# *Exploratory Data Analysis of eCommerce Transactions Dataset*

Introduction:

This report presents key findings from an Exploratory Data Analysis (EDA) conducted on an eCommerce transactions dataset. The analysis aims to extract meaningful business insights using the data comprising of customer information, product details, and transaction records. The analysis is based on an eCommerce transactions dataset, comprising of three CSV files: `Customers.csv`, `Products.csv`, and `Transactions.csv`. The Customers.csv dataset contains customer information, such as ID, name, region, and signup date. The Products.csv dataset includes product details like ID, name, category, and price. The Transactions.csv dataset stores information about individual transactions like transaction IDs, customer and product IDs, date, quantity, and total transaction value. The primary objective is to provide actionable recommendations that can inform strategic decisions and drive business growth.

Key Business Insights from the EDA:

➢ Top-Revenue Generating Categories:
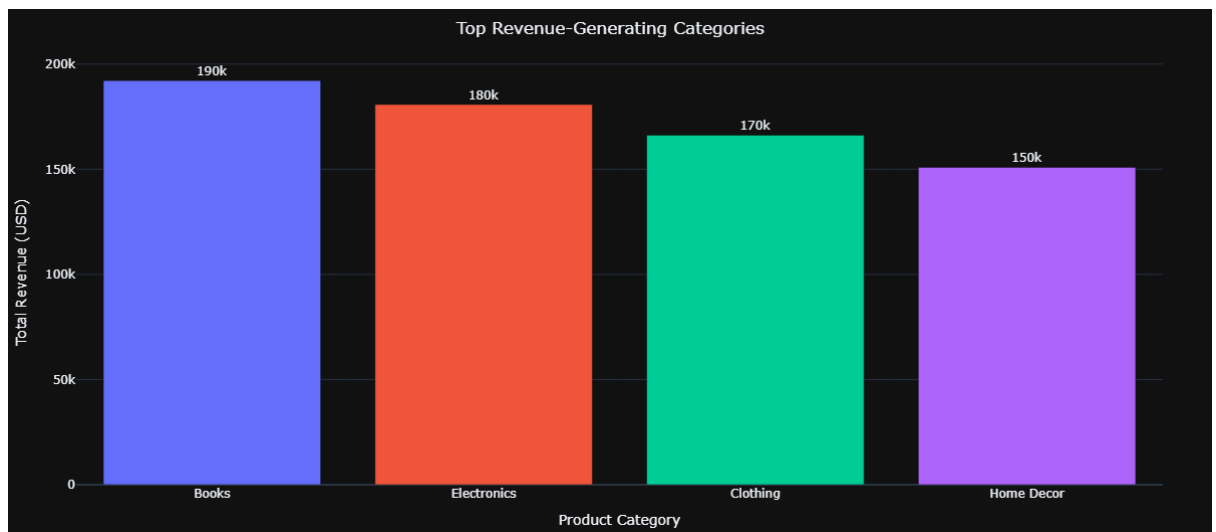
- **Finding:** The "Books" category generates the highest revenue, exceeding 190k, followed by "Electronics" with approximately 180k. Conversely, the "Home Decor" category generates the least revenue at 150k.
- Implication: This suggests that a large proportion of the total revenue is generated from Books and Electronics categories while the Home Decor category is under performing
- Recommendation: The company should prioritize marketing and inventory strategies towards 'Books' and 'Electronics', while putting further emphasis on the Home Decor category to investigate the reasons for underperformance.

➢ Region-Wise Revenue:

