Enhancing Sales Performance with Data-Driven Insights

Background

A multinational retail company that specialises in consumer goods is looking to enhance its sales performance by leveraging data analytics. The company operates in multiple countries, offering a diverse range of products across different categories. With thousands of daily transactions, customer interactions, and product shipments, the company generates vast amounts of sales data. However, they lack a clear understanding of their key revenue drivers, customer purchasing behaviour, and product demand trends.

By analysing historical sales data, the company aims to improve business decision-making, optimise inventory management, enhance customer engagement, and identify opportunities for revenue growth.

Business Challenges

The company faces several challenges that need to be addressed using data analytics:

1. **Lack of Sales Visibility Across Product Lines:**
    - The company offers multiple product lines, but it is unclear which categories generate the most revenue.
    - Understanding sales performance at the product level will help with strategic product planning.
2. **Identifying High-Value Customers and Retention Strategies:**
    - The business needs to determine which customers contribute the most to revenue.
    - A lack of personalised engagement strategies might result in losing high-value customers.
3. **Understanding Seasonal and Regional Sales Trends:**
    - Sales may vary based on time of the year, location, or market conditions.

- Without clear insights into these patterns, the company struggles with demand forecasting and inventory management.

4. **Optimising Pricing and Discount Strategies:**
   - The gap between the **Manufacturer's Suggested Retail Price (MSRP)** and the **actual sale price** affects profitability.
   - Identifying products frequently sold at a discount can guide future pricing strategies.
5. **Impact of Order Size and Deal Size on Revenue:**
   - Understanding how order quantity varies with different deal sizes (Small, Medium, Large) can help optimise pricing for bulk purchases.
   - Analysing the **relationship between order size and revenue** will help fine-tune sales strategies.

Objectives and Business Goals

The primary goal of this study is to leverage **PySpark and Data Visualisation** to gain actionable insights into sales performance. Specifically, the company aims to:

1. **Evaluate the Performance of Different Product Lines**
   - Identify which product lines contribute the most to total sales.
   - Discover whether certain categories are underperforming and need marketing attention.
2. **Identify and Target High-Spending Customers**
   - Pinpoint the top 10 customers based on total spending.
   - Develop targeted loyalty programmes or promotions to increase customer retention.
3. **Analyse Seasonal and Geographic Sales Trends**
   - Determine if certain months or locations have **higher sales activity**.
   - Use historical trends to improve inventory forecasting.
4. **Optimise Pricing and Discount Strategies**
   - Compare the **MSRP (standard price)** to **actual sale price**.
   - Identify which products are often sold at discounts and determine the impact on revenue.
5. **Understand the Relationship Between Order Size and Deal Size**
   - Analyse whether **bulk orders generate higher revenue**.
   - Optimise deal size-based pricing to increase **profit margins**.

Approach and Data Processing

To address these business challenges, we will use **PySpark** (for large-scale data processing) and **data visualisation tools (Matplotlib, Seaborn, Pandas)** to derive insights. The dataset contains the following key attributes:

- **Sales Information:** QUANTITYORDERED, PRICEEACH, SALES, ORDERDATE, DEALSIZE
- **Product Information:** PRODUCTLINE, PRODUCTCODE, MSRP
- **Customer Information:** CUSTOMERNAME, PHONE, ADDRESSLINE1, CITY, COUNTRY
- **Order Status:** STATUS (example: Shipped, Cancelled, On Hold)

Business questions:

1. What are the top 5 product lines by total sales?
2. How do sales vary by country?
3. What is the distribution of deal sizes?
4. Who are the top 10 customers by total purchase amount?
5. What is the relationship between quantity ordered and price each?