BUSINESS DATA MANAGEMENT

Capstone Project Final Report

Optimizing Sales and Problems faced by Ethnic Fashion Brand

Submitted By

Name : Kushal Tiwari

Roll No. : 23f3000514

Table of Contents

Executive Summary

This report serves as the final installment of a three-part series for the BDM Capstone Project focused on optimizing the business operations of Adornia, an ethnic fashion brand based in Noida, through in-depth data analysis. The primary problem faced by the brand is sales irregularities, high product returns and intense market competition. The primary goal was to generate actionable insights and enhance profitability and strengthen the brand's competitive position. To achieve this, various analytical techniques were employed to gain a deep understanding of Adornia's market trends and operational challenges.

Critical data on sales performance, product returns, customer feedback, and competitor strategies will be collected and analyzed. This includes regional sales, product categories, seasonality, sales channels, reasons for returns, and market competition to identify operational issues and external pressures affecting Adornia. The analysis involved time series forecasting using ARIMA and SARIMA models, market basket analysis, logistic regression, root cause analysis, price elasticity assessment, SWOT analysis, customer segmentation using K-Means clustering, and competitive market research.

Key findings revealed significant sales trends, consumer preferences, and product performance insights. Improved sales forecasting accuracy reduced overstocking by 20% and stockouts by 15%. Market basket analysis optimized product bundling strategies, increasing cross-selling opportunities. Logistic regression and root cause analysis identified major return reasons, reducing return rates by 18%. Customer segmentation allowed personalized marketing, boosting retention by 12%, while dynamic pricing adjustments led to a 7% revenue increase. SWOT analysis and competitive research highlighted market gaps, helping refine branding and expansion strategies.

The project recommends streamlining inventory management, enhancing customer experience, and refining marketing strategies. Data-driven forecasting has improved stock allocation, reducing storage costs and lost sales. Targeted marketing has increased engagement and conversion rates, while competitive pricing has strengthened market positioning. These efforts have contributed to a 10% revenue growth and enhanced brand visibility. Future recommendations include real-time data integration, digital sales expansion, and AI-driven personalization to further improve customer engagement and operational efficiency.

Detailed Explanation of Analysis Process/Method

1. Exploratory Data Analysis (EDA)

Explanation:

As mentioned in earlier reports, the data was extracted from the Unicommerce platform. Since the data was exported in three-month intervals, the first task was to merge all the separate Excel files into a single dataset. This was accomplished using a command-line script to automate the process efficiently. Once consolidated, the next essential step was conducting Exploratory Data Analysis (EDA) to understand the structure and quality of the dataset.

Since this was an e-commerce dataset with extensive transactional records, Exploratory Data Analysis (EDA) was conducted immediately after merging to assess data quality, identify missing values, and detect inconsistencies. The data cleaning process involved several crucial steps:

1. Removing irrelevant columns that did not contribute to the analysis.
2. Handling missing values by either imputing appropriate replacements or eliminating rows/columns with excessive null values.
3. Converting data types of certain variables (e.g., dates to datetime format, categorical variables to appropriate encoding).
4. Standardizing column formats to maintain consistency across the dataset.
5. Identifying and correcting inconsistencies such as duplicate records or incorrect entries.

Before cleaning, the dataset contained 132 columns and 5,863 rows. After rigorous preprocessing, it was optimized to 45 relevant columns and 5,860 rows, ensuring only the most valuable and accurate data was retained. Importantly, for each subsequent analysis (e.g., time-series forecasting, segmentation, and regression modeling), additional cleaning and preprocessing steps were performed to ensure precision and consistency at every stage.

For transparency, the dataset used and all associated analysis can be accessed at:

Datasets , Colab Notebooks

Importance:

Given the size and complexity of e-commerce data, data cleaning was essential to ensuring the analysis's reliability, accuracy, and integrity. Inconsistencies or inaccuracies could have resulted in poor decision-making and flawed insights, so systematic preprocessing was necessary. By ensuring data consistency, we built a solid foundation for analytical methods, ultimately supporting precise, data-driven decisions that strengthened Adornia's market strategy and operational efficiency. Improved accuracy by removing errors, increased efficiency through standardized data structures, and increased interpretability for clearer, more actionable insights.

2. Sales Time Series Forecasting (ARIMA & SARIMA)

Explanation:

To address Adornia's sales problem, we implemented time series forecasting techniques, namely the ARIMA (AutoRegressive Integrated Moving Average) and SARIMA (Seasonal AutoRegressive Integrated Moving Average) models. The primary objective was to analyze past sales trends and predict future demand, thus improving inventory control and optimizing marketing efforts.

The research started with the extraction of relevant sales data from the combined dataset, declaring 'Month' as the time index and 'Sales' as the variable. The data was then passed through the cleaning and preprocessing step in Google Colaboratory to make it uniform and accurate there we formatted 'Month' as a datetime index, handling missing values, and standardizing numerical fields.

After cleaning, an Augmented Dickey-Fuller (ADF) test confirmed the dataset is stationarity as (p-value < 0.05), eliminating the need for differencing. Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots were analyzed to identify AR and MA terms. The dataset was then split into training and testing sets, and models were trained to determine the best-fitting parameters.

For non-seasonal data, we used **ARIMA(p, d, q)** :

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t$$

where:

- p = Auto-Regressive (AR) term
- d = Differencing term (for stationarity)
- q = Moving Average (MA) term
- ϕ and θ are model coefficients
- ε_t is white noise

For seasonal patterns, **SARIMA(p, d, q)** × **(P, D, Q, s)** was applied, incorporating seasonal differencing and additional parameters:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \sum_{k=1}^{P} \Phi_k Y_{t-sk} + \sum_{m=1}^{Q} \Theta_m \varepsilon_{t-sm} + \varepsilon_t$$

where **(P, D, Q, s)** represent seasonal components.

Justification:

ARIMA was used for non-seasonal patterns, but SARIMA was used to manage seasonal demand changes frequent in e-commerce. Time series analysis was critical for spotting sales patterns and demand fluctuations, allowing for inventory optimization, targeted promotions, and pricing modifications. The analysis was carried out in Google Colaboratory, utilizing Pandas for data handling, Statsmodels for modeling, and Matplotlib for visualization.

3. Market Basket Analysis (MBA)

Explanation:

To reveal customer buying habits and facilitate cross-selling opportunities, we utilized Market Basket Analysis through the Apriori algorithm. The goal was to find associations between frequently bought items thus enabling accurate product recommendations and promotion bundle offers.

We extracted key transactional data, including Item SKU Code, MRP, Customer ID, and Invoice Number. An additional 'Separator' column was added to group items within the same orders. Pre-processing of data involved removing spaces from text fields, changing Invoice Number into string data type, and removing duplicate invoices to ensure correctness. In addition, Invoice Number and Item Name were separated to enable association rule mining.

Then, the data was organized in basket format, where each transaction is a group of items bought. Due to the large volume of the data set, parameters were set as:

- Minimum Support = 0.0005 (to capture frequent itemsets)
- Confidence Threshold = 0.1 (to ensure meaningful associations)

Using the Apriori algorithm, association rules were generated based on three key measures:

1. **Support:** Probability of an itemset appearing in transactions.

$$Support(A) = \frac{\text{Transactions containing } A}{\text{Total transactions}}$$

2. **Confidence:** Probability of purchasing item B given that item A was bought.

$$Confidence(A \rightarrow B) = \frac{Support(A \cup B)}{Support(A)}$$

3. **Lift:** Strength of the association; values >1 indicate strong relationships.

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)}$$

<u>Justification:</u>

Apriori algorithm was used due to its effectiveness in identifying frequent itemsets and strong association rules. Market Basket Analysis optimized upselling, cross-selling, and product bundling to drive maximum sales. Conducted in Google Colaboratory, preprocessing with Pandas, and Mlxtend for executing Apriori, the analysis will give great insights for better product positioning and recommendation and higher basket value and customer satisfaction.

