Analyzing Amazon Sales and Fulfillment Data: An End-to-End Data Analysis Project

Bogaram AkshayKumar

5 min read · Just now

Introduction:

In the rapidly evolving world of e-commerce, understanding customer behavior and optimizing sales strategies are crucial for staying ahead of the competition. This blog details an end-to-end data analysis project focused on Amazon's sales and fulfillment data, offering insights and actionable recommendations to improve performance.

Project Overview:

The primary objective of this project was to conduct a comprehensive analysis of Amazon's sales and fulfillment data. We aimed to identify key metrics impacting sales performance, understand customer purchase patterns, evaluate fulfillment efficiency, and provide data-driven recommendations to enhance overall operations.

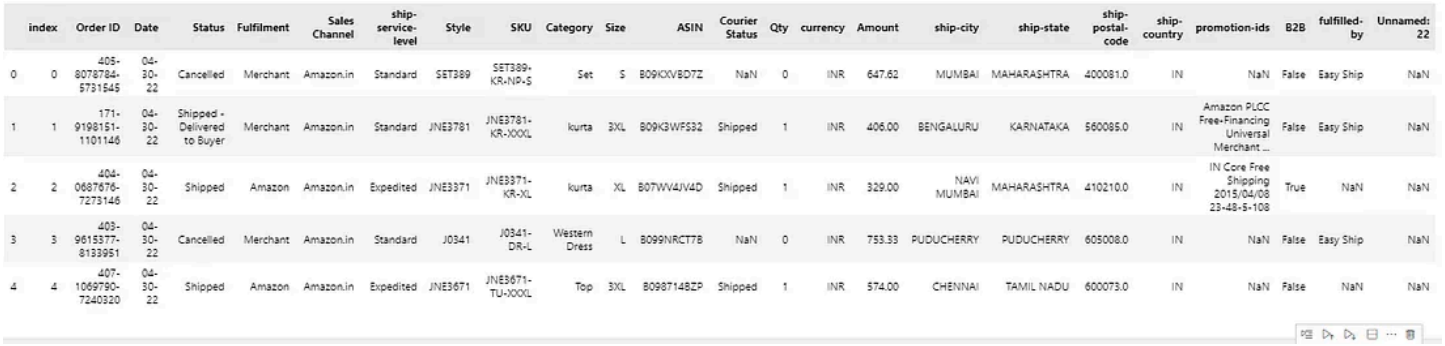
Data Sources:

E-Commerce Sales Dataset

Analyzing and Maximizing Online Business Performance

www.kaggle.com

How the data looks like:



	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Style	SKU	Category	Size	ASIN	Courier Status	Qty	currency	Amount	ship-city	ship-state	ship-postal-code	ship-country	promotion-ids	B2B	fulfilled-by	Unnamed: 22
0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standard	SET389	SET389-KR-NP-S	Set	S	B09KXVBD7Z	NaN	0	INR	647.62	MUMBAI	MAHARASHTRA	400081.0	IN	NaN	False	Easy Ship	NaN
1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	JNE3781	JNE3781-KR-XXXL	kurta	3XL	B09K3WFS32	Shipped	1	INR	406.00	BENGALURU	KARNATAKA	560085.0	IN	Amazon PLCC Free-Financing Universal Merchant …	False	Easy Ship	NaN
2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3371	JNE3371-KR-XL	kurta	XL	B07WV4JV4D	Shipped	1	INR	329.00	NAVI MUMBAI	MAHARASHTRA	410210.0	IN	IN Core Free Shipping 2015/04/08 23-48-5-108	True	NaN	NaN
3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standard	J0341	J0341-DR-L	Western Dress	L	B099NRCT7B	NaN	0	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008.0	IN	NaN	False	Easy Ship	NaN
4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedited	JNE3671	JNE3671-TU-XXXL	Top	3XL	B098714BZP	Shipped	1	INR	574.00	CHENNAI	TAMIL NADU	600073.0	IN	NaN	False	NaN	NaN

Steps involved in this project:

1. **Preliminary investigation about the data: (What i found?)**

- The data contains 12,8975 rows and 24 columns
- In which 8 columns are unnecessary, and contains more than 80k null values
- Out of all 24, most of them are of object data types and rest are numerical data types

2. **Data cleaning and manipulation:**

- The dataset included various columns such as Date, Status, Fulfillment, Sales Channel, Ship-Service-Level, Style, Category, Size, ASIN, Courier Status, Quantity

(Qty), Amount, Ship-City, Ship-State, and Ship-Postal-Code. The initial step involved data cleaning to ensure accuracy and consistency:

- Handling Missing Values: Rows with missing or zero values in critical columns like `Amount` and `Quantity` were filtered out, as these entries were deemed incorrect or incomplete.
- Formatting and Standardization: Column names were standardized, and unnecessary spaces were removed to facilitate easier analysis

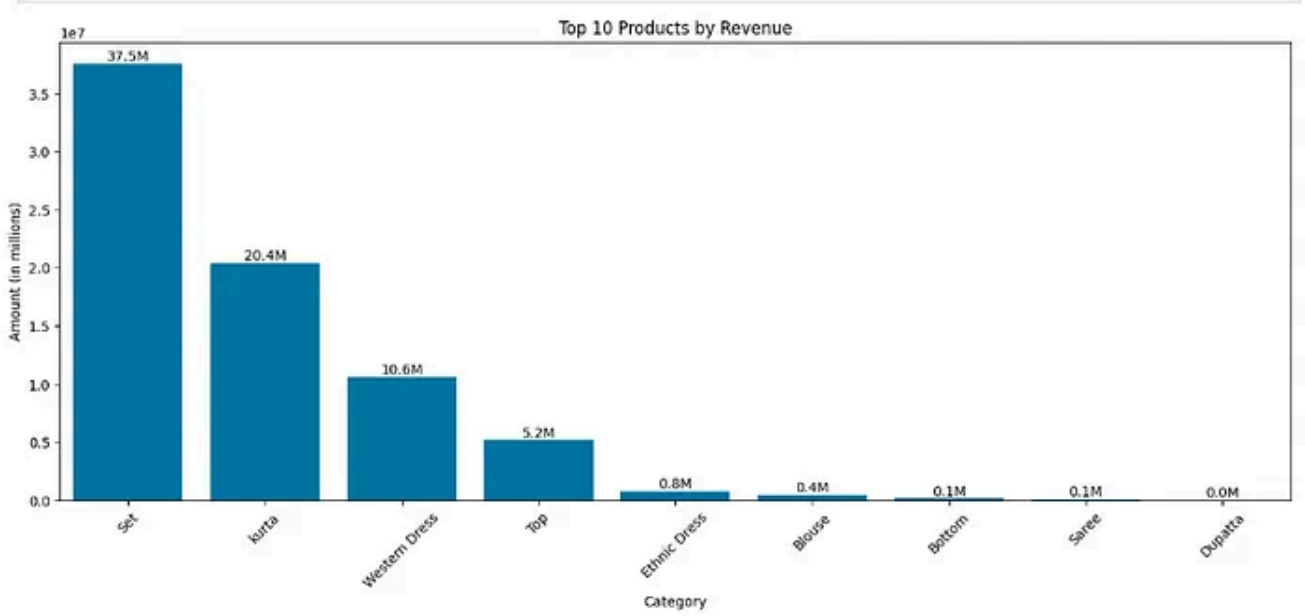
```
# Dropping rows where 'Amount' or 'Qty' is zero
df = df[(df['Amount'] != 0) & (df['Qty'] != 0)]
```

Exploratory Data Analysis (EDA)

1. **Sales Performance by Category**

- To understand which product categories were driving sales, we analyzed the quantity of items sold per category and their corresponding revenue. This analysis revealed that certain categories like ‘Set’ and ‘Kurta’ were top performers, especially during festive seasons.

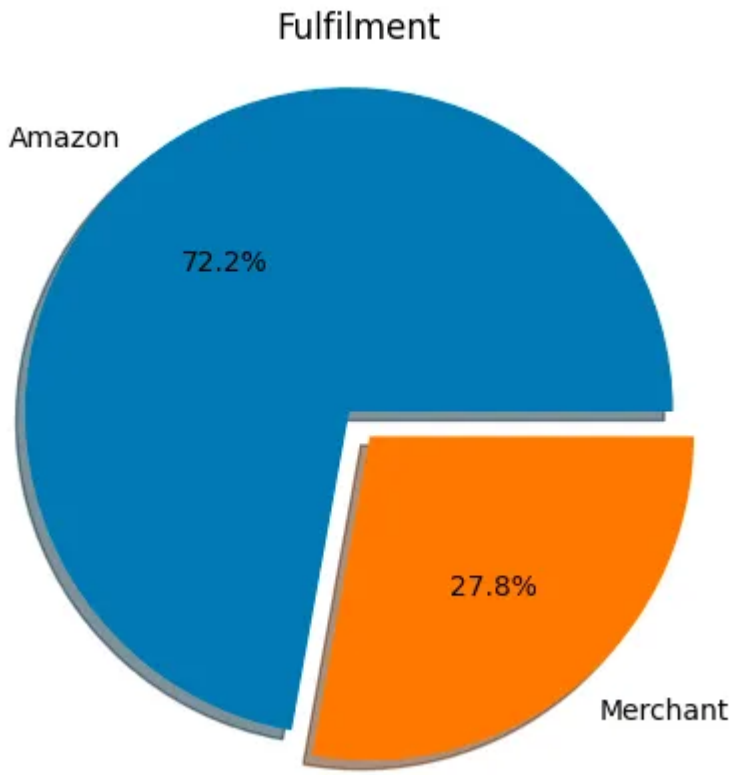
```
# Analyzing sales performance by category
category_sales = df.groupby('Category')['Amount'].sum().sort_values(ascending=F
sns.barplot(x='Category', y='Amount', data=category_sales, palette='viridis')
plt.title('Sales Performance by Category')
plt.xticks(rotation=45)
plt.show()
```