Revenue by Region
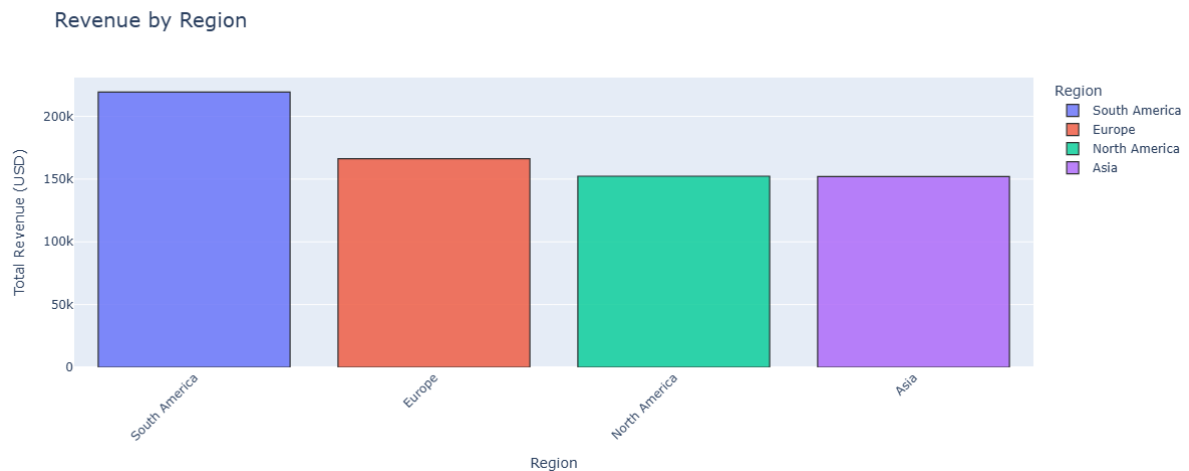
Findings: South America generates the highest revenue, at approximately 219k, which is considerably higher than Europe, which contributes approximately 166k; while North America and Asia regions both contribute the least, generating a similar revenue of about 152k.

Implication: This indicates that South America is the most significant market while the other two markets require more focused attention.

Recommendation: The company should tailor marketing strategies towards South America and further analyse North America and Asia to identify ways to boost sales.

➢ Revenue Contribution:

Revenue Contribution: Top 5% Customers vs Others
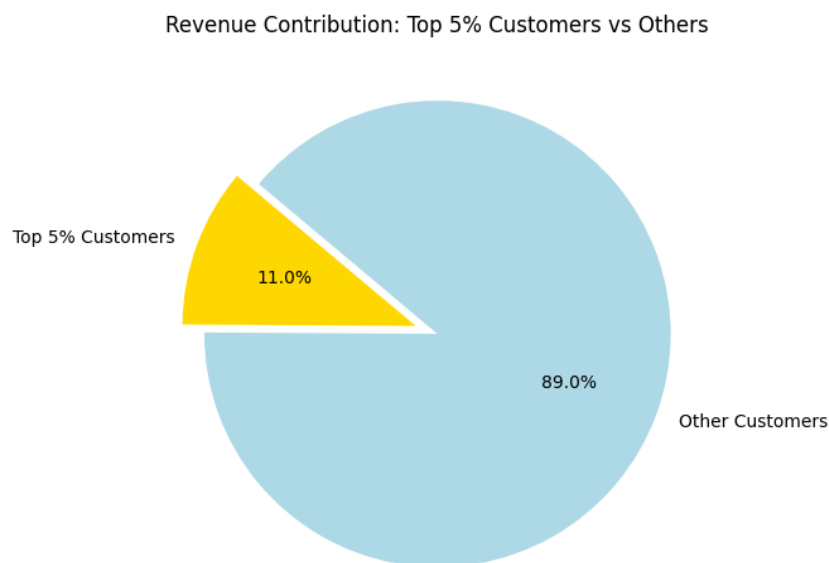
The pie chart visually represents the revenue contribution of two groups of customers: "Top 5% Customers" and "Other Customers."

- Top 5% Contribution: The "Top 5% Customers" segment contributes 11.0% of the total revenue. The segment is visually marked by the use of a different colour and by an offset.

- Other Customers Contribution: The "Other Customers" segment makes up the remaining 89.0% of the total revenue.

o **Finding:** While the top 5% of customers contribute 11.0% of the total revenue, the remaining 95% of 'Other Customers' account for most of the revenue, at 89.0%.

o **Implication:** This indicates that although a small segment of customers generates a noticeable portion of the total revenue, a large percentage of revenue is also coming from the rest of the customers. This suggests that while top customers are valuable, focusing solely on them would not generate sufficient revenue and the other segments need to be focussed on as well.

o **Recommendation:** The company should maintain focus on top customers using loyalty programs and personalized engagement to encourage their spending, while also exploring ways to increase the spending of the other segments. This may include exploring options such as targeted promotions or increased customer segmentation to cater towards the unique behaviour of different segments.

