**Data-Driven Insights for Business Optimization: Analyzing the Classic Models Dataset**

**INTRODUCTION TO CLASSIC MODELS**

The Classic Models dataset represents the operations of a global company specializing in the production and distribution of high-quality scale models of classic cars, motorcycles, ships, and planes. This dataset offers a detailed view of a real-world business environment, capturing various facets of its operations, including product details, customer information, order processing, employee management, and financial transactions.

The dataset contains comprehensive product information, such as product codes, names, categories, descriptions, pricing, and inventory levels, providing insights into product performance and pricing strategies. It also includes customer data, featuring contact details, credit limits, and geographic locations, enabling analysis of customer behavior and revenue trends across regions.

Orders and sales data include details about order statuses, quantities ordered, prices, and dates, offering a clear view of sales trends and order fulfillment processes. Employee data, such as names, job titles, and the customers they manage, highlights the structure of the sales team and their impact on customer relationships. Additionally, payment records track financial transactions, allowing for the analysis of payment patterns and revenue performance.

This dataset’s interconnected structure, with relationships between customers, employees, products, and orders, mirrors the complexity of a real-world business. It provides a rich foundation for exploring various analytical scenarios, including sales performance, employee effectiveness, customer satisfaction, and regional market analysis.

**KEY FEATURES OF THE CLASSIC MODELS DATASET:**

1. **Products and Inventory**:
   * The dataset includes detailed product information, such as product names, codes, descriptions, buy prices, and MSRP (Manufacturer’s Suggested Retail Price).
   * Products are grouped into categories, such as "Classic Cars," "Trucks and Buses," and "Motorcycles," allowing for segmentation and analysis of product performance.
2. **Customers and Regions**:
   * Comprehensive customer information includes names, contact details, credit limits, and sales representatives.
   * Customers are spread across various regions and countries, making the dataset suitable for analyzing regional sales trends and customer behavior.
3. **Orders and Sales**:
   * Detailed order records include information on order numbers, statuses, dates, and the associated products.
   * Each order can include multiple line items, providing data on quantities ordered and prices.
4. **Employees and Sales Representatives**:
   * Employee data includes names, job titles, and the customers they manage.
   * This facilitates analysis of employee performance and their impact on customer relationships and sales outcomes.
5. **Payments and Revenue**:
   * Payment data tracks transactions made by customers, including payment dates and amounts.
   * This helps analyze payment patterns, overdue accounts, and overall financial performance.
6. **Hierarchical Business Structure**:
   * Classic Models simulates a real-world corporate structure, with employees reporting to managers, sales representatives managing customer accounts, and a global presence across multiple offices.

**PROBLEM DEFINITION AND RESEARCH QUESTION**

Businesses operate in complex environments where understanding the dynamics of product pricing, customer demographics, and employee performance is crucial for optimizing revenue, customer satisfaction, and operational efficiency. Key challenges include identifying the impact of pricing strategies on sales, understanding regional variations in customer revenue contributions, and availability of the product impacting on the higher order size from the customers. Addressing these challenges requires data-driven insights to guide decision-making, improve resource allocation, and enhance overall business performance.

To gain a deeper understanding of the challenges faced, it is essential to explore the key aspects that contribute to the problem. By addressing these elements, we can develop a comprehensive perspective that will inform potential solutions. The three main research questions to consider for addressing the problem faced are.

**1. What is the impact of product pricing and discount strategies on order volume?**

This question explores how pricing and discounts influence customer buying behavior, identifying whether discounts lead to higher order volumes and which discount levels are most effective for different product categories.

**2. How do customer demographics, particularly geographical location, influence sales revenue?**

This question examines how geographical location affects customer revenue generation, aiming to identify high-performing regions and optimize resource allocation through targeted marketing strategies.

