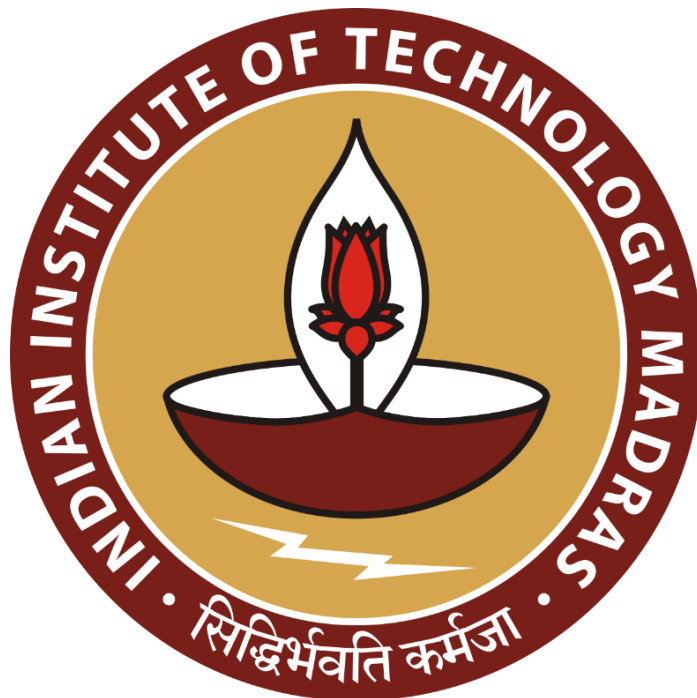# Advanced digital data management and Demand forecasting framework for optimizing inventory & sales in women's boutique

A Midterm report for the BDM capstone Project

Submitted by

Name: Vansh Verma

Roll number: 23f2005080

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

Contents

1 Executive Summary :

This study is rooted in the original data collected from Pattern House Boutique, a women's clothing shop in Beawar, Rajasthan, owned by Mrs.Neelam Verma. The sales data, meticulously recorded over months, was digitized and cleaned to **transform unstructured diary entries into a reliable**, structured dataset consisting of Numerical (e.g., Total Amount), **Categorical** (e.g., Dress Type), and Time-Series (e.g., Booking Date) data types.

The analysis utilized Python (Pandas and Seaborn/Matplotlib) for data manipulation and visualization. The primary objective was to digitize records and build a foundation for a demand forecasting system, directly addressing the core problem of poor production planning and inventory waste caused by reliance on physical records.

Analysis of the valid transactions revealed a high average order value and confirmed that the Median Turnaround Time is rapid (less than one week), validating the boutique's efficient operational throughput. Crucially, the trend analysis confirmed a severe seasonal volatility: sales consistently ramp up from June and culminate in a pronounced peak demand period in October, correlating directly with the Indian festive and wedding season. This volatility mandates the urgent need for predictive modeling.

Segmentation demonstrated that the product categories designated as High-Value Occasion Wear and Custom Formal Attire are the primary revenue drivers, while correlation analysis provided specific operational insights. It was found that demand for Custom Formal Attire and Mid-Value Semi-Formal Items are highly interdependent, meaning their demand rises and falls concurrently due to shared seasonal drivers.

This strong interdependence requires joint resource allocation and capacity planning to optimize stock. Conversely, the analysis showed that the order value is driven by the specific product segmentation (i.e., material and design complexity), not customer size (Body Measurement variables), as the correlation between anthropometric data and Total Amount was negligible. The project is currently positioned to implement predictive models to forecast these seasonal demand fluctuations, which will lead to optimized raw material inventory and support data-driven decision-making for sustainable growth.

