Project summary: Retail Sales Analysis

A retail dataset needed to be structured, cleaned, and analyzed to uncover sales performance, customer behavior, and category trends. The goal was to build a MySQL database, ensure data quality, perform exploratory analysis, and answer focused business questions that inform decisions such as seasonality, top customers, and high‑value transactions.

What was done:

* Database setup: Created a dedicated MySQL database and a retail\_sales table capturing transaction details (transaction/date/time, customer demographics, product category, quantity, pricing, COGS, and total sale).
* Data quality checks and cleaning: Assessed record counts, unique customers, and distinct categories. Identified rows with missing key fields and removed incomplete records to maintain analytical integrity.
* Exploratory analysis: Quantified overall records, customer base, and category coverage. Verified table contents and distributions to understand the dataset’s scope and structure.
* Business analysis:
  + Filtered sales by specific dates and months to support calendar‑based insights.
  + Isolated category‑specific performance (e.g., Clothing) with quantity thresholds to gauge demand.
  + Aggregated sales by category to measure revenue and order volume contribution.
  + Profiled customer demographics (e.g., average age for Beauty buyers) to understand audience characteristics.
  + Flagged high‑value transactions to spotlight premium purchases.
  + Counted transactions by gender within categories to explore behavioral differences.
  + Assessed monthly performance and identified the best‑selling month per year to capture seasonality.
  + Ranked customers by total spend to surface top buyers.
  + Measured unique customer penetration by category to gauge breadth of appeal.
  + Segmented orders by time‑of‑day shifts (Morning, Afternoon, Evening) to understand operational and demand patterns.

Results:

* Data readiness: A clean, analysis‑ready retail\_sales table with nulls removed from critical fields, enabling accurate aggregations and comparisons.
* Category performance: Clear view of category‑level revenue and order counts; identification of strong and weaker categories.
* Customer insights: Top‑spending customers highlighted; unique customer counts by category revealed reach; average age profiles informed targeting.
* High‑value sales: Transactions above key thresholds surfaced for premium focus.
* Time trends: Best‑performing months identified for each year, evidencing seasonality and supporting planning.
* Operational timing: Shift‑based order counts showed when demand concentrates across the day.
* Actionable takeaways: Findings support merchandising (focus on strong categories), marketing (target segments by age and category), customer programs (reward top buyers), and operations (staffing by busiest shifts and months).

This deliverable demonstrates end‑to‑end SQL capability in MySQL: data modeling, cleaning, EDA, and business‑oriented analytics that translate to clear, decision‑ready insights.