YYYY-MM-DD for both order and fulfilment dates.
2) Normalized categorical fields (e.g., bag colour, location, client type) to ensure uniform naming conventions.
3) Removed duplicate and irrelevant entries that did not contribute to meaningful analysis.
4) Created derived columns such as Order Month and Order Year to facilitate time-based analysis.
5) Verified numeric fields (like “Number of Bags”) for valid and positive values.
6) Ensured missing values were either imputed appropriately or excluded after cross-verification.
7) After cleaning, the dataset was validated to be free from major errors and ready for descriptive and visual analysis.

Data Format and Structure:

The final cleaned dataset was stored in Excel (.xlsx) format with 500 rows and 14 columns. Each row corresponds to a single customer order, and each column represents a specific business attribute such as client type, bag colour, or order quantity.

Feature	Description
Order ID	Unique code identifying each order record.
Client ID / Name	Identifies the customer associated with the order.
Location	District/city where the order originated.
Client Type	Indicates whether the order came from a business or individual client.
Bag Colour	Specifies the colour of the ordered bag (White, Yellow, Golden, etc.).
Bag Size	Dimension of the bag (e.g., 16×19 inch).
Bag Type	Type of bag material (Non-woven, Cotton, Cool Cotton).
Printing Colour	Number of colours used for printing (Single, Double, or Triple).
Number of Bags	Quantity of bags ordered.
Order Date / Fulfilment Date	Indicates when the order was placed and delivered.
Order Month / Year	Derived fields for time-based trend analysis.
Occasion	Describes the type of event (Festival, Regular, etc.) triggering the order.
Status	Final order outcome (Completed or Delayed).

4. DESCRIPTIVE STATISTICS

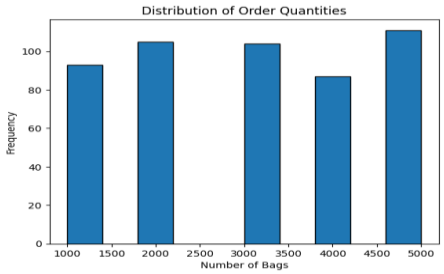
The dataset includes 500 orders from Sri Meenakshi Screen Printing between 2022 and 2025, detailing order dates, client type, location, bag features, and quantities. The key metric, Number of Bags, captures overall demand and helps assess sales consistency and client ordering patterns.

Descriptive Statistics Summary:

	Mean	Median	Mode	Standard Deviation	Minimum	Maximum	Skewness
No of Bags	3036.0	3000.0	5000	1420.824227	1000	5000	0.012017

The average order size is 3036 bags, with a median of 3000 and a mode of 5000, indicating balanced demand centered around medium to large orders. The standard deviation (1420.82) and range (1000–5000) show moderate variation in order sizes. A skewness of 0.012 confirms a nearly symmetric distribution, meaning small and large orders occur with similar frequency. The slightly flat distribution suggests diverse order volumes rather than dominance by a single size category.

The mode, being 5000, indicates that high-volume orders are the most frequently occurring, likely due to regular bulk purchases from key business clients.



The Figure illustrates the distribution of order quantities, showing that most orders cluster around the 3000 range with a nearly symmetric spread, supporting the computed skewness value of 0.012.

Descriptive Statistics by Client Type:

Descriptive Statistics by Client Type:

	Client Type	count	mean	median	std	min	max	skewness
0	Business	250	3124.0	3000.0	1443.940392	1000	5000	-0.056946
1	Individual	250	2948.0	3000.0	1394.659925	1000	5000	0.075046

The dataset includes 250 business and 250 individual client orders. Business clients show a slightly higher average order quantity (3124 bags) than individual clients (2948 bags), indicating that businesses tend to place larger or bulk orders. Both groups have a median of 3000 bags, suggesting overall balance in order size distribution.

The standard deviation values (1443.94 for business and 1394.66 for individual clients) indicate moderate variation in order quantities across both categories. The skewness values (−0.057 for business, 0.075 for individual) are close to zero, implying that order sizes are nearly symmetric for both groups with no strong bias toward smaller or larger orders.

This balance shows that Sri Meenakshi Screen Printing maintains a consistent mix of bulk and medium-sized orders across its client base, with slightly higher volumes contributed by business clients.

Descriptive Statistics for Regional Demand:

Descriptive Statistics by Location:

	Location	count	mean	median	std	min	max	skewness
0	Coimbatore	71	3154.929577	3000.0	1420.743139	1000	5000	-0.064902
1	Madurai	393	3020.356234	3000.0	1428.426019	1000	5000	0.022042
2	Theni	36	2972.222222	3000.0	1362.479066	1000	5000	0.050583

The regional data shows Madurai with the highest number of orders (393), followed by Coimbatore (71) and Theni (36). Average order quantities are close 3154 bags in Coimbatore, 3020 in Madurai, and 2972 in Theni indicating steady demand across regions. The standard deviation (1362–1428) and near-zero skewness (−0.06 to 0.05) confirm balanced distributions with moderate variation. Overall, all three regions show consistent order patterns, with Madurai remaining the key demand center.

5. DETAILED EXPLAINATION OF ANALYSIS PROCESS

The analytical process for Sri Meenakshi Screen Printing has been structured around three major problem areas: demand forecasting, high-value customer identification, and regional demand analysis.

5.1 Demand Forecasting

To forecast future demand, a Linear Regression model will be used to study the relationship between time (month index) and the number of bags ordered. Monthly totals from 2022–2025 will be used to establish a simple trend model, where the slope represents demand growth.

Linear Regression is chosen because it is easy to interpret and effective for small, stable time-series data. It captures overall trends without requiring large datasets or complex parameters, unlike models such as ARIMA or Holt-Winters, which need stronger seasonal evidence. Evaluation metrics such as MAE, RMSE, and MAPE will be applied to assess prediction accuracy. In later stages, advanced models may be explored to include seasonality, but Linear Regression provides a strong baseline for the current phase.

5.2 High-Value Customer Identification

To identify the most valuable customers, a descriptive segmentation method will be applied. The dataset will be grouped by Client ID to calculate total orders, total quantity ordered, and average order size. Customers will be segmented using statistical thresholds derived from the mean and standard deviation of total order quantities:

High-Value: > (Mean + 1 SD)

Medium-Value: within (Mean ± 1 SD)

Low-Value: < (Mean − 1 SD)

This approach is objective and data-driven, avoiding arbitrary cut-offs and ensuring transparency. It is more appropriate than clustering techniques, which require more parameters and data variety. Once implemented, this segmentation will help the company prioritize high-value clients and design suitable retention strategies.

5.3 Regional Demand Analysis

Regional variation will be studied using descriptive comparative statistics. The data will be grouped by Location and analysed using metrics such as count, mean, median, standard deviation, and skewness. These measures will reveal both the intensity and consistency of orders in each region.

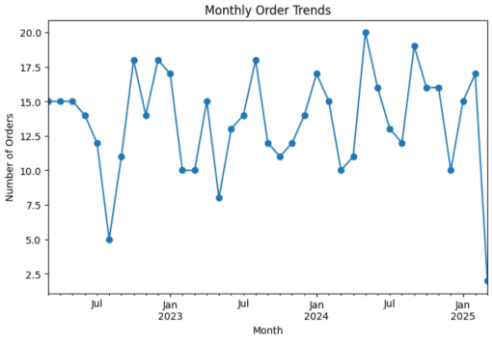
This method is suitable because the goal is comparison, not prediction. It provides clear insights into regional differences without complex testing requirements. It is more appropriate than inferential methods like ANOVA at this stage since descriptive comparison directly highlights which regions dominate demand and which show growth potential.

5.4 Data Cleaning and Pre-processing

Data pre-processing ensures reliability before analysis. Dates are standardized to YYYY-MM-DD, categorical fields such as Location and Client Type are normalized and missing or duplicate records are removed. Numeric fields like Number of Bags are converted to integers, and derived columns such as Order Month, Order Year, and Month Index are created for aggregation and trend analysis. These steps guarantee a consistent, validated dataset for accurate analytical outcomes.

6. RESULTS AND FINDINGS

Seasonal Demand Patterns:



Observation (Numeric):

The monthly number of orders fluctuates between 2 and 20, with an average of around 13–15 orders per month. Peaks are seen in certain months (e.g., mid-2023 and mid-2024), while noticeable drops occur around late 2022 and early 2025.

Trend:

The data shows clear seasonal fluctuations in demand. There are recurring peaks roughly every few months, indicating festival- or event-driven spikes in orders. Conversely, certain months consistently experience lower order volumes, reflecting off-season demand.

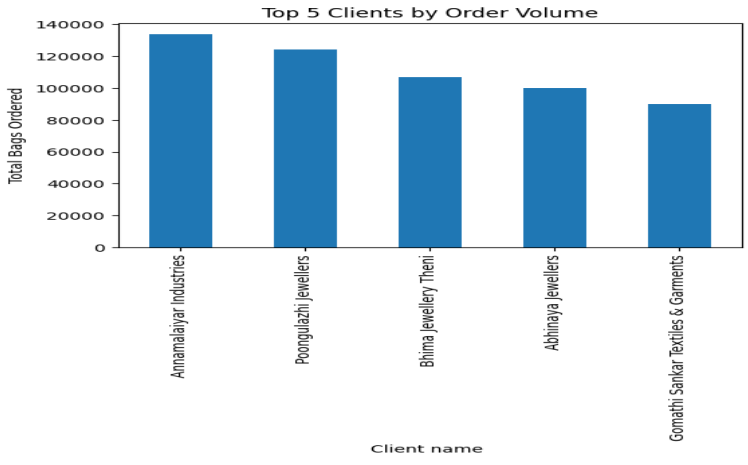
Why this pattern occurs:

For a business like Sri Meenakshi Screen Printing, demand is tied closely to occasions such as weddings, housewarmings, temple festivals, and local events. These typically cluster around particular months in Tamil Nadu (wedding seasons, festival months), which explains the demand spikes. Lean months align with periods where fewer cultural events occur, leading to lower order volumes.

Business Insight:

This seasonal pattern highlights the need for proactive raw material procurement and workforce allocation. By anticipating peak months, the company can stock adequate cotton cloth, dyes, and printing supplies in advance, while avoiding excess purchases during lean months. This ensures better inventory turnover and directly supports revenue growth by reducing missed orders due to shortages.

High-Value Customer Segmentation:



Observation (Numeric):

The top 5 clients together account for a major share of the company's orders, each placing between 90,000 and 135,000 bags over the observed period. The single largest client, Annamalaiyar Industries, contributes the highest order volume (~135,000 bags), followed closely by Poongulazi Jewellers (~125,000 bags). The remaining three clients – Bhima Jewellery Theni, Abhinaya Jewellers, and Gomathi Sankar Textiles – also show consistently high demand, each contributing above 90,000 bags.

Trend:

A small group of repeat business clients drives a disproportionately large share of total revenue, indicating customer concentration. This suggests that retaining these high-value customers is critical for stable sales.

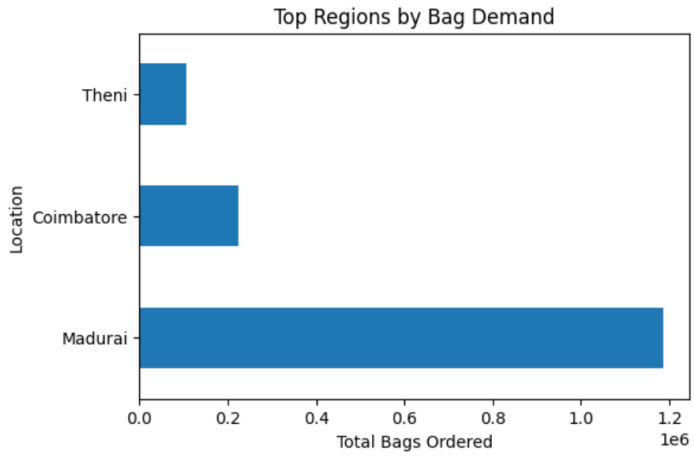
Why this pattern occurs:

These businesses likely place bulk and repeat orders, possibly linked to regular retail demand in their stores. Their sustained order volumes highlight their importance as anchor clients for Sri Meenakshi Screen Printing.

Business Insight:

Focusing on customer retention programs (faster deliveries, personalized service) for these top clients could secure recurring revenue. At the same time, over-dependence on a few customers poses risk the company should balance this by identifying and nurturing mid-tier clients to diversify its revenue base.

Regional Demand Analysis:



Observation (Numeric):

Madurai dominates regional demand with close to 1.2 million bags ordered, far exceeding other regions. Coimbatore follows with around 220,000 bags, while Theni records the lowest demand at roughly 100,000 bags.

Trend:

Madurai is clearly the central hub of demand, accounting for the majority of orders.
Coimbatore plays a secondary role, while Theni remains a smaller contributor. The demand distribution is highly skewed, with one region driving most of the business.

Why this pattern occurs:

Madurai, being a larger urban and commercial hub with more retail businesses and cultural events, naturally generates higher bag demand. Coimbatore, though significant, has comparatively less dependency on such bags, and Theni, being smaller and more rural, contributes the least.

Business Insight:

The company should prioritize Madurai for bulk production planning, marketing campaigns, and faster logistics, as it contributes the lion's share of revenue. At the same time, targeted promotions in Coimbatore and Theni can help capture untapped market potential, reducing dependency on a single region. Diversifying demand across multiple regions would also mitigate business risks tied to over-concentration.