

Example – Dataset Documentation and Specifications

Dataset Name: fact_Sales
Owner: Janet Smith – Senior Data Engineer
Created Date: 12/05/2023
Last Updated: 01/08/2025

Purpose:

The fact_Sales table stores all finalised sales transactions for reporting and analytics. It is the primary source for revenue, sales trends, and product performance dashboards in Power BI.

Source Systems:

- 1. ERP System – Orders Module (ERP_Orders) – Provides base sales transaction data.
- 2. Product Master Database (dim_Product) – Provides product details and standard pricing.
- 3. Customer Database (dim_Customer) – Provides customer information and region codes.

Update Frequency:

ETL Process: Runs daily at 02:00 AM UTC.
Load Type: Incremental load using OrderID as the unique key.

Transformation Logic:

- 1. Filter orders with OrderStatus = 'Completed'.
- 2. Join with dim_Product to add product category and standard price.
- 3. Join with dim_Customer to add customer region and segment.
- 4. Calculate TotalAmount = Quantity × UnitPrice.
- 5. Remove test transactions where CustomerID = 'TEST' or ProductID = 'DEMO'.

Data Dictionary:

Column Name	Data Type	Description	Example
SalesID	INT	Unique identifier for the sale	104593
OrderID	VARCHAR(20)	ERP order reference number	ORD-202508-001
ProductID	VARCHAR(20)	Unique product code from dim_Product	PRD-ACM-123
CustomerID	VARCHAR(20)	Unique customer code from dim_Customer	CUST-00155
OrderDate	DATE	Date when the sale occurred	01/08/2023
Quantity	INT	Number of units sold	12
UnitPrice	DECIMAL(10,2)	Sale price per unit in £ (GBP)	15.99
TotalAmount	DECIMAL(12,2)	Total value of the sale in £ (GBP)	191.88
Region	VARCHAR(50)	Customer region from dim_Customer	West Yorkshire

Quality Checks:

All OrderID values must be unique in the table.
No NULL values in ProductID, CustomerID, OrderDate, or TotalAmount.
TotalAmount must equal Quantity × UnitPrice.
Every CustomerID must exist in dim_Customer.
Every ProductID must exist in dim_Product.

Known Issues:

Historical records before 2019 may have missing Region values.
Orders with discounts greater than 50% should be reviewed for accuracy.

Change Log:

Date	Change Description	Changed By
01/08/2025	Added Region column from dim_Customer.	Janet Smith
15/12/2024	Updated ETL logic to exclude test transactions.	John Lee