Exploratory Data Analysis (EDA) with Pandas in Amazon Sales Report

This project analyzes the **Amazon Sales Report data set** to identify key trends in sales, shipments, and order fulfillment. The goal is to provide **insights into product demand, shipping efficiency, and fulfillment methods.**

**Goals of the Project :**

* **Explore the Amazon Sales data set using pandas.**
* **Perform feature engineering to derive useful insights.**
* **Visualize data distributions and trends with various plot types.**
* **Summarize key findings that can aid in business decision making.**

**Materials & Methods :**

The Amazon Sales data set includes various columns related to order details, shipping, fulfillment, and product information. It also covers product categories along with their available sizes, shipping state details, courier status, B2B (Business-to-Business) sales, and order fulfillment.

General Part :

* **Libraries Import :** Pandas, Numpy, Seaborn, Matplotlib
* **Data set Exploration :** Initial exploration of the data set, checking for missing values, duplicates and generating summary statistics
* **Feature Engineering :** Transformation of columns and creation of new features like shipping and order status
* **Visualization in Pandas :** Distribution analysis, B2B sales breakdowns

Project outcome & Insights :

The project perform Exploratory Data Analysis(EDA) on Amazon Sales Report data set to get meaningful insights into Sales & Product Demand Analysis, Shipping & State-Wise Distribution, Fulfillment & Courier Status Analysis, B2B Sales Breakdown.

Below are the key outcomes :

1. **Sales & Product Demand Analysis :**

* The “**top-selling product categories**” were identified based on the quantity (**`Qty`**) of products sold.
* The **“most frequently purchased size”** was analyzed using a “count plot”.
* A **“histogram of product sizes”** was created to observe the distribution of different sizes in the data set

.

1. Shipping & State-Wise Distribution :

* The **“Top 10 States”** with the highest number of shipments were identified using a **“bar chart and pie chart”**.
* The **“most frequent shipping state”** was highlighted in red to emphasize high-volume regions.
* The data set indicates that some states contribute significantly more to overall sales volume than others.

1. Fulfillment & Courier Status Analysis :

* A **“pie chart”** was created to analyze different fulfillment methods (`Fulfillment` column).
* A **“count plot”** of `Courier Status` against `Status` was generated to assess shipping efficiency and delays.
* The fulfillment distribution showed a “dominance of specific fulfillment methods”, indicating preferred logistics.

1. B2B Sales Breakdown :

* The **`B2B`** column was analyzed using a “**pie chart”**, showing the percentage of orders that were Business-to-Business (B2B) v/s direct-to-consumer.
* This helps understand whether Amazon's business customers significantly impact sales volume.

Feature Engineering :

* Ship-state column’s state names converted into uppercase
* Dropping null values columns from data set (New and Pending columns)
* Order Month extracted from Date
* Renaming Qty column name to Quantity

Key Questions and Insights to be Addressed :

* **Q1. Which are the top-10 shipping states?**

top\_10\_state = df['ship-state'].value\_counts().head(10)

print(top\_10\_state)

MAHARASHTRA 6236  
KARNATAKA 4550  
UTTAR PRADESH 3298  
TAMIL NADU 3167  
TELANGANA 3136  
KERALA 2213  
DELHI 1955  
WEST BENGAL 1653  
ANDHRA PRADESH 1621  
GUJRAT 1382

* **Q2. What is the total revenue generated by each product category?**

df['Total\_Sales'] = df['Amount'] \* df['Quantity']  
category\_sales = df.groupby('Category')['Total\_Sales'].sum().reset\_index()  
category\_sales = category\_sales.sort\_values(by='Total\_Sales', ascending=False)  
  
print(category\_sales.head())

Category Total\_Sales  
5 T-shirt 10103409.0  
2 Shirt 5359763.0  
0 Blazzer 4405510.0  
6 Trousers 1083032.0  
1 Perfume 179404.0

Q3. Which state has the highest number of orders?

state\_orders = df['ship-state'].value\_counts().reset\_index()  
  
*# Rename columns*state\_orders.columns = ['State', 'Order\_Count']  
  
print(state\_orders.head())

