

Task 1: Exploratory Data Analysis (EDA) and Business Insights

Steps:

1. Load the datasets.
2. Analyze patterns and relationships.
3. Derive five business insights and document them

The datasets are successfully loaded, and here's their structure:

Customers.csv

- **CustomerID**: Unique identifier for customers.
- **CustomerName**: Name of the customer.
- **Region**: The customer's region (e.g., Asia, South America).
- **SignupDate**: Date when the customer signed up.

Products.csv

- **ProductID**: Unique identifier for products.
- **ProductName**: Name of the product.
- **Category**: Product category (e.g., Books, Electronics).
- **Price**: Price of the product.

Transactions.csv

- **TransactionID**: Unique identifier for transactions.
- **CustomerID**: Links the transaction to a customer.
- **ProductID**: Links the transaction to a product.
- **TransactionDate**: Date and time of the transaction.
- **Quantity**: Number of items bought.
- **TotalValue**: Total cost (Quantity * Price).
- **Price**: Price of the product at the transaction.

Next Step

Let's proceed with Exploratory Data Analysis (EDA) for Task 1 to identify patterns and extract meaningful business insights.

Analyzed

Data Overview

Summary Statistics

- **Transactions.csv:**
 - Average quantity purchased: ~2.54 items.
 - Average transaction value: ~₹690.
 - Price range of products: ₹16 to ₹498.
- **Products.csv:**
 - Average product price: ₹268.
 - Categories: Books, Electronics, Home Decor, Clothing.

Missing Values

- No missing values in any dataset.

Unique Values

- **Regions:** South America, Asia, North America, Europe.
- **Product Categories:** Books, Electronics, Home Decor, Clothing.

Business Insights from the Data:

1. **Customer Distribution by Region**
 - The highest number of customers comes from **South America (59 customers)**, followed by **Europe (50 customers)**. Asia and North America have similar customer counts (~45-46).
 - Suggests focusing regional marketing efforts on South America and Europe.
2. **Top 5 Most Purchased Products**
 - Top products by quantity sold:
 - **SoundWave Jeans**
 - **SoundWave Cookbook**
 - **TechPro Headphones**
 - **ActiveWear Rug**
 - **HomeSense Desk Lamp**
 - These products drive volume sales and could be further promoted.
3. **Revenue Contribution by Product Category**
 - Categories generating the most revenue:
 - **Books: ₹192,147**
 - **Electronics: ₹180,783**
 - **Clothing: ₹166,171**
 - **Home Decor: ₹150,894**
 - Books are leading, suggesting significant customer preference in this segment.
4. **Average Spend Per Region**
 - South America customers have the **highest average transaction value (₹722)**, followed by Europe and Asia. North America has the lowest (₹624).
 - Upselling and premium product strategies might work best in South America.
5. **Monthly Sales Trends**
 - Sales dipped in **October and November 2024** but recovered in **December 2024**.

- Suggests seasonal factors or campaigns in December improved revenue.

If you'd like, I can provide a detailed report for these insights or proceed with Task 2: Building the Lookalike Model. Let me know!

You said: