

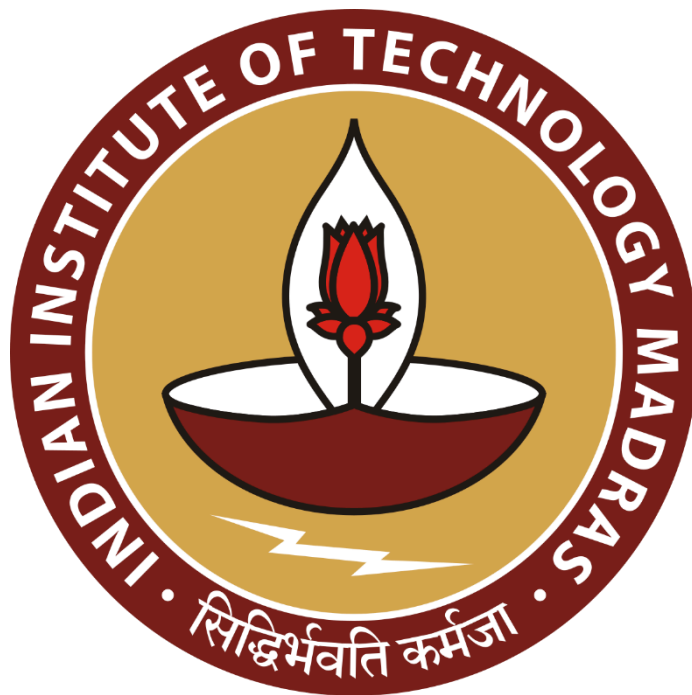
# **Amazon Sales & Fulfillment: Data-Driven Insights for Optimization & Growth**

**A Final Report for the BDM capstone Project**

Submitted by

Name: Tarun Kandarpa

Roll number: 23f2002936



IITM Online BS Degree Program,  
Indian Institute of Technology, Madras, Chennai  
Tamil Nadu, India, 600036

## Contents

1	Executive Summary	3
2	Proof of originality	3
3	Meta data and descriptive statistics	4
4	Detailed explanation of analysis process/ method	6
5	Results and findings	9
6	Interpretation of results and recommendation	18

## **Declaration Statement**

I am working on a Project titled “**Amazon Sales & Fulfillment: Data-Driven Insights for Optimization & Growth**”. I extend my appreciation to **Amazon**, for providing the necessary resources that enabled me to conduct my project.

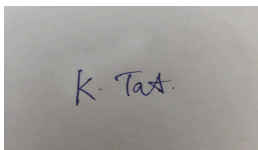
I hereby assert that the data presented and assessed in this project report is genuine and precise to the utmost extent of my knowledge and capabilities. The data has been gathered from primary sources and carefully analyzed to assure its reliability.

Additionally, I affirm that all procedures employed for the purpose of data collection and analysis have been duly explained in this report. The outcomes and inferences derived from the data are an accurate depiction of the findings acquired through thorough analytical procedures.

I am dedicated to adhering to the principles of academic honesty and integrity, and I am receptive to any additional examination or validation of the data contained in this project report.

I understand that the execution of this project is intended for individual completion and is not to be undertaken collectively. I thus affirm that I am not engaged in any form of collaboration with other individuals, and that all the work undertaken has been solely conducted by me. In the event that plagiarism is detected in the report at any stage of the project's completion, I am fully aware and prepared to accept disciplinary measures imposed by the relevant authority.

I understand that all recommendations made in this project report are within the context of the academic project taken up towards course fulfillment in the BS Degree Program offered by IIT Madras. The institution does not endorse any of the claims or comments.

A rectangular box containing a handwritten signature in blue ink that reads "K. Tarun".

Signature of Candidate

Name: Tarun Kandarpa

Date: 14/04/2025

# 1 Executive Summary

The project titled “**Amazon Sales & Fulfillment: Data-Driven Insights for Optimization & Growth**” addresses critical operational challenges within Amazon’s e-commerce ecosystem, specifically targeting high order cancellations, fulfillment inefficiencies, and suboptimal promotional strategies. A comprehensive dataset comprising transactional records, cancellation logs, and fulfillment metrics was analyzed to uncover patterns that reveal a concentrated group of high-impact products driving the bulk of revenue while also contributing disproportionately to order cancellations. Notably, orders processed via Amazon’s own fulfillment network consistently show lower cancellation rates compared to those handled by third-party merchants, underscoring the importance of a robust logistics framework. Detailed data preprocessing, exploratory data analysis, and visualization techniques were employed to reveal trends in daily and monthly sales, revenue fluctuations, and product performance across various categories.

The analysis further demonstrated that targeted promotions significantly boost revenue and dramatically reduce cancellation rates, indicating that strategic marketing initiatives are vital for sustaining customer commitment. Additionally, inventory misalignment during peak demand periods was identified as a key factor exacerbating operational inefficiencies. These insights form the basis for actionable recommendations, including optimizing fulfillment channels, enhancing inventory management through demand-driven reordering, and refining promotional strategies to focus on high-impact SKUs. By implementing these data-driven strategies, the project aims to reduce revenue losses, improve operational efficiency, and strengthen Amazon’s competitive position in the rapidly evolving e-commerce marketplace.

## 2 Proof of originality

The dataset “**Amazon Sale Report.csv**” used in this project was obtained from Kaggle and is based on Amazon India e-commerce sales data, including order IDs, product names, quantities sold, prices, order dates, and customer locations. Although SKU-level granularity is not available, the dataset is comprehensive enough to analyze order fulfillment, cancellation trends, and promotional impacts.

Originally compiled by Tehseen Sajid, this dataset is widely used in e-commerce data analysis. All data preprocessing, analysis, and visualization were conducted using Python (Matplotlib and Seaborn) and Excel. The version-controlled code and scripts are maintained in a dedicated repository, accessible at:

Source: [Kaggle Dataset Source](#)