State Order\_Count  
0 MAHARASHTRA 6236  
1 KARNATAKA 4550  
2 UTTAR PRADESH 3298  
3 TAMIL NADU 3167  
4 TELANGANA 3136

Q4. Which courier service has the most deliveries?

courier\_counts = df['Fulfilment'].value\_counts().reset\_index()  
  
courier\_counts.columns = ['Fulfilment', 'Number of Deliveries']  
  
print(courier\_counts.head())

Fulfilment Number of Deliveries  
0 Merchant 37514

Q5. What is the most common product size ordered?

size\_counts = df['Size'].value\_counts().reset\_index()  
  
size\_counts.columns = ['Size', 'Quantity']  
  
print(size\_counts.head())

Size Quantity  
0 M 6806

1 L 6646  
2 XL 6326  
3 XXL 5090  
4 S 4558

Q6.Which month has the highest sale?

df['Order Month'] = df['Date'].dt.month  
monthly\_sales = df.groupby('Order Month')['Total\_Sales'].sum().reset\_index()  
monthly\_sales = monthly\_sales.sort\_values(by='Total\_Sales', ascending=False)  
  
print(monthly\_sales)

Order Month Total\_Sales  
1 4 8117420.0  
2 5 7494038.0  
3 6 5682811.0

0 3 10797.0

Q7.Which size is selling the fastest in terms of quantity?

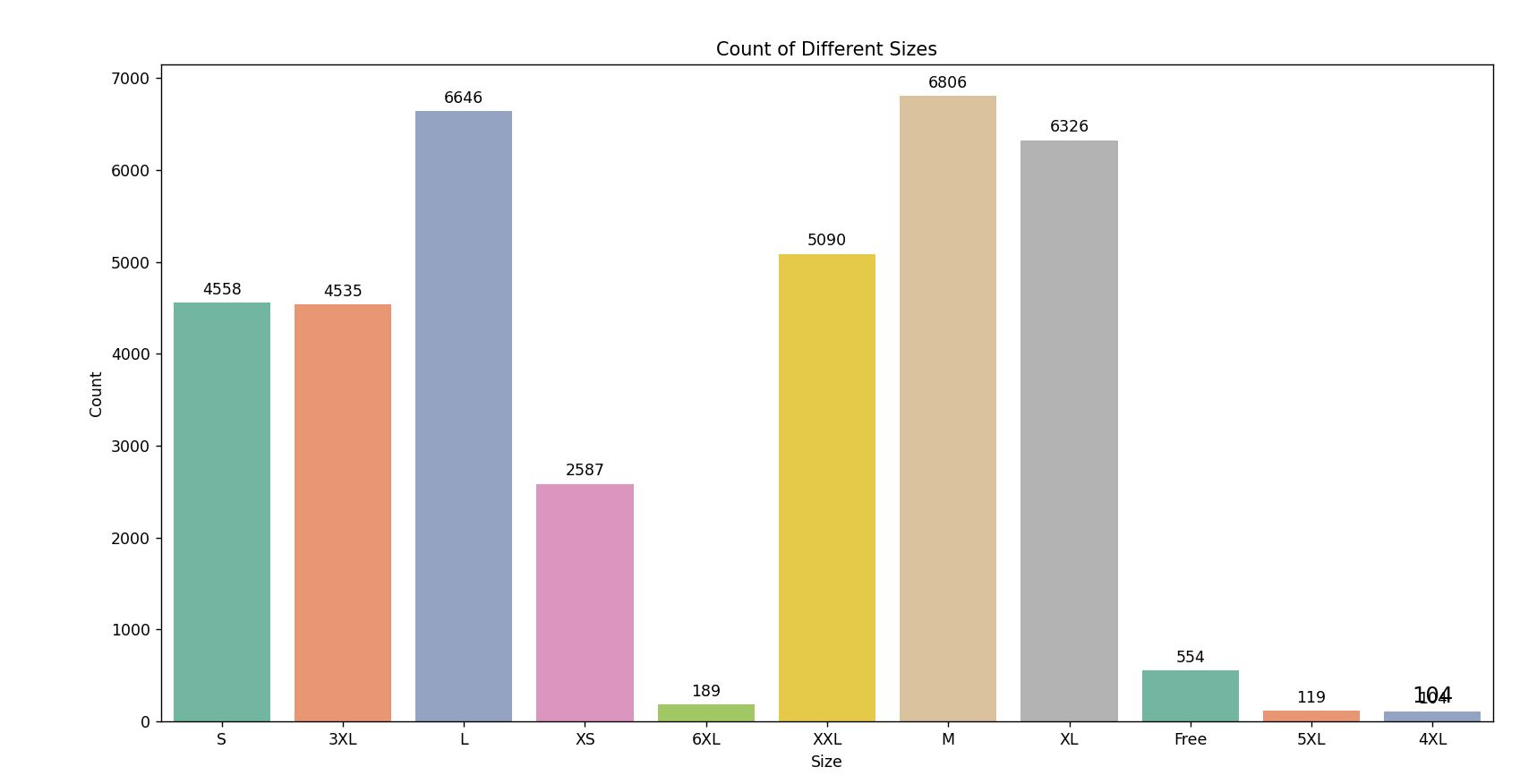
size\_sales = df.groupby('Size')['Quantity'].sum().reset\_index()  
size\_sales = size\_sales.sort\_values(by='Quantity', ascending=False)  
  
