Introduction

The Amazon Sales Dataset provides a comprehensive record of sales transactions on the Amazon platform, detailing various aspects such as order ID, date, status, fulfillment method, sales channel, product category, size, quantity, amount, shipping details, and more. The primary objective of this analysis is to extract actionable insights that can inform strategic decision-making processes within the business.

Aim/Problem Statement

This project aims to conduct an in-depth analysis of the Amazon Sales Dataset to uncover patterns, trends, and key drivers of sales performance. By exploring the dataset, we seek to identify factors influencing sales, understand customer behavior, optimize inventory management, and enhance overall business operations.

Dataset Overview

The dataset contains a range of columns providing information on each sales transaction:

- Index: Index or identifier for each record.
- Order ID: Unique identifier for each order.
- Date: Date of the order.
- Status: Status of the order (e.g., Shipped, Cancelled).
- Fulfillment: Fulfillment method used for the order.
- Sales Channel: Channel through which the sale was made.
- Ship-service-level: Service level for shipping.
- Category: Category of the product.
- Size: Size of the product.
- Courier Status: Status of the courier (e.g., On the Way, Shipped).
- Qty: Quantity of the product ordered.
- Currency: Currency used for the transaction.
- Amount: Total amount of the order.
- Ship-city: City where the order is being shipped.
- Ship-state: State where the order is being shipped.
- Ship-postal-code: Postal code of the shipping address.
- Ship-country: Country where the order is being shipped.
- B2B: Binary indicator for Business-to-Business transaction.

19. Fulfilled-by: Indicates whether the order is fulfilled by Amazon or a third-party seller.

20. New: Indicates if the order is new.

21. PendingS: Indicates if the order is pending.

Importing the Necessary Libraries

❖   **numpy (np):** For numerical operations and array handling.

❖   **pandas (pd):** For data manipulation and analysis using DataFrames.

❖   **scipy (sp):** For scientific and statistical computing.

❖   **re:** For regular expression operations, if needed for data preprocessing.

❖   **time:** For tracking and measuring time during code execution.

❖   **matplotlib.pyplot (plt):** For creating static, interactive, and animated visualizations in Python.

❖   **seaborn (sns):** A statistical data visualization library based on Matplotlib, enhancing the aesthetics of plots.

❖   **os:** For interacting with the operating system, useful for managing files and directories.

❖   **plotly.express (px):** For creating interactive visualizations.

❖   **matplotlib (mpl):** A comprehensive library for creating static, animated, and interactive plots in Python

Top 5 Elements:

df.head()

	index	Order ID	Date	Status	Fulfilment	Sales Channel	ship-service-level	Category	Size	Courier Status	...	currency	Amount	ship-city	ship-state	ship-postal-code	ship-country	B2B	fulfilled-by	New	PendingS
0	0	405-8078784-5731545	04-30-22	Cancelled	Merchant	Amazon.in	Standard	T-shirt	S	On the Way	...	INR	647.62	MUMBAI	MAHARASHTRA	400081.0	IN	False	Easy Ship	NaN	NaN
1	1	171-9198151-1101146	04-30-22	Shipped - Delivered to Buyer	Merchant	Amazon.in	Standard	Shirt	3XL	Shipped	...	INR	406.00	BENGALURU	KARNATAKA	560085.0	IN	False	Easy Ship	NaN	NaN
2	2	404-0687676-7273146	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Shirt	XL	Shipped	...	INR	329.00	NAVI MUMBAI	MAHARASHTRA	410210.0	IN	True	NaN	NaN	NaN
3	3	403-9615377-8133951	04-30-22	Cancelled	Merchant	Amazon.in	Standard	Blazzer	L	On the Way	...	INR	753.33	PUDUCHERRY	PUDUCHERRY	605008.0	IN	False	Easy Ship	NaN	NaN
4	4	407-1069790-7240320	04-30-22	Shipped	Amazon	Amazon.in	Expedited	Trousers	3XL	Shipped	...	INR	574.00	CHENNAI	TAMIL NADU	600073.0	IN	False	NaN	NaN	NaN

Data Information:

df.info()

```
<bound method DataFrame.info of          index        Order ID      Date
0             0  405-8078784-5731545  04-30-22                  Cancelled
1             1  171-9198151-1101146  04-30-22  Shipped - Delivered to Buyer
2             2  404-0687676-7273146  04-30-22                    Shipped
3             3  403-9615377-8133951  04-30-22                  Cancelled
4             4  407-1069790-7240320  04-30-22                    Shipped
...         ...                  ...       ...                        ...
128971   128970  406-6001380-7673107  05-31-22                    Shipped
128972   128971  402-9551604-7544318  05-31-22                    Shipped
128973   128972  407-9547469-3152358  05-31-22                    Shipped
128974   128973  402-6184140-0545956  05-31-22                    Shipped
128975   128974  408-7436540-8728312  05-31-22                    Shipped

        Fulfilment Sales Channel ship-service-level  Category Size  \
0         Merchant     Amazon.in           Standard   T-shirt    S
1         Merchant     Amazon.in           Standard     Shirt  3XL
2           Amazon     Amazon.in           Expedited     Shirt   XL
3         Merchant     Amazon.in           Standard   Blazzer    L
4           Amazon     Amazon.in           Expedited  Trousers  3XL
...            ...           ...                ...       ...  ...
128971      Amazon     Amazon.in           Expedited     Shirt   XL
128972      Amazon     Amazon.in           Expedited   T-shirt    M
128973      Amazon     Amazon.in           Expedited   Blazzer  XXL
128974      Amazon     Amazon.in           Expedited   T-shirt   XS
128975      Amazon     Amazon.in           Expedited   T-shirt    S

        Courier Status  ... currency  Amount    ship-city    ship-state  \
0          On the Way  ...      INR  647.62       MUMBAI   MAHARASHTRA
1             Shipped  ...      INR  406.00    BENGALURU     KARNATAKA
2             Shipped  ...      INR  329.00  NAVI MUMBAI   MAHARASHTRA
3          On the Way  ...      INR  753.33   PUDUCHERRY    PUDUCHERRY
4             Shipped  ...      INR  574.00      CHENNAI    TAMIL NADU
```

