**SAMPLE SALES DATA ANALYSIS**

**Project Overview**

In today’s competitive business environment, data-driven decisions are critical to sustaining growth and profitability. This project aims to analyze historical sales data for a fictional global company to gain a deep understanding of its **sales performance**, **customer behavior**, **product success**, and **operational efficiency**. The business is experiencing fluctuations in revenue across regions, inconsistent product performance, and unclear customer buying trends. To better inform sales strategies and optimize operations, this project explores key patterns and anomalies in the sales dataset using structured SQL queries.

By uncovering the factors that influence high sales, frequent cancellations, and product popularity, the organization can make informed decisions that support revenue growth and operational improvements.

**📌 Key Objectives**

This project is designed to achieve the following goals:

1. Assess overall sales performance by calculating total revenue, order volume, and average order value.
2. Analyze sales trends over time to detect seasonal fluctuations and growth patterns.
3. Identify top-performing products and product lines by revenue and quantity sold.
4. Evaluate customer contribution to sales through customer segmentation and deal size distribution.
5. Examine regional sales patterns to uncover the highest and lowest-performing territories or countries.
6. Analyze order fulfillment efficiency, including cancelled and disputed orders, to quantify potential revenue losses.
7. Compare actual selling prices with MSRP to assess discounting behavior and pricing strategy effectiveness.
8. Provide actionable insights and strategic recommendations to enhance sales performance and customer retention.

**Data Cleaning and Preparation**

Before diving into analysis, a thorough data cleaning and validation process was conducted to ensure accuracy, consistency, and readiness for querying.

**1. Original Condition of the Data**

To preserve data integrity, a copy of the original dataset was created. A temporary working dataset was also prepared, which included only the relevant columns required for the analysis. This approach allowed for focused data exploration without altering the source.

**2. Data Cleaning Procedures**

The following data cleaning steps were performed:

✅ **Checked for missing and empty values** across all columns to ensure completeness.

✅ **Checked for duplicate records**, although none were found in this case.

✅ **Reviewed text columns for inconsistent entries** such as varying case or spacing in categorical fields.

**3. Column Derivations**

A new **Sales\_ID** column was created by concatenating the OrderNumber and OrderDate columns. This provided a unique identifier for each transaction, which improved traceability and simplified grouping during analysis.

**4. Data Quality Checks Before Analysis**

To validate the dataset before applying queries, several checks were conducted:

🔹 **Row Count Check:** Confirmed 2,823 rows present.

🔹 **Column Count Check:** Confirmed 19 relevant columns included.

🔹 **Data Type Check:** Ensured all data types matched their expected formats (e.g., numeric, text, date).

🔹 **Duplicate Check:** Verified the absence of duplicate entries.

**5. Final Assessment**

* No null or empty values remained in the dataset.
* All text columns were standardized and consistent.
* All data types were correct and aligned with the intended analysis.
* The dataset was clean, complete, and ready for efficient querying and reporting.

**Key Analysis**

This section highlights the key SQL-based analyses performed on the sales dataset to answer important business questions and guide strategic decisions.

**1. Sales Performance Overview**

Understanding overall business health through revenue performance, volume, and trends.

* **Total Sales Revenue**: Calculated the total revenue across all orders to establish the company's gross earnings baseline.
* **Total Quantity of Items Sold**: Summed all quantities ordered to understand the volume of product movement.
* **Average Revenue per Order**: Derived by dividing total sales by the number of distinct orders to measure how much revenue each order typically contributes.
* **Sales Trend Over Time**: Analyzed monthly and yearly sales trends by converting month numbers to readable names and grouping sales accordingly.
* **Top 5 Countries by Total Sales**: Identified countries contributing the highest revenue to focus regional sales expansion efforts.

**2. Product Analysis**

Revealing which products and product lines drive the most value.

* **Top 10 Best-Selling Products**: Ranked products by quantity ordered to discover customer demand.
* **Top 10 Revenue-Generating Products**: Ranked products by total revenue to assess profitability.
* **Revenue by Product Line**: Summed sales by product categories to evaluate the strongest and weakest segments.
* **Sales vs MSRP**: Compared actual selling prices with MSRP to identify discounting patterns and evaluate pricing strategy.
* **Average Price per Product Line**: Calculated the average unit price per product line for deeper insight into value perception.

**3. Customer Analysis**

Understanding buyer behavior and customer-driven revenue.

* **Top 10 Customers by Total Revenue**: Identified high-value customers responsible for the bulk of revenue.
* **Average Order Value per Customer**: Measured customer profitability based on average order size.
* **Customer Distribution by Country**: Counted customers by country to reveal regional customer spread.
* **Customer Loyalty**: Counted number of transactions per customer to evaluate order frequency.
* **Top Deal Sizes by Customer**: Analyzed customer preference for Small, Medium, or Large deals based on revenue generated.

**4. Order Status and Operational Efficiency**

Analyzing how well orders are being fulfilled and where revenue is lost.

* **Order Status Breakdown**: Summarized the proportion of orders by status (Shipped, Cancelled, On Hold, etc.) to assess operational health.
* **Sales Loss from Cancelled Orders**: Calculated the total sales value of cancelled orders to quantify missed revenue opportunities.

**5. Regional Sales Analysis**

Evaluating where the company performs best geographically.

