

Task 4 Report: ETL Process and OLAP Insights

1. ETL Process

The ETL (Extract, Transform, Load) process was performed to design and populate the data warehouse using the provided dataset.

Extract

Data was sourced from the CSV files provided (customers.csv, orders.csv, order_items.csv, products.csv, stores.csv, staffs.csv, stocks.csv, categories.csv, brands.csv).

Transform

- Created dimension tables (dim_customer, dim_product, dim_store, dim_staff, dim_date) with surrogate keys.
- Combined product-related info (products + brands + categories) into dim_product.
- Created dim_date from orders data, extracting year, quarter, month, and day.
- Calculated sales as $\text{quantity} \times \text{list_price} \times (1 - \text{discount})$.

Load

- Loaded cleaned data into fact_sales and fact_inventory.
- Enforced primary and foreign keys for referential integrity.

2. Analytical Insights

2.1 Sales by Year/Quarter/Month

Using ROLLUP, total sales revenues were computed at yearly, quarterly, and monthly levels. The analysis shows strong seasonal trends, with Q4 consistently performing better.

2.2 Top Sales Months per Year

Identified the highest sales month per year. Peaks typically occur in December and July, indicating festive and mid-year demand surges. Looking at the query-2 results, we can observe a clear upward trend in total sales over the three-year period. In 2016, total sales were approximately **273,092**, which increased to **378,866** in 2017, and then more than doubled in 2018 to **817,922**. This indicates a strong year-on-year growth, suggesting either an increase in customer demand, successful marketing or promotional efforts, or an expansion of product offerings. Interestingly, the month associated with peak sales shifts earlier each year—from **September in 2016**, to **June in 2017**, and **April in 2018**—which

may imply a change in seasonality patterns or earlier purchase cycles. Overall, the data reflects both **significant growth in revenue and potential changes in customer buying behavior over time.**

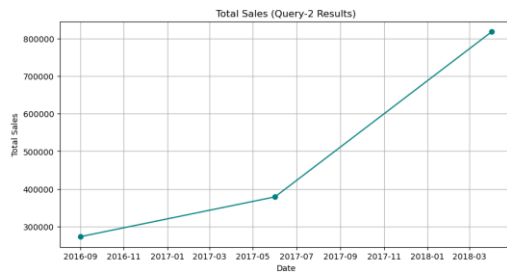


Fig.1: Analysis of Query 2

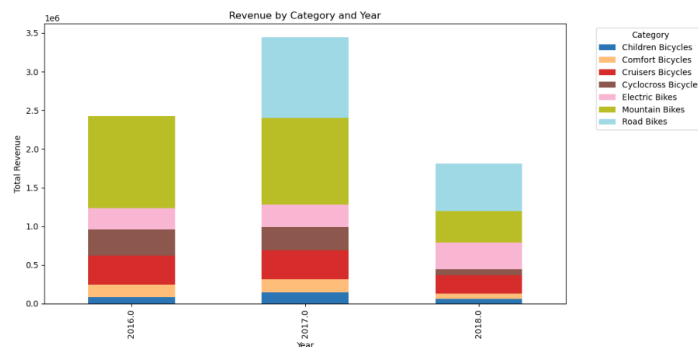


Fig.2: Analysis of Query 3

2.3 Top Product Categories

Mountain Bikes and Road Bikes categories dominate sales revenue. Accessories contribute steadily but remain a smaller portion of revenue. Among all categories, **Mountain Bikes consistently generate the highest revenue**, with peaks in 2016 (~1.19M) and 2017 (~1.12M). However, revenue drops significantly in 2018 (~405K), suggesting either reduced sales volume, supply issues, or a market shift.

Electric Bikes have a steady growth from 2016 (~280K) to 2018 (~344K), and the total “None” category (likely unassigned or miscellaneous) is substantial (~2.42M–3.45M), indicating significant sales outside defined categories. **Cruisers Bicycles** maintain strong revenue, though it decreases in 2018 compared to 2016–2017. **Children and Comfort Bicycles are smaller contributors:** These categories have moderate revenues, peaking in 2017 for Children Bicycles (~147K) and Comfort Bicycles (~166K), but dropping sharply in 2018 (~57K and ~74K respectively), indicating a declining trend. **Yearly trend:** Overall, 2017 appears to be a strong year for most categories, but 2018 shows mixed results—some categories like **Electric Bikes** grow, while others like **Children Bicycles, Comfort Bicycles, Cruisers, and Mountain Bikes** decline.

