# Description Sales Business

Matias Siles

matiassiles843@gmail.com

Version 1.0

June 30, 2025

# Contents

# 1 Executive Summary

This document provides comprehensive documentation for a business fictitious sales, a multidimensional retail transaction database spanning five years of business operations.

# 2 Dataset Overview

## 2.1 Company Background

This fictitious business is a leading multichannel retail company specializing in consumer electronics, clothing, home goods, and lifestyle products. The company operates in five major global regions and maintains a diverse sales ecosystem that includes online platforms, physical retail stores, mobile applications, and third-party marketplace partnerships.

## 2.2 Dataset Specifications

| Attribute | Value |
|---|---|
| Total Records | ~12,000 transactions |
| Time Period | January 2020 – December 2024 |
| Business Type | Multi-channel retail |
| Geographic Coverage | Global (5 regions) |
| Product Categories | 8 main categories |
| File Format | CSV |
| Data Size | ~15 MB |

Table 1: Dataset Specifications

## 2.3 Business Context

The dataset captures fictitious business performance through various market conditions including:

- COVID-19 pandemic impact (2020-2021)

- Supply chain disruptions (2021-2022)

- Economic recovery periods (2022-2024)

- Seasonal fluctuations and promotional campaigns

- Multi-channel digital transformation initiatives

# 3 Data Schema

## 3.1 Transaction Identifiers

| Column | Type | Description |
|---|---|---|
| order_id | String | Unique identifier for each sales transaction. Format: ORD_YYYYMMDD_XXXX |

| Column | Type | Description |
|---|---|---|
| date | DateTime | Date when the transaction occurred (YYYY-MM-DD format) |

<div align="center">Table 2: Transaction Identifier Columns</div>

## 3.2 Customer Information

| Column | Type | Description |
|---|---|---|
| customer_id | String | Unique customer identifier. Format: CUST_XXXXX |
| customer_age | Integer | Customer's age in years (18-80). |
| customer_gender | String | Customer's gender (Male, Female, Other). |
| customer_segment | String | Customer classification based on purchasing behavior and value |

<div align="center">Table 3: Customer Information Columns</div>

### 3.2.1 Customer Segment Classifications

**Premium** High-value customers with frequent purchases and premium product preferences

**Standard** Regular customers with moderate purchase frequency and balanced price sensitivity

**Budget** Price-sensitive customers seeking deals and promotional offers

**Enterprise** B2B customers with bulk purchases and corporate accounts

## 3.3 Product Information

| Column | Type | Description |
|---|---|---|
| product_category | String | Main product classification (8 categories) |
| product_name | String | Specific product name/model identifier |

<div align="center">Table 4: Product Information Columns</div>

### 3.3.1 Product Categories

| Category | Description |
|---|---|
| Electronics | Computers, smartphones, tablets, accessories, and consumer electronics |
| Clothing | Apparel, shoes, fashion accessories, and seasonal wear |
| Home & Garden | Furniture, home decor, appliances, and gardening supplies |
| Sports & Outdoors | Fitness equipment, outdoor gear, and recreational products |
| Books | Physical and digital books, educational materials, and publications |
| Health & Beauty | Personal care products, cosmetics, and wellness items |
| Automotive | Car accessories, maintenance tools, and automotive products |
| Toys & Games | Children's toys, board games, video games, and entertainment |

Table 5: Product Category Details

## 3.4 Financial Data

| Column | Type | Description |
|---|---|---|
| unit_price | Float | Price per individual item in USD |
| quantity | Integer | Number of items purchased in the transaction |
| subtotal | Float | Total before discounts and taxes (unit_price × quantity) |
| discount_rate | Float | Percentage discount applied (0.0 to 1.0) |
| discount_amount | Float | Dollar amount of discount applied |
| tax_rate | Float | Tax percentage applied based on region |
| tax_amount | Float | Tax amount in USD |
| shipping_cost | Float | Shipping and handling charges |
| total_amount | Float | Final transaction amount |