### 3 Meta Data and Descriptive Statistics

A	B	C	D	E	F	G	H	I	J	K	L	M
index	Order ID	Date	Status	Fulfillment	Sales Channel	ship-service-level	Style	SKU	Category	Size	ASIN	Courier Status
1	0 405-8078784-57	04-30-22	Cancelled	Merchant	<a href="#">Amazon.in</a>	Standard	SET389	SET389-KR-NP-Set	S		B09KXVBD7Z	
2	1 171-9198151-11	04-30-22	Shipped - Delive	Merchant	<a href="#">Amazon.in</a>	Standard	JNE3781	JNE3781-KR-XX kurta	3XL		B09K3WFS32	Shipped
3	2 404-0687676-72	04-30-22	Shipped	Amazon	<a href="#">Amazon.in</a>	Expedited	JNE3371	JNE3371-KR-XL kurta	XL		B07WV4JV4D	Shipped
4	3 403-9615377-81	04-30-22	Cancelled	Merchant	<a href="#">Amazon.in</a>	Standard	J0341	J0341-DR-L Western Dress	L		B099NRCT7B	
5	4 407-1069790-72	04-30-22	Shipped	Amazon	<a href="#">Amazon.in</a>	Expedited	JNE3671	JNE3671-TU-XX Top	3XL		B098714BZP	Shipped
6	5 404-1490984-45	04-30-22	Shipped	Amazon	<a href="#">Amazon.in</a>	Expedited	SET264	SET264-KR-NP-Set	XL		B08YN7XD5G	Shipped
7	6 408-5748499-68	04-30-22	Shipped	Amazon	<a href="#">Amazon.in</a>	Expedited	J0095	J0095-SET-L Set	L		B08CMHNWBN	Shipped
8	7 406-7807733-37	04-30-22	Shipped - Delive	Merchant	<a href="#">Amazon.in</a>	Standard	JNE3405	JNE3405-KR-S kurta	S		B081WX4G4Q	Shipped
9	8 407-5443024-52	04-30-22	Cancelled	Amazon	<a href="#">Amazon.in</a>	Expedited	SET200	SET200-KR-NP-Set	3XL		B08L91ZZXN	Cancelled
10	9 402-4393761-03	04-30-22	Shipped	Amazon	<a href="#">Amazon.in</a>	Expedited	JNE3461	JNE3461-KR-XX kurta	XXL		B08B3XF5MH	Shipped
11	10 407-5633625-69	04-30-22	Shipped	Amazon	<a href="#">Amazon.in</a>	Expedited	JNE3160	JNE3160-KR-G- kurta	S		B07K3YQLF1	Shipped
12	11 171-4638481-63	04-30-22	Shipped	Amazon	<a href="#">Amazon.in</a>	Expedited	JNE3500	JNE3500-KR-XS kurta	XS		B098117DJ3	Shipped
13	12 405-5513694-81	04-30-22	Shipped - Delive	Merchant	<a href="#">Amazon.in</a>	Standard	JNE3405	JNE3405-KR-XS kurta	XS		B081XCMYXJ	Shipped

This dataset contains **128,975 rows** and **22 columns**, capturing various order-related details such as identifiers, fulfillment methods, pricing, shipping information, and promotional data. A few columns, such as **Courier Status**, **promotion-ids**, and **fulfilled-by**, contain missing values, which were considered during preprocessing to ensure data quality. Each column plays a specific role in analyzing Amazon's e-commerce performance:

- **Order ID** (object): Unique identifier for each order.
- **Date** (datetime64[ns]): Timestamp of the order, crucial for time-series insights.
- **Status** (object): Indicates whether the order is Shipped, Cancelled, etc.
- **Fulfillment** (object): Specifies Amazon or merchant-managed shipping.
- **Sales Channel** (object): Platform through which the sale was made.
- **ship-service-level** (object): Describes the shipping speed or service (e.g., Expedited).
- **Style** (object): Internal product design or model code.
- **SKU** (object): Stock Keeping Unit for product-level tracking.
- **Category** (object): Broader grouping (e.g., Set, Kurta).
- **Size** (object): Product dimension or apparel sizing.
- **ASIN** (object): Amazon's product identification number.
- **Courier Status** (object): Delivery stage (e.g., Shipped, In-Transit).
- **Qty** (int64): Number of units in each order.
- **currency** (object): Transaction currency (e.g., INR).
- **Amount** (float64): Total price of the order.

- **ship-city, ship-state, ship-postal-code, ship-country**: Destination details for shipping.
- **promotion-ids** (object): Promotional offers applied.
- **B2B** (bool): Indicates if the order is a business-to-business transaction.
- **fulfilled-by** (object): Specifies whether the order is Self Ship or fulfilled by Amazon.

The numerical columns provide key insights into purchasing patterns and shipping distribution:

	Amount	Qty
Count	121180.000000	128975.000000
Mean	648.561465	0.904431
Std	281.211687	0.313354
Min	0.000000	0.000000
25%	449.000000	1.000000
50%	605.000000	1.000000
75%	788.000000	1.000000
Max	5584.000000	15.000000

**Table 3.1: Descriptive Statistics**

- **Qty**: Most orders involve a single item (**median = 1, max = 15**), with an average of **0.90** and a standard deviation of **0.31**.
- **Amount**: The average order value is **₹609** (**min = ₹0, max = ₹5,584**), with a standard deviation of **₹313.35**. The middle 50% of orders range from **₹413 to ₹771**.

Most orders consist of a single item, with the median quantity being 1 and a maximum of 15 items per order. The average order value is ₹609, indicating a moderate purchase size across transactions. Key attributes such as **Status, Fulfillment type, and B2B classification** play a crucial role in analyzing order patterns, identifying trends in cancellations, and assessing the impact of promotions. These factors help in understanding customer behavior, optimizing fulfillment strategies, and improving overall sales performance.

## 4 Detailed explanation of analysis process/ method

The data processing, cleaning, and analysis were primarily conducted using Python in Google Colab and Google Sheets (Excel). Pandas and NumPy were used for data manipulation, while Matplotlib and Seaborn facilitated key visualizations. Additional insights, such as Pareto analysis, were derived using Excel Pivot Tables.

### 4.1. Data Preprocessing

To ensure data accuracy and integrity, irrelevant columns such as **"Index"** and an unnamed 22nd column were removed. A thorough check for missing values revealed key gaps in the **"Courier Status"** (5.33%), **"Amount"** (6.04%), **"promotion-ids"** (38.11%), and **"fulfilled-by"** (69.55%) fields. Imputation strategies were applied:

- **"Courier Status"** and **"fulfilled-by"** missing values were flagged for further investigation, as their absence could indicate logistical inefficiencies.
- **"Amount"** values were checked for inconsistencies, and missing values were handled based on order trends.
- **"promotion-ids"** were analyzed separately to assess the impact of missing promotional data on sales and revenue.

Handling these missing values was necessary to prevent skewed insights, and appropriate imputation strategies were applied based on business relevance to maintain data consistency and ensure meaningful analysis.

### 4.2. Trend Analysis: Revenue and Volume Over Time

Trend analysis was performed to examine sales patterns and revenue fluctuations over time. Daily sales volume and revenue trends were analyzed using Pivot Tables to detect short-term demand variations and overall performance consistency. Additionally, monthly revenue trends were visualized using Matplotlib to identify seasonal peaks and declines, aiding in demand forecasting, inventory management, and resource allocation. These visualizations were prepared to support subsequent interpretation of demand fluctuations and temporal sales patterns.

### 4.3. Sales and Revenue Distribution Among Categories and SKUs

To determine which product categories and SKUs contributed the most to revenue and sales volume, **Pandas' groupby(), sum(), and ranking methods** were used.

