

A Comprehensive Data Analysis of Online Shopping Trends, Customer Behavior, and Sales Performance

-Divya Kothari, 240485889

1.Introduction

Online shopping has become a big part of our daily lives, changing the way we shop and how businesses operate. With the growth of e-commerce, it is important to understand the data behind online stores to improve business decisions, customer experience, and long-term success. This dataset from an online shop gives us a closer look at various aspects of the business, such as customer details, orders, products, payments, and deliveries.

By analyzing this data, we can find useful patterns about how customers behave, which products are most popular, and how the business operates. Understanding this information helps businesses make smarter decisions about which products to sell, how to market them, and how to improve customer service. As online shopping keeps growing, using data effectively is essential for staying competitive and meeting customer needs.

We will explore various parts of the data, such as customer information, order trends, product details, payment methods, and shipping. Our goal is to find insights that can help businesses improve their operations, keep customers happy, and grow in the online shopping world.

2. Objective

The objective of this analysis is to explore and derive insights from an online shopping dataset to improve business operations, customer satisfaction, and sales performance. The analysis focuses on understanding customer behavior, order patterns, product performance, and delivery processes to help make informed decisions that improve efficiency and enhance the shopping experience. The specific objectives are as follows:

1. Customer Analysis:
 - Understanding Customer Behavior: explore customer demographics, purchasing frequency, and spending patterns to identify customer segments and their preferences for targeted marketing.
 - Retention and Loyalty: By analyzing repeat customers, we aim to enhance retention through loyalty programs, personalized offers, and tailored communication.
2. Sales and Order Trend Analysis:
 - Tracking Sales and Order Trends Analyzing monthly sales and order volume will highlight seasonal fluctuations, aiding in inventory, staffing, and marketing planning.
 - Order Value Distribution: Examining order value distribution will offer insights into customer spending, helping refine pricing, promotions, and product bundling strategies.
3. Product and Supplier Performance:
 - Identifying Top-Selling Products: Analyzing sales data will reveal popular items, informing inventory management, marketing, and stock replenishment.
 - Supplier Performance: Evaluating supplier contributions will help identify key suppliers, optimize relationships, and negotiate better terms.
4. Customer Feedback and Review Insights:
 - Analyzing Product Ratings and Reviews: Understanding how customer ratings and reviews influence product purchases
 - Review Volume and Product Engagement: Evaluating review volume per product will identify popular items, creating opportunities for increased visibility and engagement.

5. **Payment Behavior and Trends:**
 - **Preferred Payment Methods:** Analyzing payment choices reveals customer preferences and helps improve the checkout process.
 - **Payment Failures and Issues:** Identifying payment issues helps address technical problems and ensure smoother transactions.
6. **Delivery and Fulfillment Efficiency:**
 - **On-Time Delivery Performance:** Evaluating delivery rates and delays helps improve fulfillment processes and customer experience.
 - **Carrier Performance:** Comparing carriers helps choose the most reliable and cost-effective delivery partners.
7. **Geographical Insights:**
 - **Customer Distribution and Demand:** Analyzing customer locations helps target marketing and optimize logistics.
 - **Regional Sales Performance:** Assessing regional sales allows tailoring strategies and optimizing inventory.
8. **Order and Product Analysis:**
 - **Order Size and Frequency:** Analyzing order quantity and frequency helps improve inventory, packaging, and fulfillment.
 - **Product Pricing Trends:** Understanding product prices in orders aids in refining pricing strategies to meet demand.

In summary, this analysis aims to use data insights to make decisions that boost efficiency, improve customer satisfaction, and support business growth. By optimizing key areas like customer engagement and inventory management, the business can achieve long-term success.

3. Methodology:

3.1 Data Source

For this project, I used a dataset from Kaggle, which can be found here:
<https://www.kaggle.com/datasets/marthadimgba/online-shop-2024/data>.

The dataset consists of 8 CSV files that cover various parts of an online shopping business. These files include *customers.csv*, *order_items.csv*, *orders.csv*, *payment.csv*, *products.csv*, *reviews.csv*, *shipments.csv*, and *suppliers.csv*. Each file provides unique information, like customer details, order histories, payment methods, product details, reviews, delivery statuses, and supplier info.