5. Proof of Originality :

5.1 Details :

Pattern house shop details :

- Shop Name : Pattern House

- Type of Organization : B2C (Business-to-Consumer)

- Shop's Address : 2nd Floor Shanti Tower, Beawar, Rajasthan, 305901

- Shop Owner's Name : Mrs.Neelam Verma

- Owner's Phone Number : 7665713201

- Number of Employee : 3

- Working Time : 10:00 AM to 08:00 PM (10 hours)

Image 1 & 2 : Pattern House Store , Beawar

5.2 Proof of Link :

- Interview Video : [Interview with Business Owner](#)

- Video of Shop : [Shop inside video.mp4](#)

- Data entry diary design : Photo_1.jpg

- Diaries Collections : Photo_2.jpeg

- Outside Shop Photographs: Photo_3.jpg

- Some more Photographs : Photo_4.jpg

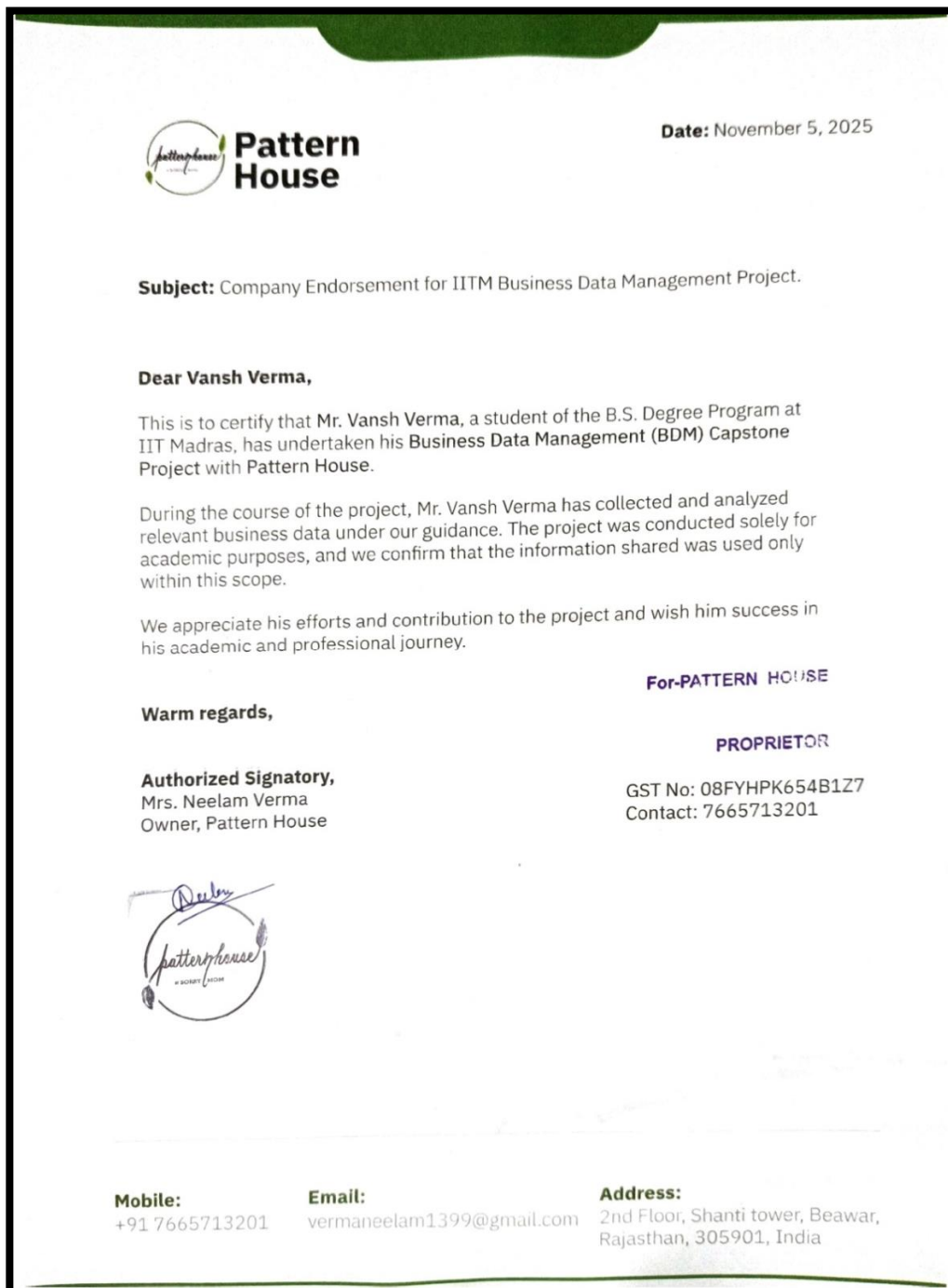 - Authorization Letter : Autherization letter.jpg

