Project Report

Business and Campaign Analysis for Simple Threads

Project Title: "Business and Campaign Analysis for Simple Threads"
Course: Basics of Data Analytics (BDA)
Name: Daksh Khandelwal
Date: 18th December 2025

1. Executive Summary

This report presents a comprehensive business analysis for Simple Threads, utilizing internal datasets covering Customer, Product, Purchase, and Discount information. The objective is to derive actionable insights into business performance for Q1 2025 and evaluate the effectiveness of the year-end promotional campaign. The analysis follows a structured approach, progressing through Customer Demographics, Sales Performance, Campaign Effectiveness, and Advanced Statistical Analysis to support strategic decision-making and future growth.

2. Data Overview

This section outlines the four primary datasets provided for the analysis of Simple Threads' business performance and campaign effectiveness.

2.1 Customer_Information.csv

https://docs.google.com/spreadsheets/d/1EsA5qM5KuYM38mlm0pFTdZj52K4DgygYixhboQjqqMA/edit?usp=sharing

Purpose: This dataset serves as the primary repository for customer demographics, enabling the segmentation of the customer base.

Data Types: Categorical (Gender, City), Numerical (Age), and Identifiers.

Support for Analysis: It allows for "Customer Analysis" by identifying unique customers and understanding their distribution across different regions (cities) and demographics (age/gender).

2.2 Product_Information.csv

https://docs.google.com/spreadsheets/d/1AOQJFq4Kaaa2NCll8QKBfJlHPeATvO4LautF8V-sbck/edit?usp=sharing

Purpose: This file acts as the product catalog, detailing the attributes of all items available for sale.

Data Types: Categorical (Product Name, Color) and Numerical (Price).

Support for Analysis: It is essential for "Sales Analysis" to determine product popularity and analyze sales trends across different categories and price points.

2.3 Purchase_Information.csv

https://docs.google.com/spreadsheets/d/1_rGGVODLwHasdjA-Rk6js7It3fyw8oHkDnoJjgj4Obg/edit?usp=sharing

Purpose: This dataset captures the transactional history of the business for the first quarter of 2025 (January - March).
Data Types: Date/Time, Numerical (Amount), and Identifiers (Customer/Product/Discount IDs).
Support for Analysis: It forms the backbone of the "Sales" and "Campaign" analyses by linking purchases to specific products and discount usage.

2.4 Discount_Information.csv

https://docs.google.com/spreadsheets/d/1-Y2ivdZ3HNAqxVwI27TqprQuadMZPVWXqBGVA_8MvWo/edit?usp=sharing

Purpose: This dataset records the details of the year-end promotional email campaign, linking specific customers to the offers they received.
Data Types: Date, Identifiers, and Discount Percentages.
Support for Analysis: It is critical for "Campaign Analysis" to verify which customers received coupons and measure campaign effectiveness.

3. Data Cleaning & Preparation

This section details the specific operations performed to transform the raw datasets into a consistent and usable format for analysis.

3.1 Handling Missing Value

Action: In the Purchase Information dataset, 'Discount Code' placeholders ('.') were replaced with "NO_DISCOUNT".

Reasoning: DIFFERENTIATING transactions without a promotional code from actual missing data facilitates accurate categorical analysis.

Action: Rows with missing or invalid identifiers in Customer Information were dropped to maintain a clean master list.

3.2 Duplicate Removal

Action: Identified and removed duplicate rows in Customer Information.

Reasoning: Preventing the inflation of customer metrics ensures each individual is counted only once.

3.3 Data Type Formatting

Action: Standardized date columns across Purchase Information and Discount Information to a uniform YYYY-MM-DD format.

Reasoning: Uniform dates are essential for accurate time-series analysis.

Action: Cleaned numerical columns by stripping "Rs." from Price and "%" from Discount Percentage.

Reasoning: Converting these to numeric types enables mathematical aggregation.

3.4 Data Consistency and Validation

Action: Corrected spelling inconsistencies in categorical data (e.g., "Mumbay" -> "Mumbai", "t-shrit" -> "t-shirt").

Action: Converted negative values in the Age column to absolute values.

3.5 Data Readiness for Analysis

Once cleaning was completed, all datasets were deemed suitable for analysis. Since the project focuses on descriptive and comparative business analysis, datasets were analyzed independently using logical relationships rather than full physical integration.

This cleaning and preparation process ensured that the subsequent analysis was based on high-quality, reliable data and that insights derived were both accurate and meaningful.

Dataset:
https://docs.google.com/spreadsheets/d/1dN9cO46lblJhTpVbG2PobMF__Szb6XAeFzxadkJh-hM/edit?usp=sharing

4. Phase 1: Customer Analysis

Objective:

To understand the size and demographics of the customer base and evaluate their spending behavior during Q1 2025. This analysis segments the "Registered Base" (from Customer Information) versus the "Active Purchasing Base" (from Purchase Information) due to dataset linkage constraints.

	Customer ID
2	AB1234
3	CD5678
4	EF9012
5	GH3456
6	IJ7890
7	KL2345
8	MN6789
9	OP1234
10	QR5678
11	ST9012
12	UV3456
13	WX7890
14	YZ2345
15	BA6789
16	DC1234
17	EF5678
18	GH9012
19	IJ3456
20	KL7890
21	MN2345
22	OP6789
23	QR1234

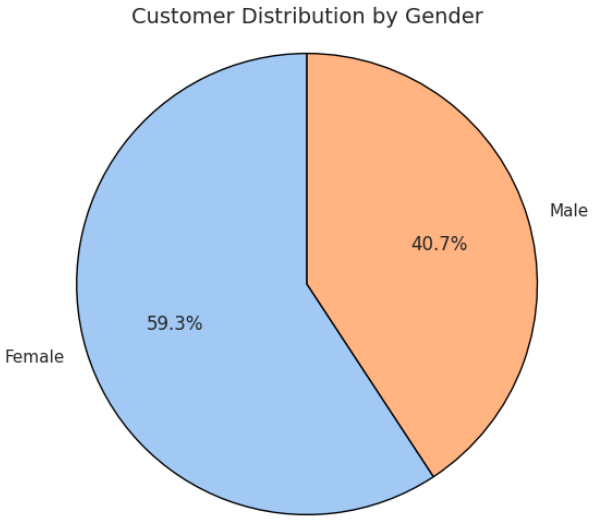
4.1 Unique Customers

Total Registered Customers: 54
Unique Customers who made purchases (Q1 2025): 22
Insight: Only ~40% of the registered customer base was active in the first quarter.

