

# **CAPSTONE PROJECT**

## **RETAIL ANALYSIS**



- **THANU SREE**

## **OVER VIEW**

The capstone project, titled " **Retail Analysis**"

The objective of this project is to develop a Power BI dashboard using the Retail Database to provide comprehensive insights into the retail business's performance. The dashboard will focus on sales, product, customer, and demographic analysis, aiming to facilitate data-driven decision-making, optimize sales strategies, and enhance customer experiences to improve business strategies and achieve success in the competitive retail market.

# **THE PROCESS**

## **1. Data Acquisition from GitHub:**

Retrieve the related data from GitHub respiratory provided link. Retrieve the meteorological dataset from a specified source that contains crucial information about weather patterns, encompassing diverse regions and their climatic performance across distinct meteorological systems.

## **2. Data Cleaning and Transformation:**

In this phase, we focus on handling missing values, implementing data normalization procedures, and effectively managing outliers within the meteorological dataset. Dealing with missing values includes employing techniques such as imputation or exclusion based on the extent of missing data. Normalizing the data is crucial to standardize the scale of meteorological variables, ensuring equitable comparisons across diverse regions. Managing outliers involves identifying and appropriately addressing data points that significantly deviate from the overall climatic patterns, minimizing their impact on the analysis.

The Transformation of this weather analysis related data has been gone through several steps. Based on the date time common every table has been changed and the data sets has converted as per convenient to the problem statement, using SQL and Power Bi.

## **3. Connection with Tools:**

Incorporating the datasets involves connecting them seamlessly with analytical tools and establishing interfaces with various data processing platforms. This integration includes linking the dataset with Power BI, MS-Excel, and SQL Workbench to ensure a smooth flow of data and accurate processing. By establishing these connections, we enable efficient data

integration, allowing for comprehensive analysis and exploration of the dataset across different analytical tools.

#### **4. Problem Solving in Power Bi:**

Utilize Power BI's versatile visualization capabilities to address a diverse set of problem statements by employing a variety of charts and visualizations. Through detailed data manipulation and analysis, unveil crucial insights and trends that provide valuable information for making well-informed decisions.

#### **5. Exploratory Data Analysis (EDA):**

Conducting EDA utilizing either Excel or MySQL Workbench, the choice of which depends on the specific requirements of the analysis. The primary objective is to extract meaningful insights from the analysis and effectively utilize them for informed decision-making.

#### **6. Creating of Visual and Insightful Power Point Presentation (PPT):**

Creating a comprehensive and insightful PowerPoint presentation that effectively communicates the key information related to the entire problem statement. The presentation will include data relevant to the given questions, detailing the appropriate problem-solving processes. Additionally, it will encompass conclusions, insights, and other pertinent aspects of the analysis.

#### **7. Detailed Documentation:**

Prepare a comprehensive report that thoroughly documents the entire project lifecycle. The report should include sections covering data collection, transformation, formulation of problem statements, integration of tools, solutions developed using Power BI, insights derived from Exploratory Data Analysis (EDA), and visualizations presented in PowerPoint. This detailed documentation will provide a holistic overview of the project's progression and outcomes.

## **OBJECTIVES**

The objective of this project is to develop a Power BI dashboard using the Retail Database to provide comprehensive insights into the retail business's performance. The dashboard will focus on sales, product, customer, and demographic analysis, aiming to facilitate data-driven decision-making, optimize sales strategies, and enhance customer experiences. The goal is to empower retail stakeholders with actionable insights, enabling them to identify top-selling products, customer preferences, and target demographics. The dashboard will offer valuable insights and recommendations for targeted marketing, inventory optimization, and personalized customer experiences. The final deliverables will include a report and presentation showcasing the dashboard's findings and significance, serving as a powerful tool for retail stakeholders to improve business strategies and achieve success in the competitive retail market.

## Significance

retail analytics to explain past operational and financial performance, diagnose what might have gone wrong, suggest alternative approaches that would have been more productive, forecast demand, and offer suggestions, sometimes in real time, that store associates, customer service agents, and others can use to cross-sell, upsell, or improve the customer's experience. In all cases, the tools are intended to help retailers boost sales, profits, and customer satisfaction.

- **In-store analytics tools** use data generated from POS systems and in-store video cameras to help retailers analyze customer shopping patterns so they can place products more effectively in aisles, ensure appropriate inventory levels, and reduce theft. Video footage, for example, can show whether customers are slowing down to look at a given display, while POS system data can show the effectiveness of merchandising on customers who use their loyalty cards.
- **Customer analytics** uses data from systems that customers interact with, including POS systems, websites, phone logs, and customer service chats. Analyzing this data helps retailers determine which and where certain items are most popular, why certain items are being returned or exchanged, or what promotions or suggestions are most effective with customers. For example, it can help determine what marketing language is most effective over the phone, as opposed to in chats, to promote a new item.
- **Inventory analytics**, as the name suggests, assesses inventory levels for goods a retailer has on offer. It's used to prescribe more efficient warehousing and distribution strategies, such as when a distribution center is preferable to a more local storehouse, and when to replenish items based on inventory levels and projected demand. Inventory analytics can, for example, reduce the labor and shipping costs associated with carrying too much safety stock.
- **Merchandise analytics** helps retailers determine whether they're displaying their wares effectively, mostly in physical stores, with the goal of enticing consumers to make a purchase by using compelling assortments or offers. Merchandise analytics also helps retailers adjust prices to increase profit margins across products.
- **Web analytics** tracks the digital footprint of consumers as they linger over certain parts of a web page or click from one page to another. It follows them from the source that led them to the site to the moment they leave. This type of analytics helps online retailers decide how and where to display their goods on the site, the prices they charge, and the marketing promotions they should run.

- **Business intelligence (BI)** reports, often presented in the form of dashboards, are preset to show certain key performance indicators, such as inventory turns and sell-through rate. They are used mainly to share top-line trends with peers and senior management.
- **Demand forecasting** Forecasting demand for particular items sold online based on the path customers followed to view those items, move them to their shopping cart, remove those items, or abandon the cart entirely. While those actions aren't counted as sales, they can extrapolate future demand.
- **Sales forecasting** helps retailers predict future sales based on actual sales figures and other factors. Used in tandem with demand forecasting, it can predict what total demand will be for an item across all channels and can help retailers ensure they have the necessary inventory to fulfil that demand.

# DATA DICTIONARY

## Dataset Description

The retail dataset provided contains valuable information about a retail company, capturing various aspects of its operations. This dataset is essential for understanding and analyzing the company's offices, employees, customers, products, product lines, orders, order details, and payments.

### Table Explanations

#### 1. Table name - Office Table

**Purpose:**

This table stores information about the different offices of the retail company

**Columns:**

Office Code

City

Phone number

Address

State

Country

Postal Code

Territory

**Usage:**

Each office is uniquely identified by its office code.

#### 2. Table name – Employees Table

**Purpose:**

The employees table holds data about the company's employees.

**Columns:**

Employee Number

Last Name

First Name

Extension

Email Address

Office Code

**Usage:**

Used to know about Employees details



3. **Table name** – *Customers Table*

**Purpose:**

This table contains information about the retail company's customers.

**Columns:**

Customer Number

Customer Name

Contact Last Name

Contact First Name

Phone Number

City

State

Postal code

Country

Retail company's customers.

**Usage:**

Used to know about customers details

4. **Table name** – *Products Table*

**Purpose:**

This table contains information about products.

**Columns:**

Product code

Product Name

Product line

Product Scale

product vendor

product description

quantity in stock

buy price

Manufacturer's Suggested Retail Price (MSRP).

**Usage:**

Used to know about Product details

5. **Table name** – *Product lines Table*

**Purpose:**

This table is used to describe the different product lines available in the company's inventory.

**Columns:**

Product line

product line name,

text description,

HTML description (for web-based content)

an image (stored as a BLOB) Product Lines Table

**Usage:**

Used to know about Product details

6. **Table name** – *Orders Table*

**Purpose:**

This table contains information about Orders.

**Columns:**

order number (a unique identifier for each order)

order date

required date

shipped date

order status (using an ENUM for predefined statuses)

comments

customer number (identifying the customer who placed the order).

**Usage:**

Used to know about Orders

7. **Table name** – *Order details Table*

**Purpose:**

This table contains information about Order details.

**Columns:**

order number (linking to the orders table)

product code (linking to the products table)

quantity ordered

price per item

order line number.

**Usage:**

Used to know about Orders

8. **Table name** – *Payments Table*

**Purpose:**

This table contains information about Payments.

**Columns:**

customer number (linking to the customers table)

