|  |
| --- |
| 1 |
| Capstone Project – Retail Analysis |
| Data Analytics |

|  |
| --- |
| Ankit Kumar  3/26/2024 |

Retail analysis involves the comprehensive evaluation of various aspects within the retail industry to gain insights into performance, trends, consumer behavior, and market dynamics

Overview

THE PROCESS

1. Data Acquisition from GitHub:

Obtain the requisite dataset from a designated GitHub repository, containing essential information on sales, customers, products, and demographics encompassing various countries and their performance across distinct ranking systems.

2. Data Transformation and Enhancement:

If necessary execute data transformation procedures to ensure data quality and consistency. Additionally consider augmenting the dataset with new problem statements to enrich the analysis potential.

3. Connecting with tools:

Establish connection between the dataset and various analytical tools. Interface the dataset with Power BI, Excel, and MySQL Workbench, facilitating seamless data integration and processing.

4. Problem Statement Solution in Power BI:

Utilize Power BI to delve into the specified problem statements. Employ its robust features for data visualization, exploration and analysis effectively deriving insights and solutions.

5. Exploratory Data Analysis (EDA) :

Perform exploratory data analysis using either Excel or SQL workbench, depending on the complexity of the analysis. Extract meaningful patterns, relationships and trends from the data to inform subsequent decision- making.

6. Creation of Visual and Insightful PowerPoint:

Develop a comprehensive PowerPoint presentation that encapsulates the project’s objectives, methodologies, problem statement solutions and key visualization. Each problem statement should be accompanied by a dedicated section with pertinent conclusions and insights.

7. Detailed Documentation:

Compile a detailed report that meticulously documents the entire project lifecycle. Include sections on data collection, transformation, problem statement formulation, tools integration, Power Bi solutions, EDA insights and PowerPoint Visualizations.

Objective

# Retail analysis is like taking a close look at how stores work to make smart decisions. It involves checking out lots of things, such as what people buy, how much they buy, and when they buy it. By doing this, stores can figure out the best prices for stuff, know what customers like, and make sure they always have enough things in stock.

# This analysis also helps stores understand what people prefer to buy and how they like to shop. For example, some people might prefer going to a physical store, while others like buying things online. By figuring out what works best, stores can improve their ways of selling things to make customers happier.

# Stores also compare themselves with other stores to see how well they're doing and learn from them. They keep an eye on what's happening in the market and how the economy is doing to make smart decisions.

# In short, retail analysis helps stores do better. It helps them sell things smarter, make customers happier, and keep up with what people want. It's like a roadmap that guides stores to make good choices and grow better."

Significance

Retail analysis holds immense significance in documentation for various reasons. Firstly, it provides a comprehensive understanding of consumer behavior, preferences, and buying patterns, which are crucial for strategic decision-making. By documenting and analyzing retail data, businesses gain insights into which products are selling well, at what price points, and in which locations or demographics. This information helps in optimizing inventory management, ensuring that the right products are stocked in the right quantities, reducing excess inventory costs, and meeting consumer demands efficiently.

Moreover, documentation of retail analysis aids in evaluating the performance of different sales channels, such as online platforms versus physical stores, enabling businesses to allocate resources effectively. This documentation also facilitates the identification of trends and seasonal variations, allowing retailers to plan promotions, marketing campaigns, and product launches strategically.

Furthermore, retail analysis documentation supports the assessment of pricing strategies' effectiveness. By tracking sales volumes, profit margins, and customer responses to pricing changes, businesses can fine-tune their pricing strategies for maximum profitability and competitiveness in the market.

Additionally, documented retail analysis serves as a valuable reference for future planning and forecasting. It helps in setting realistic sales targets, understanding market dynamics, and making informed predictions about future consumer behavior, thereby aiding in the development of sustainable growth strategies.

Overall, comprehensive documentation of retail analysis is instrumental in providing a data-driven foundation for informed decision-making, optimizing operations, and ultimately driving business growth and profitability in the highly competitive retail landscape.

Data Dictionary:

1. Table: customers

**Customer number** (int): Unique identifier for each customer.

**Name** (varchar): First name of the customer.

**Phone** (varchar): Contact phone number of the customer.

**Address** **customer name** (varchar): Full name of the customer.

**Last name** (varchar): Last name of the customer.

**First**

(varchar) : Customer's address.

