**Project Title:**  
**Sales, Customer & Discount Dynamics: A Data-Driven Story using SQL + Tableau**

**Problem Statement:**  
This project aims to understand the overall business health of the Superstore by analyzing sales, discounts, profitability, regional performance, and customer contributions. Using SQL for robust data extraction and Tableau for dynamic visualization, we explore how discounts impact profitability, identify key customers and product lines, and evaluate regional performance. The insights help inform data-driven decisions on pricing, customer targeting, and regional strategy.

## **Executive Summary:**

This project presents a comprehensive analysis of Superstore’s transactional data, leveraging SQL for data integrity checks and core analytics, and Tableau for dynamic, interactive visualization of performance patterns.

Through the examination of 2,823 transactions, the project confirms the data quality — with no duplicates or missing critical values — ensuring reliable insights. The SQL analysis uncovers key revenue drivers, top customers, regional dynamics, and loss-making segments. Complementary Tableau dashboards offer interactive views on sales performance, profitability, discount strategies, and customer-level contributions.

Key takeaways include:

* Seasonality patterns, with November emerging as a peak sales month.
* Classic Cars identified as a high-revenue yet loss-making product line.
* EMEA region generating substantial sales but delivering negative or negligible profit, signalling a need for strategic review.
* High discount bands contributing to sales volume but eroding overall profitability.
* Customer-level insights highlighting accounts that drive sales but underperform in profit margin.

Together, these findings enable the business to make informed, data-backed decisions on pricing, discounting, product prioritization, and regional strategies to improve both top-line and bottom-line performance.

### ****Methodology:****

This project combined robust SQL querying with advanced Tableau visualization to analyze sales, discount patterns, profitability, and customer contributions.

SQL was used for:

* Cleaning data (checking duplicates, missing values).
* Aggregating metrics (revenue, profit, discount levels).
* Segmenting results by product line, region, discount band, and customer.

Tableau was used for:

* Creating interactive dashboards with dynamic filters (year, country, product line, territory).
* Visualizing relationships between discount levels and profitability.
* Allowing business users to explore sales trends, customer value, and regional dynamics.

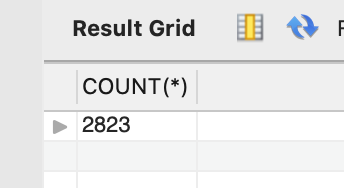
Connection  
A live connection between Tableau Desktop and the MySQL database ensured that all dashboards reflected up-to-date transactional data. The connection was configured with standard authentication, using localhost as the host and port 3306. Filtering was managed entirely within Tableau to retain maximum flexibility at the visualization layer. However Tableau public doesn’t allow live connections so cleaned csv file exported from mysql was uploaded.

# **SQL ANALYSIS AND INSIGHTS:**

## **Data Sanity and Cleaning Checks**

### Row Count Check

SELECT COUNT(\*) FROM superstore\_sales;

****

**Result Summary:**  
The dataset contains a total of 2823 rows.

**Business Interpretation:**  
The dataset consists of 2823 transactional records. This forms the base upon which all further analysis is performed. The size of the dataset is sufficient to derive meaningful business insights related to sales, customers, products, and regional performance.

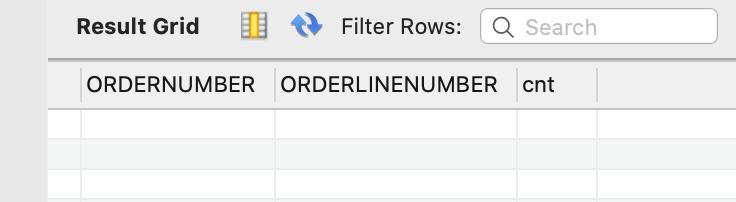
### Duplicate Order Check

SELECT ORDERNUMBER, ORDERLINENUMBER, COUNT(\*) AS cnt

FROM superstore\_sales

GROUP BY ORDERNUMBER, ORDERLINENUMBER

HAVING cnt > 1;

****

**Result Summary:**  
No duplicate records were identified at the order number and order line number level.

**Business Interpretation:**  
The absence of duplicate order records confirms the data integrity at the transactional level. This ensures that subsequent analyses will not be skewed due to repeated entries.