- **Category-wise revenue contribution** was calculated, ranking categories from highest to lowest revenue generators.
- **SKU-level sales performance** was analyzed using **data filtering and aggregation**, detecting high-performing and underperforming products.

A **Pareto analysis (80/20 rule)** was conducted in **Excel using Pivot Tables**, with **SKUs in rows and aggregated amount (sum) as values**, confirming whether a small percentage of SKUs contributed to the majority of revenue, guiding inventory and promotional strategies.

### 4.4. Fulfillment Analysis

To evaluate the efficiency of different fulfillment methods, revenue and delay data were aggregated by fulfillment type and analyzed using a grouped column chart with Matplotlib. This visualization allowed for a direct comparison between Amazon Fulfillment and Merchant Fulfillment, highlighting differences in performance. By assessing revenue generation and delay occurrences, this analysis provided insights into logistical efficiency and its impact on customer experience and overall business performance.

### 4.5. Cancellations and Delays Analysis

To assess the impact of cancellations, cancellation rates were calculated for different fulfillment methods using a simple formula:

$$\text{Cancellation Rate} = (\text{Number of Cancelled Orders} / \text{Total Orders}) \times 100$$

This methodology allows for a more meaningful comparison between Amazon-fulfilled and Merchant-fulfilled orders by normalizing cancellation counts relative to total order volume.



Initially, there was no **"Delay"** column in the dataset. It was created using **Pandas** based on specific conditions involving **"Status"** and **"Courier Status"**:

Status	Courier Status	Delay Flag
Pending	Unshipped, UNKNOWN	Yes
Shipping	Unshipped, UNKNOWN	Yes
Pending - Waiting for Pick Up	Unshipped, UNKNOWN	Yes
Shipped - Out for Delivery	Unshipped, UNKNOWN	Yes

A strong correlation was observed between **delays and cancellations**, as orders experiencing prolonged delays had a significantly higher chance of being canceled. Merchant-fulfilled orders were more prone to delays due to slower shipping times and inconsistent tracking updates, leading to increased customer dissatisfaction and order cancellations.

Additionally, **category-wise cancellation rates** were analyzed and visualized using **Seaborn bar charts**, identifying high-risk product segments contributing to revenue loss.

#### 4.6. Revenue Lost and Order Status Distribution

Revenue loss from **cancellations and returns** was quantified using **Pandas' aggregation functions**, allowing for a structured assessment of financial impact. Order statuses were categorized and analyzed to measure their contribution to overall revenue fluctuations.

The distribution of order statuses was examined using **Seaborn heatmaps and alternative visualizations**, ensuring a clear representation of fulfillment performance. This methodology helped identify patterns in **order completion, cancellations, and returns**, providing insights into potential revenue leakage and operational inefficiencies.

#### 4.7. Fulfillment and Promotional Impact Comparison

Orders were segmented by promotion status and fulfillment type using **Pandas' filtering and groupby() operations**.

- Separate visualizations were created using Matplotlib/Seaborn to analyze revenue and cancellation rates between promotional and non-promotional orders.
- Total revenue and cancellation rates were calculated for each segment, while order counts were compared between fulfillment types to assess promotional effectiveness. This evaluation highlighted how fulfillment efficiency and targeted promotions work together to enhance performance.

Source of my analysis :-

- [Colab Notebook 1](#)
- [Colab Notebook 2](#)
- [Google Sheet 1](#)
- [Google Sheet 2](#)

## 5 Results and Findings

### 5.1 General Trends: Sales & Revenue

Sales volume between April and June 2022 was highly volatile. Early April witnessed significant fluctuations with a major mid-April drop—likely due to supply chain disruptions or operational challenges.

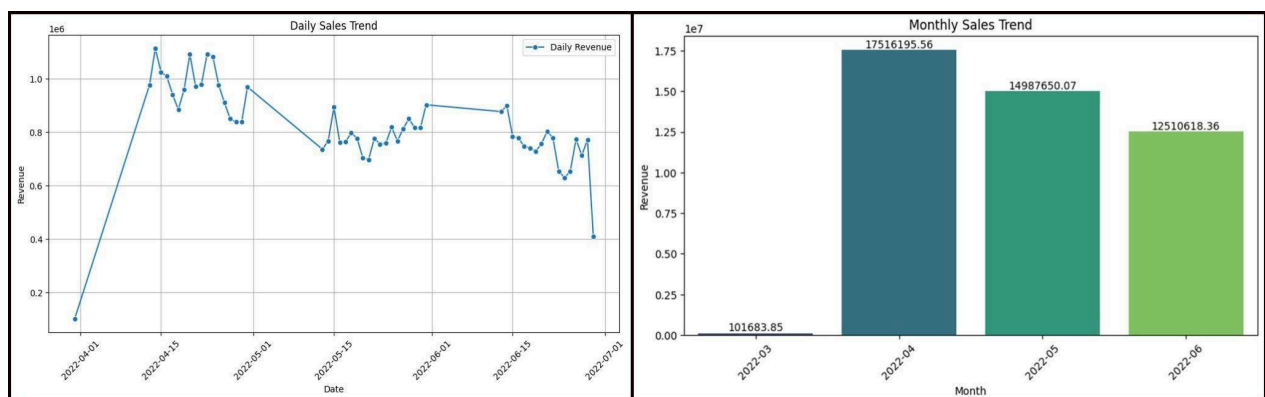


fig. 5.1.1

Although volumes stabilized in May, a gradual decline through June suggests waning customer interest or stock shortages impacting order fulfillment. Similarly, revenue trends mirrored these patterns. March registered minimal revenue (approximately ₹101K), possibly

due to early-stage operational constraints. In contrast, April peaked at around ₹17.5 million, but revenue then fell to about ₹14.9 million in May (a roughly 15% decline) and further to approximately ₹12.5 million in June (a 16% drop). The sharp mid-April dip followed by a sustained decline indicates that while strong demand existed initially, external factors—such as market disruptions or inventory issues—subsequently contributed to revenue loss.