**City state** (varchar): City and state of the customer.

**Postal code** (varchar): Postal code of the customer's location.

**Country** (varchar): Country of the customer.

**Sales rep. employee number** (int): Employee number of the sales representative.

**Credit limit** (decimal): Maximum credit limit for the customer.

2. Table: Employees

**last name** (varchar): Last name of the employee.

**first name** (varchar): First name of the employee.

**Extension** (varchar): Extension number for contact.

**Email** (varchar): Email address of the employee.

**Office code** (int): Unique identifier for the office.

**Report to** (int): Employee ID to whom the employee reports.

**Job title** (varchar): Job title of the employee.

3. Table: offices

**Office code** (int): Unique identifier for each office.

**City** (varchar): City where the office is located.

**Address** (varchar): Address of the office.

**State** (varchar): State of the office location.

**Country** (varchar): Country of the office.

**Postal code** (varchar): Postal code of the office.

**Territory** (varchar): Territory covered by the office.

4. Table: order details

**Order number** (int): Unique identifier for each order.

**Product code** (int): Unique identifier for each product.

**Quantity ordered** (int): Quantity of the product ordered.

**Price each** (decimal): Price per unit of the product.

**Order line number** (int): Line number for the order.

5. Table: orders

**Order number** (int): Unique identifier for each order.

**Order date** (date): Date when the order was placed.

**Required date** (date): Date by which the order is required.

**Shipped date** (date): Date when the order was shipped.

**Status** (varchar): Current status of the order.

**Comments** (varchar): Additional comments related to the order.

6. Table: payments

**Customer number** (int): Unique identifier for each customer.

**Check number** (varchar): Unique identifier for each check payment.

**Payment date** (date): Date when the payment was made.

**Amount** (decimal): Amount of the payment.

7. Table: product lines

**Product line** (varchar): Category of the product line.

**Text description** (varchar): Description of the product line.

8. Table: products

**Product code** (int): Unique identifier for each product.

**Product name** (varchar): Name of the product.

**Product line** (varchar): Category of the product line.

**Product scale** (varchar): Scale of the product.

**Product vendor** (varchar): Vendor of the product.

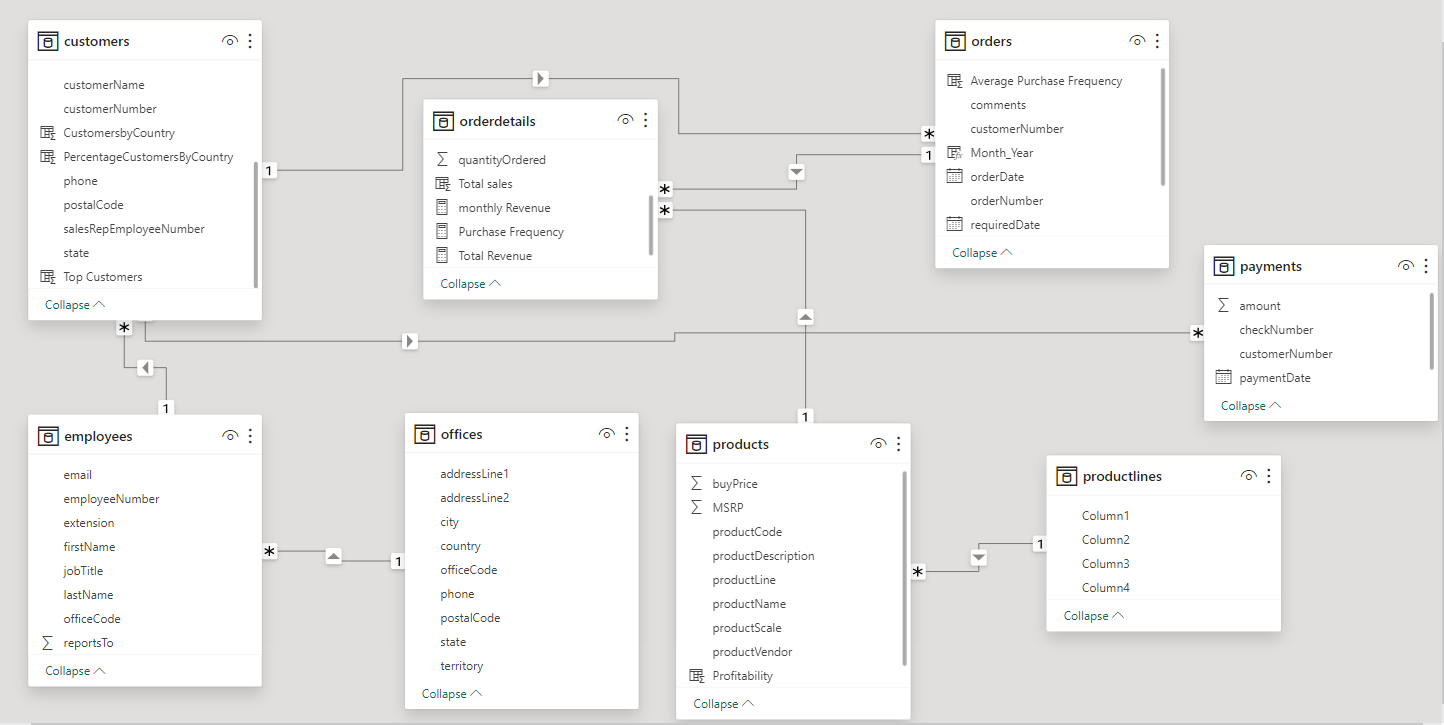
**Product description** (varchar): Description of the product.

