**Data Analysis Project**

**1.Title**

* Project Title: Amazon Sales Analysis
* Name: Ananya
* Date:

**2.Introduction**

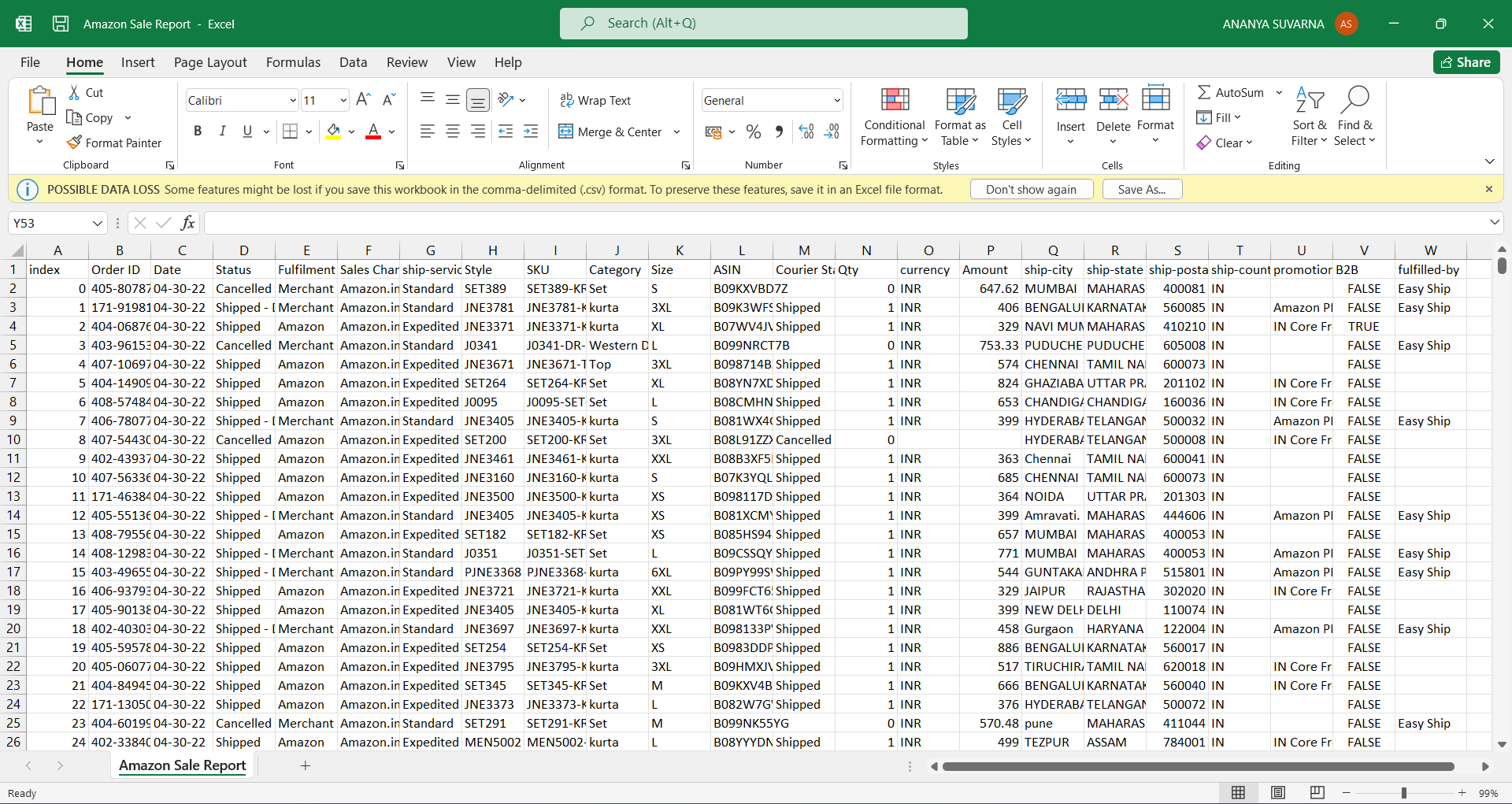
* This project focuses on analysing Amazon's sales data to gain insights into sales performance, order fulfilment, customer behaviour, and product trends. By examining patterns in revenue, top-performing product categories, and seasonal sales variations, this analysis aims to provide actionable insights that can help improve strategic decision-making for inventory, marketing, and customer engagement.

**3. Aim**

* The aim of this project is to analyse Amazon’s sales data to identify the key factors that impact the revenue growth, product category performance, and customer purchase patterns.

