# Analytical Study of an Online Commerce Brand

**A Mid-Term report for the BDM capstone Project**


Submitted by

Name: Arnav Mehta

Roll number: 21f3002647

IITM Online BS Degree Program,

Indian Institute of Technology, Madras, Chennai

Tamil Nadu, India, 600036

# Contents

# 1 Executive Summary

Molcha by Babita Singh is a women's ethnic apparel brand with a significant online presence, primarily operating through platforms like Myntra. The brand specializes in curated ethnic wear, including saree blouses, palazzos, and contemporary Indian apparel. While the business has achieved considerable market reach, it faces serious operational inefficiencies, chiefly high product return rates, inventory mismatches, and courier-related losses that have adversely impacted profitability.

To assess these challenges and quantify their financial impact, transactional and operational data was collected from the Myntra Seller Portal over a nine-month period (June 2024 to February 2025). Key variables included order status, article type, transaction value, payment mode, return reason, courier partner, and customer location.

The analysis revealed a total revenue of ₹92.7 lakh, with a net realized profit of ₹28.3 lakh, resulting in an actual profit margin of 30.5%. In the absence of returns and related deductions, the estimated profit margin could have been as high as 75.2%. This represents an **absolute margin reduction of 44.7 percentage points**, or a **relative decline of nearly 59.5%** due to return-related losses. These losses, comprising return handling costs, lost revenue, and tax deductions amounted to over ₹41 lakh. The return rate was alarmingly high at 52.01%.

wSpecific product categories such as cushion covers and churidars displayed high profitability and low return rates, whereas kurtis, co-ords, and saree blouses emerged as high-risk segments with frequent returns and lower margins. Based on these findings, the report recommends rebalancing inventory toward more profitable, low-return categories, optimizing courier allocation in high-RTO regions, and introducing demand forecasting mechanisms to better align inventory with actual market demand.

# 2 Proof of originality of the Data

This report is based on **primary data** collected directly from a real-world small business — *Molcha by Babita Singh*, which operates on online platforms like Myntra. The data was obtained with full consent and authorization from the business owner and is unique, not sourced from any open or secondary datasets.

## Business Details

- Business Name: Molcha by Babita Singh
- Address: 52a/9 Kishangarh ,Vasant Kunj ,New delhi-110070
- Owner's Name: Ms. Babita Singh

## Authorization & Validation

Video of Interaction with Business Owner: https://drive.google.com/file/d/1AGcPsy2tgl40cSEbo9rp2s_p3iNB9gVT/view?usp=sharing

Letter: https://drive.google.com/file/d/1op-dAJZr2Lv0g1HUeHv6ebuyC8t7x09B/view?usp=sharing

## Source of Data

The data was collected using **automated digital systems provided by the Myntra Seller Portal**, thereby eliminating manual entry errors. The data came from the following modules:

### 1. Monthly Sales Reports

- **Source**: Myntra Seller Portal
- **Format**: Structured Excel sheets (.xlsx)
- **Contents**: Order values, return rates, commissions, platform fees, category-wise profits

### 2. Real-Time Performance Dashboards

- **Source**: Myntra Analytics Interface
- **Metrics Tracked**:
  - Daily return trends
  - Inventory turnover
  - Sales distribution by geography

# 3 Metadata

This project is based on primary data collected from the Myntra Seller Portal, covering continuous nine-month period from June 1, 2024 to February 28, 2025. The dataset was downloaded in structured spreadsheet formats and consists of three key modules: Order Flow Reports, Return Reports, and Return to Origin (RTO) Reports.