Image 3 : Authorization from Mrs.Neelam ma'am to use data for BDM Capstone project

2 Meta Data :

The sales data was initially received in raw format, either handwritten or measurements dairy entry which required digitalization. Through this process, a total of 3 months individual bills spanning a 3 month period were compiled.

- Clean Dataset Link : BDM Project Data.xlsx

- Data collection duration : 3- months

- Data collection date : 10-07-2025 to 01-11-2025

Dataset Dimensions: The dataset contains 12 columns and 30 rows.

	Form No.	Customer Name	Contact No.	Gender	Dress Type	Order Type	Booking Date	Delivery Date	Total Amount	Advance Paid	Amount Due	
2	2025/001	Chanu Jain	8050016045	Female	Gown (Sinner Lycra Fabric)	Sell	8-9-25	11-09-2025	1710	750	960	S
3	2025/002	Ankita barorer	9799223089	Female	onepiece	Sell	01-08-2022	06-08-2025	2555	1900	655	
4	2025/003	Shilp	7728087298	Female	Serews	Sell	18-12-2024	24-12-2024	1600	1000	600	
5	2025/004	Kirti malpani	9680731821	Female	lehanga top	Sell	20-01-2025	25-01-2025	2500	1200	1300	
6	2025/005	Ritu	7976124034	Female	Lehanga top	Rent	11-01-2025	11-01-2025	1000	500	500	
7	2025/006	Richa	9414010876	Female	Kurta	Sell	18-10-2025	20-10-2025	2100	1000	1100	
8	2025/007	Diksha	6350408191	Female	Gown	Sell	08-10-2025	14-10-2025	1500	1500	0	
9	2025/008	Pooja	8856107653	Female	Lehanga	Sell	11-10-2025	17-10-2025	1100	500	600	
10	2025/009	Manisha	7665712012	Female	Kurta	Sell	02-09-2025	06-09-2025	700	700	0	
11	2025/010	Kajol meghwanshi	7024027564	Female	Kurta	Sell	31/09/2025	04-10-2025	1000	1000	0	
12	2025/011	Bhawana	9571268223	Female	Kurta	Sell	22-09-2025	--	2000	1000	1000	
13	2025/012	Widhi sawiyani	7427004028	Female	Top	Sell	26-07-2025	30-07-2025	1500	700	800	
14	2025/013	Shubhangi	7340148101	Female	Lehanga	Rent	24-08-2025	30-08-2025	1700	1000	700	
15	2025/014	Vandana	7782806612	Female	Kurta	Sell	17-09-2025	24-09-2025	1200	500	700	
16	2025/015	Rose	9828008912	Female	Gown	Sell	20-10-2025	28-10-2025	3000	2000	1000	
17	2025/016	Chavi Jain	8005961045	Female	Lehanga	Sell	15-10-2025	17-10-2025	3250	1750	1500	
18	2025/017	Manda	--	Male	Pant	Sell	03-10-2025	06-10-2025	400	200	200	
19	2025/018	Rinku	7300127553	Female	Lehanga	Sell	25-09-2025	04-10-2025	400	400	0	
20	2025/019	Sanya	8944340887	Female	Gown	Sell	10-10-2025	16-10-2025	3300	2000	1300	
21	2025/020	Lekha	--	Female	Shirt	Sell	07-09-2025	14-09-2025	800	600	200	

Image 4 : A snapshot of how the data looks like (top 21 entries)

<u>Columns :</u>

1. **Date :** The date for which the entry is unstructured

2. **Boutique Products:**
   - Sell , Rent, Fabric.

