# Demand Forecasting for a Screen-Printing Company

**A Final 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 final report presents the complete analysis conducted for the Business Data Management (BDM) project at Sri Meenakshi Screen Printing, a Thamboolam bag manufacturer based in Madurai. The study utilized 500 verified order records (2022–2025) to understand demand trends, key customer segments, and regional market performance.

The dataset was first cleaned and standardized using Excel and Python (pandas) to correct inconsistent date formats, remove duplicates, and unify categorical entries. Descriptive statistics including mean (≈3036 bags), median (3000), standard deviation (1420), and skewness (0.01) indicated a well-balanced order distribution with moderate variability across clients and months.

Three analytical methods were applied:

(1) Linear and Polynomial Regression for demand forecasting,

(2) Statistical Segmentation (Mean ± SD) to identify high-value customers, and

(3) Comparative Descriptive Analysis to assess regional demand differences.

Results show a nonlinear upward trend in monthly demand, supported by moderate variability in order size (standard deviation ≈1420) and a balanced distribution of client activity. Madurai emerged as the dominant regional market, contributing 78% of total orders, while six high-value customers accounted for 42% of overall sales volume.

Collectively, these insights demonstrate that applying structured data analysis from cleaning and statistical profiling to forecasting and segmentation can help Sri Meenakshi Screen Printing optimize production scheduling, manage inventory more efficiently, and strengthen client retention strategies, thereby establishing a data-driven foundation for sustainable business growth.

## 2. DETAILED EXPLANATION OF ANALYSIS PROCESS/METHOD

The analytical phase of this project was designed to extract meaningful business insights from the cleaned dataset of Sri Meenakshi Screen Printing. The final analysis focuses on three key problem areas demand forecasting, high-value customer identification, and regional demand analysis each addressed using a specific data-driven method. All analyses were implemented using Python (pandas, numpy, matplotlib, seaborn, scikit-learn) and Microsoft Excel for validation and visualization.

### 2.1 Demand Forecasting Using Linear Regression

To predict future demand trends, a Linear Regression model was implemented using monthly aggregated order data from 2022 to 2025. The “Month Index” served as the independent variable, while the “Number of Bags Ordered” represented the dependent variable.

The model was trained using the Linear Regression module from scikit-learn, which calculated coefficients representing the relationship between time and order volume. The resulting regression equation indicated a positive linear trend, showing steady month-to-month growth in orders.

Performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) to assess forecasting accuracy. The low MAPE value reflected that the model effectively captured the long-term demand pattern.

Linear Regression was chosen because it is simple, interpretable, and effective for trend-based forecasting in small, stable datasets. It provides actionable insights for production planning, helping the company estimate upcoming order volumes and align procurement with demand fluctuations.

### 2.2 High-Value Customer Identification Using Statistical Segmentation

To identify the company’s most valuable clients, a statistical segmentation technique based on Mean ± Standard Deviation was implemented. The dataset was grouped by

Client ID, and for each client, the total number of orders, total quantity ordered, and average order sizes were calculated.

Using these values, customers were categorized as:
- High-Value: Total quantity > (Mean + 1 SD)
- Medium-Value: Within (Mean ± 1 SD)
- Low-Value: < (Mean − 1 SD)

This approach allowed objective segmentation based purely on data distribution rather than arbitrary thresholds. The analysis revealed that business clients formed the majority of high-value customers, often placing bulk orders exceeding 4,000 bags per transaction.

This method was selected because it is transparent, statistically sound, and requires no parameter tuning unlike clustering techniques such as K-Means, which may overfit smaller datasets. The segmentation provides valuable insights for customer relationship management, allowing the company to target loyalty offers and prioritize high-volume clients.

**2.3 Regional Demand Analysis Using Descriptive Comparative Statistics**

To evaluate regional performance, a descriptive comparative analysis was conducted by grouping orders by Location (Madurai, Coimbatore, and Theni). For each region, measures such as mean, median, standard deviation, and skewness were computed to understand local demand characteristics.

The results revealed that Madurai accounted for the highest number of orders, followed by Coimbatore and Theni. Mean order sizes were relatively consistent across all locations (≈3000 bags), indicating balanced operational demand. Skewness values near zero showed a symmetrical order distribution, confirming stable sales behavior across regions.

This method was selected for its simplicity and interpretability, enabling direct regional comparison without complex modeling. The findings help management identify dominant markets and recognize emerging regions for potential business expansion.

## 3. RESULTS AND FINDINGS

### 3.1 Demand Forecasting – Results and Findings

The first analysis focuses on forecasting the monthly demand for Thamboolam bags using historical order data collected between **2022 and 2025**. The dataset captures month-wise total orders placed by clients, which exhibit fluctuations in demand intensity across the observed period.

