

Zeotap – Data Scientist Intern EDA assignment

Submission By:- Ansh Phutela (IIT Patna)

Contact:- anshphutela.2012@gmail.com

Dataset Insights:

1. In the given dataset there are **no missing values** in any of the features/input vector.
2. Except **SignupDate feature in customers dataset and TransactionDate feature in transactions dataset**, every feature is in desired and correct datatype. Changing them into pandas's datetime format to allow date-based operations.
3. No important feature's values are duplicated in the given dataset. The Customer_ID's values in customers dataset are all unique similarly ProductId in products dataset and TransactionID in transaction's dataset. There are some CustomerID which are not unique in transaction's dataset but It is valid because same person can buy multiple products and multiple times.

Business Insights:

1. Regional Customer Activity:

South America leads with **59 active customers**, followed by Europe (50) and North America (46). The top region's dominance reflects its strong customer base, providing an opportunity to expand marketing efforts in nearby regions like North America and Asia.

2. Purchasing Trends in South America:

In South America, "**Books**" generate the highest sales(**69,752 USD**), then "Electronics" (58846 USD). Customers in this region purchase an average of 2.5–2.7 units per transaction, indicating a preference for moderate quantities across all categories.

3. Top-Performing Categories:

"**Books**" contribute the most to revenue (**27.85%**), followed by "Electronics" (26.20%). Combined with "Clothing" (24.08%) and "Home Decor" (21.87%), these categories dominate sales, suggesting potential in expanding these product lines for further growth.

4. Seasonal Sales Peaks:

The months of **July, September, and January witness the highest sales**, with **July 2024 leading at \$71,366**. This pattern highlights seasonal opportunities for promotional campaigns and inventory planning during these peak periods.

5. Customer Retention Insights:

Repeat customers dominate revenue generation, contributing 99.08% of total sales, while **one-time buyers contribute less than 1%**. This underscores the importance of loyalty programs and retention strategies to sustain and grow the existing customer base.