check number

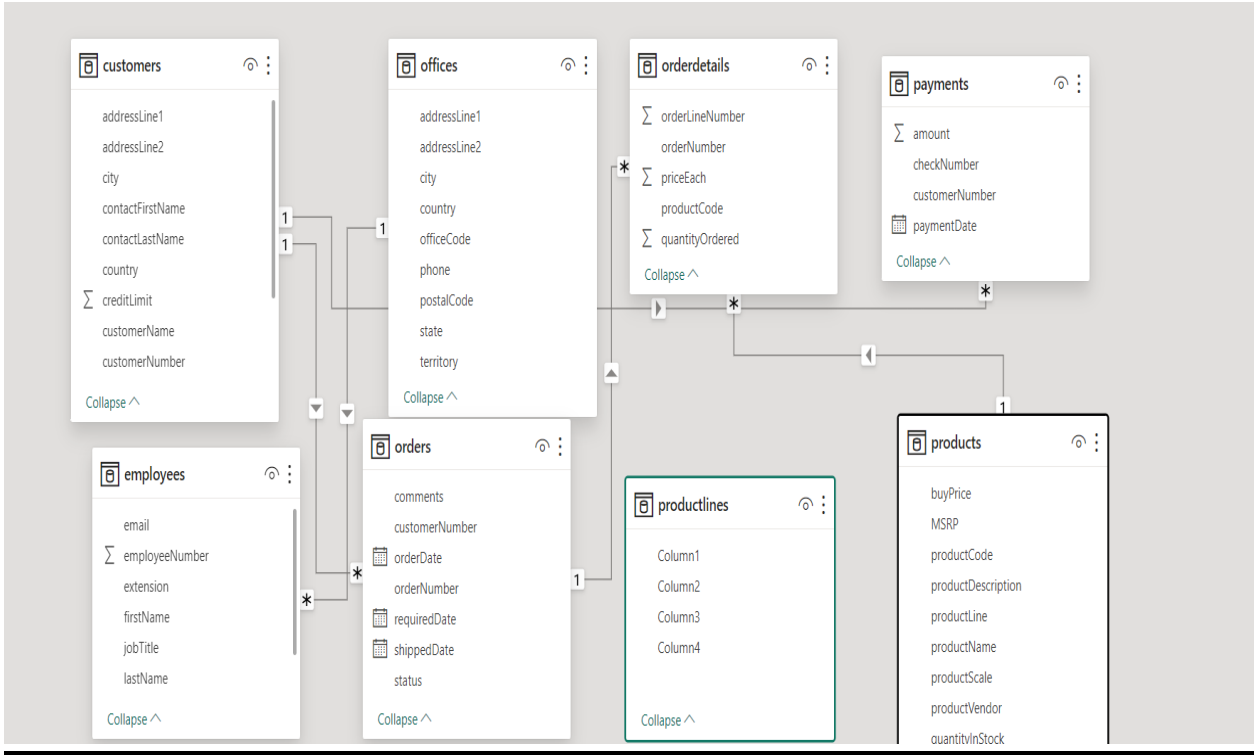
payment date

payment amount.

**Usage:**

Used to know about Payments

# ER DIAGRAM



# **POWERBI PROBLEM STATEMENTS**

## **Retail analysis is the combination of Sales Analysis, Product Analysis, Customer Analysis, Demographic Analysis**

**Sales Analysis:** Sales analysis is reviewing your sales data to identify trends and patterns. Sales data can help you make better decisions about your product, pricing, promotions, inventory, customer needs other aspects of your business. Sales analysis can be as simple as reviewing your sales figures regularly.

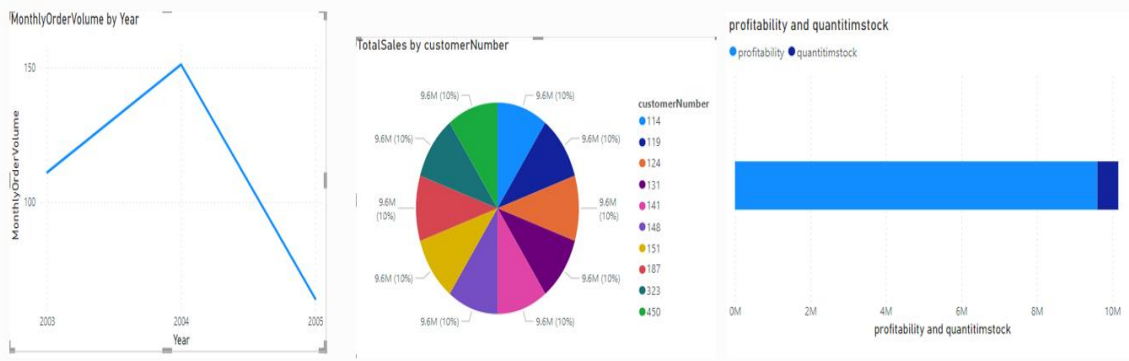
**Product Analysis:** Product analysis is the process of understanding user behavior in regards to your product or service. Product analytics data is used to determine what kind of user experience your product or service offers, using real customer feedback and behavioral data to inform future product updates or improvements.

**Customer Analysis:** Customer analytics are a form of data that shows what customers are doing and why they are doing it. This includes information like where customers shop, why they visit those stores, what they buy, and which factors influence their purchase decisions.

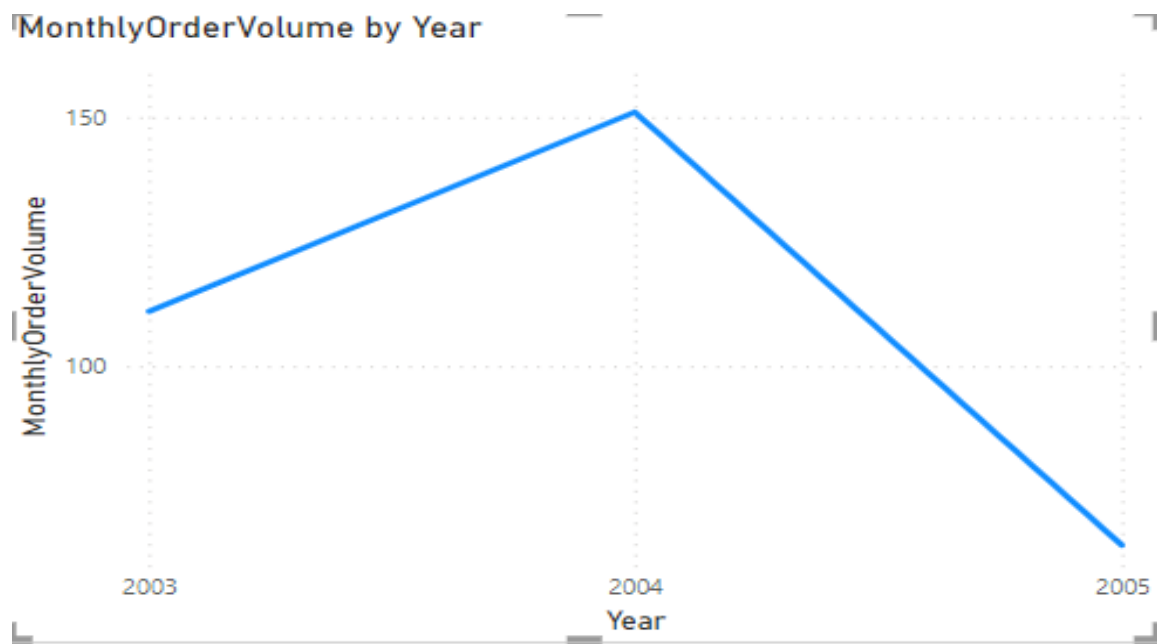
**Demographic analysis:** Demographic analysis is the collection and analysis of broad characteristics about groups of people and populations. Demographic data is very useful for businesses to understand how to market to consumers and plan strategically for future trends in consumer demand.

# Sales Analysis

## Sales Analysis

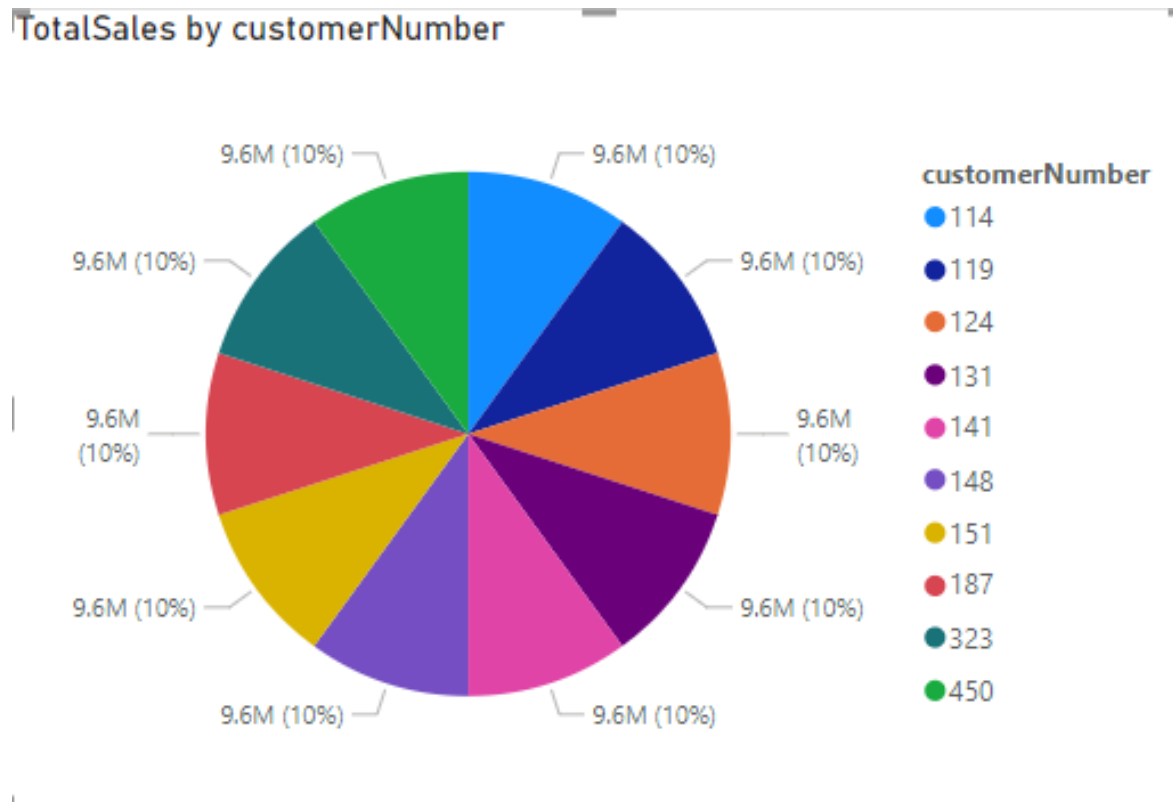


Problem Statement 2Q: What is the trend in customer order volume over the past year? (Visual: Monthly order volume trend)



Explanation: In this line chart. the year is on X – axis and the monthly order value is on y-axis in the chart. Where, 2003 the trend was increased and 2004 it increased compared to 2003 but when comes to 2005 Monthly order value it got completely down. So the trend was like this.

3Q. How does the sales performance of top customers compare to the rest?



Explanation: The sales performance of customer was very happy and all the 10 customers they have equal contribution, having 10% of sales equally here in that particular chert



5Q. How does the profitability of different products compare based on their quantity in stock?

profitability and quantitimstock

● profitability ● quantitimstock

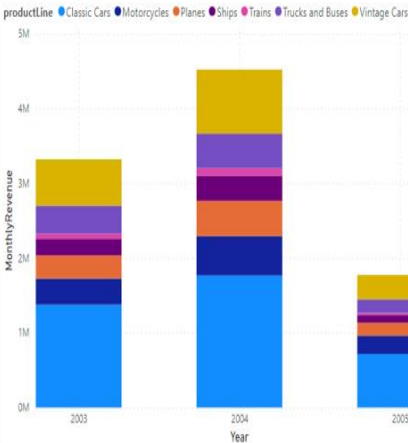


Explanation : Here in this graph the profitability the , the profitability has more approach and the quantity stock has less, the profitability of some products can based on the quantity in the stock only.

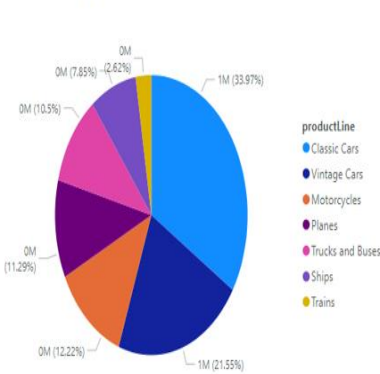
# Product Analysis

## Product Analysis

MonthlyRevenue by Year and productLine



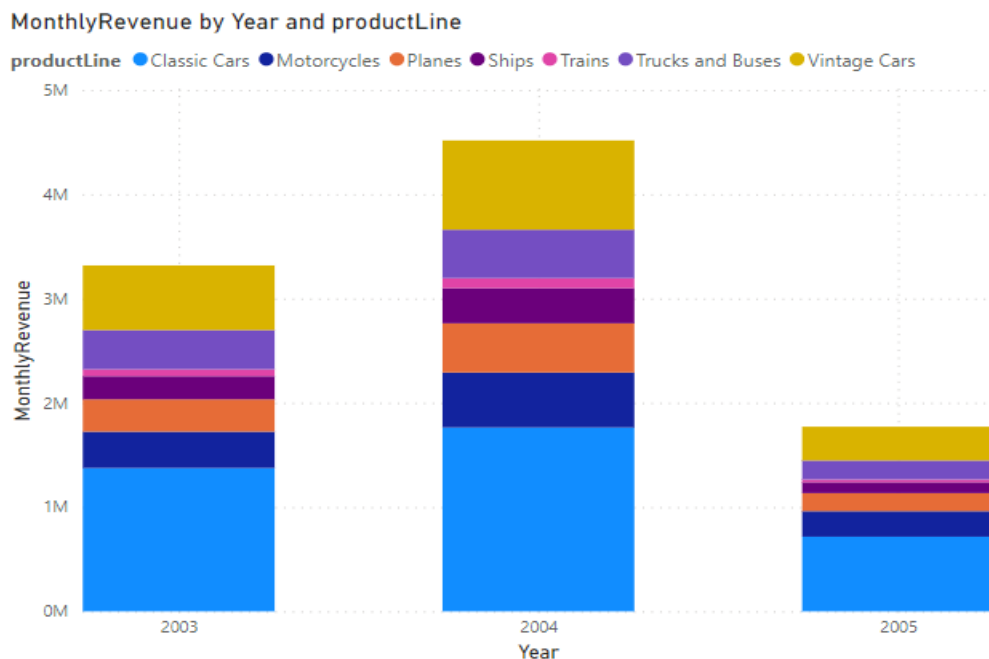
ProductSales by productLine



Count of Totalsales and Sum of priceEach



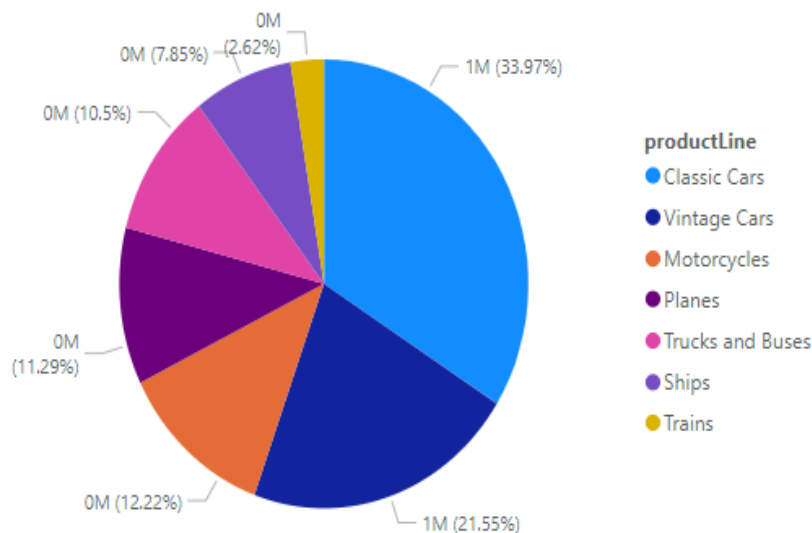
1Q. How does monthly revenue vary across different product categories? (Visual: Monthly revenue by product category)



Explanation: This graphs shows that monthly revenue among different product lines (categories) ,on 2004 the monthly revenue is increased on various product lines, and then decreased in 2005.

4Q. What is the distribution of product sales across different product lines? (Visual: Product sales by product line)

ProductSales by productLine

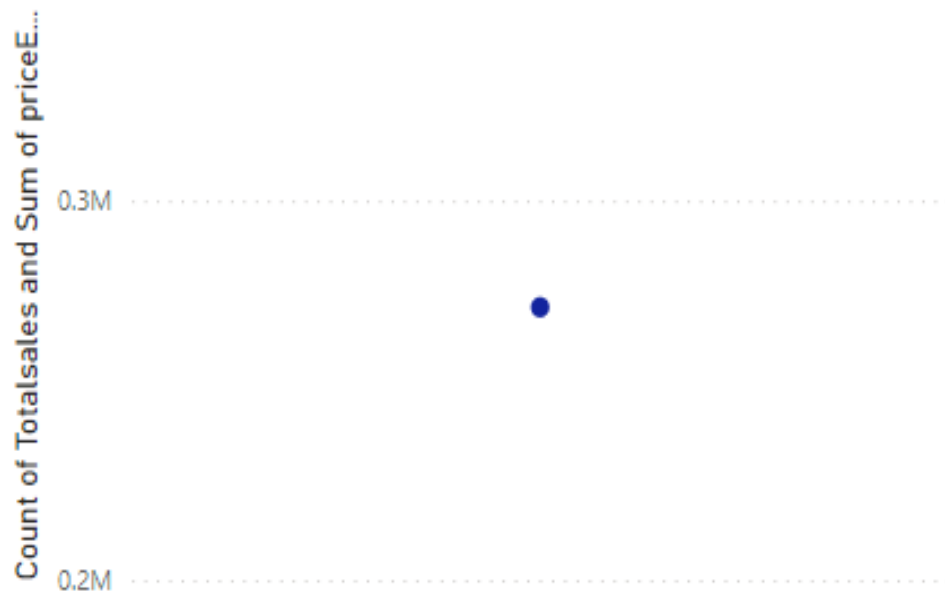


Explanation: This graph show that Classic cars product line has the highest distribution of 33.97% among the others and also the second place on Vintage cars of product line has the distribution of 21.55%.

6Q. How does product pricing impact sales volume?

Count of Totalsales and Sum of priceEach

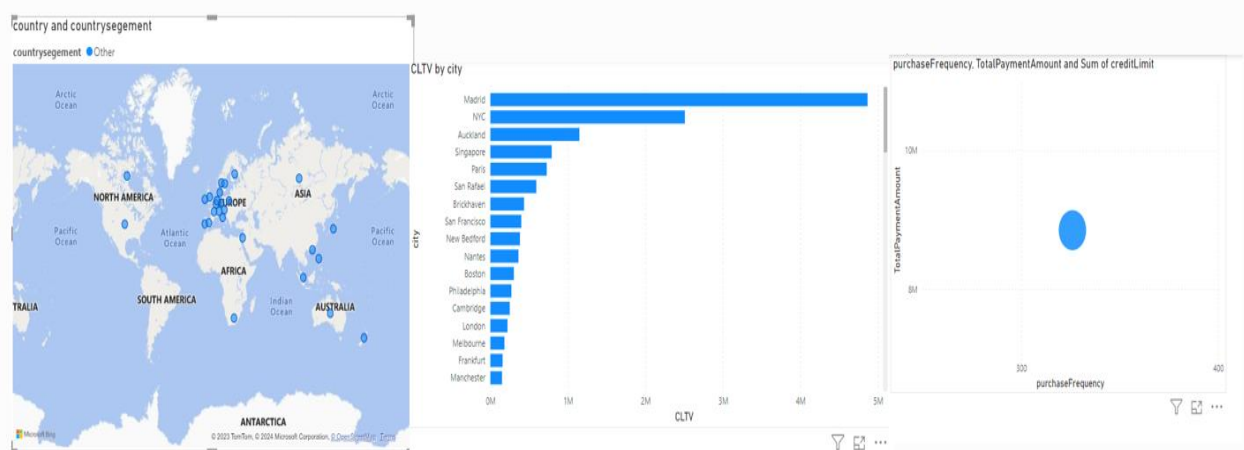
● Count of Totalsales ● Sum of priceEach



Explanation: This above scatter plot shows that the product pricing impact the sales volume if the pricing is less then the sales can be high

# Customer Analysis

## Customer Analysis

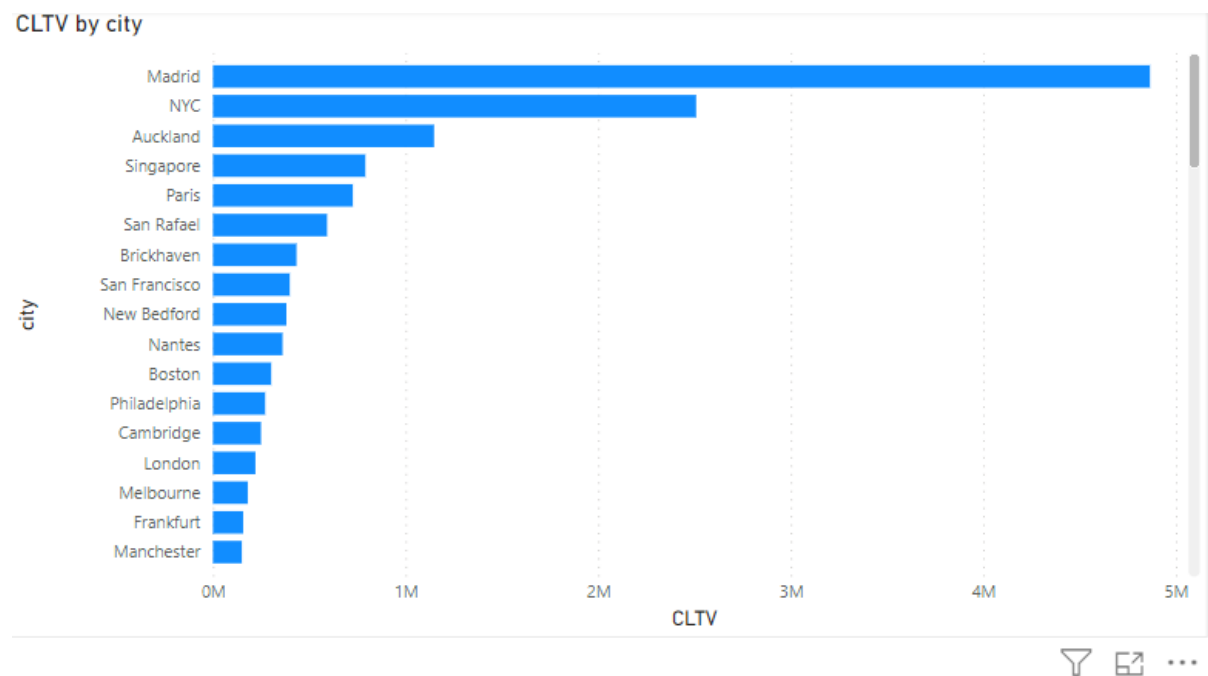


7Q. What is the distribution of customers across different demographic segments? (Visual: Customer segmentation by demographics)



Explanation: In this map chart , this shows about Here the distribution of customers can be among different countries like Asia, Australia etc.. as shown in the chart

8Q. How does customer lifetime value vary for different customer acquisition channels?

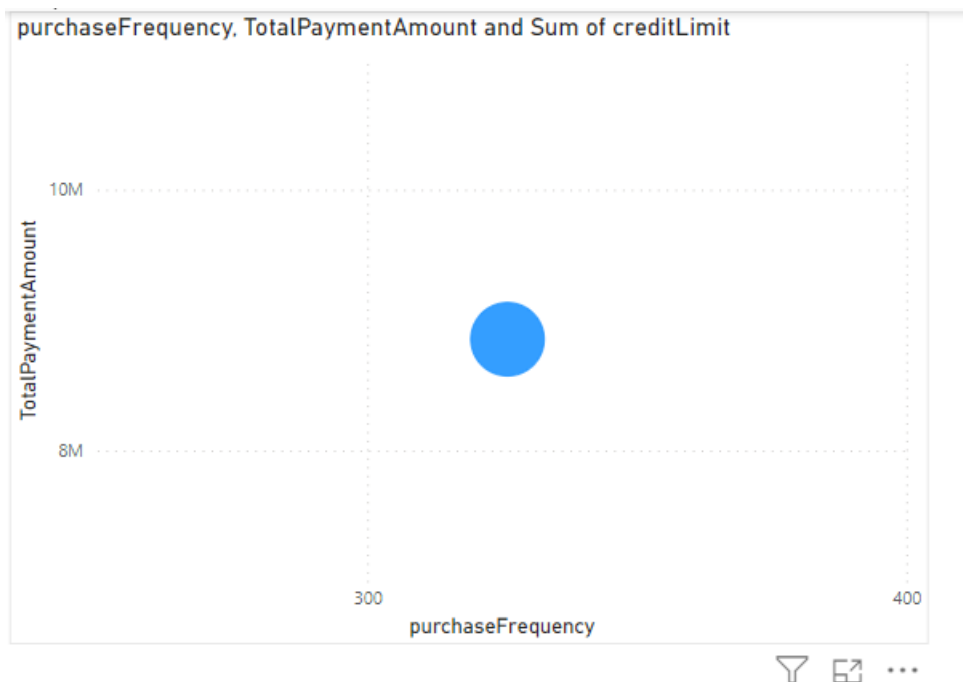


Explanation :

In this graph ,the CLTV can be definitely vary from different acquisition channels, but the channels not in data , but the CLTV has more in Madrid city



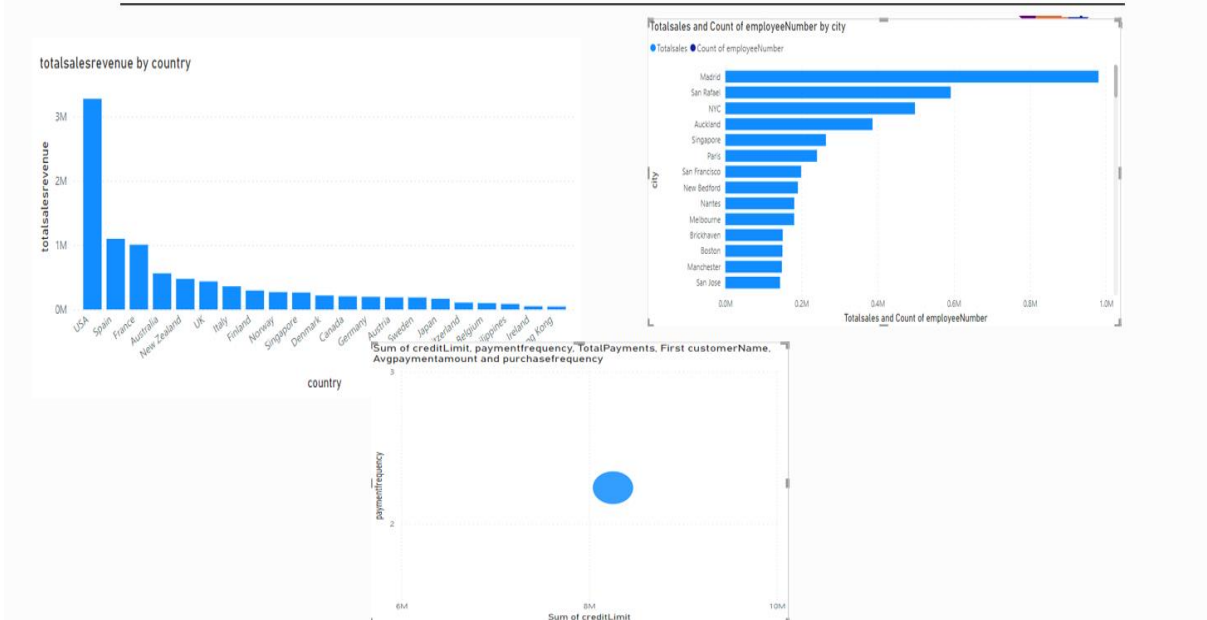
9Q. What is the correlation between customer age and purchase frequency?



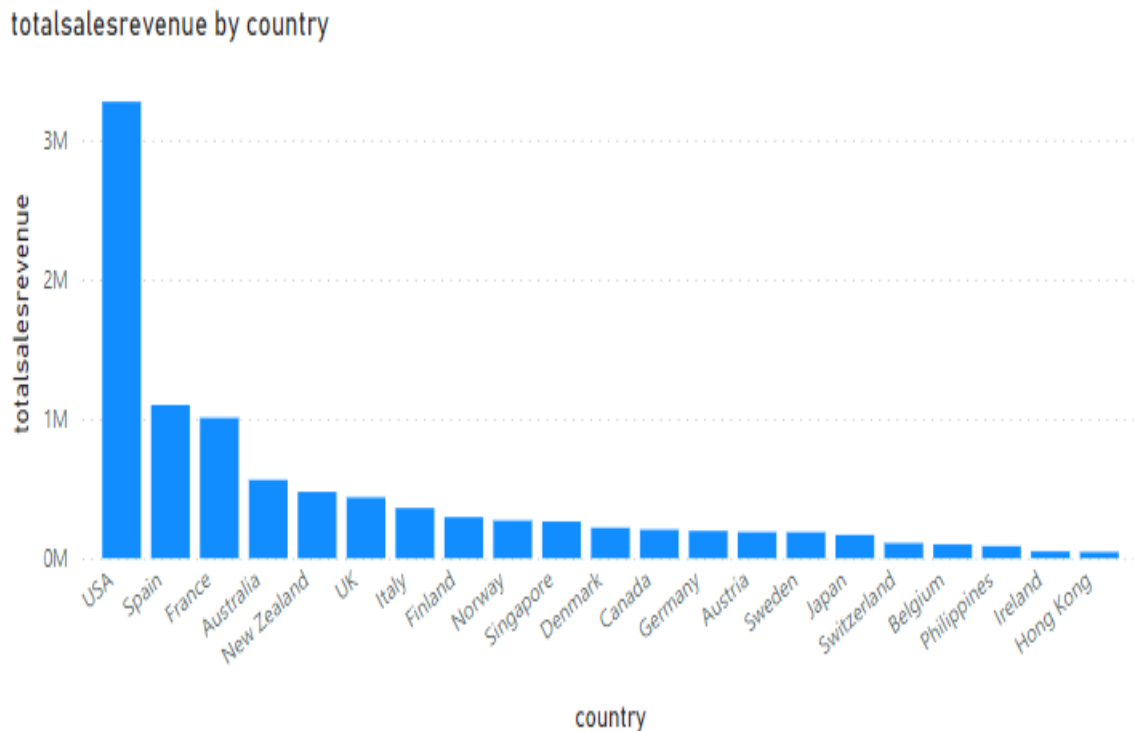
Explanation: Here the customer age can be factor for purchase frequency, but there is no apparent linear relation between them

# Demographic Analysis

## Demographic Analysis

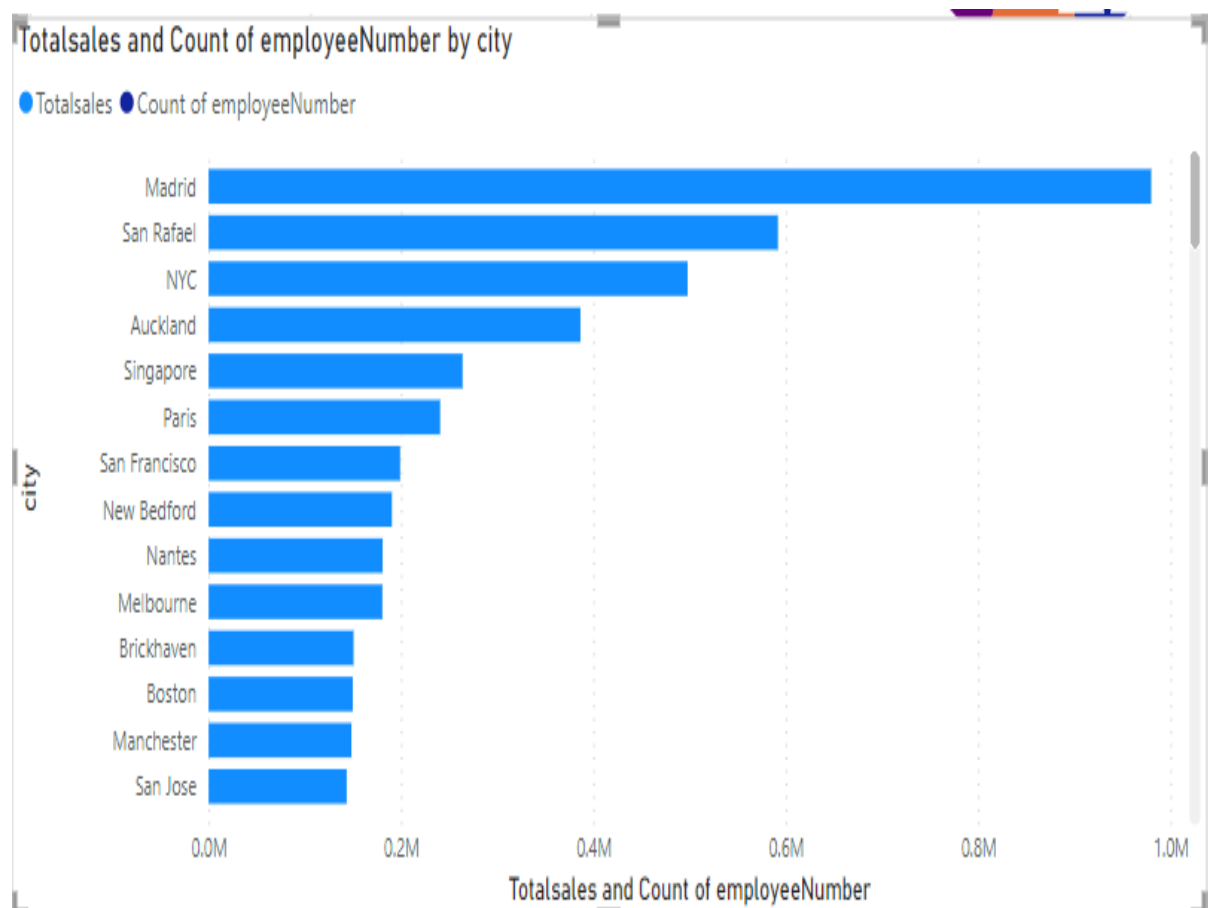


10Q. What are the top regions in terms of sales revenue? (Visual: Sales revenue by region)



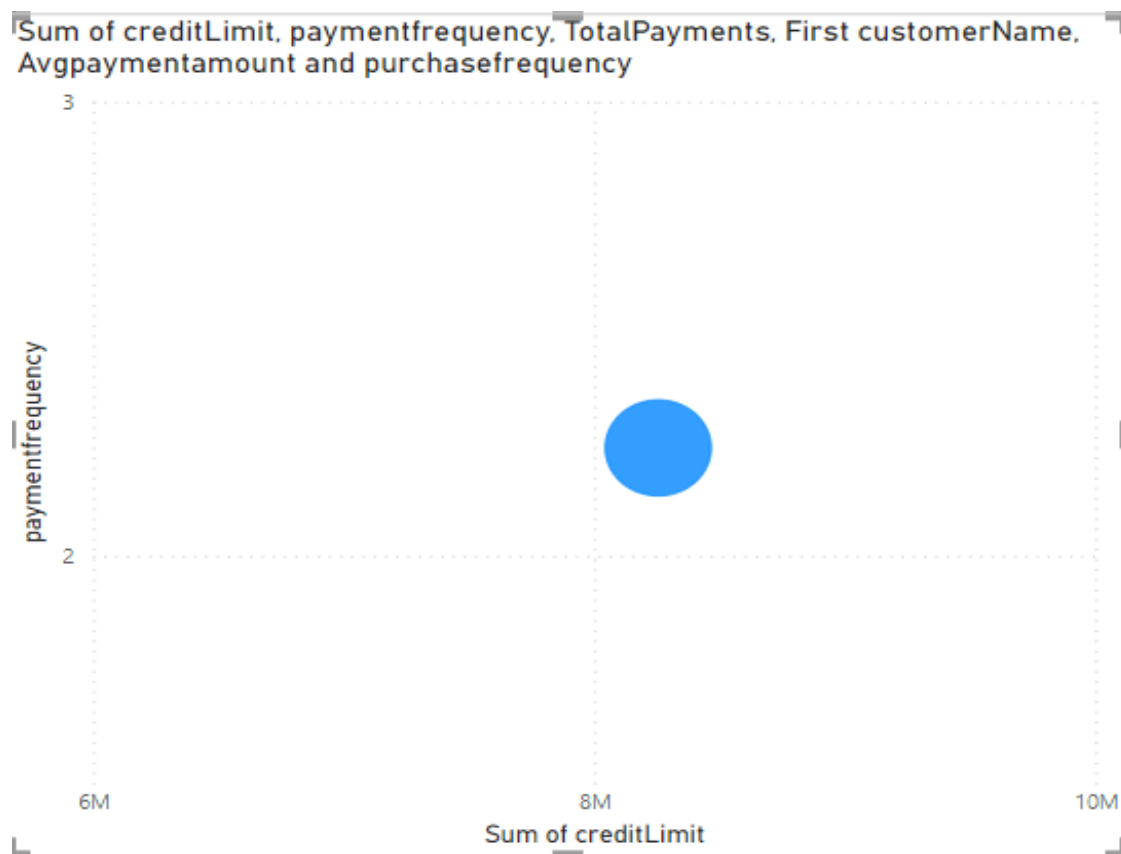
Explanation : this above graph is showing USA has in the top position in terms of sales revenue and the next is Spain and next is France s shown in diagram

11Q. How does the performance of sales employees vary across different regions? (Visual: Employee performance by region)



Explanation : In this above graph , that shows the sales may vary across the regions and , Madrid City is having the high sales in here in the data take and in the chart.

12Q. What is the correlation between customer demographics and purchase frequency?



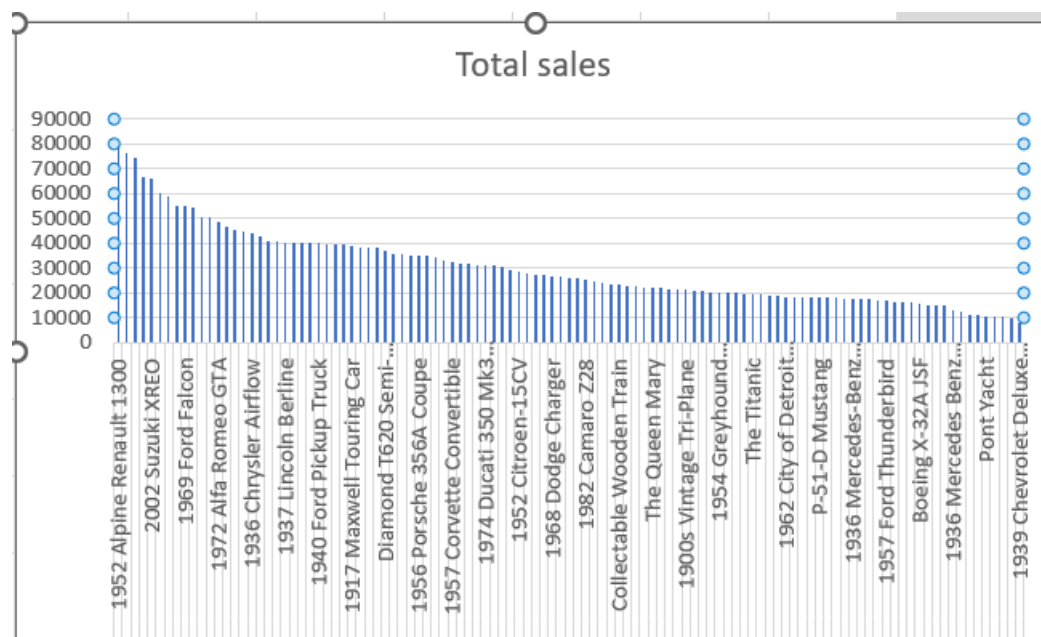
Explanation : the correlation as shown in diagram that it is placed in middle, there is a relation because based on demographics the purchase frequency can be there

**EDA**

**PROBLEM**

**STATEMENTS**

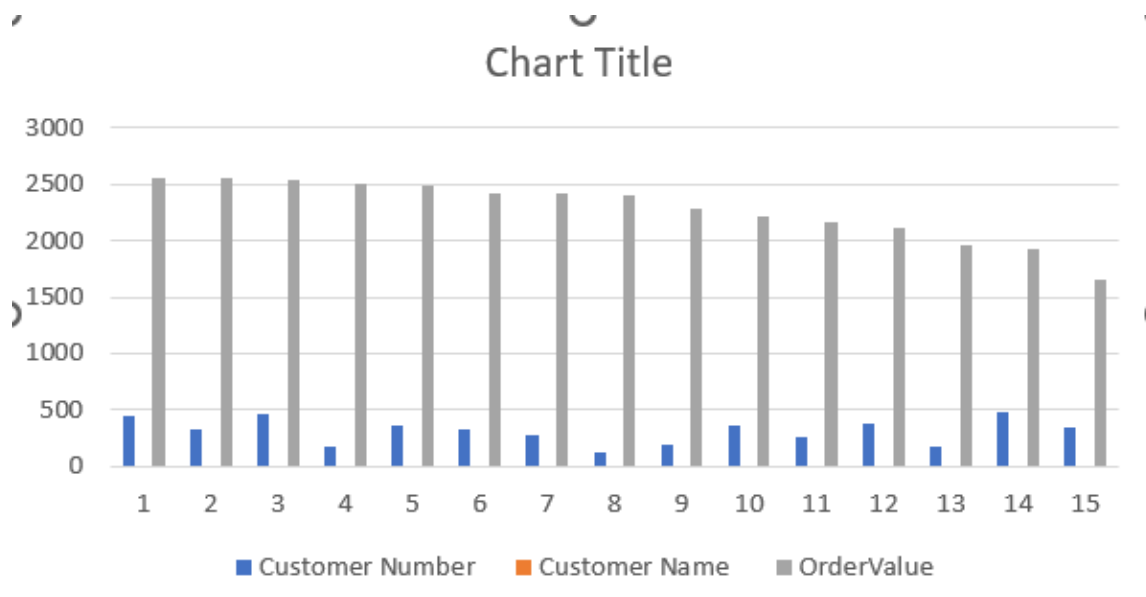
Q1. Which factors contribute to the highest sales in a particular region?



```
select c.country, p.productName, sum(od.quantityOrdered * od.priceEach) as
    Totalsales
    from customers as c
    join orders as o on c.customerNumber = o.customerNumber
    join orderdetails as od on o.orderNumber = od.rderNumber
    join products as p on od.productCode = p.productCode
    where c.country = 'USA'
    group by c.country, p.productName
    order by Totalsales desc
```

Explanation : the 1952 Alpine Renault 1300 is the product having high sales

2Q. How can customer purchasing patterns be influenced to increase average order value?

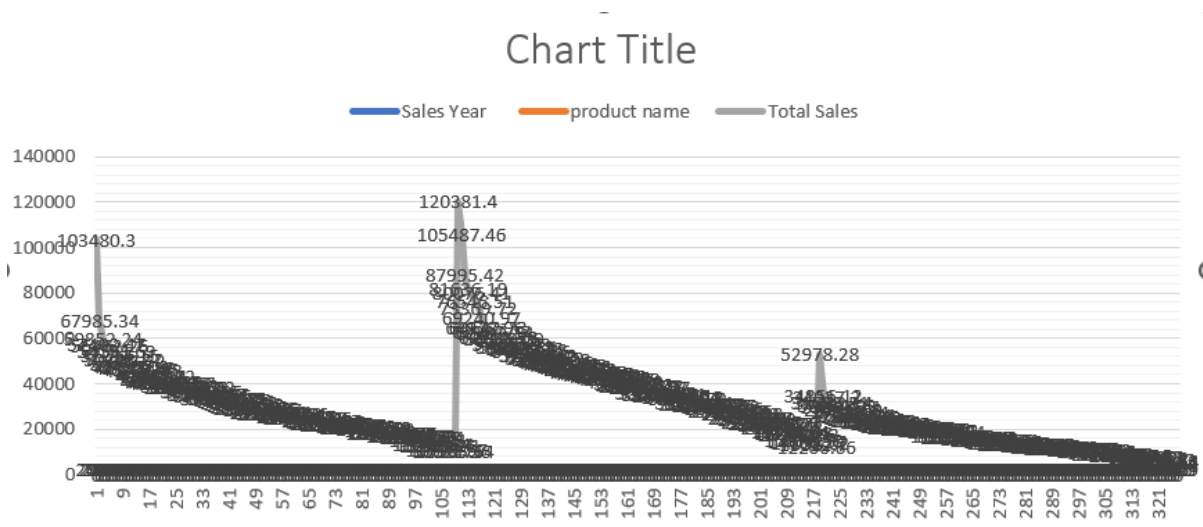


```
select c.customerNumber, c.customerName, avg(od.quantityOrdered * od.priceEach)
      as AverageOrderValue
      from customers as c
    join orders as o on c.customerNumber = o.customerNumber
    join orderdetails as od on o.orderNumber = od.orderNumber
    join products as p on od.productCode = p.productCode
    where p.productLine = 'motorcycles'
    group by c.customerNumber, c.customerName
    order by AverageOrderValue desc
```

Explanation : The customer purchasing patterns can be changed by order value as shown in the chart



3Q. what are the key drivers of sales growth, and how can they be leveraged for future success?

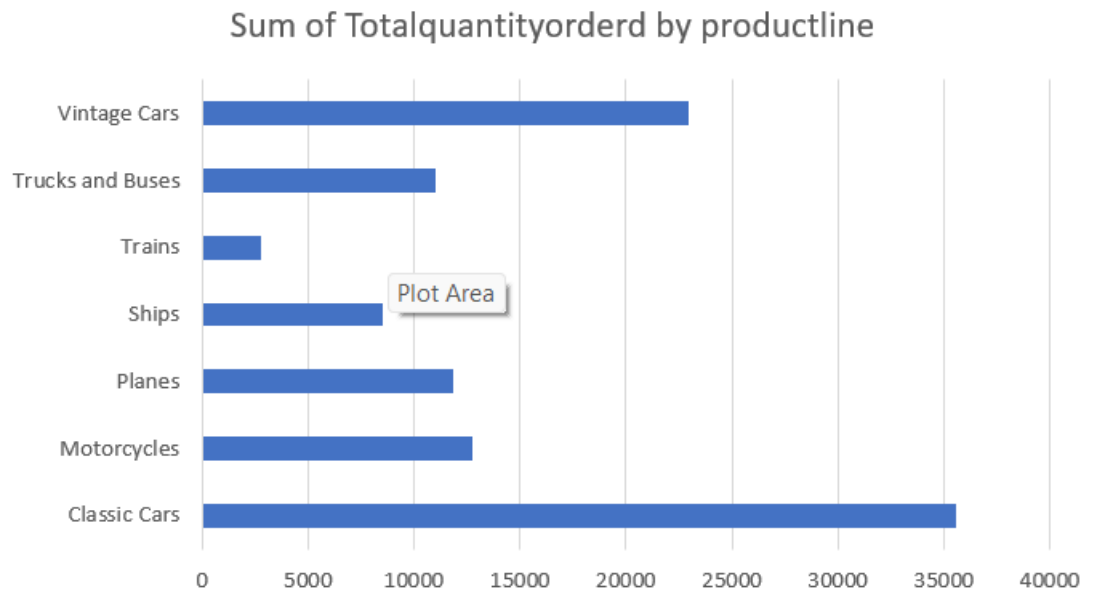


```
SELECT YEAR(o.orderDate) AS SalesYear,
       P.ProductName,
       sum(OD.quantityOrdered * OD.priceEach) AS TotalSales
FROM orders as o
JOIN orderdetails OD ON o.orderNumber = OD.orderNumber
JOIN Products P ON OD.ProductCode = P.ProductCode
GROUP BY SalesYear, P.ProductName
ORDER BY SalesYear, TotalSales DESC;
```

Sales Year	product name	Total Sales
2003	1992 Ferrari 360 Spider red	103480.3
2003	1952 Alpine Renault 1300	67985.34

Explanation: the products and related factors are important for sales growth

4Q. Which product features or attributes are most appealing to customers?



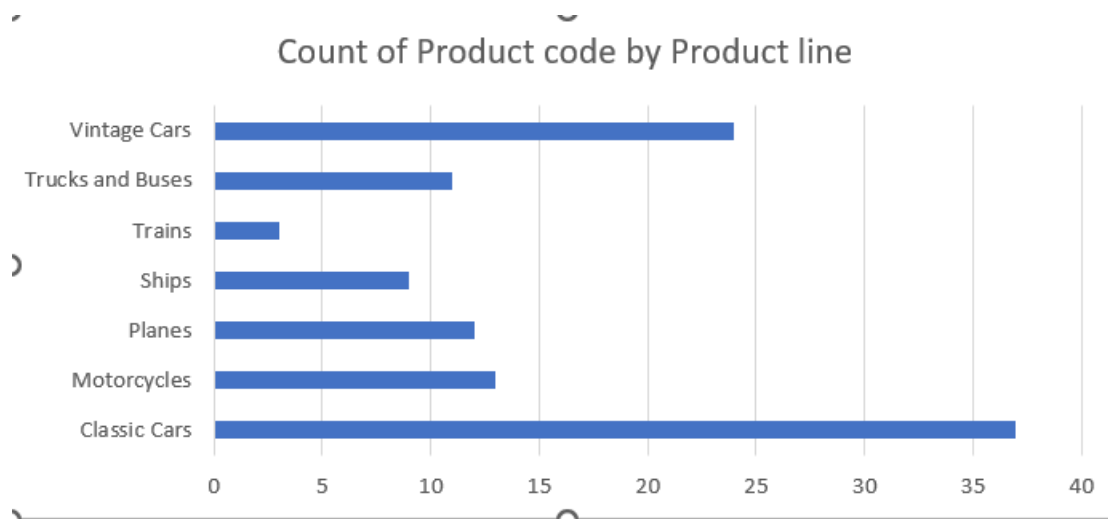
```
select p.productCode, p.productName, p.productLine, sum(od.quantityOrdered) as TotalQuantityorderd
from products as p
join productlines as pl on p.productLine = pl.productLine
join orderdetails as od on p.productCode = od.productCode

group by p.productCode , p.productname, pl.productLine

order by TotalQuantityorderd
```

Explanation: the Vintage cars product line has more quantity orders, which are more appealing to customers

## 5Q.How can the product mix be optimized to cater to changing market demands

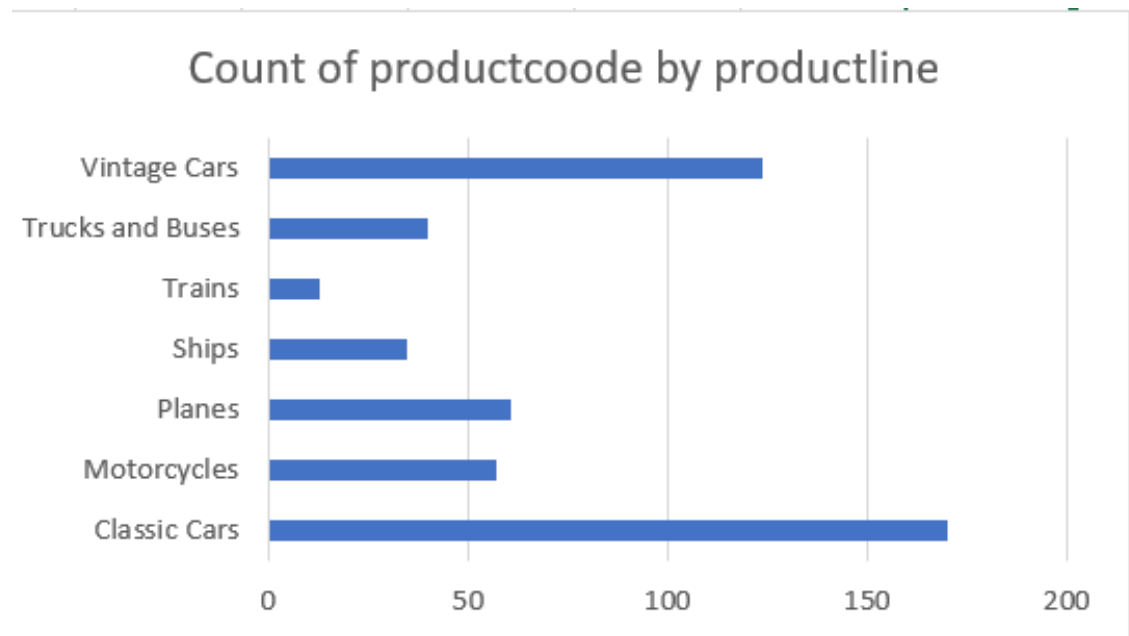


```
select p.productCode, p.productName, p.productLine, sum(od.quantityOrdered) as TotalQuantityorderd
from products as p
join productlines as pl on p.productLine = pl.productLine
join orderdetails as od on p.productCode = od.productCode

group by p.productCode , p.productname, pl.productLine
order by TotalQuantityorderd desc
```

Explanation : Here the Classic cars has the more count of product code , this changes the market demands

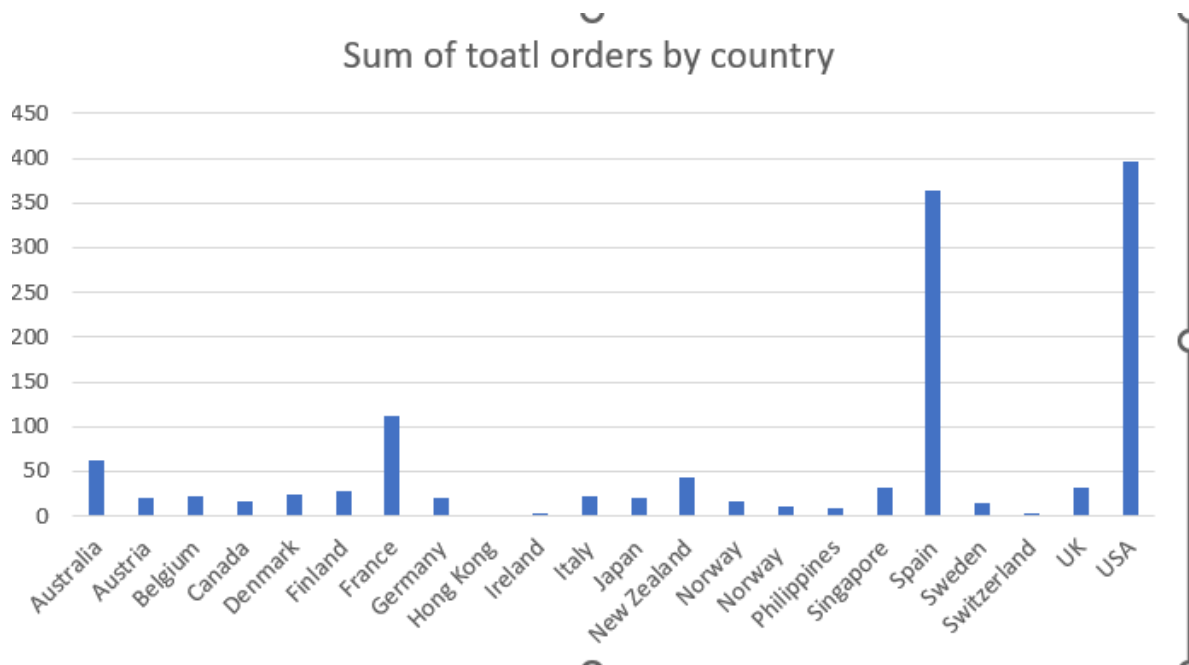
6Q. Are there any specific market segments where a particular product is underperforming, and how can it be improved?



```
select p.productCode, p.productName,  
pl.productLine,o.customerNumber,c.customerName,sum(od.quantityOrdered) as  
TotalQuantityorderd  
from products as p  
join productlines as pl on p.productLine = pl.productLine  
join orderdetails as od on p.productCode = od.productCode  
join orders as o on od.orderNumber= o.orderNumber  
join customers as c on o.customerNumber = c.customerNumber  
group by p.productCode , p.productname, pl.productLine, o.customerNumber, c.customerName  
order by TotalQuantityorderd asc
```

Explanation : The Trains is performing least performance in the market

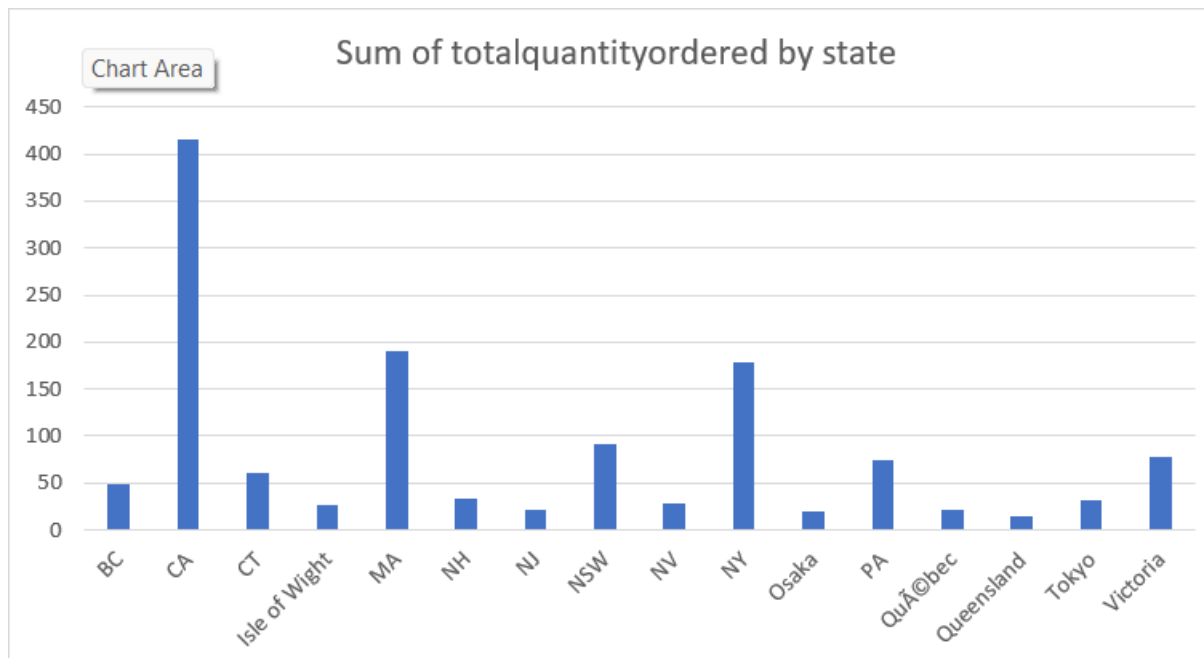
7Q. What are the main factors that influence customer loyalty and repeat purchases?



```
select c.customerNumber, c.customerName, c.city, c.state, c.country, count(o.orderNumber) as Totalorders
from customers as c
join orders as o on c.customerNumber = o.customerNumber
left join payments as p on c.customerNumber = p.customerNumber
group by c.customerNumber, c.customerName, c.city, c.state, c.country
order by Totalorders desc
```

Explanation : the product lines and important factors are mainly helpful for the loyalty and repeat purchases

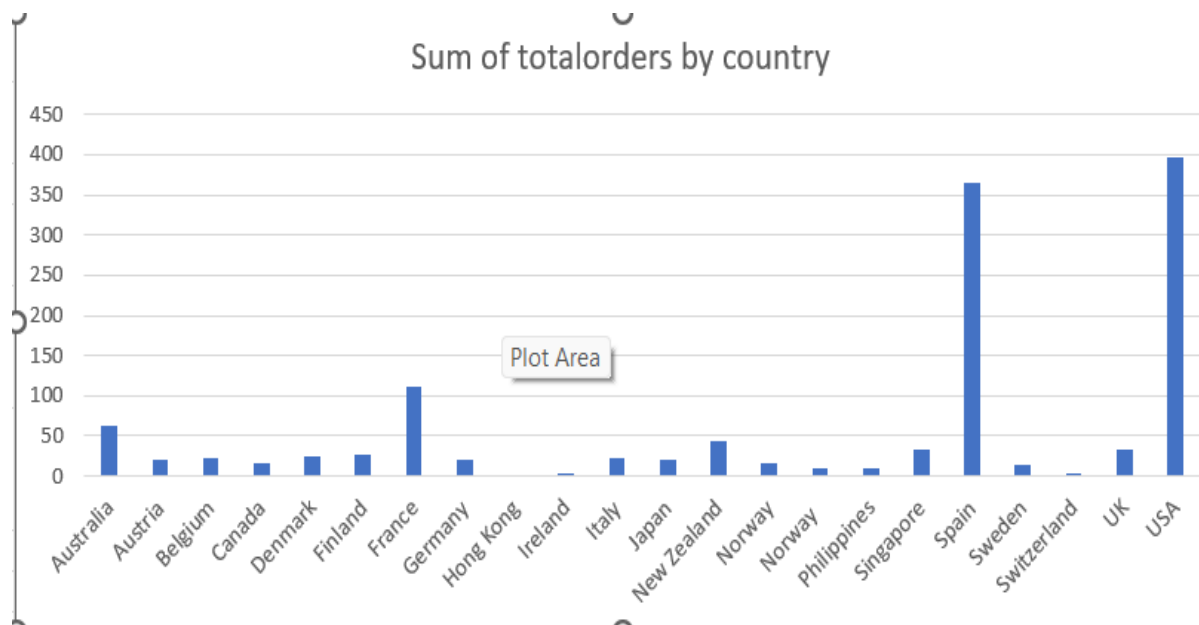
8Q. How do customer preferences differ based on geographic location, and how can marketing campaigns be customized accordingly?



```
select c.state, c.country, count(od.quantityOrdered) as totalquantityordered
from customers as c
join orders as o on c.customerNumber = o.customerNumber
join orderdetails as od on o.orderNumber = od.orderNumber
join products as p on od.productCode = p.productCode
where c.state is not null
group by c.state, c.country, p.productLine
order by c.state, c.country, totalquantityordered
```

Explanation : customer preferences can change, the CA state more quantity ordered

9Q. What are the characteristics of high-value customers, and how can similar customers be targeted for acquisition

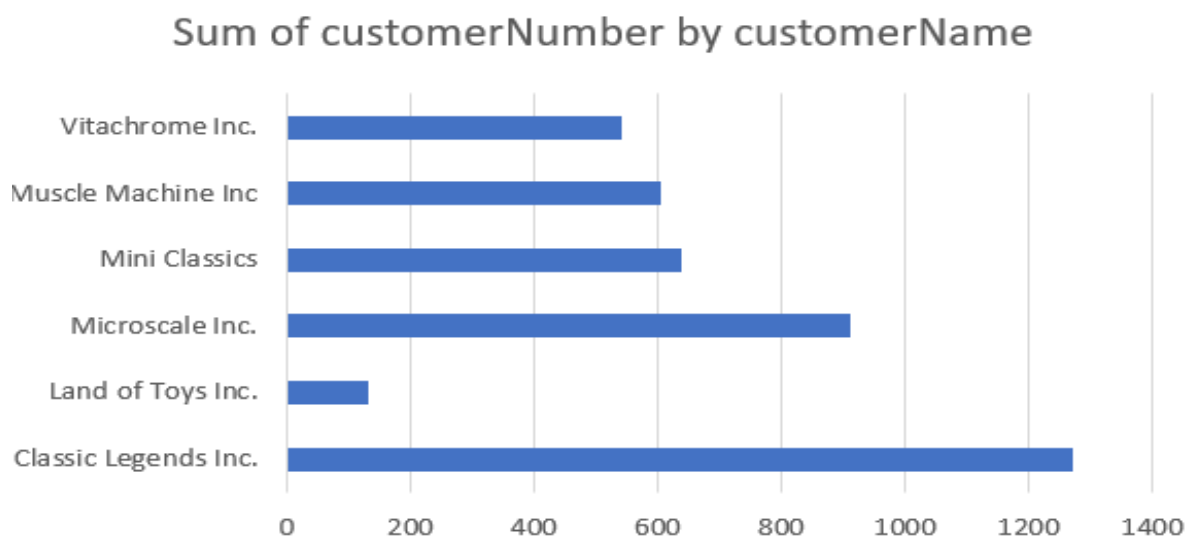


```
select c.customerNumber, c.customerName, c.city, c.state, c.country, count(o.orderNumber) as totalorderds, sum(p.amount) as Totalamountspent
from customers as c
join orders as o on c.customerNumber = o.customerNumber
left join payments as p on c.customerNumber = p.customerNumber

group by c.customerNumber, c.customerName, c.city, c.state, c.country
order by Totalamountspent
```

Explanation: total orders are high from USA and the customer has high value

10Q. How can marketing strategies be tailored to target specific demographic segments in different regions?

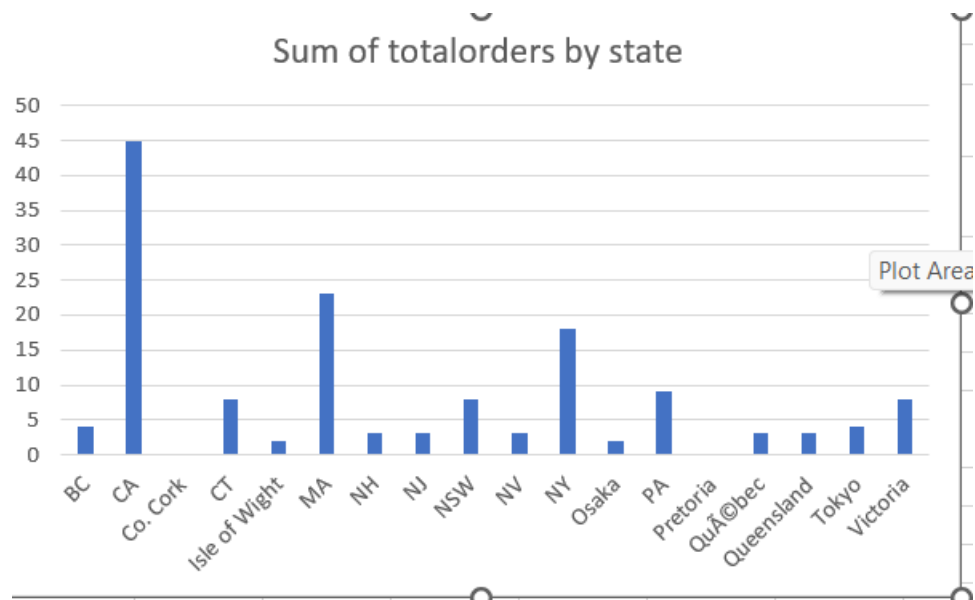


```
select c.customerNumber,c.customerName, c.city, c.state, c.country, o.orderNumber, o.orderDate, o.status
from customers as c
join orders as o on c.customerNumber = o.customerNumber
where c.state= 'NY'
```

Explanation : The marketing strategies can be tailored to target some segments in regions , where the classic legends has high



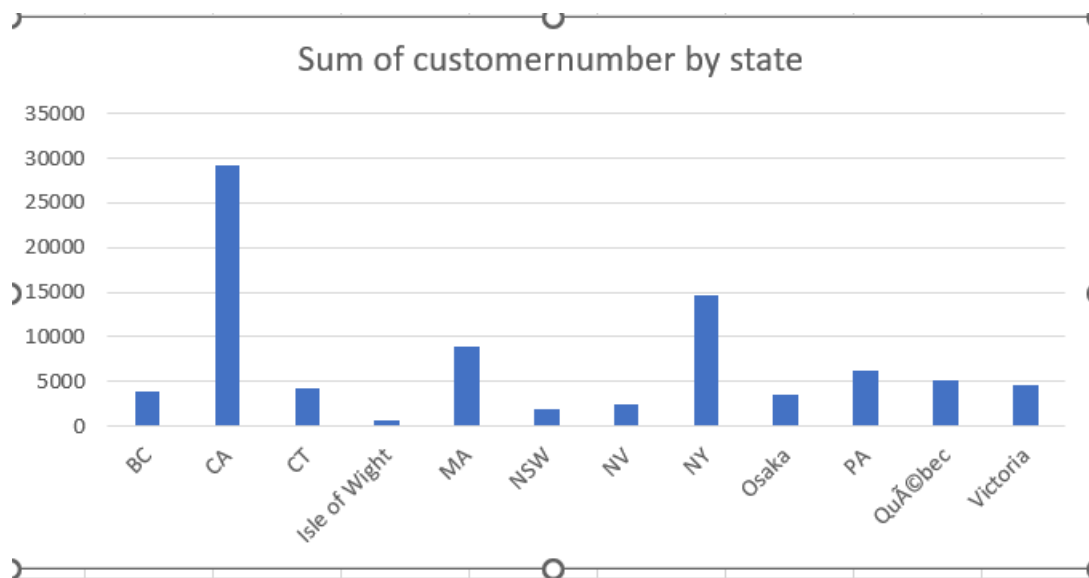
11Q.What are the potential untapped markets based on demographic indicators, and how can market penetration be increased?



```
select c.customerNumber, c.customerName, c.city, c.state, count(o.orderNumber) as totalorders
from customers as c
left join orders as o on c.customerNumber = o.customerNumber
WHERE c.state is not null
GROUP BY c.customerNumber, c.customerName, c.city, c.state
having totalorders = 0 or avg(o.status) = 'Inactive'
```

Explanation: This graph shows that the total orders has highest in the CA state

12Q. How do customer preferences and behavior differ based on demographic factors, and how can they be leveraged for personalized marketing campaigns?



```
select c.customerNumber, c.customerName, c.city, c.state,c.country, od.productcode, count(*) as purchasecount
from customers as c
join orders as o on c.customerNumber = o.customerNumber
join orderdetails as od on o.orderNumber = od.orderNumber
where state is not null
group by c.customerNumber, c.customerName, c.city, c.state,c.country, od.productcode
order by purchasecount desc
```

Explanation : the customer preferences can be differ from regions and the CA region has highest

# OVERALL DASHBOARD

