**Amazon Sales Analysis**

**Problem Statement**

**Analyse and provide Insights on Amazon Sales.**

**Problem Description**

**The provided dataset contains information about sales transactions on Amazon, including details such as order ID, date, status, fulfilment method, sales channel, product category, size, quantity, amount, shipping details, and more. The objective is to conduct a comprehensive analysis of the data and extract actionable insights to support business decision-making.**

**Key Objectives:**

**1. Sales Overview:** Understand the overall sales performance, trends, and patterns over time.

**2. Product Analysis:** Analyse the distribution of product categories, sizes, and quantities sold to identify popular products.

**3. Fulfilment Analysis:** Investigate the fulfilment methods used and their effectiveness in delivering orders.

**4. Customer Segmentation:** Segment customers based on their buying behaviour, location, and other relevant factors.

**5. Geographical Analysis:** Explore the geographical distribution of sales, focusing on states and cities.

**6. Business Insights:** Provide actionable insights and recommendations based on the analysis to optimize sales strategies, improve customer satisfaction, and enhance overall business performance.

* **Analysis:**

Analysis of Amazon sales can be carried out in three Steps. These steps are essential to ensure accurate analysis of data which is an important aspect for correct insights. These Major steps are:

1. **Data Collection:**

The data is collected, explored, interpreted accurately. The data collection is then followed by ETL (Extract Transform Load), Extract refers as the collection and extraction of data from Dat source, it varies as per the requirements it may be organisational data, sample data etc. The extracted data contains incorrect data, faulty data, inconsistent data and many more that may lead to faulty and inaccurate data analysis, thus we make use of data Transformation. The transformed data is then used for analysis this is called as load.

1. **Exploratory Data Analysis:**

Exploratory data analysis step is performed after ETL. EDA is necessary to explore data and understand its dimensions, datatypes, relations between fields, distribution and more. EDA helps us to retrieve insights which is useful to understand the problem statement, the reason of its occurrence and its impact on the company. Additionally, it is very useful to draw the solutions to overcome the problems.

1. **Insights:**

Insights can be considered as the conclusion which is drawn through Exploratory data Analysis. Insights help to identify the reason and the impact of the problem. The Insights then help to find out the solution, in some cases it is also very useful to predict the future outcome thus playing a huge role in Data Analysis.

With the help pf these three steps, Amazon Data analysis is performed.

1. Data Collection:
   * Importing required Libraries:

Libraries used for Analysis is imported for better computations, data manipulation, statistical computations and Data Visualizations.

* + Understanding the distribution of data:

1. **Qty:** The data indicates that the majority of the Qty column values are concentrated around 1, proved by the fact that 75% of the values are exactly 1. This tight clustering around 1 is further supported by the small standard deviation (0.313), which indicates minimal variability in the dataset. The presence of a maximum value of 15 suggests a few instances of significantly larger quantities.
2. **Amount:** The differences in the central tendencies highlight a higher spread of values in the Amount column with a mean value of 645 and standard deviation of 281. The noticeable difference between the mean (648.56) and the median (605) suggests a dynamic and skewed distribution of values.
   * Data Cleaning and Manipulation:

Data and its types are explored and manipulated as per the requirements.

**Drop null columns:**

**Null columns which do not contain any data is dropped.**

**Drop the rows having zero quantity that represents faulty and incorrect data:**

**Removing the rows having zero quantity since the orders that have no quantity are faulty entries.**

**Standardizing the different varied forms of States and cities names into a consistent format across all rows:**

**Each states names are converted into uppercase to prevent case sensitiveness. The names containing any characters and symbols other than alphabets are removed thus, making it more accurate.**

**Amount specified with cancelled orders are excluded:**

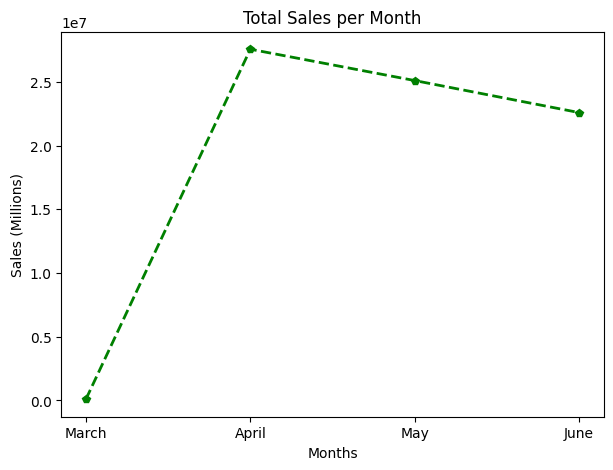
**Amount of cancelled orders are excluded to avoid inaccurate calculation of Revenue, since amounts of cancelled orders are refunded.**

1. Exploratory Data Analysis:

### ****Sales Overview:**** Understanding the overall sales performance, trends, and patterns over time

### Sales performance over Months.

From the chart below, we can observe that sales started with little to no activity on March, 2022. There was an impressive surge in sales leading up to April. However, following this peak in April, sales began to decline steadily, decreasing by approximately 1 million each month.



### Sales by Channel

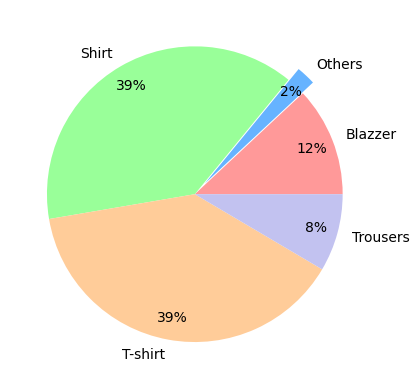
The table below concludes that the Amazon sales channel was dominant, with no orders recorded from any other channels.

|  |  |  |
| --- | --- | --- |
| Index | Sales Channel | Amount |
| 1. | Amazon | **75381835.0** |
| 2. | Non-Amazon | 0,0 |

### ****Product Analysis:**** Analysing the distribution of product categories, sizes, and quantities sold to identify popular products.

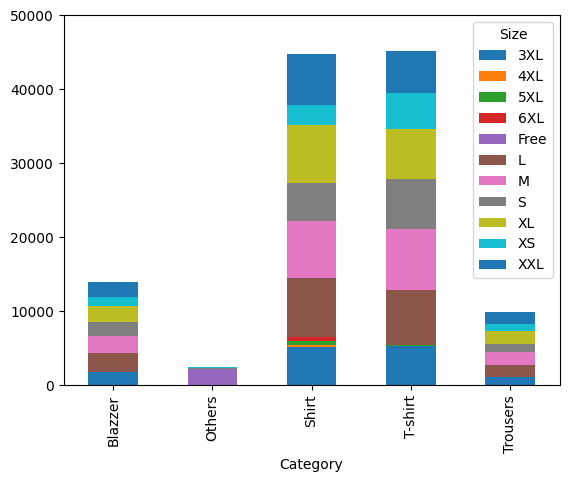
### Category Distribution

Pie Chart below shows the proportion of sales for each product category. The pie chart tells us that Tshirt and shirt yields the maximum revenue than other category products.



