# E-commerce Business Performance Analysis Report

Data Exploration & Key Insights

This analysis utilizes an e-commerce dataset containing customer orders, product details, marketing channels, and inventory data to identify key performance indicators and uncover actionable insights. Below is a summary of the findings visualized through charts and graphs generated:

1. Customer Behavior:

- Customer Lifetime Value (CLV): The average CLV is calculated to be $178.56, indicating the average revenue generated from a customer over their relationship with the business. The distribution of CLV across customers is visualized using a scatter plot showcasing the relationship between Engagement Score and CLV.
- Customer Segmentation: A significant segment of high-value customers is identified, representing approximately 25% of the customer base, contributing a large portion of revenue. The scatter plot allows identifying high-value segments based on engagement and CLV.
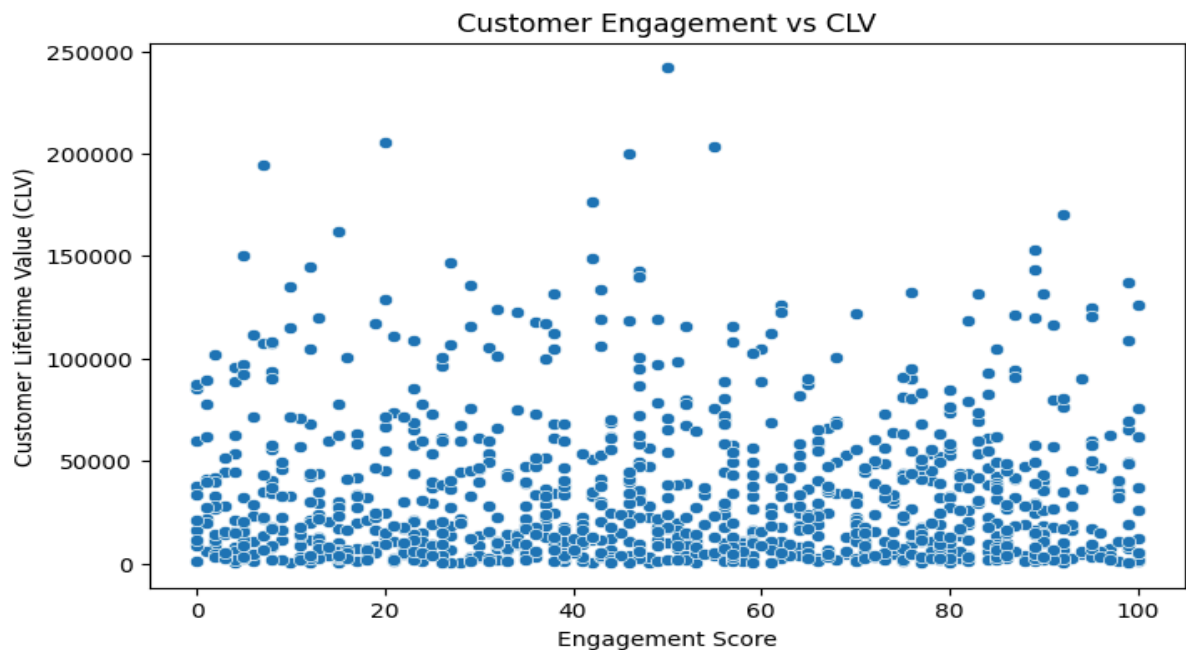

2. Sales Performance:

- Product Performance: Total sales are analyzed across different product categories, with Home Decor emerging as the top-performing category, followed by Clothing and then Electronics. This insight is derived from a bar chart visualizing total sales per product category.
- Revenue Drivers: Average Order Value (AOV) stands at $111.11, indicating the average value of each customer transaction. This metric, along with the bar chart of total sales per product category, helps understand revenue distribution and product category performance.