**4.Data Description**

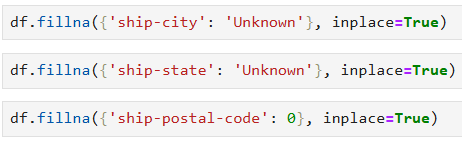
* **Source of Data:** This dataset was sourced from Amazon’s publicly available sales data on Kaggle.
* **Dataset Size:** The dataset contains around 1,28,000 rows and 24 columns.
* **Data Types:** This dataset has Numerical values such as index, quantity, amount, ship-postal-code, Categorical values such as sales channel, SKU, size, category, ship-city, ship-state etc and Boolean values for B2B.
* **Key Variables (Columns):**
  + - Order ID: A unique identifier for each transaction.
    - Date: Date of purchase.
    - Fulfilment: Indicates the method used to fulfil the order (e.g., Fulfilled by Amazon, Fulfilled by Merchant).
    - SKU: Stock Keeping Unit is a unique identifier used by retailers to track inventory.
    - Category: The product category to which the purchased item belongs.
    - Qty: The number of units sold for each order.
    - Amount: Total revenue generated for each order.
    - Ship-city, ship-state: Location where the order was shipped.



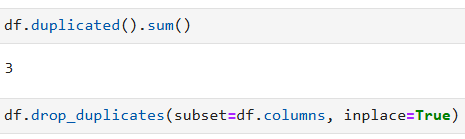
**5.Preprocessing**

* Handling missing values: Dropped all null values in Amount column, filled missing values in ship-city and ship-state as unknown and ship-postal-code as 0.



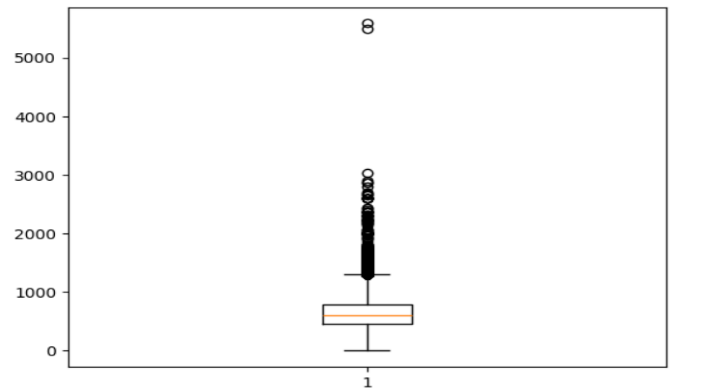


* Removing Duplicates: Duplicated order entries were removed.

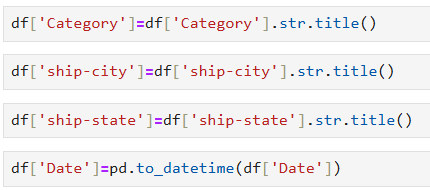


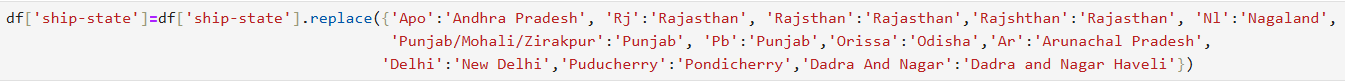
* Handling Outliers: Here outliers reflect the real business sales. Hence not removed.