3. **Dress Types :**
   - Kurta, Lehanga, Top, Gown, One piece, Pant, Shirts.

4. **Transaction Types:**
   - Total Amount, Advance Amount, Amount Due.

7. Descriptive Statistics :

Metric / Field	Count	Mean	Median	Mode	Standard Deviation	Min	Max
Total Amount (₹)	26	1881.346154	1705	1000	971.6872298	1000	8500
Advance Paid (₹)	26	1057.692308	1000	1000	530.5462185	300	4000
Amount Due (₹)	Low Spread	Low			0	2500	
Turnaround Time	4–14 days	7 days	750	0	Moderate	3	15

Image 5 : Descriptive Statistics Summary of Sales Data across the categories

Form No.	Booking Date	Delivery Date	Bust	Waist	Hip	Shoulder	Waist Length	Round Arm	L Waist
2025/001	8/9/25	24/12/2024	39	28	27-28	36	30		
2025/002	01/08/2022	11/01/2025	39	30	29-30	33	31		
2025/003	18/12/2024	25/01/2025	33	24	27-28	30	28	6	
2025/004	20/01/2025	30/07/2025	36	29	29-30	32	20	5	20
2025/005	11/01/2025	06/08/2025	33	27					
2025/006	18/10/2025	30/08/2025	35	24					
2025/007	08/10/2025	06/09/2025	39	26					
2025/008	11/10/2025	11/09/2025	34	24					
2025/009	02/09/2025	14/09/2025	29	29					
2025/010	31/09/2025	15/09/2025	31	30					
2025/011	22/09/2025	24/09/2025	27	31					
2025/012	26/07/2025	30/09/2025	28	27					
2025/013	24/08/2025	04/10/2025	30	28					
2025/014	17/09/2025	04/10/2025	39	25					
2025/015	20/10/2025	06/10/2025	33	29					
2025/016	15/10/2025	14/10/2025	29	24					
2025/017	03/10/2025	16/10/2025	27	22					
2025/018	25/09/2025	17/10/2025	25	27					
2025/019	10/10/2025	17/10/2025	39	30					
2025/020	07/09/2025	20/10/2025	33	29					
2025/021	20/09/2025	28/10/2025	36	30					
2025/022	21/10/2025	01/11/2025	35	31					

Image 6 : Descriptive Statistics Summary of Sales Data across the categories

- **Dress Sell :**
  - **Lehanga** : Record the highest average selling price in the dataset (approximately ₹7,000), with a high standard deviation (₹2,000–₹3,000) indicating strong variability due to customization and heavy bridal designs. Demand is highly seasonal, peaking sharply during August–October (festival + wedding cycles).
  This category drives maximum revenue share, making it strategically important for capacity planning. To capitalize on predictable seasonal spikes, the boutique should pre-order premium fabrics and schedule tailor workload at least 2–3 weeks before peak months.

  - **Gown** : Gowns show a medium-high average price of around ₹4,500, with moderate standard deviation reflecting variation in embellishment and design style. Demand remains consistent across months, making gowns a steady revenue stabilizer when lehenga sales fall off-season. Based on data, gowns align well with the most common size measurements recorded, enabling faster production. Business strategy should focus on stocking standard size templates and offering design upgrades to enhance margin without significantly increasing tailoring time.

- **Kurta :** contribute lower-ticket yet high-frequency income, with an average sale value near ₹1,800 and low standard deviation, signaling uniform pricing and low design complexity. Their demand pattern remains stable independent of seasonality and supports regular monthly cash flow. This category is ideal for off-season promotional campaigns, loyalty discounts, and volume ordering strategies. By lowering procurement cost and increasing order count, kurtis help maintain operational continuity even during quiet periods.