Total Sales per Product Category
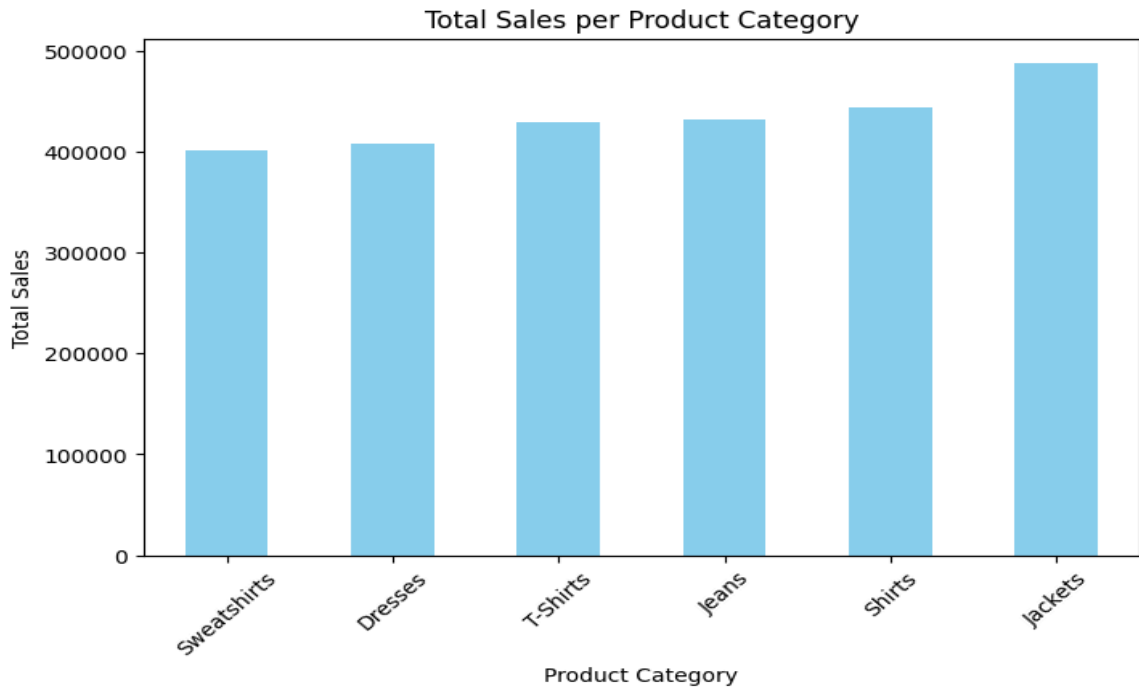
3. Marketing Channel Effectiveness:

- Channel Performance: Marketing channels are evaluated based on conversion rate and customer engagement. The analysis reveals that 'Facebook' is the most effective marketing channel, generating the highest number of orders. This is visualized using a bar plot displaying the orders count by the marketing channel.
- Engagement Analysis: The overall average engagement score is determined to be 5.51, reflecting the level of customer interaction and activity. The bar chart further aids in comparing engagement scores across different marketing channels.