- Data Format: Excel/Sheets (XLSX)
- Range: June 01, 2024 to February 28, 2025
- **Currency & Units**: All financial data is reported in Indian Rupees (₹); categorical variables are labeled as text, and quantitative features are in numerical form (integers, percentages, floats).
- **Data Source**:
    - Myntra Seller Portal
    - Sections used:
        - **Order_Flow_Report** (Order-level data)
        - **Return_Report** (Returns/refunds)
        - **Packed_RTO_Report** (Courier and logistic returns)
- **Drive Link to Original Data**: https://drive.google.com/drive/folders/11ixA6wQKDWSQ7XMH6mSj85wEu4Hih-FS?usp=sharing

## 3.3 Features Collected and Justification :

The dataset includes a range of categorical, numeric, and derived fields that were carefully selected to analyze the brand's profitability, return behavior, and operational efficiency. Each feature contributes directly to understanding the core business issues of return losses, courier inefficiencies, and inventory misalignment.

| Feature | Type | Justification |
|---|---|---|
| **Article Type** | Categorical | Categorizes products (e.g., Saree Blouse, Cushion Covers) to analyze sales, margins, and return rates by type |
| **Order Count** | Integer | Measures demand concentration; used for segment-wise planning (e.g., 8,205 women's vs. 74 unisex orders) |
| **Net Profit & Margin** | Float / % | Indicates post-deduction profitability; e.g., net profit was ₹28.3 lakh with margin dropping from 75.2% to 30.59% |
| **Return Rate** | Percentage | Captures inefficiency in categories (e.g., Kurtis had an 80% return rate) |
| **Profit Calculation** | Derived Variable | Computed as customer payment minus platform deductions, logistics cost, tax loss; highlights value leakage |
| **Pricing Strategy** | Categorical Buckets | Analyzes profitability across price ranges (e.g., ₹501–₹1000 had highest order count but high return rate) |
| **Losses** | Float (₹) | Quantifies impact from returns (₹41.3 lakh total loss from RT & RTO orders) |

# 4 Descriptive Statistics:

## Overall Order Flow Statistics:

| ARTICLE TYPE | COUNT | MARGIN% | RETURN % |
|---|---|---|---|
| **Saree Blouse** | 5481 | 28.46 | 54.90 |
| **Palazzos** | 777 | 37.09 | 44.27 |
| **Salwar** | 676 | 40.81 | 44.53 |
| **Dresses** | 298 | 30.65 | 48.32 |
| **Co-ords** | 210 | 17.50 | 64.76 |
| **Churidar** | 165 | 47.40 | 30.91 |
| **Shirts** | 153 | 40.93 | 45.10 |
| **Harem Pants** | 106 | 29.02 | 45.28 |
| **Shorts** | 94 | 12.23 | 68.09 |
| **Skirts** | 84 | 25.73 | 52.38 |
| **Cushion Covers** | 74 | 69.25 | 12.16 |
| **Tops** | 67 | 25.27 | 65.67 |
| **Ethnic Dresses** | 43 | 28.28 | 53.49 |
| **Trousers** | 28 | 38.67 | 39.29 |
| **Kurtas** | 12 | 44.87 | 33.33 |
| **Earrings** | 5 | 53.30 | 20.00 |
| **Kurtis** | 5 | 2.84 | 80.00 |

**Table 4.1: Product Categories with Order Count, Margin, and Return Rate**

Insight: Saree Blouses dominate order volume but suffer from high return rates (54.9%). Cushion Covers and Churidars are high-margin, low-return categories. Kurtis, despite having the highest return rate (80%), are the least profitable (2.84% margin).

| PRICE RANGE | RETURN RATE | AVG PRICE | AVG PROFIT | PROFIT MARGIN | TAX PER ORDER | ORDER COUNT |
|---|---|---|---|---|---|---|
| **0-500** | 39.34 | **439.90** | **166.69** | **37.89** | **0.74** | **122** |
| **500-1000** | **52.38** | **902.64** | **278.40** | **30.84** | **1.79** | **5678** |
| **1000-2000** | **51.83** | **1646.48** | **494.07** | **30.01** | **3.19** | **2458** |
| **2000-5000** | **47.62** | **2362.33** | **1034.51** | **43.79** | **3.57** | **21** |

**Table 4.2: Price Range vs Return Rate, Profitability and Tax Impact**

**Insight**: The ₹501–₹1000 price range had the highest number of orders but also a high return rate (52.38%). Higher price segments (₹2000–₹5000) offered better margins and lower return rates, though at lower order volumes. Profitability improves with price, but customer retention risk increases.