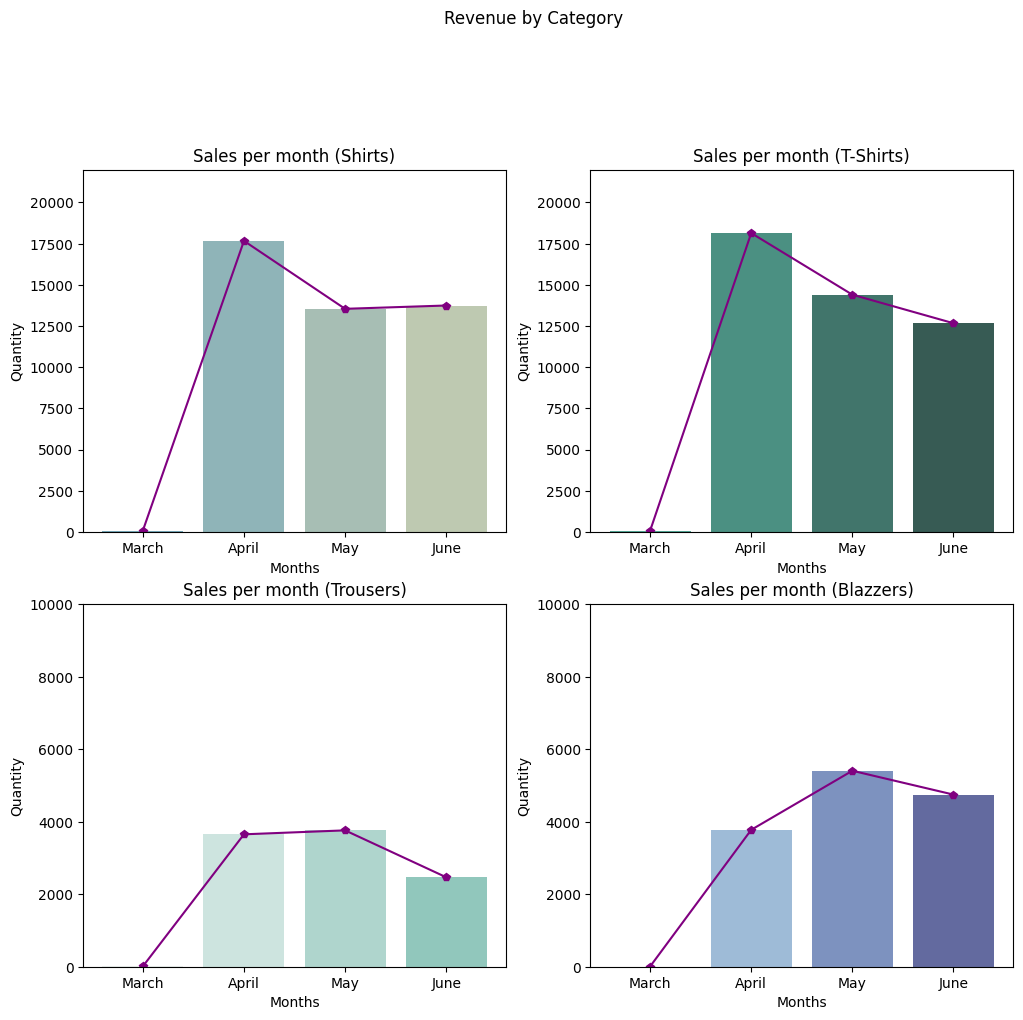
### Size Distribution

The stacked chart below displays the sales of various sizes across different product categories. It highlights that T-shirts in size M have the highest number of orders, followed by shirts, blazers, and other items.



### Revenue by Category

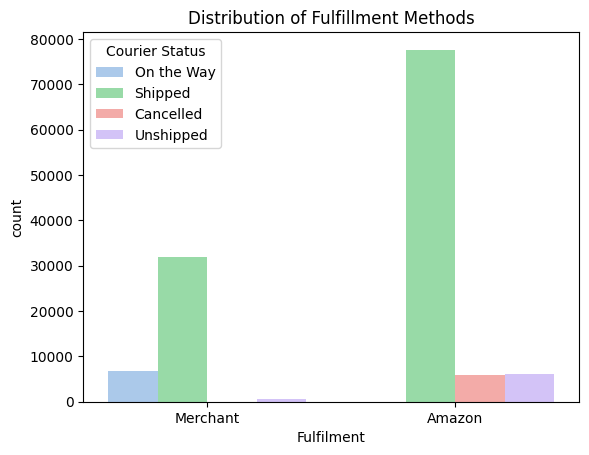
The bar chart below displays the sales of various categories over months. It helps us to identify the increase and decrease in the sales of each category.



### ****Fulfilment Analysis:**** Investigating the fulfilment methods used and their effectiveness in delivering orders.

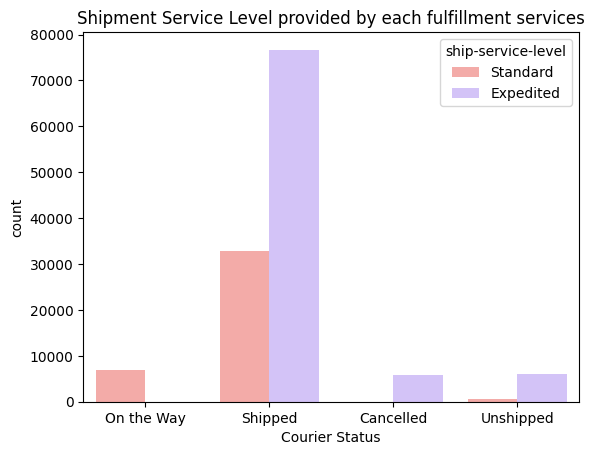
### Distribution of Fulfilment Methods

The chart below illustrates the distribution of services provided by each fulfilment agencies. According to the chart below, services provided by merchant shows minimal or no cancelled orders, with the majority of orders being shipped and some currently in transit. The data also indicates that Amazon fulfilment service dominates, although there are instances of unshipped and cancelled orders handled by Amazon.