This dataset is a great fit for analyzing many aspects of online shopping, such as understanding customer behavior, tracking sales trends, looking at product popularity, payment habits, and delivery processes. It offers a lot of useful information, including customer demographics, order, and payment history, and how well products are doing. The data covers several factors and is ideal for research, offering insights that can help improve business strategies in online retail.

3.2 Database Creation

The database for the online shopping platform was designed to store and analyze key business aspects efficiently. It includes interconnected tables for customer details, orders, payments, products, reviews, shipments, and suppliers. Here is an overview:

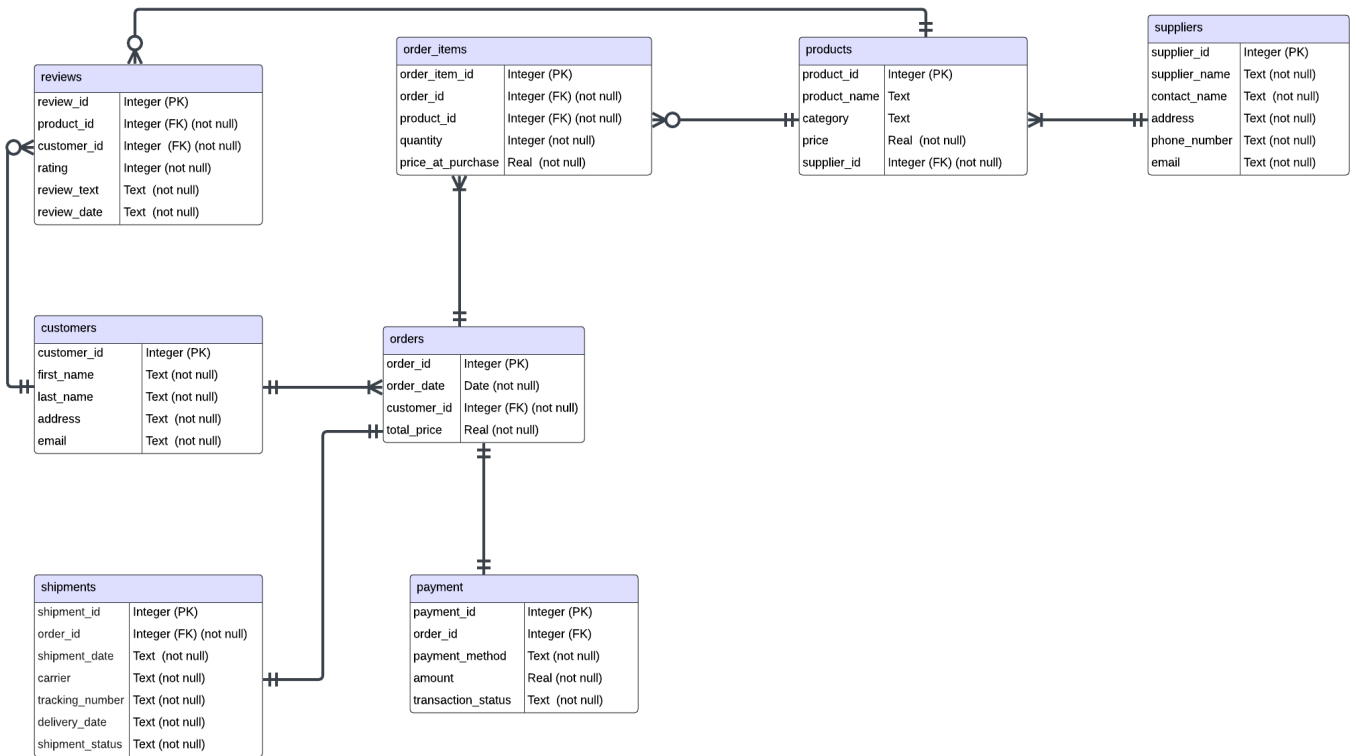
- **Customers Table:** Stores customer details (customer_id(PK), name, address, email, phone).
- **Orders Table:** Records order information (orders_id(PK), date, total price, customer ID).
- **Order_Items Table:** Tracks product details in each order (order_items_id(PK), product ID(FK), quantity, price).
- **Payment Table:** Stores payment details (payment_ID(PK), order ID(FK), method, amount, status).
- **Products Table:** Contains product info (product_id(PK), name, category, price, supplier ID).
- **Reviews Table:** Captures customer reviews (review_idf(PK), product ID(FK), customer ID(FK), rating, text).
- **Shipments Table:** Tracks shipment details (shipment_id(PK), order ID(FK), carrier, tracking number, delivery status).
- **Suppliers Table:** Stores supplier info (suppliers_ID(PK), name, contact, address, phone, email).

