

capstone project

Retail Analytics



Rahul SINGH MARKAM

Accio job

Data Analyst

Overview

The retail industry is a highly competitive and data-rich environment that offers vast opportunities for businesses to leverage insights for improved decision-making. This project focuses on performing an in-depth analysis of retail data, examining multiple dimensions such as sales, customer behavior, product performance, and regional preferences. The aim is to help retail businesses identify growth opportunities, optimize product offerings, and enhance customer satisfaction.

The project is structured into two primary phases. The first phase involves conducting exploratory data analysis (EDA) using SQL and Excel, where key questions are addressed by querying the available data. This data encompasses a range of variables such as customer demographics, product categories, sales trends, and payment information. The EDA process uncovers insights about customer purchasing patterns, regional performance, and product attributes that drive sales. The second phase involves developing a comprehensive Power BI dashboard, enabling interactive visualizations to track critical metrics and trends across different business dimensions.

Ultimately, this project is designed to empower decision-makers with data-driven insights that will help improve marketing strategies, increase average order values, optimize product portfolios, and penetrate untapped markets. By answering key business questions and visualizing trends, this retail analysis offers practical recommendations to boost revenue, enhance customer loyalty, and achieve long-term business growth.

Objective

The primary objectives of this retail analysis project are as follows:

* Identify Factors Contributing to Sales Performance: Examine the variables that drive sales in different regions and highlight the factors contributing to the highest revenue generation.
* Analyze Customer Purchasing Patterns: Explore how customer behavior influences purchasing decisions and identify strategies to increase average order value.
* Pinpoint Key Drivers of Sales Growth: Investigate the factors that contribute to increased sales and assess how they can be leveraged to achieve sustainable growth.
* Evaluate Product Features and Attributes: Determine which product features are most appealing to customers and assess their influence on sales performance.
* Optimize Product Mix: Identify gaps in the product portfolio and explore opportunities for optimizing the product mix to meet evolving market demands.
* Segment Underperforming Markets: Examine market segments where certain products are underperforming and suggest improvement strategies.
* Enhance Customer Loyalty and Retention: Identify factors that drive repeat purchases and customer loyalty, offering insights to improve retention.
* Tailor Marketing Strategies to Regional Preferences: Customize marketing efforts based on geographic preferences and demographic indicators to maximize regional sales.
* Target High-Value Customers for Acquisition: Analyze the characteristics of high-value customers and develop strategies for acquiring similar customer segments.
* Increase Market Penetration in Untapped Markets: Identify potential markets with low penetration and devise strategies to expand business in those areas.

Significance

The significance of this retail analysis project lies in its potential to transform raw data into actionable insights that can directly impact business performance. The retail landscape is highly dynamic, with consumer preferences and market conditions constantly evolving. As such, businesses must continuously adapt their strategies to stay competitive. This project offers several key contributions:

Data-Driven Decision Making: By providing clear, data-backed insights, the analysis helps businesses make more informed decisions about product management, customer engagement, and regional sales strategies.

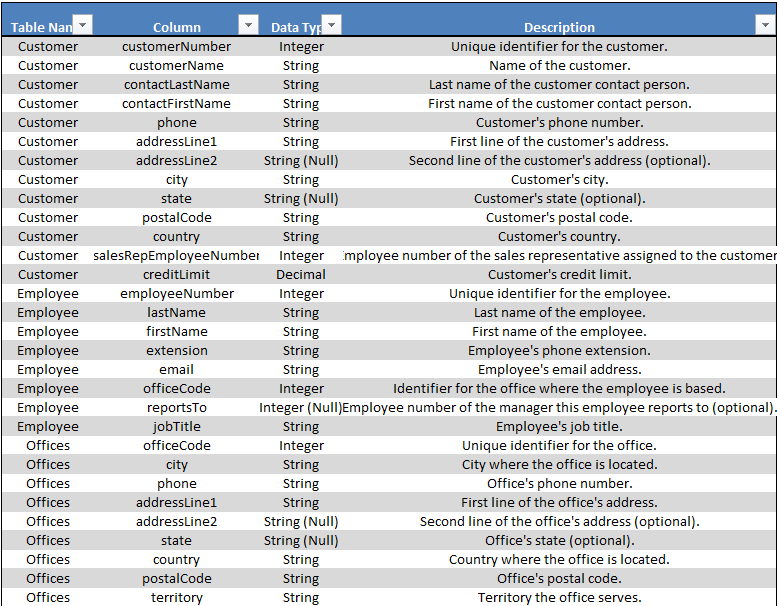
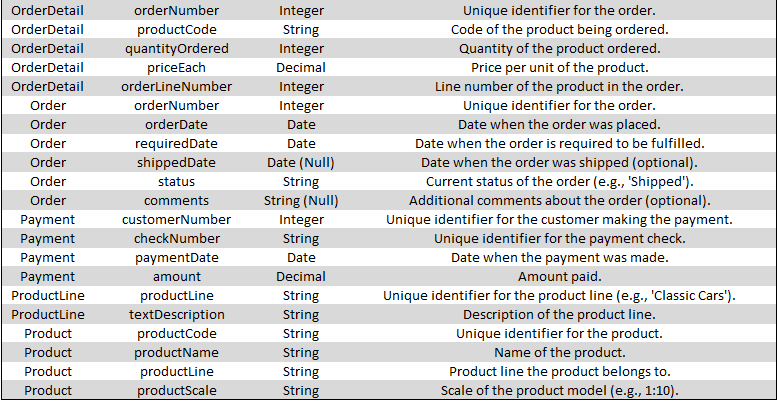
Enhanced Customer Understanding: Through the exploration of customer behavior and preferences, the project enables businesses to tailor their marketing campaigns and product offerings to better meet customer needs.

Optimized Sales and Marketing Strategies: The project highlights areas where marketing efforts can be refined to increase reach and effectiveness, particularly in underperforming regions or market segments.

Improved Business Growth and Profitability: By identifying key drivers of sales growth and uncovering opportunities for revenue optimization, the analysis offers practical recommendations to boost profitability.

Comprehensive Business Overview: The Power BI dashboard provides an interactive platform for tracking business performance across multiple dimensions, making it easier for stakeholders to monitor key metrics and trends in real-time.

Data Dictionary



# Table Name: customers

Column Names: customerNumber, customerName, contactLastName, contactFirstName, phone, addressLine1, addressLine2, city, state, postalCode, country, salesRepEmployeeNumber, creditLimit

Description: Contains customer details including customer number, name, contact information, address, sales representative, and credit limit.

# Table Name: employees

Column Names: employeeNumber, lastName, firstName, extension, email, officeCode, reportsTo, jobTitle

Description: Details of employees such as employee number, name, contact details, office code, reporting manager, and job title.

# Table Name: offices

Column Names: officeCode, city, phone, addressLine1, addressLine2, state, country, postalCode, territory

Description: Information about office locations including office code, address, phone, country, and territory.

# Table Name: order\_details

Column Names: orderNumber, productCode, quantityOrdered, priceEach, orderLineNumber

Description: Details of each order, including product code, quantity ordered, price per item, and order line number.

# Table Name: orders

Column Names: orderNumber, orderDate, requiredDate, shippedDate, status, comments

Description: Order-level information including order date, required date, shipped date, order status, and comments.

# Table Name: payments

Column Names: customerNumber, checkNumber, paymentDate, amount

Description: Customer payments, including customer number, check number, payment date, and payment amount.