4.2 Customer Demographic Distribution

A. Gender Distribution

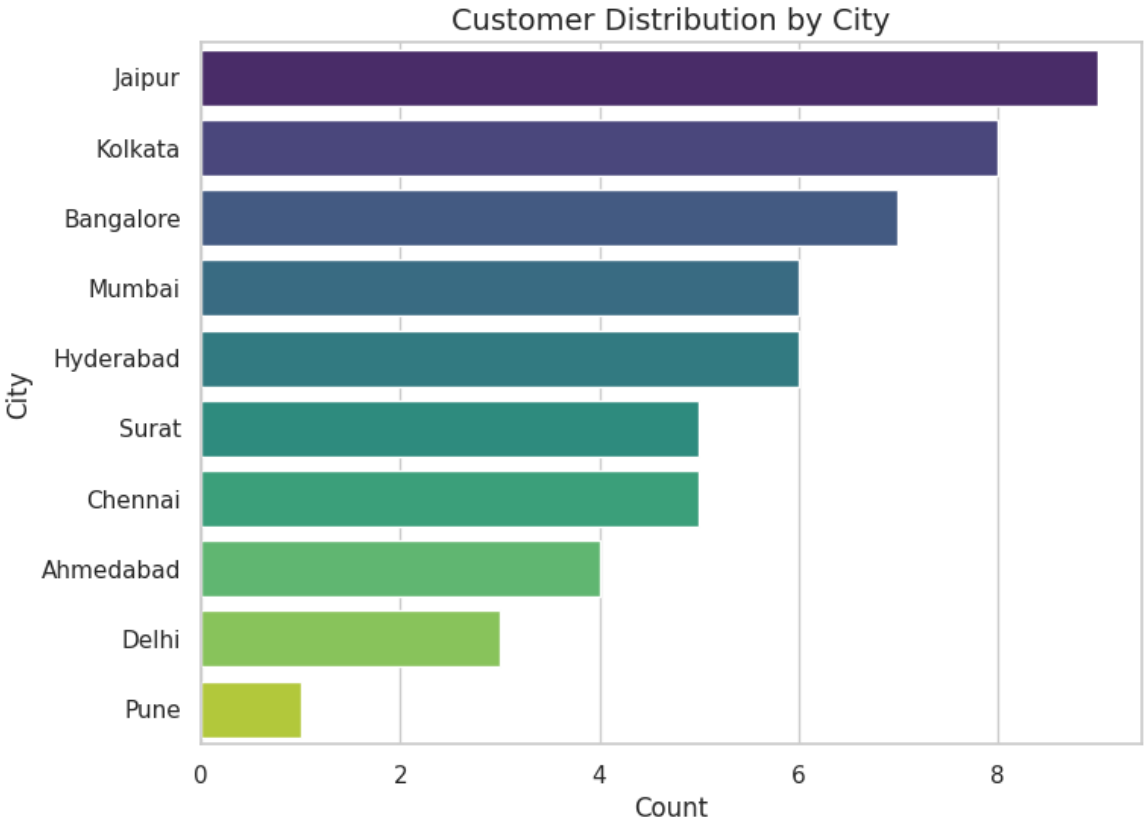
- **Female:** 32 (59%)
- **Male:** 22 (41%)
- **Interpretation:** The customer base is skewed towards females, suggesting a stronger engagement or product appeal within this segment.



B. Regional Distribution (Top 5 Cities)

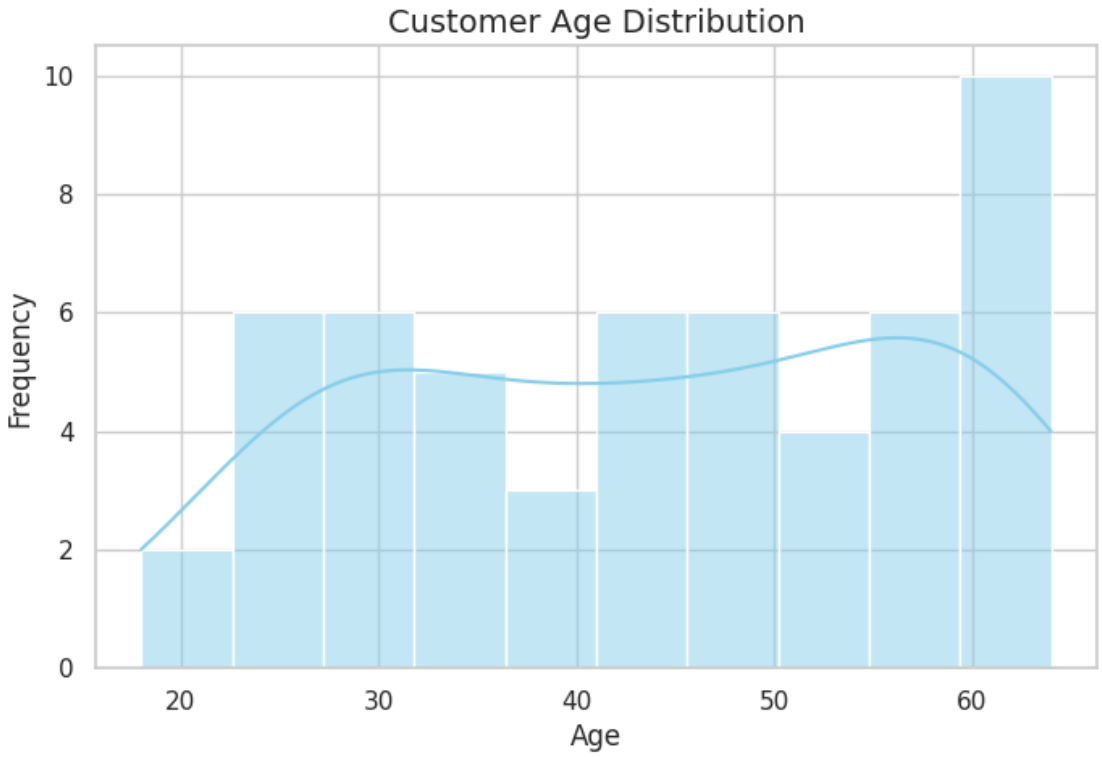
1. **Jaipur:** 9
2. **Kolkata:** 8
3. **Bangalore:** 7
4. **Mumbai:** 6
5. **Hyderabad:** 6

- **<u>Interpretation</u>:** Jaipur serves as the top location for the customer base, followed closely by major metros like Kolkata and Bangalore.



C. Age Distribution

- **Mean Age:** 43.7 years
- **Median Age:** 43.5 years
- **Range:** 18 - 64 years
- **<u>Interpretation</u>:** The customer base is mature, with an average age in the mid-40s. The distribution is fairly balanced without extreme skews toward very young or very old demographics.