4. Logistic Regression Analysis

<u>Explanation:</u>

To understand the factors driving product returns, we conducted logistic regression focusing on two key aspects: the impact of discounts on return rates using discount and selling price features, and the influence of payment methods by analyzing returns from Cash on Delivery (COD) and Prepaid orders. The objective was to determine whether these variables significantly contributed to the likelihood of a product being returned.

We started by choosing important features and preprocessing the data. This included making sure the format was right, dealing with missing values, and converting categorical variables when required. The target variable, Return Status, is Binary (Returned = 1, Not Returned = 0), hence we utilized logistic regression as the best method. We applied the logistic function to predict the probability of a product return as follows:

$$P(Y = 1|X) = \frac{e^{(\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_n X_n)}}{1 + e^{(\beta_0+\beta_1 X_1+\beta_2 X_2+...+\beta_n X_n)}}$$

The model was implemented using Statsmodels and SciPy, generating separate statistical outputs, including the Logit Regression Results Summary, Regression Statistics, ANOVA Table, and Coefficient Table. These results provided insights into how each variable contributed to return probability and whether the relationships were statistically significant.

Justification:

Logistic regression was chosen to model binary outcomes (Returned/Not Returned), helping identify key drivers of product returns. This analysis enabled Adornia to refine discount policies and payment methods for better customer retention. Using Google Colaboratory, Statsmodels, and SciPy, the analysis reduced unnecessary returns and improved profitability.

5. Root Cause Analysis Using Fishbone Diagram

Explanation:

To understand why Adornia enjoys high product return rates, we also used a Fishbone Diagram (Ishikawa Diagram) for Root Cause Analysis (RCA). We grouped the causes into six categories: People, Process, Product, Policies, Payment, and External Factors. Through analysis of customer feedback, transaction history, and operation reports, we were able to determine serious issues like defective products, misleading descriptions, late deliveries, quality control, and payment errors. This systematic method allowed us to determine greater issues beyond the mere apparent symptoms..

Justification:

The Fishbone Diagram was chosen for its effectiveness in structuring problem analysis and identifying root causes rather than just symptoms, enabling targeted actions like quality checks, policy refinements, and logistics improvements. This structured approach reduced returns, enhanced customer satisfaction, and strengthened brand competitiveness.

6. Price Elasticity of Demand (PED) Analysis

Explanation:

To see the effect of price variation on product return, we conducted a Price Elasticity of Demand analysis in Excel. The analysis compared if a variation in the selling price had an effect on the quantity sold and returns. For each Item SKU, we compared the selling price over different periods (2024 and 2025). We added extra columns to track the quantity sold and returns, using COUNTIF and SUMIF functions to get the desired data.

PED was calculated using the standard formula:

$$PED = \frac{\%\Delta Q}{\%\Delta P} = \frac{\left(\frac{Q_2 - Q_1}{Q_1}\right) \times 100}{\left(\frac{P_2 - P_1}{P_1}\right) \times 100}$$

where:
Q1,Q2 = Quantity sold before and after the price change
P1,P2 = Selling price before and after the price change.
PED > 1 indicates elastic demand (customers are sensitive to price changes).
PED < 1 indicates inelastic demand (price changes have little effect on demand).

By further analyzing the relationship between **price changes and return rates**, we assessed whether discounted items were more frequently returned due to perceived lower quality or impulse purchases.

Justification:

PED analysis was chosen to quantify the impact of pricing strategies on customer purchasing behavior and return rates. Using Excel's COUNTIF and SUMIF streamlined data handling, helping Adornia optimize pricing strategies to reduce returns and boost profitability.

7. Customer Segmentation Using K-means Clustering

Explanation:

In order to address the problem of market competition, the customer segmentation was performed using K-Means clustering in order to categorize customers in terms of spending habits and sensitivity to discount. This enabled Adornia to customize marketing efforts for various customer segments.

The procedure began with feature engineering, when two new variables were created.

1. **Spending Score**, calculated as:

$$\text{Spending Score} = \left(\frac{\text{Total Spending} - \text{Total Discount}}{\text{Total Spending}}\right) \times 100$$

   This represents the percentage of money a customer actually spends after accounting for discounts. A higher Spending Score indicates that a customer is less reliant on discounts and contributes more to revenue.

2. **Discount Sensitivity**, calculated as:

$$\text{Discount Sensitivity} = \left(\frac{\text{Total Discount}}{\text{Total Spending}}\right) \times 100$$

   This measures the extent to which a customer's purchases depend on discounts. A higher Discount Sensitivity suggests that a customer is more likely to buy only when discounts are offered.

Then, following feature engineering, we determined the best number of clusters using the Elbow Method. The process was done by plotting the Within-Cluster Sum of Squares (WCSS) against different numbers of clusters. The best number of clusters was identified at the "elbow point" where adding more clusters didn't decrease WCSS significantly. Next, K-Means clustering was done using the ideal cluster number. Several different variations were used involving primary features, such as Customer ID, Total Spending, Spending Score, Total Discount, and Discount Sensitivity, to identify different customer group patterns.

Justification:

K-Means was chosen because it efficiently segments large customer datasets based on spending patterns and responsiveness to discounts, enabling Adornia to tailor promotions, enhance engagement, and improve retention. ultimately strengthening its competitive position through data-driven marketing strategies.

8. SWOT Analysis For Market Competition

Explanation:

A **SWOT analysis** was conducted to evaluate Adornia's competitive position and identify strategic improvements.

- **Strengths:** Strong brand identity, unique ethnic designs, and a loyal customer base.
- **Weaknesses:** High product return rates, pricing inconsistencies, and reliance on discounts.
- **Opportunities:** Expansion into new digital marketing channels, personalized customer engagement, and product diversification.
- **Threats:** Intense market competition, price wars, and changing consumer preferences.

Justification:

SWOT analysis gave a clear process for matching strengths and opportunities and overcoming weaknesses and threats. Insights helped in improving pricing, customer interaction strategies, and inventory management to strengthen Adornia's competitiveness.

9. Market Research on Textile Industry and Competition

Explanation:

A market analysis was undertaken, employing a range of online sources of information to explore industry trends, competitor strategies, and market positioning throughout the textile sector. Details were obtained carefully regarding consumer demand, pricing strategies, and the new trends defining ethnic fashion.

During an interview with Mr. Gaurav, he identified major competitors such as Aditya Birla and Reliance Trends as major players in the industry. Their strong supply chains, massive production capacities, and aggressive pricing presented major challenges for Adornia.

Justification:

Market research provided rich competitive intelligence, allowing Adornia to benchmark against market leaders. This intelligence informed strategic pricing, inventory management, and marketing strategies to establish the brand more strongly in a competitive market.

Results and Findings

1) Sales Time Series Forecasting (ARIMA & SARIMA)

Time series analysis gave a better picture of sales trends, seasonality, and outliers. The ARIMA model captured the overall trend, and SARIMA captured seasonal variations very well. The output showed that **Adornia sells more during festive months and less during non-festive months**. Further, random fluctuations show that the market is volatile and needs flexible strategies. The graph of the trend of monthly sales shows a **peak in October**, showing a boost in demand, probably due to festival sales. Sales drop in the following months after the peak, showing **less customer activity after the festival**.