# Table Name: product\_lines

Column Names: productLine, textDescription

Description: Product line descriptions providing details about product categories.

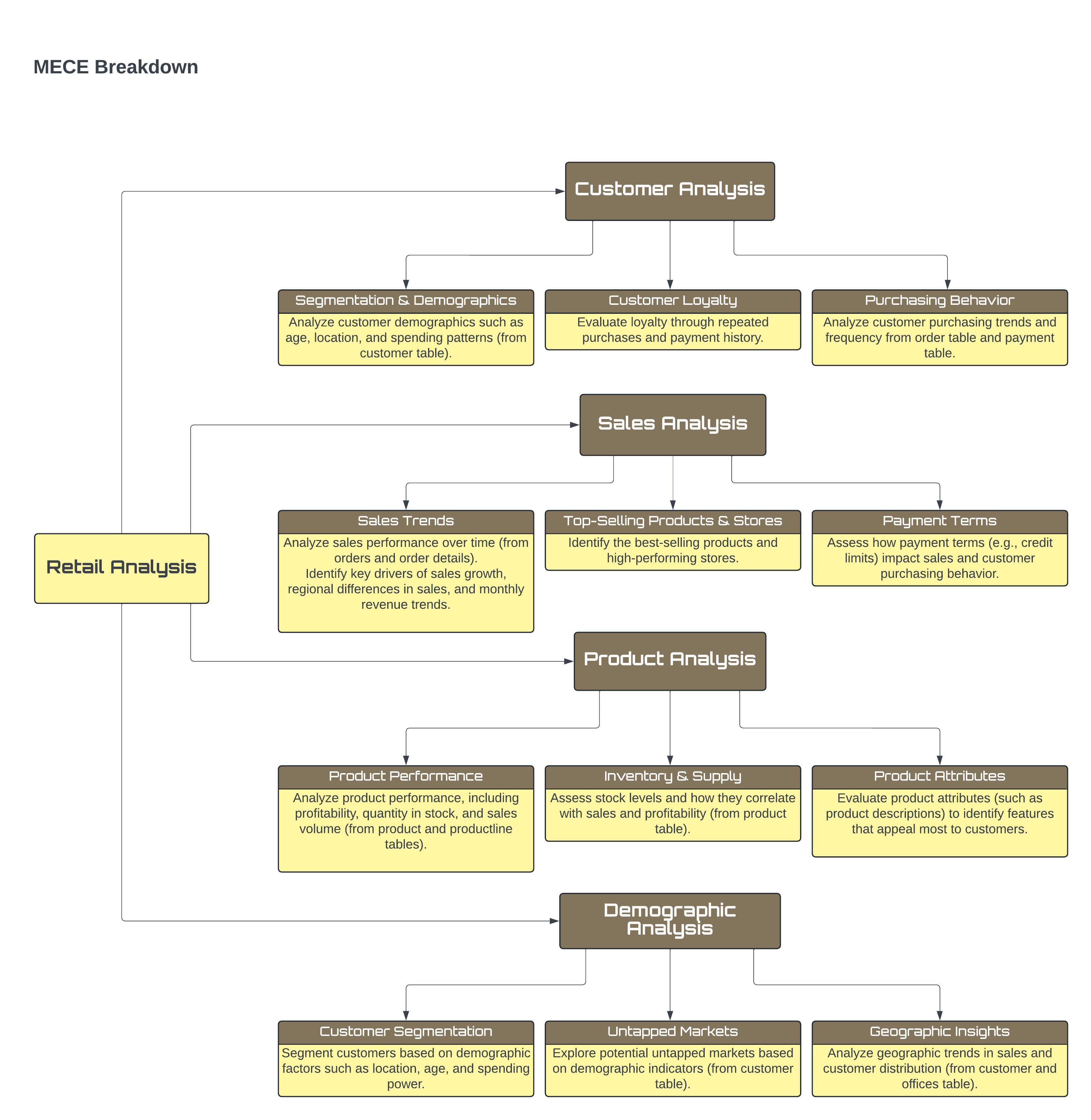
# Table Name: products

Column Names: productCode, productName, productLine, productScale, productVendor, productDescription, quantityInStock, buyPrice, MSRP

Description: Product information such as product code, name, line, scale, vendor, description, quantity in stock, buying price, and MSRP.

**Note :**

This data dictionary outlines the structure of several key tables used in a retail analysis database, which provides comprehensive information on customers, employees, orders, products, payments, and offices. These tables, when combined, enable the tracking and analysis of sales, customer relationships, product inventory, and operational logistics.

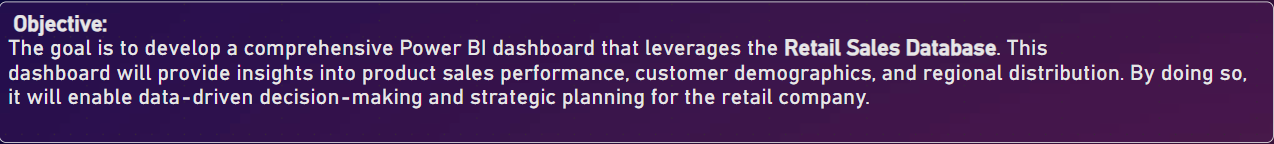
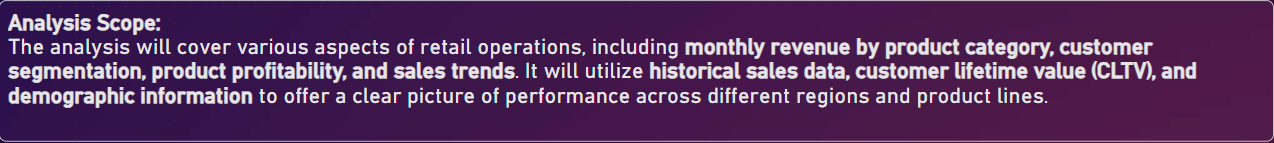
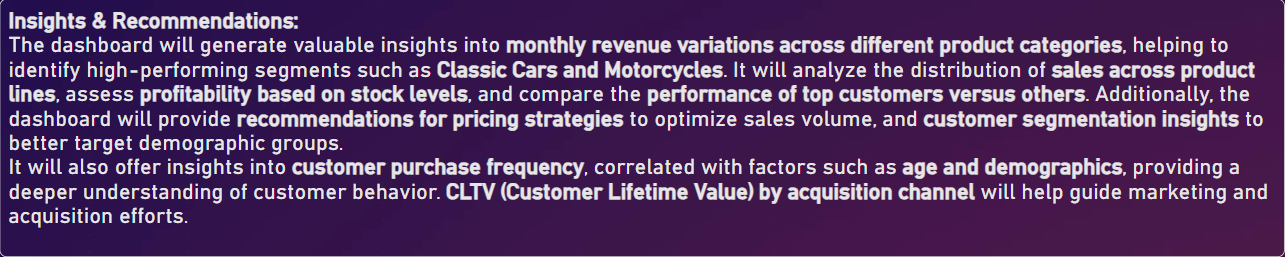
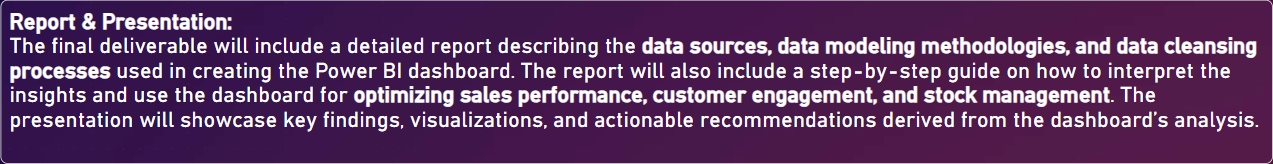
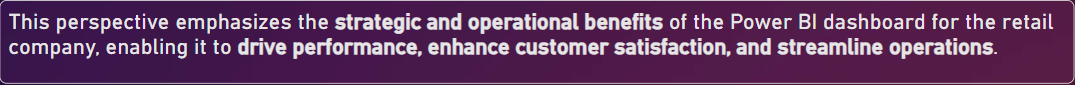