### Missing Values Check

SELECT

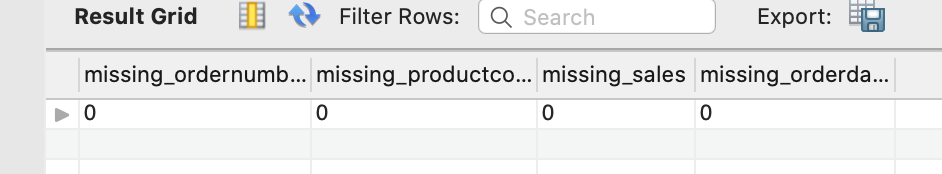
SUM(ORDERNUMBER IS NULL) AS missing\_ordernumber,

SUM(PRODUCTCODE IS NULL) AS missing\_productcode,

SUM(SALES IS NULL) AS missing\_sales,

SUM(ORDERDATE IS NULL) AS missing\_orderdate

FROM superstore\_sales;



**Result Summary:**  
All key fields, including order number, product code, sales, and order date, have zero missing values.

**Business Interpretation:**  
The dataset is complete for critical fields required for sales and profitability analysis. No data cleaning intervention is required for these fields.

### Comprehensive Null Check

SELECT

SUM(ORDERNUMBER IS NULL) AS missing\_ordernumber,

SUM(QUANTITYORDERED IS NULL) AS missing\_quantityordered,

SUM(PRICEEACH IS NULL) AS missing\_priceeach,

SUM(ORDERLINENUMBER IS NULL) AS missing\_orderlinenumber,

SUM(SALES IS NULL) AS missing\_sales,

SUM(ORDERDATE IS NULL) AS missing\_orderdate,

SUM(STATUS IS NULL) AS missing\_status,

SUM(QTR\_ID IS NULL) AS missing\_qtr\_id,

SUM(MONTH\_ID IS NULL) AS missing\_month\_id,

SUM(YEAR\_ID IS NULL) AS missing\_year\_id,

SUM(PRODUCTLINE IS NULL) AS missing\_productline,

SUM(MSRP IS NULL) AS missing\_msrp,

SUM(PRODUCTCODE IS NULL) AS missing\_productcode,

SUM(CUSTOMERNAME IS NULL) AS missing\_customername,

SUM(PHONE IS NULL) AS missing\_phone,

SUM(ADDRESSLINE1 IS NULL) AS missing\_addressline1,

SUM(ADDRESSLINE2 IS NULL) AS missing\_addressline2,

SUM(CITY IS NULL) AS missing\_city,

SUM(STATE IS NULL) AS missing\_state,

SUM(POSTALCODE IS NULL) AS missing\_postalcode,

SUM(COUNTRY IS NULL) AS missing\_country,

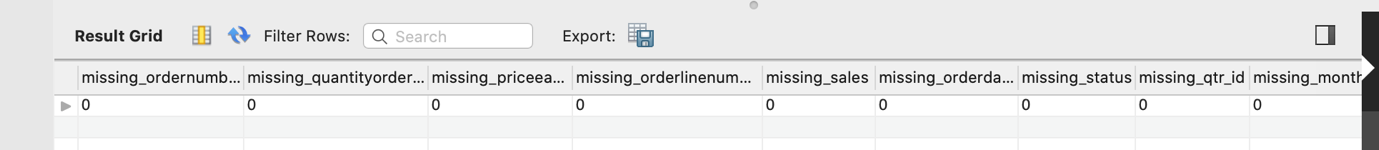
SUM(TERRITORY IS NULL) AS missing\_territory,

SUM(CONTACTLASTNAME IS NULL) AS missing\_contactlastname,

SUM(CONTACTFIRSTNAME IS NULL) AS missing\_contactfirstname,

SUM(DEALSIZE IS NULL) AS missing\_dealsize

FROM superstore\_sales;

****

**Result Summary:**  
All fields in the dataset have zero missing values.

**Business Interpretation:**  
The dataset is robust and reliable. No data gaps exist that could affect our analytical outputs.