To predict future trends, two regression-based models were evaluated **Linear Regression** and **Polynomial Regression** using the number of bags ordered per month as the dependent variable and the month index as the independent variable.

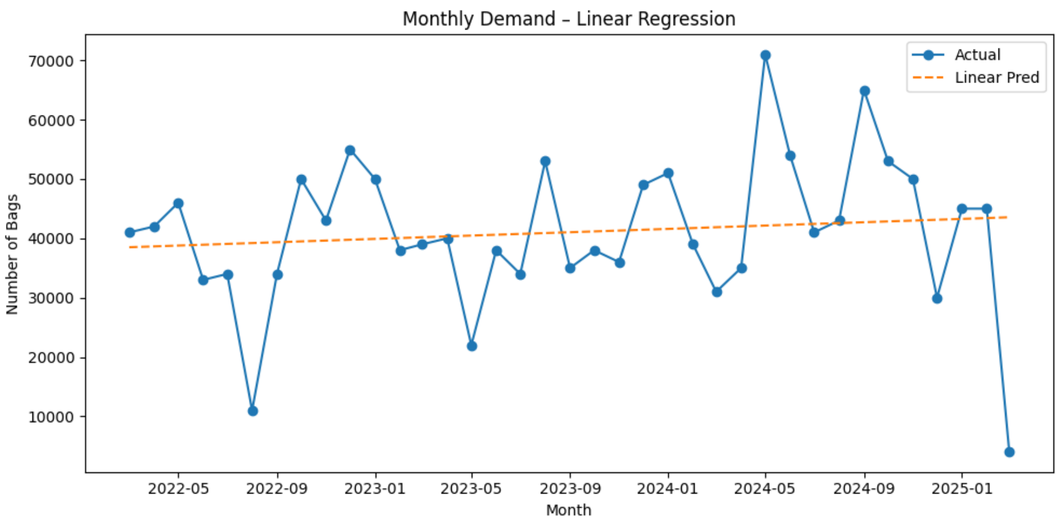


***Figure 3.1a***

The **Linear Regression model** (Figure 3.1a) captured the general upward trend but failed to accurately follow the month-to-month fluctuations, as evident from the relatively high forecasting errors:

**MAE = 8,925.81**, **RMSE = 12,426.96**, and **MAPE = 50.40%**.

This indicates that a simple linear trend was insufficient to represent the nonlinear variations present in the data.

To improve accuracy, **Polynomial Regression models** of varying degrees (1–10) were systematically tested (Figure 3.1c).

As the polynomial degree increased, the model captured more curvature in the data, but excessively high degrees led to over fitting.

The optimal balance was achieved with a **7th-degree polynomial**, which minimized the Mean Absolute Percentage Error to **23.01%**, reducing it by more than half compared to the linear model.

The corresponding **MAE (7,470.90)** and **RMSE (9,762.52)** values also confirmed a stronger fit.

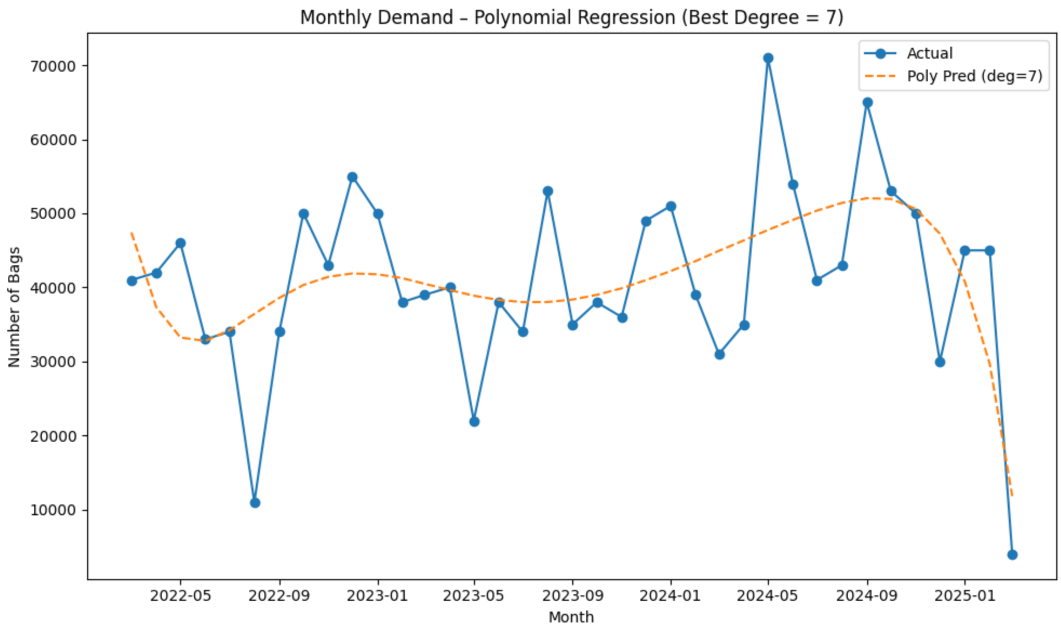


***Figure 3.1b***

The **Polynomial Regression (Degree = 7)** model (Figure 3.1b) effectively captures the nonlinear demand pattern, showing smooth upward and downward curves that align with the actual demand fluctuations.