**Quantity in stock** (int): Quantity of the product in stock.

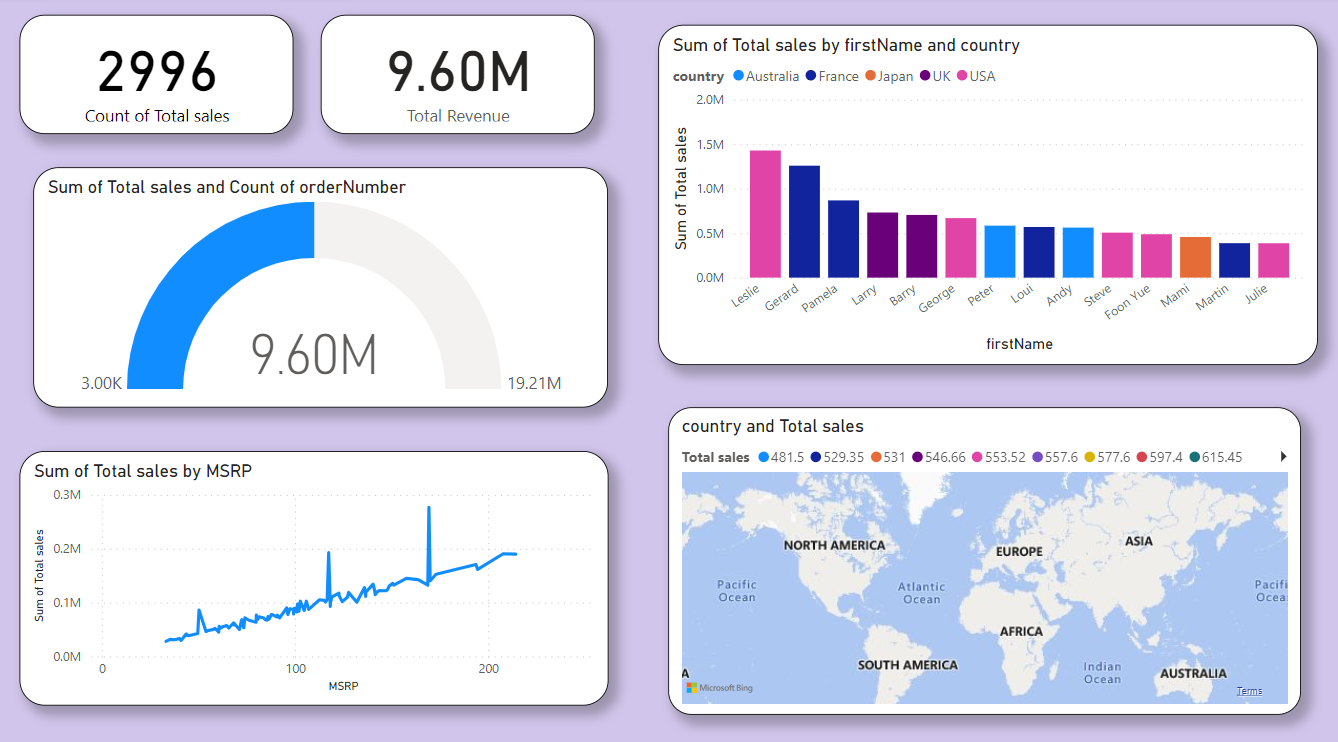
**Buy price** (decimal): Purchase price of the product.