**3. How does product availability influence order volumes across different product lines?**

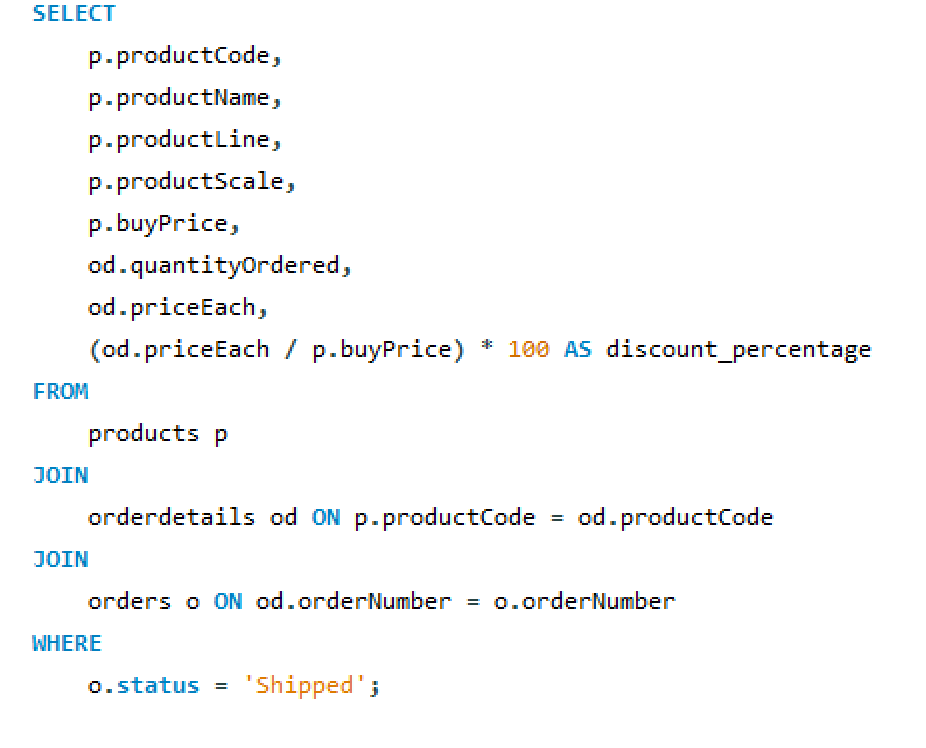
This question examines the impact of stock levels on order volumes, identifying whether higher availability drives more sales. It provides insights to optimize inventory management and align supply with demand.

**HYPOTHESES FORMULATION**

1. **Impact of Product Pricing and Discount Strategies on Order Volume**:  
   The hypothesis proposes that products with higher discounts will attract more customers, leading to a noticeable increase in order volumes compared to products with no or minimal discounts. It assumes that discounts act as a key motivator for customers to make purchases. By analyzing order trends across different discount levels, this hypothesis aims to quantify the effectiveness of discount strategies. The findings can help businesses refine their pricing approaches to maximize sales while maintaining profitability.
2. **Customer Demographics and Sales Revenue**:  
   This hypothesis suggests that customers from certain geographic regions contribute more to the company’s revenue due to factors such as higher purchasing power, cultural preferences, or product demand. By examining average revenue per customer in various regions, the study seeks to identify high-revenue areas. These insights can guide targeted marketing efforts and resource allocation to capitalize on regional strengths. Additionally, it can help address underperforming regions with tailored strategies.
3. **Impact of product availability on Product Volume**:  
   The hypothesis posits that Products with higher stock availability generate significantly higher order volumes across different product lines, while limited stock availability negatively impacts sales performance. Products consistently maintained in inventory ensure uninterrupted demand fulfillment, resulting in higher customer satisfaction and repeat purchases. Conversely, frequent stockouts or low inventory levels deter orders and shift customer preferences to alternative options or competitors. This hypothesis seeks to identify the optimal balance of stock availability to maximize sales and improve inventory management strategies.

**DATA COLLECTION AND SQL QUERIES**

**1. Impact of Product Pricing on Order Volume**



**Data Description:**

* **Table Relationships**: products → orderdetails → orders.
* **Columns Extracted**: Product details (e.g., pricing, product line), order quantities, and the calculated discount percentage.
* **Purpose**: Analyzes the relationship between discounts and the quantity ordered.