Relational Structure and Constraints:

- Primary Keys: Ensure unique identification of records (e.g., customer_id, order_id).
- Foreign Keys: Connect related tables (e.g., product_id links order_items and products).
- Unique Constraints: Prevent duplicate entries (e.g., email, tracking_number).
- Not Null Constraints: Ensure critical fields like price, order_date, or shipment_status are always populated.

The database for the online shopping platform stores key information like customers, orders, products, payments, and shipments. It helps organize data and makes it easy to analyze customer behavior, sales, payments, and deliveries.

I used Lucidchart to create the ERD (Entity-Relationship Diagram) that shows how these tables are connected. The database was built using SQLite and Python’s sqlite3 library, and it is designed to grow as the business expands.



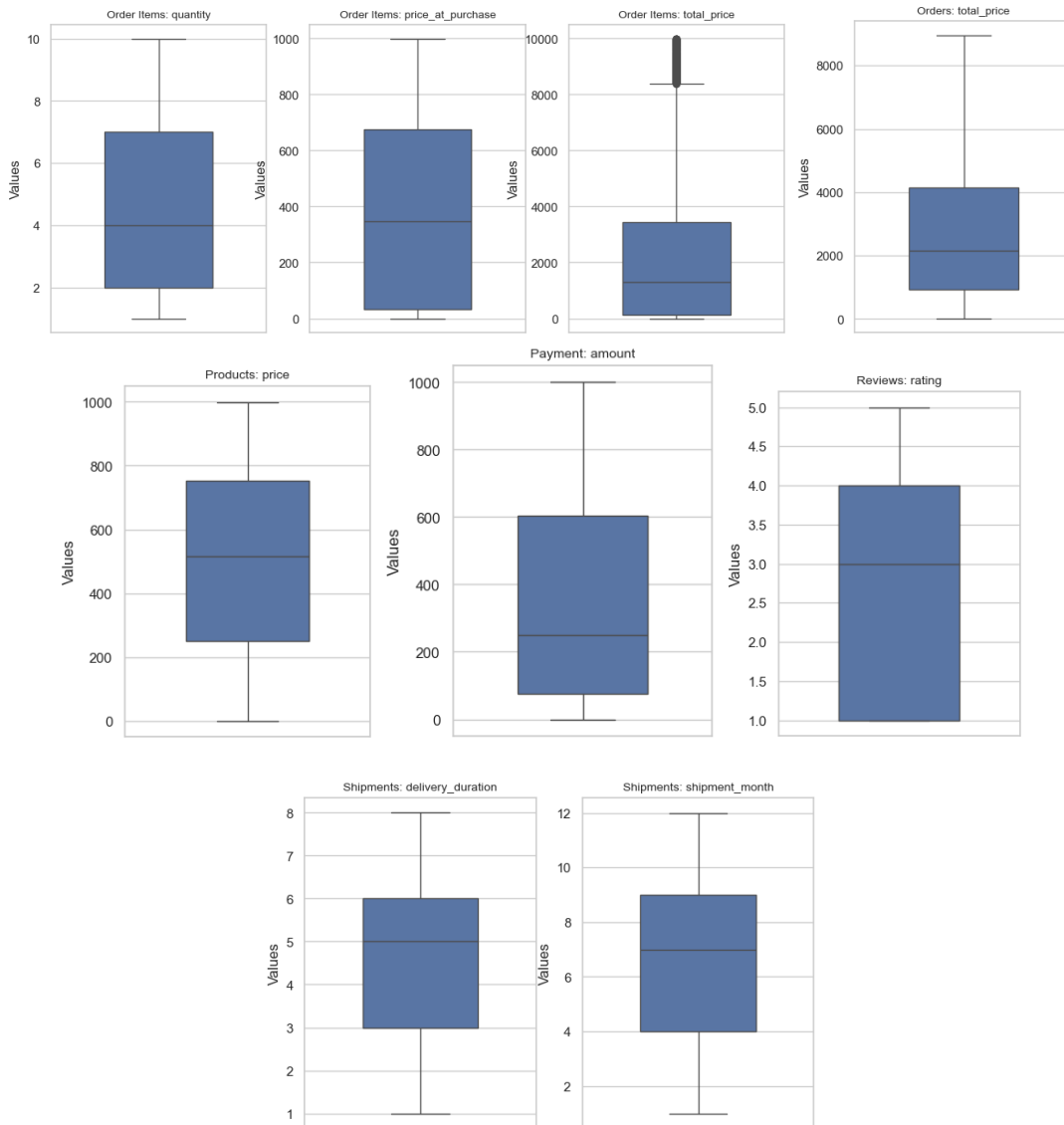
3.3 Data Preprocessing

Data preprocessing is an essential step in preparing the dataset for analysis, ensuring its consistency and reliability. Below is an overview of the key steps undertaken.

- Initially, columns involving dates, such as `order_date`, `review_date`, `shipment_date`, and `delivery_date`, were converted to the *datetime* format. This was done to facilitate time-based analysis, enabling a deeper understanding of trends and patterns over time.
- Duplicate entries were identified and removed from critical tables like `customers`, `orders`, `order_items`, and others to eliminate redundancy.
- Outlier detection was carried out using box plots, focusing on columns like `quantity`, `price`, `total_price`, `amount`, and `rating`. These plots visually display the distribution, median, and interquartile range (IQR), with values outside the whiskers considered outliers.

The formula to detect outliers is:

- Lower Bound = $Q1 - 1.5 * IQR$
- Upper Bound = $Q3 + 1.5 * IQR$



The box plot for quantity showed some extremely high values, which could be due to bulk purchases or data entry mistakes that need further review. The "amount" column had a wide range, with certain values much higher than typical, suggesting potential outliers. The box plot for total price also revealed large outliers, which may be due to bulk orders or unusually expensive items, indicating that while most orders are within a normal price range, a few are much higher. The rating box plot showed that most ratings were around the middle, but a few extremely low ratings could point to product issues or inaccuracies.

The datasets were successfully merged to connect related information, such as customer details with orders, product details with order items, and payment data with orders, giving us a complete view of the data. Feature engineering added useful insights, including Customer Lifetime Value (CLV), total order quantities, payment method frequencies, product review counts, and total revenue by product category. A flag was also created to identify whether deliveries were on time. These steps prepared the data for deeper analysis, allowing us to understand online shopping behavior, customer purchasing patterns, and product performance.