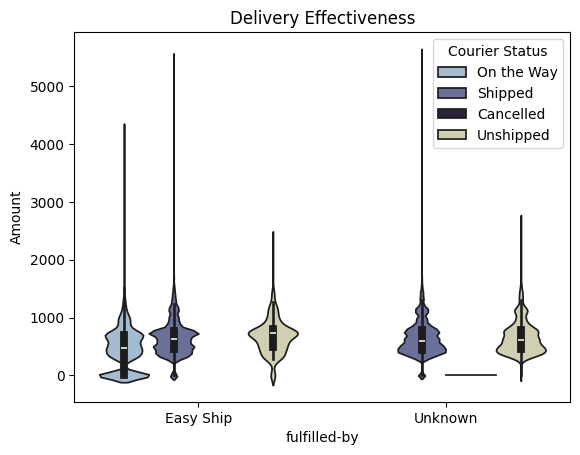
### Service Level

The chart clearly illustrates the difference in shipping service levels provided by each fulfillment agency. It shows that Expedited Shipping had significantly more orders shipped compared to Standard Shipping. Additionally, it's evident that a larger number of customers preferred Expedited shipping over Standard shipping.



### Delivery Effectiveness

**The chart below illustrates the delivery effectiveness of each fulfilment services. The violin plot helps also helps us to tell us the most preferred fulfilment service and the courier statuses of each service.**



### Time Taken to deliver the Orders

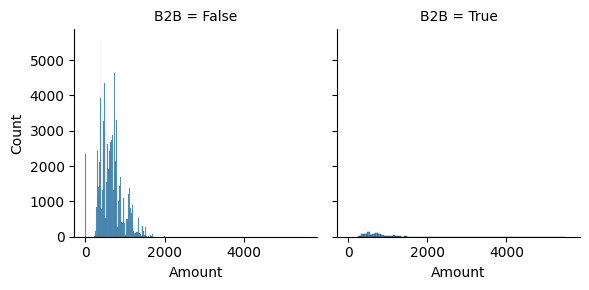
The table indicates that no courier statuses have been updated, resulting in zero records.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Order\_Id | | Date\_ship | | Date\_deliver | | Days\_to\_deliver | |
|  |  | |  | |  | |  |
|  |  | |  | |  | |  |

### ****Customer Segmentation:**** Segment customers based on their buying behaviour, location, and other relevant factors.

### Buying Behaviour

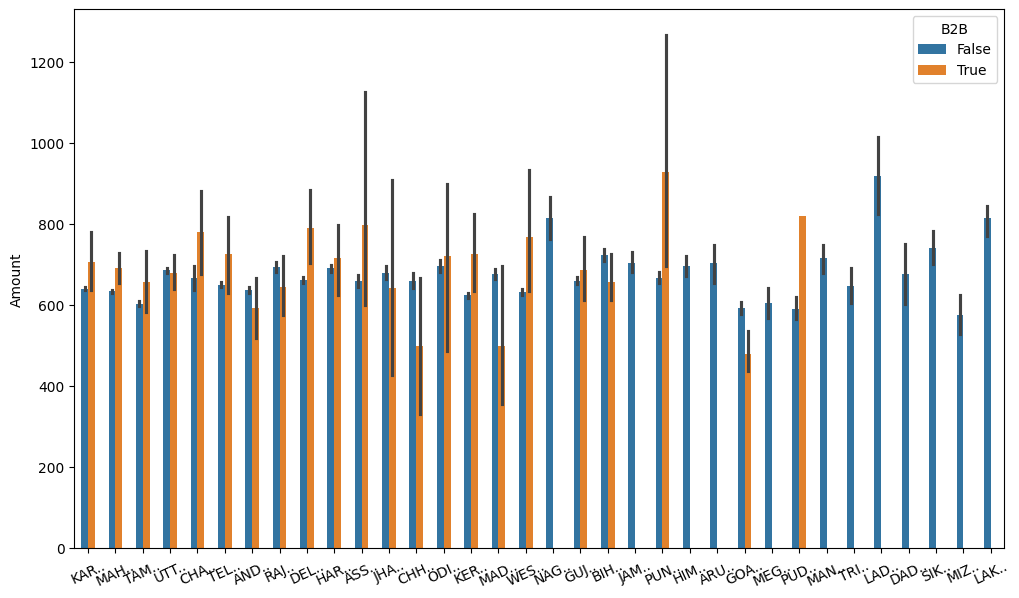
The histogram below reveals that most transactions are conducted as business-to-consumer rather than business-to-business. This indicates that there are minimal or no business-to-business deals.