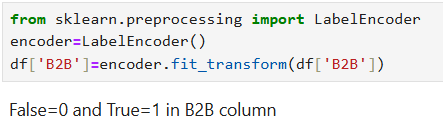


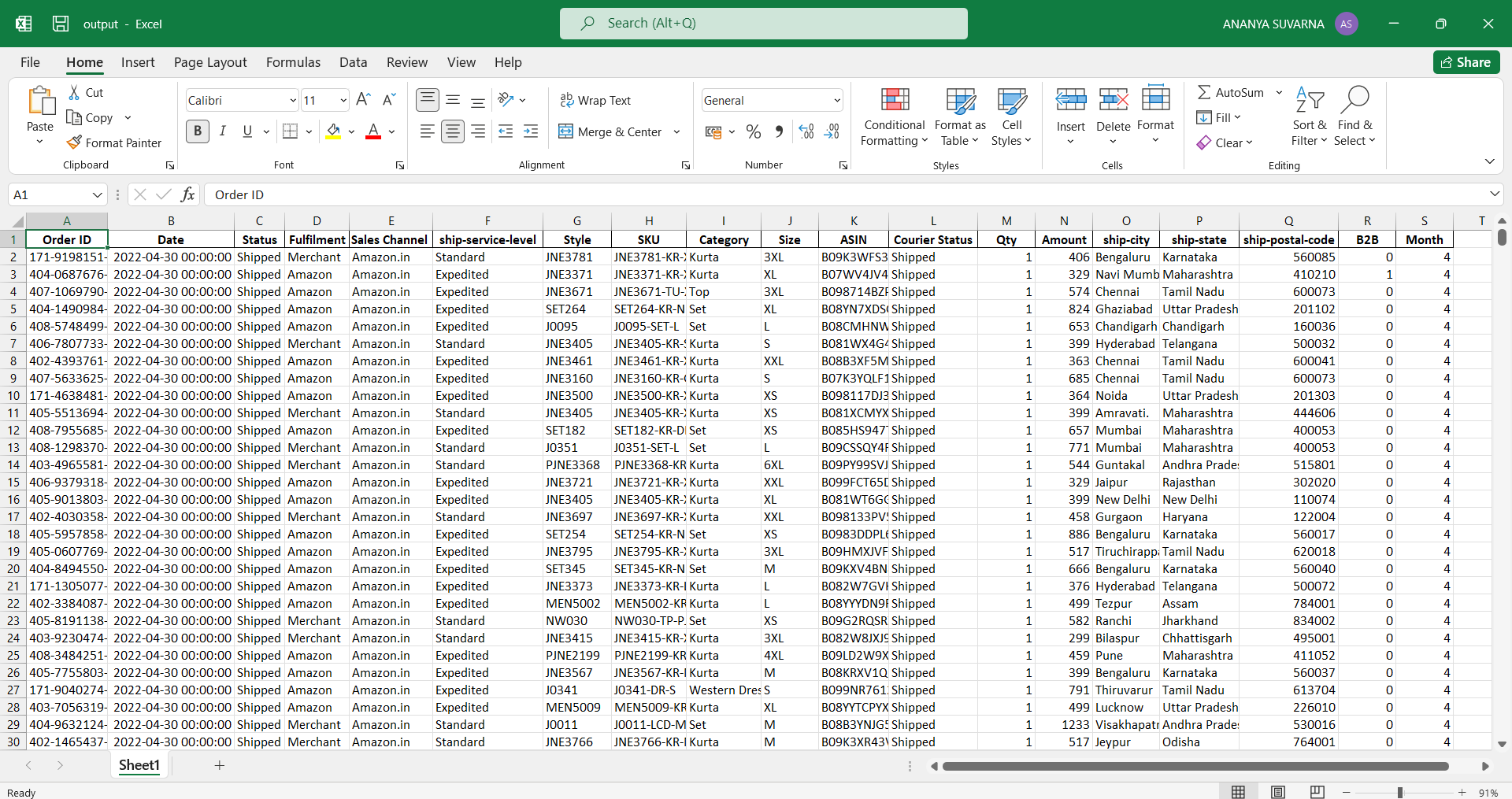
* Handling Inconsistencies: Fixed the inconsistent cases, and shipping states were standardized by converting all values into their full state names.





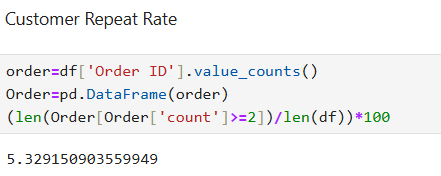
* Encoding Categorical Variable: B2B column encoded using label encoding.



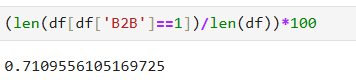


**6.Data Analysis**

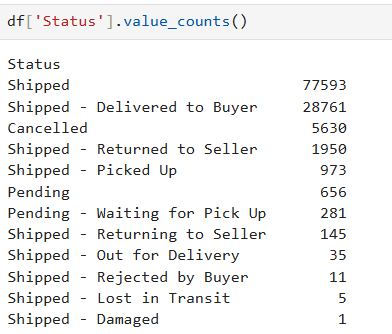
* A customer repeat rate of only 5%.



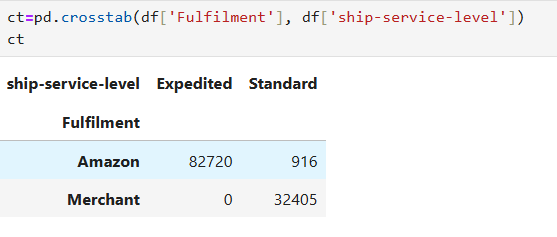
* Only about 0.7% sales are B2B, remaining are B2C



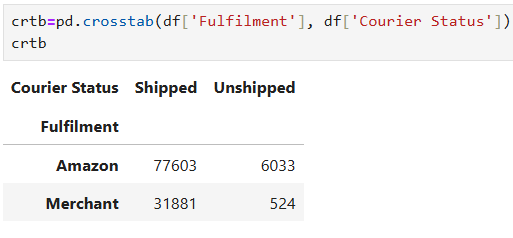
* Majority of the orders are being processed and shipped. There is relatively high number of orders cancelled which could be due to supply chain disruption or delay in shipping.