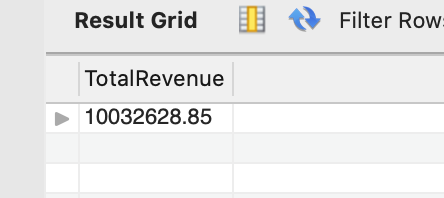
## **Analytical Queries**

### Total Revenue

SELECT

ROUND(SUM(SALES), 2) AS TotalRevenue

FROM superstore\_sales;



**Result Summary:**  
Total revenue generated from all transactions across all years is $10,032,628.85.

**Business Interpretation:**  
This value represents the gross revenue accumulated during the period covered in the dataset. It serves as a reference point to evaluate the contribution of various products, regions, and customer segments.

### Top Ten Best-Selling Products by Revenue

SELECT

PRODUCTCODE,

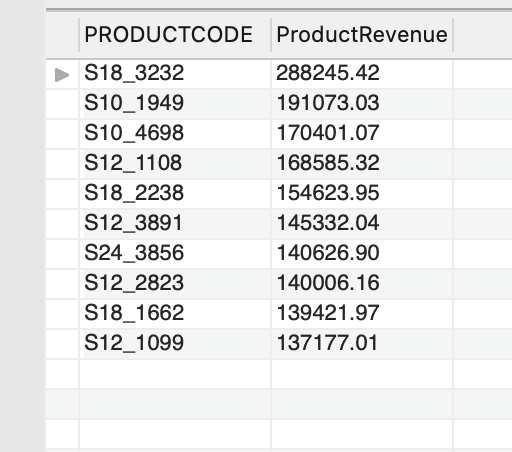
ROUND(SUM(SALES), 2) AS ProductRevenue

FROM superstore\_sales

GROUP BY PRODUCTCODE

ORDER BY ProductRevenue DESC

LIMIT 10;

****

**Result Summary:**  
The highest-grossing product is S18\_3232 with a revenue of $288,245.42, followed by S10\_1949 with $191,073.03, and other products as listed in the result.

**Business Interpretation:**  
These top ten products account for a significant portion of the total revenue. They are key drivers of business performance, and their sales trends warrant close monitoring for inventory and marketing decisions.

### Region-Wise Revenue Summary

SELECT

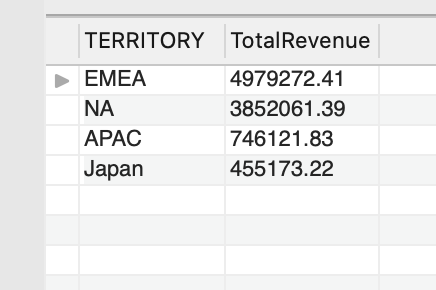
TERRITORY,

ROUND(SUM(SALES), 2) AS TotalRevenue

FROM superstore\_sales

GROUP BY TERRITORY

ORDER BY TotalRevenue DESC;

****

**Result Summary:**  
The EMEA region leads with a revenue of $4,979,272.41, followed by North America at $3,852,061.39, APAC at $746,121.83, and Japan at $455,173.22.

**Business Interpretation:**  
EMEA is the most lucrative market segment, contributing nearly half of the total revenue. Regional strategies could focus on sustaining EMEA performance while exploring growth in the smaller regions.

### Average Discount by Product Line

SELECT

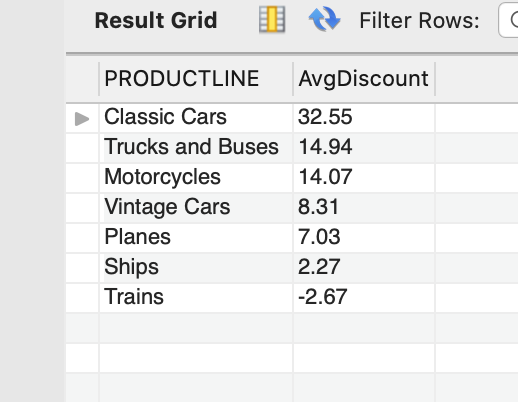
PRODUCTLINE,

ROUND(AVG(MSRP - PRICEEACH), 2) AS AvgDiscount

FROM superstore\_sales

GROUP BY PRODUCTLINE

ORDER BY AvgDiscount DESC;

****

**Result Summary:**  
Classic Cars experience the highest average discount of 32.55%. Trains display a negative average discount, implying price hikes over MSRP for some transactions.