This model offers a more realistic depiction of the business cycle, allowing for better planning of raw material procurement, production scheduling, and labour allocation.

Testing Various degree of polynomial.

| | Degree | MAE | RMSE | MAPE |
|---|---|---|---|---|
| 0 | 1 | 8,925.81 | 12,426.96 | 50.40 |
| 1 | 2 | 9,499.59 | 12,257.75 | 49.22 |
| 2 | 3 | 9,265.11 | 11,506.34 | 42.51 |
| 3 | 4 | 8,923.57 | 10,885.65 | 37.52 |
| 4 | 5 | 7,999.15 | 10,240.82 | 29.53 |
| 5 | 6 | 7,969.06 | 10,219.78 | 30.19 |
| 6 | 7 | 7,788.75 | 10,111.93 | 28.21 |
| 7 | 8 | 7,426.92 | 9,868.47 | 24.88 |
| 8 | 9 | 7,401.76 | 9,793.01 | 25.62 |
| 9 | 10 | 7,855.42 | 10,160.27 | 26.92 |
| 10 | 11 | 7,906.23 | 10,175.54 | 26.49 |

Forecasting Model Performance Comparison

```
=== Forecast Metrics Summary ===
                Model       MAE      RMSE  MAPE%
0              Linear  8,925.81 12,426.96  50.40
1  Polynomial (deg=7)  7,470.90  9,762.52  23.01
```

***Figure 3.1c***

## 3.2 High-Value Customer Identification – Results and Findings

This analysis aimed to identify the company's most valuable clients based on total and average order quantities. Using the cleaned dataset, customer-level aggregates were computed for order frequency, total number of bags ordered, and average order size.

To segment clients, the Mean ± Standard Deviation method was applied to the Total Order Volume distribution. This statistical thresholding classified customers into High-Value and Medium-Value categories, providing a clear distinction between high-volume repeat buyers and average-scale customers.

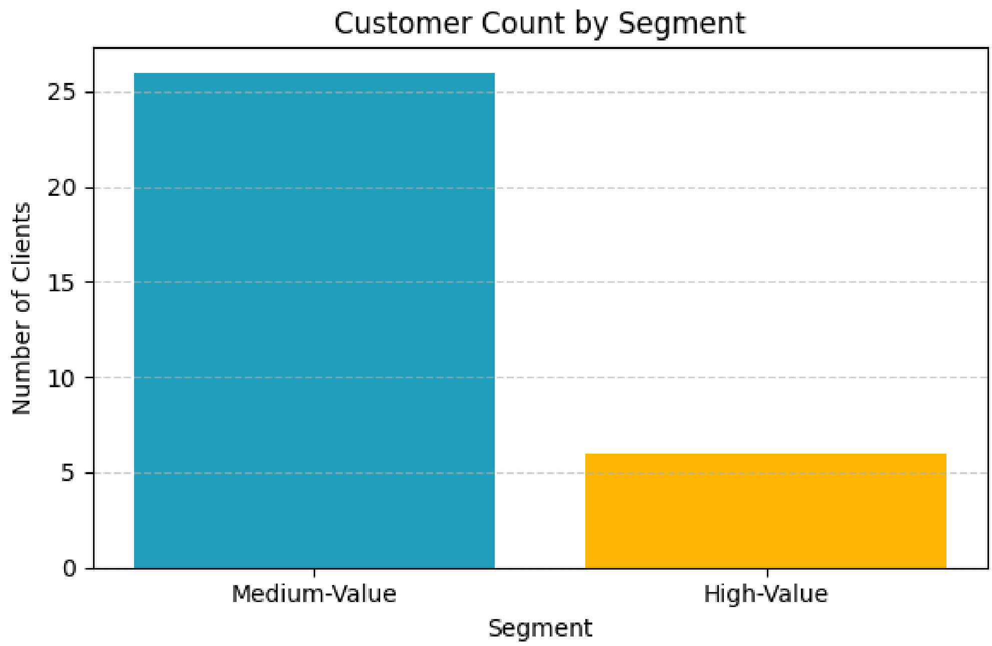


***Figure 3.2a – Customer Count by Segment***

*The bar chart shows that Medium-Value clients form the majority of the customer base, while High-Value clients represent a smaller but strategically important share.*