3.4 Exploratory Data Analysis

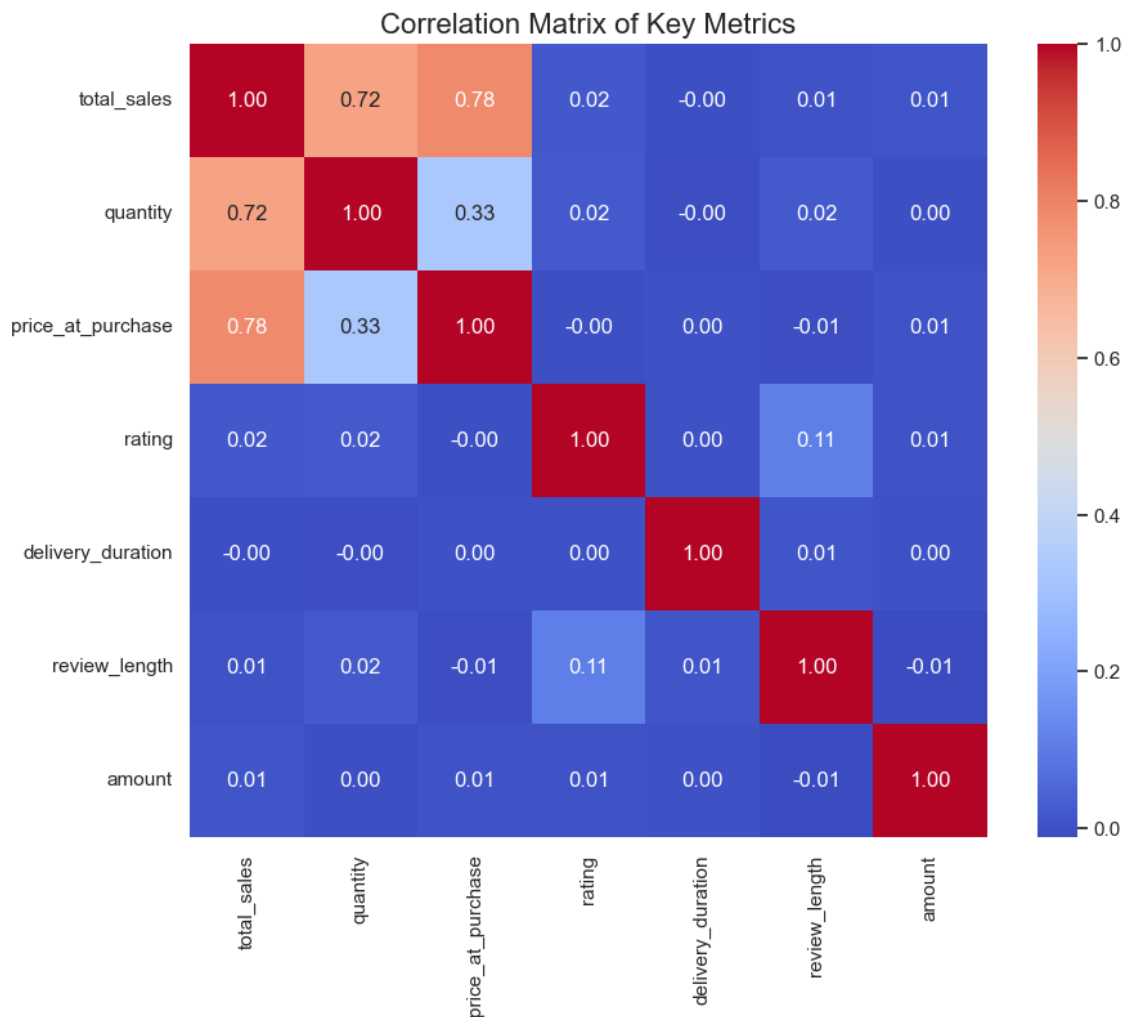
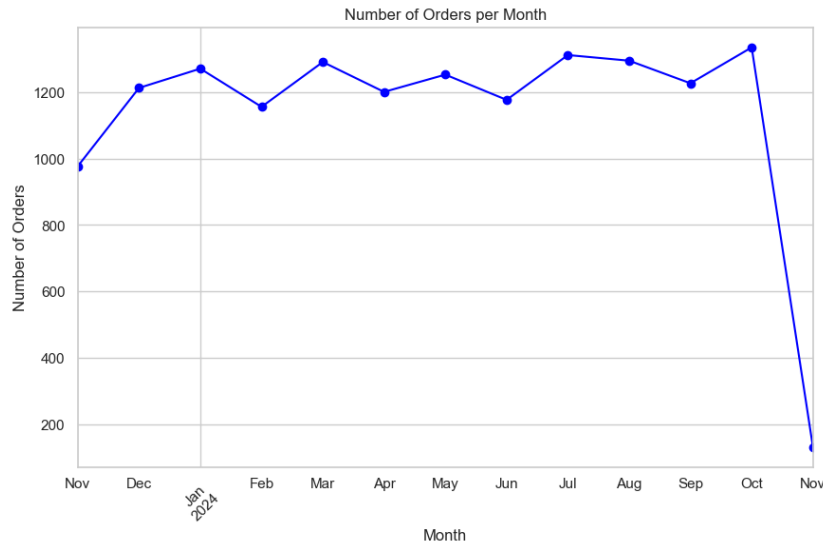


Figure: Correlation Matrix of Key Metrics

Sales and Order Trends:



- The line plot of total orders per month helped track the trend of orders over time, highlighting seasonal patterns and shifts in customer demand. This analysis helps predict peak months, aiding in resource allocation and planning.
- The line plot of total sales per month visualized sales fluctuations, revealing high and low sales periods, which can be used to assess sales performance and forecast future sales.

