**Overview of E-Commerce Datasets**

Below is a comprehensive description of the six datasets used in this e-commerce analysis. Each dataset focuses on a distinct aspect of the business, providing insights into transactions, customers, products, competitors, market trends, and marketing campaigns. This document outlines the typical columns (features) in each dataset, along with a brief description and data type guidance.

**1. transactions.csv**

• Purpose:  
Captures information about every transaction made within the e-commerce platform.

• Typical Columns:

1. Transaction\_ID  
   – Description: Unique identifier for each transaction (primary key)  
   – Example: "TX12345"  
   – Data Type: String (categorical identifier)
2. DateTime  
   – Description: Date and time of the transaction  
   – Example: "2025-01-15 14:35:22"  
   – Data Type: DateTime
3. Customer\_ID  
   – Description: Unique identifier for the customer making the transaction (foreign key to Customers dataset)  
   – Example: "CUST1001"  
   – Data Type: String
4. Product\_ID  
   – Description: Unique identifier for the purchased product (foreign key to Products dataset)  
   – Example: "PROD567"  
   – Data Type: String
5. Quantity\_Purchased  
   – Description: Number of product units purchased in the transaction  
   – Example: 2  
   – Data Type: Integer
6. Base\_Price  
   – Description: Regular (list) price of the product at the time of purchase  
   – Example: 49.99  
   – Data Type: Float
7. Final\_Price  
   – Description: Actual price paid by the customer after discounts or promotions  
   – Example: 39.99  
   – Data Type: Float
8. Discount\_Offered  
   – Description: Percentage or amount of discount provided on the purchase (if any)  
   – Example: 20 (%)  
   – Data Type: Numeric
9. Gross\_Revenue  
   – Description: Total revenue (Quantity\_Purchased × Final\_Price)  
   – Example: 79.98 (when 2 × 39.99)  
   – Data Type: Float
10. Net\_Profit  
    – Description: Estimated profit after deducting cost of goods sold and other direct transaction-related costs  
    – Example: 35.00  
    – Data Type: Float
11. Category  
    – Description: Product category classification (e.g., “Electronics”, “Apparel”, etc.)  
    – Example: "Electronics"  
    – Data Type: String (categorical)
12. Region  
    – Description: Geographical region or market where the transaction took place (if applicable)  
    – Example: "North America"  
    – Data Type: String (categorical)

**2. customers.csv**

• Purpose:  
Contains demographic, behavioral, and segmentation data about customers.

• Typical Columns:

1. Customer\_ID  
   – Description: Unique identifier for each customer (primary key)  
   – Example: "CUST1001"  
   – Data Type: String
2. Name  
   – Description: Customer’s name (optional if privacy is a concern)  
   – Example: "John Doe"  
   – Data Type: String
3. Email  
   – Description: Customer’s email address (optional if privacy is a concern)  
   – Example: "[john.doe@example.com](mailto:john.doe@example.com)"  
   – Data Type: String
4. Region  
   – Description: Geographical location/region of the customer  
   – Example: "North America"  
   – Data Type: String (categorical)
5. Loyalty\_Score  
   – Description: A numeric measure of the customer's loyalty (can be based on purchase frequency, total spend, etc.)  
   – Example: 85  
   – Data Type: Integer or Float
6. Purchase\_Frequency  
   – Description: Estimated average number of purchases per month/quarter  
   – Example: 2.5  
   – Data Type: Float
7. Average\_Order\_Value  
   – Description: Average monetary value of the customer's orders  
   – Example: 62.5  
   – Data Type: Float
8. Created\_At  
   – Description: Date the customer account was created  
   – Example: "2023-05-10"  
   – Data Type: DateTime
9. Segment  
   – Description: Customer segment label (e.g., “VIP”, “Regular”, “Occasional Buyer”)  
   – Example: "VIP"  
   – Data Type: String (categorical)

**3. products.csv**

• Purpose:  
Lists all products available on the platform, including key details such as cost, category, and pricing.