STATEMENT

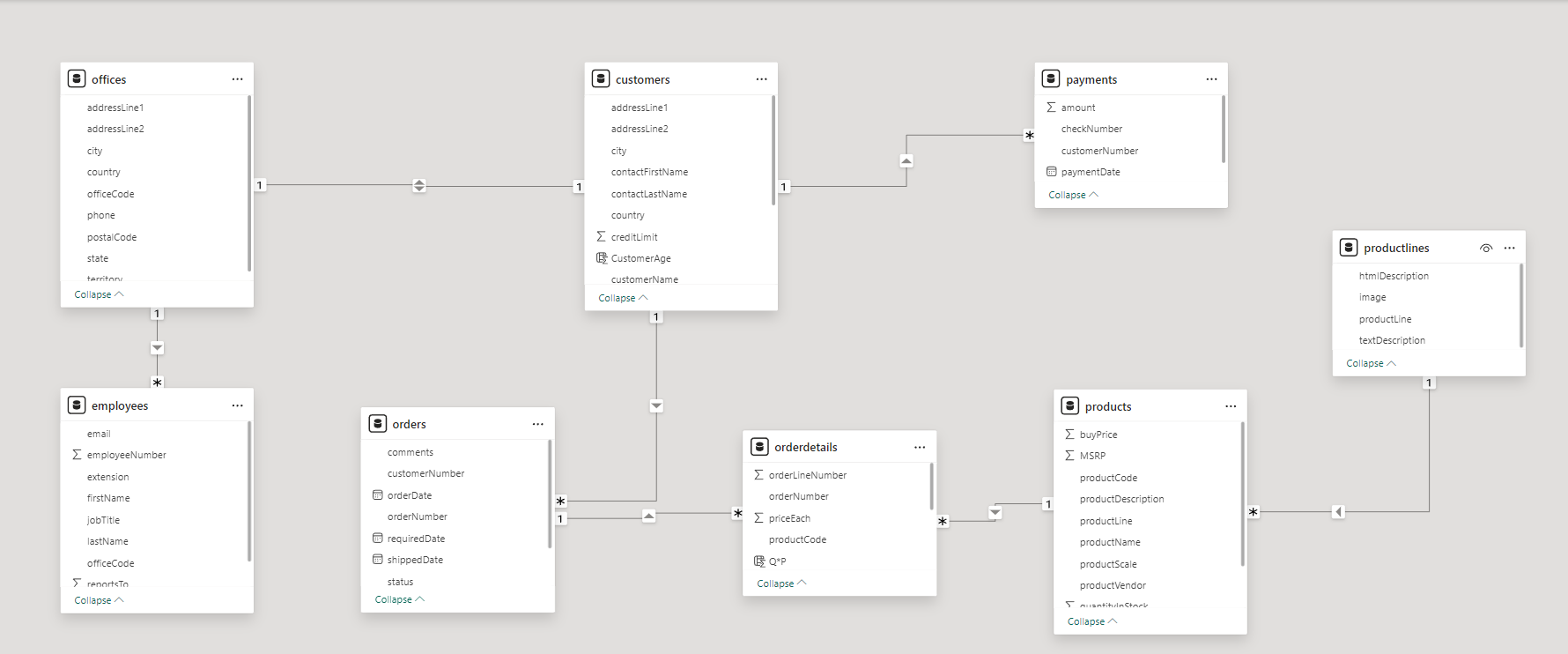
POWER BI – PROBLEM



ER Diagram

An **ER Diagram** (Entity-Relationship Diagram) in the context of Power BI is a visual representation of the relationships between different tables within a data model. It is primarily used to model how data flows between tables and how they are connected through relationships, usually by key fields.

In Power BI, this is referred to as the **Data Model View**, where tables from various sources (like databases, Excel files, etc.) are displayed, and relationships between them can be created or edited. The relationships in Power BI can be:

* **One-to-Many**: A single record in one table relates to multiple records in another table (e.g., one customer having multiple orders).
* **Many-to-Many**: Multiple records in one table relate to multiple records in another table (requires careful modeling).
* **One-to-One**: A single record in one table relates to a single record in another table.

Customer Analysis

# Segmentation & Demographics:

Analyze customer demographics such as age, location, and spending patterns (from customer table). Segment customers based on geographic location, credit limits, or purchase frequency.

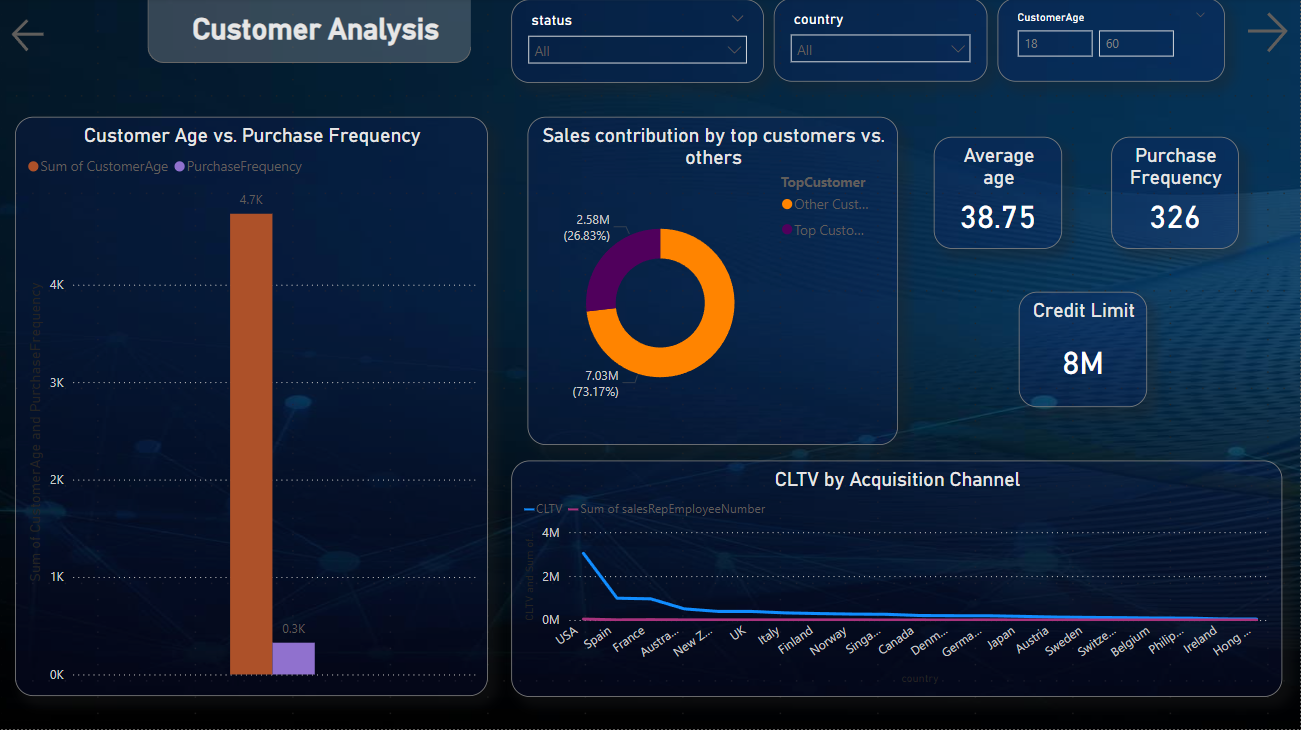
# Insights:

* Which demographics contribute most to sales in specific regions?
* What are the characteristics of high-value customers (repeat purchases, loyalty)?
* How do customer preferences vary by location, and how can this inform targeted marketing?