## Return (RT) orders Statistics:

| Metric | Value |
|---|---|
| **Refunded Orders** | 3,593 |
| **Unique SKUs** | 1,524 |
| **Total Items** | 3,591 |
| **Packet Count** | 3,369 |
| **Average Refunded Shipment Value** | Rs. 1,137.36 |
| **Average Order Value(mean)** | Rs. 1,119.19 |
| **Standard Deviation (Order Value)** | Rs. 373 |
| **Maximum Shipment Value** | Rs. 2,530 |
| **Minimum Shipment Value** | Rs. 216 |

**Table 4.3: Summary of Returned Orders and Refund Metrics**

| Return Status | Count | Revenue Lost |
|---|---|---|
| **DSL** | 2,577 | 29,02,024.68 |
| **RL** | 947 | 10,41,422.29 |
| **LPI** | 63 | 71,342 |
| **RRC** | 6 | 6,451 |
| **TOTAL** | 3,593 | 40,21,240.97 |

**Table 4.4: Revenue Loss by Return Type**

Insight: DSL alone accounts for ~72% of total return-related revenue loss. It is the dominant return category and a major operational concern. RRC and LPI are relatively minor but indicate occasional courier-side inefficiencies.

| Mode | Orders | Total Loss |
|---|---|---|
| **Prepaid (ON)** | 2,611 | 29,41,151.97 |
| **COD** | 982 | 10,80,088.00 |

**Table 4.5: Return Losses by Payment Type**

**Insight**: Prepaid orders result in a higher volume of returns and contribute nearly **73% of the total revenue loss** due to returns. COD returns are fewer but still significant, especially considering logistics complexity and cash handling.

| Refund Mode | Qty |
|---|---|
| **OR** | 2,526 |
| **NEFT** | 707 |

| | |
|---|---|
| **Myntra Credit** | 360 |

**Table 4.6: Distribution of Refund Modes Among Returned Orders**

**Insight**: Most customers opt for direct-to-source refund (Original). NEFT and Myntra Credit are minor but still important to track for user experience optimization.

## Return to Origin (RTO) and Failed Order (F) Statistics:

| Metrics | Value |
|---|---|
| **Total Orders Packed** | 844 |
| **Packed IDs** | 784 |
| **Unique SKUs** | 559 |
| **Unique Customer Locations** | 32 |
| **Order Status – Failure (F) Orders** | 554 |
| **Order Status – RTO Orders** | 290 |
| **Total Orders Returned (F+RTO)** | 844 |

**Table 4.7: Summary of Packed Orders Resulting in RTO or Failure**

| Article Type | Distinct Order Count |
|---|---|
| **Saree Blouse** | 625 |
| **Salwar** | 66 |
| **Palazzos** | 51 |
| **Dresses** | 18 |
| **Co-ords** | 14 |
| **Shirts** | 13 |
| **Churidar** | 13 |
| **Shorts** | 13 |
| **Tops** | 11 |
| **Skirts** | 7 |
| **Ethnic Dresses** | 4 |
| **Harem Pants** | 3 |
| **Cushion Covers** | 2 |
| **Kurtas** | 2 |
| **Kurtis** | 1 |
| **Trousers** | 1 |

**Table 4.8: Article-Wise Impact of RTO/Failed Orders (Top categories)**

**Insight**: While **Saree Blouses** accounted for nearly **74% of all RTO/failed orders**, indicating a critical issue with quality or sizing, categories like **Cushion Covers**, **Kurtas**, and **Trousers** showed **minimal RTO impact**, making them more reliable and operationally efficient.

| Payment Method | Order Status | Distinct Order Count |
|---|---|---|
| **COD** | F | 308 |
| **COD** | RTO | 256 |
| **ON** | F | 246 |
| **ON** | RTO | 34 |

**Table 4.9: Payment Mode vs Order Status for Returns**

| Courier | RTO Volume |
|---|---|
| **Flipkart Logistics** | HIGH |
| **EK_E2E** | HOGH |
| **Ecom Express** | LOW |
| **Delhivery** | LOW |

**Table 4.10: Courier Partner Efficiency – RTO Volume Trends**

**Insight**: Flipkart Logistics and EK_E2E consistently show higher RTOs, possibly due to regional gaps or logistical errors. These partners require further audit and coordination

# 5 Detailed Explanation of Analysis Process & Methods

The analytical approach in this project was structured to address key business inefficiencies such as high product return rates, courier failures, and unoptimized inventory flow. A combination of spreadsheet pre-cleaning, Python-based data transformation, and visual analytics using Pandas, Matplotlib, and Seaborn was employed. Each technique was selected based on its suitability for the specific problem it addressed.

