

Zeotap Data Science Internship, EDA Documentation

Mahesh Reddy

1 EDA

1.1 Distribution of signup dates

The histogram displayed above showcases the distribution of signup dates over time. From an exploratory data analysis (EDA) perspective, it provides insights into user acquisition trends. The frequency of signups is segmented into specific time intervals, revealing fluctuations in activity. Peaks in signup frequency suggest periods of increased user acquisition, possibly influenced by marketing campaigns, promotions, or seasonal demand. Conversely, dips in the chart highlight periods of lower activity, which may warrant further investigation to uncover underlying causes. This visualization helps identify temporal patterns and trends, offering actionable insights for optimizing customer acquisition strategies.

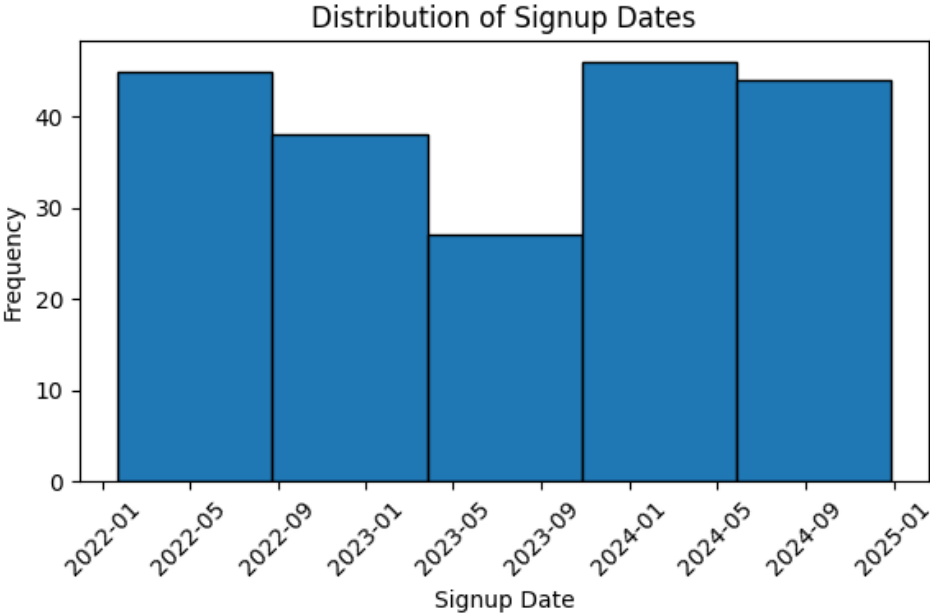


Figure 1: Sales Revenue by Region

1.2 Total Price by Product Category

The bar chart illustrates the total price contributions of various product categories. It highlights that books have the highest cumulative sales value among the categories, followed by electronics and clothing. Home decor has the lowest total value, indicating

comparatively less revenue generation. This comparison helps in understanding which product categories drive the most revenue and can guide future strategies for inventory prioritization and sales focus.

