DATACO SUPPLY CHAIN

Python – Cleaning

1. change the column name form Upper with spaces to lower case with underscore (snake_casing)
2. Checking empty column
3. Dropping columns not usable for analysis and dashboard
4. To check number of duplicate rows
5. checking number of duplicate rows
6. Standardising date columns in date format
7. Validating actual shipping days column
8. New column for delay in shipment
9. Standardize categorical text

SQL – Analysis

Theme 1: Supply Chain & Logistics Performance

Focus: Efficiency, delays, and shipping logic.

1. **Geographic Bottlenecks:** Identify the **City** and **Country** with the highest *average* delay variance.

   o *Refinement:* Filter out locations with fewer than 10 orders to avoid statistical noise.

2. **Shipment Mode reliability:** Compare the percentage of Late delivery vs. Shipping on time across different Shipping Mode types (Standard, First Class, etc.).

Theme 2: Financial Health & Profitability

Focus: Margins, losses, and high-value targets.

5. **Market Profitability Ranking:** Identify the most profitable Market (e.g., Europe, LATAM), but rank them by **Profit Margin %** (Profit / Sales), not just total raw profit. This is a better measure of efficiency.

6. **The "Category Star":** Find the Category Name that ranks highest in *both* total Sales and total Benefit (Profit).

7. **Department Performance Extremes:** List the top 10 highest revenue Department Names and the bottom 10, side-by-side.

8. **The "Loss Leader" Analysis (New):** Identify which Category Name has high Sales but negative or very low Benefit per order. These are products that move volume but lose money.

Theme 3: Customer & Segment Behaviour

Focus: Who buys, how much, and retention.

9. **Segment Value Analysis:** Determine which Customer Segment (Consumer, Corporate, etc.) drives the most Revenue and has the highest Order Frequency.

10. **Customer Lifetime Value (CLV) Snapshot:** Calculate the total Sales per Customer Id and sort to find your top 1% of customers.

11. **Fraud Geography:** Identify which Order Country or Order City has the highest percentage of orders marked as SUSPECTED_FRAUD.

Theme 4: Time-Series & Trends (Advanced SQL)

Focus: Growth and seasonality.

12. **MoM and YoY Growth:** Calculate Month-over-Month and Year-over-Year growth for both Revenue and Profit.

   o *Skill Check:* This requires SQL Window Functions (LAG()).

13. **Seasonality Patterns (New):** Calculate average Sales per month across all years to identify peak seasons (e.g., Is December always the highest?).