# Purchasing Behavior:

Analyze customer purchasing trends and frequency from order table and payment table.

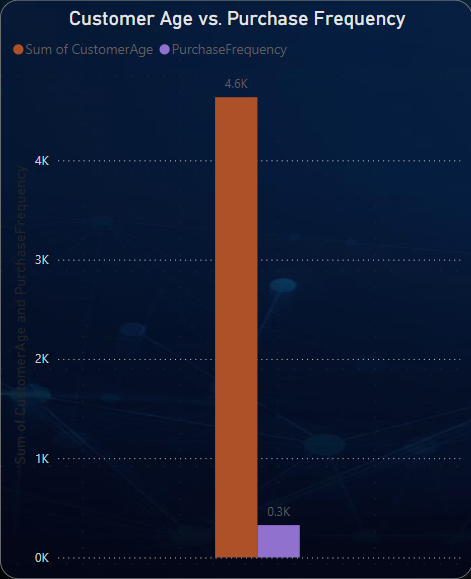
# Customer Loyalty:

Evaluate loyalty through repeated purchases and payment history.

Customer Age vs Purchase Frequency

What is the correlation between customer age and purchase frequency?

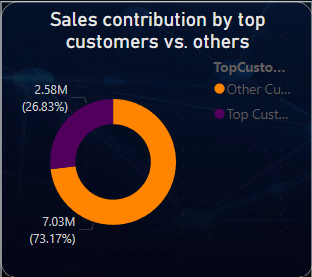
* The purchase frequency seems to be concentrated around customers aged 30-50.
* Interestingly, there is no clear increase in purchase frequency with age, suggesting that buying habits
* do not strongly correlate with age within the company's customer base.



Sales contribution by Top Customer Vs Other Customer

How does the sales performance of top customers compare to the rest?

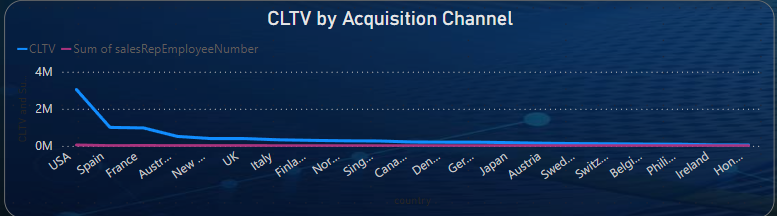
The "Sales Contribution by Top Customers vs. Others" doughnut chart clearly illustrates that top customers account for the majority of sales, contributing 7.03M (73.17%) of the total revenue. In contrast, other customers contribute 2.58M (26.83%). This highlights that top customers are crucial to the company's sales, significantly outperforming the general customer base and demonstrating their influence on overall sales performance.



CLTV by Acquisition Channel

How does customer lifetime value vary for different customer acquisition channels?

* The USA and Spain lead in Customer Lifetime Value (CLTV), with significantly higher values compared to other countries.
* The trend drops off sharply after the top-performing regions, showing a wide disparity in CLTV between major and minor markets.
* Channels contributing to regions like New Zealand and Germany see lower lifetime values.



Product Analysis

**Product Performance**:

Analyze product performance, including profitability, quantity in stock, and sales volume (from product and productline tables).

**Insights:**

Which product lines (e.g., Motorcycles, Classic Cars) are the most profitable?

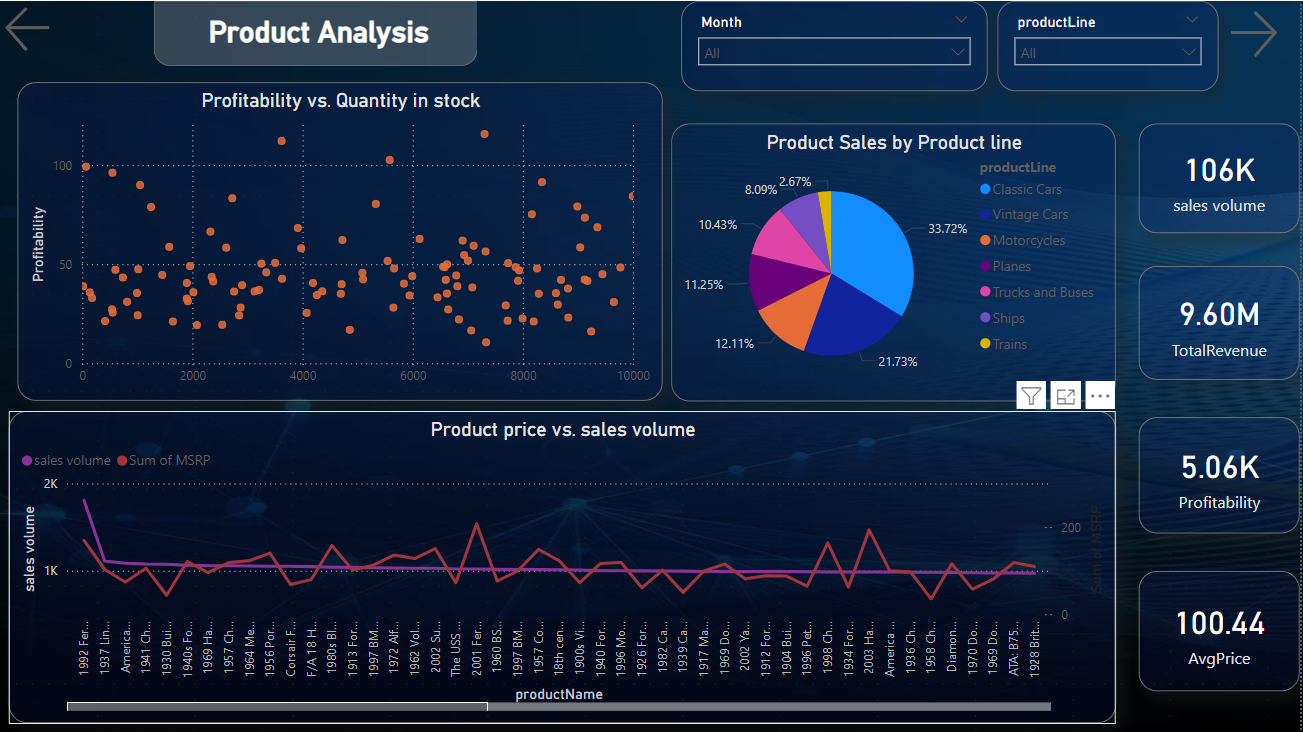
How does product pricing impact sales volume?