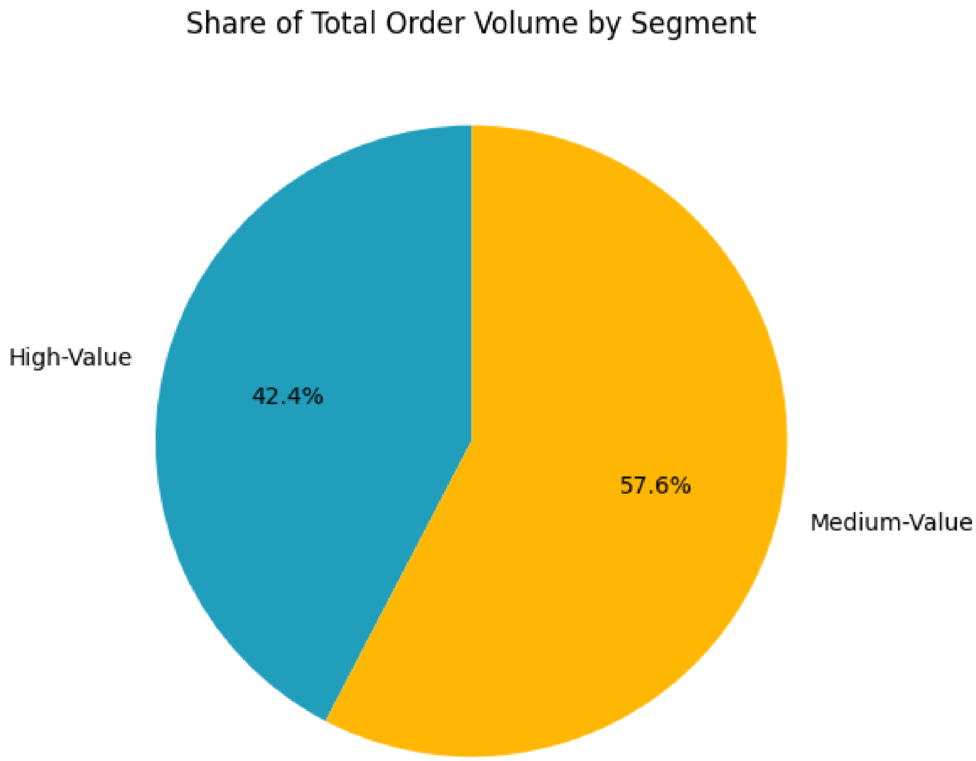


***Figure 3.2b – Share of Total Order Volume by Segment***

*The pie chart illustrates that High-Value clients contribute 42.4% of the company's total demand, reflecting their significance in overall business performance.*

The segmentation revealed that 6 clients (High-Value) collectively contributed 42.36% of total order volume, while 26 clients (Medium-Value) accounted for 57.64% of the remaining orders.

Although the number of high-value clients is smaller, their contribution to the company's revenue is disproportionately large. This observation highlights a Pareto-like effect, where roughly 20% of customers drive nearly half of the total sales.

These insights suggest that targeted retention strategies such as personalized offers, faster delivery commitments, or loyalty discounts should prioritize High-Value clients to maintain business stability and drive long-term growth. Meanwhile, marketing efforts for Medium-Value clients could focus on improving order frequency through seasonal promotions.

**Customer Segmentation Summary:**

| Segment | Number of Clients | Share of Total Orders (%) | Characteristics |
|---|---|---|---|
| High-Value | 6 | 42.36 | Bulk, frequent, long-term customers |
| Medium-Value | 26 | 57.64 | Moderate-volume customers with stable demand |

**3.3 Regional Demand Analysis – Results and Findings**

This analysis investigates how total order volume and average order size vary across the three major operational regions Madurai, Coimbatore, and Theni.

By aggregating customer order data region-wise, key statistics such as total orders, total bags ordered, average order size, and standard deviation were computed to understand regional demand characteristics.