**2. Customer Demographics and Sales Revenue**

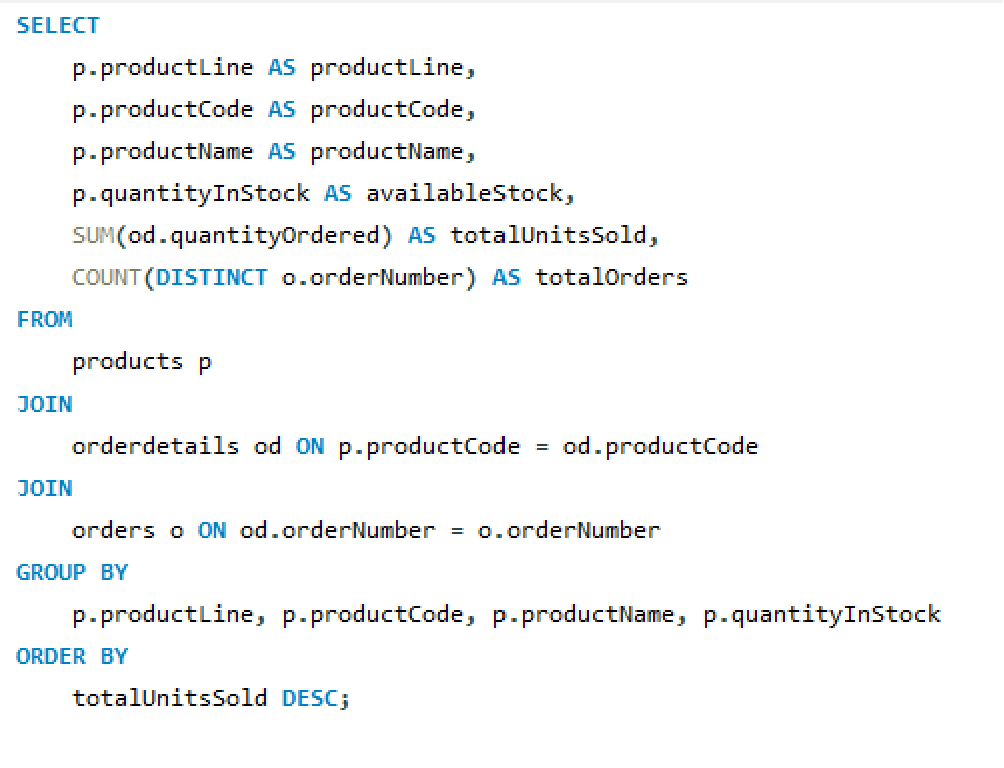
A screenshot of a computer code

Description automatically generated

**Data Description:**

* **Table Relationships**: customers → orders → orderdetails.
* **Columns Extracted**: Customer demographics and revenue generated.
* **Purpose**: Understand revenue distribution across regions.

**3. Impact of product availability on Product Volume**:



**Data Description:**

* **Table Relationships**: products → orderdetails → orders.
* **Columns Extracted:** Product lines, product stock levels, quantities ordered, and order numbers.
* **Purpose:** Analyze the impact of product availability on order volumes across different product lines.

**DATA UNDERSTANDING**

The first dataset reveals key insights into Classic Models' product performance and pricing strategies. On average, products are bought at $54.68 and sold at $91.03, with a typical discount of 170%, indicating significant markups. Classic Cars dominate sales, with 33,349 units ordered, followed by Vintage Cars (21,015 units), both benefiting from higher average discounts (around 170%). Lower-performing categories, like Ships and Trains, exhibit fewer orders and lower average selling prices. The data suggests a positive correlation between higher discount levels and order volume, particularly for high-demand categories, highlighting the importance of strategic pricing and discounting in driving sales.