**Product Attributes:**

Evaluate product attributes (such as product descriptions) to identify features that appeal

**Inventory & Supply:**

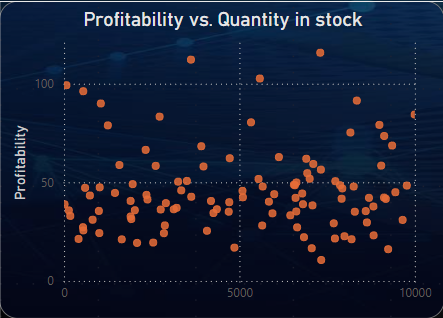
Assess stock levels and how they correlate with sales and profitability (from product table).



Profitability vs Quantity in Stock

How does the profitability of different products compare based on their quantity in stock?

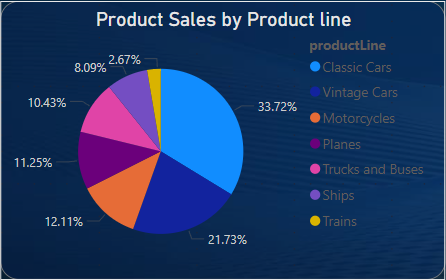
* There is a moderate spread of profitability across different products based on their stock.
* High profitability is not necessarily linked to high stock, as some products with lower stock quantities still demonstrate strong profitability.
* This suggests strategic pricing or high demand for certain items regardless of their availability.



Product Sales by Product Line

What is the distribution of product sales across different product lines?

* Classic Cars dominate with 33.72% of the total sales, followed by Motorcycles at 21.73% and Planes at 12.11%.
* Other categories like Trucks and Buses, Vintage Cars, and Ships make up smaller portions of the total sales.



Product Price vs Sales Volume

How does product pricing impact sales volume?

* There seems to be an inverse relationship: as MSRP (Manufacturer's Suggested Retail Price) increases, sales volume tends to decrease.
* Products with a lower price point tend to sell in higher volumes, indicating that customers are price-sensitive.
* Notably, sales for products priced above a certain threshold experience a sharp drop.

Sales Analysis

**Sales Trends:**

Analyze sales performance over time (from orders and order details).

Identify key drivers of sales growth, regional differences in sales, and monthly revenue trends.

**Insights:**

Which regions or periods contribute most to sales spikes?

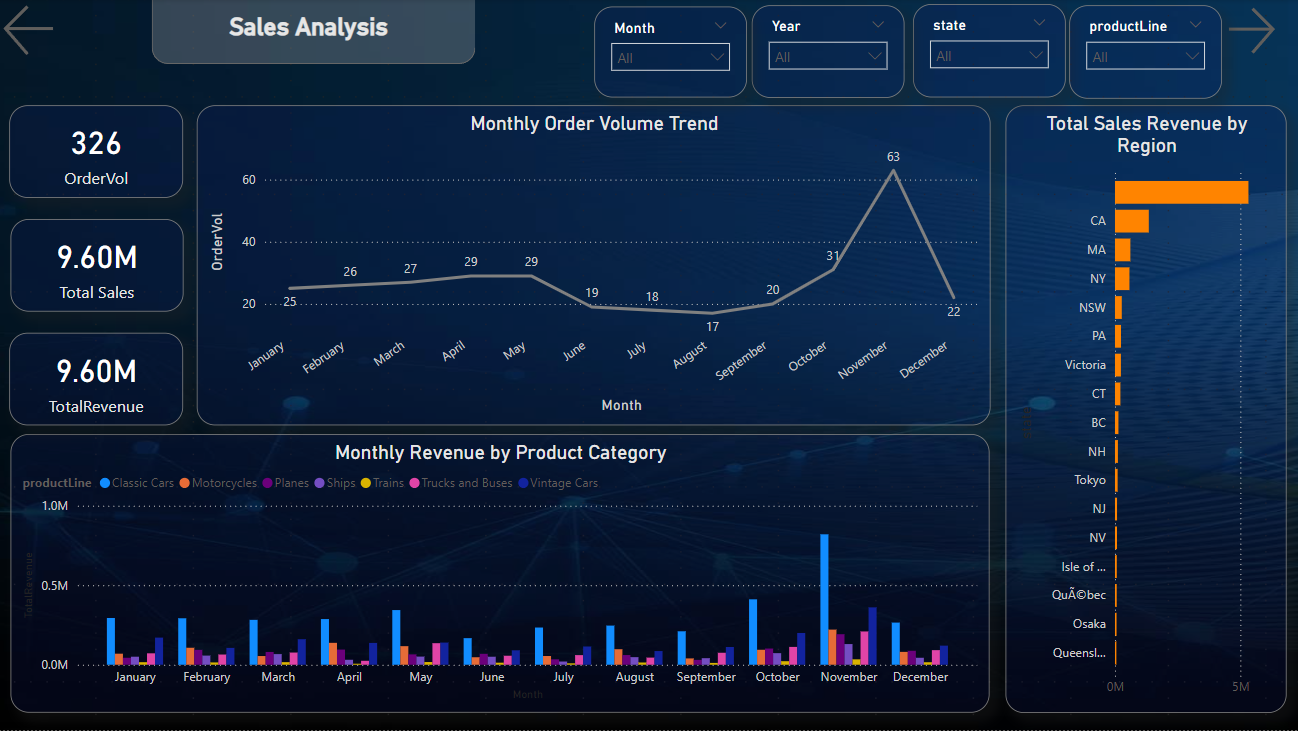
How can top-performing stores and sales representatives be leveraged?

**Top-Selling Products & Stores:**

Identify the best-selling products and high-performing stores.

**Payment Terms:**

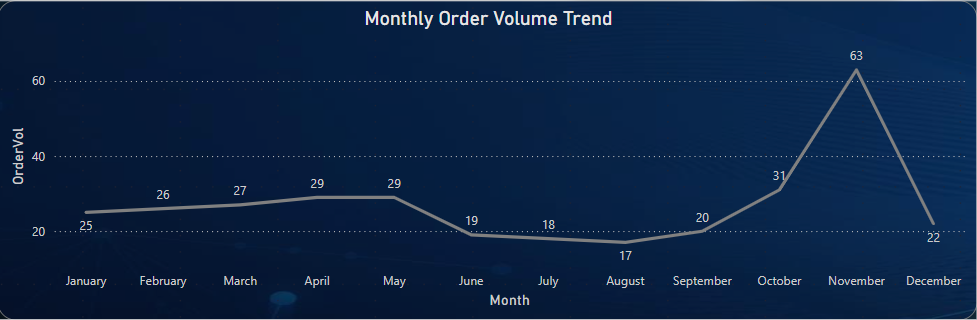
Assess how payment terms (e.g., credit limits) impact sales and customer purchasing behavior.



Monthly Order Volume Trend

What is the trend in customer order volume over the past year?

From the "Monthly Order Volume Trend" graph, it is evident that customer orders show a gradual fluctuation throughout the year. January through June maintain a relatively steady order volume ranging from 20 to 30 orders each month. However, a sharp increase is noticeable in December, where the order volume peaks at 63, suggesting a seasonal surge in orders, possibly driven by holiday promotions or year-end demand. There is a slight dip during July and September, indicating off-peak periods.



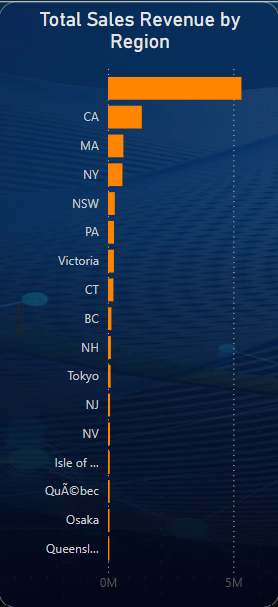
Monthly Revenue by Product Category

How does monthly revenue vary across different product categories?