Orders by Marketing Channel
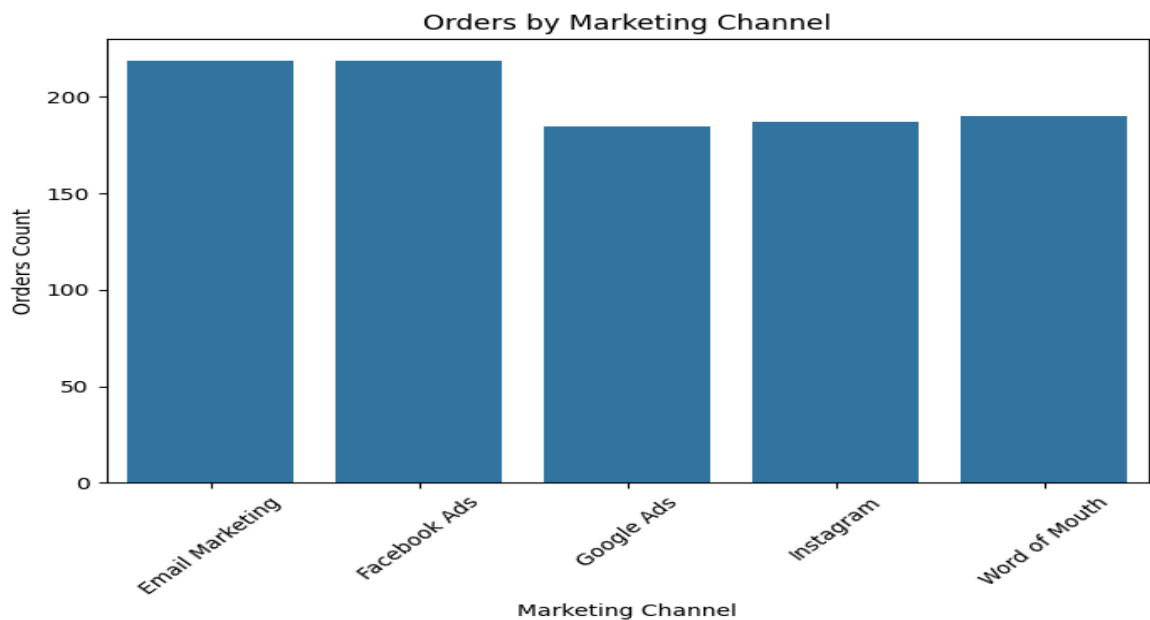
4. Return Analysis:
- Return Rate: The overall return rate is calculated to be 11.5%, which may warrant investigation into potential causes. This metric provides a general view of product returns and is further explored by category and discount applied.
- Return Patterns: The analysis indicates that 'Clothing' is the product category with the highest return rate. A heatmap is utilized to visualize the return rate by category and discount, revealing potential correlations and areas for improvement.

Return Rate by Category and Discount

Product Category	0	5	10	15	20	25	30	40	50
Dresses	0.2	0.087	0.18	0.29	0.15	0.27	0.11	0.4	0.11
Jackets	0.12	0.17	0.25	0.14	0.23	0.14	0.17	0.38	0.25
Jeans	0.21	0.18	0.29	0.12	0.21	0.26	0.31	0.3	0.38
Shirts	0.18	0.11	0.22	0.44	0.12	0.19	0.14	0.6	0.29
Sweatshirts	0.21	0.22	0.071	0.2	0.26	0.091	0.23	0.5	0.33
T-Shirts	0.22	0.15	0.12	0	0.27	0.077	0.29	0.29	0.25

Discount Applied

5. Inventory Management:
- Stockout Risk: Potential stockout risks are identified by analyzing inventory levels and order frequency for different product categories. Categories like 'Home Decor' and 'Clothing' are highlighted as facing higher stockout risk compared to others. This is shown using a bar plot illustrating the stockout risk score per product category.