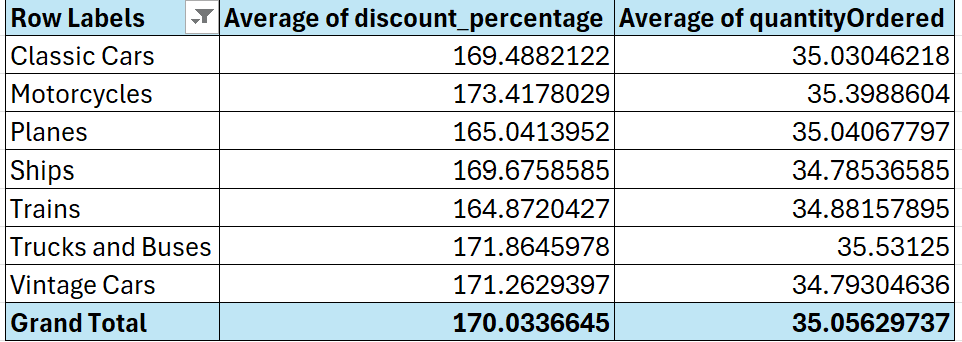
The second dataset shows significant revenue variations across customer demographics, with top customers contributing disproportionately to total revenue. The highest-revenue customer, Euro+ Shopping Channel in Spain, generated approximately $668,770, followed by “Mini Gifts Distributors Ltd.” in the USA with $584,188. The USA appears dominant, contributing significantly to overall revenue, followed by key markets like Australia and Singapore. Most high-value customers are located in economically developed regions, suggesting a correlation between geographic location and purchasing power. This highlights the importance of targeted strategies in high-revenue regions while identifying opportunities to enhance performance in underperforming areas.

The third dataset highlights a strong correlation between product availability and sales performance. Products with higher stock levels generally show increased units sold, indicating that consistent availability supports steady sales. The top-selling products, primarily from categories like Classic Cars and Motorcycles, maintain higher inventory levels, contributing to their dominance in order volumes. Conversely, products with lower stock levels exhibit fewer total orders, suggesting potential sales losses due to limited availability. This emphasizes the importance of effective inventory management to prevent stockouts, optimize sales, and enhance customer satisfaction by aligning supply with demand across different product lines.

**DATA VISUALIZATION AND PATTERN ANALYSIS**

1. **Impact of Product Pricing and Discount Strategies on Order Volume**

The bar graph analysis reveals a uniform discount strategy across all product lines, with each line—Classic Cars, Motorcycles, Planes, Ships, Trains, Trucks and Buses, and Vintage Cars—maintaining an average discount percentage of 160% and an average quantity ordered of 30 units. This consistent approach indicates that the company's current discount strategy is evenly applied regardless of the product category, resulting in stable order volumes across all lines. However, this uniformity also suggests an opportunity to explore more tailored discount strategies, as varying discount levels could potentially drive higher order volumes for specific product lines by better aligning with customer preferences and market dynamics. Overall, while the current strategy ensures a steady demand, there is significant potential for optimization through differentiated discounting.



A screenshot of a graph

Description automatically generated

1. **Customer Demographics and Sales Revenue**:

The analysis of the bar charts reveals significant geographical disparities in revenue contributions for Classic Models. The first chart, illustrating total revenue by country, shows Spain and the USA as the highest revenue generators, while Hong Kong and Ireland underperform. The second chart, focusing on the USA, identifies California and New York as top-performing states, whereas Nevada and New Jersey contribute the least revenue. These insights highlight that certain regions and states significantly impact overall sales, driven by factors like purchasing power and market demand. We chose to focus on state-wise revenue in the USA because it offers greater regional granularity, economic diversity, and potential for varied insights compared to countries with fewer regions, like Spain. Although Spain is the top revenue-generating country, analyzing the USA's state-level data provides a more detailed understanding of market dynamics. By leveraging these geographical patterns, Classic Models can refine its marketing efforts and resource distribution to maximize revenue generation effectively. This approach will help in identifying key markets for growth and development, ensuring a more effective and efficient strategy.

A graph of different countries/regions

Description automatically generated