2. Fulfillment Status Distribution

Understanding fulfillment efficiency was crucial. The data showed that 94.4% of orders were successfully shipped, while 5.6% faced issues such as cancellations or returns. This metric highlighted the reliability of Amazon's fulfillment process and areas needing improvement.

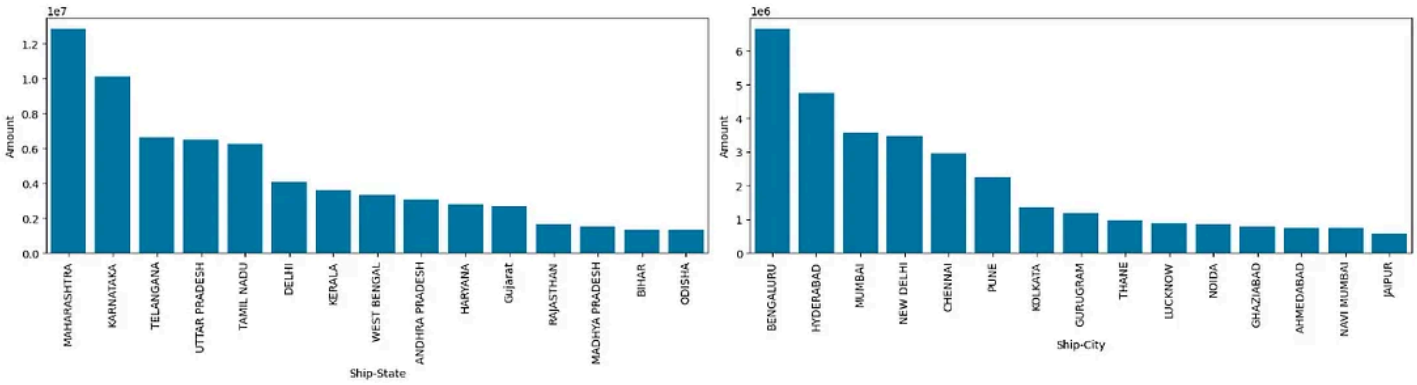
```
# Fulfillment status distribution
status_counts = df['Status'].value_counts()
status_counts.plot.pie(autopct='%1.1f%%', colors=['#4CAF50', '#FF9999', '#FFCC0
plt.title('Fulfillment Status Distribution')
plt.ylabel('')
plt.show()
```



3. Geographical Analysis of Sales

Analyzing sales data based on regions provided insights into customer demographics and preferences. States like Maharashtra, Karnataka, and Tamil Nadu were leading in sales volumes, with cities like Bengaluru and Hyderabad topping the charts.