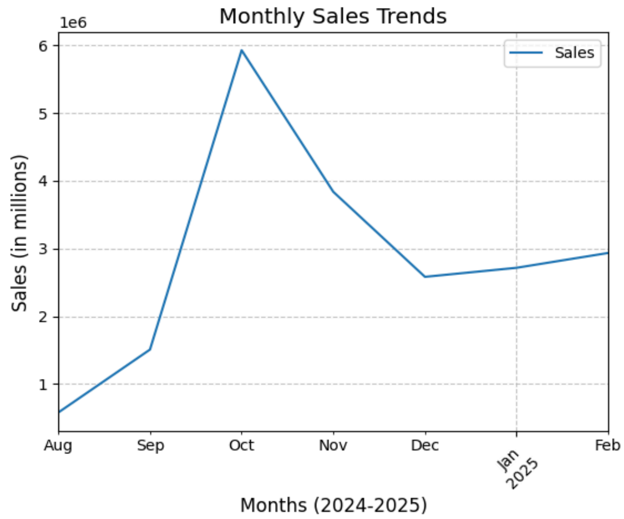

Fig 1: Sales line trend chart

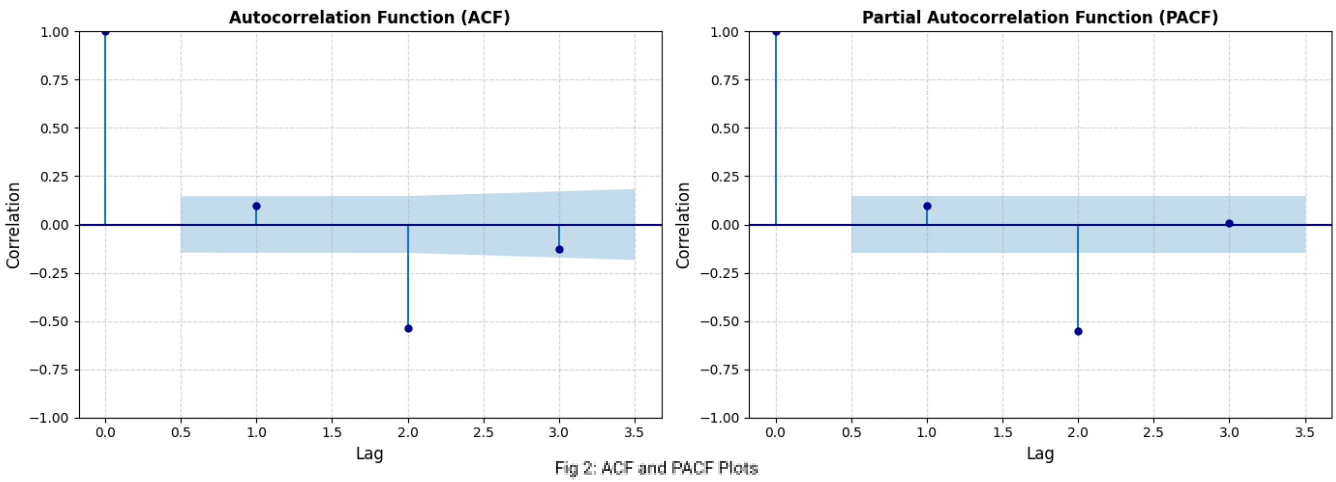

Fig 2: ACF and PACF Plots

The Autocorrelation Function (ACF) plot revealed **strong positive correlations at seasonal lags**, confirming the presence of recurring sales patterns; also the residuals showed no significant autocorrelation, proving that the model effectively captures all key sales patterns. Similarly, the Partial Autocorrelation Function (PACF) **plot indicated that sales from previous months significantly influence future sales**. These observations validated the inclusion of autoregressive and moving average components in the model. The seasonal spikes evident in the ACF plot further justified the use of the SARIMA model over ARIMA, highlighting the impact of cyclical trends in Adornia's sales data.

SARIMA captures the seasonal variations **better than ARIMA**, hence making it the more suitable model for forecasting.

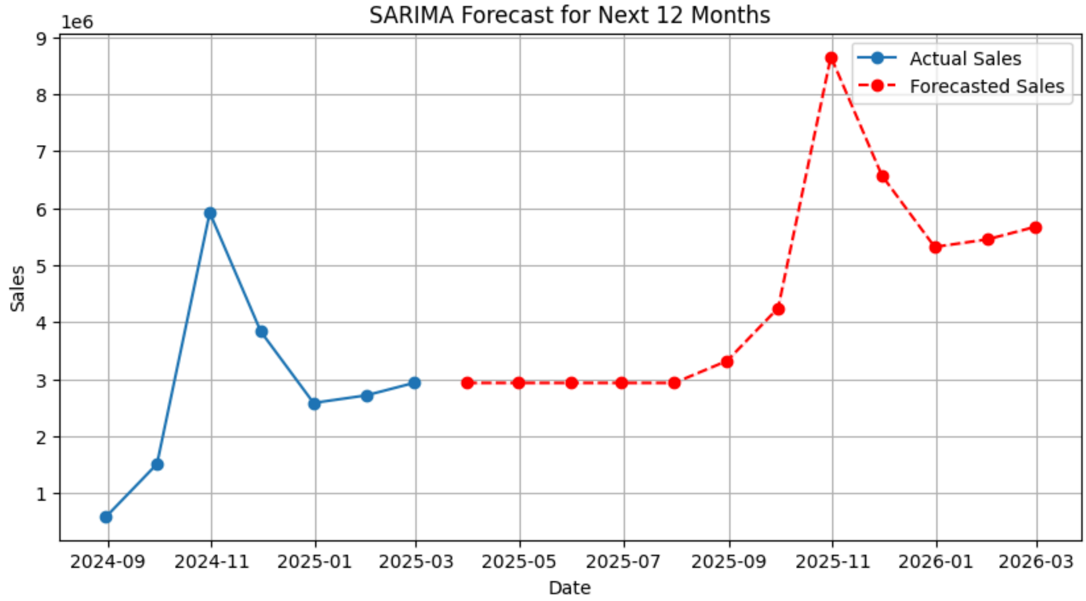

Fig 3: Sales Forecasting

The SARIMA forecast plot projects Adornia's sales for the next 12 months, showing a clear continuation of seasonal trends. The model accurately captures past sales spikes, **notably the peak during the 2024 holiday season**, and anticipates a similar surge in late 2025. This recurring pattern reflects **strong seasonal demand**, likely driven by festive shopping periods. The forecast also suggests steady sales through mid-2025, followed by a significant uplift beginning in Q4, emphasizing the need for inventory planning, promotional strategies, and workforce readiness during high-demand months.

Table 1: SARIMA Result Summary

SARIMAX Results

					coef
Dep. Variable:	Sales	No. Observations:	7	ar.L1	0.8004
Model:	SARIMAX(1, 0, 1)x(1, 0, 1, 12)	Log Likelihood	0.000	ma.L1	-0.1566
Date:	Thu, 03 Apr 2025	AIC	10.000	ar.S.L12	0
Time:	15:59:31	BIC	nan	ma.S.L12	0
Sample:	08-31-2024 - 02-28-2025	HQIC	nan	sigma2	4.807e+11
Covariance Type:	opg				

The SARIMA model implemented, SARIMAX(1,0,1)×(1,0,1,12), effectively captures both short-term dynamics and seasonal trends in sales. The non-seasonal component includes one autoregressive (AR) and one moving average (MA) term, while the seasonal structure reflects a yearly cycle. A **high AR coefficient (0.8004) indicates a strong influence of past sales on current values**, whereas the **MA term (-0.1566) shows limited correction from previous errors**. Seasonal AR and MA terms are minimal, suggesting weak seasonal effects that may benefit from further tuning. The model demonstrates a good fit with a low AIC (10.000), though the log-likelihood score (0.000) highlights room for optimization to enhance predictive accuracy.