**Business Interpretation:**  
Products with higher discounts may need a profitability review to ensure margins are protected. Classic Cars discounts are notably large and may indicate aggressive pricing strategies.

### Monthly Sales Trend

SELECT

YEAR\_ID,

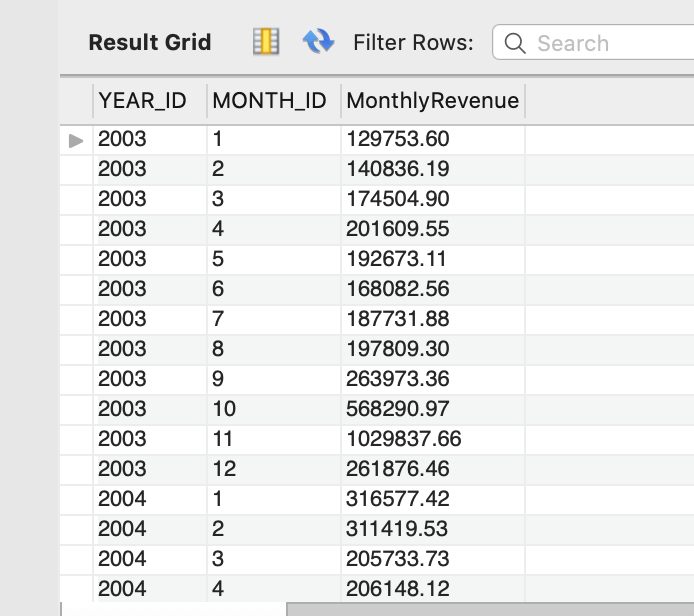
MONTH\_ID,

ROUND(SUM(SALES), 2) AS MonthlyRevenue

FROM superstore\_sales

GROUP BY YEAR\_ID, MONTH\_ID

ORDER BY YEAR\_ID, MONTH\_ID;

****

**Result Summary:**  
Monthly revenues exhibit significant peaks during October and November across years, suggesting seasonal demand patterns.

**Business Interpretation:**  
The business experiences strong seasonality. Sales and marketing teams should align promotions and stock levels to capitalize on these peak months.

### Top Customers by Revenue

SELECT

CUSTOMERNAME,

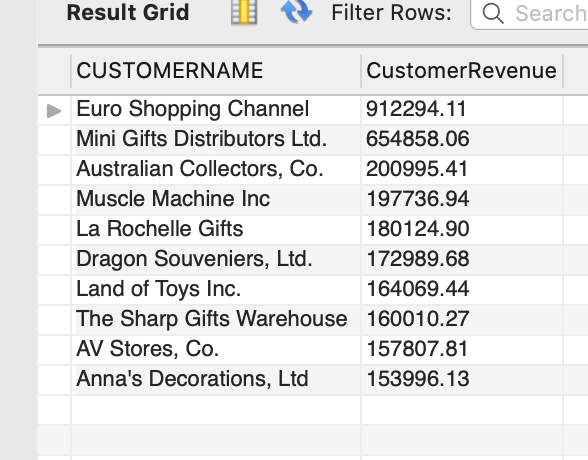
ROUND(SUM(SALES), 2) AS CustomerRevenue

FROM superstore\_sales

GROUP BY CUSTOMERNAME

ORDER BY CustomerRevenue DESC

LIMIT 10;



**Result Summary:**   
The top customer is Euro Shopping Channel with $912,294.11 in revenue. The next significant customer is Mini Gifts Distributors Ltd. at $654,858.06.

**Business Interpretation:**  
A small group of customers generates a substantial portion of revenue. Customer retention strategies should focus on these high-value clients. Further revenue analysis from these customers should be done to check the customer contribution to profit.