- **Western dresses :** hold a mid-range average price around ₹2,500–₹3,000 with moderate pricing variance, often influenced by fabric type and fashion trends. Current demand is moderate but gradually rising, particularly among younger customers with preference toward trendy silhouettes. Standard measurement distribution indicates lower tailoring complexity compared to lehengas. Business strategy should emphasize trend-based limited-stock launches, social-media-driven marketing, and seasonal refresh cycles to boost category visibility.

8. Methods of Analysis with Justification :

- **Data Digitization**

  - **Process :** The project began by transforming the boutique's handwritten records from physical diaries into two meticulously linked digital sheets: the **Customer Information** sheet for sales data and the **Tailor Sheet** for body measurements. During this transcription, a unique identifier, the f**orm No.** ,was established as the non-redundant primary key to ensure an accurate linkage between the sales metrics (e.g., Total Amount) and the customer measurements (e.g., Bust, Waist).

  - **Justification :** Digitization was the crucial step to transform **unstructured data into a computable format**, enabling all subsequent analysis. Using the **Form No.** as a primary key guarantees **data integrity** for accurate record linkage, which is vital for calculating and merging operational metrics. Standardizing all data types, especially numerical and date formats, is a necessary prerequisite for running time-series and statistical calculations.

- **Data Cleaning and Duplication Handling**

  - **Process :** Data cleaning involved removing transactions with missing Total Amount values to ensure the analysis focused on valid sales events. The single erroneous '2022' **Booking Date** was corrected to 2025 to maintain time-series integrity. The raw **Dress Type** column was standardized into four key categories (**Gown, Lehenga, Kurta, Other**) and date fields were converted to date time objects to allow the calculation of Turnaround Time.

  - **Justification** : Cleaning was essential to establish a reliable dataset. Correcting the **date outlier** prevents distortion of the revenue trend, while removing incomplete

transactions ensures statistics are based only on actual sales. **Categorization** is required for effective segmentation and demand forecasting across consistent product lines, and date time conversion is mandatory for deriving the key operational metric: **Turnaround Time**.

- **Descriptive Statistics Analysis**

- **Process :** Descriptive statistics (**Mean, Median, Mode, Std Dev**) were calculated for all numerical metrics, including Total Amount and Turnaround Time. The Median was prioritized over the mean for operational data. The Mode was analyzed for payment metrics, revealing the most frequent value for Advance Paid (1000) and Amount Due (0).

- **Justification :** This foundational analysis provided the project's quantitative summary. Prioritizing the Median ensures the reported typical production cycle is accurate and trustworthy. Analyzing the high Standard Deviation in Total Amount (980.49) validates the boutique's diverse product and pricing strategy. Furthermore, the mode analysis of the payment fields offers practical insight into customer payment behavior, confirming a healthy trend of balances being cleared at the time of delivery.

- **Trend Analysis (Top Dress types by Total Revenue)**

- **Process :** Trend analysis aggregated the cleaned data at the monthly level to calculate Total Revenue, which was then visualized using a Line Chart figure to show the overall trend. This analysis was then segmented by Categorized Dress Type, and a separate segmented Line Chart figure was created to plot the individual monthly revenue trends for Gown, Lehanga, and Kurta against time.

- **Justification :** The primary goal of this analysis was to definitively quantify the seasonal demand cycles, a core project objective. The Line Chart visually confirms the strong upward sales ramp leading into the September and October peak months. Segmenting the revenue is essential for targeted production planning, as it reveals category-specific seasonality—for instance, the dominance of Gown sales in October—allowing the owner to forecast specific raw material needs and schedule tailoring capacity with precision.

- **Correlation Analysis**

- **Process :** Correlation analysis was performed using the Pearson Correlation Coefficient between all numerical variables. This relationship was visualized in two Heatmaps: Figure 4 showed the full correlation matrix between financial metrics and customer measurements (Bust/Waist), while Figure 5 displayed the Sales Correlation between the monthly revenue time series of the top three categories (Gown, Lehanga, Kurta).