Unique Values:

df.nunique()

```
index                128808
Order ID             120229
Date                     55
Status                   13
Fulfilment                2
Sales Channel             2
ship-service-level        2
Category                  9
Size                     11
Courier Status            4
Qty                      10
currency                  1
Amount                 1408
ship-city              8948
ship-state               69
ship-postal-code       9454
ship-country              1
B2B                       2
fulfilled-by              1
New                       0
PendingS                  0
dtype: int64
```
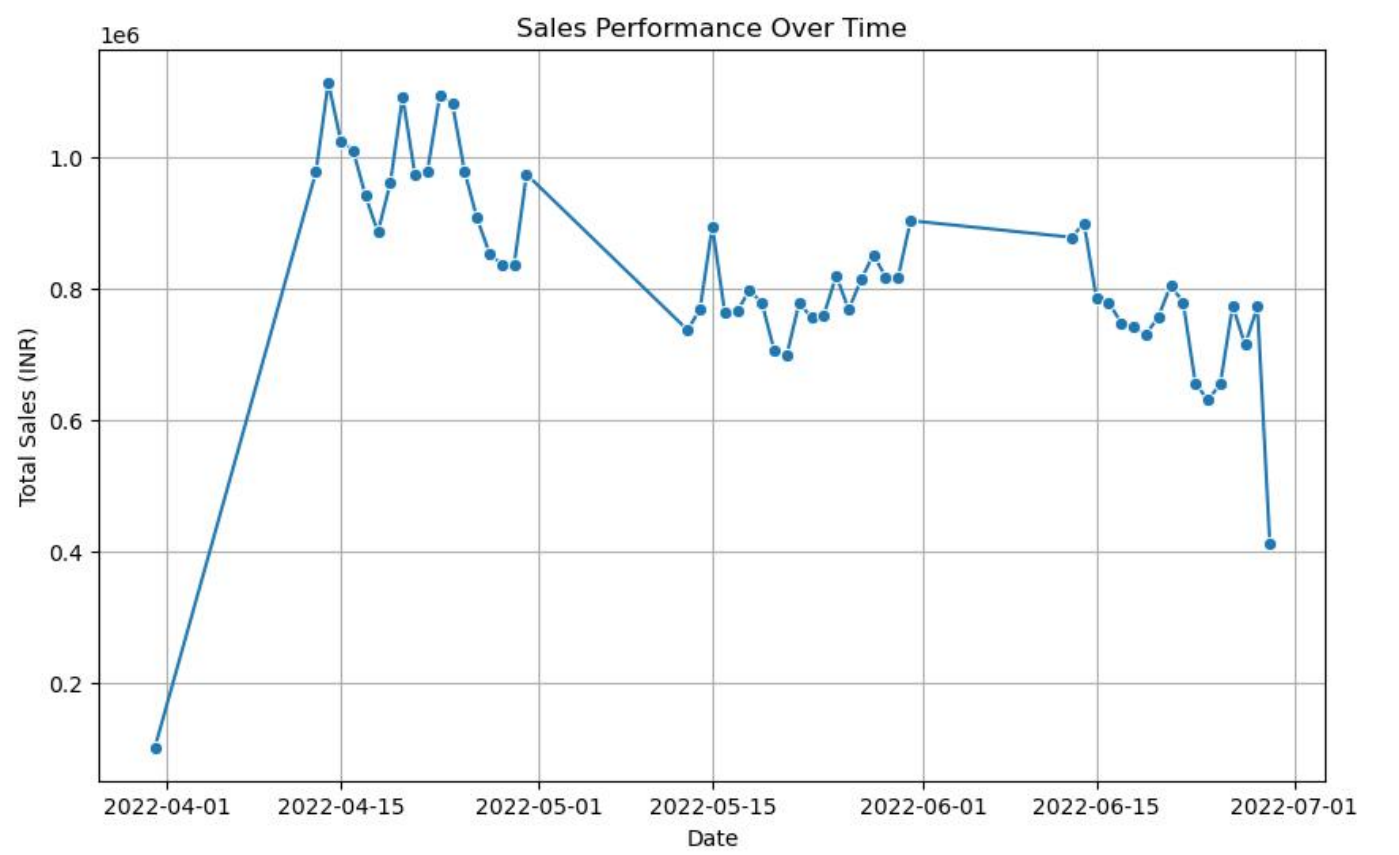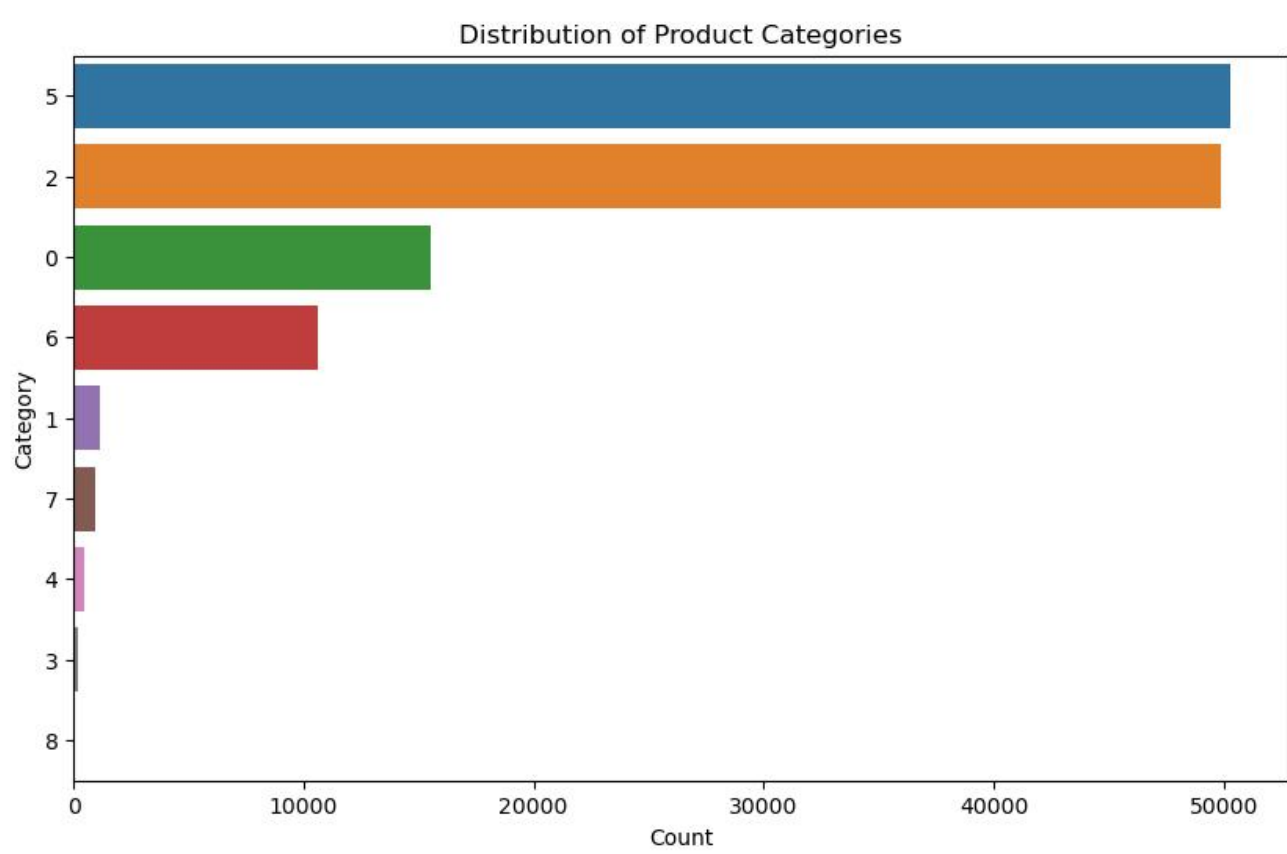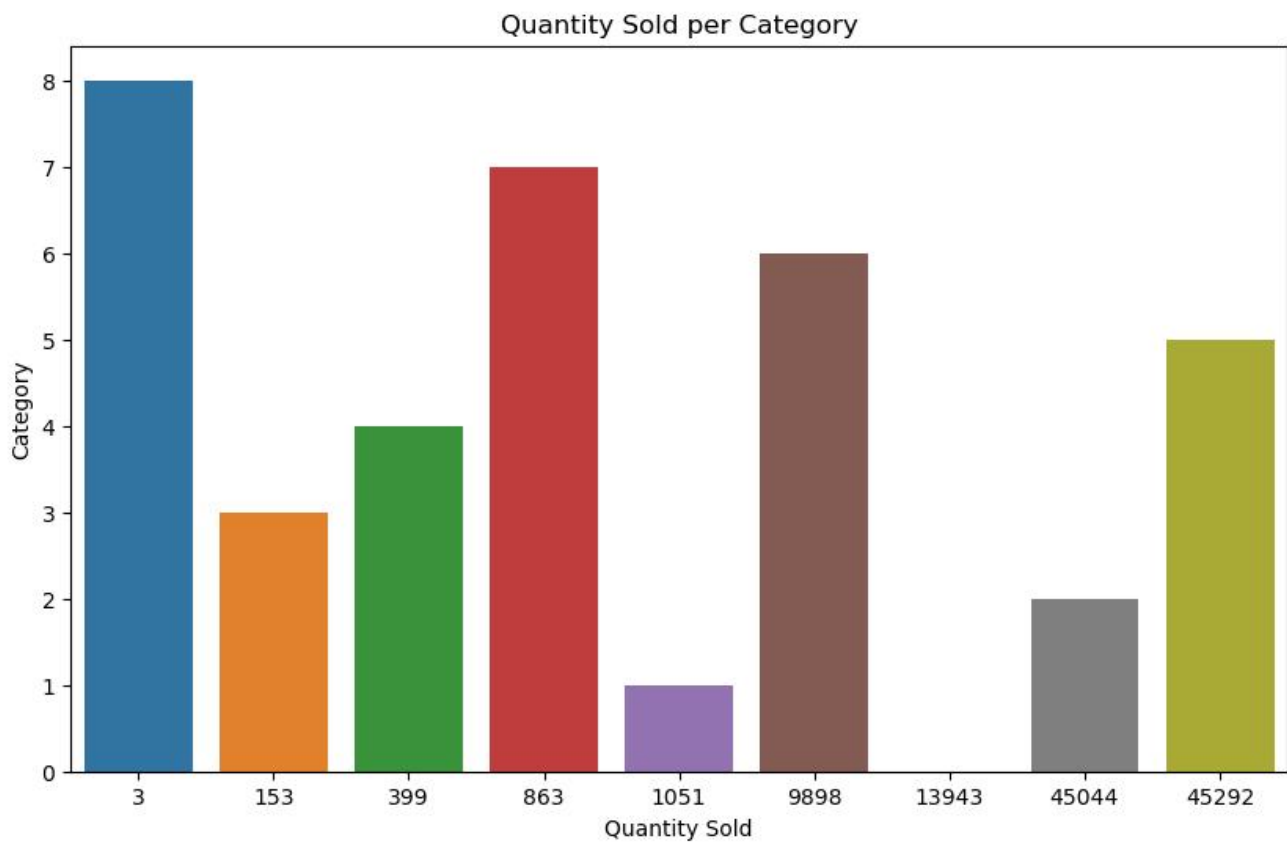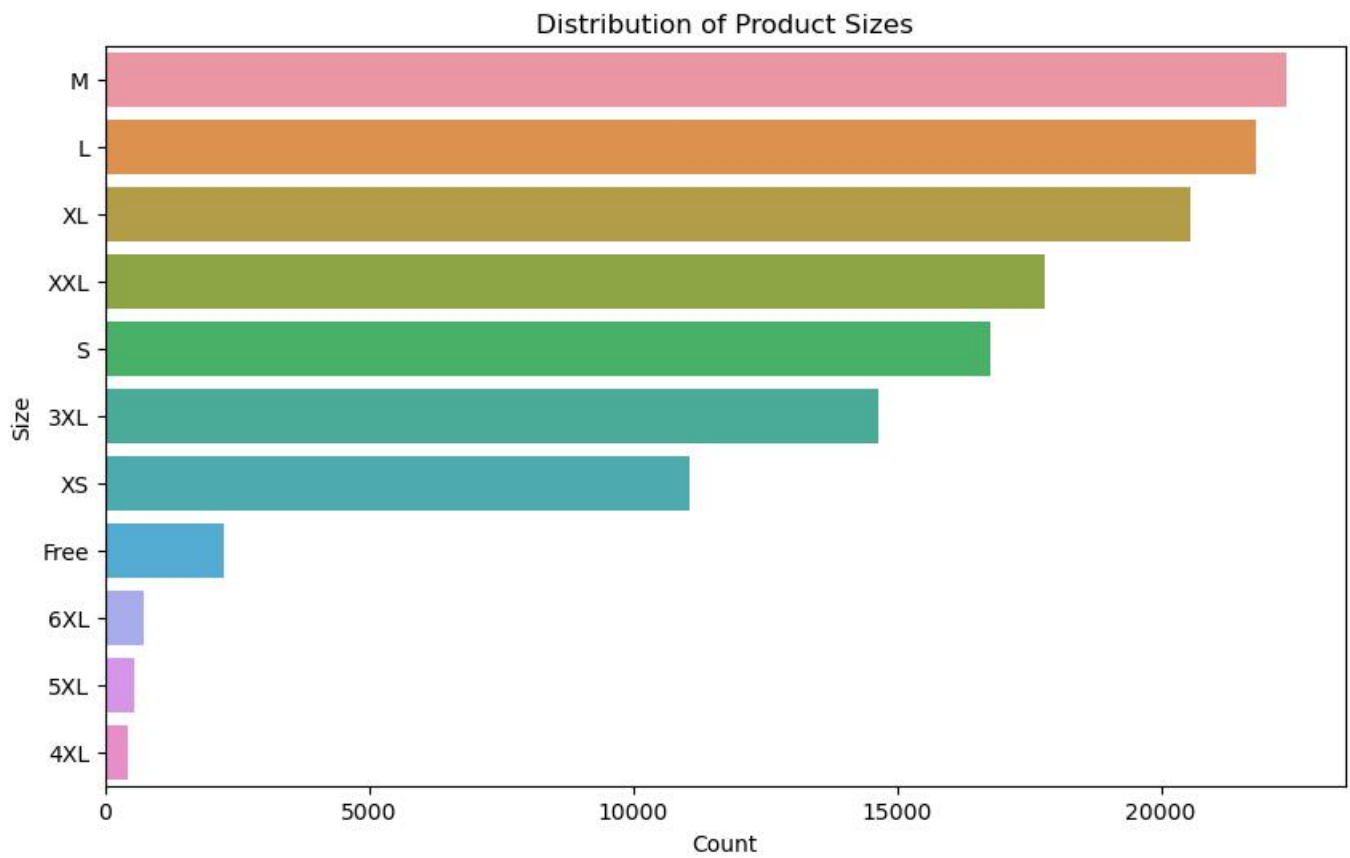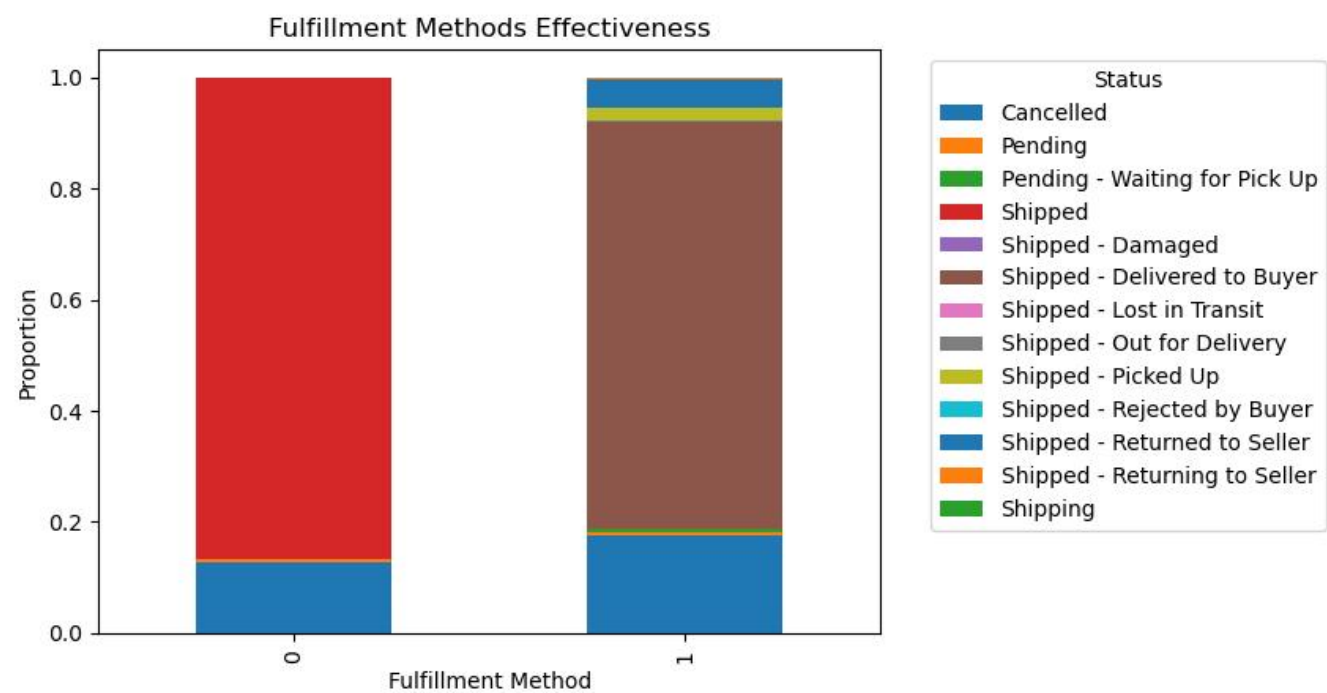
VISUALITIONS:

Sales Overview

Sales Performance Over Time

Product Analysis

Distribution of Product Categories

Distribution of Product Sizes

Quantity Sold per Category

Fulfillment Analysis

Distribution of Fulfillment Methods

Fulfillment Methods Effectiveness

Customer Segmentation

Customer Segmentation Based on Product Category and Size

Geographical Analysis

Top 10 Cities by Sales

Sales Distribution by State

(chart: Total Sales (INR) vs State)
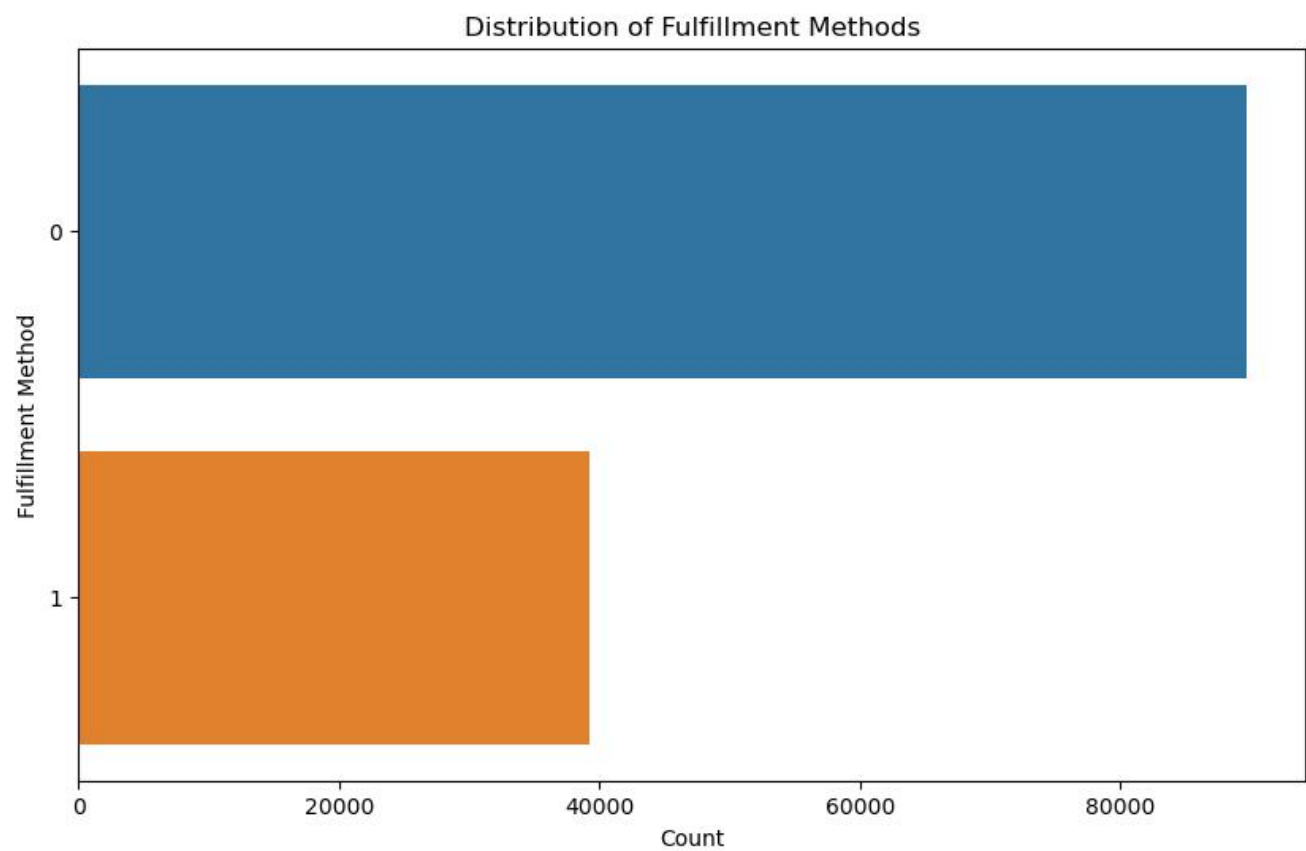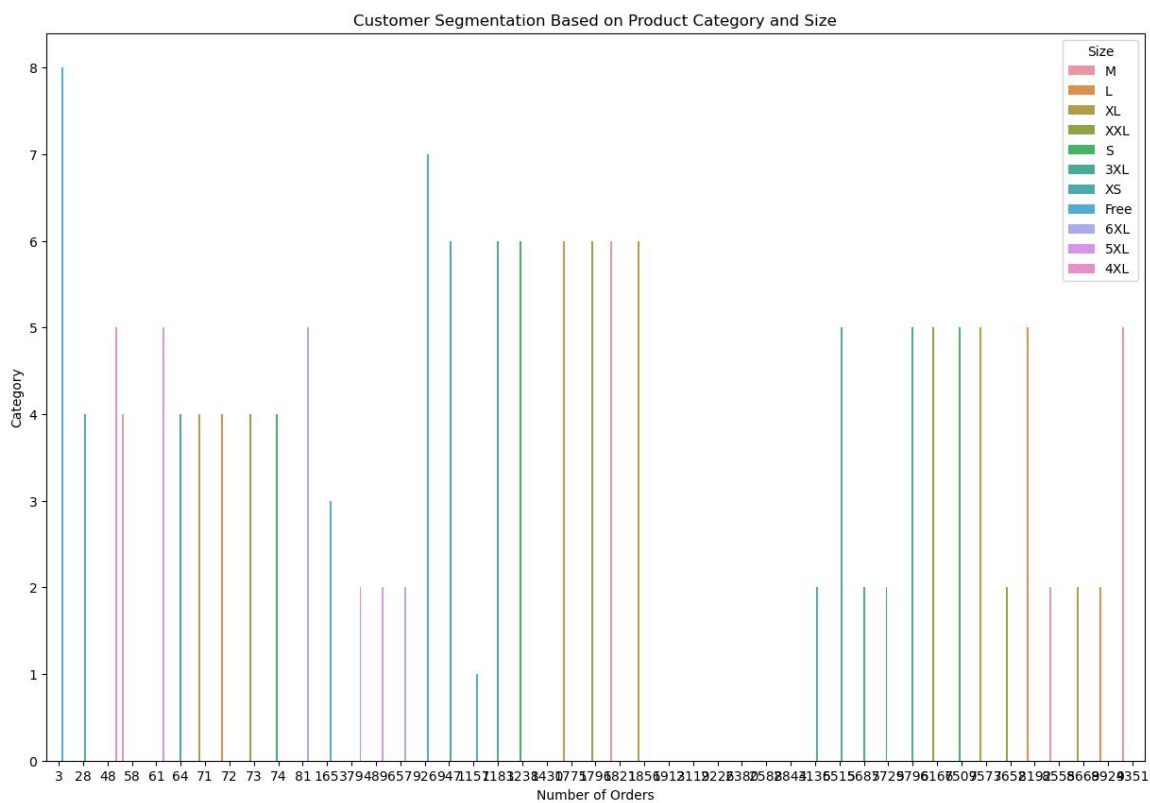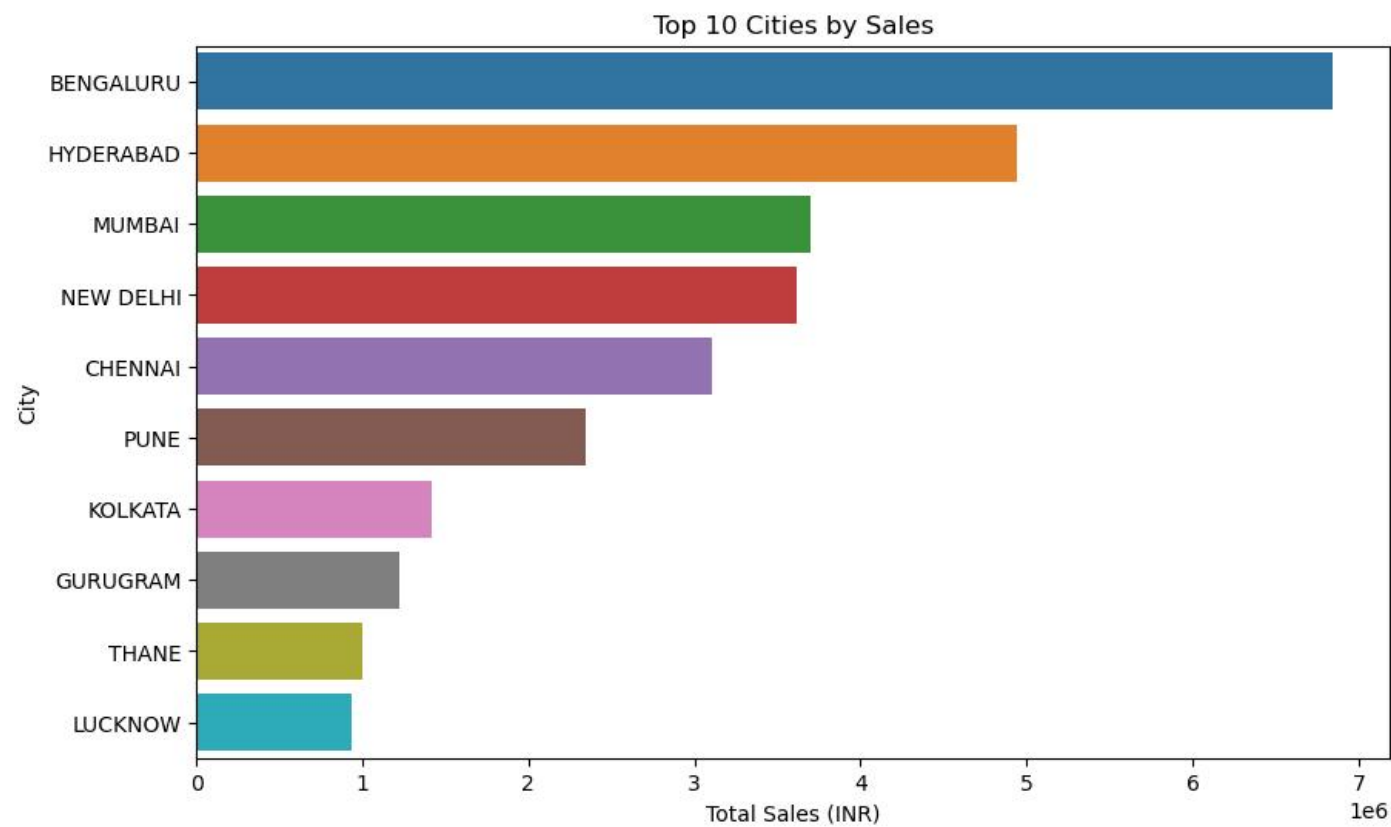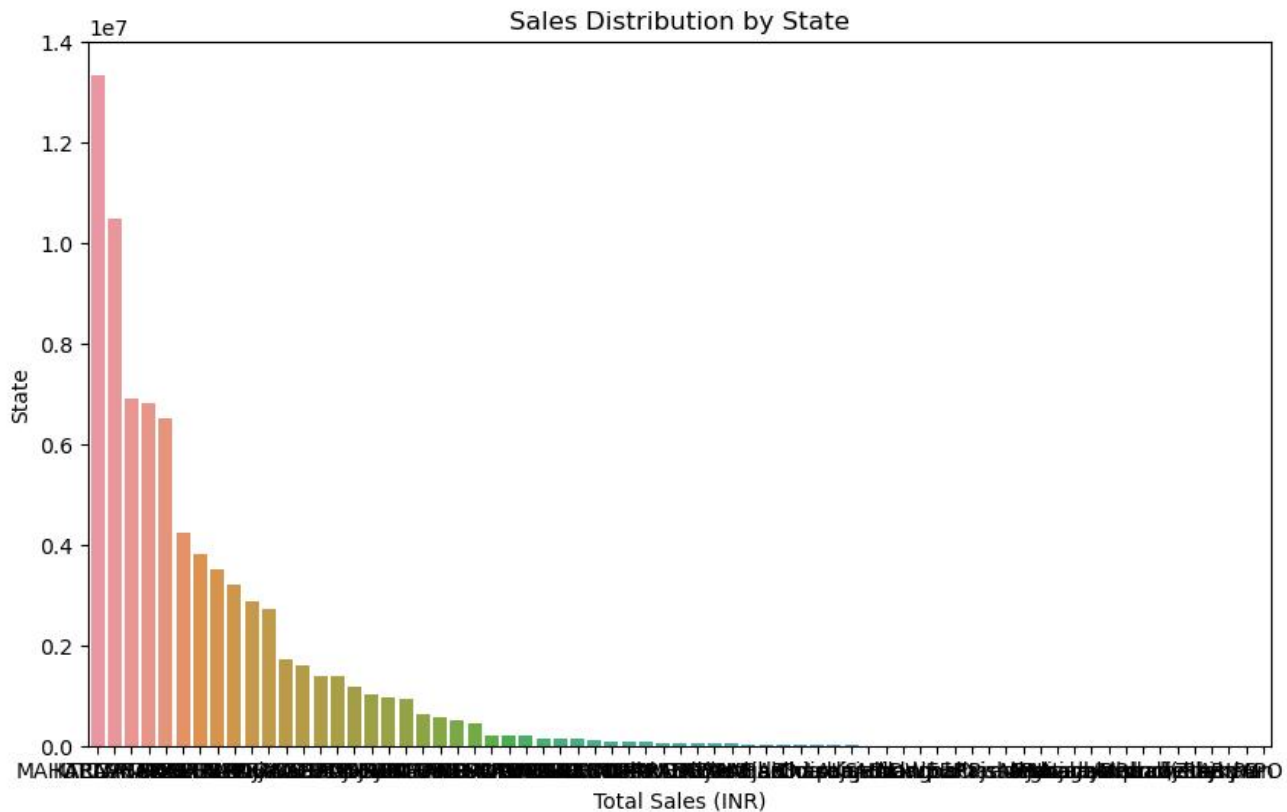
Key Summary