* The Monthly Revenue by Product Category chart shows distinct variations in revenue across product categories like Classic Cars, Motorcycles, Planes, and others.
* Classic Cars consistently generate the highest revenue throughout the year, peaking around July and November.
* Categories like Planes, Trucks and Buses, and Vintage Cars contribute less to the total monthly revenue.
* Revenue across all categories tends to fluctuate month by month, with noticeable increases in revenue during July and December.

Sales Revenue by Region

What are the top regions in terms of sales revenue?

* From the "Total Sales Revenue by Region" bar chart, California (CA) leads as the top-performing region with over 5M in sales revenue. Massachusetts (MA) and New York (NY) follow, though their contributions are significantly smaller in comparison. California’s dominance in sales indicates its importance as a major revenue-driving region for the company, while other regions contribute more modest amounts.

Demographic Analysis

**Customer Segmentation:**

Segment customers based on demographic factors such as location, age, and spending power.

**Insights:**

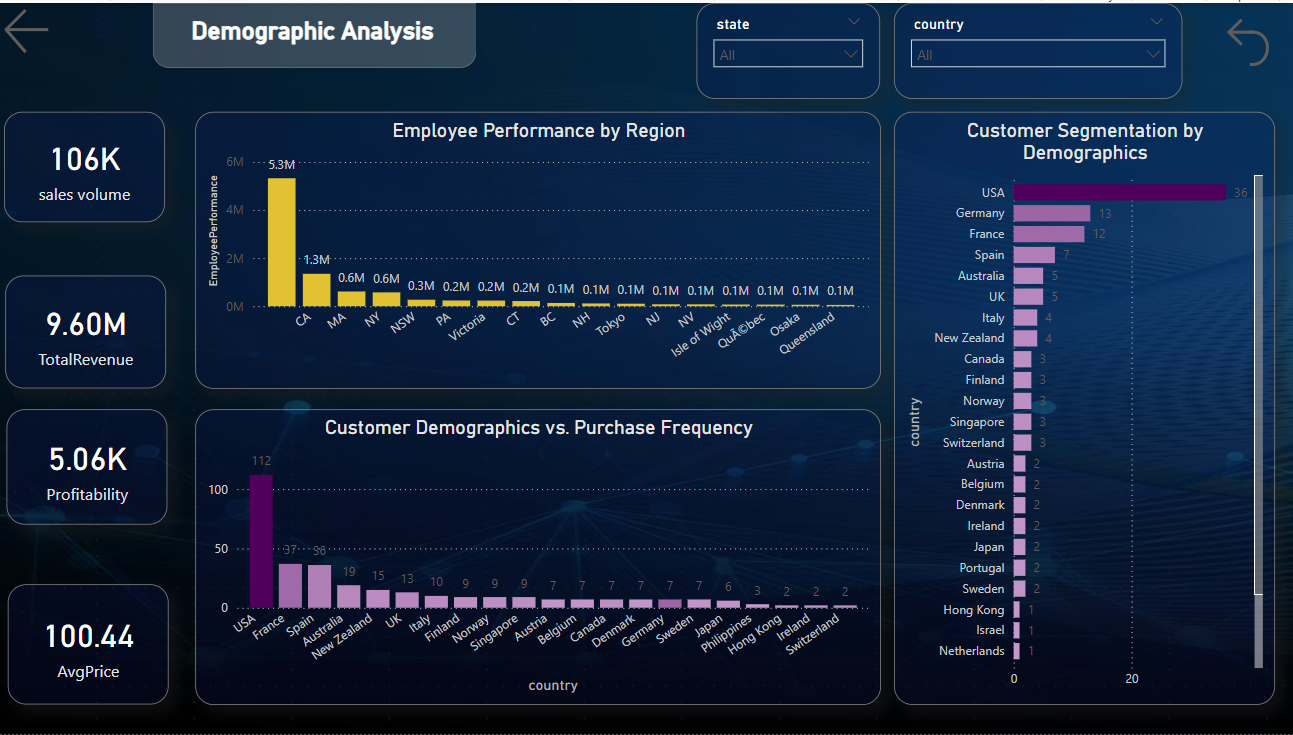
How does customer lifetime value (CLTV) vary by demographic segment?

How do preferences vary across different regions, and what marketing strategies can be employed?

**Geographic Insights:**

Analyze geographic trends in sales and customer distribution (from customer and offices table).

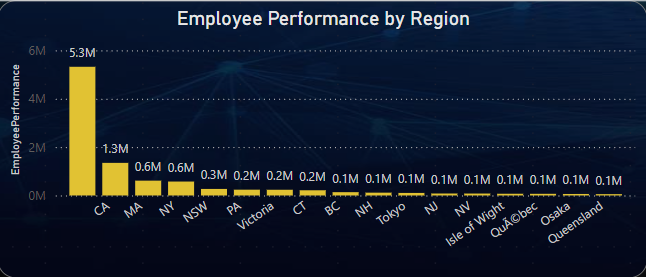
**Untapped Markets**:

Explore potential untapped markets based on demographic indicators (from customer table).

Employee Performance by Region

How does the performance of sales employees vary across different regions?

* The "Employee Performance by Region" chart reveals a substantial difference in performance across regions. California employees generate the highest revenue, with a notable figure of 5.3M, while employees in regions like Massachusetts (1.3M) and New York (0.6M) also perform well but lag far behind California. Other regions like New South Wales (NSW), Pennsylvania (PA), and Victoria show significantly lower performance, which could suggest the need for improved sales strategies or resource allocation in those areas.

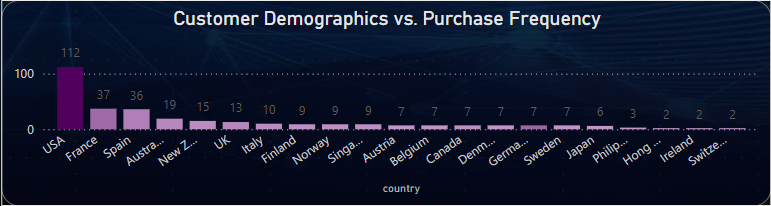


Customer Demographics vs.

Purchase Frequency

What is the correlation between customer demographics and purchase frequency?

* There’s a clear correlation between regions with a larger customer base and higher purchase frequencies. For example, USA has both a high number of customers and higher purchase frequency.
* Smaller customer bases in regions like Ireland, Poland, and Israel correspond to lower purchase frequencies, indicating that demographic size might influence how often customers purchase.



Customer Segmentation by Demographics

What is the distribution of customers across different demographic segments?

* The USA has the highest number of customers, followed by Germany and France.
* There is a clear concentration of customers from Europe and North America, with minimal representation from smaller countries like Japan, Singapore, and Israel.
* This indicates that the company’s customer base is primarily from developed, high-income regions.