## 5.1 Return Rate & Refund Pattern Analysis

- To evaluate product-level return behavior, descriptive statistics were applied to identify articles with disproportionately high return rates. The `groupby()` function in Pandas was used to aggregate metrics by article type, return type, and payment method. Visualization tools (bar plots, heatmaps) were used to highlight SKUs with >50% returns.

- **Justification**: Return behavior is categorical and ratio-based, making **grouped aggregation** and **percentage comparisons** the most appropriate over clustering or regression at this stage.

- *Tool Used: Python (Pandas, Matplotlib), Notebook: RTO_Colab*

## 5.2 Courier & RTO Performance Evaluation

- Courier partner efficiency was measured by analyzing RTO orders by region and courier name. Seaborn bar charts and grouped pivot tables helped identify courier services (e.g., Flipkart Logistics, EK_E2E) contributing to higher failure rates.
- **Justification**: Since courier data is segmented and hierarchical (by region and status), **pivot-based trend comparison** was more effective than general correlation techniques.
- *Tool Used: Python (Seaborn, Pandas), Notebook: RTO_Courier_Returns*

## 5.3 Product Category Profitability Analysis

- Product-wise revenue, net profit, and return-adjusted margin were analyzed using calculated fields like `net_profit = revenue - (returns + shipping + platform fee + tax)`. Python's descriptive functions (`mean`, `sum`, `std`) helped derive average order values, margins, and losses per category.
- **Justification**: Margin analysis is **ratio-based**, so using **custom-engineered columns** allowed more flexibility than pre-built statistical models or regressions.
- *Tool Used: Python (Pandas), Notebook: Category_Profitability.ipynb*

## 5.4 Pricing Strategy Assessment

- The dataset was bucketed into price ranges (e.g., ₹0–₹500, ₹501–₹1000) to compare return rates and profit margins by pricing tier. Visual comparisons helped identify unprofitable price ranges despite high volume (e.g., ₹501–₹1000).
- **Justification**: Binning via `pd.cut()` enabled comparative analysis. A predictive model was not used because the problem required **diagnosis, not forecasting**.
- *Tool Used: Python (Pandas, Matplotlib), File: Pricing_Analysis_Colab*

## 5.5 Packed vs. Fulfilled Order Flow Evaluation

- Using order logs and packing data, the flow of items from dispatch to delivery was mapped using Python and spreadsheet-based aggregation (Google Sheets). Delays and failures were highlighted using status counts and conditional filters.
- **Justification**: Initial cleaning in Google Sheets allowed **fast validation** of SKU-level anomalies before deeper analysis in Python. This hybrid method was better suited than only coding or only Excel.
- *Tools Used: Google Sheets, Python (Pandas), Notebook: Packed_OrderFlow.ipynb*

These targeted, problem-specific methods allowed for actionable insights into return behavior, courier performance, and operational bottlenecks — all while balancing speed, clarity, and accuracy