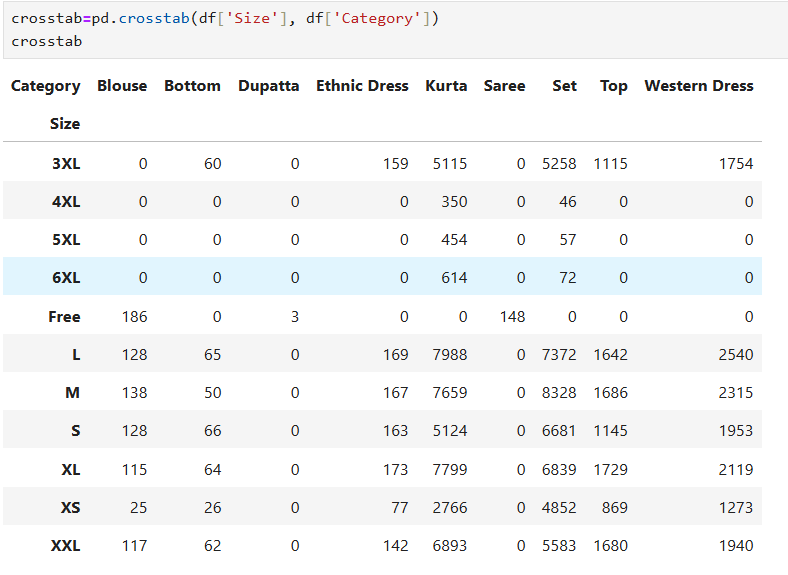
* Amazon focus mainly on expedited shipping to deliver products faster and improve customer satisfaction. While orders fulfilled by merchant rely solely on standard shipping.



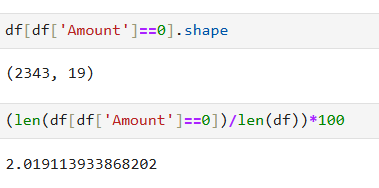
* About 92% of orders fulfilled by Amazon are shipped, where as 98% orders fulfilled by Merchant are shipped.



* This provided the category wise sales in size.



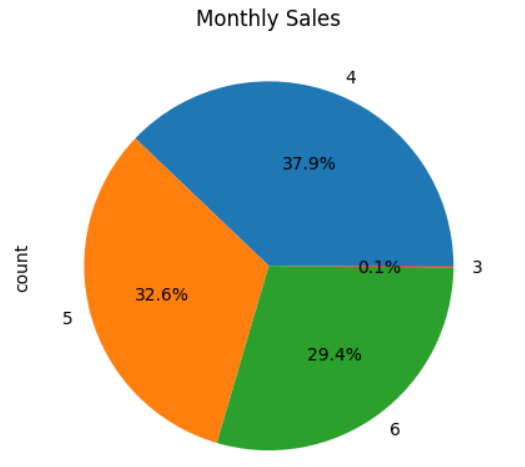
* About 2% of Transaction have an amount of zero (Promotional giveaways, coupon codes).