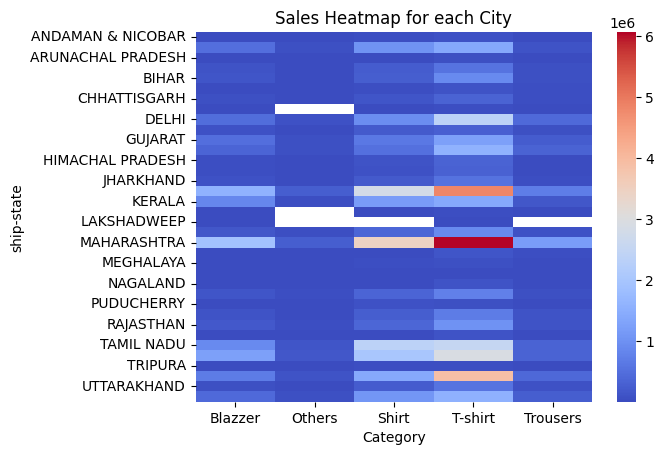
### ****Geographical analysis:**** Explore the geographical distribution of sales, focusing on states and cities.

### Sales by region

The plot below illustrates the types of transactions for each state, showing that Punjab has the highest B2B sales, while Ladakh has the highest B2C sales.

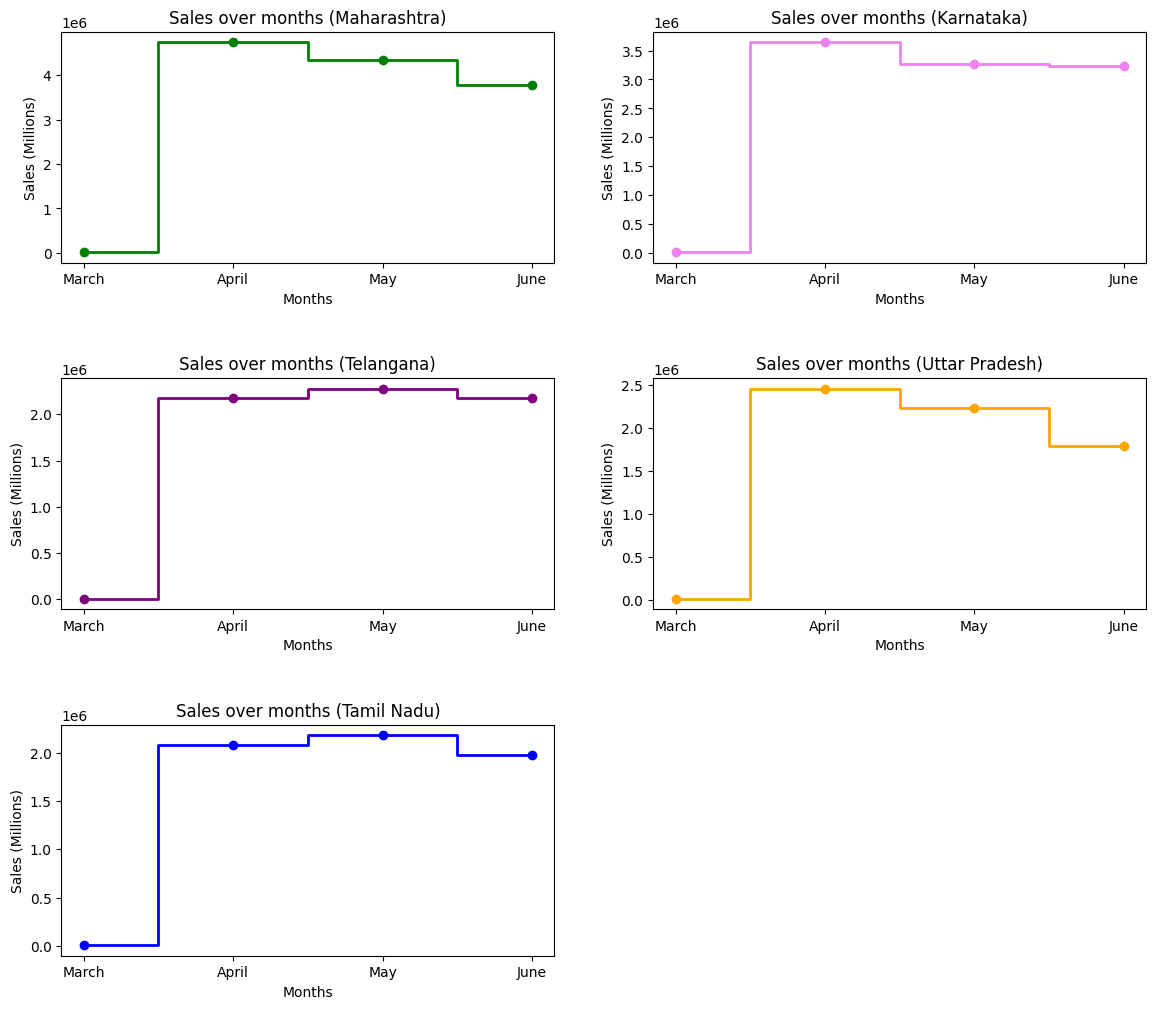


### Sales of each category at each State



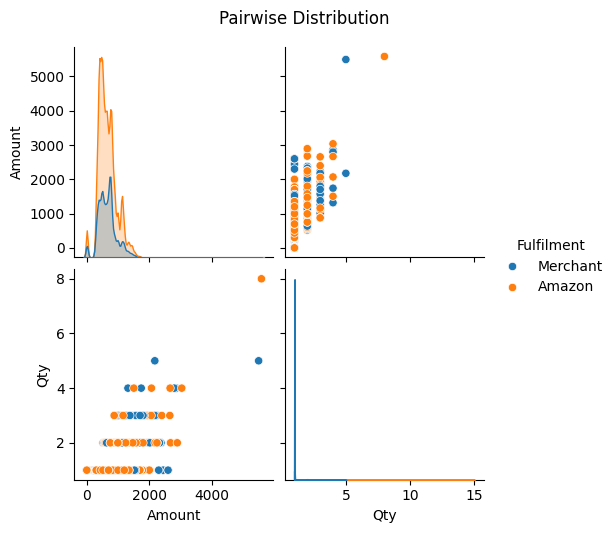
### Top 5 Sales Region (City)

The table below portrays that Bengaluru has the highest number of Amazon customers, followed by Hyderabad, Mumbai, and other cities. Similarly, Maharashtra is the state with the most Amazon customers, followed by Karnataka, Telangana, and others.



### Pairwise Distribution of fulfilment services

The distribution below illustrates the pairwise distribution of data for Amount and Quantity.



# Insights:

**From the above Analysis:**

1. Sales initially showed minimal activity but experienced a significant increase until April. After this peak, sales began to decline steadily by approximately 1 million per month.
2. The Amazon sales channel is dominant, with no orders recorded from other channels.
3. T-shirts, especially in size M, have the highest number of orders and generate the most revenue, followed by shirts, blazers, and trousers.
4. The dominance of T-shirts in terms of popularity and revenue remains consistent across different regions.
5. Expedited shipping is significantly more popular than standard shipping, indicating a customer preference for faster delivery options.
6. Amazon's fulfilment service is the most prominent, with most orders being shipped or on the way. However, there are still some unshipped or cancelled orders within the Amazon service.
7. Bengaluru has the highest number of Amazon customers, followed by Hyderabad and Mumbai.
8. Maharashtra has the most Amazon customers at the state level, followed by Karnataka and Telangana.