## 6 Results and Findings

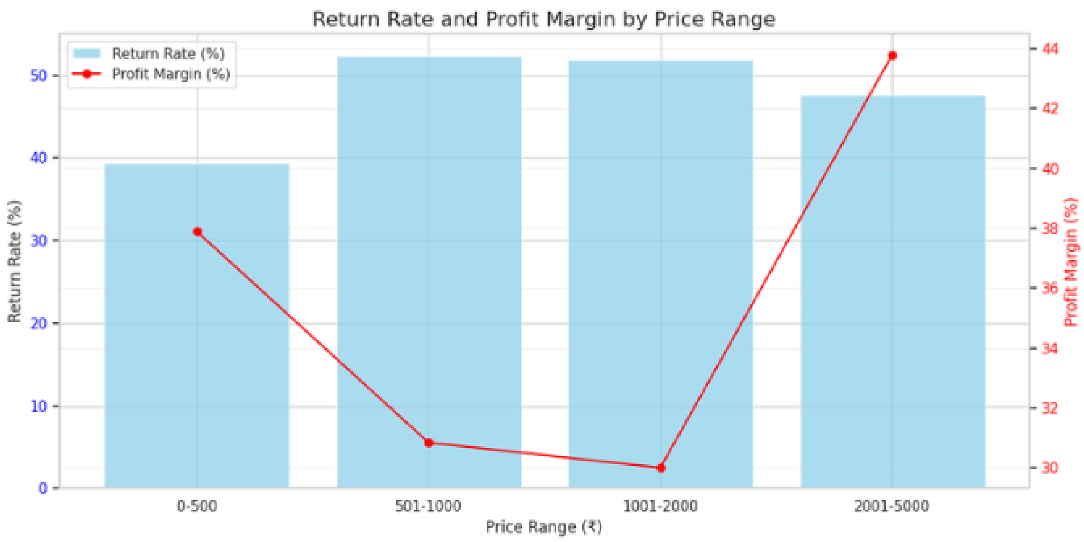
This section presents visual evidence and key trends derived from the data analysis. Each insight is directly tied to a business problem and visualized using Python (Matplotlib/Seaborn). All charts have labeled axes, units, and are arranged by theme for clarity.

### 6.1 Return Behavior by Product Category

**Insight**: Kurtis and Co-ords show return rates above 65%, with Kurtis reaching 80%. Cushion Covers had the lowest return rate (12.16%) and highest margin (69.25%), suggesting strong quality and fit alignment.

### 6.2 Return Rate and Profit Margin by Price Range

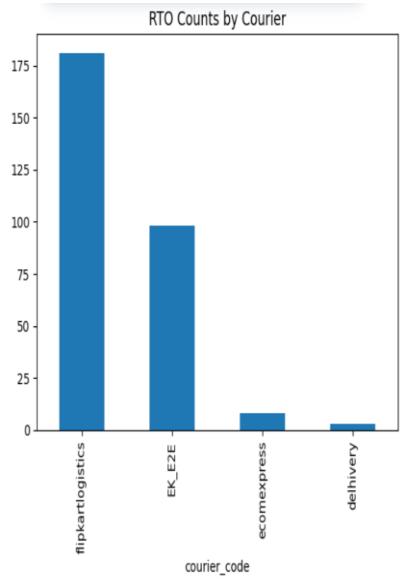
**Figure 6.X: Return Rate and Profit Margin by Price Range**




**Insight**: **Insight**: The ₹501–₹1000 and ₹1001–₹2000 price segments have the **highest return rates (over 50%)** and the **lowest profit margins (around 30%)**, indicating poor performance despite high order volumes. In contrast, the ₹2001–₹5000 range offers the **highest profitability (43.79%)** with comparatively lower returns, making it a strategically favorable segment for premium inventory focus

### 6.3 RTO and Failed Orders by Courier Partner

**Figure 6.3: RTO Volume by Courier**




**Insight**: Flipkart Logistics and EK_E2E contributed the highest RTOs. Delhivery and Ecom Express had better delivery outcomes. This highlights the need to reallocate delivery zones.