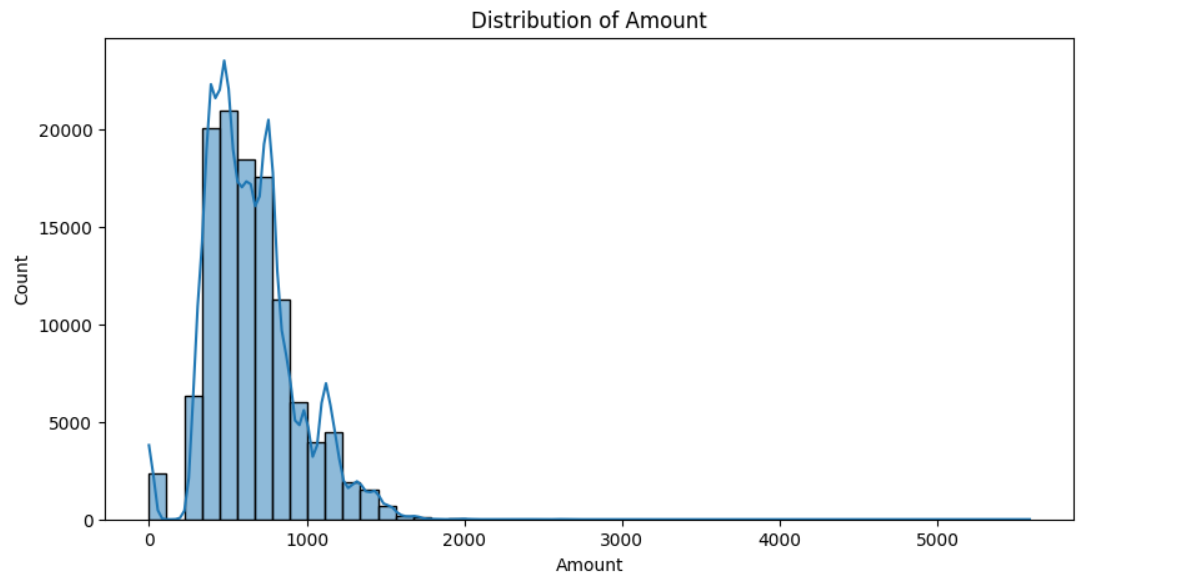
**7.Exploratory Data Analysis**

* **Univariate Analysis**

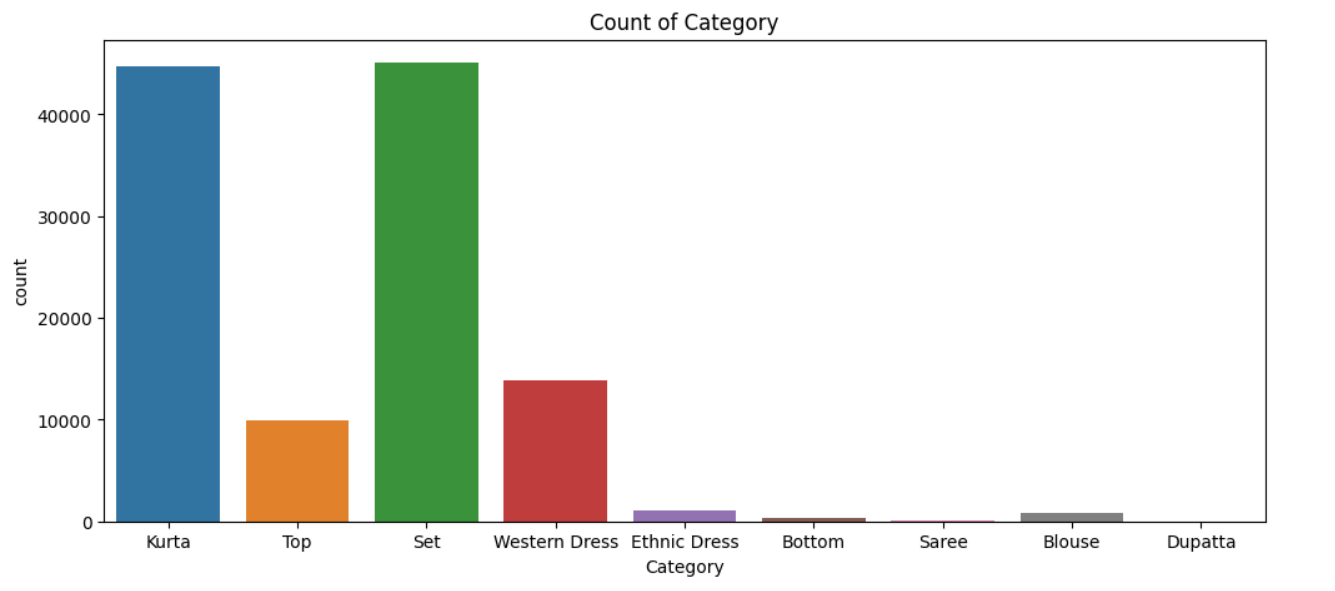
1. Maximum orders are placed in the month of April.



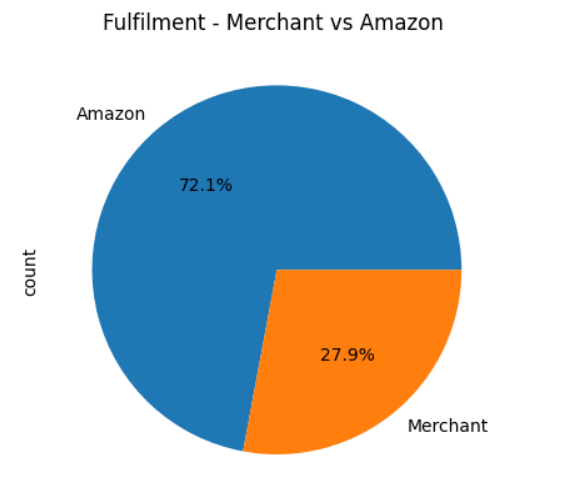
1. Distribution of the amount is skewed. Many transactions are under 1000.