Distribution of Revenue and Orders:

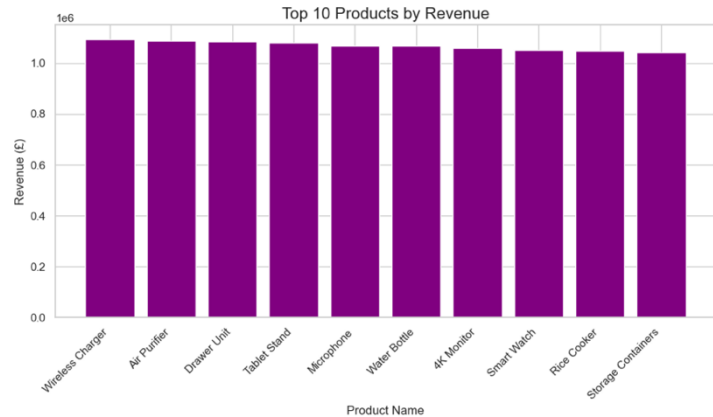
- The order values histogram showed that most customers spend within a certain range, with a few spending much more. This helps us understand customer spending behavior.
- The number of orders per customer histogram revealed how often customers make purchases, helping identify repeat buyers and one-time shoppers for targeted marketing.

Customer Spending Patterns:

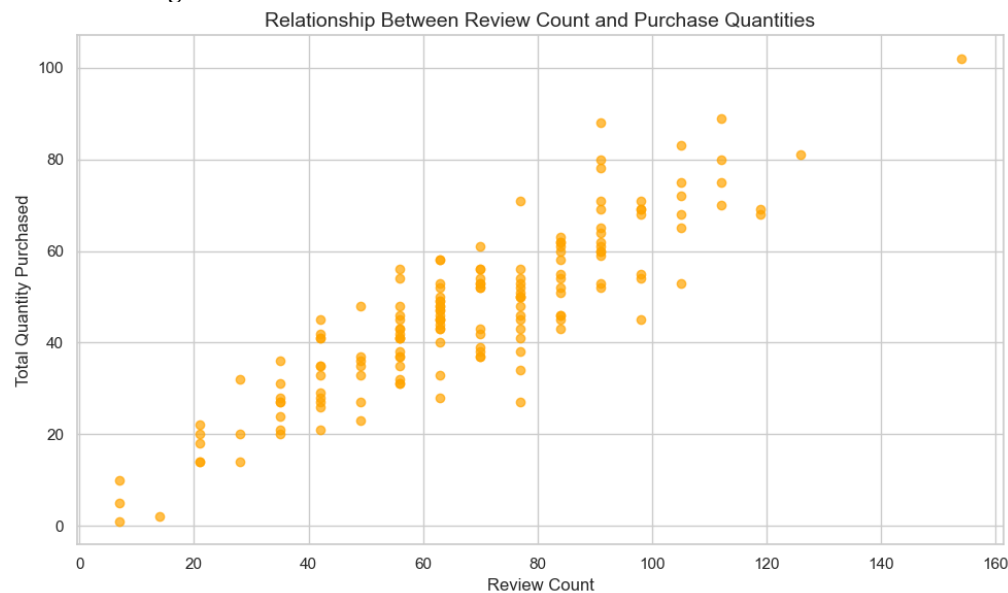
- The box plot for Customer Lifetime Value (CLV) demonstrated the spread of customer spending, highlighting high-value customers. This helps tailor strategies for retaining these customers with loyalty programs or exclusive offers.
- The scatter plot between the number of orders and total spending indicated that frequent buyers tend to spend more, revealing a clear relationship between customer order frequency and their spending level.