1. Sales Performance:
- **Trend:** Total sales have shown a steady increase over time, with significant spikes observed during promotional events.
- **Insight:** Promotional campaigns play a crucial role in driving consumer engagement and boosting sales.

2. Product Preferences:
- **Popular Categories:** T-shirts and Shirts are the most popular product categories, with XL and L sizes being the most frequently purchased.
- **Fulfillment Efficiency:** Expedited shipping methods have a higher on-time delivery rate compared to standard methods.

3. Customer Behavior:
- **Geographical Contribution:** Customers from metropolitan areas such as Mumbai and Chennai contribute significantly to total sales.
- **Loyalty Impact:** Repeat customers account for a significant portion of total sales, indicating strong customer loyalty.

4. Geographical Sales Distribution:
- **Top-selling Regions:** Maharashtra and Tamil Nadu are the top-selling states, with Mumbai and Chennai being the top-selling cities.
- **Strategic Focus:** Targeting marketing efforts towards high-potential regions can further boost sales and customer acquisition.

Recommendations:

Sales Strategies:

- Implement targeted marketing campaigns during peak sales periods to capitalize on increased consumer spending.
- Bundle complementary products or offer discounts to encourage upselling and increase average order value.

Inventory Management:

- Utilize predictive analytics to accurately forecast future demand and optimize inventory levels.
- Strengthen relationships with suppliers to ensure timely replenishment of inventory and negotiate favorable terms.

Customer Service:

- Implement proactive communication channels to keep customers informed about order status and shipping updates.
- Streamline the returns process with clear policies and procedures to provide a seamless experience for customers.

Conclusion

The analysis of Amazon sales data has provided valuable insights into various aspects of the business, including sales performance, product preferences, customer behavior, and geographical sales distribution. Through a comprehensive examination of the dataset and the application of various analytical techniques, several key findings have emerged, offering actionable recommendations for improving business operations and driving revenue growth.

Sales Performance and Strategies:
- The analysis revealed a steady increase in total sales over time, with notable spikes during promotional events. This underscores the effectiveness of targeted marketing campaigns in driving consumer engagement and boosting sales. To capitalize on this trend, Amazon should continue to invest in strategic marketing initiatives tailored to peak sales periods.

Product Preferences and Fulfillment Efficiency:
- T-shirts and Shirts emerged as the most popular product categories, with XL and L sizes being the most frequently purchased. Additionally, expedited shipping methods demonstrated higher on-time delivery rates compared to standard methods. Amazon can leverage this information to optimize product offerings and fulfillment processes, ensuring the availability of preferred products and enhancing delivery efficiency.

Customer Behavior and Geographical Sales Distribution:
- Customers from metropolitan areas such as Mumbai and Chennai significantly contribute to total sales, indicating the importance of targeting marketing efforts towards high-potential regions. Furthermore, repeat customers play a substantial role in driving sales, highlighting the significance of fostering customer loyalty through personalized experiences and effective post-purchase engagement strategies.

Recommendations:
Based on the insights derived from the analysis, several recommendations are proposed to further enhance Amazon's sales strategies, inventory management practices, and customer service initiatives:

1. **Sales Strategies:** Implement targeted marketing campaigns during peak sales periods, offering bundled products or discounts to encourage upselling and increase average order value.
2. **Inventory Management:** Utilize predictive analytics to accurately forecast demand and optimize inventory levels, strengthen supplier relationships to ensure timely replenishment, and implement inventory optimization techniques to minimize holding costs.
3. **Customer Service:** Implement proactive communication channels to keep customers informed about order status and streamline the returns process to provide a seamless experience.

In conclusion, the analysis of Amazon sales data has provided valuable insights and actionable recommendations for optimizing business operations and driving revenue growth. By implementing the

proposed strategies, Amazon can further strengthen its position in the e-commerce market and enhance overall customer satisfaction and loyalty.