## 5.2 Category & SKU-Level Performance

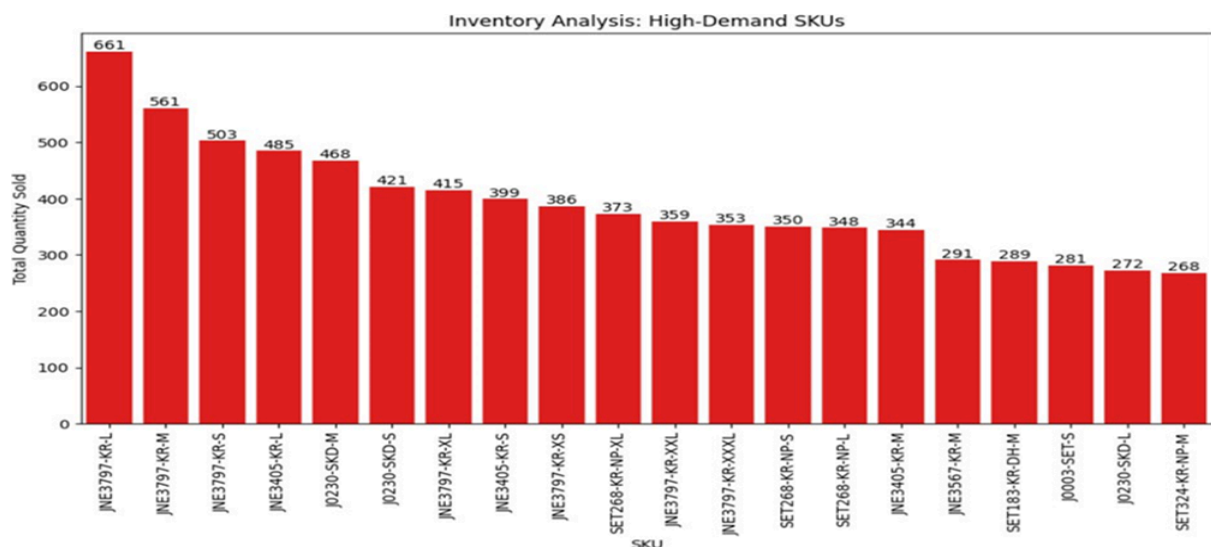


fig. 5.2.1

High-demand SKUs show a clear customer preference. The SKU **"JNE3797-KR-L"** leads with **661** units sold, followed by **"JNE3797-KR-M"** (**561** units) and **"JNE3797-KR-S"** (**503** units), emphasizing the importance of these sizes. This clustering suggests that maintaining optimal stock for these items is critical to prevent stockouts.

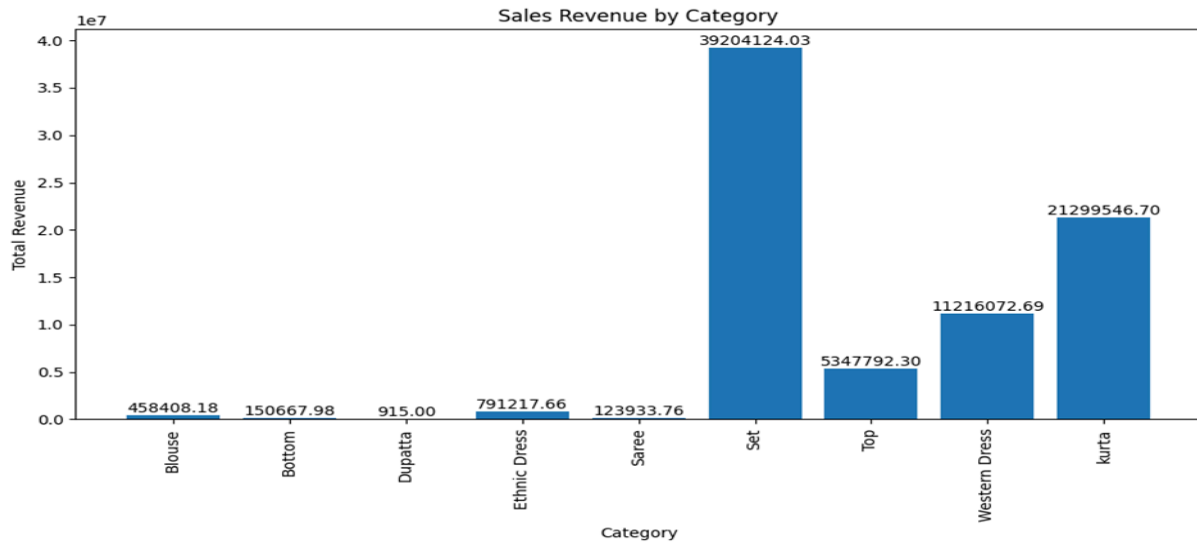


fig. 5.2.2

At the category level, the **"Set"** category generates the highest revenue (**₹39.2M**), followed by **"Kurta"** (**₹21.3M**) and **"Western Dress"** (**~₹11.2M**), indicating that high price points in some categories drive significant revenue despite lower volumes.

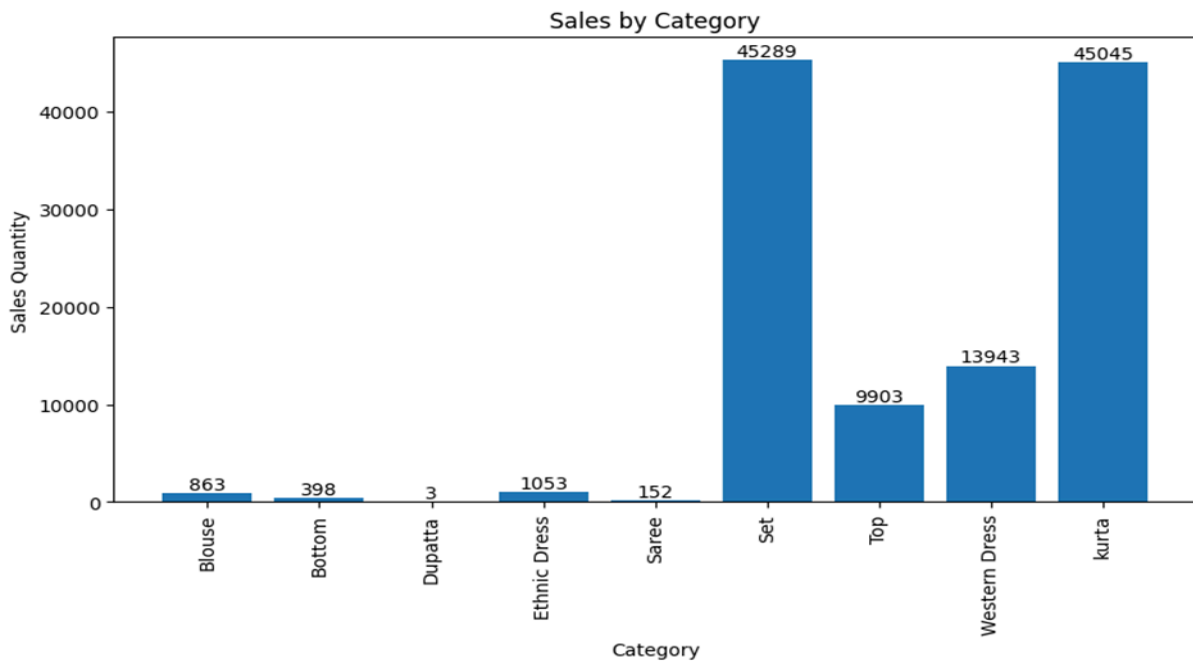


fig. 5.2.3

In terms of sales volume, **"Set"** and **"Kurta"** each achieve around **45,000** units, while **"Western Dress"** registers approximately **13,943** units; categories like **"Dupatta"** and **"Saree"** have minimal sales.