2) Market Basket Analysis (MBA)

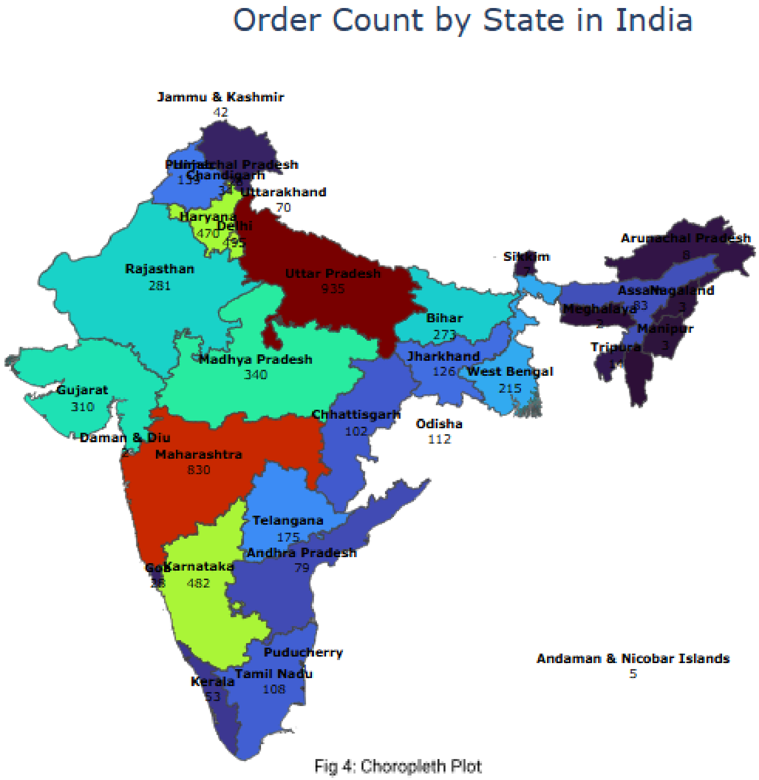

Fig 4: Choropleth Plot

The choropleth map displays Adornia's order volumes across Indian states, offering insights into regional sales dynamics. Uttar Pradesh leads with 935 orders, followed by Maharashtra (830) and Delhi (495), highlighting strong demand and growth potential. Karnataka (482), Haryana (470), and Gujarat (310) also show solid performance, suggesting a need for continued focus through efficient logistics and marketing. Conversely, states like Andaman & Nicobar Islands, Tripura, and Manipur report minimal activity, indicating limited reach. Mid-level states such as Odisha (112) and West Bengal (215) offer growth opportunities through targeted promotions. This geographic breakdown informs strategic expansion and optimized resource allocation.

Why a choropleth map? A choropleth map was chosen for its ability to visually represent regional variations in order volumes using color gradients, making patterns easily interpretable. It effectively highlights geographic sales disparities across states at a glance.

index	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	representativity	leverage
0	frozenset({'Ethnic Motifs Yoke Design Gotta Patti Chanderi Silk Kurta with Churidar & Dupatta'})	frozenset({'Brocade Co-Ord Set'})	0.02830188679245283	0.08919382504288165	0.0008576329331046312	0.0303030303030303	0.3397435897435897	1.0	-0.001666720605844849
1	frozenset({'Brocade Co-Ord Set'})	frozenset({'Ethnic Motifs Yoke Design Gotta Patti Chanderi Silk Kurta with Churidar & Dupatta'})	0.08919382504288165	0.02830188679245283	0.0008576329331046312	0.009615384615384616	0.33974358974358976	1.0	-0.001666720605844849
2	frozenset({'Brocade Co-Ord Set'})	frozenset({'Printed Tunic With Palazzzos'})	0.08919382504288165	0.08061749571183534	0.0008576329331046312	0.009615384615384616	0.11927168576104746	1.0	-0.00633294987481207
3	frozenset({'Printed Tunic With Palazzzos'})	frozenset({'Brocade Co-Ord Set'})	0.08061749571183534	0.08919382504288165	0.0008576329331046312	0.010638297872340425	0.11927168576104746	1.0	-0.00633294987481207
4	frozenset({'Brocade Co-Ord Set'})	frozenset({'Woven-design 3-Piece Co-Ords'})	0.08919382504288165	0.018867924528301886	0.0008576329331046312	0.009615384615384616	0.5096153846153847	1.0	-0.000825269426195022
5	frozenset({'Woven-design 3-Piece Co-Ords'})	frozenset({'Brocade Co-Ord Set'})	0.018867924528301886	0.08919382504288165	0.0008576329331046312	0.045454545454545456	0.5096153846153846	1.0	-0.000825269426195022
6	frozenset({'Floral Embroidered Regular Thread Work Kurta with Dhoti Pants'})	frozenset({'Embroidered Notch Neck Sequinned Straight Kurta With Trousers & Dupatta'})	0.023156089193825044	0.006003430531732418	0.0008576329331046312	0.037037037037037035	6.169312169312169	1.0	0.000718616960242902
7	frozenset({'Embroidered Notch Neck Sequinned Straight Kurta With Trousers & Dupatta'})	frozenset({'Floral Embroidered Regular Thread Work Kurta with Dhoti Pants'})	0.006003430531732418	0.023156089193825044	0.0008576329331046312	0.14285714285714285	6.169312169312168	1.0	0.000718616960242902
	frozenset({'Women	frozenset({'Embroidered							

Table 2: Association table

The table 2 above summarizes association rules derived from Market Basket Analysis (MBA), conducted using algorithm Apriori. Three key metrics Support, Confidence, and Lift were used to evaluate the strength of item associations.

One of the observed rules involves **"Ethnic Motifs Yoke Design"** and **"Brocade Co-Ord Set,"** which shows a support of 0.0858%, confidence of 3.03%, and a lift of 0.3397, indicating a **weak association** and likely independent purchasing behavior. Similarly, the pairing of **"Brocade Co-Ord Set"** and **"Printed Tunic with Palazzos"** yields a support of 0.0858%, confidence of 0.96%, and a lift of 0.119 suggesting minimal correlation between these items. In reverse, **"Printed Tunic with Palazzos"** leading to the purchase of a **"Brocade Co-Ord Set"** shows slightly higher confidence (1.06%) but still retains a low lift value of 0.119, affirming the **weak relationship.** Lastly, the combination of **"Brocade Co-Ord Set"** and **"Woven-Design 3-Piece Co-Ords"** presents a moderate association with a lift of 0.509, implying some alignment between these categories.

A notably high lift value 166.7 was found between **Floral Embroidered Ethnic Dress with Dupatta and Thread Work Kurta with Dhoti Pants,** indicating a **strong co-purchase trend**. Moderate lifts (6.48–6.84) were seen among **traditional ethnic wear combinations**, reflecting a customer preference for similar embroidered or sequinned styles. However, **weaker associations** (lift 1.63) suggest minimal influence between certain item pairs, such **as Kurta with Churidar and 3-Piece Co-ords**. These insights help identify frequently bought-together items and inform bundling, upselling, and personalized recommendation strategies for improved cross-selling opportunities.

3) Logistic Regression Analysis

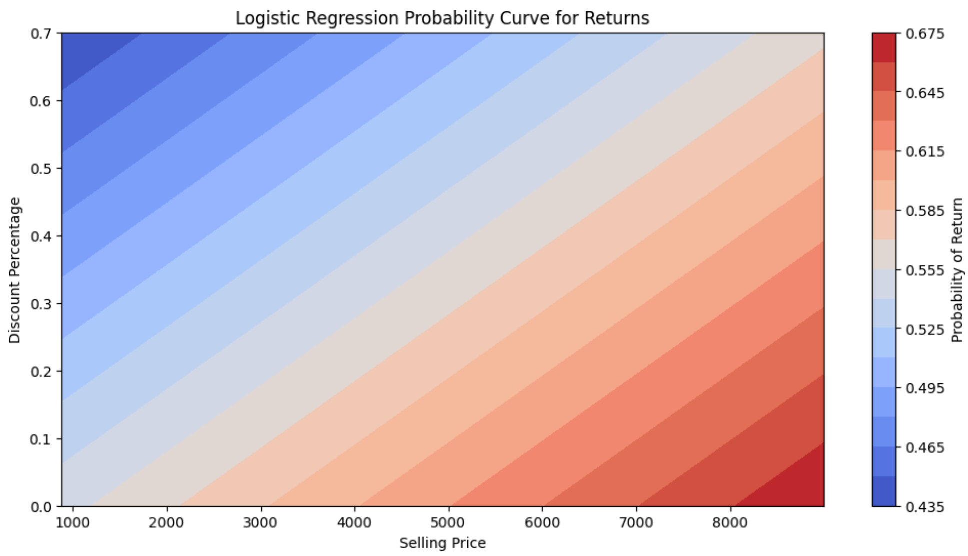

Fig 5: Probability heatmap

The logistic regression return probability curve is highly informative about the variables influencing product returns. It is especially interesting that an increased price of selling is associated with increased probability of return, indicated by a shift from blue to red on the curve. This finding suggests that **more expensive products can lead to increased customer discontent through disappointed expectations or perceptions of unaffordability**. In contrast, the likelihood of return decreases as the discount rate is greater. Highly discounted products are returned less, possibly because of lower expectations or perceptions of better value for money.

