Data Analysis

Customer & Product Analysis

Presented by Alice Chen

Project Overview

This project leverages SQL and Power BI to analyse a retail company's sales data, aiming to uncover insights into customer behaviours and product performance. Through meticulous data analysis, trends, patterns, and opportunities to drive strategic decisions are explored.

Project Objective

The goal of this project is to support stakeholders in answering critical business questions related to sales performance, customer engagement, and product popularity, providing insights for decision-making.

- Question 1: Which product line should the company prioritise for restocking?
- Question 2: How should the company tailor marketing and communication strategies to customer behaviours?
- Question 3: How much can the company spend on acquiring new customers?

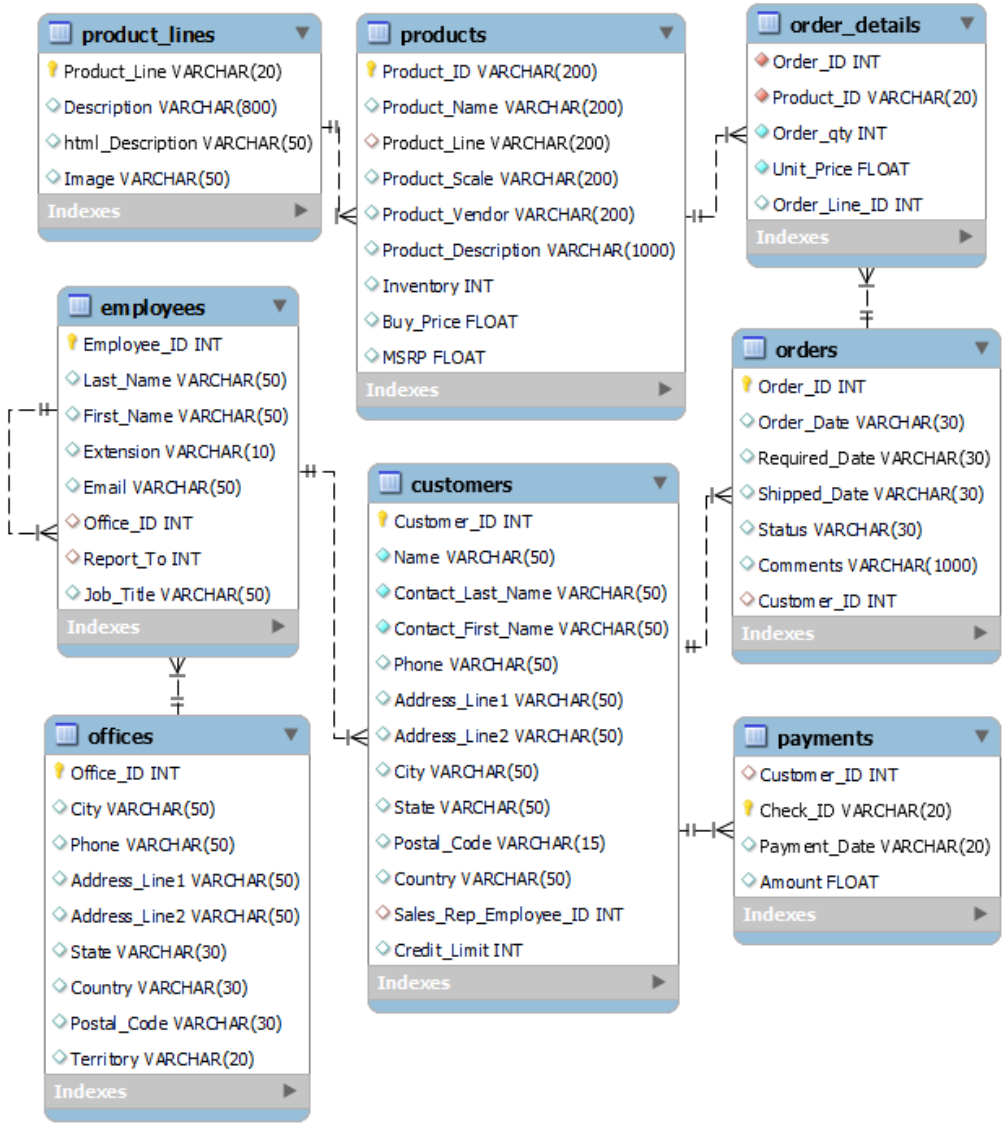
Data Source

- The dataset used for this analysis is sourced from Dataquest's Guided Project: "Customers and Products Analysis Using SQL".
- It is publicly available and can be accessed directly through Dataquest's platform.
- This dataset has been instrumental in providing a comprehensive view of the retail company's sales data, allowing for a detailed exploration of customer behaviours and product performance.

Tools Used

- MySQL Workbench - For data querying, manipulation, and analysis
- Power BI - For visualising data and uncovering insights

Table Overview



Table_name	Number_of_Attributes	Number_of_Rows
Customers	13	122
Products	9	110
Product_Lines	4	7
Orders	7	326
Order_Details	5	2996
Payments	4	273
Employees	8	23
Offices	9	7

Analysis & Results

Question 1:

Which product line should the company priority for restocking?

SQL Query

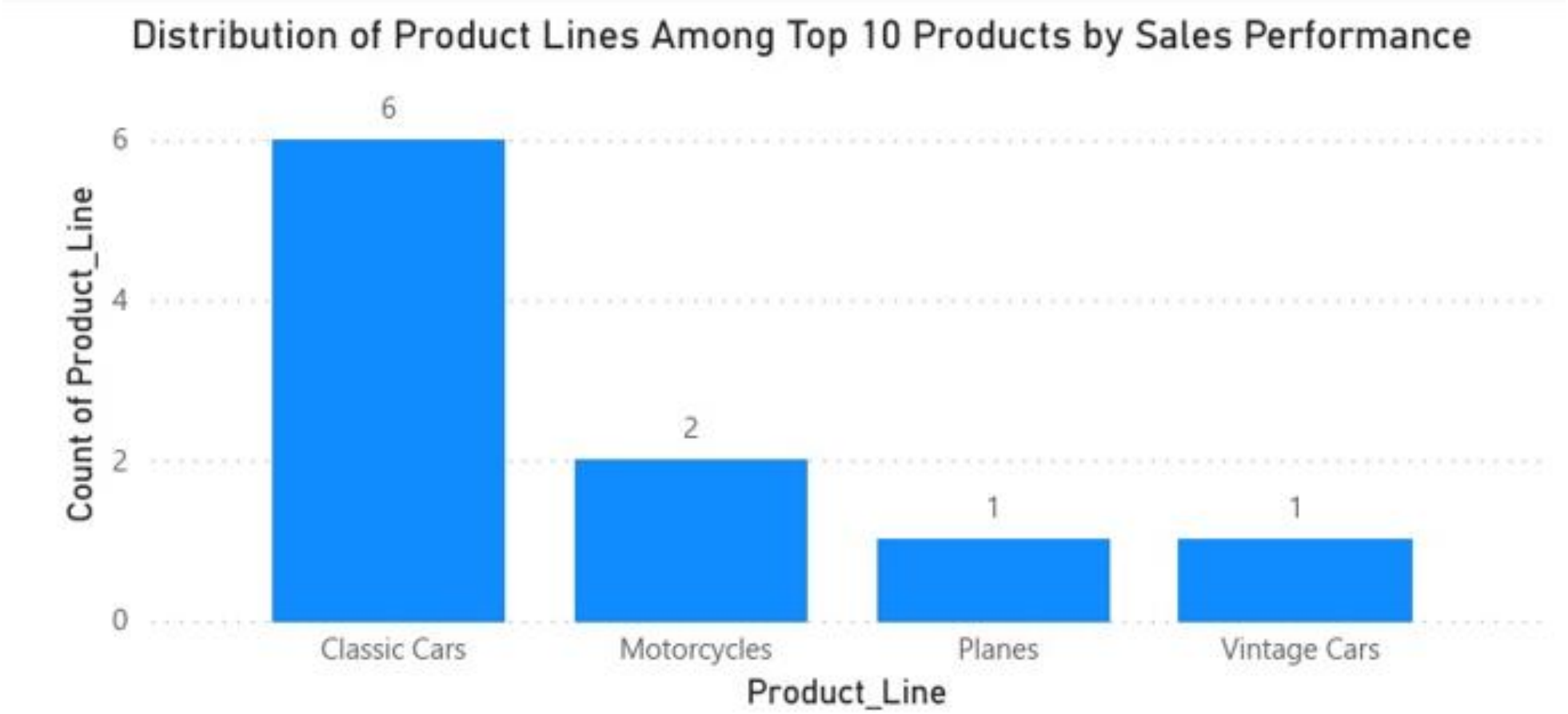
```
-- Create CTE for Low_Stock and Product_Performance tables --
WITH
Low_Stock AS (
SELECT p.Product_ID, p.Product_Line, p.Product_Name,
            ROUND(SUM(o.Order_qty/p.Inventory),2) AS Low_Stock
  FROM Products p
 INNER JOIN Order_Details o
      ON p.Product_ID = o.Product_ID
GROUP BY p.Product_ID
ORDER BY Low_Stock
),
Product_Performance AS (
SELECT p.Product_ID, p.Product_Line, p.Product_Name,
            ROUND(SUM(o.Order_qty * o.Unit_Price),2) AS Product_Performance
   FROM Order_Details o
  INNER JOIN Products p
       ON p.Product_ID = o.Product_ID
GROUP BY p.Product_ID
ORDER BY Product_Performance DESC
)
```

```
-- Show the table of Top 10 highest performances products
and the inventory level --
SELECT p.Product_Name, p.Product_Line,
            p.Product_Performance, l.Low_Stock
  FROM Product_Performance p
 INNER JOIN Low_Stock l
      ON p.Product_ID = l.Product_ID
WHERE Product_ID IN (SELECT Product_ID
                              FROM Low_Stock)
 GROUP BY Product_ID
 ORDER BY Product_Performance DESC
 LIMIT 10;
```

Product_Name	Product_Line	Product_Performance	Low_Stock
1992 Ferrari 360 Spider red	Classic Cars	276839.98	0.22
2001 Ferrari Enzo	Classic Cars	190755.86	0.28
1952 Alpine Renault 1300	Classic Cars	190017.96	0.13
2003 Harley-Davidson Eagle Drag Bike	Motorcycles	170686	0.18
1968 Ford Mustang	Classic Cars	161531.48	13.72
1969 Ford Falcon	Classic Cars	152543.02	0.92
1980s Black Hawk Helicopter	Planes	144959.91	0.20
1998 Chrysler Plymouth Prowler	Classic Cars	142530.63	0.21
1917 Grand Touring Sedan	Vintage Cars	140535.6	0.34
2002 Suzuki XREO	Motorcycles	135767.03	0.10

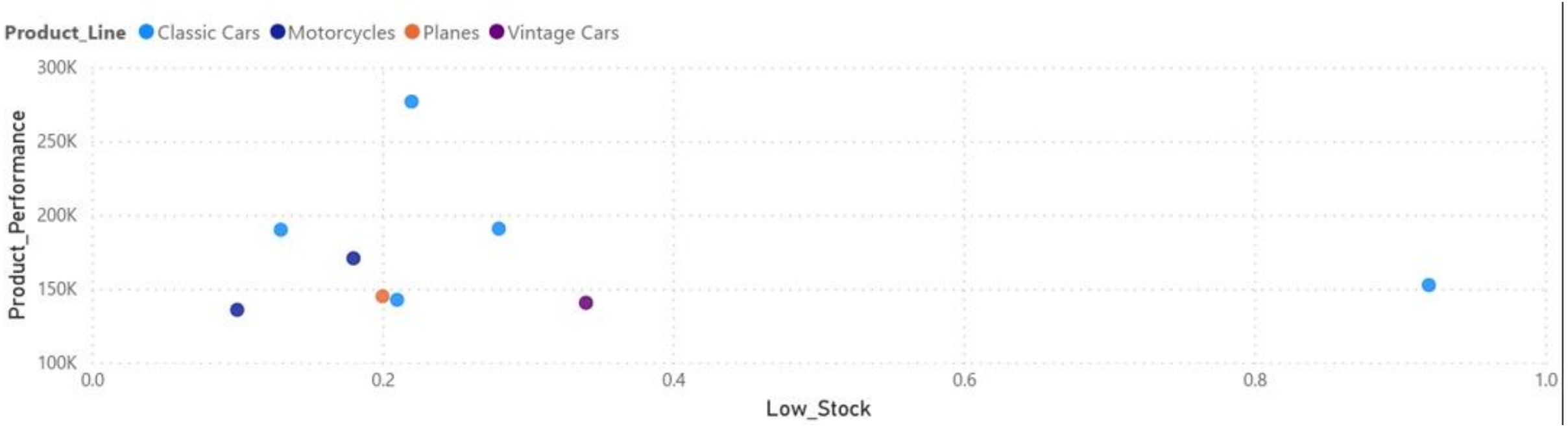
Data Visualisation

Within the Top high-performing products, with “Classic Cars” showing a dominant market need.



Data Visualisation

In most cases, these high-performance products maintain a stock level that is not yet critical but warrants close monitoring to prevent stockouts.



Insights

- These findings imply that specific product lines, particularly “Classic Cars”, have a strong market demand.
- Most of them are not in urgent of restocking, except for the 1968 Ford Mustang.
- The presence of motorcycles, vintage cars, and planes suggests varied consumer interests, catering to both practical uses and speciality hobbies.

Analysis & Results

Question 2:

How should the company tailor marketing and communication strategies to customer behaviors?

SQL Query

```
-- Create CTE FOR Top 5 VIP customers --
WITH
Profit_Table AS (
SELECT o.Customer_ID, od.Order_Qty, od.Unit_Price, p.Buy_Price,
            c.Contact_First_Name  AS  First_Name,
            c.Contact_Last_Name AS Last_Name, c.Country, c.City
  FROM Orders o
 INNER JOIN Order_Details od
       ON o.Order_ID = od.Order_ID
  INNER JOIN Products p
        ON od.Product_ID = p.Product_ID
   INNER JOIN Customers c
         ON o.Customer_ID = c.Customer_ID
)
-- Show the table --
SELECT Customer_ID, First_Name, Last_Name,
            ROUND(SUM(Order_Qty * (Unit_Price-Buy_Price)),3) AS Profit,
            Country,  City
  FROM Profit_Table
 GROUP BY Customer_ID
 ORDER BY Profit DESC
 LIMIT 5;
```

Customer_ID	First_Name	Last_Name	Profit	Country	City
141	Diego	Freyre	326519.661	Spain	Madrid
124	Susan	Nelson	236769.391	USA	San Rafael
151	Jeff	Young	72370.089	USA	NYC
114	Peter	Ferguson	70311.069	Australia	Melbourne
119	Janine	Labrune	60875.3	France	Nantes

SQL Query

```
-- Create CTE FOR Top 5 least engaging customers --
WITH
Profit_Table AS (
SELECT o.Customer_ID, od.Order_Qty, od.Unit_Price, p.Buy_Price,
            c.Contact_First_Name AS First_Name,
            c.Contact_Last_Name AS Last_Name, c.Country, c.City
 FROM Orders o
 INNER JOIN Order_Details od
      ON o.Order_ID = od.Order_ID
 INNER JOIN Products p
      ON od.Product_ID = p.Product_ID
  INNER JOIN Customers c
       ON o.Customer_ID = c.Customer_ID
)
-- Show the table --
SELECT Customer_ID, First_Name, Last_Name,
             ROUND(SUM(Order_Qty * (Unit_Price-Buy_Price)),3) AS Profit,
             Country, City
  FROM Profit_Table
 GROUP BY Customer_ID
 ORDER BY Profit
 LIMIT 5;
```

Customer_ID	First_Name	Last_Name	Profit	Country	City
219	Mary	Young	2610.87	USA	Glendale
198	Leslie	Taylor	6586.02	USA	Brickhaven
473	Franco	Ricotti	9532.93	Italy	Milan
103	Carine	Schmitt	10063.80	France	Nantes
489	Thomas	Smith	10868.04	UK	London

Insights

- The contrast between the profit contributions of VIPs and least engaging customers suggests that a small segment of customers is driving a large portion of the company's revenue.

- There is no specific country or city that uniquely characterises either VIP or least engaging customers. High and low engagement levels are found across diverse geographic locations, from the USA and Australia to France and the UK, suggesting that customer engagement is influenced by factors other than geographic region.

Analysis & Results

Question 3:

How much can the company spend on acquiring new customers?

SQL Query

```
-- Create CTE --
WITH
Payment_With_Month_Table AS (
SELECT *, CAST(SUBSTRING(Payment_Date,7,10) AS SIGNED) * 100 +
          CAST(SUBSTRING(Payment_Date,4,5) AS SIGNED) AS `year_month`
  FROM Payments
),
Customer_Num_Table AS (
SELECT `year_month` , COUNT(*) AS Customer_Number,
          SUM(Amount) AS Total
  FROM Payment_With_Month_Table p1
 GROUP BY `year_month`
),
New_Customer_Num_Table AS (
SELECT p1.`year_month` , COUNT(*) AS New_Customer_Num,
          SUM(p1.Amount) AS New_Customer_Total,
          (SELECT Customer_Number
             FROM Customer_Num_Table c
            WHERE p1.`year_month` = c.`year_month`) AS Customer_Number,
           (SELECT Total
              FROM Customer_Num_Table c
             WHERE p1.`year_month` = c.`year_month`) AS Customer_Total
  FROM Payment_With_Month_Table p1
WHERE p1.Customer_ID NOT IN (SELECT Customer_ID
                               FROM Payment_With_Month_Table p2
                              WHERE p1.`year_month` > p2.`year_month`)
GROUP BY P1.`year_month`
)
```

```
-- Show the table --
SELECT `year_month` ,ROUND(New_Customer_Num *100 /
          Customer_Number,1) AS New_Customer_Props,
          ROUND(New_Customer_Total *100/ Customer_Total,1) AS
          New_Customer_Total_Props
  FROM New_Customer_Num_Table
 ORDER BY `year_month` ;
```

Year_Month	New_Customer_Props	New_Customer_Total_Props
200301	100	100
200302	100	100
200303	100	100
200304	100	100
200305	100	100
200306	100	100
200307	75	68.3
200308	66.7	54.2
200309	80	95.9
200310	69.2	69.3
200311	57.9	53.9
200312	60	54.9
200401	33.3	41.1
200402	33.3	26.5
200403	54.5	55
200404	40	40.3
200405	12.5	17.3
200406	33.3	43.9
200407	10	6.5
200408	18.2	26.2
200409	40	56.4

SQL Query

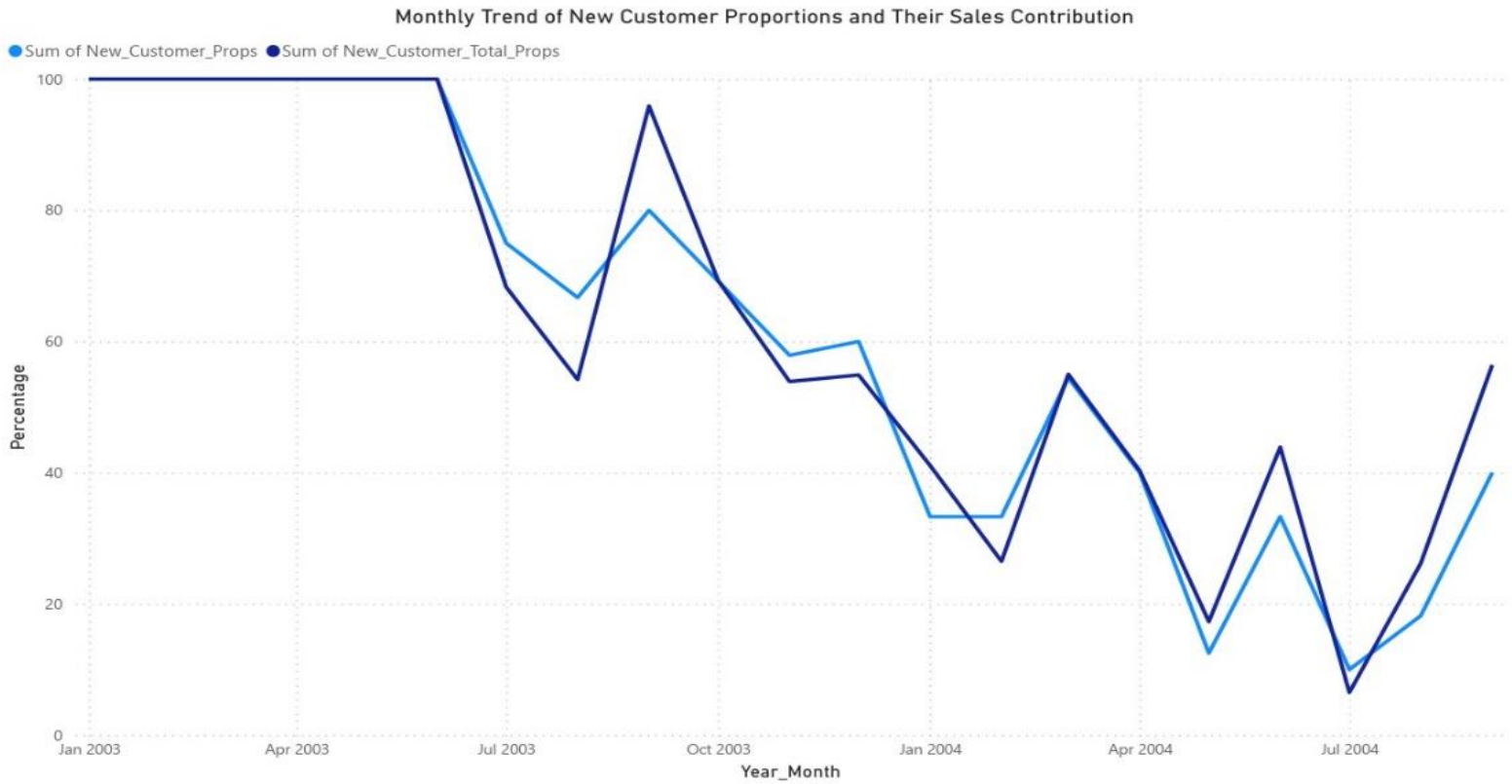
```
WITH
Money_In_By_Customer_Table AS (
SELECT o.Customer_ID, SUM(od.Order_qty * (od.Unit_Price - p.Buy_Price)) AS Revenue
  FROM Orders o
 INNER JOIN Order_Details od
       ON o.Order_ID = od.Order_ID
 INNER JOIN Products p
       ON od.Product_ID = p.Product_ID
GROUP BY o.Customer_ID
)
SELECT ROUND(AVG(Revenue),2) AS LTV
   FROM Money_In_By_Customer_Table;
```

LTV_Value
39039.59

Data Visualisation

The initial 100% new customer proportions indicate the early stages of the company, where all customers were newcomers.

There is a noticeable decline in new customer proportions and their contribution to sales as the company moves into 2004, with the lowest new customer proportion in July 2004.



Insights

- The high LTV suggests that customers, once acquired, have the potential to contribute significant revenue over time.

- The decreasing trend in new customer proportions and their sales contribution over time could suggest market saturation, increased competition, or challenges in attracting new customers.

- The line graph underlines critical periods where customer acquisition strategies may need reassessment. For instance, the sharp decrease after the initial surge points to the necessity of enhancing marketing efforts and customer retention strategies.

Conclusion

1. Restocking Priority: The data analysis identified key product lines that the company should prioritise for restocking, with a significant focus on the classic cars and motorcycles categories due to their high demand and low inventory levels.

2. Marketing and Communication Strategies: The assessment of customer engagement levels has highlighted the importance of personalised marketing strategies, particularly for the top-tier customers who contribute the most to the company's profit, as well as the necessity to re-engage less active customers.

3. Customer Acquisition Cost: By establishing the average Customer Lifetime Value (LTV), the company can now make informed decisions about the investment in acquiring new customers, ensuring that the cost of acquisition is aligned with the long-term revenue potential.