

Sales Data Pipeline & Analytics

End-to-end data pipeline and analytics for a global spice-export company

Overview

I built an end-to-end analytics pipeline for a global spice exporter whose sales data lived as raw, messy exports out of an ERP accounting system. The goal was to turn that scattered data into a clean, queryable warehouse that leadership could use to understand sales performance across dozens of export markets and make better pricing and product-mix decisions.

The Problem

Raw ERP sales exports were inconsistent and not analysis-ready — making it slow and error-prone to answer basic questions about how the business was performing across countries, customers, and products.

What I Built

- Wrote Python scripts to parse and clean the raw ERP exports — standardizing formats, handling inconsistencies, and validating the data.
- Designed and loaded the cleaned data into a DuckDB warehouse using a star-schema structure built for analytical querying.
- Wrote SQL (in DBeaver) to analyze sales across dozens of export markets, and built strategic analysis workbooks on top of the warehouse.
- Surfaced insights on revenue trends over time, customer concentration, and shifts in product mix — framed to support pricing and market-mix decisions.

Outcome

Transformed manual, scattered ERP data into a reliable, queryable analytics layer — enabling faster, data-driven answers to strategic questions about the export business. The project covered the full lifecycle: ingestion, cleaning, modeling, querying, and insight delivery.

Tools & Methods

Python (data parsing, cleaning, automation) · SQL · DuckDB (star-schema data warehouse) · DBeaver · Excel (analysis workbooks). Built end-to-end.