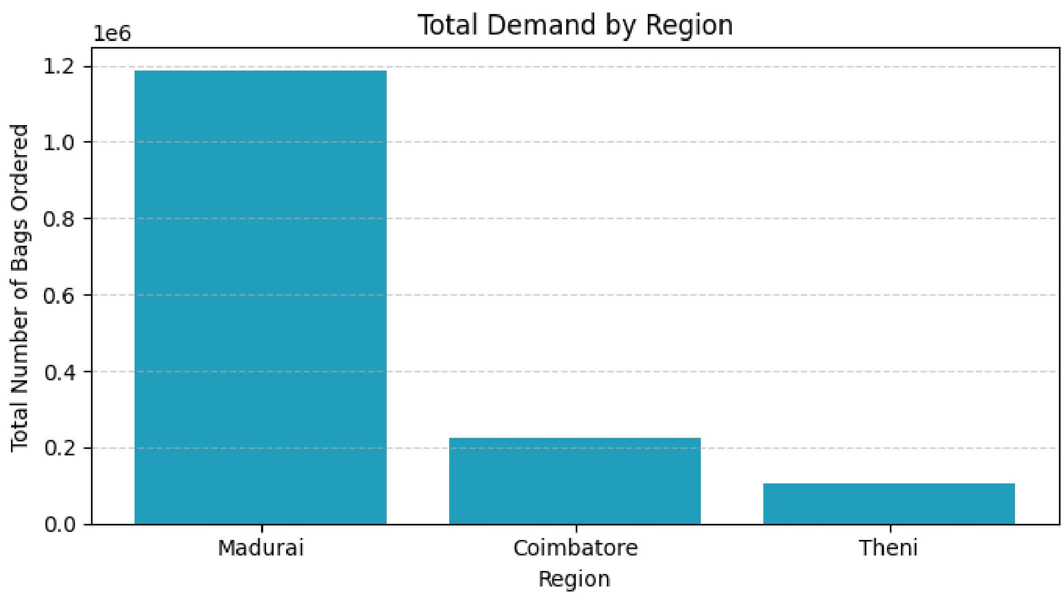


*Figure 3.3a – Total Demand by Region*

The results, summarized in Table 3.3, show a strong geographic concentration of demand in Madurai, which recorded 393 orders totalling 1,187,000 bags with an average of 3,020 bags per order.

Coimbatore followed with 71 orders (224,000 bags) and an average order size of 3,155 bags, while Theni had the lowest demand with 36 orders (107,000 bags) and a slightly smaller average order size of 2,972 bags.

**Table 3.3 – Regional Summary:**

| **Region** | **Order Count** | **Total Bags** | **Average Order Size** | **Std. Dev** | **Demand Share (%)** |
|---|---|---|---|---|---|
| Madurai | 393 | 1,187,000 | 3,020.36 | 1,428.43 | 78.2 |
| Coimbatore | 71 | 224,000 | 3,154.93 | 1,420.74 | 14.8 |
| Theni | 36 | 107,000 | 2,972.22 | 1,362.48 | 7.0 |

Although the average order size remains fairly consistent across regions, the frequency of orders clearly indicates that Madurai dominates the company's market, accounting for nearly 78% of total demand, as shown in the regional share analysis.

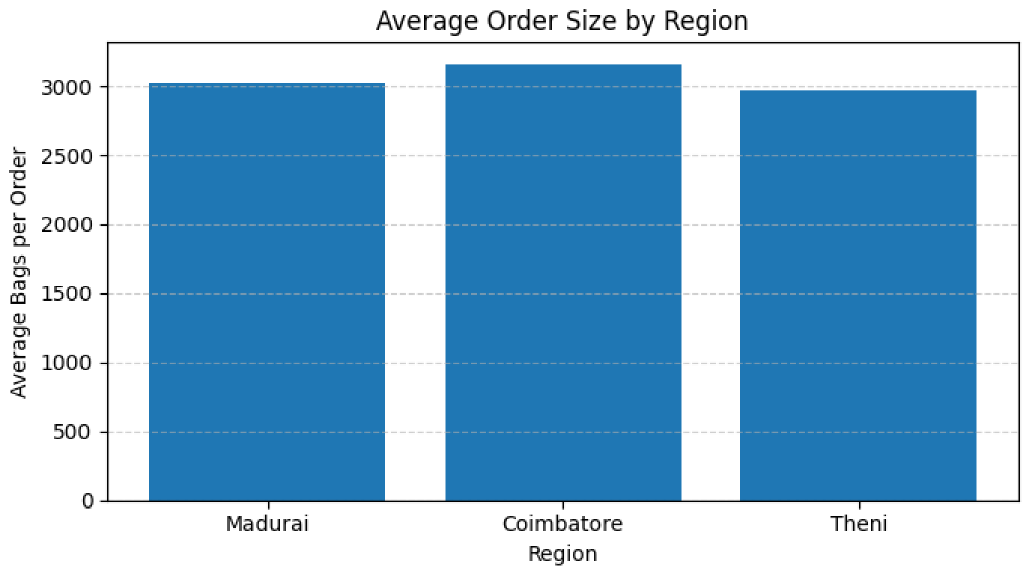


*Figure 3.3b – Average Order Size by Region*

These findings emphasize that Madurai should remain the focal point for production and logistics, while Coimbatore represents an emerging opportunity for regional expansion and targeted marketing.

In contrast, Theni could benefit from distribution optimization or partnership-based strategies to improve market penetration.