## 6.4 Return Losses by Refund Status

**Figure 6.4: Revenue Loss by Return Type (DSL, RL, LPI, RRC)**


**Insight**: Delayed shipments (DSL) alone caused over ₹29 lakh in losses — 72% of all return-related losses.

## 6.5 High-Value Article Types Among RTOs

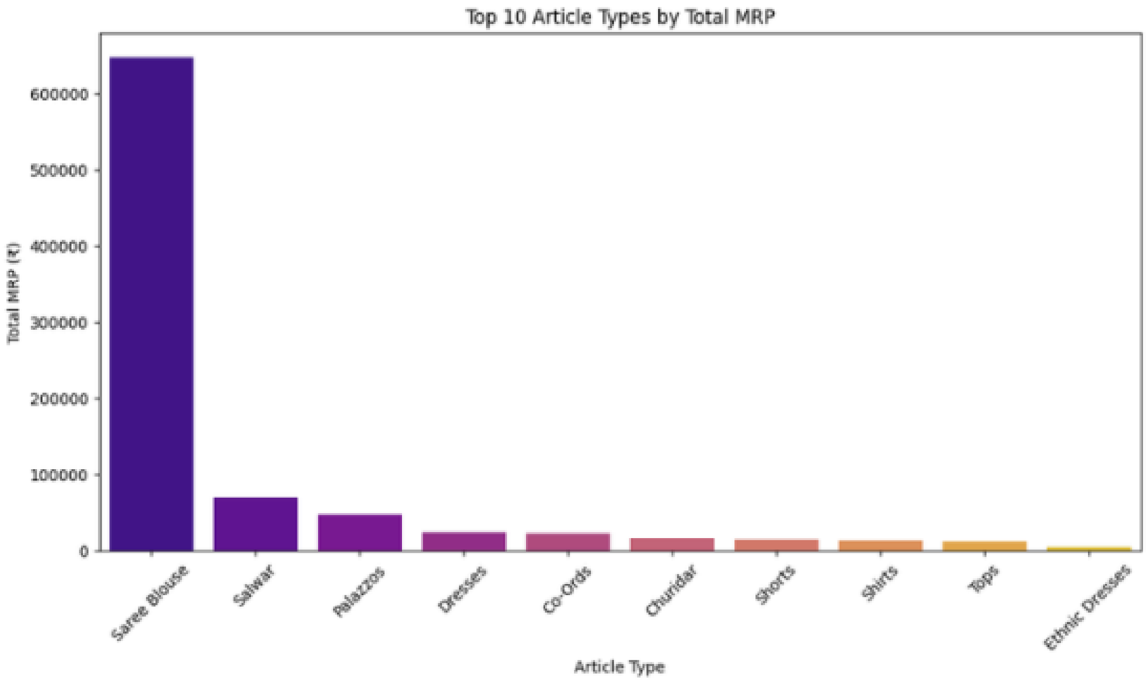



**Figure 6.5**: Bar chart showing the top 10 article types with the highest **total MRP (₹)** among RTO (Return to Origin) orders. The X-axis lists the article types, while the Y-axis shows cumulative total MRP lost.

**Insight**: **Saree Blouse** stands out overwhelmingly with over ₹6.4 lakh in total MRP loss, far exceeding the next-highest category (Salwar). This suggests that Saree Blouses contribute disproportionately to high-value RTOs and should be the **primary focus for quality checks, size standardization, or catalog revision**. Other categories like Palazzos, Dresses, and Co-ords also show consistent value losses and may require **refined inventory or courier strategies**.

## 6.6 Distribution of Revenue Lost per Returned Item

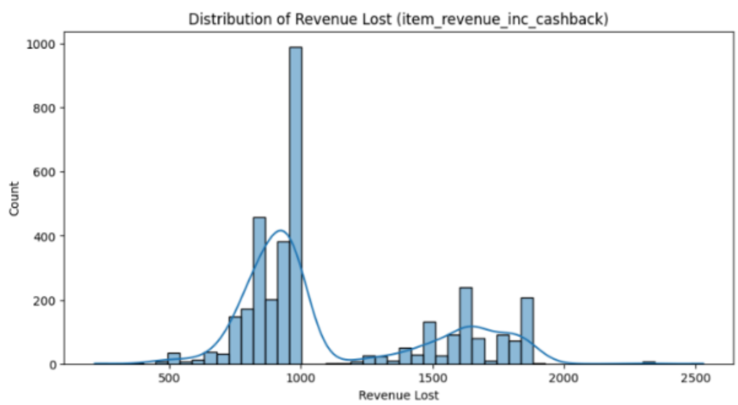
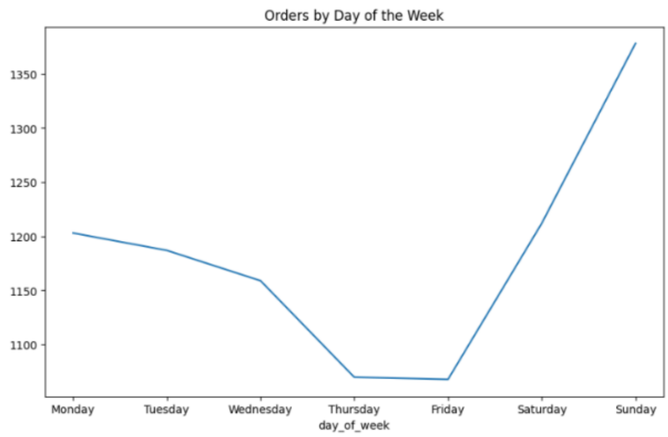