1. Most of the orders are placed in the category of Kurta and Set.

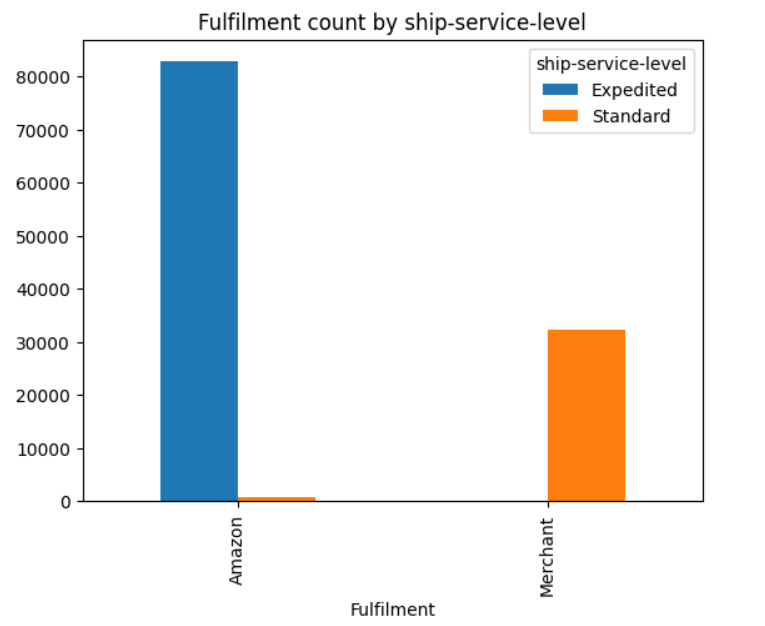


4. Most of the orders are fulfilled by Amazon.

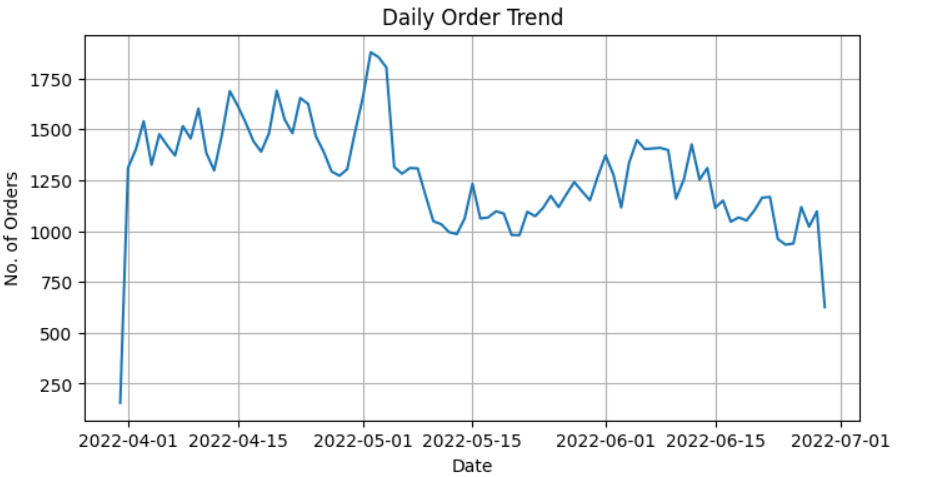


* **Bivariate Analysis**

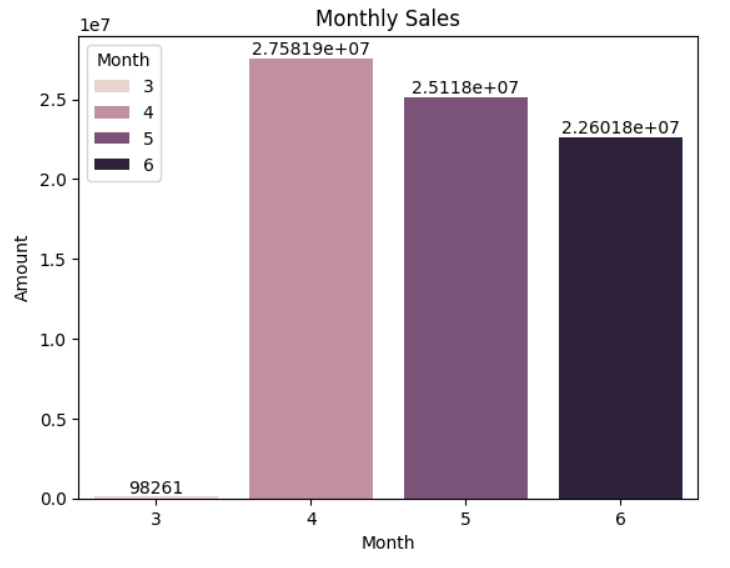
1. A significant number of customers prefer expedited shipping for their orders fulfilled by Amazon.