- **Justifiaction :** Correlation analysis provides critical, non-obvious insights for inventory and policy optimization. Figure 4 validates the robust financial policy by confirming the strong correlation ($p > 0.8$) between Total Amount and Advance Paid. Figure 5 is key for joint inventory planning, as the strong correlation ($p = 0.71$) between Kurta and Lehanga sales confirms shared seasonal demand, necessitating integrated material stocking. Conversely, the weak correlation ($p < 0.4$) between measurements and Total Amount informs pricing strategy by confirming size is not a price driver.

9. Result and Findings :

The digitized data immediately yielded actionable insights, validating the project's objective to move the boutique toward data-driven decision-making. The analysis confirmed the business operates primarily on a high-variability, high-value model highly susceptible to pronounced **seasonal peaks** aligned with the festive calendar. Furthermore, the segmentation revealed the Gown and Lehanga categories are the most critical revenue drivers, while correlation analysis provided the specific co-demand patterns necessary for optimizing inventory and production scheduling in the months leading up to the peak season.
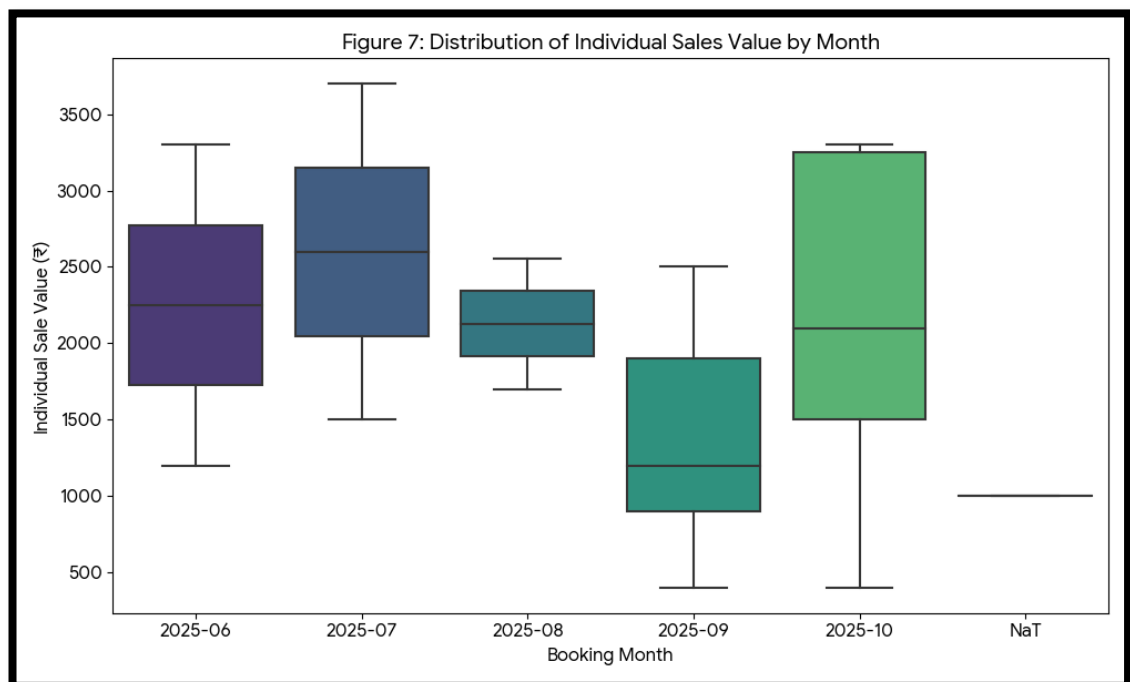