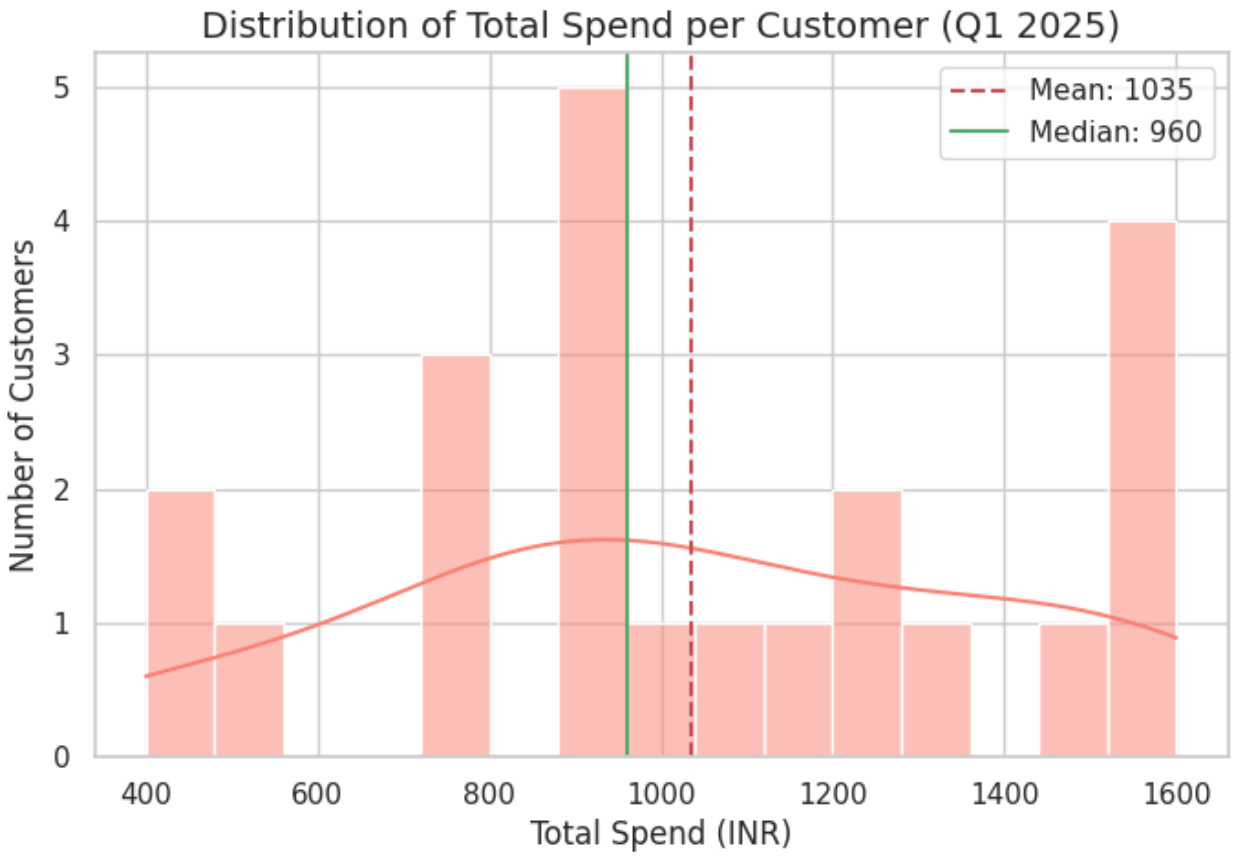
4.3 Average Customer Spend

Average Spend per Customer: ₹1,035

Median Spend per Customer: ₹960

Interpretation: The spending distribution is relatively consistent, with the mean slightly higher than the median, indicating a few higher-value shoppers pulling the average up.

Customer ID	SUM of Amount
AB1234	1,020.00
BA6789	900.00
CD5678	400.00
DC1234	765.00
EF5678	500.00
EF9012	900.00
GH3456	400.00
GH9012	1,145.00
IJ3456	900.00
IJ7890	1,280.00
KL2345	900.00
KL7890	1,530.00
MN2345	900.00
MN6789	1,080.00
OP1234	765.00
OP6789	1,600.00
QR1234	1,230.00
QR5678	1,275.00
ST9012	720.00
UV3456	1,500.00
WX7890	1,530.00
YZ2345	1,530.00
Average	1,035.00



Key Insights:

1. **<u>Conversion Opportunity</u>:** With only 22 out of 54 registered customers active in Q1, there is a significant opportunity to re-engage the remaining 60% of the user base.
2. **<u>Target Demographic</u>:** The "Core Customer" profile is a **Female, roughly 44 years old, living in Jaipur or Kolkata**. Marketing campaigns tailored to this persona may yield higher ROI.
3. **<u>Revenue Baseline</u>:** An average spend of ~₹1,000 suggests that most customers are buying 1-2 items per quarter (given product prices range from ₹300-900). Strategies to increase basket size could boost this metric.

5. Phase 2: Sales Analysis

Objective:

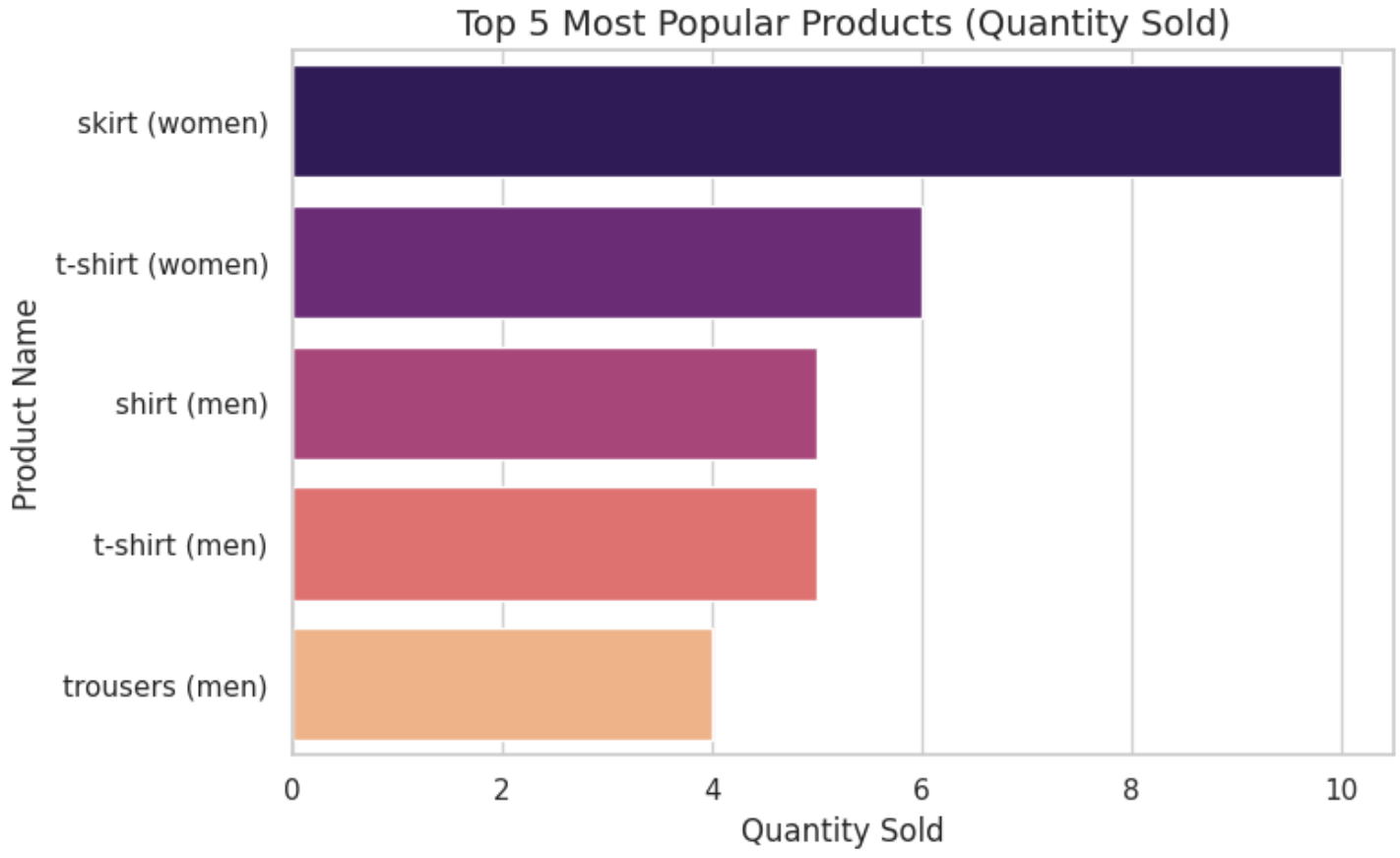
To evaluate the overall sales performance, identify high-performing products, and analyze revenue trends across different product categories during the first quarter of 2025.