Product and Supplier Insights:

- The bar plot for top products by revenue highlighted the highest-performing products in terms of revenue, aiding decisions about promotions, stock replenishment, and product focus.
- The bar plot for top suppliers by revenue displayed which suppliers contributed the most to sales, helping prioritize supplier relationships and improve negotiations or partnerships.



Product Ratings and Reviews:



- The scatter plot showing product ratings vs. purchase quantity confirmed that higher-rated products tend to see more purchases. This emphasizes the importance of positive reviews in influencing customer behavior.
- The bar plot for the number of reviews per product indicated which products received the most attention, helping identify products that need more reviews or marketing efforts to boost visibility.

Customer Retention and Repeat Purchases:

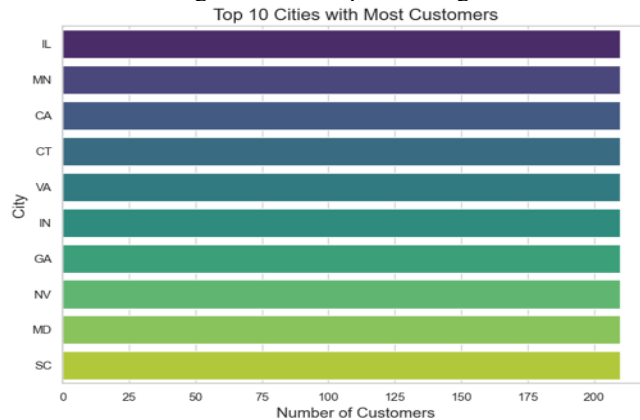
- The scatter plot for total spend vs. order count illustrated repeat customer behavior, showing customers who made multiple purchases and spent more. This insight can help target loyalty offers and improve retention strategies.
- The line plot for repeat purchase rate per month tracked how repeat purchases varied over time, offering insights into the success of customer retention initiatives.

Delivery and Fulfillment:

- The delivery delays plot showed which months or carriers had more delays, helping improve planning with delivery partners.
- The on-time delivery rate plot showed how well deliveries met expectations, guiding better delivery strategies.

Geographical Insights:

- The bar plot for the number of customers per city revealed which cities had the most customers, helping refine marketing efforts and optimize logistics for those locations.



Order Items Analysis:

- The histogram for order item quantity provided insights into the typical number of items purchased per order. This data can inform packaging decisions, inventory management, and order fulfillment strategies.
- The bar plot for product price range per order displayed the common price points of products in each order, offering valuable information for adjusting pricing strategies based on customer preferences.

These visualizations provide valuable insights into customer behavior, spending habits, product performance, and delivery efficiency. Each plot helps us understand the data better, guiding decisions in areas like inventory management, marketing, customer retention, and supplier relationships.

4. Key Insights and Findings

1. Customer Behavior Patterns

Findings:

- Order Size:** Approximately 70% of orders consist of 1-3 items, indicating a preference for smaller purchases.
- Seasonal Trends:** Sales peaked in October, driven by festive offers, but declined significantly in November.
- Revenue Concentration:** The top 10 customers account for 25% of total revenue, each spending between £500 and £1,000 annually.

Implications:

Festive sales drive revenue, highlighting the need for ongoing customer engagement, while high-value customers play a crucial role in overall earnings.

Recommendations:

1. **Sustain Seasonal Momentum:** Introduce targeted campaigns, such as Black Friday or Cyber Monday promotions, to mitigate the post-festival dip in sales.
 2. **Enhance Loyalty Programs:** Implement loyalty rewards, offering benefits like exclusive discounts or early sales access to high-value customers.
-

2. Product Performance Analysis

Findings:

- **Top Products:** Wireless Chargers, Drawer Units, and Air Purifiers generate the highest revenue.
- **Category Leaders:** Home & Kitchen products drive revenue, while Electronics lead in sales volume.
- **Delivery Concerns:** While Furniture is a high-performing category, delayed deliveries negatively affect customer satisfaction.

Implications:

Consistent availability of top-performing products is critical, and addressing delivery delays can significantly improve customer trust and retention.