### Loss-Making Segments

SELECT

PRODUCTLINE,

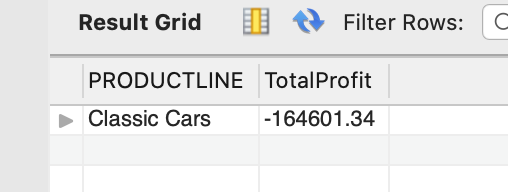
ROUND(SUM(SALES - (QUANTITYORDERED \* MSRP)), 2) AS TotalProfit

FROM superstore\_sales

GROUP BY PRODUCTLINE

HAVING TotalProfit < 0

ORDER BY TotalProfit;

****

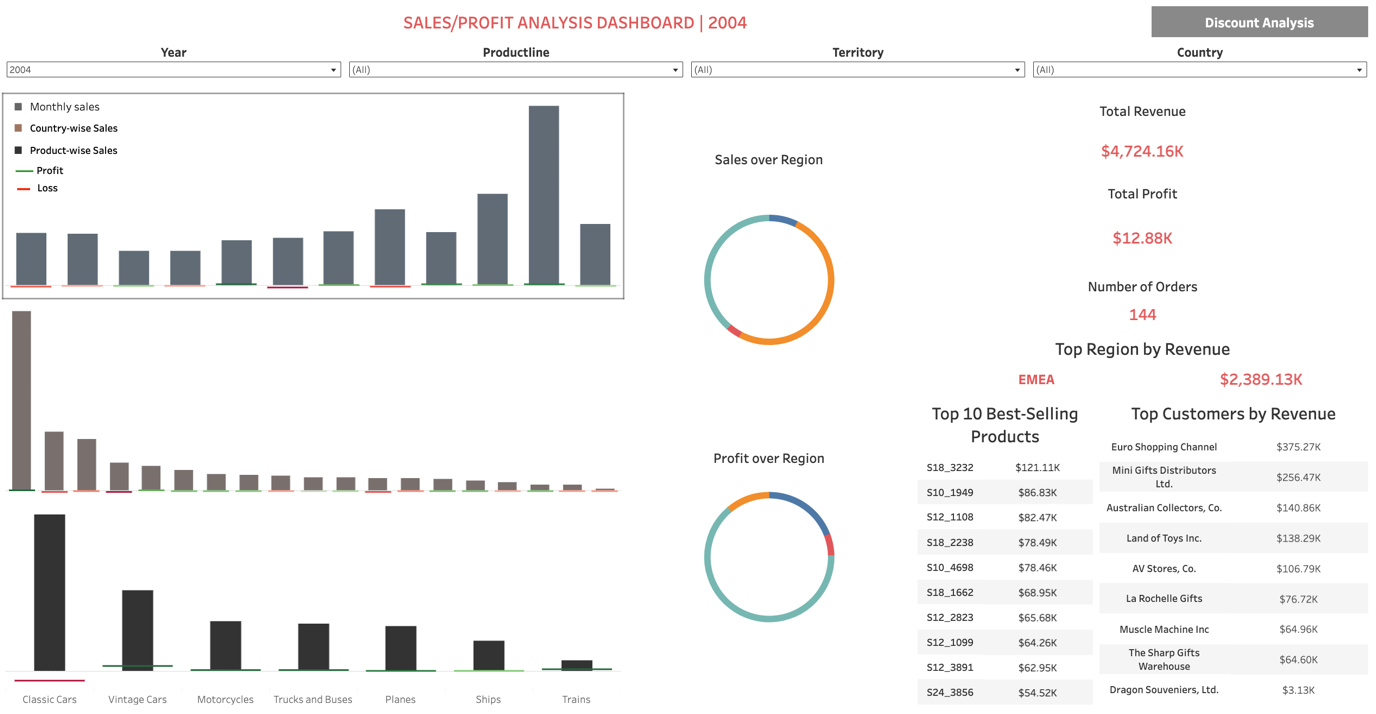
**Result Summary:**  
Classic Cars generated a negative total profit of $164,601.34, indicating the product line operated at a loss.

**Business Interpretation:**  
Classic Cars may be sold at unsustainable discounts or high cost-to-serve levels. This segment should be examined for pricing and cost structure optimization.