➢ **Seasonal Sales Patterns:**

Monthly Sales Trends
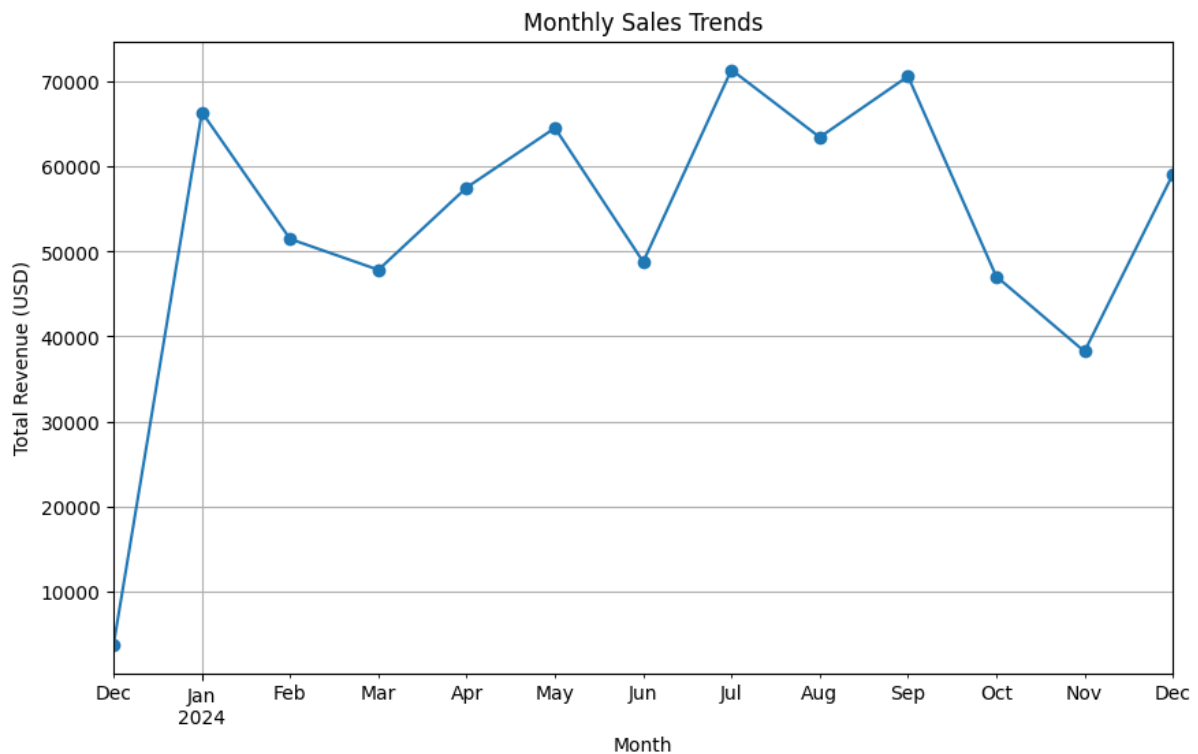
The line plot depicts the total revenue generated each month, spanning from December 2023 to December 2024. Based on the plot and data observations:

- Starting Point: The sales are at its lowest in December 2023 with only a revenue of approximately 3.7K.

- Initial Peak: In January 2024, there is a sharp increase with sales reaching a high of approximately 66.4K, followed by a drop in sales from February to March.

- Mid-Year Highs: The revenue starts picking up again in the subsequent months and stays considerably high from May to July. Sales peak in July 2024, reaching a high of approximately 71.4K.

- Year-End Drop and Recovery: The sales drop again after July and then increases in December 2024 reaching a high of approximately 59K.

"Monthly sales show a clear seasonal pattern, starting low in December 2023 before a strong increase in January 2024, followed by a mid-year peak in July 2024, and another recovery towards the end of December 2024. There are dips in sales during February-March and August to November."

◯ **Finding:** Monthly sales show a clear seasonal pattern, starting low in December 2023 before a strong increase in January 2024, followed by a mid-year peak in July 2024, and another recovery towards the end of December 2024. There are dips in sales during February-March and August to November.

◯ **Implication:** This implies the presence of a seasonal trend in consumer demand for the products, requiring the company to plan inventory and marketing accordingly.

◯ **Recommendation:** The company should ensure sufficient inventory is available before the high-demand months (January and July), and to further plan for strategic marketing campaigns during the periods of low sales.