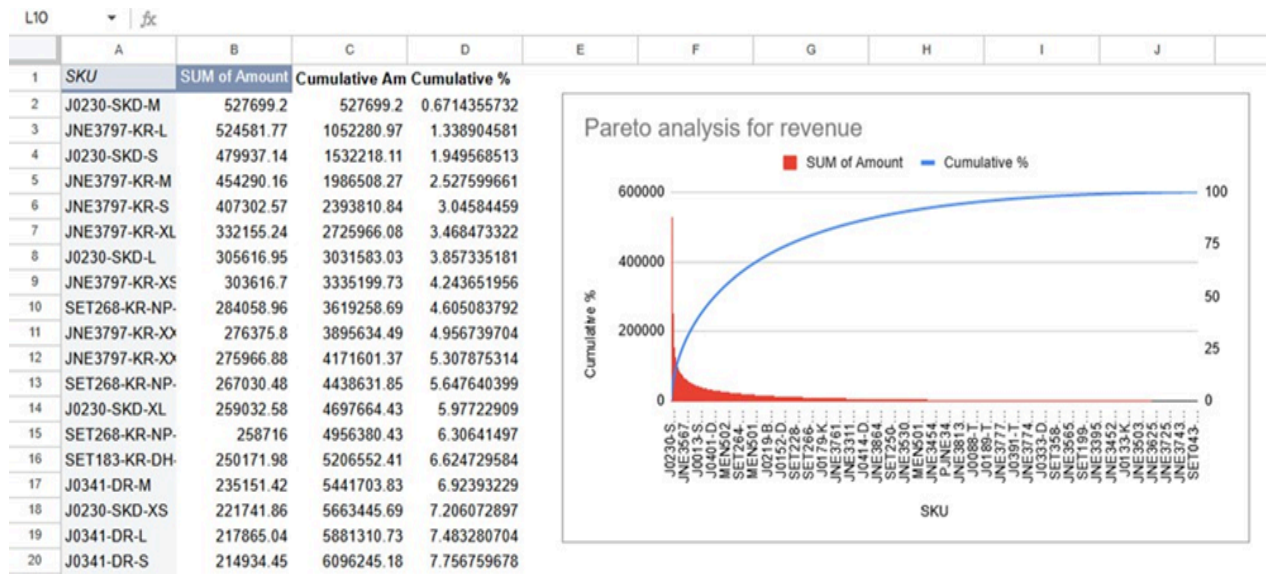


fig. 5.2.4

A Pareto analysis further confirms that a small percentage of SKUs such as **JNE3797-KR-L** (₹524,581.77), **JNE3797-KR-M** (₹459,240.16), and **J0230-SKD-M** (₹527,699.2) account for the majority of revenue, underscoring the 80/20 rule.

The category-wise analysis reveals a clear concentration of both revenue and sales volume in a few key segments, particularly "Set" and "Kurta," which together account for the majority of the performance. While "Western Dress" also contributes notably to revenue, it does so with relatively lower sales volume, suggesting higher average order values. In contrast, categories like "Dupatta" and "Saree" show negligible performance, highlighting underperforming areas that may require strategic reassessment. These insights can guide inventory planning, promotional focus, and category-level investments to optimize overall business outcomes.

### 5.3 Fulfillment Analysis

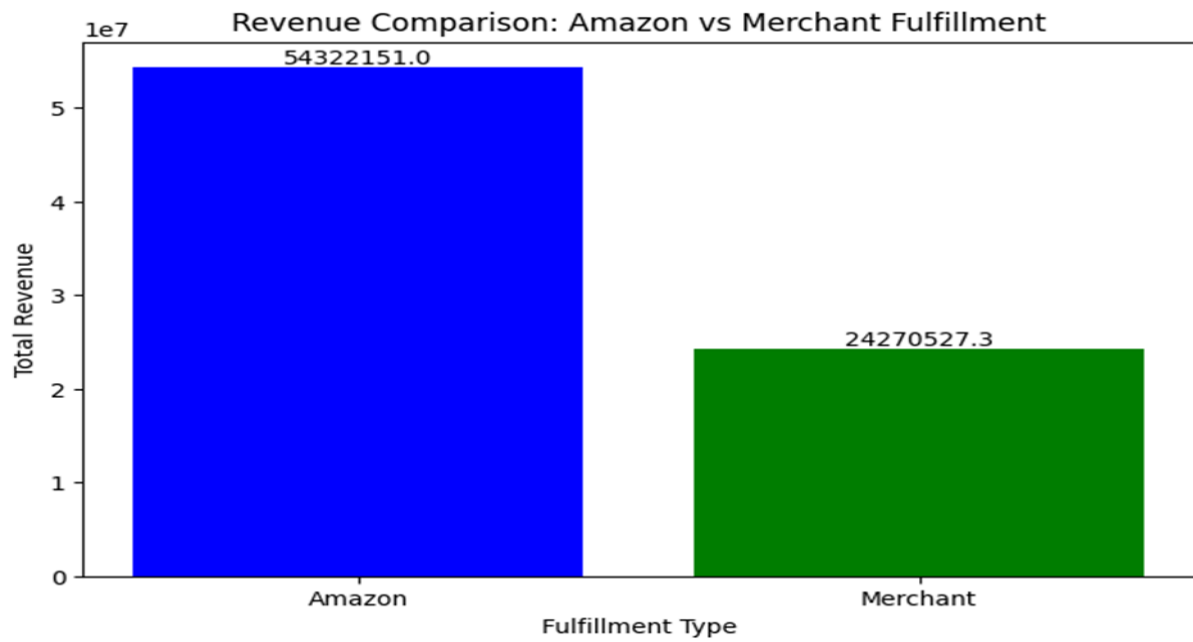


fig. 5.3.1

The revenue generated through Amazon Fulfillment (₹54.3M) is 2.24 times that of Merchant Fulfillment (₹24.3M), highlighting the impact of a robust fulfillment system on overall sales.