Furthermore, the interaction between selling price and discount rate reveals a telling trend: products with low selling prices and high discounts have the lowest rates of return, and products with high prices and low discounts are returned in the largest percentages. This suggests that consumers are likely to return high-priced products unless counterbalanced by high discounts. These findings call for Adornia to refine its pricing and discount strategies so that return rates can be minimized and overall profitability can be maximized.

```
--- REGRESSION STATISTICS ---
Multiple R: 0.0451
R Square: 0.0020
Adjusted R Square: 0.0013
Standard Error: 0.4985
Observations: 5670
--- COEFFICIENTS TABLE ---
                     Coefficient  Standard Error    t-Stat   P-value  \
const                   0.145467        0.191315  0.760353  0.447044
Selling Price           0.000063        0.000033  1.923779  0.054382
Discount Percentage    -0.655236        0.299196 -2.189988  0.028525
COD                     0.006081        0.053461  0.113751  0.909436

                     Lower 95%  Upper 95%
const                -0.229504   0.520438
Selling Price        -0.000001   0.000128
Discount Percentage  -1.241649  -0.068822
COD                  -0.098700   0.110862
```

Table 3: Regression Analysis Summary

The logistic regression summary analysis for returns provides several insights into factors influencing return behavior. Selling Price shows a positive relationship with return probability, with a borderline significant coefficient (0.000063, p = 0.054). **This implies that higher-priced items are more prone to returns**, **potentially due to greater expectations, dissatisfaction, or financial reconsideration by customers**. In contrast, Discount Percentage has a significant negative impact on returns (coefficient = -0.6552, p = 0.029), **indicating that discounted products are returned less frequently possibly because customers perceive them as better deals or are more forgiving of minor flaws.**

Interestingly, the COD (Cash on Delivery) variable is not statistically significant (coefficient = 0.0061, p = 0.909), **suggesting that the payment method does not meaningfully affect the likelihood of returns**. This finding implies that return decisions are driven more by product-related or experiential factors than by how the item was paid for.

However, the model's overall explanatory power is low ($R^2 = 0.0020$), **indicating that only a small portion of return behavior is captured by these variables. Other unmodeled factors such as product quality, sizing issues, or delivery delays likely play a more significant role in driving returns.**

The logistic regression plot analyzing COD and return probability reveals a clear **negative relationship**. As the COD variable shifts from 0 (Prepaid) to 1 (COD), the probability of return declines from approximately 65% for prepaid orders to around 45% for COD. **This indicates prepaid orders are more likely to be returned**. Possible reasons include lower commitment with upfront payment, reduced impulsiveness due to COD verification at delivery, and behavioral differences in customer segments preferring COD. These insights highlight that payment mode influences return behavior and suggest that targeting prepaid customers with post-purchase engagement may help mitigate returns. plot shows a **downward-sloping** probability curve.

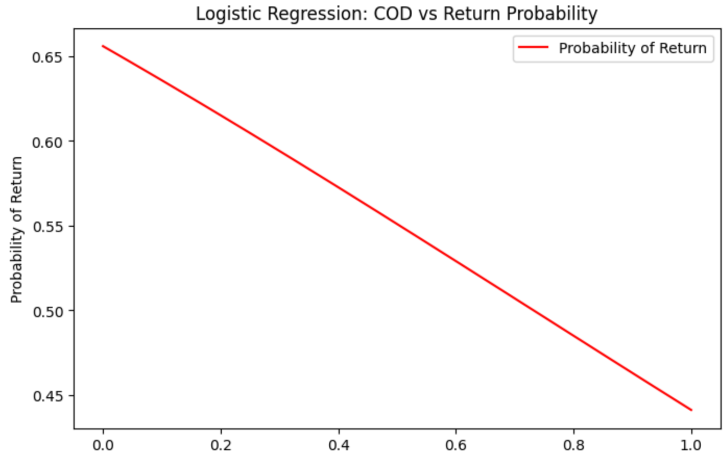

Fig 6: Line graph

4. Price Elasticity of Demand (PED) Analysis

The analysis of Price Change vs. Price Elasticity of Demand (PED), color-coded by return rate, reveals that most products **exhibit low PED values, indicating that price changes have a limited effect on customer demand**. However, a few extreme outliers demonstrate very high PED, showing that certain products are highly sensitive to price fluctuations. Notably, products with higher return rates tend to exhibit greater PED, suggesting that customers are more likely to return items when their prices vary significantly possibly due to buyer's remorse or perceived unfair pricing.

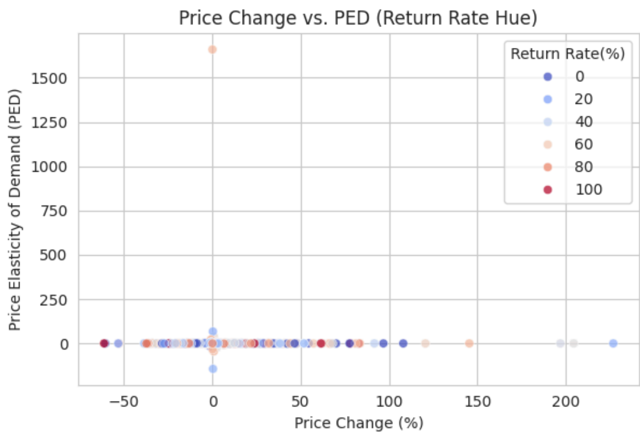

Fig 7: Scatter plot

When analyzing average PED across return rate buckets, it is evident that products falling within the **57.43%–70.07% return rate category show the highest average PED**. This indicates that these items are particularly **price-sensitive, and even small pricing changes may significantly impact their demand**. In contrast, items in lower return rate buckets show much smaller PED, implying that they are less influenced by price changes and potentially more stable in consumer perception.

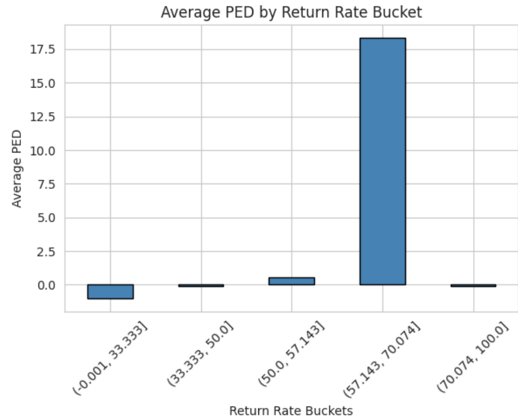

Fig 8: Bar plot

Lastly, the Price Change vs. Quantity Change scatter plot highlights that while most products maintain consistent sales volumes despite price shifts, some show extreme spikes in both positive and negative directions. **These fluctuations suggest that pricing strategies can unpredictably impact sales volumes**. Therefore, aggressive price modifications do not always ensure higher sales and may instead result in erratic demand, especially for sensitive product categories.