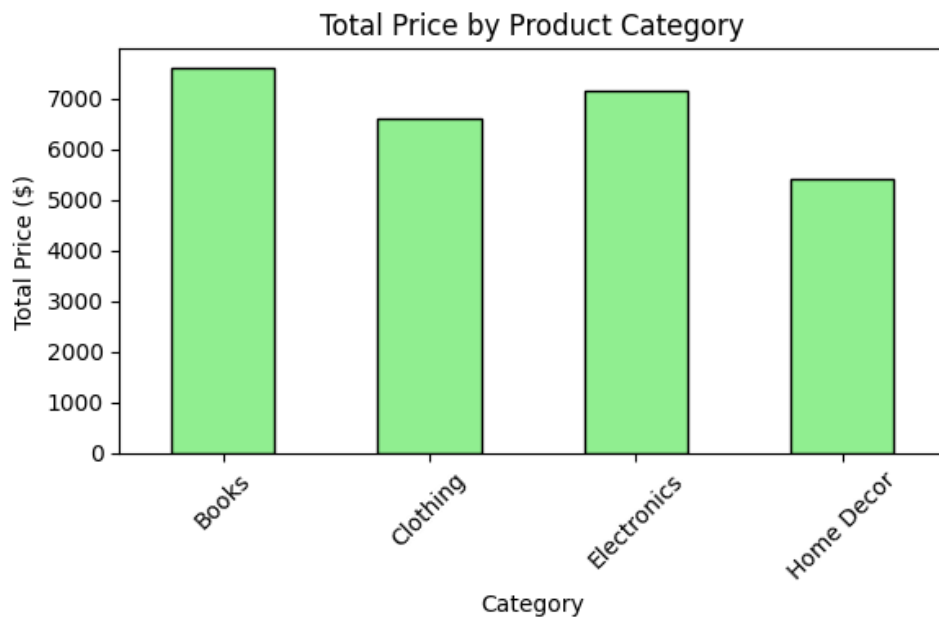


Figure 2: Sales Revenue by Region

1.3 Total transactions value by region

The bar chart showcases the total transaction values across different regions. South America leads with the highest transaction value, indicating strong sales performance in this region. Europe, Asia, and North America show relatively similar transaction values, with South America standing out significantly. This regional breakdown provides a clear understanding of where the majority of revenue is generated, offering valuable insights for regional-focused strategies and resource allocation.

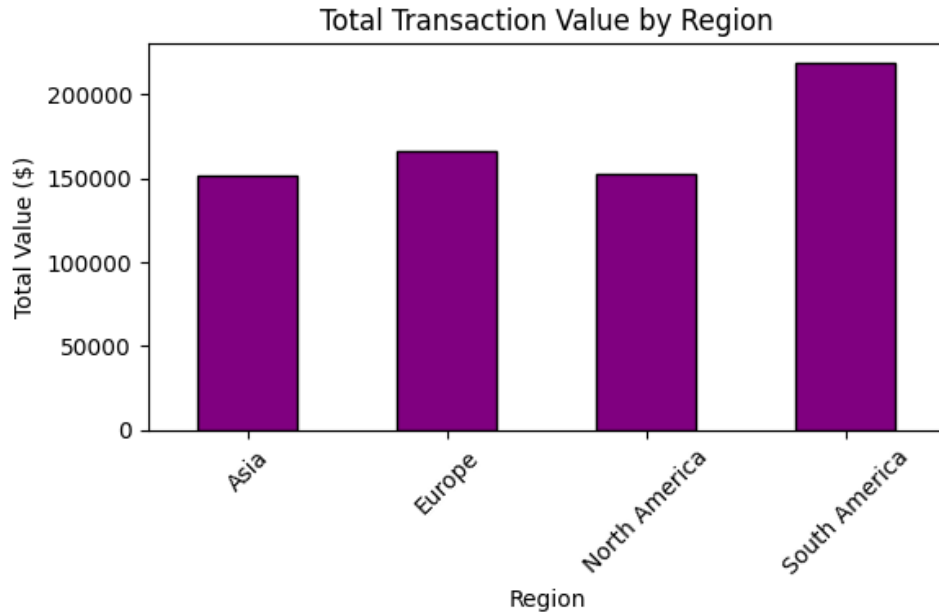


Figure 3: Sales Revenue by Region

2 Business Insights

2.1 Sales revenue by region

The sales revenue distribution across regions reveals South America's dominance, contributing 31.61 percent of the total revenue, the highest among all regions. Europe follows with 23.96 percent, showing a significant but smaller share. Meanwhile, North America and Asia exhibit nearly identical contributions, accounting for 21.96 percent and 21.94 percent, respectively. This data highlights the substantial revenue share from South America, surpassing North America's and Asia's combined contributions, showcasing its vital role in overall sales performance.