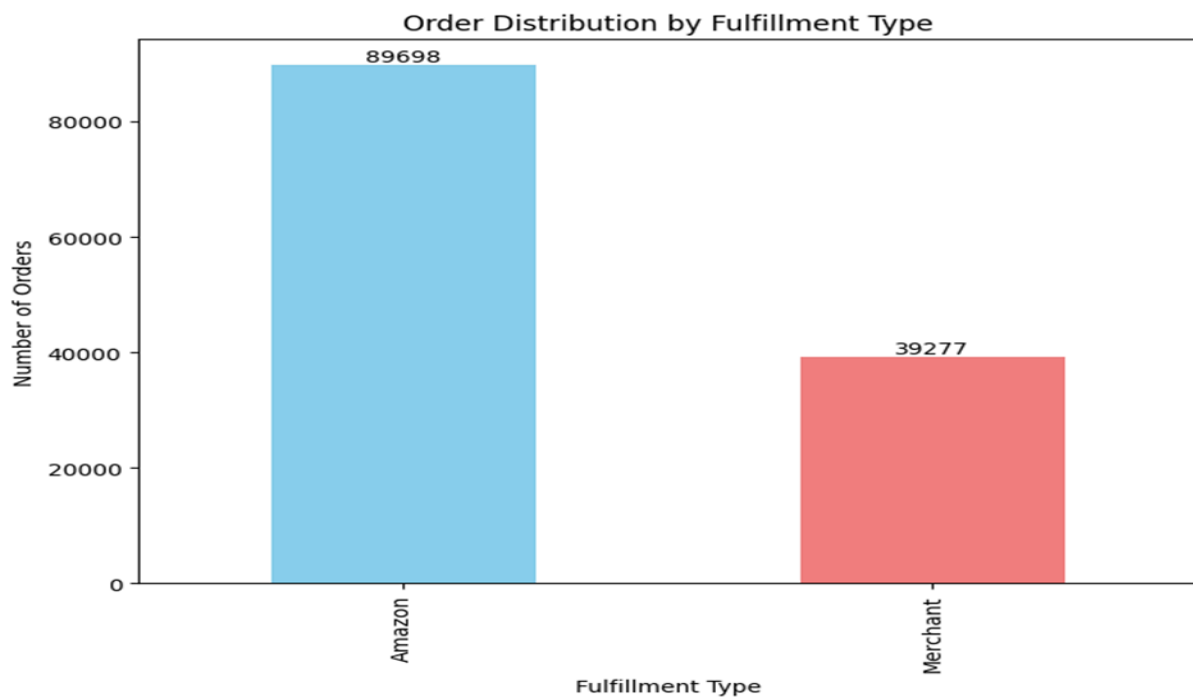
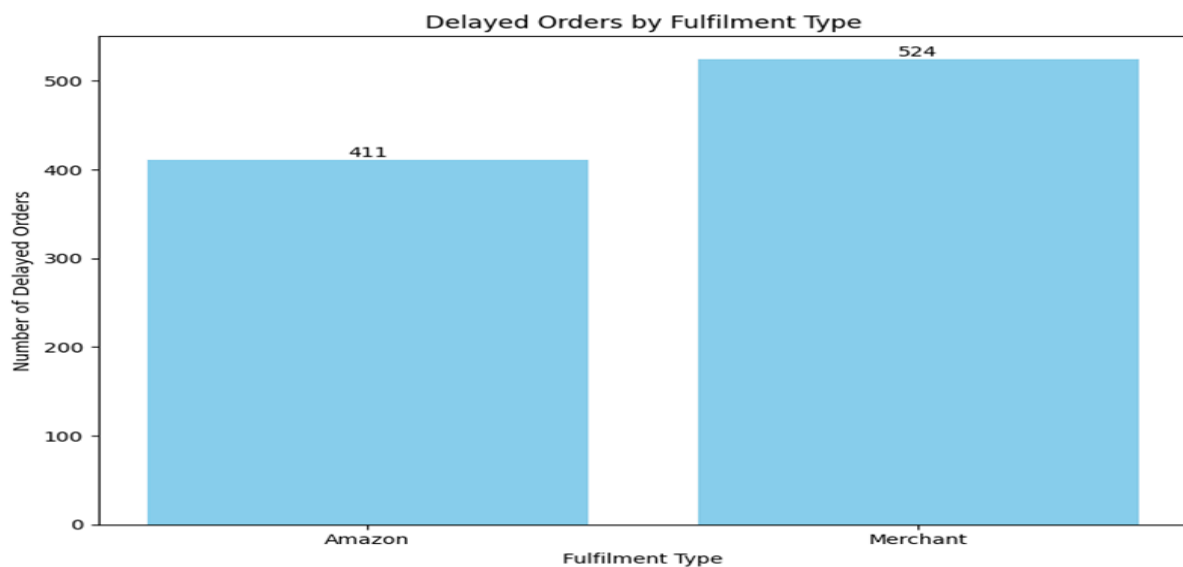


fig. 5.3.2

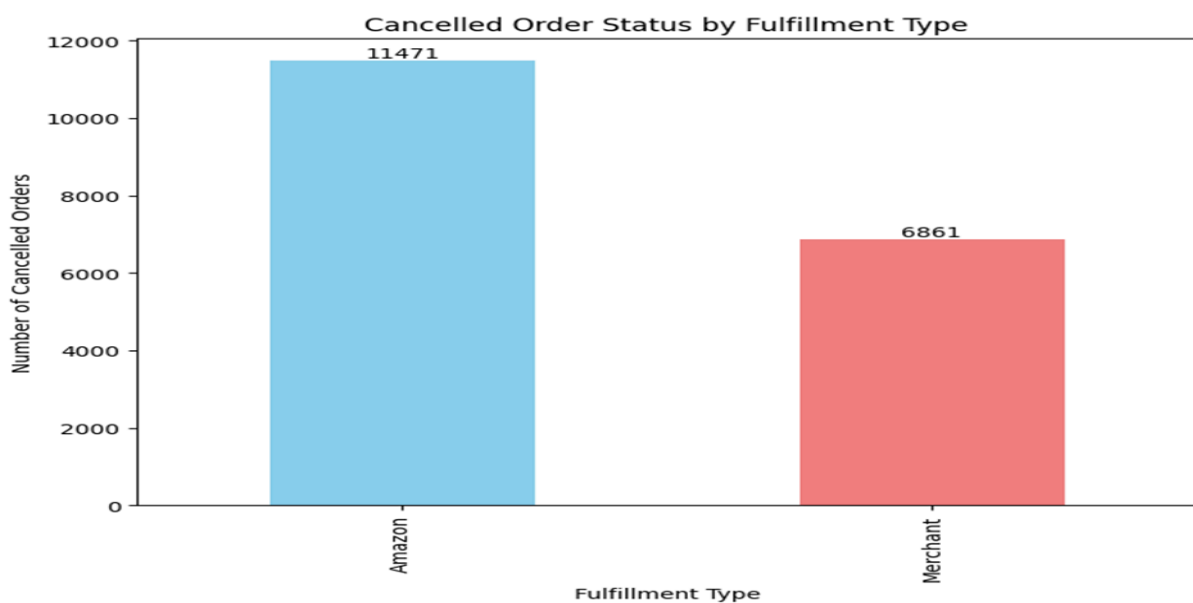
Moreover, Amazon processed 89,698 orders, more than double the 39,277 handled by Merchant Fulfillment, reflecting a strong customer preference.