print(size\_sales.head())

Size Quantity  
6 M 5905  
5 L 5795  
8 XL 5481  
10 XXL 4465  
0 3XL 3972

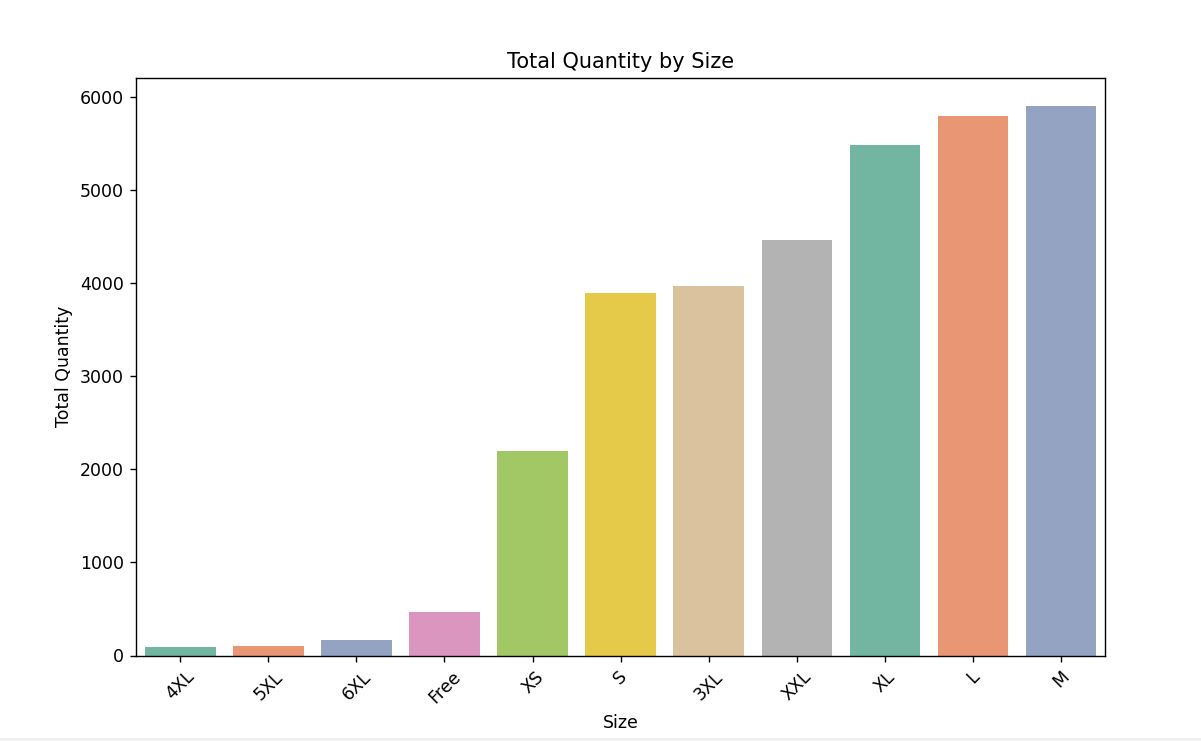
Visualization :