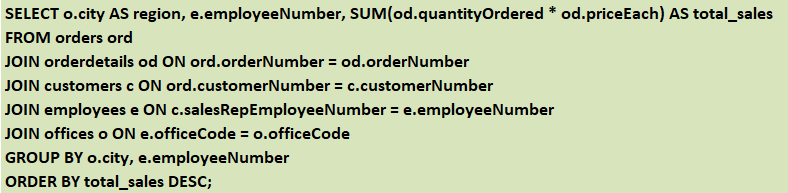
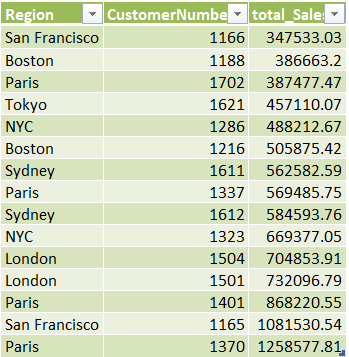
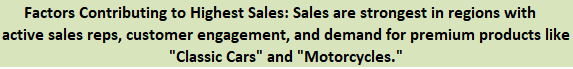


STATEMENT

EDA– PROBLEM

Question 1

Which factors contribute to the highest sales in a particular region?



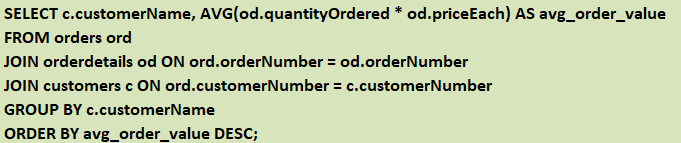
Conclusion

OUTPUT

QUERIES

Question 2

How can customer purchasing patterns be influenced to increase average order value?



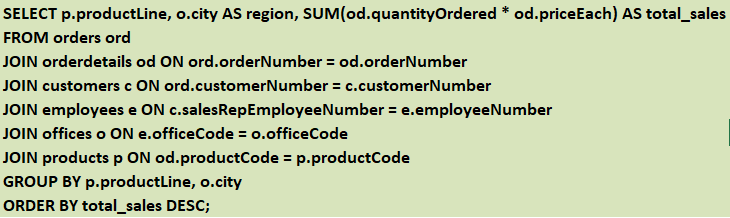
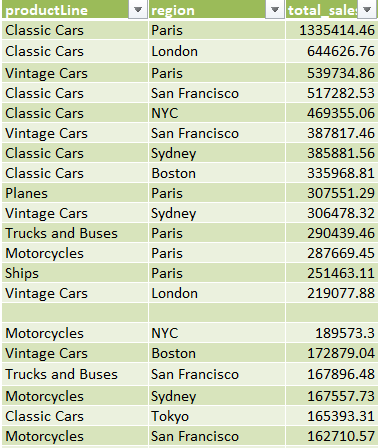
OUTPUT SAMPLE

QUERIES

Conclusion

Question 3

What are the key drivers of sales growth, and how can they be leveraged for future success



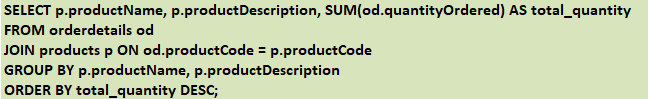
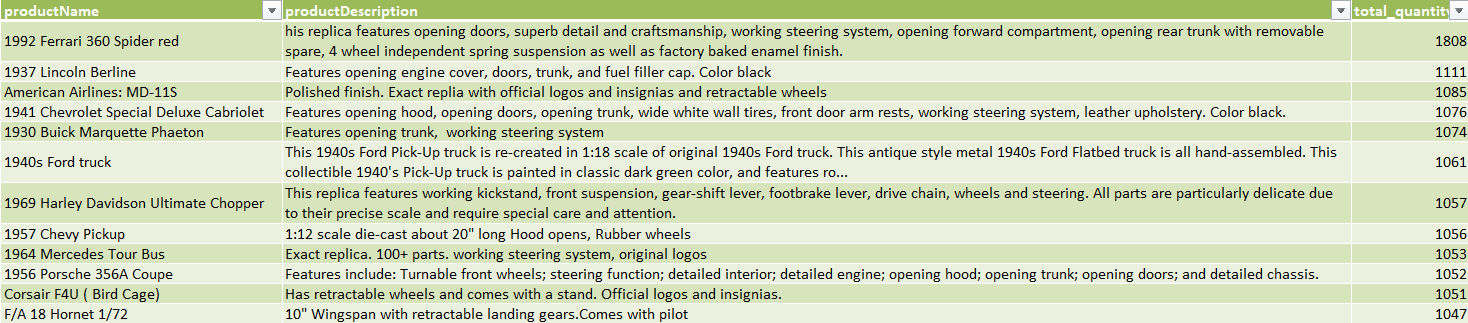
Conclusion

OUTPUT SAMPLE

QUERIES

Question 4

Which product features or attributes are most appealing to customers?



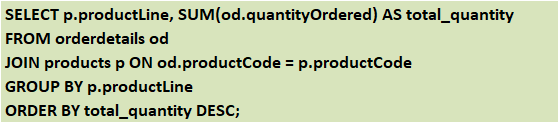
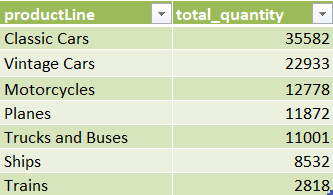
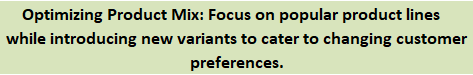
Conclusion

OUTPUT SAMPLE

QUERIES

Question 5

How can the product mix be optimized to cater to changing market demands?



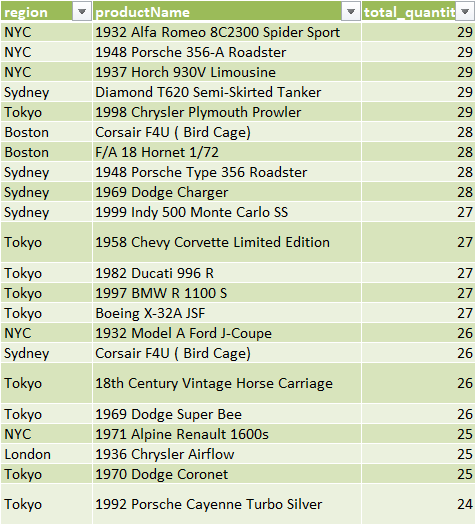
Conclusion

OUTPUT

QUERIES

Question 6

Are there any specific market segments where a particular product is underperforming, and how can it be improved?



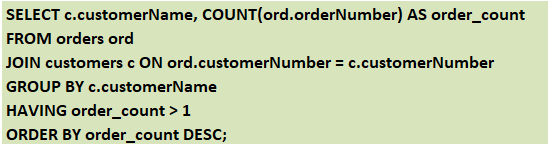
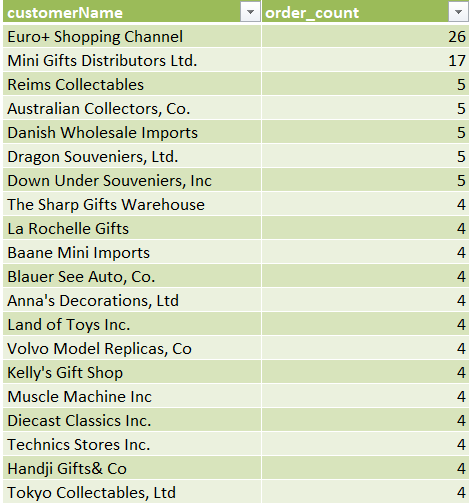
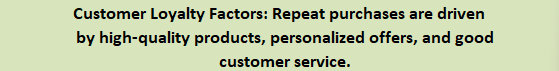
QUERIES