**fig. 5.3.3**

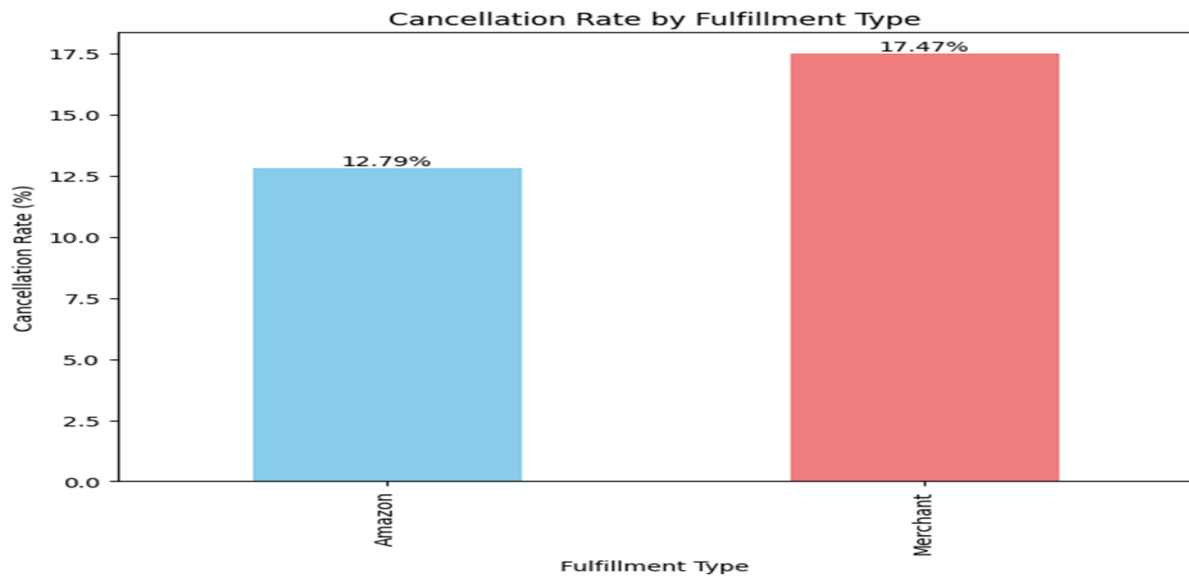
Efficiency is also evident: Amazon-fulfilled orders had 411 delayed orders compared to 524 for Merchant Fulfillment. These differences suggest that Amazon's logistics and tracking capabilities are more efficient, leading to fewer delays and, consequently, higher customer satisfaction and revenue.

#### **5.4 Cancellations & Delays Analysis**



**fig. 5.4.1**

**Fig. 5.4.1** shows that Amazon has a higher absolute number of cancelled orders compared to Merchant. This is expected, considering the much larger volume of orders processed by Amazon. However, to better understand the severity of cancellations, it is essential to analyze the cancellation rate instead of just absolute values.



**fig. 5.4.2**

**Fig. 5.4.2** presents the cancellation rate, which is calculated as:

$$\text{Cancellation Rate} = (\text{Number of Cancelled Orders} / \text{Total Orders}) \times 100$$

Using this formula:

- **Amazon:**  $(11,471 / 89,698) \times 100 \approx 12.79\%$
- **Merchant:**  $(6,861 / 39,277) \times 100 \approx 17.47\%$

**Fig. 5.4.2** provides this rate-based perspective, revealing that Merchant-fulfilled orders have a higher cancellation rate compared to Amazon-fulfilled ones. This suggests that despite having fewer total cancellations, Merchant orders are more prone to being canceled on a per-order basis. This discrepancy is likely due to slower shipping times and inconsistent tracking updates associated with Merchant fulfillment.

Additionally, delays are significant. As previously defined in **Section 4.5 (Cancellations and Delays Analysis)**, orders were flagged as delayed based on specific conditions involving order status and courier tracking information.:



```

delay_conditions = (
    ((df['Status'] == 'Pending') & (df['Courier Status'].isin(['Unshipped', 'UNKNOWN']))) |
    ((df['Status'] == 'Shipping') & (df['Courier Status'].isin(['Unshipped', 'UNKNOWN']))) |
    ((df['Status'] == 'Pending - Waiting for Pick Up') & (df['Courier Status'].isin(['Unshipped', 'UNKNOWN'])))
    ((df['Status'] == 'Shipped - Out for Delivery') & (df['Courier Status'].isin(['Unshipped', 'UNKNOWN'])))
)

```

**fig. 5.4.3**

- **Pending Orders:** Remain pending with courier status “Unshipped” or “UNKNOWN”.
- **Shipping Orders:** Lack confirmation of courier pickup.
- **Out for Delivery:** Remain “Unshipped” or “UNKNOWN” despite being near delivery.
- **Pending – Waiting for Pick-Up:** Have not been collected.

These delay conditions highlight operational inefficiencies that can often contribute to challenging fulfillment experiences. In general, delays have the potential to influence customer satisfaction and may lead to cancellation scenarios, particularly in time-sensitive or customer-centric environments. As outlined in **Section 5.3 (Fulfillment Analysis)** and referenced in **Figure 5.3.3**, Merchant-fulfilled orders recorded **524** delayed orders, compared to **411** under Amazon Fulfillment. This difference underscores how fulfillment methods influence delivery reliability—and by extension—customer satisfaction. Merchant delays, driven by opaque courier tracking and slower dispatch, appear more frequently associated with cancellations, highlighting a potential operational risk that warrants closer attention.