Several charts created to present inside including :

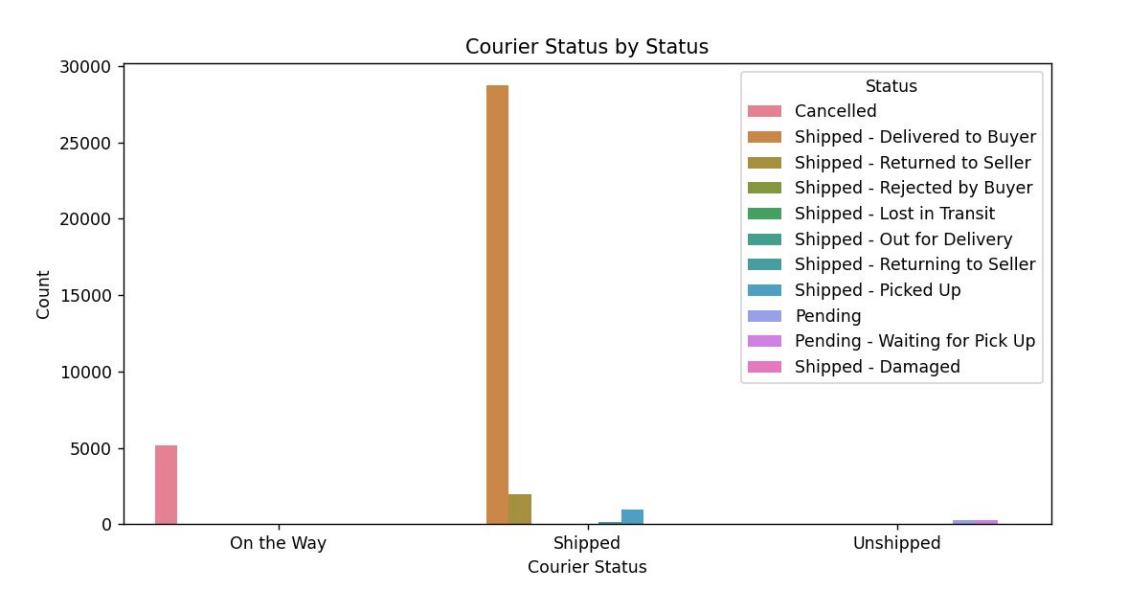
* **Count of Different sizes**