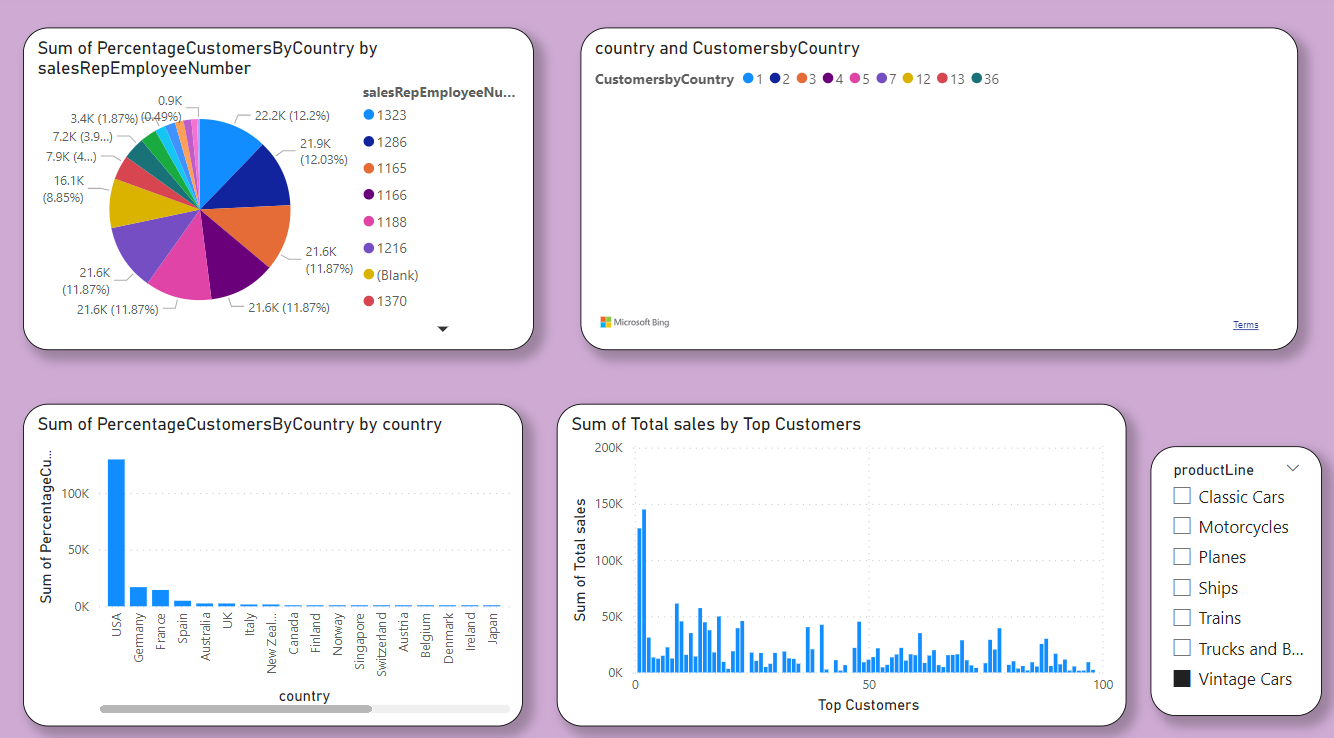
**MSRP** (decimal): Manufacturer's Suggested Retail Price.