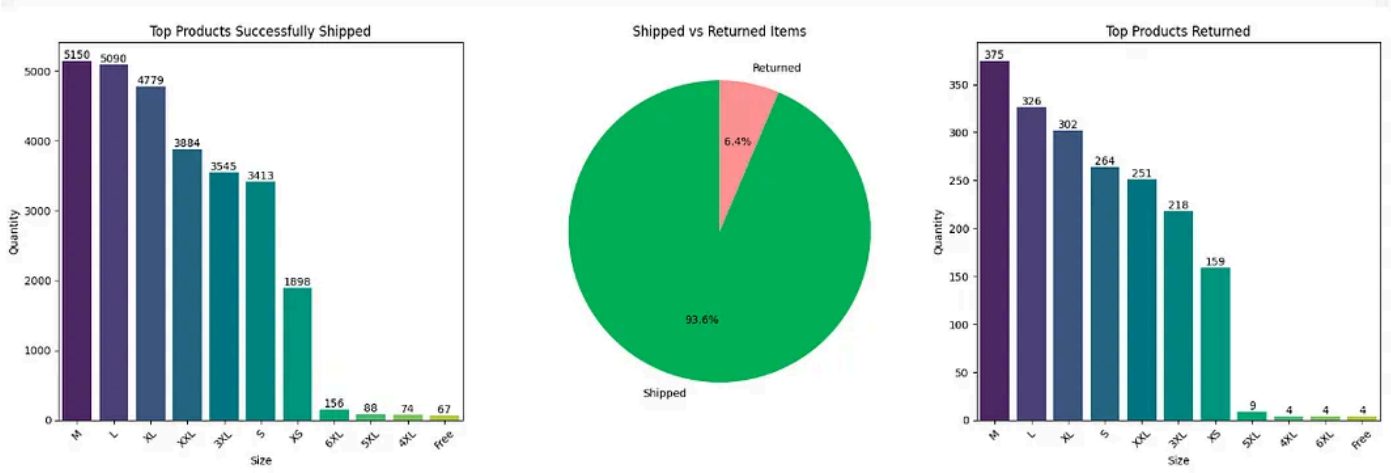
```
# Sales performance by state
state_sales = df.groupby('Ship-State')['Amount'].sum().sort_values(ascending=Fa
sns.barplot(x='Ship-State', y='Amount', data=state_sales, palette='viridis')
plt.title('Sales Performance by State')
plt.xticks(rotation=45)
plt.show()
```



4. Return Rates Analysis

Identifying products with high return rates helped in understanding customer dissatisfaction. The analysis showed that ‘Set’ and ‘Kurta’ had higher return rates, indicating possible issues with product quality or customer expectations.

```
# Return rates by category
returned_items = df[df['Status'] == 'Shipped - Returned to Seller']
return_rates = returned_items.groupby('Category')['Qty'].sum().sort_values(asce
sns.barplot(x='Category', y='Qty', data=return_rates, palette='viridis')
plt.title('Return Rates by Category')
plt.xticks(rotation=45)
plt.show()
```



Key Findings and insights:

1. Purchase Patterns and Customer Preferences:

- High Sales Volumes: Maharashtra and Karnataka lead in sales volumes, with — Bengaluru and Hyderabad being the top cities.
- Popular Products: Sets and Kurtas are the highest revenue-generating products, indicating a strong preference for these items during festive seasons like Ugadi.
- Size Preferences: Medium size products are the most purchased, reflecting a common customer preference.

2. Fulfillment Efficiency:

- High Fulfillment Rate: 94.4% of orders are shipped successfully, demonstrating effective fulfillment processes. Dominant Fulfillment Method: Amazon's fulfillment service dominates with 72.2% of orders, compared to 27.8% by merchants.
- Return Rates: The overall return rate is low, with only 6.5% of products being returned, indicating high customer satisfaction.
- Commonly Returned Items: Sets and Kurtas have the highest return rates, suggesting potential issues with these products, possibly related to size or quality.
- Dominant Sales Channel: B2B sales account for around 90% of total sales, highlighting the importance of bulk purchases in the business model.
- Product Pricing: Products like Sets and Kurtas are priced affordably, which drives their high sales volumes
- High Return Regions: Maharashtra and Uttar Pradesh have the highest return rates, with Bangalore and Hyderabad being the top cities for returns.
- Size and Return Rates: Medium-sized products are both the most purchased and most returned, indicating potential issues with size accuracy.

3. Geographical Insights:

- Maharashtra and Karnataka emerged as top states in sales, with Bengaluru and Hyderabad leading among cities.

4. Customer Preferences:

- Medium size products were the most popular, indicating customer preferences that can guide inventory management.

Recommendations:

Based on the analysis, several actionable recommendations were made:

1. Improve Product Quality: Addressing issues in top-returned categories like ‘Set’ and ‘Kurta’ can reduce return rates and enhance customer satisfaction.
2. Optimize Fulfillment Processes: Focusing on the reasons behind unshipped orders can improve the overall fulfillment rate.
3. Targeted Marketing Campaigns: Leveraging geographical insights to design targeted marketing strategies can boost sales in high-performing regions.
4. Enhance Size Guides: Providing detailed size guides can reduce the return rates related to size mismatches.

Conclusion:

This project provided a deep dive into Amazon’s sales and fulfillment data, offering valuable insights into customer behavior, product performance, and operational efficiency. The comprehensive analysis not only highlighted areas for improvement but also empowered stakeholders with data-driven strategies to optimize sales and customer satisfaction.

This end-to-end analysis exemplifies the power of data-driven decision-making in e-commerce, demonstrating how detailed insights can lead to significant business improvements.

Data Analysis

Edit profile

Written by Bogaram AkshayKumar

0 Followers