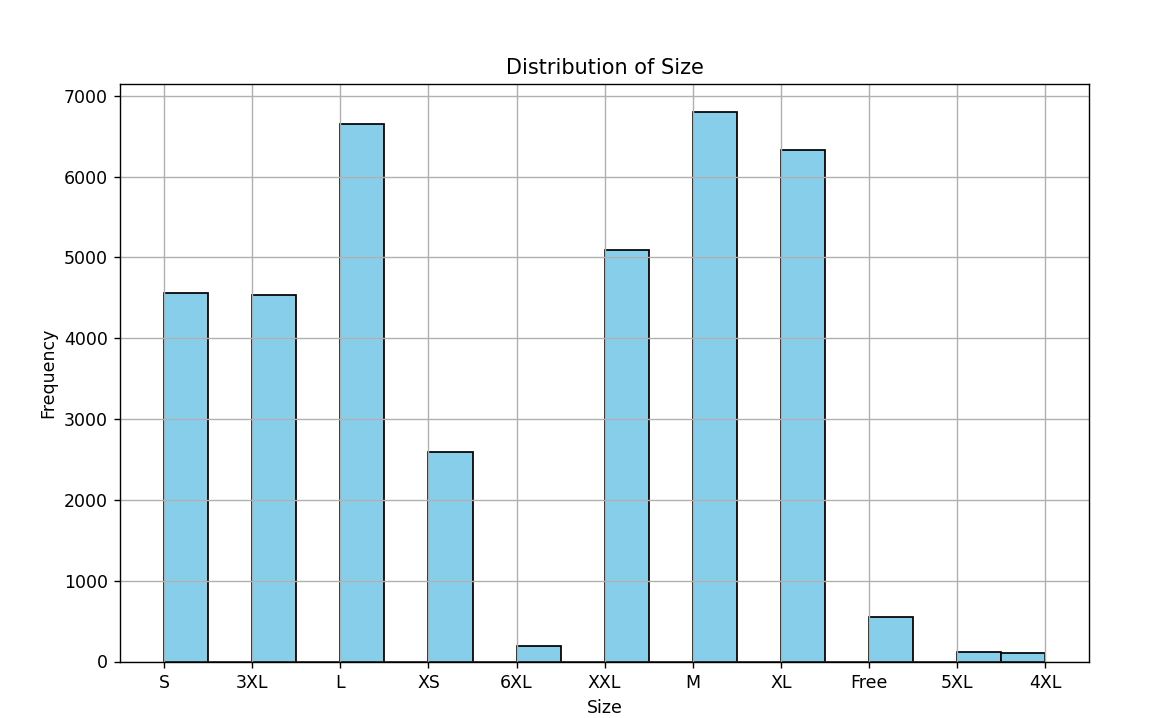
* **Total Quantity by size**



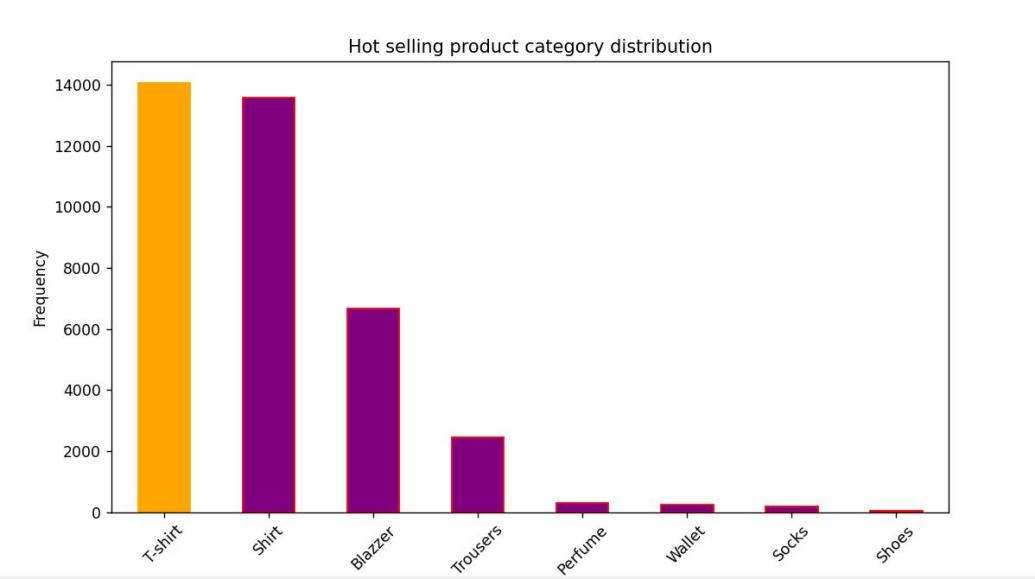
* **Courier status by status**