Table 6: Financial Data Columns

**Financial Calculation Formula:**

$$\text{total\_amount} = \text{subtotal} - \text{discount\_amount} + \text{tax\_amount} + \text{shipping\_cost} \tag{1}$$

## 3.5 Sales & Marketing

| Column | Type | Description |
|---|---|---|
| sales_channel | String | Channel through which the sale was made |
| payment_method | String | Method used for payment |
| region | String | Geographic region of the sale |
| sales_rep | String | Name of the sales representative handling the transaction |

| Column | Type | Description |
| --- | --- | --- |
| lead_source | String | Original source that brought the customer |

Table 7: Sales & Marketing Columns

### 3.5.1 Sales Channel Distribution

| Channel | Weight | Description |
| --- | --- | --- |
| Online | 40% | Company website e-commerce platform |
| Retail Store | 30% | Physical brick-and-mortar locations |
| Mobile App | 15% | Native mobile application purchases |
| Phone | 10% | Telephone sales and customer service |
| Third Party | 5% | External marketplaces and partner platforms |

Table 8: Sales Channel Distribution

### 3.5.2 Geographic Regions

| Region | Tax Rate | Countries Included |
| --- | --- | --- |
| North America | 8% | USA, Canada, Mexico |
| Europe | 20% | EU countries, UK, Norway, Switzerland |
| Asia Pacific | 10% | Japan, Australia, South Korea, Singapore |
| Latin America | 15% | Brazil, Argentina, Chile, Colombia |
| Middle East & Africa | 5% | UAE, Saudi Arabia, South Africa |

Table 9: Regional Tax Structure

## 3.6 Customer Experience

### 3.6.1 Return Analysis

- **Return Rate:** Approximately 8% of all transactions

- **Return Reasons:** Defective, Wrong Item, Not as Described, Changed Mind, Damaged in Shipping

- **Satisfaction Correlation:** Returned items typically have satisfaction scores of 1-3

## 3.7 Derived Fields

| Column | Type | Description |
| --- | --- | --- |
| year | Integer | Year extracted from transaction date |
| month | Integer | Month extracted from transaction date (1-12) |
| quarter | Integer | Quarter of the year (1-4) |

| Column | Type | Description |
|---|---|---|
| day_of_week | String | Day name (Monday, Tuesday, etc.) |
| is_weekend | Boolean | Whether transaction occurred on weekend |
| profit_margin | Float | Estimated profit margin percentage (0.15-0.45) |
| profit | Float | Estimated profit amount in USD |

Table 11: Derived Fields

# 4 Data Quality Assessment

## 4.1 Missing Data Analysis

The dataset intentionally includes realistic data quality issues commonly found in business datasets:

| Column | Missing % | Business Reason |
|---|---|---|
| customer_age | 5% | Privacy concerns, optional field |
| customer_gender | 3% | Optional demographic information |
| satisfaction_score | 10% | Survey non-response, system issues |
| return_reason | 92% | Only applicable to returned items |

Table 12: Missing Data Summary

## 4.2 Data Quality Issues

1. **Duplicate Records:** Approximately 1% duplicate transactions (common in real-world data)

2. **Outliers:** Occasional extreme values in quantity (>10 items) and pricing (>$1000)

3. **Data Entry Errors:** Rare instances of unusual price points (0.1x or 10x normal values)

4. **Seasonal Patterns:** Built-in seasonality effects requiring careful analysis

5. **Business Logic Constraints:** Some combinations may violate expected business rules

# 5 Business Intelligence Insights

## 5.1 Seasonal Patterns

- **Holiday Season:** 25-40% sales increase in November-December

- **Summer Peak:** Sports & Outdoors category shows 20% increase June-August

- **Back-to-School:** Electronics and Books categories peak in August-September

- **Quarter-End Effects:** Enterprise segment shows increased activity at quarter boundaries

## 5.2 Channel Performance

- **Online Dominance:** 40% of all transactions, highest average order value

- **Mobile Growth:** 15% share with fastest year-over-year growth rate

- **Retail Store:** 30% share, highest customer satisfaction scores

- **Cross-Channel Behavior:** 25% of customers use multiple channels