5.1 Most Popular Products

Top Selling Product: Skirt (Women) is the clear leader with **10 units** sold.

Runners Up: T-shirt (Women) (6 units), Shirt (Men) (5 units), and T-shirt (Men) (5 units).

Insight: Women's apparel, specifically skirts, is driving the volume of sales.

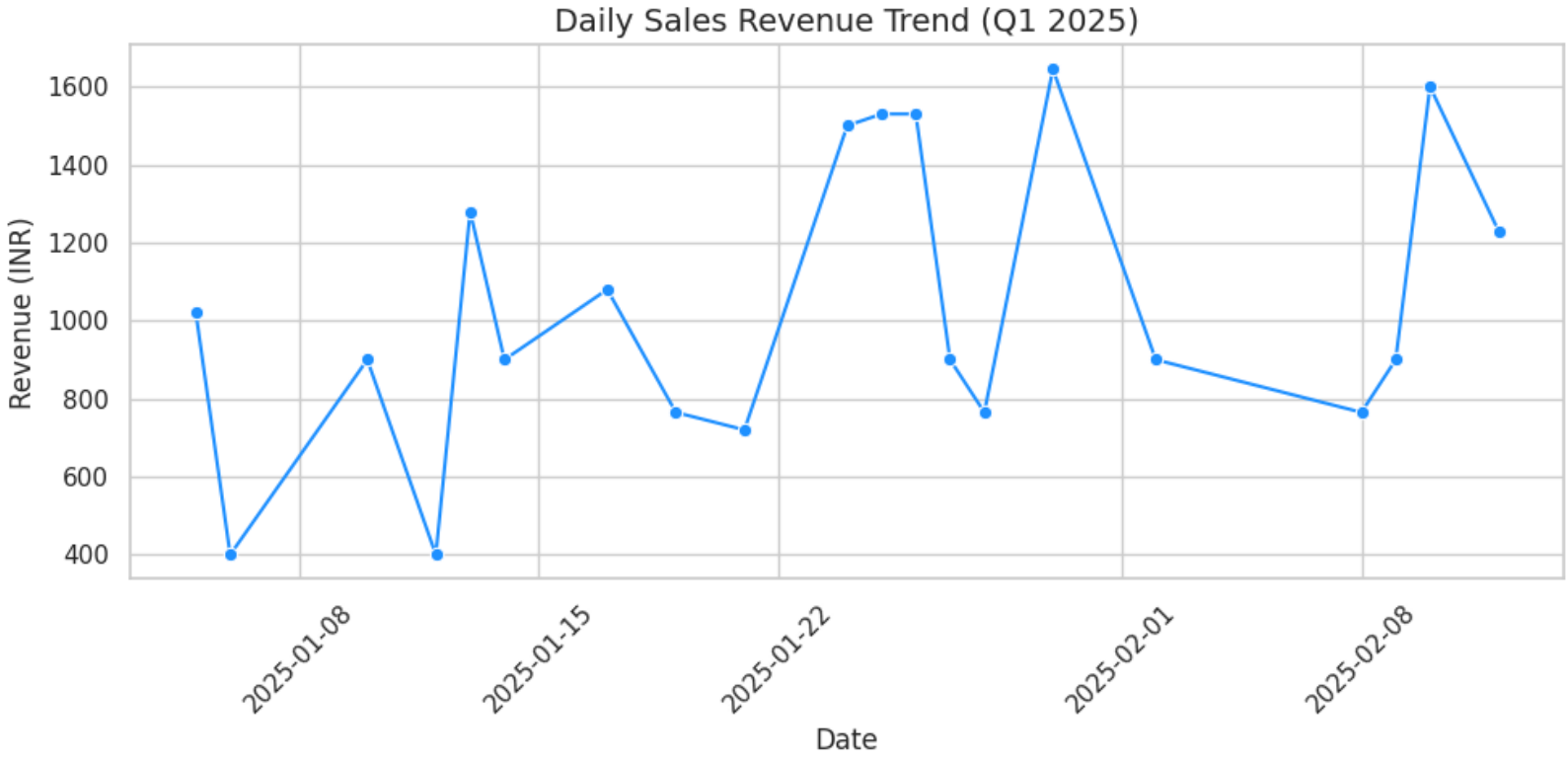


5.2 Total Sales Revenue

<u>Total Revenue (Q1 2025)</u>: INR 22,770

It was calculated by summing all transaction amounts recorded during the period. This metric provides an overall measure of business performance and reflects the company's revenue generation during the quarter.

Product ID	SUM of Amount
1001	740.00
1002	760.00
1003	400.00
1004	950.00
1005	875.00
1006	950.00
1007	960.00
1008	360.00
1009	1,050.00
1010	1,440.00
1011	1,480.00
1012	640.00
1013	1,440.00
1014	255.00
1015	255.00
1016	255.00
1018	720.00
1019	1,665.00
1020	1,530.00
1021	500.00
1022	500.00
1023	500.00
1024	1,530.00
1025	765.00
1026	1,530.00
1027	720.00
Grand Total	**22,770.00**