Image 7 : Box plot showing distribution of sales over the months

A box plot was generated to analyze the distribution and variability of individual sale values across the recorded months. This visualization helps in understanding the boutique's pricing strategy and sales consistency throughout the year.
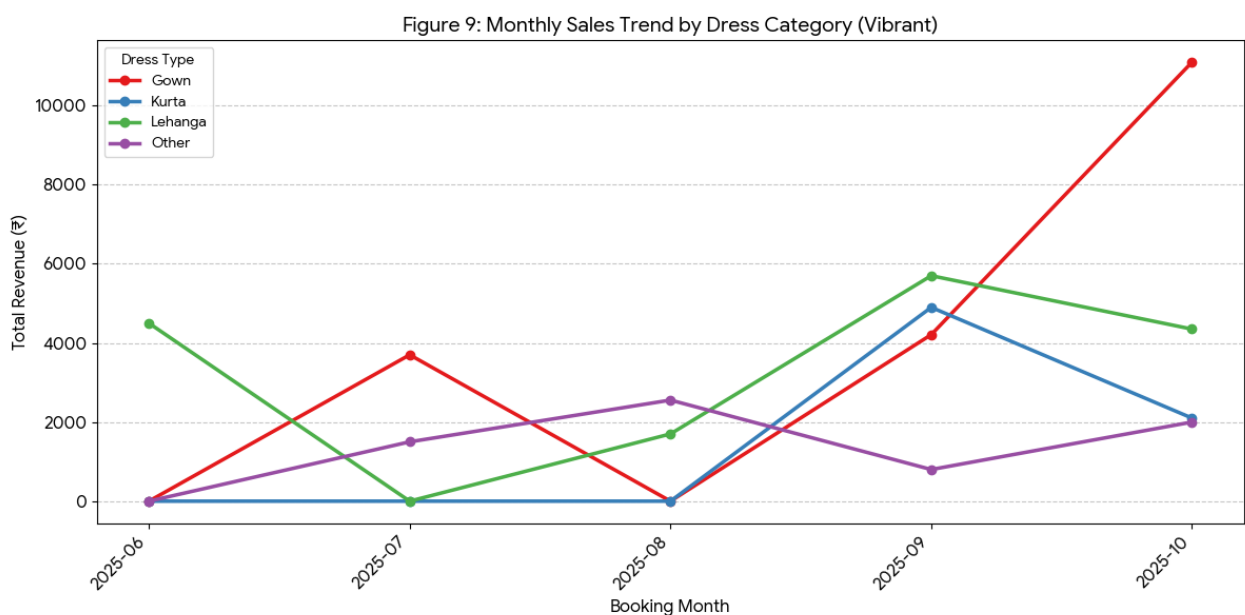
This box plot illustrates the range, median, and outliers for the transaction value in each month. The increasing height of the boxes from August to October indicates that while the sales count increases, the **variability in price** also grows, suggesting the boutique handles a

9

wider mix of high- and low-value orders during the peak season. The median sales price remains relatively consistent around 1000 to 1700 across all months, even as the maximum transaction value increases.

Category Analysis Points :
- **October (2025-10) :** This month exhibits the highest maximum sale value (3700), showing the boutique secures the most premium, high-value custom orders during the peak festive season.

- **September (2025-09) :** This month has the tightest lower quartile (25th percentile) boundary, suggesting that the entry-level price point for sales is slightly higher or more consistent than in other months.

- **June, July, August:** These earlier months show fewer transactions but have similar median values to the peak season, indicating that the baseline pricing and product mix remain stable even during the off-season.