➢ **Analysis of Revenue by Category and Region**

Revenue by Category and Region
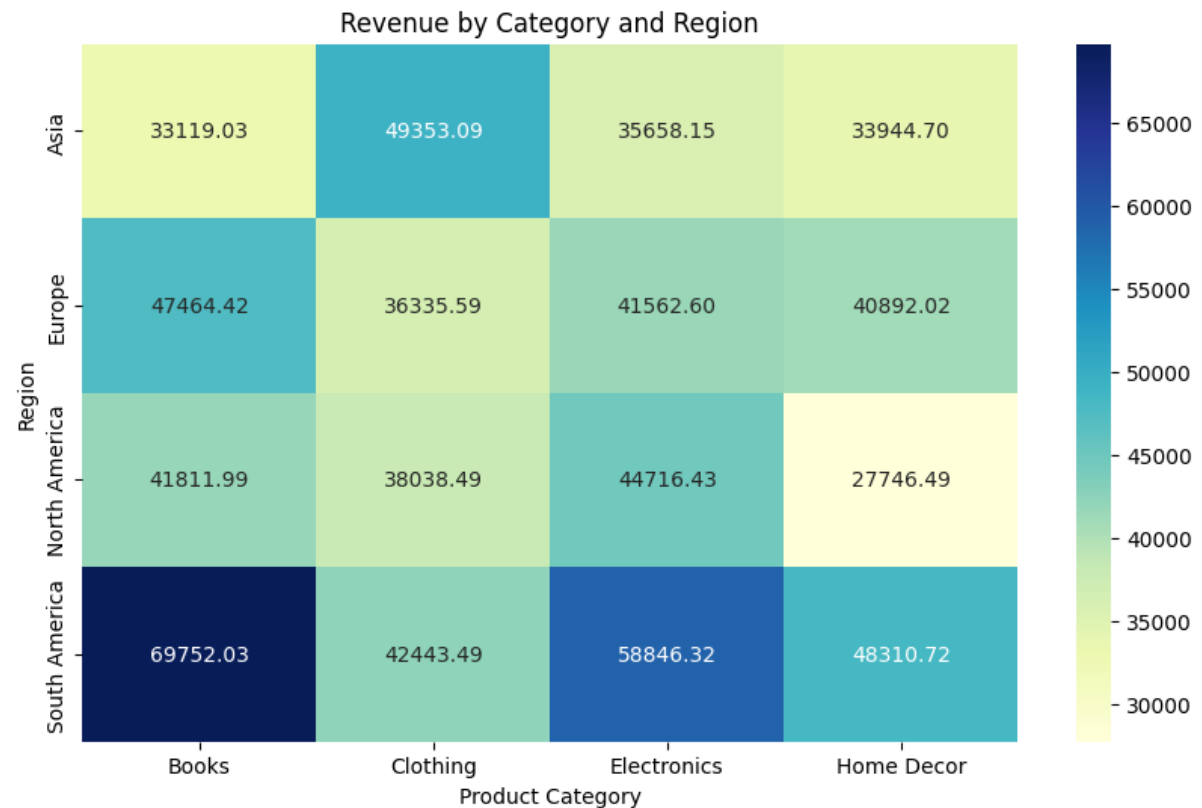
The data represents a breakdown of revenue for each product category within each region. Key observations are:

- **South America's Strength:** South America consistently generates higher revenue across all categories compared to other regions. This suggests a strong market presence in South America.

- **Regional Category Preferences:**

  o **Books**: South America generates the highest revenue with Books as the top-performing category. Europe is also a strong market for books.

  o **Electronics**: South America is the highest revenue generator for Electronics and other categories. North America and Europe both contribute a substantial amount of revenue in this category

  o **Clothing**: Asia has the best-performing market for Clothing with South America as a close second.

  o **Home Decor**: Europe is the second strongest market for Home Decor.

◯ **Finding:** While South America has the highest revenue across all categories, a more granular analysis of the regional performance shows that Books and Electronics are the top revenue generating categories in South America and Europe, while Asia and North America have slightly different preferences with Clothing and Electronics as top performers, respectively.

◯ **Implication:** This implies that different regions have different preferences for the type of products they buy, requiring a tailored approach to cater towards these varied demands.

◯ **Recommendation:** The company should focus on region-specific marketing campaigns that highlight the best-selling categories for that particular region.

Conclusion:

These insights underscore the need for the company to adopt a targeted approach to its sales strategy, marketing, and inventory management. It highlights both key areas of success and also areas where strategic improvement is needed. Based on this analysis, the company should prioritize its efforts in the following areas: investment towards top performing categories (Books and Electronics), targeted marketing towards South America, loyalty programs for top customers, and stock up strategically before peak sales periods. Finally, a granular analysis of categories by region is crucial to target regional preferences.