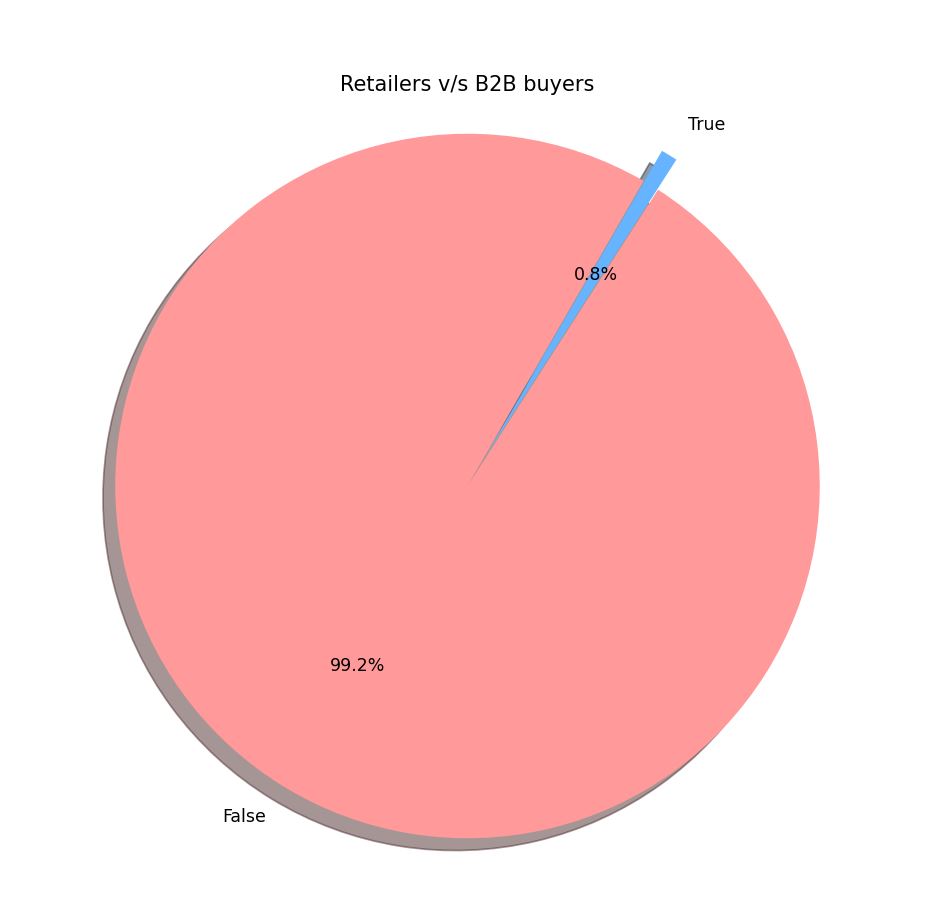
* **Distribution of size**



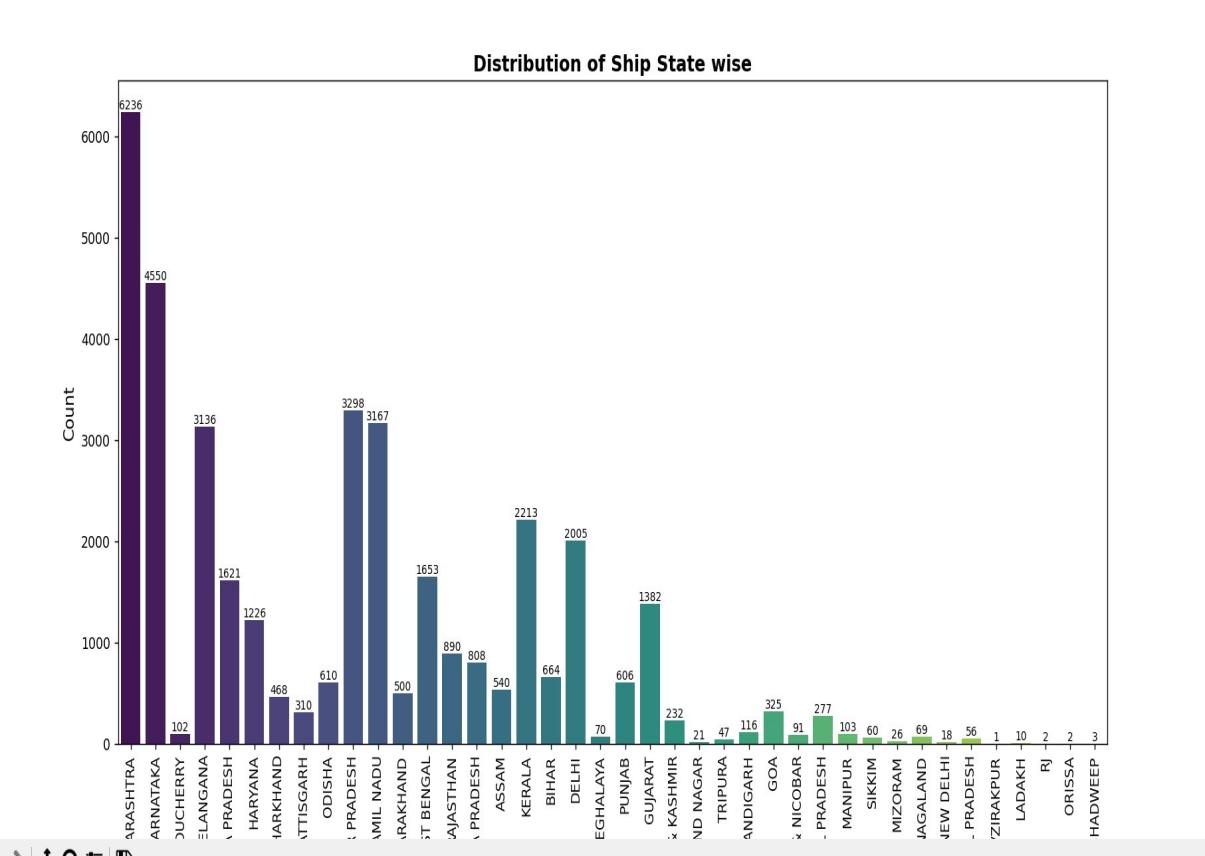
* **Hot selling product category distribution**