## **Tableau analysis :**

## **Link:** <https://public.tableau.com/views/SALES_ANALYSIS_SAMPLE/Dashboard2?:language=en-US&:sid=&:display_count=n&:origin=viz_share_link>

## **Dashboard 1: Sales and Profit Analysis (Dynamic for Year 2004)**



### ****Overview****

This dashboard provides a comprehensive view of sales and profit performance across time periods, regions, countries, product lines, and customers. It enables business users to assess how various dimensions contribute to top-line and bottom-line results, supporting data-driven decisions on pricing, targeting, and product strategy.

### ****Key Filters****

* Year
* Product Line
* Territory
* Country

All filters apply globally across the dashboard to allow multi-level drill-down.

### ****Key Performance Indicators****

* Total Revenue
* Total Profit
* Total Orders
* Top Region by Revenue

### ****Charts and Visual Elements****

1. **Bar Chart 1: Sales and Profit by Month**
   * Sales represented as bars
   * Profit represented as a Gantt bar overlay (green for profit, red for loss)
   * Allows seasonality analysis
2. **Bar Chart 2: Sales and Profit by Country**
   * Highlights country-level performance
   * Red profit bars indicate loss-making countries
3. **Bar Chart 3: Sales and Profit by Product Line**
   * Shows product-level contribution to revenue and profitability
4. **Donut Pie Charts**
   * Sales by Region
   * Profit by Region
5. **Top 10 Products (List Format)**
   * Dynamically updates based on filter selection
   * Identifies key products driving sales and profit
6. **Top 10 Customers (List Format)**
   * Identifies major customers and their contribution to performance

### ****Key Insights & Business Interpretations****

**Seasonality**

* November consistently emerges as the peak sales month across years.

Business decision: The organization should prepare for inventory build-up, promotional campaigns, and logistics scaling ahead of this seasonal spike to maximize revenue and efficiency.

**Product Line Performance**

* **Classic Cars** is a major contributor to sales but is loss-making across most regions.

Business decision: Immediate action is needed to reassess the pricing, discounting, or cost structure for Classic Cars. Alternatively, consider repositioning or rationalizing this product line.

**Regional Insights**

* **EMEA** is the top sales region (50.8 percent of total sales at $2.38 million) but contributes negligibly or negatively to profit (-$1.83k).
* **NA** and **APAC** show better balance between sales and profit.

Business decision: Regional pricing and sales strategies should be realigned. EMEA’s cost structure, pricing, or customer mix needs in-depth review. Perhaps the region is overly discount-dependent or facing competitive pressure that erodes margins.

**Country-Level Patterns**

* **United States** is the top country by sales and shows favorable profit.
* **France** and **Spain** are in the top three for sales but are loss-making.

Business decision: Focus on replicating the US model (where pricing, discounts, or operations are yielding profit) in France and Spain. Consider tightening discounts or reevaluating customer focus in these countries.