* **Top Performing Countries & Cities**: Ranked locations by total sales to discover strategic hotspots.
* **Regional Deal Size Distribution**: Broke down deal size patterns (Small, Medium, Large) by country to understand regional buying behavior.

**📊 Insights & Discussion**

The analyses performed using SQL provided actionable insights into various dimensions of the business including overall sales performance, customer behavior, product popularity, operational efficiency, and regional strengths. These findings are directly linked to the project’s key objectives.

**1. Sales Performance Overview**

* The business generated a total revenue of $10.03 million from 99,067 products sold, with an average revenue per order of $3,553.89. This indicates strong order value, especially for a product-based retail model.
* Monthly trend analysis revealed that November, October, and May are the highest revenue-generating months, suggesting seasonal or promotional influences.
* In terms of yearly growth, sales peaked in 2004 ($4.72M), showing strong momentum from the previous years.
* Regionally, the USA leads all countries with over $3.6M in sales, followed by Spain and France, confirming North America and Western Europe as the company’s strongest markets.

**2. Product Analysis**

* The top-selling product by quantity was S18\_3232 with 1,774 units, while the lowest-selling products still performed decently with 700+ units sold.
* In terms of revenue, the following product codes such as the S18\_3232, S10\_1949, and S10\_4698 led the pack, contributing significantly to the total sales.
* The Classic Cars product line dominated with nearly $3.92 million in revenue, far surpassing other lines.
* The Sales vs MSRP analysis showed that the company heavily discounts many items especially Classic Cars, often sold at $114 below MSRP. This could indicate a pricing strategy to boost volume, but it may also point to potential lost margins.
* The average price per product line hovered between $75–$88, with Trucks and Buses commanding the highest average price.

**3. Customer Analysis**

* The top customers include Euro Shopping Channel and Mini Gifts Distributors Ltd., contributing $912K and $654K respectively, highlighting their significance in overall revenue.
* Many customers exhibit high Average Order Values (AOV), with several above $3,000, indicating a loyal and high-value customer base.
* Customer distribution shows a strong presence in the USA, with over 1,000 transactions, followed by Spain and France.
* In terms of customer loyalty, Euro Shopping Channel leads with 259 orders, reaffirming its role as a key account.
* Deal size analysis indicates that the majority of revenue is generated from Medium deals, but notable customers like Euro Shopping Channel and The Sharp Gifts Warehouse also contribute significantly through Large deals, showing a healthy mix of high-value engagements.

**4. Operational Efficiency**

* The order status breakdown shows excellent fulfillment performance, with 92.7% of orders successfully shipped. However, 2.13% were cancelled, resulting in a $194,487.48 revenue loss — a figure that should be monitored closely.
* Only 0.5% of orders were disputed, and around 3% were either on hold or in process, indicating minor operational friction points that can be optimized.

**5. Regional Sales Analysis**

* Madrid (Spain) and several U.S. cities (San Rafael, NYC) are top revenue hubs, which should be targeted for future campaigns and loyalty programs.
* The regional deal size distribution reveals that most regions, especially the USA and Spain, lean heavily toward Small and Medium deals, while Large deals remain a minority in most countries. This may indicate an opportunity to upsell or bundle products to increase order sizes in key markets.

**💡 Recommendations**

To address the original business problems and support future growth, the following recommendations are proposed:

**1. Strengthen Pricing Strategy**

* Re-evaluate deep discounts offered on high-demand items like Classic Cars, where excessive markdowns may be unnecessarily cutting into profit margins.
* Implement tiered discounting based on volume or loyalty status to protect profitability.

**2. Reduce Cancellations and Revenue Leakage**

* Investigate root causes of cancelled orders to reduce the $194K+ lost revenue. Common issues may include stockouts, delivery delays, or payment failures.
* Introduce order confirmation or pre-shipment checks for high-value orders.

**3. Expand Large Deal Opportunities**

* Focus on converting medium deal customers into large deal clients through bundled offers, loyalty tiers, or personalized upselling.
* Launch enterprise packages or B2B bulk-buy programs in high-performing regions like the USA and Spain.

**4. Targeted Regional Campaigns**

* Run localized campaigns in top-performing cities (e.g., Madrid, NYC, Paris) and test retargeting strategies for low-engagement regions.
* Align regional inventory and promotions with seasonal trends observed in the sales trend analysis.

**5. Customer Loyalty and Retention**

* Reward loyal customers like Euro Shopping Channel with exclusive offers or early access deals.
* Use the AOV and order frequency data to design personalized retention campaigns and reduce churn.

**✅ Conclusion**

This project successfully leveraged SQL to analyze a real-world sales dataset and extract valuable business insights. By exploring key areas such as sales performance, product trends, customer behavior, order efficiency, and regional activity, we were able to identify both strengths and gaps in the company’s sales operations. The analysis revealed that the company is generating strong revenue, with high-value customers and consistent sales across top product lines like Classic Cars. Regions such as the USA and Spain are driving a large portion of the revenue, while seasonal trends highlight peak months like November and October.

However, the analysis also uncovered opportunities for improvement. A significant portion of revenue is lost through cancelled orders, and many products are being sold below MSRP, which may impact profitability. Additionally, large deal sizes are underrepresented across most regions, suggesting untapped potential in scaling customer transactions.

Overall, this project demonstrates the power of SQL as a tool for business intelligence, supporting smarter, data-driven decision-making across sales, operations, and customer engagement.