## 5.5 Promotions Impact

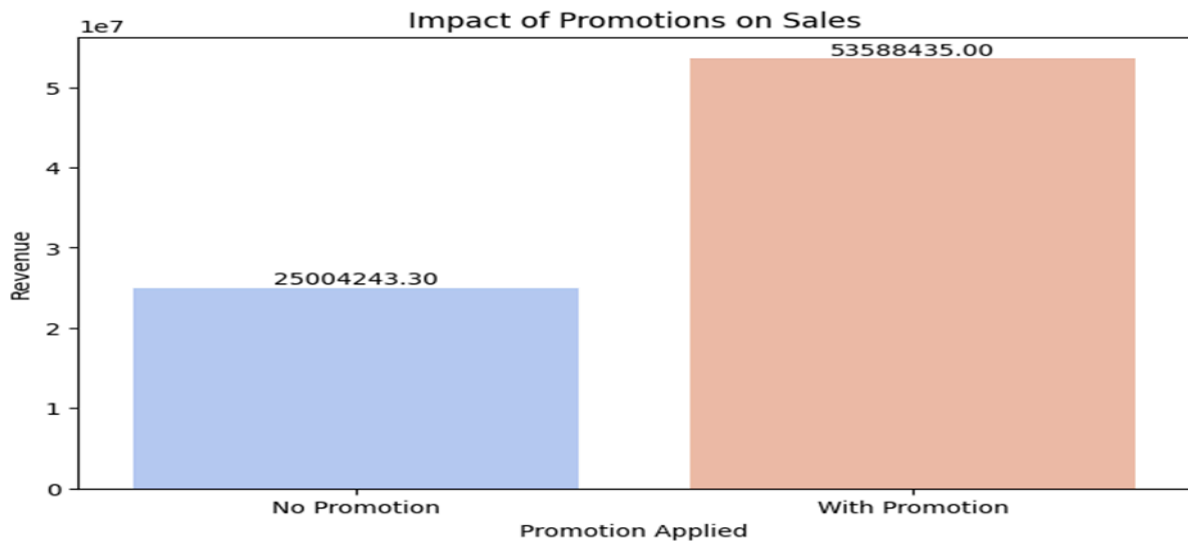


fig. 5.5.1

Promotions dramatically influence revenue and cancellation rates. Without promotions, revenue is around ₹25 million, while with promotions, it increases to approximately ₹53.6 million—a 114.4% uplift.

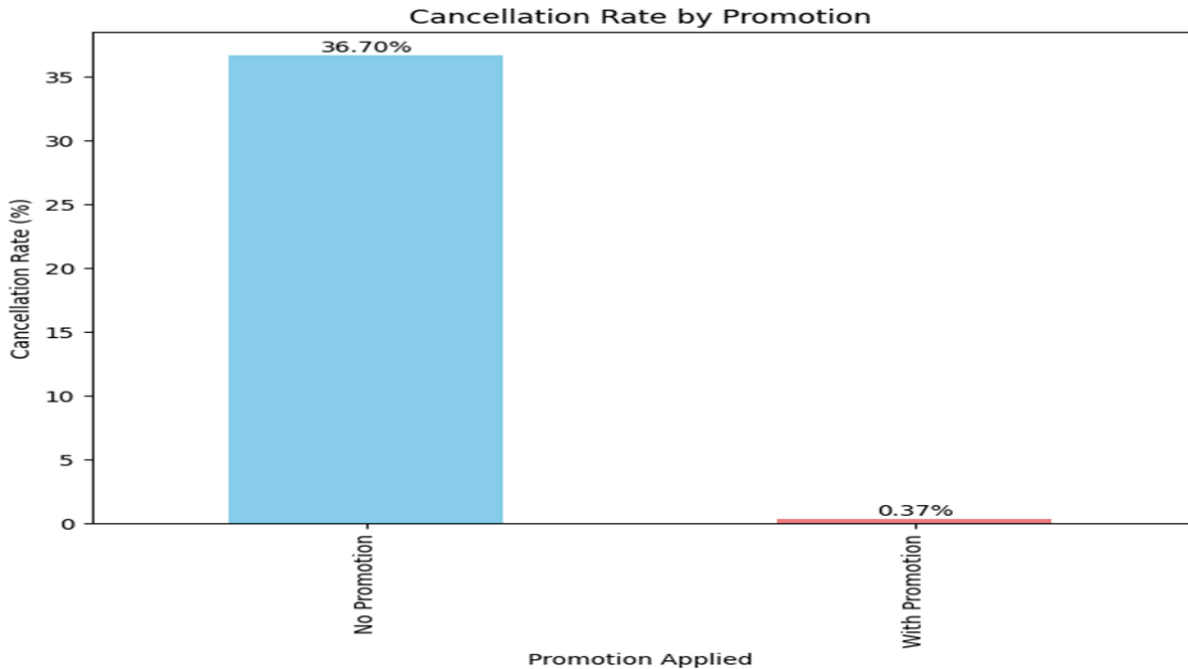


fig. 5.5.2

Moreover, orders without promotions have a cancellation rate of 36.70%, compared to a near-negligible rate of 0.37% with promotions as evident from **fig. 5.5.2**.

## **6 Interpretation of results and recommendation**

### **6.1. Interpretations**

- The analysis shows that while strong sales and revenue peaks—such as April’s performance—occur, they are followed by significant declines in subsequent months, indicating that initial success is undermined by underlying issues. Daily sales volatility reflects highly variable customer demand, and the mirrored revenue trends suggest that even minor disruptions in order processing or inventory can have immediate financial impacts.
- At the category level, the “Set” and “Kurta” segments dominate both revenue and sales volume, with a Pareto analysis confirming that a small percentage of SKUs drive the majority of revenue. This means that any inefficiencies affecting these key products disproportionately impact overall performance.
- Fulfillment analysis reveals that Amazon’s in-house system outperforms merchant-fulfilled orders, achieving higher revenue with lower cancellation rates and fewer delays. In contrast, merchant logistics suffer from slower shipping and inconsistent tracking, leading to higher cancellations and lost revenue.
- Promotions play a critical role by more than doubling revenue—from approximately ₹25 million to ₹53.6 million—and drastically reducing cancellation rates from 36.70% to 0.37%. These results underscore that targeted promotional strategies not only boost sales but also foster stronger customer commitment, demonstrating that an integrated approach is essential for enhancing overall performance.

## 6.2. Recommendations

**(1) *Fulfillment Optimization:*** Increase orders processed via Amazon's fulfillment network—leveraging its lower cancellation rate (12.79%) and higher revenue—while standardizing merchant procedures through real-time tracking and clear SLAs.

**(2) *Inventory Management:*** Implement dynamic reordering based on real-time data to ensure high-demand SKUs such as JNE3797-KR-L, JNE3797-KR-M are always stocked, and regularly perform SKU rationalization to focus on the top 20% driving 80% of revenue.

**(3) *Promotional Strategies:*** Develop consistent, targeted promotional campaigns for key categories (like Set and Kurta) and incentivize prepayment with discounts or loyalty rewards to substantially reduce cancellation rates.

**(4) *Customer Retention:*** Enhance order tracking and communication systems for real-time updates, and refine product listings with improved descriptions and quality control to minimize cancellations.

**(5) *Operational Efficiency:*** Automate order processing to reduce errors and expedite dispatch, and strengthen courier coordination through real-time tracking and performance incentives to cut delays.

Implementing these strategic recommendations will enable Amazon's e-commerce division to transform its operational framework and achieve significant gains in efficiency and profitability. By capitalizing on the strengths of Amazon's fulfillment network and enhancing the processes within merchant fulfillment, the business can effectively reduce cancellations and minimize delays. Coupled with dynamic inventory management and targeted promotional strategies, these measures will ensure optimal stock levels, improved customer retention, and robust revenue growth. Ultimately, this integrated approach not only mitigates the critical issues identified in the analysis but also positions Amazon to lead the competitive landscape with streamlined operations and an enhanced customer experience.