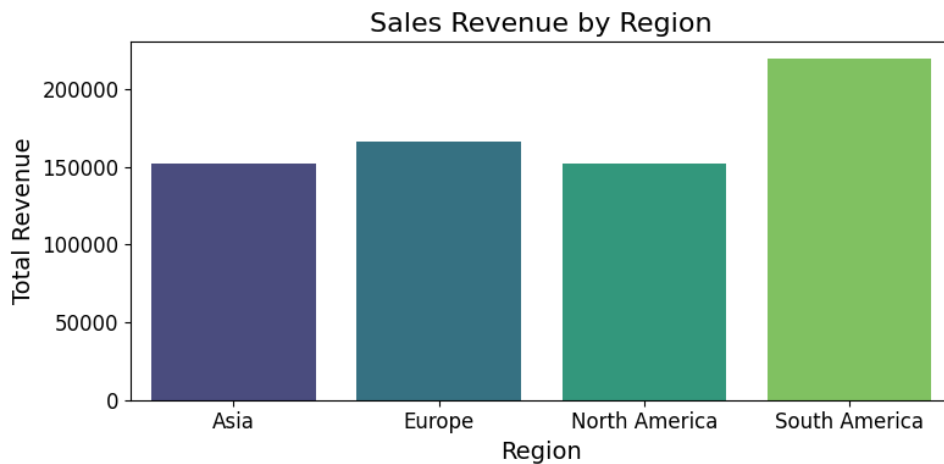


Figure 4: Sales Revenue by Region

2.2 Quantity vs Revenue by Category

The bubble chart illustrates the relationship between quantity sold and revenue generated across various categories, providing key insights into their contributions. The Books category stands out as the most significant contributor, with both higher sales volume and revenue, emphasizing its strong market position. In contrast, Home Decor, despite substantial sales volume, generates relatively lower revenue, likely due to lower price points or profit margins. Clothing demonstrates a unique trend, with a smaller sales quantity but higher revenue, suggesting higher unit prices or profitability. Finally, Electronics achieves higher revenue than Home Decor despite selling fewer units, underscoring its premium pricing or profitability potential. This analysis highlights distinct performance dynamics among the categories.

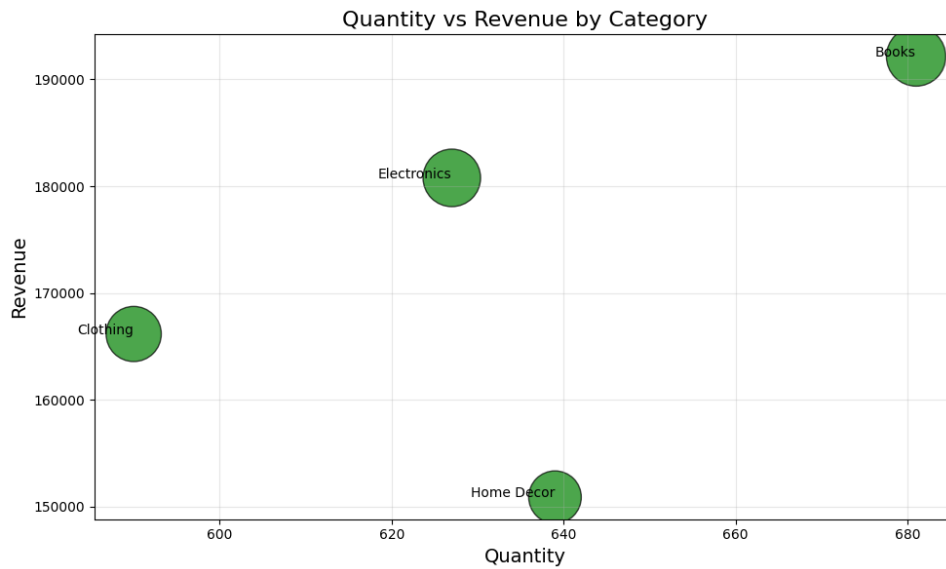


Figure 5: Quantity vs Revenue by Category

2.3 Average time difference between signup and first transaction user

Out of a total of 200 users who signed up on a specific date, the time difference between the first signup and the first transaction was analyzed. However, certain data points were deemed invalid, as they indicated transactions occurring before the signup date, which is logically incorrect. These invalid data points were excluded from the mean calculation. Among the 200 users, 137 valid points were identified and considered, resulting in a calculated average time difference of approximately 403.03 days between the signup and the first transaction. This analysis highlights the importance of data validation in ensuring the accuracy of insights.

2.4 Repeat Customer Rate by Region

The analysis of repeat customer rates across regions reveals valuable insights into customer behavior. Europe leads with the highest repeat customer rate of 96.00 percent, indicating strong customer retention and loyalty in this market. South America follows closely at 94.92 percent, showcasing consistent customer engagement. North America,

with a repeat customer rate of 93.48 percent, also demonstrates solid performance, while Asia has a slightly lower rate of 90.91 percent, suggesting room for improvement in customer retention strategies. These findings emphasize the importance of tailored approaches to strengthen customer loyalty in regions with lower rates and maintain effective strategies in high-performing markets.