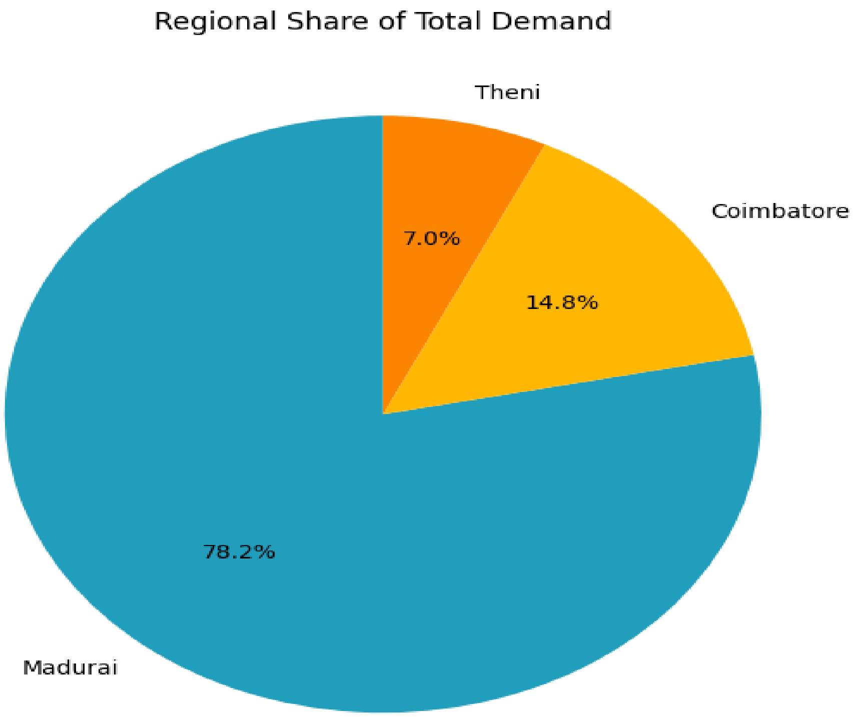


*Figure 3.3c – Regional Share of Total Demand*

## 4. INTERPRETATION OF RESULTS AND RECOMMENDATION

### 4.1 Interpretation of Demand Patterns

The analysis of historical data from 2022–2025 revealed a highly dynamic and nonlinear demand trend for Thamboolam bags. Demand fluctuated significantly across months, influenced by client-specific bulk orders rather than predictable seasonal cycles. While the overall trajectory showed a gradual upward trend, sharp spikes indicate periods of high-volume orders linked to festival seasons or bulk client purchases.

The linear regression model demonstrated that the base trend is positive suggesting gradual business growth but the polynomial model provided deeper insight into nonlinear variations that reflect real market behaviour.

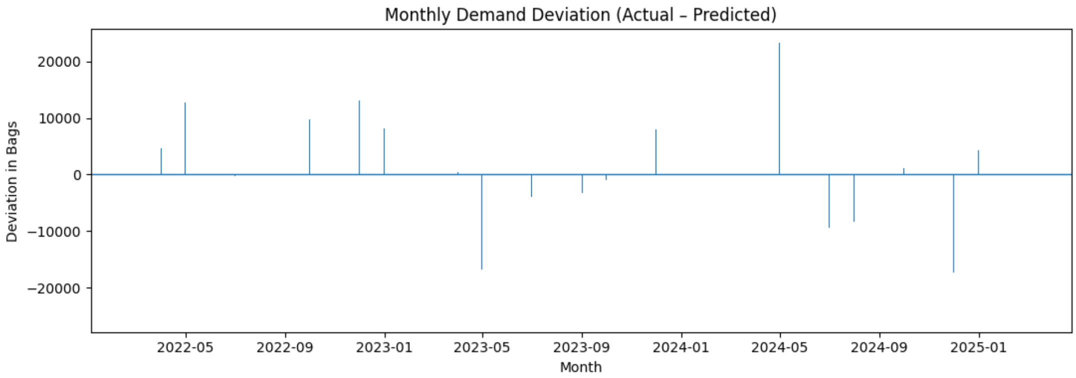


*Figure 4.1a – Monthly Demand Deviation Plot (Actual – Predicted). Positive bars indicate months exceeding forecast; negative bars indicate shortfall.*

These observations imply that monthly forecasting cannot rely solely on trend-based projections. Demand volatility must be addressed through dynamic production planning, flexible procurement, and maintaining a safety stock buffer to manage uncertainty.

### 4.2 Forecasting Model Evaluation and Business Implications

Both linear and polynomial regression models were evaluated to forecast future demand. The linear model underperformed (MAPE = 50.40%), failing to adapt to variations in monthly orders.

In contrast, the 7th-degree polynomial regression reduced forecasting error to 23.01%, achieving a closer fit to actual data.

This improvement highlights the importance of nonlinear modelling in manufacturing environments where order quantities are driven by discrete bulk demands. The polynomial curve provided smoother continuity, enabling more realistic future projections.