Conclusion

OUTPUT SAMPLE

Question 7

What are the main factors that influence customer loyalty and repeat purchases



QUERIES

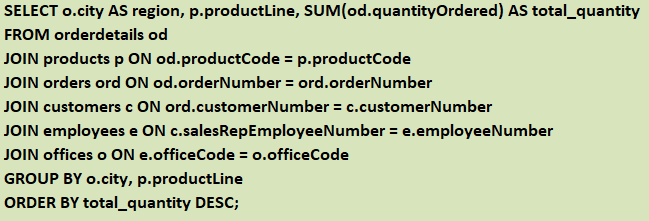
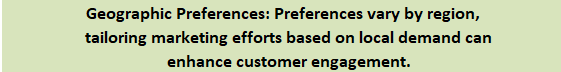
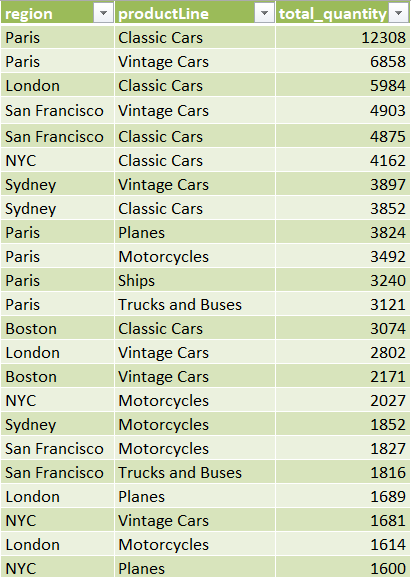
Conclusion

OUTPUT SAMPLE

Question 8

How do customer preferences differ based on geographic location, and how can marketing campaigns be customized accordingly?

QUERIES

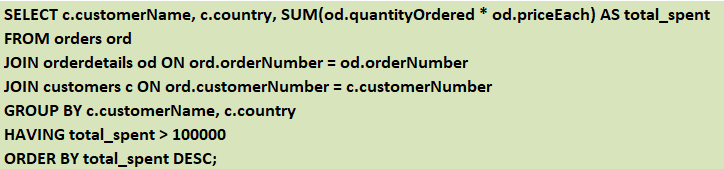
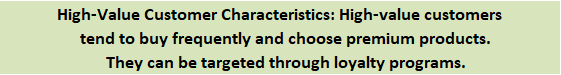


Conclusion

OUTPUT SAMPLE

Question 9

What are the characteristics of high-value customers, and how can similar customers be targeted for acquisition?



Conclusion

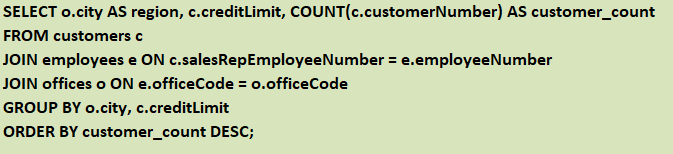
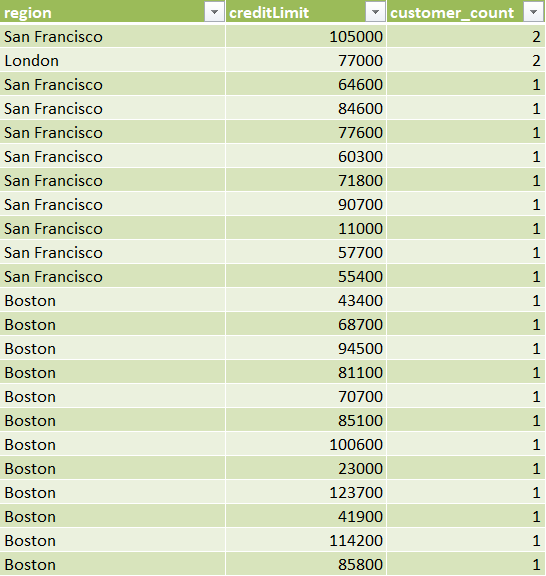
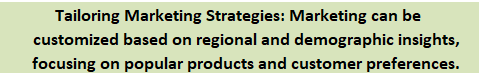
OUTPUT SAMPLE

QUERIES

Question 10

How can marketing strategies be tailored to target specific demographic segments in different regions?

QUERIES

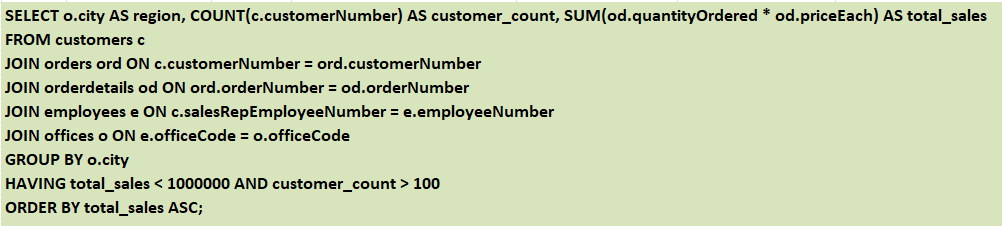
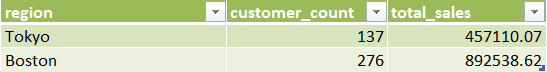
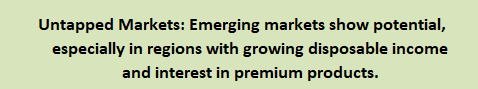


Conclusion

OUTPUT SAMPLE

Question 11

What are the potential untapped markets based on demographic indicators, and how can market penetration be increased?



Conclusion

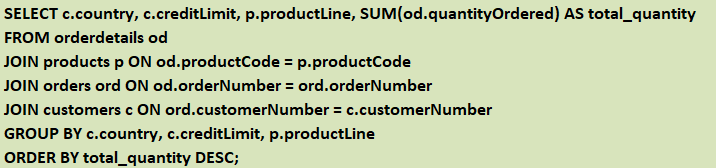
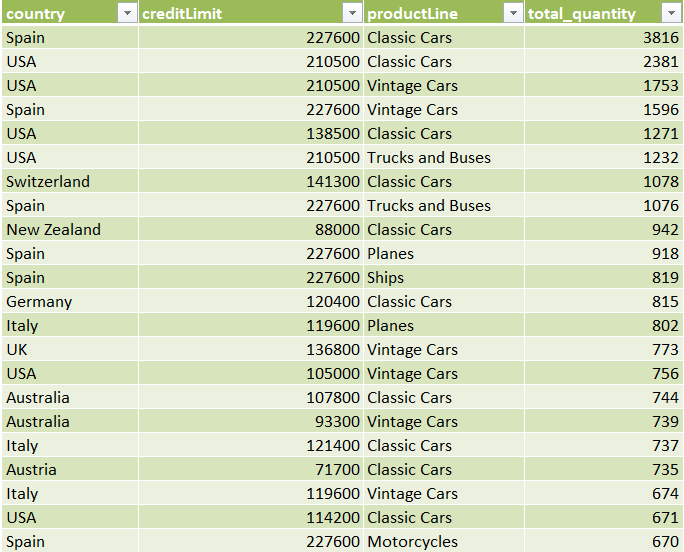
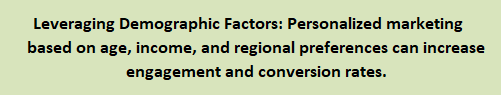
OUTPUT

QUERIES

Question 12

How do customer preferences and behavior differ based on demographic factors, and how can they be leveraged for personalized marketing campaigns?

QUERIES



Conclusion

OUTPUT SAMPLE