**E-commerce Dataset Documentation**

**Overview**

This synthetic dataset simulates a comprehensive e-commerce platform's operations over a two-year period. It contains interrelated data tables covering transactions, customers, products, competitor information, market trends, and marketing campaigns. The dataset is designed to enable sophisticated business analytics and machine learning applications.

**Dataset Structure**

The dataset consists of six main tables:

**1. Transactions Table**

Primary table containing all purchase events and related metrics.

**Key Fields:**

* **Transaction\_ID**: Unique identifier for each transaction
* **Customer\_ID**: Links to Customers table
* **Product\_ID**: Links to Products table
* **DateTime**: Timestamp of the transaction
* **Base\_Price**: Original product price
* **Final\_Price**: Price after discounts
* **Quantity\_Purchased**: Number of items bought

**Financial Metrics:**

* **Gross\_Revenue**: Total revenue before deductions
* **Net\_Revenue**: Revenue after processing fees
* **Total\_Cost**: Combined costs (product, shipping, processing)
* **Net\_Profit**: Profit after all costs
* **Profit\_Margin**: Percentage profit margin
* **Processing\_Fee**: Payment processing costs

**Operational Metrics:**

* **Shipping\_Method**: Chosen delivery method
* **Shipping\_Cost**: Cost of delivery
* **Estimated\_Delivery\_Days**: Expected delivery timeframe
* **Return\_Status**: Whether item was returned
* **Return\_Reason**: Reason for return if applicable
* **Return\_Processing\_Cost**: Cost of handling returns

**Customer Interaction:**

* **Device\_Used**: Purchase platform
* **Time\_on\_Site**: Session duration
* **Abandoned\_Cart\_Items**: Items left unpurchased
* **Cart\_Abandonment\_Reason**: Why cart was abandoned
* **Rating\_Given**: Product rating
* **Customer\_Service\_Contacted**: Support interaction flag

**2. Customers Table**

Contains customer profiles and behavioral data.

**Demographics:**

* **Customer\_ID**: Unique identifier
* **Age\_Group**: Customer age range
* **Gender**: Customer gender
* **Region**: Geographic location
* **Language**: Preferred language

**Value Metrics:**

* **Loyalty\_Score**: 0-100 loyalty rating
* **Customer\_Segment**: High/Medium/Low value
* **Customer\_Lifetime\_Value**: Predicted total value
* **Average\_Order\_Value**: Typical purchase amount
* **Purchase\_Frequency**: Purchase regularity

**Behavioral Data:**

* **Discount\_Sensitivity**: Price sensitivity level
* **Preferred\_Payment**: Payment method preference
* **Email\_Subscription**: Marketing opt-in status
* **Mobile\_App\_User**: App usage flag
* **Social\_Media\_Follower**: Social media engagement

**3. Products Table**

Detailed product information and metrics.

**Basic Information:**

* **Product\_ID**: Unique identifier
* **Category**: Product category
* **Brand**: Manufacturer/Brand name
* **Launch\_Date**: Product introduction date

**Pricing and Costs:**

* **Base\_Price**: Standard price
* **Cost\_Price**: Manufacturing/acquisition cost
* **Storage\_Cost**: Warehousing expense
* **Shipping\_Cost**: Standard shipping cost
* **Warranty\_Cost**: Warranty expense

**Inventory Management:**

* **Supplier\_ID**: Supplier identifier
* **Supplier\_Lead\_Time**: Resupply timeframe
* **Reorder\_Point**: Stock level for reordering
* **Min\_Order\_Quantity**: Minimum order amount
* **Stock\_Level**: Current inventory

**Product Details:**

* **Weight\_KG**: Product weight
* **Dimensions\_CM**: Product dimensions
* **Country\_of\_Origin**: Manufacturing location
* **Eco\_Friendly**: Sustainability flag
* **Has\_Warranty**: Warranty availability

**4. Competitor Data Table**

Tracks competitor pricing and offerings.

**Pricing Information:**

* **Competitor\_Base\_Price**: Standard price
* **Competitor\_Final\_Price**: Price after discounts
* **Competitor\_Discount**: Active discount percentage
* **Price\_Difference\_Pct**: Price comparison percentage

**Competitive Metrics:**

* **Market\_Share**: Estimated market share
* **Competitor\_Rating**: Product rating
* **Competitor\_Review\_Count**: Number of reviews
* **Competitor\_Trend**: Price trend direction
* **Delivery\_Time**: Shipping timeframe

**5. Market Trends Table**

Economic and market condition indicators.