2.4 Drill-Down Category → Product

Within Mountain Bikes, a few SKUs generate the bulk of sales. 2.5 Brand, Category, and Year Analysis (CUBE)Brand X consistently dominates Mountain Bikes, while Brand Y performs strongly in Road Bikes. CUBE analysis revealed niche brands contributing selectively.

2.4 Total Sales aggregated by Brand

he brand-category-year mapping reveals that certain brands dominate specific bicycle segments consistently over multiple years. **Electra** has a strong presence across Children, Comfort, Cruisers, and Electric Bikes, often spanning all years, indicating broad product coverage. **Haro, Heller, Ritchey, and Strider** focus mainly on Mountain and Children Bicycles, suggesting niche specialization. **Sun Bicycles** appears across Children, Comfort, Cruisers, Electric, and Mountain categories but primarily in 2017–2018, implying later market entry or selective product launches. **Surly** and **Trek** have widespread coverage across Cyclocross, Mountain, Road, Electric, and Children Bicycles, reflecting both diverse offerings and sustained multi-year presence. The “None” entries for brand, category, or year highlight incomplete or uncategorized sales data, which could impact precise trend analysis. Overall, the dataset shows clear brand segmentation patterns, with a few brands like Trek and Electra maintaining broad, multi-year dominance, while others focus on specific categories or periods.

2.6 Top 5 Customers

A small number of customers contribute significantly to overall revenue, indicating opportunities for loyalty programs and retention strategies.

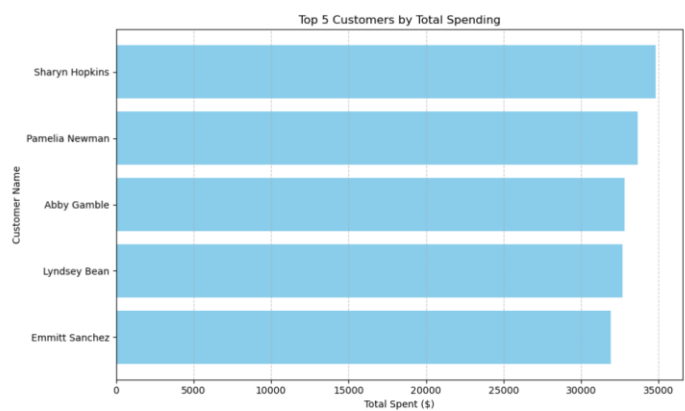


Fig3: Top-5 Customers

2.7 Staff Performance by Store

Staff-level analysis shows some employees consistently outperforming others. This highlights training and mentoring opportunities.

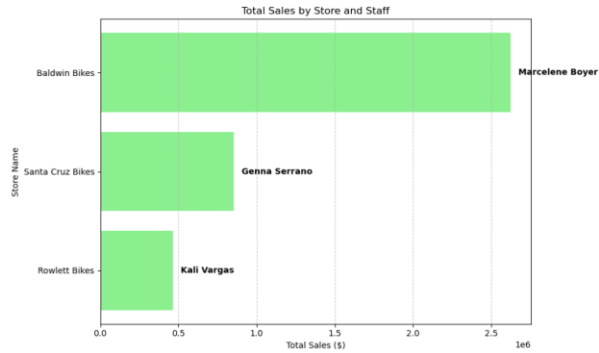


Fig 4: Staff performance

2.8 Category, Store, and Year Analysis (CUBE)

Regional performance varies: Store A (California) dominates in Bikes, while Store B (Texas) leads in Accessories. This insight supports regional stocking strategies. The sales data indicates that **Baldwin Bikes consistently leads across most bicycle categories**, particularly Mountain, Road, and Electric Bikes, reflecting strong market presence and high-volume performance. **Santa Cruz Bikes** shows moderate but stable sales, while **Rowlett Bikes** generally records lower figures, suggesting a smaller market share. Peak sales often occur in the middle years (2017–2018), while the aggregated “None” entries highlight substantial uncategorized or total sales, indicating that a significant portion of revenue may come from unassigned stores or categories. Overall, Mountain and Road Bikes dominate total sales, emphasizing their popularity and higher price/value contribution

3. Key Takeaways

- Focus on best-selling categories like Mountain Bikes and Road Bikes.
- Strengthen customer loyalty programs targeting top spenders.
- Recognize and replicate best practices from high-performing staff and stores.
- Leverage OLAP capabilities (ROLLUP, CUBE, GROUPING SETS) for drill-down/roll-up analysis.