**Figure 6.X**: Distribution of revenue lost per item due to returns (including cashback impact). X-axis shows the amount of loss per item (in ₹), while Y-axis shows the frequency of such losse

**Insight**: The distribution is **bimodal**, indicating two distinct clusters of returned item losses — one near ₹900–₹1,000 (most common), and another around ₹1,500–₹2,000. A peak occurs just below ₹1,000, suggesting most returned orders result in nearly full-value loss. The long right tail indicates a few high-ticket returns with up to ₹2,500 lost. This highlights the need to flag **high-value SKUs** for enhanced return protection measures or pre-shipment quality checks

## 6.7 Weekday-Wise Profit Margin

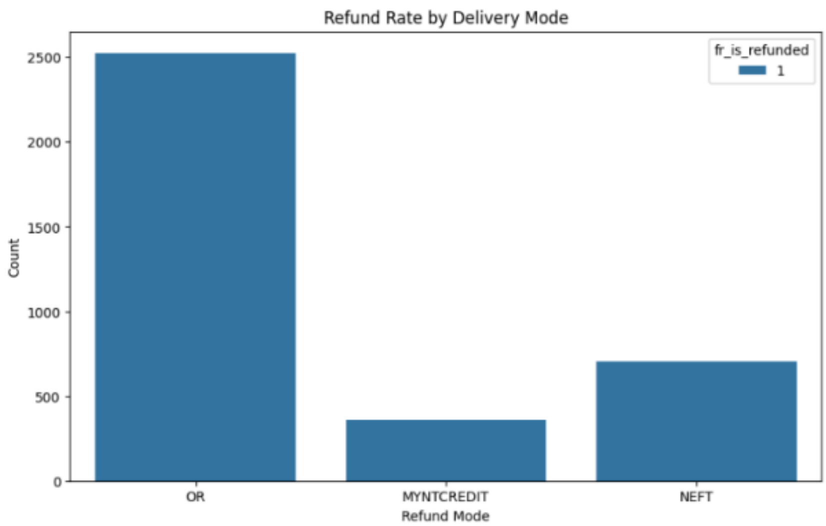
**Figure 6.7: Profit Margin by Weekday**




**Insight**: Friday has the best margin (~34.48%), while Monday is the worst (~27.53%), possibly due to post-weekend buyer returns

## 6.8 Refund Preference

**Figure 6.8: Refund Mode Distribution**




**Insight**: Majority (2526) prefer refunds to original source, meaning banking systems must be prioritized for refund speed.

## 6.9 Summary and Conclusion

This analysis uncovered substantial inefficiencies in return management, courier operations, and product-level profitability that are directly affecting business performance. Categories like **Saree Blouses** and **Kurtis** contributed heavily to return-driven revenue loss, while **Cushion Covers** and

**Earrings** emerged as highly profitable with minimal returns, highlighting the need for strategic inventory shifts. The **₹501–₹1000 price segment**, despite being the most ordered, showed elevated return rates, signaling pricing mismatch or customer dissatisfaction. Courier performance varied significantly, with **Flipkart Logistics** and **EK_E2E** showing high RTO volumes, whereas **Delhivery** and **Ecom Express** demonstrated operational reliability. Additionally, patterns in **refund preferences**, **weekday profit margins**, and the **bimodal distribution of return losses** offered deeper insight into buyer behavior and shipment vulnerabilities.

These findings underline the importance of data-driven decision-making across operations — from **courier reassignment** and **product design audits** to **dynamic pricing** and **quality checks**. Addressing these high-impact areas can significantly reduce financial leakage and lay the groundwork for more scalable and profitable operations in the future.

Additional Information: https://drive.google.com/drive/folders/1Ovn31Ox-A-Q70J0iKKTZGxaBzBupR5E-?usp=sharing