• Typical Columns:

1. Product\_ID  
   – Description: Unique identifier for each product (primary key)  
   – Example: "PROD567"  
   – Data Type: String
2. Product\_Name  
   – Description: Descriptive product name  
   – Example: "Wireless Earbuds"  
   – Data Type: String
3. Category  
   – Description: Product category classification  
   – Example: "Electronics"  
   – Data Type: String (categorical)
4. Base\_Price  
   – Description: Standard retail price for the product  
   – Example: 49.99  
   – Data Type: Float
5. Cost\_of\_Goods\_Sold  
   – Description: Direct cost incurred for producing or sourcing the product  
   – Example: 25.00  
   – Data Type: Float
6. Stock\_Quantity  
   – Description: Current inventory level  
   – Example: 350  
   – Data Type: Integer
7. Supplier  
   – Description: Source or supplier of the product  
   – Example: "Tech Supply Inc."  
   – Data Type: String
8. Created\_At  
   – Description: Date the product was added to the catalog  
   – Example: "2022-11-01"  
   – Data Type: DateTime

**4. competitor\_data.csv**

• Purpose:  
Tracks competitor pricing, offerings, and market positioning to assist in competitive analysis.

• Typical Columns:

1. Competitor\_ID  
   – Description: Unique identifier for the competitor  
   – Example: "COMP101"  
   – Data Type: String
2. Competitor\_Name  
   – Description: Name of the competitor  
   – Example: "ShopSmart"  
   – Data Type: String
3. Product\_ID  
   – Description: Product identifier aligned to our own product catalog for comparison (if known)  
   – Example: "PROD567"  
   – Data Type: String
4. Competitor\_Price  
   – Description: The competitor's price for a specific product (matched by Product\_ID or product characteristics)  
   – Example: 45.99  
   – Data Type: Float
5. Last\_Updated  
   – Description: The date/time this competitor data was last refreshed  
   – Example: "2025-01-14 10:20:00"  
   – Data Type: DateTime
6. Region  
   – Description: If region-specific competitor data applies  
   – Example: "Europe"  
   – Data Type: String (categorical)
7. Additional\_Notes  
   – Description: Observations regarding competitor promotions, brand positioning, etc.  
   – Example: "Seasonal discount"  
   – Data Type: String

**5. market\_trends.csv**

• Purpose:  
Provides macro-level market information, seasonal trends, and external factors influencing customer buying behavior.

• Typical Columns:

1. Trend\_ID  
   – Description: Unique identifier for each recorded market trend  
   – Example: "TD202501"  
   – Data Type: String
2. Trend\_Description  
   – Description: Brief summary of the observed trend  
   – Example: "Holiday Shopping Surge"  
   – Data Type: String
3. Start\_Date  
   – Description: The date from which the trend started to take effect  
   – Example: "2024-12-01"  
   – Data Type: DateTime
4. End\_Date  
   – Description: The date the trend is expected to end (if applicable)  
   – Example: "2025-01-10"  
   – Data Type: DateTime
5. Impact\_Region  
   – Description: Geographic region impacted by the trend  
   – Example: "Global"  
   – Data Type: String (categorical)
6. Impact\_Metric  
   – Description: A measure indicating the impact level or magnitude (could be an index or percentage change)  
   – Example: 1.2 (representing a 20% increase in sales due to the trend)  
   – Data Type: Float or custom metric scale
7. Notes  
   – Description: Additional context on how the trend might affect demand, pricing, or consumer behavior  
   – Data Type: String

**6. marketing\_campaigns.csv**

• Purpose:  
Documents marketing campaign details, including promotions, targeted customer segments, and performance metrics.

• Typical Columns:

1. Campaign\_ID  
   – Description: Unique identifier for each marketing campaign  
   – Example: "MC202513"  
   – Data Type: String
2. Campaign\_Name  
   – Description: Brief name or title of the campaign  
   – Example: "New Year Sale"  
   – Data Type: String
3. Start\_Date  
   – Description: Launch date of the marketing campaign  
   – Example: "2025-01-01"  
   – Data Type: DateTime
4. End\_Date  
   – Description: Last date the campaign remains active  
   – Example: "2025-01-10"  
   – Data Type: DateTime
5. Target\_Segment  
   – Description: Intended audience of the campaign (e.g., “High Loyalty Customers”, “New Customers”)  
   – Example: "VIP"  
   – Data Type: String (categorical)
6. Discount\_Offer  
   – Description: Percentage or amount of discount/promotion in the campaign  
   – Example: 15 (%)  
   – Data Type: Numeric
7. Budget  
   – Description: Monetary allocation for the campaign  
   – Example: 5000.00  
   – Data Type: Float
8. ROI  
   – Description: Return on Investment, can be computed after campaign completion  
   – Example: 1.4 (representing 140% return)  
   – Data Type: Float
9. Engagement\_Metric  
   – Description: A measure of campaign engagement (could be clicks, conversions, impressions, or other specialized metrics)  
   – Example: 200 (clicks)  
   – Data Type: Numeric
10. Notes  
    – Description: Any additional observations or details regarding the campaign performance  
    – Data Type: String