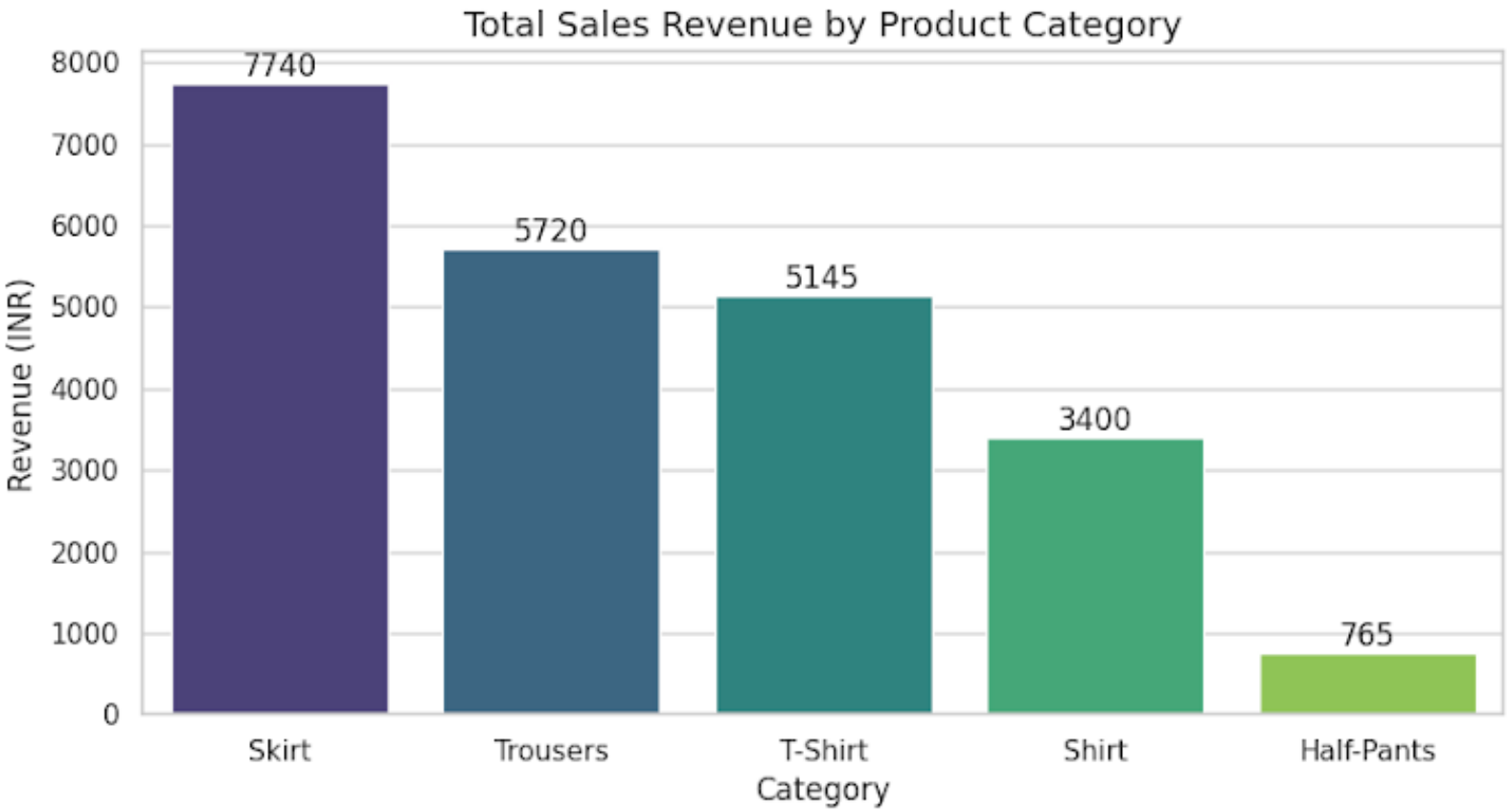


5.3 Sales Trends Across Product Categories

Top Revenue Category: Skirts (INR 7,740) contribute the most to the bottom line, followed by **Trousers** (INR 5,720) and **T-Shirts** (INR 5,145).

Low Performers: Half-Pants generated the least revenue (INR 765), indicating either low demand or a lower price point impacting total value.

Category Trend: While "Skirts" dominate in both volume and value, "Trousers" perform well in revenue despite lower volume (4 units vs 10 skirts), suggesting a higher price point per unit efficiency.



Key Insights:

1. **Product Mix:** The business relies heavily on **Women's Skirts** for both volume and revenue. Diversifying the best-sellers list (perhaps by pushing Men's Trousers which have high value) could reduce risk.
2. **Revenue Drivers:** High-ticket items like Trousers are efficient revenue drivers. Selling one pair of trousers is roughly equivalent to selling 2-3 half-pants in terms of revenue.
3. **Inventory Focus:** Ensure stock levels for "Skirt (Women)" are always sufficient, as stockouts here would significantly impact Q1 performance.

6. Phase 3: Campaign Analysis

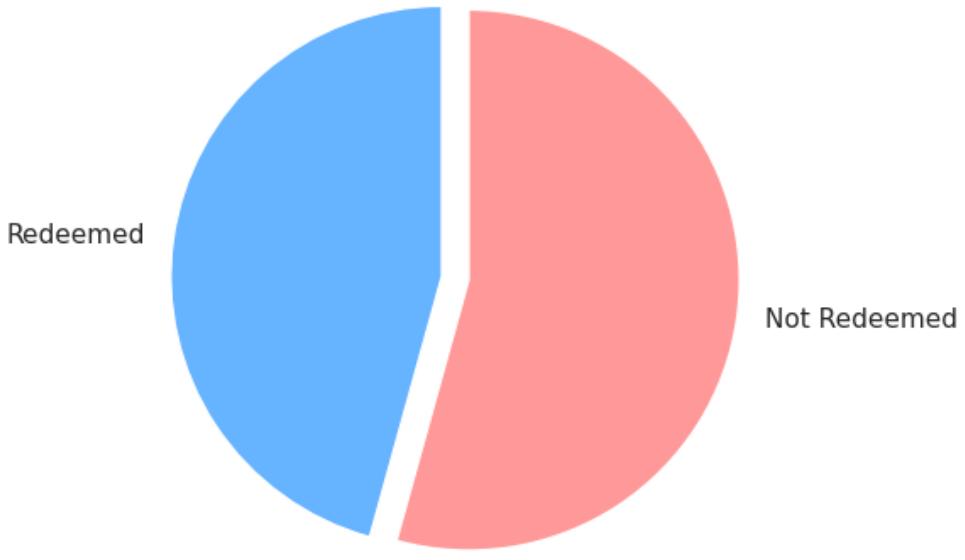
Objective:

To evaluate the success of the year-end discount email campaign by measuring customer participation, the impact of incentives on spending, and the comparative effectiveness of different discount tiers.

6.1 Discount Campaign Participation

Coupons Sent: 35 unique discount codes were distributed via email.
Coupons Redeemed: 15 unique codes were used in Q1 transactions.
Redemption Rate: 42.9%

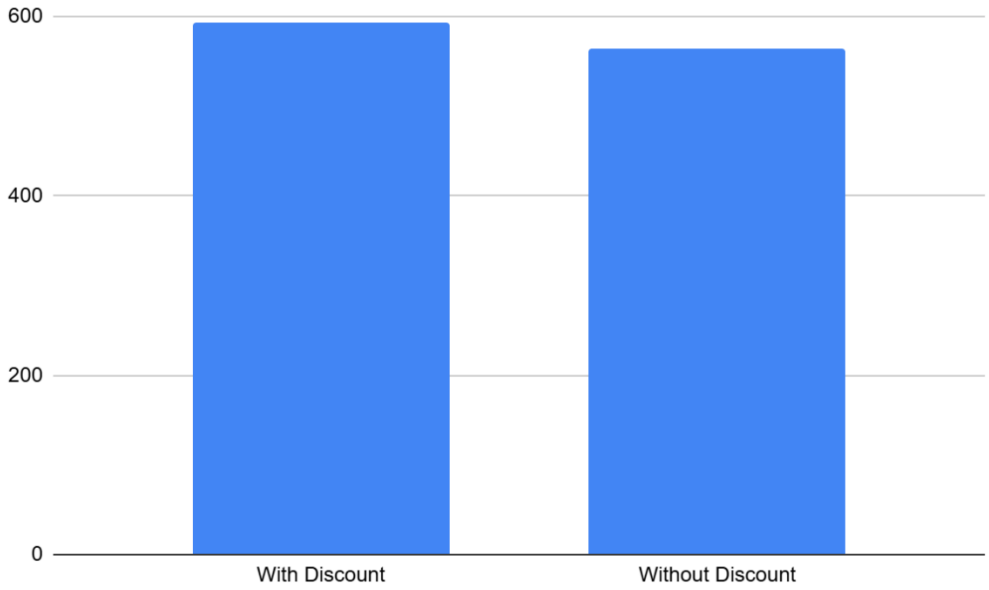
Insight: A redemption rate of nearly 43% is exceptionally high for email campaigns (industry benchmarks often hover around 2-5%), indicating highly engaged customers or very attractive offers

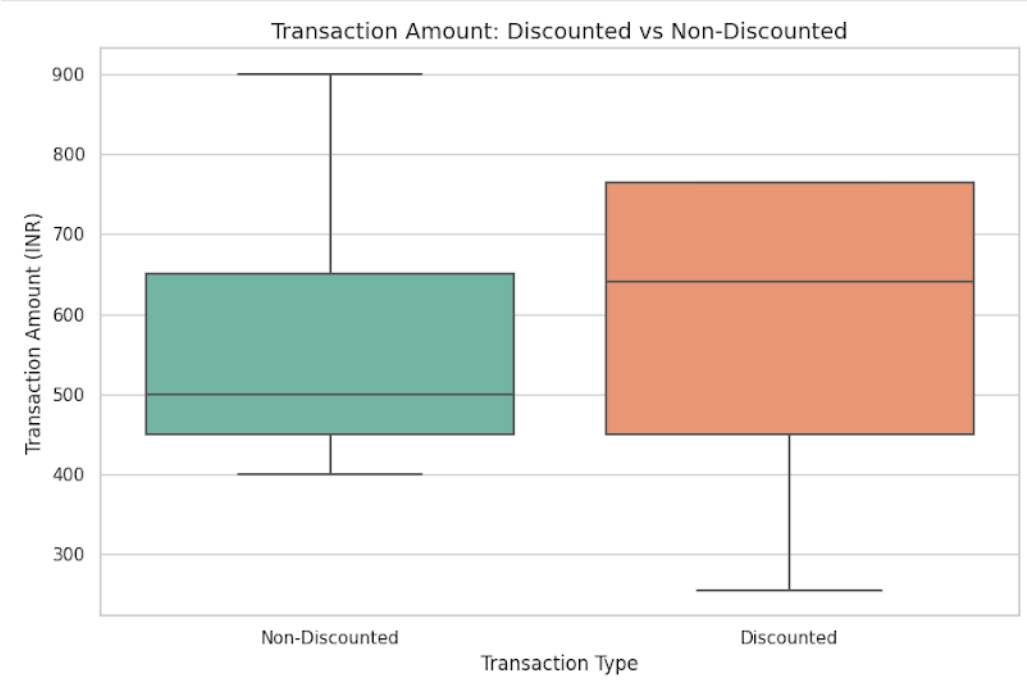


6.2 Impact of Discounts on Customer Spending

Average Spend (Discounted Transaction): ₹571.38
Average Spend (Non-Discounted Transaction): ₹563.64

Insight: There is a marginal positive difference (+1.4%) in spending when a discount is applied. While discounts successfully drove volume (as seen in participation), they did not significantly increase the average value of a single basket compared to full-price shoppers.

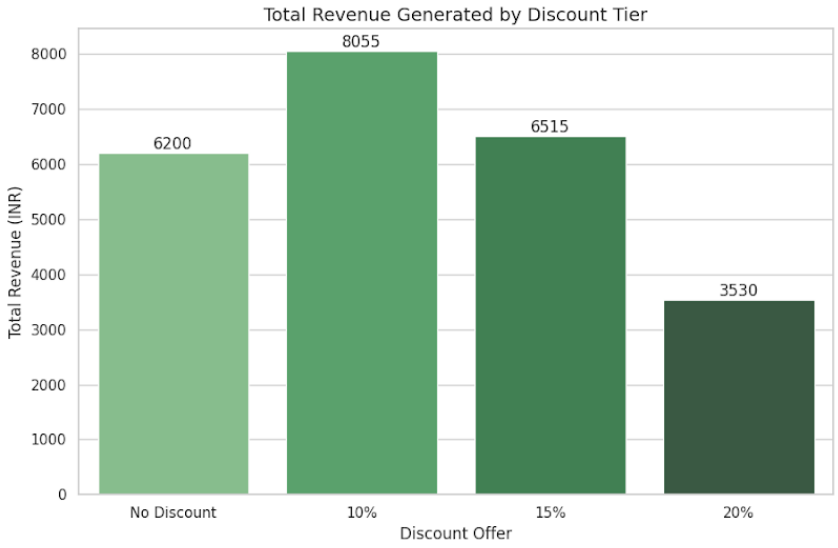


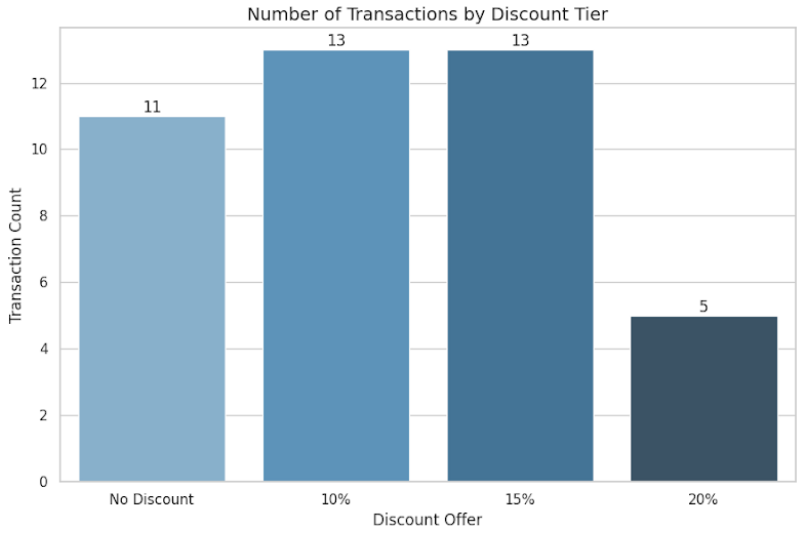


6.3 Effectiveness of Discount Offers

We analyzed performance across the three discount tiers: **10%, 15%, and 20%**.

- **Effective Discount Offers**: ASDFGH, TGBYHN & ZXCVBN
- **Most Revenue Generated: 10% Discount** (₹8,055 Revenue, 13 Transactions).
- **Highest Volume (Tie):** Both **10%** and **15%** tiers drove 13 transactions each.
- **Highest Average Order Value: 20% Discount** (₹706/transaction).
    - ***Observation****:* The steepest discount (20%) attracted the highest spend per order but had the lowest volume (only 5 transactions). This suggests it was used for larger, perhaps planned, purchases.
    - ***Observation****:* The **10% discount** was the "Sweet Spot"—it generated the most total revenue and high volume without giving away as much margin as the 20% offer.





Key Insights:

1. **<u>High Engagement</u>:** The campaign was highly successful in driving traffic, with nearly half of the recipients converting.
2. **<u>Optimal Discount Level</u>:** The **10% discount** proved to be the most effective for total revenue. It drove as much traffic as the 15% offer but retained more margin and generated more total sales.
3. **<u>Strategic Use of High Discounts</u>:** The 20% discount resulted in higher basket sizes (₹706 vs ₹620 for 10%). This tier could be reserved for high-value bundles or specific loyalty rewards rather than broad distribution, as it drives value but lower volume.

7. Phase 4: Advanced Analysis

Objective:

To statistically explore the relationships between key variables and identify any anomalies (outliers) that could distort the business understanding.

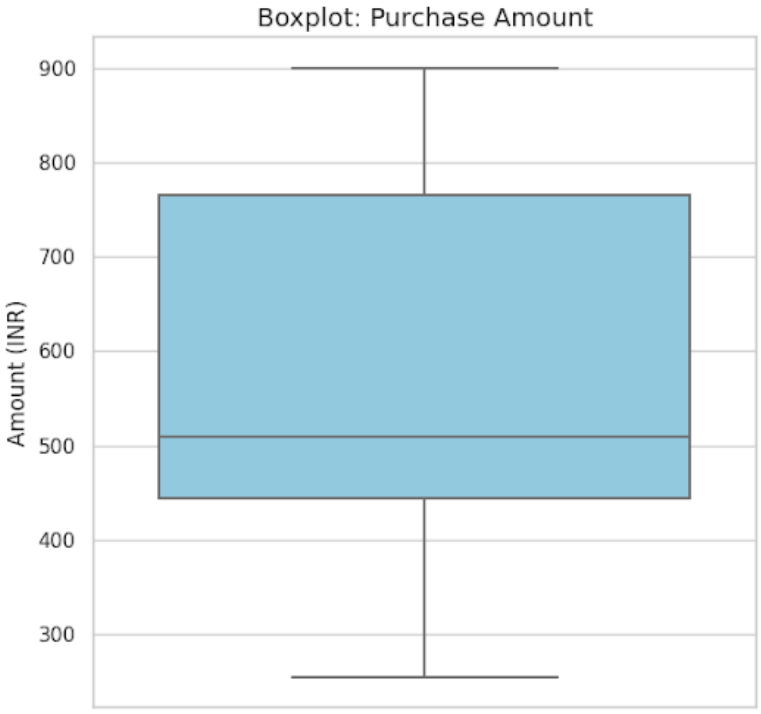
7.1 Box Plot Analysis

Box plots were created to visualize the distribution of key numerical variables, including transaction amounts, product prices, and discount percentages.

a) Purchase Amount

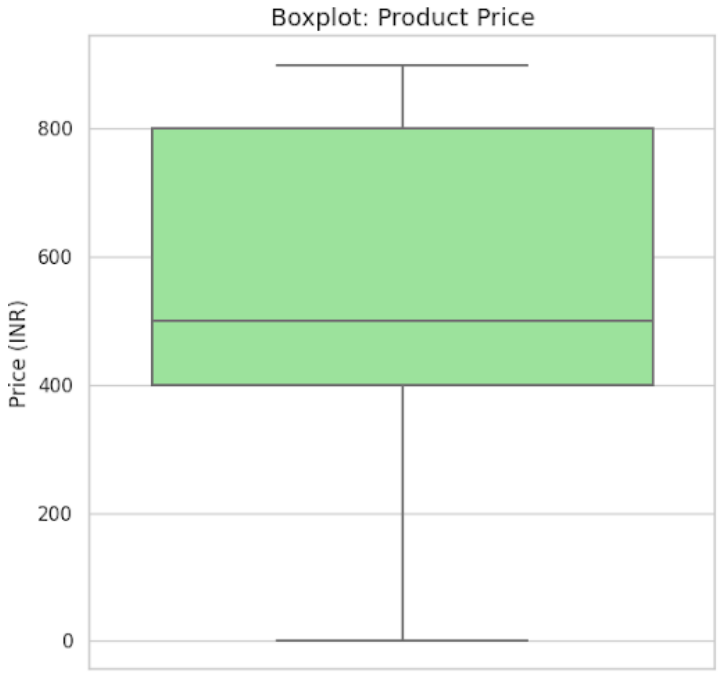
Interpretation: The boxplot shows a consistent transaction value range. The median purchase amount sits around ₹500-600. The distribution is fairly symmetrical with no extreme tails.

Spread: Most transactions fall between ₹400 and ₹800.



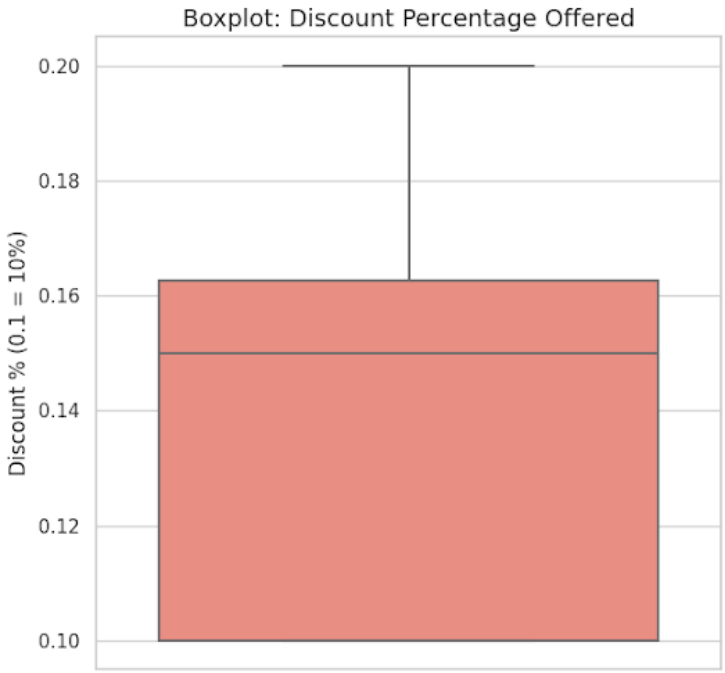
b) Product Price

Interpretation: Prices are clustered in defined tiers (₹300, ₹400, ₹500, ₹800, ₹900). The lack of continuous spread confirms that products are priced using a fixed-tier strategy rather than variable costing.



c) Discount Percentage

Interpretation: The discounts offered are strictly categorical: 10%, 15%, and 20%. The median offer is 15%.



7.2 Outlier Detection

Using the Interquartile Range (IQR) method (1.5 * IQR), we analyzed the numerical data for anomalies.

- **Purchase Amount Outliers: None Found.** All transaction values fall within the statistical expected range (-₹38 to ₹1,246). This indicates stable pricing and purchasing behavior without bulk-buying anomalies.
- **Product Price Outliers: None Found.** All product prices are within standard bounds.

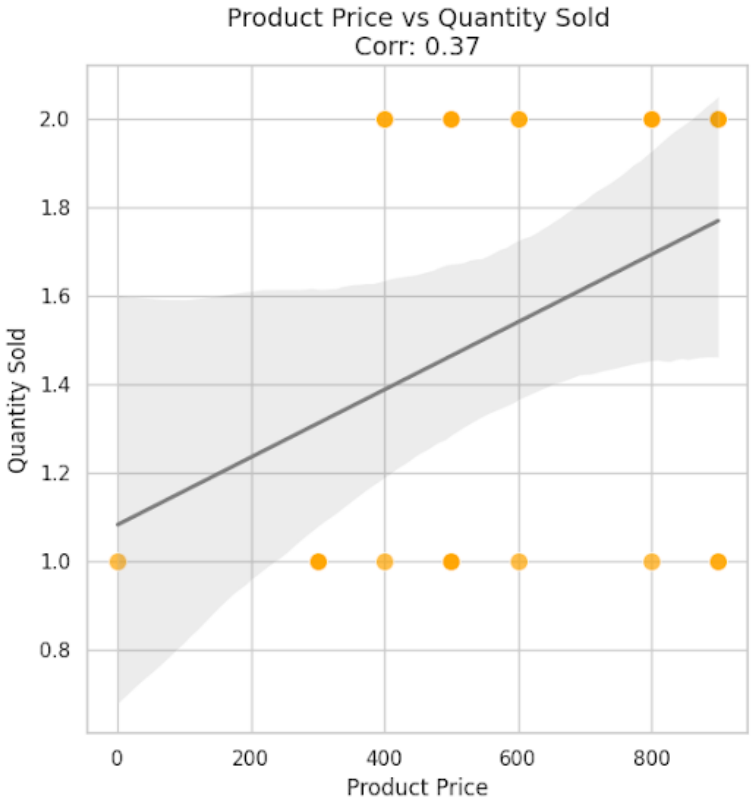
7.3 Correlation Analysis

a) Quantity Sold vs. Product Price

Correlation Coefficient: 0.37 (Moderate Positive)

Interpretation: This is an interesting, counter-intuitive finding. Typically, higher prices lead to lower demand (negative correlation). Here, we see a positive correlation, driven largely by high-volume sales of Skirts (₹900) and Trousers (₹800).