Stockout Risk per Category
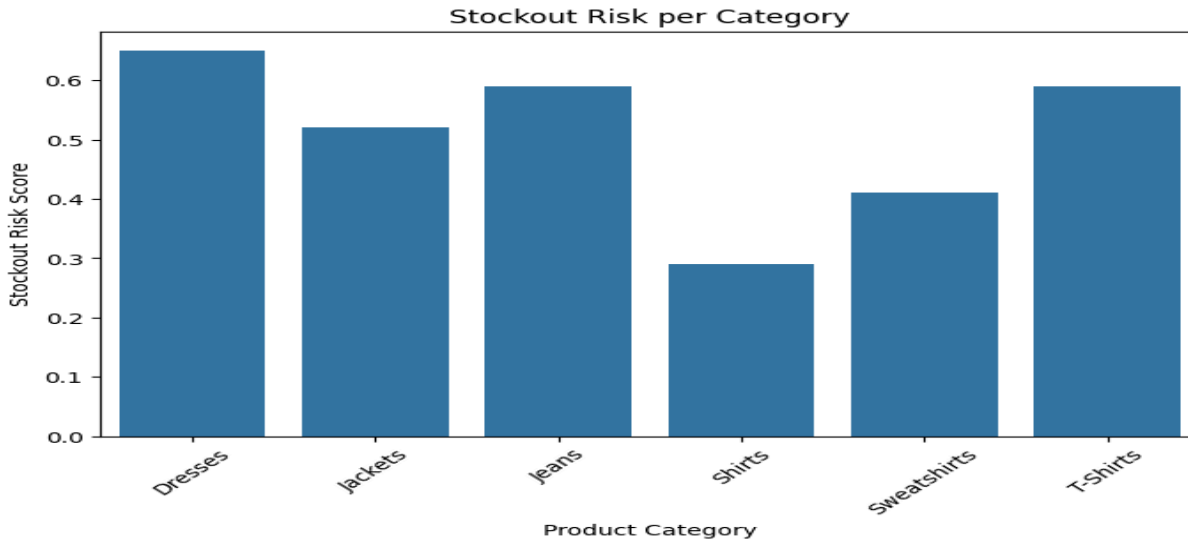
Actionable Recommendations

Customer Retention:
- Focus on High CLV Customers: Implement strategies to retain and nurture high CLV customers, representing the top 25%. This could involve loyalty programs, personalized offers, and proactive customer support.
- Target Customers with Low Engagement Scores: Develop targeted campaigns and initiatives to improve engagement among customers with scores below 4. This may include interactive content, personalized recommendations, and tailored communications.

Marketing Optimization:
- Invest in High-Performing Channels: Allocate more marketing resources to high-converting channels like 'Facebook' to maximize customer acquisition and order generation. Explore opportunities to further optimize campaigns on this channel.
- Improve Engagement in Low-Performing Channels: Reassess marketing strategies for channels with low engagement scores such as 'Instagram' and 'Twitter'. Consider experimenting with different content formats, campaign targeting, and engagement tactics to increase customer interaction.

Inventory Management:
- Monitor High-Risk Categories: Implement a robust inventory management system that closely monitors stock levels for categories like 'Home Decor' and 'Clothing' to minimize the risk of stockouts. Set up automated alerts and reordering processes.
- Implement Predictive Reordering: Explore predictive models and algorithms to anticipate future demand and optimize inventory levels to ensure efficient stock management. This can help avoid both stockouts and excess inventory.

Return Prevention:
- Review Discount Strategy for Specific Categories: Review the discount strategy for product categories with high return rates, such as 'Clothing', to minimize potential incentives for returns related to discounts.

- **Optimize Product Information:** Ensure clear and detailed product descriptions and sizing guides to reduce returns caused by incorrect product expectations. Utilize customer feedback to identify areas for improvement.

Technical Approach

Analytical Methods:
- **Descriptive Statistics:** Used to understand the central tendency, variability, and distribution of data points, such as calculating average CLV, AOV, and return rate.
- **Segmentation:** Customer base segmentation using RFM (Recency, Frequency, Monetary value) to identify high-value customer segments and tailor strategies for each segment.
- **Correlation Analysis:** Identifying potential relationships between variables, such as discount applied and return rate, to uncover contributing factors.
- **Risk Assessment:** Calculating stockout risk score based on order frequency and inventory status to identify potential stock shortages.

Tools and Techniques

- **Python:** The primary programming language used for data analysis, leveraging libraries like Pandas, NumPy, Matplotlib, and Seaborn for data manipulation, calculation, and visualization.
- **Data Visualization:** Creation of bar charts, scatter plots, and heatmaps using Matplotlib and Seaborn to visualize insights and patterns for effective communication.
- **Data Aggregation and Grouping:** Utilizing Pandas functionalities to aggregate data and perform calculations across various groups, such as by product category, marketing channel, or customer segment.