ER-Diagram

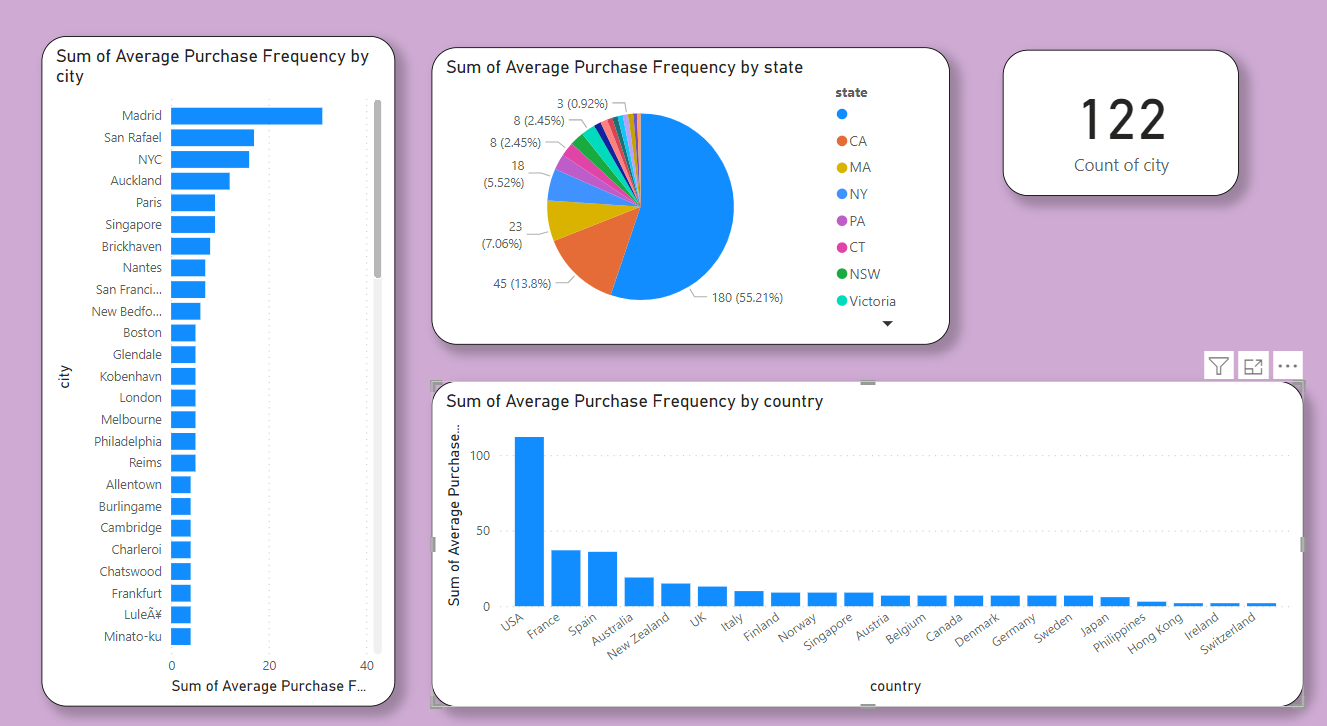


Power BI Problem Statements

 Sales Analysis

 Customer Analysis

 Product Analysis

 Demographic Analysis

How does monthly revenue vary across different product categories? (Visual: Monthly revenue by product category)

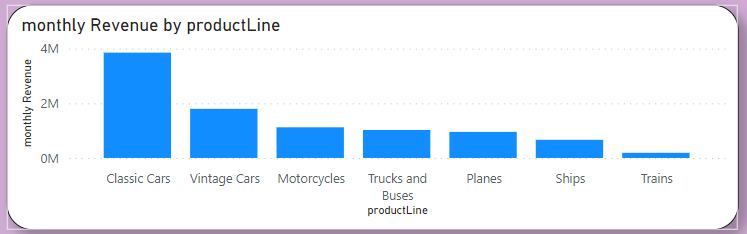
Based on the visual analysis of monthly revenue by product category, you can draw conclusions such as:

Identify the top-performing product categories in terms of monthly revenue.

Recognize trends or seasonal patterns in sales for different product lines.

Compare the revenue performance across various categories to assess which contribute most significantly to overall revenue.

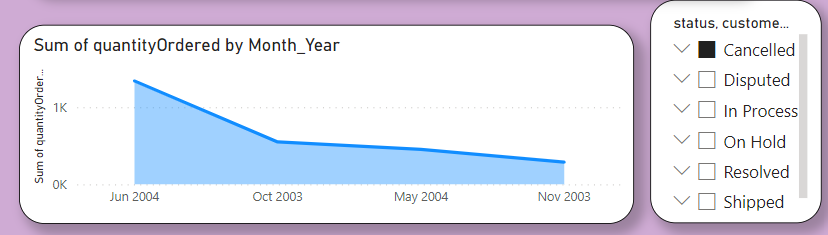
Highlight any fluctuations or notable changes in sales within specific product lines across different months.