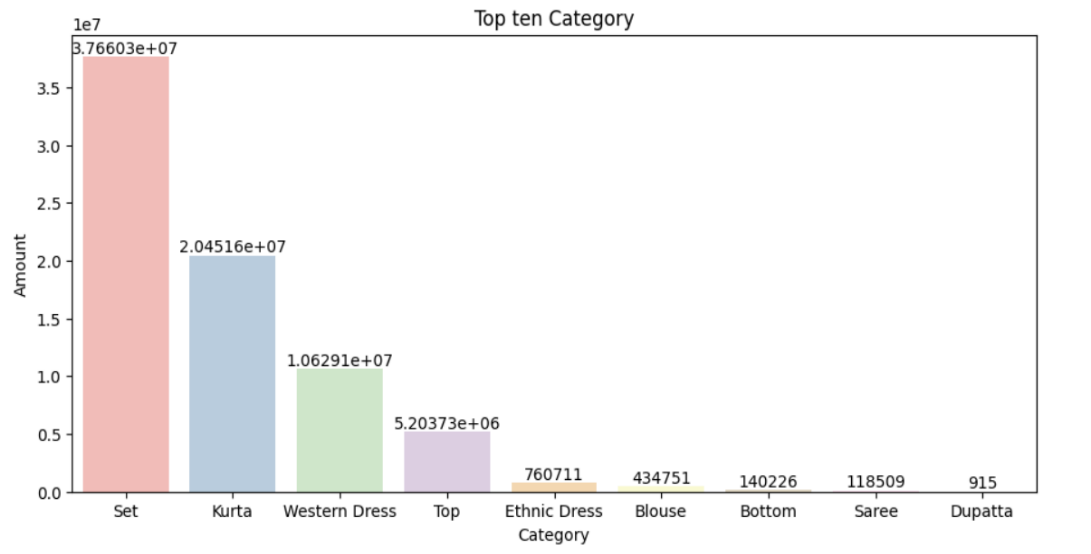
1. Maximum orders are placed in the 1st week of May.



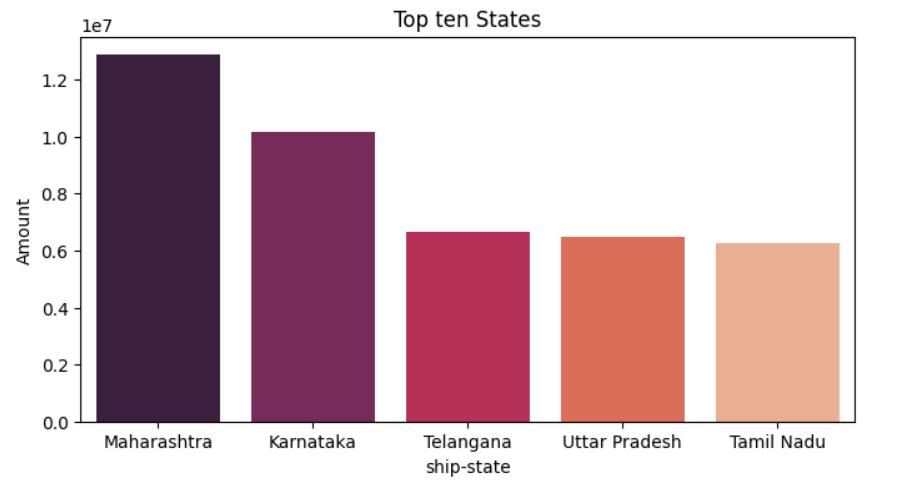
1. About 27.5 million sales in the month of April followed by 25.1 million in May and 22.6 million in June. Only Rs. 98,261 sales in March.