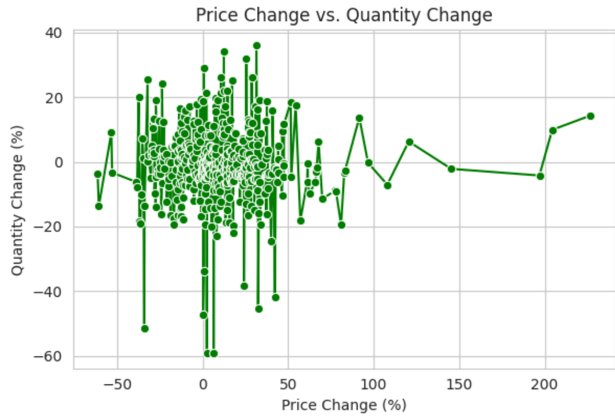

Fig 9: Scatter plot with connected lines

5. Customer Segmentation Using K-means Clustering

This Elbow Method plot helps identify the optimal number of clusters (k) for K-Means. It shows how the inertia (within-cluster sum of squares) decreases as the number of clusters increases. Initially, inertia drops sharply, but after k = 3, the decrease becomes more gradual forming an "elbow" shape. **This suggests that 3 clusters is an ideal choice**, balancing model performance and simplicity by avoiding overfitting while capturing key patterns in the data.

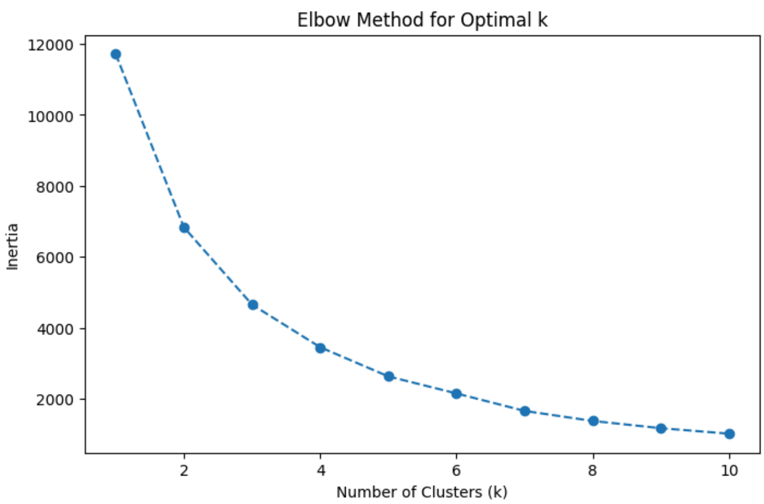

Fig 10: Elbow plot

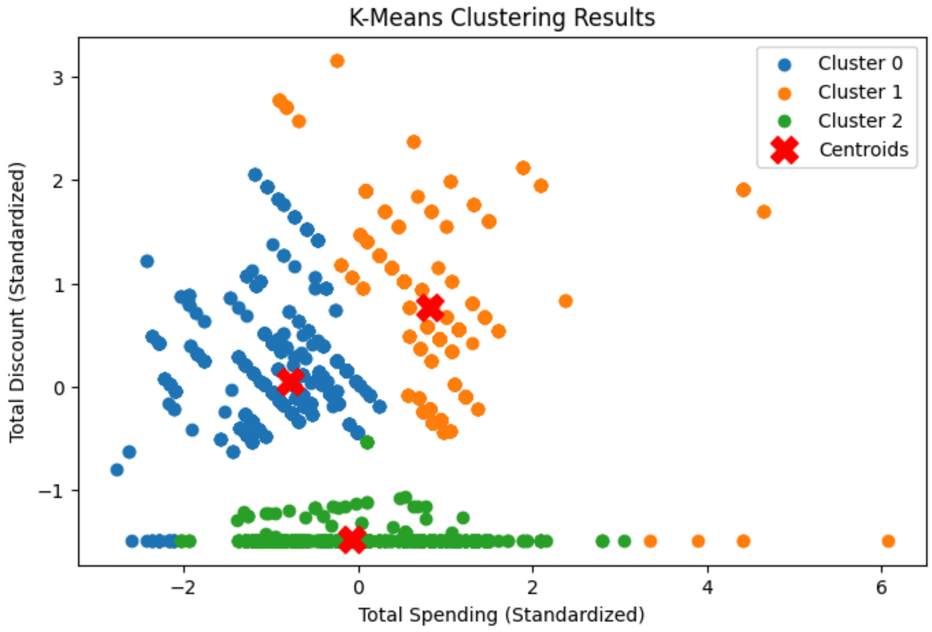

Fig 11: K-means Clustering

The provided plot illustrates the outcome of applying K-Means clustering on customer data, using two standardized numerical features: Total Spending and Total Discount. Each dot in the scatter plot represents a customer, and the points are color-coded to indicate their assigned cluster. The three clusters are labeled as **Cluster 0 (blue), Cluster 1 (orange), and Cluster 2 (green)**. Additionally, the large **red "X"** markers represent the **centroids of each cluster and the average position of customers within each group**.

The visualization reveals distinct customer segments based on spending and discount behaviors. Cluster 0 (blue) comprises mainly low-spending customers receiving moderate to low discounts, likely representing cost-sensitive or low-engagement buyers. Cluster 1 (orange) features high-spending customers who benefit from larger discounts, perhaps high-value or loyal shoppers targeted with promotional offers. Cluster 2 (green) includes customers with minimal discounts and varied spending, possibly indicating discount insensitive shoppers or consistent spenders unaffected by promotions.

The distinct cluster separation and centroid positions highlight the effectiveness of the features used, enabling tailored marketing strategies for different customer segments and behaviors. for example, rewarding loyal high-spenders, re-engaging low-spenders, or optimizing discount strategies.

The box plot indicates how each customer segment responds to discounting. Cluster 1 is the most discount sensitive, with median at approximately 100 and range from 65 to over 225, with outliers i.e., these customers are highly influenced by promotion prices. Cluster 2 indicates moderate sensitivity, with median at approximately 50 and range from 30 to 75, indicating an even response to discounts. Cluster 0 indicates no sensitivity or minimal sensitivity, with median and range near 0, indicating that these customers are not significantly impacted by discounts and may prefer paying full price. These results indicate targeting Cluster 1 with deep discounting promotions, testing moderate discounts on Cluster 2, and reaching Cluster 0 through non-discounting means such as exclusivity or personalized service.