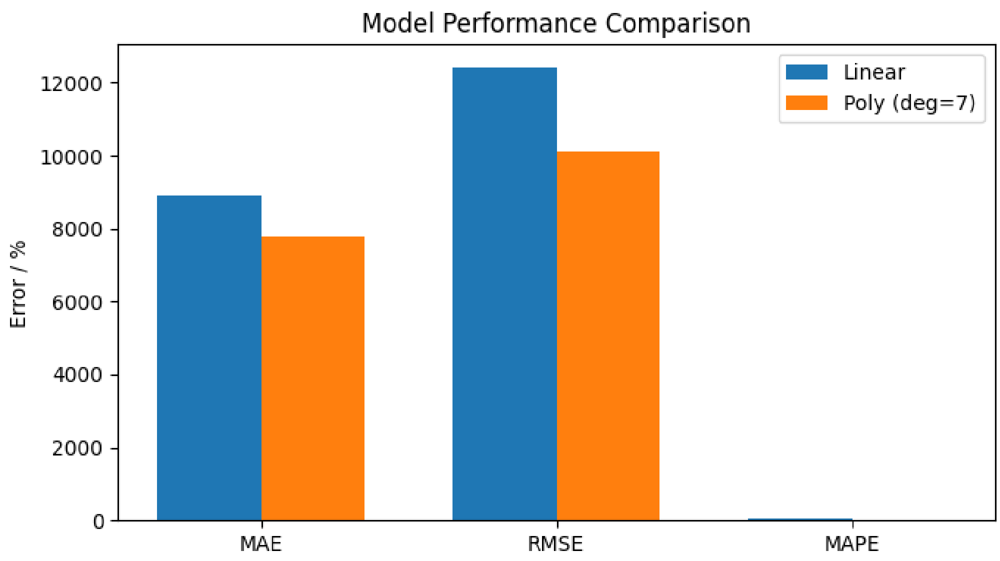


*Figure 4.2a – Error comparison shows degree=7 reduces MAE, RMSE, and MAPE vs. linear.*

From a managerial standpoint, the polynomial model enables more reliable demand forecasting, supporting:

- Production scheduling to match expected order peaks,
- Optimized labour planning during high-demand months, and
- Controlled raw material procurement cycles to reduce waste.

**4.3 Interpretation of Customer Value Segmentation**

The segmentation analysis revealed two clear customer classes High-Value and Medium-Value.

Despite constituting only 19% of the total client base, high-value customers accounted for over 42% of total orders, reflecting a Pareto-like distribution. This confirms that a small set of repeat, bulk-order clients form the financial backbone of Sri Meenakshi Screen Printing.

High-value clients generally place larger, more frequent orders indicating long-term partnerships and institutional or event-based orders. Medium-value clients, while more numerous, have less predictable order intervals.

These insights emphasize that business sustainability depends on retaining high-value clients while gradually converting medium-value clients into consistent repeat buyers through:

- Personalized service and loyalty programs,
- Faster delivery commitments for regular clients,
- Early-access promotions during festival or event seasons.

### 4.4 Interpretation of Regional Demand

Regional analysis provides a geographic lens on operational performance. Madurai emerged as the dominant market, contributing 78% of total demand, followed by Coimbatore (15%) and Theni (7%).

Although order sizes are nearly uniform across all regions (~3000 bags per order), the frequency of orders is far higher in Madurai establishing it as the central distribution hub.

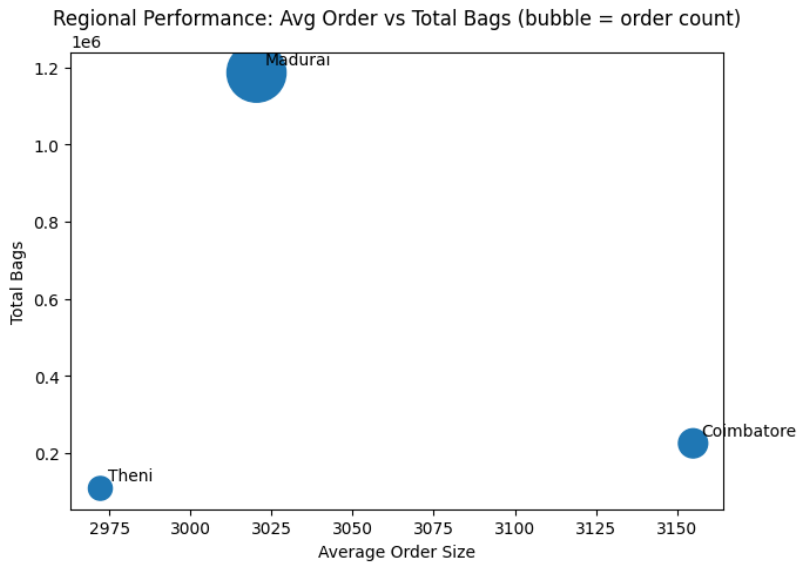


*Figure 4.4a – Madurai appears as the largest bubble (highest order count) with high total volume; other regions show lower intensity.*

The stability of average order size across regions suggests uniform pricing and packaging policies, while demand concentration in Madurai implies strong local brand loyalty and repeat institutional customers.