9. Punjab leads in B2B sales, while Ladakh has the highest B2C sales.
10. The majority of transactions are business-to-consumer, with minimal business-to-business deals.

# Recommendations:

1. Given that T-shirts, especially in size M and L, are highly popular, ensure adequate stock levels to meet demand. Consider expanding the T-shirt product line with more designs, colors, and sizes.
2. Focus on increasing the inventory and variety of other popular items like shirts, blazers, and trousers for increase sales.
3. Since Amazon is the dominant sales channel, continue to strengthen this channel. Explore promotional strategies and exclusive deals to maintain and grow this channel.
4. While Amazon is dominant, investigate potential in other sales channels to reduce dependency and tap into new customer bases.
5. Given the preference for expedited shipping, consider offering more incentives for this service, such as discounts or bundled deals whereas for standard shipping enhance the attractiveness of standard shipping by offering free shipping for orders above a certain value or loyalty rewards.
6. Investigating and resolving the reasons behind unshipped and cancelled orders within the Amazon service, improving logistics and customer communication to reduce these occurrences can be very helpful to prevent the cancelling of orders.
7. Enhance the efforts and marketing strategies to increase sales on regions having minimal sales.
8. Leverage the strong B2B sales in Punjab by offering tailored business packages and bulk purchase discounts. Expand B2B marketing efforts in regions with potential for growth. Additionally continue to support and enhance B2C sales in regions like Ladakh with targeted advertising and promotions.
9. Regularly gather customer feedback to understand preferences and pain points. Use this data to adjust product offerings, improve service levels, and enhance the overall customer experience.