**Top Products and Customers**

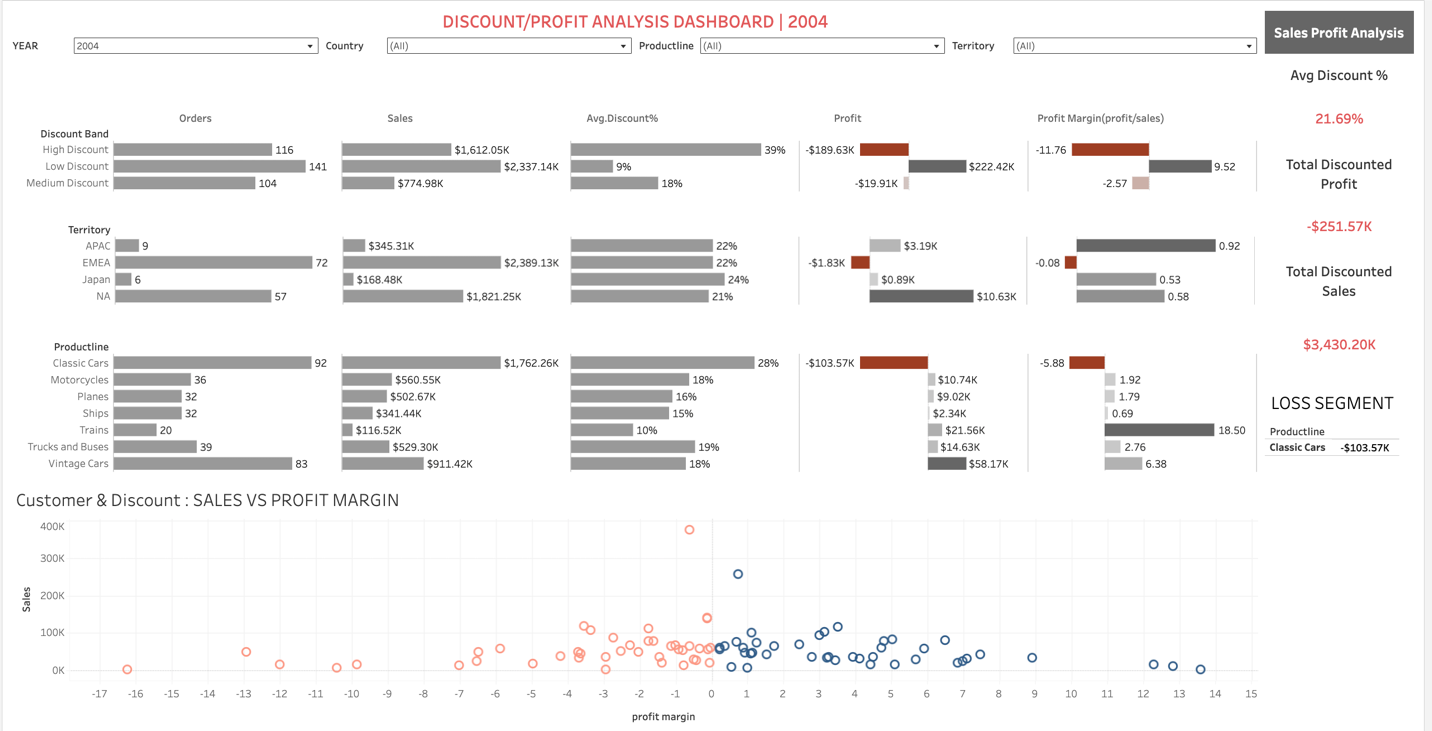
* The Top 10 Product and Customer lists are dynamic tools that help identify high-value products and accounts per selected filters.
* These lists are crucial for targeting sales efforts, cross-sell, and upsell strategies.

Business decision: Build account plans for top customers, focusing on improving their profitability where necessary. Also, prioritize marketing and inventory planning around top-performing products.

### ****Broader Recommendations****

* Use the monthly pattern to guide marketing campaigns and operational planning.
* Revisit regional discount structures, especially in EMEA.
* Prioritize margin improvement strategies for Classic Cars or evaluate exiting unprofitable product lines.
* Institutionalize account-level and product-level profitability reviews using the dashboard outputs.

## **Dashboard 2: Discount and Profitability Analysis (Dynamic for Year 2004)**



### ****Overview****

This dashboard presents an in-depth analysis of the relationship between discounting strategies and business performance across territories, product lines, and customers in the year 2004. It enables stakeholders to identify discount levels that erode profitability and regions or product lines where discounting is misaligned with profit objectives.

### ****Key Filters****

* Year
* Country
* Product Line
* Territory

All charts dynamically respond to these filters, allowing granular insights.

### ****Key Performance Indicators****

* Average Discount Percentage
* Total Discounted Sales
* Total Discounted Profit
* Loss Segment (highlighting the loss-making product line for selected filters)

### ****Charts and Visual Elements****

1. **Discount Band Analysis (Bar Charts)**  
   Compares high, medium, and low discount bands across orders, sales, average discount, profit, and profit margin. Negative metrics (losses) are highlighted in red.
2. **Territory Analysis (Bar Charts)**  
   Displays performance metrics by territory, revealing how discounts impact regional profitability.
3. **Product Line Analysis (Bar Charts)**  
   Evaluates product lines on key metrics, showcasing which lines benefit or suffer under current discounting.
4. **Customer-Level Profitability (Scatter Plot)**  
   Maps customers by sales and profit margin, with color coding (green for profit, red for loss) to highlight profitable versus loss-making customers.

### ****Key Insights & Business Interpretations****

**Discount Band Effectiveness**

* **Low discount band (avg 9%)** is highly effective, contributing the largest sales ($2.34 million) and solid profit ($222k).
* **High discount band (avg 39%)** drives substantial sales ($1.61 million) but results in a loss of $189k.
* **Medium discount band (avg 18%)** generates $774k in sales but still yields a loss (-$19k).

Business decision: The organization should revisit its high and medium discount strategies. It appears that higher discounts are not translating into proportional sales uplift and instead are damaging profitability. A targeted reduction in discount levels or tightening discount approval processes would be advisable.

**Territory Dynamics**

* EMEA drives large sales but is a significant loss-making region under current discounting.
* NA contributes strongly to profit despite also offering discounts.
* APAC achieves the highest profit margin, indicating better balance between pricing and profitability.

Business decision: Discounting policies should be region-specific rather than one-size-fits-all. EMEA requires immediate attention — either reduce discounts or reassess product-market fit. The APAC model could serve as a benchmark for discounting discipline.

**Product Line Insights**

* **Trains**: Although low in sales volume, this line generates the highest margin (18.5%) and strong profit relative to its size.
* **Vintage Cars**: Performs well under medium discounting, contributing significant profit.
* **Classic Cars**: Despite being a top seller, this line is heavily discounted and the largest loss-making segment.

Business decision: The Classic Cars product line strategy needs a major overhaul. Either reposition the product, reprice it, or limit discounting. Trains and Vintage Cars could be prioritized for promotion as they offer better returns on sales.

**Customer-Level Discounting**

* Euro Shopping Channel, while contributing significant sales, does so at a negative profit margin despite a moderate discount rate (14.87%).
* Similar patterns are likely visible across other major customers in the scatter plot.

Business decision: Introduce customer-level profitability analysis into the sales process. Key accounts should have tailored pricing strategies based on their actual contribution to the bottom line, not just their sales volume.

### ****Broader Recommendations****

* Establish stricter discount governance — link discounting to clear volume or profitability thresholds.
* Develop region- and product-specific pricing models instead of blanket discount rates.
* Identify and reward sales of high-margin products and customers through incentive plans.
* Use the scatter plot routinely in sales reviews to monitor customer-level profitability.

## **Conclusion:**

This analysis demonstrates the power of combining SQL-based data validation and extraction with Tableau’s visual storytelling to translate raw transactional data into actionable business intelligence.

The project highlights critical focus areas:

* Reviewing discount policies, especially for high discount bands and loss-making regions.
* Optimizing Classic Cars pricing and cost structures.
* Leveraging strong customer and product performance patterns for targeted sales strategies.
* Revisiting regional approaches, particularly for EMEA.

Future work could extend this analysis to include cost-to-serve metrics, customer lifetime value, and predictive modelling. The dashboards and queries built provide a scalable foundation for continuous monitoring and strategic planning.

**Challenges & Learnings**

* **Challenge:** Balancing dashboard complexity with performance while managing dynamic filters across large transactional data.  
  **Learning:** Careful index use in queries and optimizing Tableau filter actions significantly improved responsiveness.
* **Challenge:** Translating complex SQL outputs into intuitive visualizations without overwhelming the user.  
  **Learning:** Focused on clean design principles and incremental interactivity for better end-user experience.

**About Me**  
I am a passionate data analyst with a unique blend of government, startup, and business experience. With hands-on skills in SQL, Tableau, Python, and Excel, I specialize in translating complex data into clear business insights.

My focus is on empowering businesses to make smarter, data-driven decisions — whether through deep database audits, dynamic dashboards, or customized analytics solutions.

I believe data tells a story, and my mission is to help businesses read that story, act on it, and grow stronger through it. Let’s turn your data into your competitive edge!