**Market Indicators:**

* **Google\_Trend\_Score**: Search interest (0-100)
* **Market\_Sentiment**: Overall market sentiment
* **Social\_Media\_Sentiment**: Social media sentiment
* **Consumer\_Confidence\_Index**: Consumer confidence
* **Online\_Shopping\_Index**: E-commerce activity level

**Economic Indicators:**

* **Inflation\_Rate**: Current inflation rate
* **Disposable\_Income\_Change**: Income trend
* **Unemployment\_Rate**: Current unemployment
* **Raw\_Material\_Cost\_Index**: Material cost trends
* **Exchange\_Rate\_Impact**: Currency effect

**6. Marketing Campaigns Table**

Marketing initiative performance data.

**Campaign Metrics:**

* **Campaign\_ID**: Unique identifier
* **Campaign\_Type**: Campaign category
* **Budget**: Allocated budget
* **Spend**: Actual expenditure
* **Revenue**: Generated revenue
* **ROAS**: Return on ad spend

**Performance Metrics:**

* **Impressions**: Views count
* **Clicks**: Click count
* **CTR**: Click-through rate
* **Conversions**: Successful conversions
* **Conversion\_Rate**: Conversion percentage
* **CPC**: Cost per click
* **CPM**: Cost per thousand impressions

**Data Patterns and Relationships**

**Seasonal Patterns:**

1. Holiday season (November-December): Higher transaction volume
2. Black Friday/Cyber Monday: Peak discounts and sales
3. Summer sale period: Moderate increase in activity
4. Regular weekly patterns: Weekend vs. weekday variations

**Customer Behavior Patterns:**

1. High-value customers: Lower discount sensitivity, higher loyalty
2. Price-sensitive customers: Higher cart abandonment, more discount usage
3. Mobile users: Shorter session times, smaller cart values
4. Desktop users: Longer sessions, larger purchases

**Product Relationships:**

1. Category correlations: Related product purchases
2. Price sensitivity by category
3. Return rates varying by product type
4. Seasonal product popularity

**Analytical Capabilities**

**Customer Analytics:**

* Customer segmentation
* Lifetime value prediction
* Churn risk assessment
* Purchase pattern analysis
* Channel preference analysis

**Product Analytics:**

* Product profitability analysis
* Inventory optimization
* Category performance
* Price elasticity analysis
* Return rate analysis

**Marketing Analytics:**

* Campaign effectiveness
* Channel performance
* ROI analysis
* A/B test results
* Customer acquisition cost analysis

**Market Analysis:**

* Competitor price monitoring
* Market share analysis
* Trend impact assessment
* Seasonal pattern analysis

**Financial Analytics:**

* Revenue forecasting
* Profit margin analysis
* Cost structure analysis
* Payment method analysis
* Pricing optimization

**Technical Details**

**Data Format:**

* All tables available in CSV format
* DateTime fields in ISO format
* Numerical fields use period as decimal separator
* Categorical fields use string values
* Boolean fields use True/False values

**Data Volume:**

* Transactions: 100,000+ records
* Customers: 500,000+ records
* Products: 5,000+ records
* Competitor Data: 300,000+ records
* Market Trends: Daily records for 2 years
* Marketing Campaigns: 100+ campaigns

**Data Quality:**

* No missing values in key fields
* Realistic correlations between metrics
* Consistent data types across tables
* Referential integrity maintained
* Realistic noise and variations

**Usage Guidelines**

**Recommended Analysis Tools:**

* Python (pandas, numpy, scikit-learn)
* R (tidyverse, ggplot2)
* SQL for data querying
* Tableau/Power BI for visualization
* Excel for basic analysis

**Best Practices:**

1. Start with data exploration and cleaning
2. Check for seasonal patterns
3. Consider customer segments in analysis
4. Account for market trends
5. Cross-reference multiple tables for insights

**Common Analysis Scenarios:**

1. Customer Segmentation
2. Demand Forecasting
3. Price Optimization
4. Campaign Performance Analysis
5. Product Performance Analysis
6. Market Basket Analysis
7. Customer Journey Analysis
8. Churn Prediction
9. Inventory Optimization
10. Competitor Analysis