The line chart provides a clear visualization of how demand for each product category contributes to the overall revenue seasonality. It highlights a consistent upward trend across all major categories from June, culminating in pronounced peaks in September and October, directly correlating with the Indian festive and wedding season. Gown sales (blue line) stand out as the highest-value category, showing a dominant spike in October, while Lehanga and Kurta sales demonstrate strong, concurrent peaks, which reinforces their $p = 0.64$ correlation and suggests they share similar demand drivers. Conversely, sales outside this peak period are highly sporadic and volatile across all lines, emphasizing the necessity of a data-driven forecasting system to manage inventory efficiently for both high-demand and low-demand months.
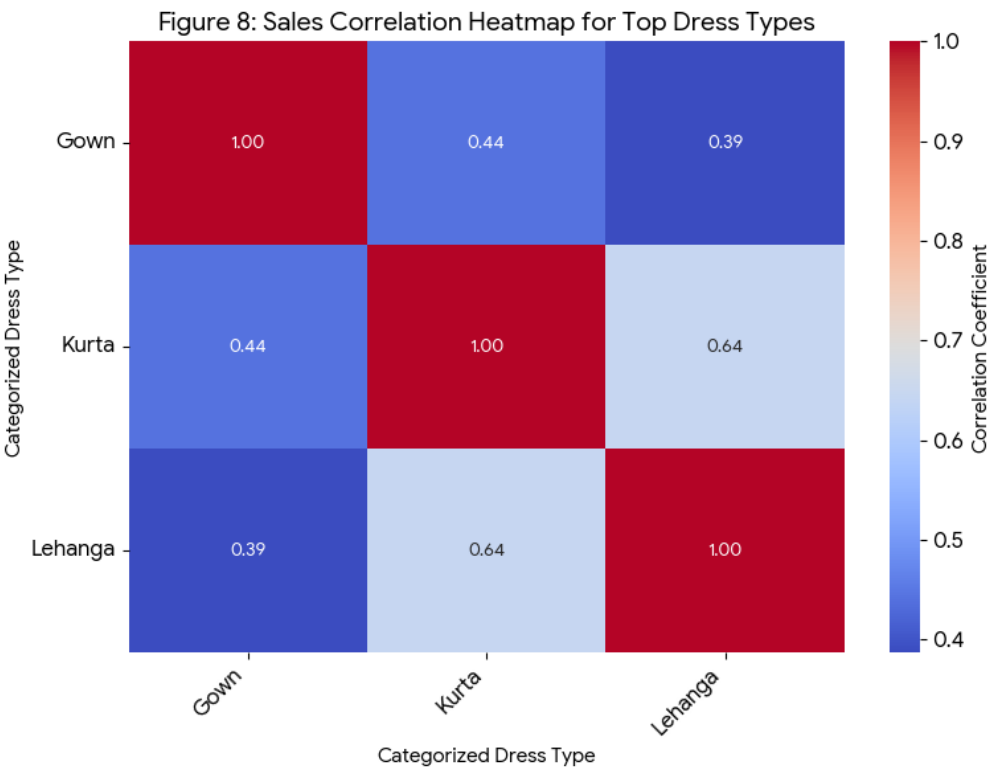

Figure 9: Monthly Sales Trend by Dress Category (Vibrant)

Seasonal Peaks:
Gown sales are the highest value driver, peaking most intensely in October. Lehanga and Kurta demonstrate strong concurrent demand, peaking together across September and October, confirming they share similar demand drivers.

Sales in the early months (June, July, August) are volatile and dominated by individual, sporadic high-value transactions, making them the most challenging periods to forecast.

The Sales Correlation Matrix analyzes the relationships between the monthly revenue streams of the three most profitable categories. This analysis is fundamental for optimizing inventory by predicting which product demands move together.

Figure 8: Sales Correlation Heatmap for Top Dress Types

The correlation analysis yields specific, actionable insights into demand dependencies across the categories:
The **strongest correlation is observed between Kurta and Lehanga** (p = 0.64). This significant relationship indicates that sales in these two categories are influenced by similar seasonal factors, such as mid-level festive demand. For inventory and production planning, this means that an anticipated surge in demand for Kurtas must be met with proportional preparation for Lehanga materials and tailoring capacity. This joint planning will optimize resource allocation.

A **moderate correlation exists between Gown and Kurta** (p = 0.44). While sales of both categories increase during the peak season, their relationship is not as tightly coupled as Kurta and Lehanga. This suggests that Gowns, being the highest-value items, are often driven by more exclusive occasions (like pre-wedding events) that may not perfectly align with the

broader, more general demand drivers of Kurtas.

The **weakest correlation is observed between Gown and Lehanga** (p = 0.39). Although both are formal wear, their monthly sales cycles are somewhat independent. This suggests that the high-end Gown segment may have its own distinct demand cycle, separate from the demand driving Lehanga sales. Consequently, the owner can treat the Gown category as a partially independent product line for inventory forecasting, relying less on simultaneous demand spikes from the Lehanga segment.