1. **Objective:**

The objective of this case study is to study the customer behaviour, sales treads and apply machine learning techniques to segment the customers into different groups and gain valuable insights.

1. **Methodology:**

* Data Preparation Steps and Assumptions:

Data Import and Cleaning: - Combined the data from two years (2009-2011). Removed cancelled transactions from StockCode column and records with missing or zero values for price and quantity.

Values from StockCode columns that contained DOT, M, and BankCharges (etc) without a customer ID were removed.

In total 23.11% of missing values were removed from the dataset.

* + - * + Data Quality Assessment:

Completeness: Ensured no missing values in crucial fields like Quantity, Price, InvoiceDate.

Validity: Removed transactions with negative prices or quantities.

Accuracy: Verified logical consistency, e.g., no sales with zero or negative prices.

Consistency: Standardized the categorical fields

* Data Transformation

Added derived columns like TotalPrice (Quantity \* Price), Month, and Year from InvoiceDate

Handled outliers in Price and Quantity through visual inspection but retained them (No negative values) for their potential significance in analysis.

* Analytical Techniques and Tools Used:

Tools: Analysis conducted using Python with libraries such as Pandas, NumPy, Matplotlib, and Seaborn.

Techniques: Data cleansing, EDA, integration, and RFM , K-means analysis for customer segmentation.

1. **Results:**

* Key Actionable Insights

Sales Overview:

* + Total sales over the period: £17,743,429.18
  + Total products sold: 10,706,167 units
  + Number of unique invoices: 36,975
  + Number of unique products: 5,283
  + With 64531 items sold, November 2011 had the highest number of sales.
  + Thursday has the highest sales volume(161664 units), followed by Tuesday, Wednesday

Top Products and Customers:

* Most sold product: WORLD WAR 2 GLIDERS ASSTD DESIGNS (109,169 units)
* Most frequent customer: Customer ID 17841 with 12,890 purchases.

Geographical Insights:

* Country with the highest sales: United Kingdom (8,705,313 units).
* Top 5 countries in terms of sales: UK, Netherlands, EIRE, France, Denmark.

Return item Analysis:

* Most returned product: REGENCY CAKESTAND 3 TIER (347 returns).
* Highest returns by country: United Kingdom, Germany, France.

1. **Data Visualizations:**

* Sales by country: - The United Kingdom leads with 8,705,313 units, followed by the EIRE, Netherlands, Germany, France, Australia, Spain, Switzerland, , Sweden and Denmark with descending sales figures

A graph with blue bars

Description automatically generated

* Top 10 Products: - The top-selling products include World War 2 Gliders Asstd Designs, White Hanging Heart T-Light Holder, Paper Craft Little Birdie, Assorted Colour Bird Ornament, Medium Ceramic Top Storage Jar, and Jumbo Bag Red Retrospot, with descending quantities.

A graph of blue and white bars

Description automatically generated

* Monthly Sales Trend: - Sales peaked in November 2011 with 64,531 units, followed closely by November 2010 with 60,312 units. There's a noticeable trend of higher sales towards the end of the year, with December also showing significant sales figures. Sales generally increased from 2009 to 2011, with occasional fluctuations.

A graph with blue lines and red dots

Description automatically generated

* Weekly Sales Trend: - Thursday leads with 161,664 units sold, followed by Tuesday and Wednesday with approximately 136,800 units each. Sunday and Monday both hover around 136,300 and 128,400 units, respectively.

A graph of blue bars with white text

Description automatically generated

* Customer Segmentation (RFM Analysis): - The segments range from Hibernating with 1,639 customers to Can't Lose with 64. Loyal Customers and Champions represent sizable segments, while Promising and New Customers have the fewest members. About to Sleep and Need Attention fall in between, indicating potential for re-engagement or retention efforts.

1. **Recommendations to the CEO**

* Customer Retention and Engagement:

Target At-Risk and Hibernating Customers: Implement re-engagement campaigns such as personalized offers and reminders to customers who have not made recent purchases.

Loyalty Programs for Top Segments: Develop exclusive loyalty programs for Champions and Loyal Customers to maintain their engagement and encourage higher spending.

* Inventory Management and Product Focus:

Focus on Best-Selling and High-Margin Products: Increase inventory for top-selling items and ensure adequate stock levels during peak seasons to maximize sales.

Analyze Return Reasons: Investigate why certain products, like the REGENCY CAKESTAND 3 TIER, have high return rates and address quality or description issues.

* Geographical Expansion and Marketing:

Expand Market Presence in High-Performing Regions: Strengthen marketing and distribution efforts in countries with high sales like the UK, Netherlands, and France

Tailored Regional Campaigns: Customize marketing campaigns based on regional preferences and buying behaviors observed in the data.

These insights and recommendations are designed to enhance customer satisfaction, streamline operations, and drive revenue growth.