What is the trend in customer order volume over the past year? (Visual: Monthly order volume trend)

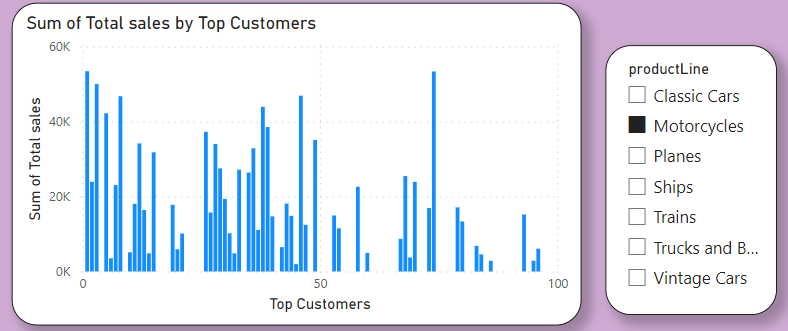
To analyze the trend in customer order volume over the past year, a visual representation of the monthly order volume can offer valuable insights. By aggregating and plotting the total number of orders placed each month, one can observe patterns and fluctuations in customer buying behavior.

The graph displaying the monthly order volume trend may reveal seasonal spikes or dips in purchases. For instance, it might show increased order volumes around holiday seasons or promotional periods. Identifying such patterns helps in understanding customer behavior and planning marketing or sales strategies accordingly.



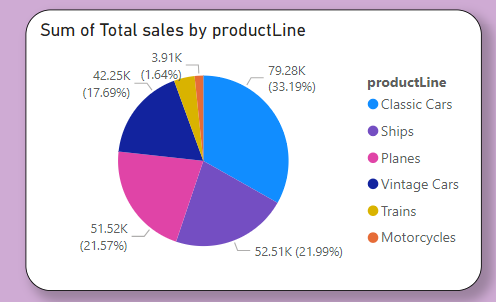
How does the sales performance of top customers compare to the rest? (Visual: Sales contribution by top customers vs. others)

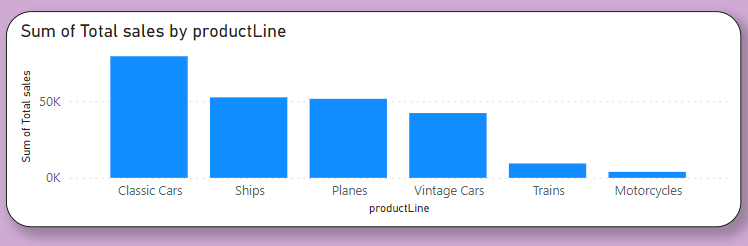
Analyzing the sales performance between the top customers and the rest unveils crucial insights into revenue contribution and customer segmentation. By evaluating the sales figures of the top customers against the collective sales generated by the remaining customer base, a clearer understanding of revenue concentration and customer significance emerges. Typically, the top customers, albeit a smaller subset, tend to exert a substantial impact on total sales revenue, often contributing a disproportionately higher percentage compared to the rest of the customer base. This phenomenon highlights the significance of cultivating and nurturing relationships with these key clients to sustain business growth. Moreover, understanding the purchasing behavior, preferences, and patterns of these top customers becomes imperative for businesses aiming to optimize their sales strategies and enhance overall performance. Visual representations, such as charts or graphs illustrating sales contributions from top customers juxtaposed with the aggregate sales from others, can effectively underscore the importance of these key accounts in driving business success while emphasizing the need for strategic focus on customer retention and expansion efforts.



What is the distribution of product sales across different product lines? (Visual: Product sales by product line)

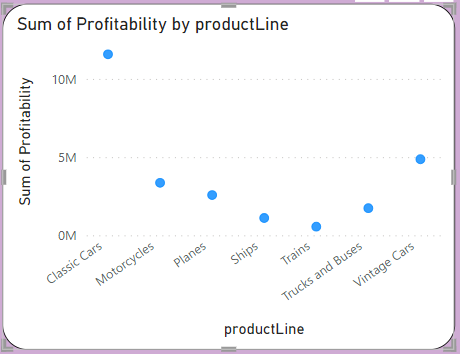
Analyzing the distribution of product sales across different product lines provides crucial insights into the performance and popularity of various product categories. By aggregating sales data from the 'orders' and 'order details' tables linked through 'product code', it's possible to visualize the sales distribution. The 'product lines' table categorizes products into distinct lines, enabling a comprehensive breakdown of sales figures per product line. This analysis can reveal which product categories are top performers, identifying trends and potential areas for growth or diversification. Visualizing the product sales by product line could display a comparative bar or pie chart, showcasing the relative contribution of each product line to the overall sales volume. This visualization would offer a quick, informative snapshot of sales distribution, aiding decision-making processes related to inventory management, marketing strategies, and resource allocation across different product lines for optimized business performance.





