

Documentation

E-Commerce Sales Data Analysis: Insights and Findings

Project Overview

This project analyzed e-commerce sales data to uncover key insights into customer behavior, sales patterns, and geographic revenue distribution. The analysis was performed using Python libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Folium for map visualization. Outliers were removed using the IQR method to ensure accurate results.

Key Findings

1. Customer Retention and Acquisition

High Customer Retention:

The majority of sales come from regular customers, indicating a strong retention rate.

Low New Customer Orders:

Orders from new customers are significantly fewer compared to regular customers, contributing less to overall revenue.

2. Customer Segments and Revenue Contribution

Customers were grouped into four segments: Low, Medium, High, and Very High, based on their purchase value.

Low Segment (11% of customers): Contributes only 0.6% of revenue.

Medium Segment (43% of customers): Contributes 10.4% of revenue.

High Segment (19% of customers): Contributes 12.3% of revenue.

Very High Segment (25% of customers): Contributes a massive 76% of total revenue.

Conclusion:

The Very High segment, though only 25% of the customer base, generates the majority of revenue (76%), highlighting the importance of focusing on these high-value customers.

3. Sales Trend by Month

Peak Sales in November:

Sales reach their highest point in November, starting to rise in September and dropping off after November. This indicates that holidays or festivals may drive significant sales during this period.

4. Top-Selling Products

Identified top-selling products based on StockCode, visualized using a bar chart.

5. Top Revenue-Generating Countries

United Kingdom (UK)

Germany

France

EIRE (Ireland)

6. Countries with the Largest Customer Base

United Kingdom (UK)

Germany

France

7. Map Visualization of Revenue by Country

Using Folium, I visualized the geographic distribution of revenue. This highlighted the concentration of high-value customers in key countries, especially in EIRE.

Specific Analysis: High-Value Customers

8. Monthly Revenue Trend of Very High Segment

Like the overall trend, sales peak in November, but there is also a moderate spike in July, suggesting a secondary sales event.

9. Top Countries for Very High Segment Customers

United Kingdom (UK)

Germany

France

10. Countries with the Highest-Spending Customers

EIRE (Ireland): Despite having a smaller customer base, it generates significantly more revenue.

United Kingdom (UK)

Germany

France

Specific Analysis: High-Value Regular Customers

11. Top Countries for High-Value Regular Customers

United Kingdom (UK)

Germany

France

Belgium

EIRE (Special Focus)

12. Revenue per Customer by Country

EIRE: Contributes 63% of total revenue, despite having a smaller customer base.

United Kingdom (UK): 11% of revenue

Belgium: 6% of revenue

Germany: 8% of revenue

France: 9% of revenue

EIRE stands out as a critical market, with customers contributing a disproportionately high share of revenue compared to their numbers.

Tools and Techniques Used

Data Cleaning: Used the IQR method to remove outliers.

Analysis Libraries: Pandas, NumPy

Visualization Libraries: Matplotlib, Seaborn, Folium for geographic mapping

Recommendations

Focus on High-Value Customers:

Implement loyalty programs or exclusive offers for Very High and High segments to maintain retention.

Personalize marketing campaigns to engage these customers more effectively.

Increase New Customer Acquisition:

Invest in targeted advertising campaigns, especially during the September to November peak sales period.

Offer discounts or promotions for first-time buyers to encourage purchases.

Optimize for Peak Sales Period (November):

Ensure sufficient inventory and staffing to handle increased demand during the holiday season.

Launch special promotions or holiday-themed campaigns starting in September.

Leverage Opportunities in EIRE:

Focus on nurturing high-value customers in EIRE with personalized offers, since they generate the most revenue per customer.

Explore expanding marketing efforts to increase the customer base in EIRE.

Explore Secondary Sales Spikes (July):

Investigate potential reasons for the moderate sales spike in July (e.g., summer promotions or local events) and capitalize on this trend.

Strengthen Medium Segment Revenue:

Encourage Medium segment customers to increase their spending through tiered rewards, upselling, or cross-selling strategies.

Data-Driven Decision Making:

Use data insights to continuously refine marketing, inventory, and customer engagement strategies.

Conclusion

This analysis emphasizes the importance of Very High customers, particularly in EIRE, which contributes the most revenue per customer. Sales peak in November, likely due to holidays or festivals, presenting an opportunity to optimize marketing strategies during that period.