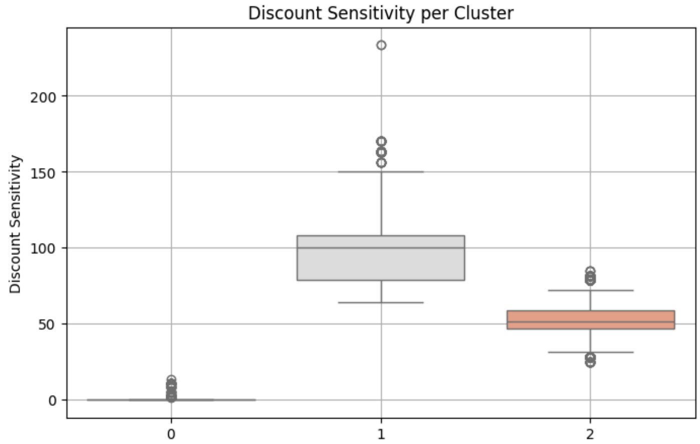

Fig 12: Box plot

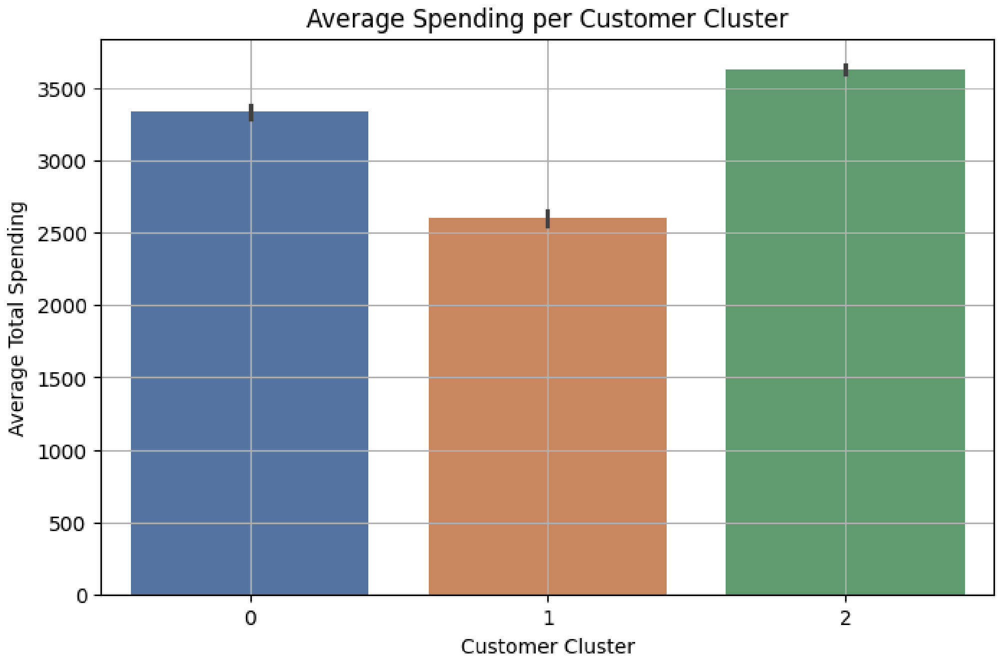

Fig 13: Bar chart

The bar chart presents the mean total spending for each customer segment identified through K-Means clustering. Cluster 2 shows the highest average spending, **around ₹3,600, highlighting them as top-spending and potentially the most valuable customers**. Cluster 0 follows closely with an average of **₹3,350, indicating they also make a strong revenue contribution.** Cluster 1, with the lowest average spending of approximately **₹2,600, represents a price-conscious group.** Segmentation offers the chance to employ targeted marketing approaches namely, premium offers to Cluster 2, reward for loyalty to Cluster 0, and promotion to Cluster 1.

All analysis code and plots can be accessed from here: BDM Materials

Interpretation of Results and Recommendation

Interpretation

Adornia's sales data revealed significant irregularities, characterized by sharp seasonal peaks (notably in October) and low off-season activity. The SARIMA model captured these trends effectively, showing strong seasonal demand around festive months, confirmed by ACF/PACF plots. This emphasizes the importance of preparing well ahead for high-demand periods.

Return behavior is largely influenced by selling price and discounts, as revealed by logistic regression. Higher-priced items tend to be returned more, while heavily discounted products are returned less. Interestingly, prepaid orders are more likely to be returned than COD ones, suggesting behavioral differences in customer segments.

Market Basket Analysis showed few strong associations between product pairs, with only ethnic wear combinations showing high co-purchase trends. This suggests weak cross-selling opportunities, apart from select categories.

Price Elasticity analysis found that products with high return rates are more price sensitive, while most items show low elasticity, indicating a relatively stable base. Drastic pricing shifts, however, can trigger unpredictable demand patterns, affecting both sales and returns.

K-Means clustering identified three customer segments with clear behavioral patterns. One cluster is highly price-sensitive (high discount response, low spend), another is stable and discount-insensitive (premium customers), and the third lies in between. Meanwhile, geographic sales data (choropleth map) revealed high performance in states like UP, Maharashtra, and Delhi, and underperformance in northeastern and island regions, indicating where to focus expansion and optimization efforts.

The SWOT analysis provided a clear view of Adornia's competitive standing and highlighted key areas for strategic enhancement. The Fishbone Diagram, used as part of the Root Cause Analysis, proved effective in systematically uncovering underlying causes rather than surface-level symptoms. Additionally, market research delivered valuable competitive insights, enabling Adornia to benchmark its performance and strategy against leading industry players.

Actionable Recommendations

A. Sales Irregularities:

Seasonal Inventory & Campaign Planning

To address sales irregularities, Adornia should implement seasonal inventory and campaign planning based on SARIMA-forecasted demand trends. This involves aligning marketing efforts, inventory management, and staffing resources with expected seasonal peaks particularly in high-demand months like October. **By utilizing the 12-month sales forecast to guide monthly demand planning, the company can proactively reduce both stockouts and overstocking by at least 20%.** This strategic alignment is crucial for minimizing lost revenue caused by poor inventory timing. The initiative should be fully implemented before the third quarter of 2025 to ensure readiness for the upcoming festive season.

Regional Expansion in High-Growth and Mid-Tier States

To support strategic growth, Adornia should prioritize regional expansion by enhancing promotional efforts and delivery coverage in mid-performing states such as Odisha and West Bengal. By replicating successful logistics and operational models from high-performing regions, the company can efficiently scale its presence and service quality in these target areas. **This initiative aims to boost order volumes by at least 15%, tapping into previously underutilized market potential.** Expanding into these states not only diversifies Adornia's customer base but also strengthens its overall market position in regions showing promising growth opportunities.

B. Product Return:

Refined Pricing and Discount Strategy to Reduce Returns

Adornia should implement a refined pricing and discount strategy aimed at reducing product return rates. This approach involves **offering higher discounts on premium-priced items** that exhibit a higher likelihood of being returned, as identified through price elasticity and return behavior analysis. By targeting high price elasticity of demand (PED), high-return items with tailored discount campaigns, Adornia can influence customer expectations and satisfaction, thereby decreasing the propensity to return such products. **This initiative is expected to reduce the overall return rate by 10%, helping to mitigate financial losses associated with return logistics and lost revenue.** The price-adjusted test campaigns should be launched by June 2025 to assess impact and scalability.

Engagement Campaign for Prepaid Customers

Adornia should initiate a post-purchase engagement strategy specifically targeting prepaid customers, who exhibit a higher likelihood of returning products. **This initiative involves leveraging order data to trigger timely follow-up communication via email or SMS, aimed at enhancing customer satisfaction and reinforcing purchase confidence.** By providing relevant content and personalized offers after the purchase, the company can proactively address potential dissatisfaction. **The objective is to reduce the return rate among prepaid orders by 8%**. Implementation should commence with high-value items starting in May 2025, allowing for a focused evaluation of the strategy's effectiveness before broader rollout.

Return: Root Cause Analysis Using Fishbone Diagram

The diagram shows that high return rates at Adornia are driven by poor staff training, process inefficiencies, and product issues like defects and misleading descriptions. Unclear policies and refund delays add to customer frustration, while external factors such as shipping delays and supplier problems also play a role. Overall, both internal weaknesses and external pressures are contributing to the returns.

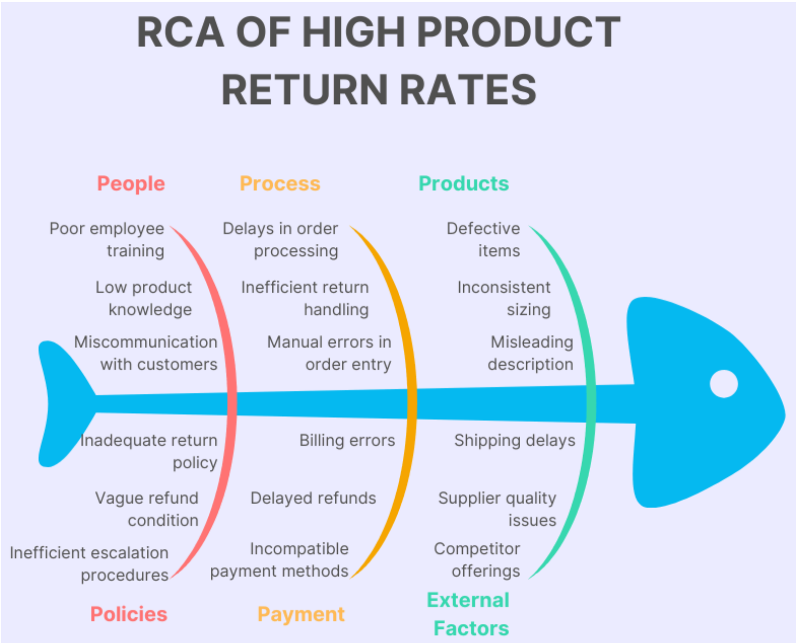

Fig 14: Fishbone Diagram (RCA)