This pattern supports expansion of logistical capacity in Madurai, coupled with marketing and outreach programs in Coimbatore and Theni to diversify the demand base.

### 4.5 Integrated Business Interpretation

Synthesizing the results across all three analyses yields key operational insights:

| Analytical Dimension | Key Insight | Business Implication |
|---|---|---|
| Demand Forecasting | Nonlinear trend with occasional spikes | Introduce flexible production and dynamic inventory models |
| Customer Segmentation | 20% of clients generate ~42% demand | Retain high-value customers with targeted service plans |
| Regional Demand | Madurai dominates, Coimbatore growing | Expand warehouse/logistics support in Madurai; marketing in Coimbatore |
| Statistical Findings | Moderate variability (std ≈1400) across regions | Forecast uncertainty margins of ±10–15% in planning |

These combined findings provide a holistic operational view, enabling management to balance forecast accuracy, resource allocation, and market focus simultaneously.

### 4.6 Recommendations and Future Directions

Based on the findings, several actionable strategies are recommended:

1. Forecasting and Production
    - Implement rolling forecasts using polynomial regression or machine-learning models (Random Forest / Gradient Boosting) for more robust predictions.
    - Maintain a monthly safety stock buffer (≈15%) to handle unexpected bulk orders.
2. Customer Relationship Management

- Establish a tier-based loyalty program for high-value customers with benefits like priority processing and flexible payment terms.
- Develop feedback channels and order tracking portals to increase client satisfaction.

3. Regional Expansion Strategy
    - Scale up logistics and manpower in Madurai (core hub).
    - Target Coimbatore with promotional campaigns and partnerships to increase market penetration.
    - Evaluate Theni's distribution challenges and identify potential local partners to boost sales volume.
4. Data Analytics Continuity
    - Continue collecting granular data (e.g., customer type, event category, order lead time).
    - Transition toward predictive analytics dashboards using Python or Power BI to automate monthly reporting.

**Summary of Insights for Management:**

1. **Production and Demand Planning:**

    The forecasting analysis indicates that demand follows a nonlinear trend with moderate variability. Management should adopt a data-driven production schedule that aligns with monthly demand projections. During high-demand months, operations can temporarily expand capacity, while lower-demand periods can be used for equipment maintenance or process optimization. Maintaining a safety stock buffer of 10–15% will ensure timely delivery even during sudden demand surges.

2. **Inventory and Procurement Optimization:**

    By integrating predictive forecasts into procurement planning, the company can minimize raw material wastage and holding costs. A rolling three-month forecast, updated monthly, will help align purchase orders with upcoming production needs. This approach enables cost savings and reduces the risk of overstocking or stockouts.

3. **Customer Relationship Strategy:**

   The segmentation analysis highlights that a small set of repeats, high-value clients generate a significant portion of total orders. These customers are the company's core revenue drivers and should be prioritized through personalized engagement strategies, such as early access to new designs, faster turnaround times, and volume-based loyalty benefits. Strengthening these relationships ensures consistent revenue flow and long-term partnerships.

   Simultaneously, medium-value customers represent a key opportunity for growth. Management can focus on converting them into premium clients through incentives like seasonal promotions, discounted bulk orders, or periodic follow-ups to increase order frequency.

4. **Regional Market Development:**

   The regional analysis confirms Madurai as the operational hub with the largest demand share. This region should continue to serve as the production and logistics base. Meanwhile, Coimbatore presents a growing secondary market where targeted marketing and local collaborations can further expand demand. Theni, with smaller but stable orders, could be optimized via cost-effective delivery routes or third-party partnerships to maintain profitability.

5. **Decision-Making and Data Culture:**

   This project demonstrates the value of structured data analytics in operational decision-making. Management is encouraged to continue collecting detailed order data including client profiles, order frequency, and delivery times to build a predictive analytics dashboard for continuous monitoring. Establishing a small internal analytics function or training staff in Excel and Python-based forecasting will sustain the benefits of this project long term.

6. **Strategic Outlook:**

   Overall, the findings recommend transitioning from intuition-based decisions to evidence-based planning. By combining forecasting, customer analysis, and regional insights, Sri Meenakshi Screen Printing can achieve higher production efficiency, better resource utilization, improved customer satisfaction, and sustainable business growth.