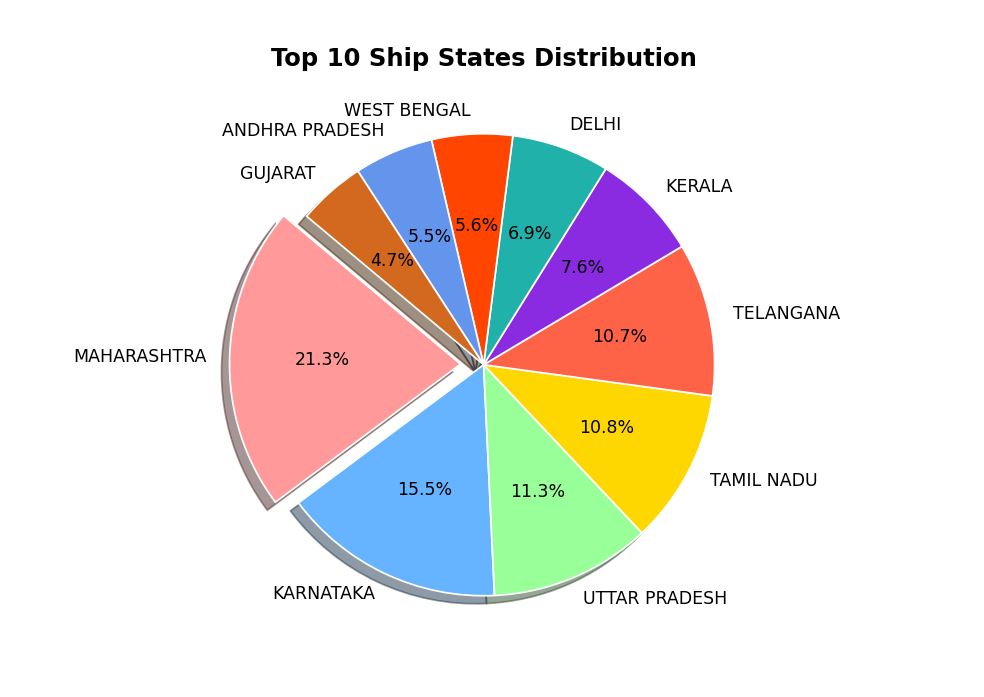
* **Retailers v/s B2B buyers**



* State wise shipping distribution



* Top 10 states Distribution



Conclusion :

The data analysis reveals that the business has a significant customer base in **Maharashtra state**, mainly serves retailers, fulfills orders through Amazon, experiences high demand for **T-shirts** and sees **M-Size** as the preferred choice among buyers.