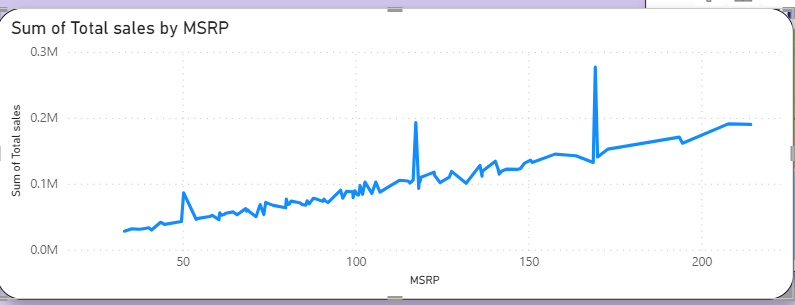
How does the profitability of different products compare based on their quantity in stock? (Visual: Profitability vs. quantity in stock)

To evaluate the profitability of different products based on their quantity in stock, a visual analysis can provide valuable insights. Plotting a graph with profitability on the y-axis and quantity in stock on the x-axis can reveal correlations and patterns. Generally, products with higher quantities in stock might suggest a higher investment in inventory. However, profitability may not be solely dependent on stock levels; it also relates to sales performance, pricing strategies, and production costs. Products with higher stock quantities might have lower margins if they move slower or have higher carrying costs. Conversely, certain products might have lower quantities in stock but higher profitability due to higher demand, unique features, or premium pricing. Identifying trends from the visual representation can help in optimizing inventory management, adjusting pricing strategies, or focusing on promoting high-profit, low-stock items. This analysis enables businesses to balance inventory levels while maximizing profitability, ensuring a more efficient use of resources and enhancing overall financial performance.



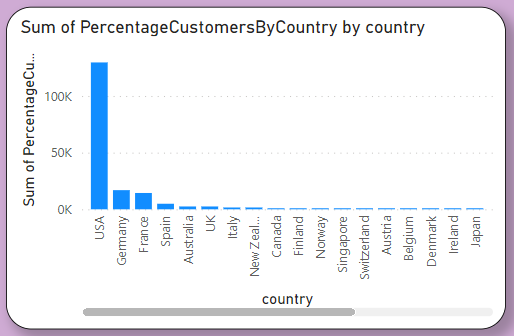
How does product pricing impact sales volume? (Visual: Product price vs. sales volume)

Product pricing plays a pivotal role in determining sales volume as it directly influences consumer behavior and purchasing decisions. When setting prices, finding the right balance is crucial: pricing products too high may deter potential buyers, while pricing them too low could compromise profitability. An optimal pricing strategy aligns with customer perceptions of value, competition analysis, and market demand. Higher-priced products often signal quality and exclusivity, attracting certain segments of consumers, while competitively priced items might capture a broader market share. Additionally, promotions, discounts, and pricing strategies such as value bundling or dynamic pricing can impact sales volume significantly. Visualizing the relationship between product price and sales volume often demonstrates patterns where moderate price points strike the best balance between sales volume and revenue generation, reflecting the delicate interplay between pricing strategies and consumer demand. Adjusting prices based on market conditions, customer preferences, and product lifecycle stages can dynamically impact sales volume, making continuous evaluation and adaptation crucial for sustained success in today's competitive marketplace.



What is the distribution of customers across different demographic segments? (Visual: Customer segmentation by demographics)