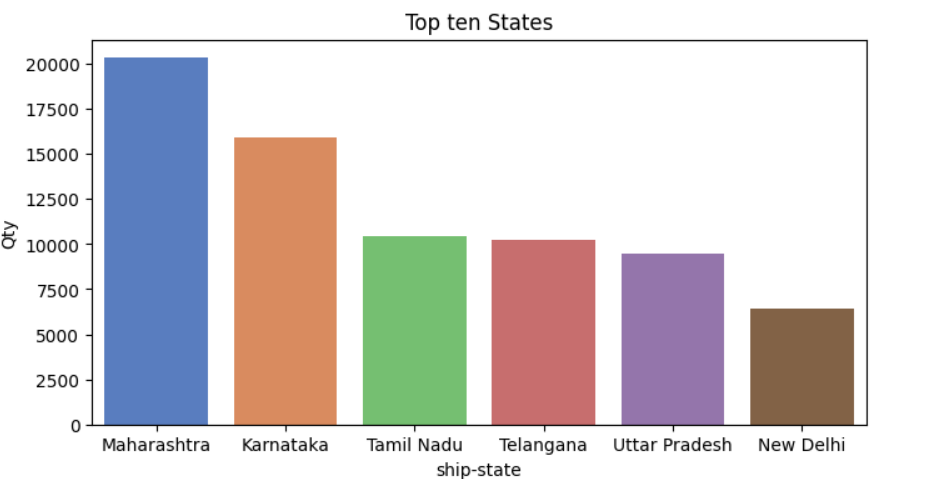
1. Highest revenue is generated from purchases in the set category followed by kurta.



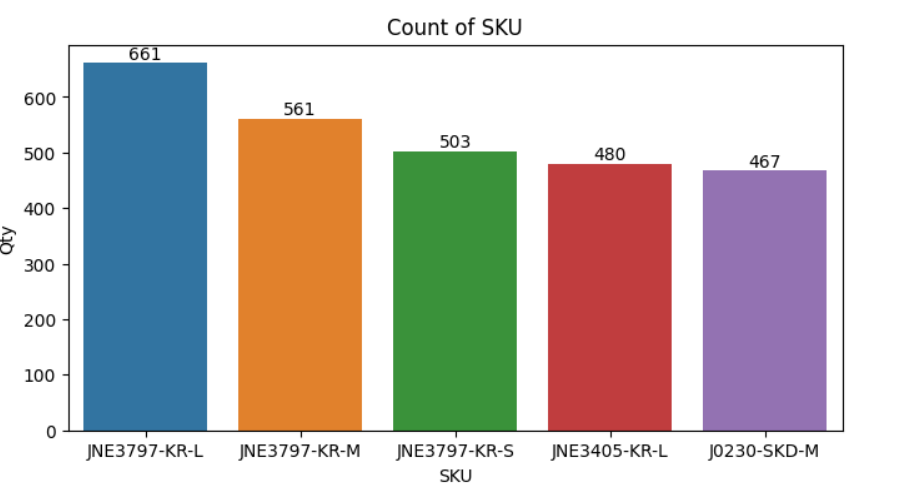
1. Top 5 cities with highest spending are Maharashtra, Karnataka, Telangana, Uttar Pradesh and Tamil Nadu.



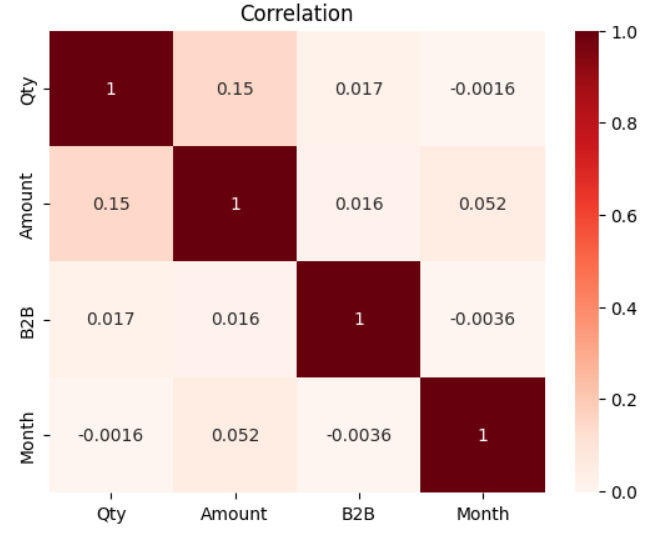
1. The highest quantity of purchase is recorder in Maharashtra, Karnataka, Tamil Nadu, Telangana and Uttar Pradesh.



1. Top 3 SKUs are JNE3797-KR-L, JNE3797-KR-M, JNE3797-KR-S.



* **Multivariate Analysis**

****

**8.Conclusion and Future scope**

* The analysis reveals that set and kurta contribute maximum to the total revenue generated with peak sales in the month of April and May. Furthermore, Maharashtra and Karnataka are the top performing regions, contributing the highest orders. By knowing top SKUs, it can help maintain the inventory for further orders.

In terms of customer preferences, it is evident that fast shipping play a crucial role in driving customer satisfaction and sales for Amazon fulfilled orders.

* **Future scope**

Predictive models can help in price optimization during the peak sales and can also help manage inventory. Advanced Machine learning models can help predict future sales trend.

**Real-World Applications:** Insights can be applied to optimize inventory management, targeted marketing, and promotional strategies.