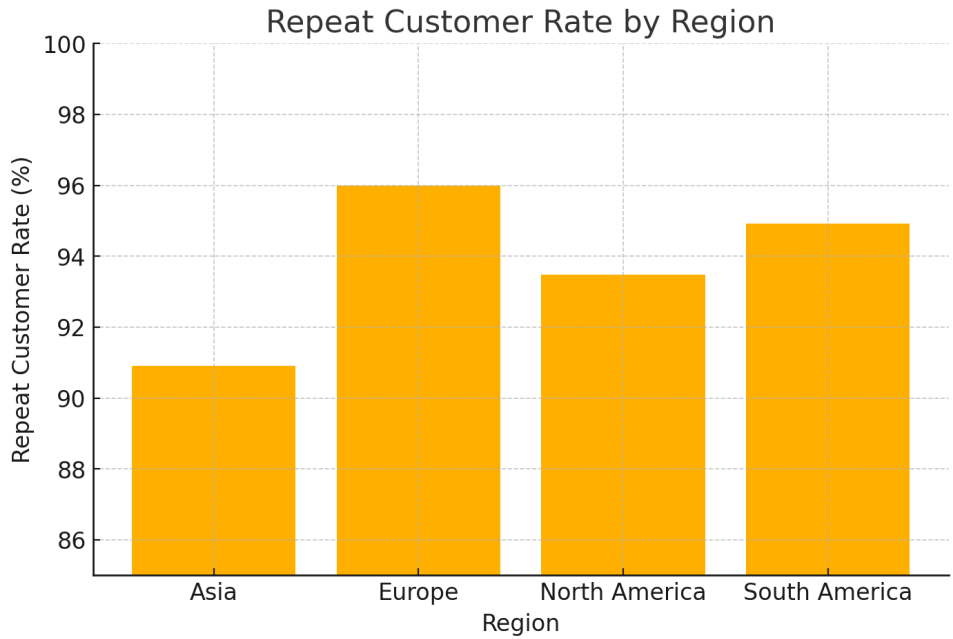


Figure 6: Quantity vs Revenue by Category

Region	Repeat Customer Rate (%)
Asia	90.91
Europe	96.00
North America	93.48
South America	94.92

Table 1: Repeat Customer Rates Across Regions

2.5 CLV (Customer Lifetime Value) by region

The chart and data highlight significant regional differences in Customer Lifetime Value (CLV). South America leads with the highest average CLV at 3,717.84 [dollars], indicating a strong customer retention and revenue generation potential in the region. This could result from effective marketing strategies, higher average customer spending, or longer customer relationships. Asia, Europe, and North America show slightly lower but relatively close CLV figures, ranging between 3,311and3,379. These regions may benefit from targeted strategies to close the gap with South America, such as personalized customer engagement, loyalty programs, or enhanced product offerings. The consistency across regions reflects a relatively uniform market approach but suggests room for optimization in lower-performing regions. Understanding the underlying drivers of South America’s success could provide valuable insights for improving performance in other markets.

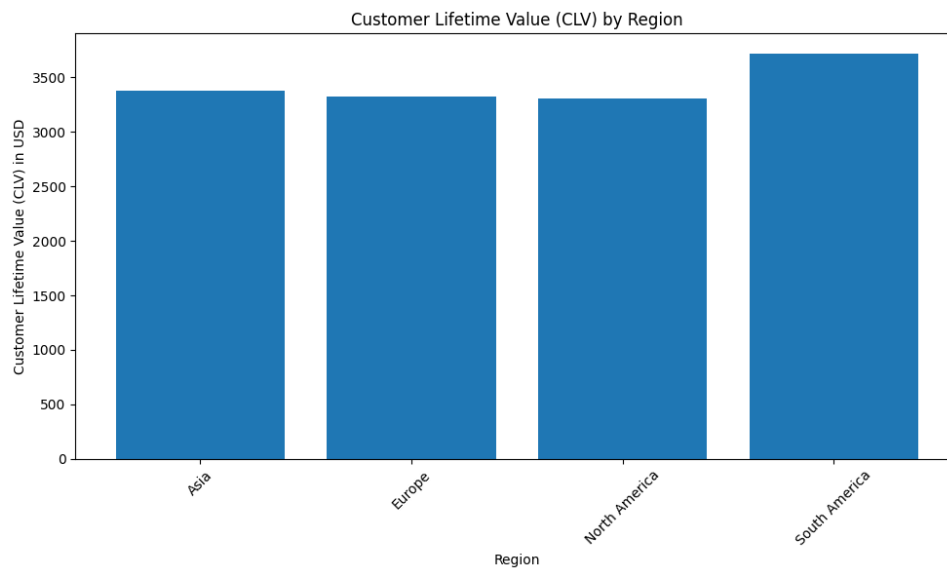


Figure 7: Quantity vs Revenue by Category

Region	Customer Lifetime Value (CLV)
Asia	3379.44
Europe	3325.09
North America	3311.16
South America	3717.84

Table 2: Customer Lifetime Value (CLV) by Region