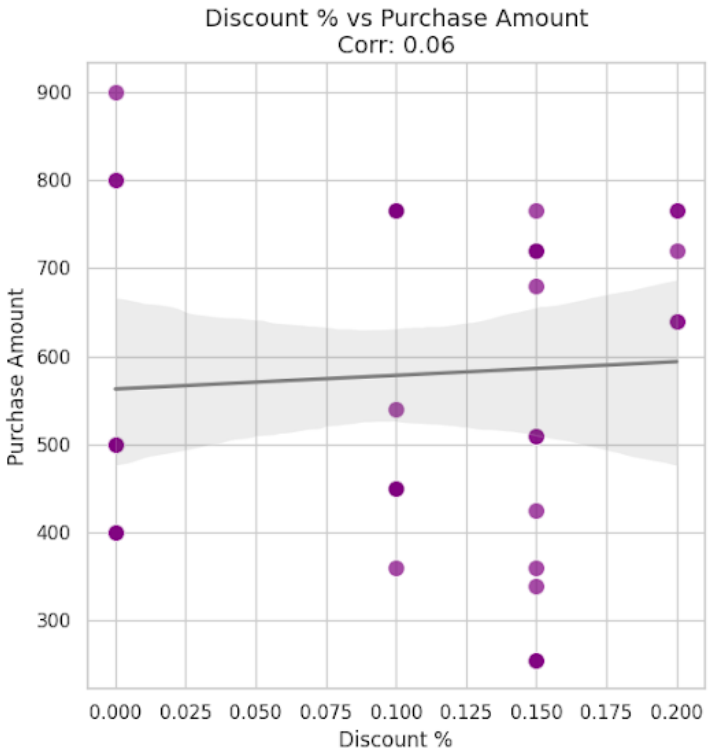
Insight: Simple Threads' customers are value-insensitive for specific high-demand items (like Skirts), preferring them over cheaper alternatives like Half-Pants (₹300).



b) Discount % vs. Purchase Amount

<u>Correlation Coefficient</u>: 0.06 (Weak Positive)

<u>Interpretation</u>: There is essentially no linear relationship between the size of the discount and the total amount spent. Offering a 20% discount does *not* lead to significantly larger basket sizes compared to a 10% discount.



Key Insights:

1. **Pricing Power:** The positive correlation between Price and Quantity suggests the brand has pricing power. The "Bestsellers" are the expensive items.
2. **Ineffective Discount Scaling:** Since higher discounts don't drive higher spend (Correlation ~0), the strategy should pivot to **frequency-based** incentives (get them to buy *more often*) rather than value-based incentives (giving deeper cuts).
3. **Stable Operations:** The absence of outliers suggests operational stability—no data entry errors in pricing or erratic bulk orders that would skew inventory planning.

8. Key Findings & Business Insights

This section synthesizes the analysis of Simple Threads' Q1 2025 performance, highlighting strategic opportunities in customer engagement, product mix, and promotional efficiency.

- **Customer Engagement Gap:** Only 22 of the 54 registered customers (41%) made purchases in Q1 2025, highlighting a significant opportunity to reactivate the dormant majority. The core active demographic is female, approximately 44 years old, primarily located in Jaipur and Kolkata.
- **Premium Product Dominance:** Contrary to typical trends, higher-priced items sold in higher volumes. The "Bottoms" category (Skirts and Trousers, ₹800–900) drove the most revenue, significantly outperforming lower-cost "Tops" and Half-Pants.
- **Optimal Discount Strategy:** The campaign achieved an exceptional 42.9% redemption rate. The 10% discount tier was the most effective, generating the highest total revenue and matching the traffic volume of the 15% offer, whereas the 20% offer failed to drive significant volume.
- **Operational Consistency:** Statistical analysis confirmed a stable operational environment with no pricing or transaction outliers, indicating consistent customer purchasing behavior around an average spend of ₹1,035.

9. Strategic Recommendations

Based on the quantitative findings from the Q1 2025 analysis, the following actionable steps are recommended to drive growth and efficiency for Simple Threads:

1. Optimize Inventory for High-Value "Bottoms"

- **Rationale:** Data reveals a positive correlation between price and sales volume, with Skirts (₹900) and Trousers (₹800) being top performers. Conversely, low-priced items like Half-Pants (₹300) yielded the lowest revenue.
- **Action:** Shift inventory procurement budgets to prioritize Skirts and Trousers. Consider reducing stock depth for Half-Pants or bundling them (e.g., "3 for ₹800") to clear slower-moving, lower-margin inventory.

2. Standardize Promotional Discounts at 10%

- **Rationale:** The 10% discount tier generated the highest total revenue (₹8,055) and equal transaction volume to the 15% tier. The 20% tier failed to drive significant volume (only 5 transactions).
- **Action:** Cap standard email promotional offers at 10%. This preserves an additional 5-10% margin without negatively impacting conversion rates. Reserve higher discounts (20%+) exclusively for high-threshold loyalty rewards (e.g., "Spend ₹2000 to unlock 20% off") to force larger basket sizes.

3. Launch a Targeted Re-Engagement Campaign

- **Rationale:** 59% of the registered customer base (32 out of 54 users) did not purchase in Q1, representing a significant lost opportunity.
- **Action:** Segment these 32 inactive users and deploy a specific "We Miss You" email campaign. Given the demographic data, tailor the imagery and messaging to the core persona (Females, 40s) and utilize the proven 10% incentive to reactivate them.

4. Focus Regional Marketing on Proven Hubs

- **Rationale:** Jaipur, Kolkata, and Bangalore account for the highest concentration of customers.
- **Action:** Allocate the majority of the digital marketing budget to these specific geolocations. Operational logistics (shipping partners) should also be optimized for these routes to ensure faster delivery times, enhancing customer satisfaction in key markets.

10. Conclusion

This analysis of Simple Threads' business data for the first quarter of 2025 has provided critical visibility into the company's current performance and customer dynamics. By integrating and examining customer demographics, transaction history, and promotional data, the report has successfully identified key revenue drivers—specifically the dominance of high-value "bottoms" categories—and validated the efficacy of the 10% discount strategy.

The findings indicate that Simple Threads possesses a stable operational foundation with a clearly defined, value-resilient core customer base. However, significant opportunities remain in reactivating dormant users and optimizing inventory mix to align with actual demand patterns. Implementing the recommended strategies will enable the management team to transition from data-backed observations to proactive growth initiatives, ensuring a stronger competitive position in the upcoming quarters.

Power Point Presentation

https://docs.google.com/presentation/d/186ck78s0cx0eAsmxCR59U2K2u7Fj44nZjygoRmlFxHw/edit?usp=sharing