

Amazon Sales Dataset – Documentation

This document describes the structure, scope, and analytical intent of the Amazon sales dataset. The dataset is designed for exploratory analysis, business intelligence reporting, and SQL-based analytical workflows. It is suitable for use with automated RAG systems, dashboards, and evaluation pipelines.

Dataset Overview

The dataset contains transactional sales records from Amazon, capturing order-level information such as products sold, pricing, quantities, customer segments, shipping details, and revenue. Each row represents a single transaction event.

Business Metrics Supported

- Total Revenue (sum of sales amount)
- Total Orders (count of distinct order IDs)
- Units Sold (sum of quantities)
- Average Order Value (revenue divided by orders)
- Category-level and product-level performance
- Time-based trends (daily, monthly, yearly)

Analytical Constraints

Customer lifetime value, churn analysis, profitability modeling, and marketing attribution are not supported unless explicitly represented in the dataset. All analysis must remain strictly grounded in the provided columns without external assumptions.

Intended Use

This dataset is intended for SQL analytics, dashboarding, reporting automation, and evaluation of retrieval-augmented generation (RAG) systems that operate on structured data. It supports clean aggregation logic and deterministic query execution.