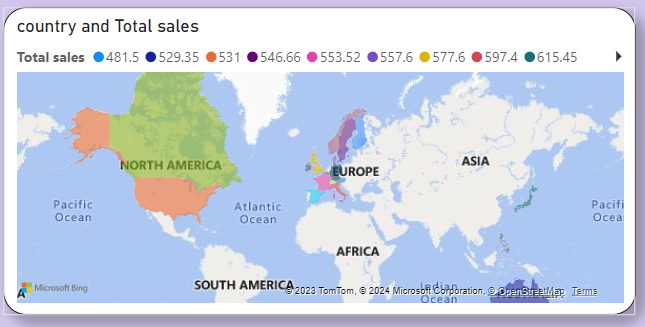
Customer segmentation by demographics involves categorizing customers based on various demographic factors such as age, gender, location, income, and more. Analyzing the provided tables, we can derive valuable insights into the distribution of customers across different demographic segments. The 'Customers' table contains crucial demographic information like customer names, addresses, phone numbers, and countries. By leveraging this data, segmentation can occur based on geographical location, enabling the identification of customer clusters in specific cities, states, or countries. Additionally, attributes like 'credit limit' may assist in segmenting customers by income levels or purchasing power. Furthermore, associating 'salesrep employee number' with customer data can aid in understanding customer preferences based on sales representatives. By employing analytical techniques, businesses can create visual representations like geographic heat maps, bar charts, or pie charts to illustrate the distribution of customers across demographic segments. Such insights facilitate targeted marketing strategies, personalized customer experiences, and tailored product offerings, ultimately enhancing customer satisfaction and boosting business growth.

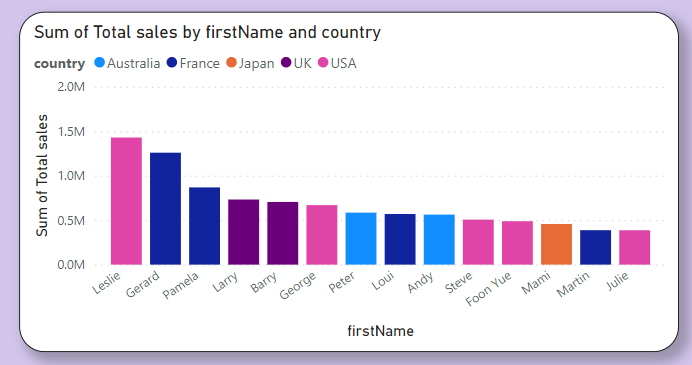


What are the top regions in terms of sales revenue? (Visual: Sales revenue by region)

To determine the top regions in terms of sales revenue, we'll need to analyze the data across various tables like orders, order details, customers, and products. By joining these tables, we can calculate the total sales revenue generated from different regions. The 'orders' and 'order details' tables provide information about each transaction, including the product code, quantity ordered, and price each. Connecting this with the 'customers' table will link sales to specific regions based on customer information like city or country. Additionally, the 'products' table offers insights into product details and prices.

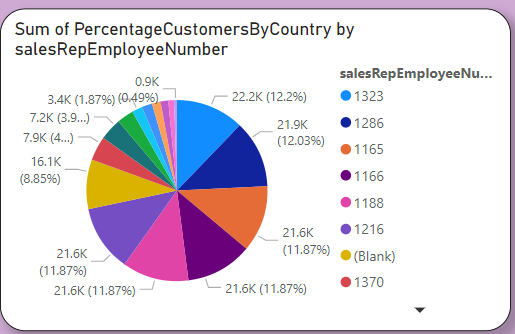
Once the data is aggregated and analyzed, a visual representation like a bar chart or a geographic map showcasing sales revenue by region can be generated. This visual will highlight regions that contribute the most to overall sales revenue. Typically, regions with higher sales volumes or where customers with higher credit limits are concentrated might indicate stronger revenue streams.





How does the performance of sales employees vary across different regions? (Visual: Employee performance by region)

Analyzing the performance of sales employees across different regions involves a comprehensive assessment of various factors. One key aspect is to evaluate sales figures, customer acquisition, and revenue generated by each employee within specific territories or regions. By correlating sales data with geographic regions, patterns may emerge regarding employee effectiveness in different areas. Factors influencing performance could include customer demographics, market demand, competition, and even cultural nuances impacting sales approaches.



What is the correlation between customer demographics and purchase frequency? (Visual: Customer demographics vs. purchase frequency)

Understanding the correlation between customer demographics and purchase frequency involves analyzing various factors such as customer attributes and their buying habits. Examining customer demographics like age, location, income, and purchase history against their purchase frequency can reveal essential insights. For instance, segments of younger customers might demonstrate higher purchase frequency, while those with higher income levels might exhibit larger transaction amounts but lower frequency. Additionally, analyzing regional patterns can unveil preferences or tendencies specific to certain areas. By correlating customer demographics with purchase frequency, businesses can tailor marketing strategies, personalize offers, and enhance customer experiences. Visual representation of this correlation could showcase trends in buying behavior across different demographic segments, allowing for more targeted and effective marketing campaigns aimed at specific customer groups.