Recommendations:

1. **Inventory Optimization:** Maintain adequate stock levels for top-selling products to prevent shortages.
 2. **Improve Logistics:** Partner with reliable organization providers to ensure timely delivery, particularly for Furniture.
-

3. Supplier and Delivery Performance

Findings:

- **Supplier Quality:** Precision Suppliers LLC contributes high revenue but keeps a low product rating of 3.2/5.
- **Delivery Excellence:** DHL has the highest on-time delivery rate (95%), while Furniture deliveries often face delays.

Implications:

While high-revenue suppliers like Precision Suppliers LLC are valuable, improving their product quality is essential. Reliable logistics partners like DHL should be prioritized.

Recommendations:

1. **Supplier Improvement Plans:** Collaborate with Precision Suppliers LLC to enhance product quality through regular feedback and performance reviews.
 2. **Expand Reliable Partnerships:** Assign a larger share of deliveries to DHL to reduce delays and improve customer satisfaction.
-

4. Reviews and Ratings

Findings:

- **Insufficient Feedback:** Popular products like the Wireless Charger lack adequate reviews.
- **Correlation Between Reviews and Sales:** Products with higher reviews consistently achieve better sales performance.

Implications:

Encouraging customer reviews can enhance product credibility, positively affecting purchase decisions and overall sales.

Recommendations:

1. **Incentivize Reviews:** Offer discounts or loyalty points to customers who provide reviews.
2. **Automated Follow-Ups:** Use email reminders post-purchase to request feedback, ensuring seamless customer engagement.

5. 1 Conclusion

This analysis highlights the key factors that influence the success of an e-commerce business: seasonal promotions, top-selling products, efficient logistics, and strong customer engagement. By addressing issues like delivery delays, supplier performance, and customer reviews, the business can create a solid foundation for growth and customer loyalty.

Strategic Focus Areas:

1. **Sustainable Sales:** Keep sales steady throughout the year by planning promotions that go beyond just seasonal peaks. This will help maintain consistent revenue.
 2. **Operational Efficiency:** Build strong relationships with reliable suppliers and logistics partners to improve efficiency and reduce delays.
 3. **Customer Engagement:** Use data to understand customer behavior and create targeted marketing campaigns. This will help build long-term customer loyalty.
-

5.2 Recommendations for Implementation:

1. **Launch Targeted Campaigns:** Offer special promotions during slower months, focusing on encouraging repeat customers.
2. **Monitor Supplier Performance:** Regularly check supplier performance and give feedback to maintain quality standards.
3. **Improve Logistics:** Allocate more resources to top-performing delivery partners to reduce delays.
4. **Build Customer Trust:** Encourage customers to leave reviews by sending automated reminders and offering small incentives.

Expected Outcomes:

- **Revenue Growth:** A 15%-20% increase by improving product availability and targeted promotions.
 - **Customer Satisfaction:** A 10%-15% improvement through faster deliveries and better product quality.
 - **Operational Efficiency:** Cost savings and improved performance by improving logistics and supplier management
-

6. Reference list

- Gelder, K. van (2024). *E-commerce Worldwide - Statistics & Facts*. [online] Statista. Available at: <https://www.statista.com/topics/871/online-shopping/#topicOverview>.
- Huseynov, F. and Yildirim, S.Ö. (2019). "Online Consumer Typologies and Their Shopping Behaviors in B2C E-Commerce Platforms." ("Online Consumer Typologies and Their Shopping Behaviors in B2C E ...") *SAGE Open*, 9(2), p.215824401985463. <https://journals.sagepub.com/doi/10.1177/2158244019854639>.
- Rosário, A. and Raimundo, R. (2021). "Consumer Marketing Strategy and E-Commerce in the Last Decade: a Literature Review." ("Consumer Behavior and the Evolution of Marketing Strategies in the Age ...") *Journal of Theoretical and Applied Electronic Commerce Research*, [online] 16(7), pp.3003–3024. <https://doi.org/10.3390/jtaer16070164>.
- www.ripublication.com. (n.d.). *GJMBS, Global Journal of Management and Business Studies, Computer Science Journals, Journals Publishers, Computer Science Journals in India, Indian Journals Subscription Agency, Indian Books Distributors*. [online] Available at: <https://www.ripublication.com/gjmbs.htm>.