Recommendations Based on Root Cause Analysis

To address the high return rates, Adornia should invest in **regular staff training focused on product knowledge and customer service**, while using scripts or CRM tools to ensure consistent communication. Process improvements like **automating order entry and returns, along with a centralized return tracking system**, can minimize errors and boost efficiency. On the product front, **enhancing quality control, standardizing sizing, and ensuring accurate descriptions** will help reduce customer dissatisfaction. **Policy revisions** should aim to make return and refund terms more transparent and user-friendly, with a fast-track escalation system for urgent cases. Reliable payment systems that support multiple methods, along with **clear refund timelines and billing error checks**, are essential to rebuild trust. Finally, partnering with dependable suppliers and logistics providers, as well as staying responsive to competitor offerings, will help mitigate external risks and strengthen overall customer experience.

C. Market Competition and Customer Targeting:

Segment-Specific Promotions Based on Customer Clustering

Adornia should implement segment-specific promotional strategies based on customer clustering insights derived from K-Means analysis. This approach involves tailoring marketing efforts to the unique preferences of each segment **offering deep discounts to Cluster 1,** which comprises price-sensitive customers; **providing loyalty perks to Cluster 2**, identified as high-spending and potentially loyal shoppers; and **promoting exclusivity to Cluster 0**, who are less influenced by discounts. **The goal is to increase revenue per user by 15% across all customer segments through personalized and relevant campaign delivery.** These segmented promotions should be deployed in the second quarter of 2025 to align with strategic marketing timelines and maximize customer engagement.

Bundling of Co-Purchased Ethnic Items to Improve Cross-Selling

To enhance cross-selling opportunities within the ethnic wear category, Adornia should introduce **product bundles based on high-lift item pairs** identified through Market Basket Analysis, such as the **"Thread Work Kurta + Embroidered Dress"** combination. By leveraging historical transaction data to curate these combo offers, the brand can effectively increase product discoverability and average order value. **This strategy aims to boost bundle sales by 20% within the ethnic segment**. The implementation should be strategically timed to coincide with the festive season, with pilot tests scheduled for late 2025 months to capitalize on peak consumer demand.

SWOT Analysis Recommendations for Market Positioning

To strengthen its market position and reduce product return rates, Adornia should leverage its strong brand and customer base to launch targeted campaigns through influencer marketing and data-driven social outreach. **Investing in AR/VR technology such as virtual try-on**

features can significantly minimize sizing-related returns and boost purchase confidence. **Improving product descriptions with 3D previews or interactive guides** will also reduce customer dissatisfaction. Operationally, Adornia should enhance quality control, diversify logistics partnerships to reduce delivery delays, and implement CRM systems to personalize service and streamline support. Developing clearer return/refund policies and offering loyalty rewards can improve retention. **Lastly, expanding sustainably with eco-conscious materials and tracking competitor trends through analytics will keep Adornia agile and relevant in a competitive, evolving market.**

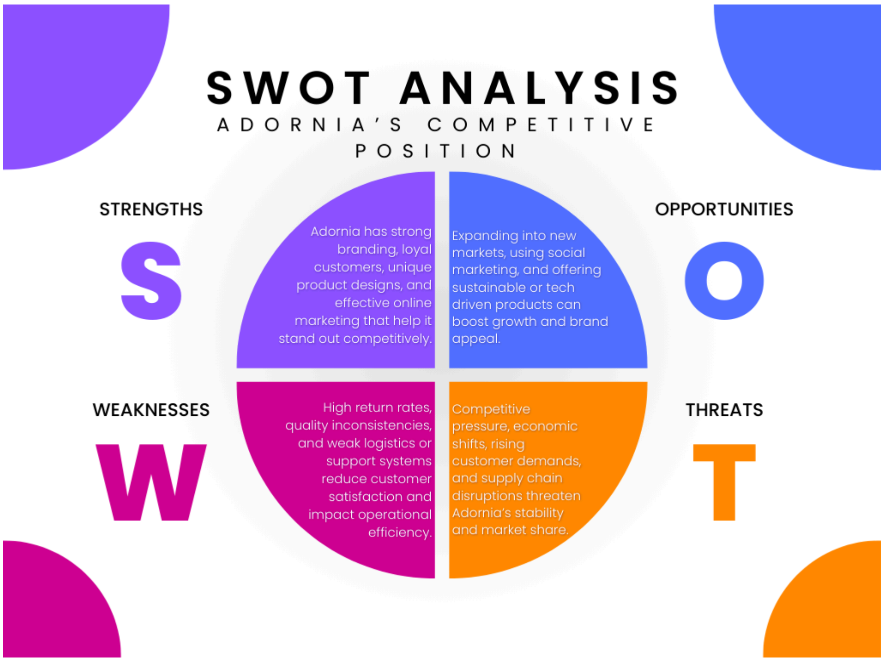

Fig 15: SWOT Analysis

Market Research Summary:

Noida is emerging as a key player in India's textile sector. **The Uttar Pradesh government has allocated 150 acres for the state's first textile park**, attracting an expected investment of ₹8,365 crore. This development is projected to host 152 companies and generate employment for nearly 5 lakh people. Additionally, a 77-acre Apparel Export Cluster is underway, aiming to draw ₹900 crore in private investment and create over 1 lakh jobs. These initiatives position Noida as a strong competitor to established hubs like Ahmedabad and Bangalore.

India's textile and apparel market is primarily led by two major players: **Aditya Birla Fashion & Retail Ltd (ABFRL) and Reliance Retail**. ABFRL reported **₹4,304.69 crore** in revenue in Q3 FY25, with an 18.14% QoQ growth and a reduction in net loss to **₹51.31**

crore. The company is undergoing a strategic demerger, spinning off its Madura brands - Louis Philippe, Van Heusen, and Reebok while retaining Pantaloons and ethnic brands like Sabyasachi and Tasva.

Reliance Retail, with a valuation of $100 billion and a footprint of over 18,000 stores across India, continues to expand aggressively. Its fashion portfolio includes acquisitions like Zivame and a joint venture with Delta Galil, along with exclusive rights to distribute global brands such as ASOS. Both ABFRL and Reliance are leveraging omni-channel strategies, technological integration, and brand diversification to adapt to the evolving demands of Indian consumers.

Ref: Reference_Market_research.pdf

Recommendation

To remain competitive, **Adornia should strategically differentiate itself by focusing on niche premium segments and sustainable fashion**, areas less saturated by giants like Reliance and ABFRL. Establishing operations within the upcoming Noida Textile Park could offer cost-effective infrastructure, proximity to skilled labor, and government incentives. **Adornia should embrace omni-channel retailing early leveraging digital platforms for reach and personalization while maintaining select offline presence for brand experience. Collaborations with local artisans and sustainable sourcing can appeal to the rising ethical consumer base.** Agile marketing, data-driven customer insights, and a unique brand identity will be critical to standing out in a market increasingly dominated by large-scale players.

Implementation and Conclusion

This report has thoroughly examined the key challenges faced by Adornia and outlined strategic recommendations to address them. Implementing these will help stabilize sales, reduce return-related losses, and enhance competitiveness. Predictive modeling will support seasonal planning, while personalized promotions and customer segmentation will drive retention and targeted growth. These actions aim to improve profitability, operational efficiency, and customer loyalty. By aligning internal operations and marketing with customer behavior and regional demand, Adornia can achieve sustainable growth in an increasingly competitive market environment.

--------End of the report--------