**E-Commerce Data Analysis**

**Executive Summary**

The "E-Commerce Data Analysis" project leverages a rich dataset containing actual transactions from a UK-based online retailer. The data spans from December 2010 to December 2011 and includes critical variables such as invoice numbers, product descriptions, quantities, invoice dates, unit prices, customer IDs, and countries of origin. This dataset, maintained by the UCI Machine Learning Repository, offers a rare opportunity to explore detailed transactional data for a retailer specialising in unique, all-occasion gifts. The retailer serves a mix of retail and wholesale customers, with a significant focus on domestic markets.

The primary objective of this project is to extract actionable insights that inform and enhance the retailer’s operational and strategic decisions. By analysing customer behaviour, product performance, and sales trends, this project provides a comprehensive understanding of the factors driving revenue and growth. The findings are aimed at optimising pricing strategies, inventory management, and marketing campaigns while identifying opportunities for international expansion and customer retention.

Key areas of analysis include:

1. **Sales Trends**: Identifying seasonal patterns and sales peaks over daily, weekly, and monthly periods.
2. **Customer Segmentation**: Clustering customers based on purchasing behaviour to highlight the most profitable segments.
3. **Purchase Drivers**: Analysing correlations between customer demographics, purchase frequency, and sales value.
4. **Product Performance**: Determining the characteristics of bestselling products and understanding what drives their success.
5. **Customer Churn**: Using classification models to predict churn and uncover factors influencing customer retention.
6. **Price-Sales Relationship**: Investigating the correlation between product prices and sales volumes, including the impact of discounts and price variability.

The project is executed using Python and libraries such as Pandas, NumPy, Matplotlib, and Seaborn, ensuring robust data processing, analysis, and visualisation. Key deliverables include detailed visualisations, statistical analyses, and data-driven recommendations that address the retailer’s most pressing challenges and opportunities. This project not only demonstrates technical proficiency but also showcases the ability to apply data-driven strategies to real-world business scenarios.

Question 1: Trends in Product Sales Over Time

Below are the images of the daily, weekly, and monthly sales trends.

**Figure 1**: Daily sales trend from December 2010 to December 2011. **Figure 2**: Weekly sales trend highlighting baseline and peak patterns. **Figure 3**: Monthly sales evolution with seasonal fluctuations.

A graph showing a line of blue lines

Description automatically generated with medium confidence

Figure 1: Daily sales trend from December 2010 to December 2011

A graph with orange lines

Description automatically generated

Figure 2: Weekly sales trend highlighting baseline and peak patterns

A green line graph with white text

Description automatically generated

Figure 3: Monthly sales evolution with seasonal fluctuations

**Trends in Product Sales Over Time: Analysis and Recommendations**

**Analysis**

**Daily Sales Trends:** Between December 2010 and December 2011, daily sales exhibited considerable volatility, ranging from near-zero to peaks of approximately 110,000 units (**Figure 1**). The average daily sales during this period were around 30,000 units. Notable sales spikes occurred in early January (68,000 units), mid-May (70,000 units), and late October (108,000 units). A consistent weekly cyclical pattern was observed, with sales peaking on business days and dropping over weekends. Notably, Q4 2011 showed increased volatility, with more frequent high-volume sales days exceeding 60,000 units. The standard daily operating range shifted from 20,000–40,000 units in Q1-Q3 to 30,000–60,000 units in Q4.

**Weekly Sales Patterns:** Weekly sales revealed a baseline of approximately 150,000 units during Q1-Q3 2011 (**Figure 2**). The lowest weekly sales were recorded during the first week of January (5,000 units), reflecting a typical post-holiday slump. However, a steady growth trajectory was observed throughout the year. Sales stabilised at around 150,000–170,000 units weekly in Q1-Q3, with a dramatic uptick in Q4, reaching peaks of 375,000 units per week. Compared to daily sales, weekly trends showed reduced volatility. Sustained growth was particularly evident between September and November 2011, where weekly sales climbed from 150,000 to over 350,000 units.

**Monthly Sales Evolution:** Monthly sales exhibited distinct seasonal patterns (**Figure 3**). The average monthly sales during Q1-Q3 were approximately 650,000 units. The dataset revealed several key phases:

* Post-holiday decline: Sales dropped from 750,000 units in December 2010 to 500,000 units in February 2011.
* Spring recovery: Sales rebounded to 650,000–700,000 units from March to May.
* Summer plateau: Sales stabilised at around 680,000 units between June and August.
* Autumn surge: A significant increase occurred from September to November, with sales climbing from 1,000,000 to 1,450,000 units.
* December decline: Sales dropped sharply to 450,000 units, likely reflecting a partial month's data.

Year-over-year growth was substantial, with an 87% increase when comparing December 2010 to the peak in November 2011.

**Strategic Recommendations**

1. **Inventory and Supply Chain Management:**
   * Adopt dynamic inventory models that account for quarterly sales variations.
   * Increase safety stock by 40% from August to prepare for the Q4 demand surge.
   * Establish supplier agreements with flexible delivery schedules during Q4.
   * Conduct weekly stock reviews during peak seasons (September to December).
2. **Operational Efficiency:**
   * Scale warehouse operations to handle weekly sales volumes of 350,000+ units in Q4.
   * Implement automated picking systems to manage daily peaks exceeding 60,000 orders.
   * Develop contingency staffing plans to address volume spikes of up to 100%.
   * Schedule maintenance and upgrades during low-activity periods, such as early January.
3. **Marketing and Sales Strategies:**
   * Create a seasonal marketing calendar tailored to sales phases:
     + Q1: Post-holiday retention campaigns.
     + Q2: Spring revival promotions.
     + Q3: Early-bird holiday offers.
     + Q4: Peak season optimisation.
   * Introduce loyalty programmes to mitigate the January sales drop.
   * Launch B2B wholesale programmes to drive consistent off-season revenue.
   * Diversify product offerings with counter-seasonal items to stabilise Q1-Q3 sales.
4. **Financial Planning:**
   * Allocate 40% of the marketing budget to Q4, with 20% for each of the remaining quarters.
   * Increase working capital to support inventory expansion during peak seasons.
   * Implement seasonal pricing strategies to maximise profit margins.
   * Develop robust cash flow projections to accommodate the 87% year-over-year growth.
5. **Customer Experience Enhancement:**
   * Upgrade website infrastructure to support up to 3x normal traffic during sales peaks.
   * Introduce automated customer service solutions to handle high inquiry volumes.
   * Optimise shipping processes to guarantee timely deliveries during Q4.
   * Provide transparent tracking systems to reassure customers during peak-season delays.

This analysis provides a comprehensive understanding of sales trends and actionable strategies to enhance performance across operations, marketing, finance, and customer experience. These insights position the business to capitalise on seasonal demand fluctuations effectively.

Question 2: E-Commerce Customer Segmentation Analysis

A graph with colored dots

Description automatically generated

Figure 4: Cluster analysis scatterplot illustrating customer purchasing behaviour

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | TotalSales | TransactionCount | TotalQuantity |
| 0 | 1601.825493 | 83.146562 | 937.420468 |
| 1 | 225721.652500 | 711.750000 | 103007.250000 |
| 2 | 72682.466000 | 2395.133333 | 47137.600000 |

Figure 5: Cluster Table illustrating customer purchasing behaviour

**E-commerce Customer Segmentation Analysis**

To identify the most profitable customer segments, clustering techniques were employed to group customers based on their purchasing behaviour. The analysis reveals three distinct clusters, each characterised by unique purchasing patterns and contributions to revenue. The results provide actionable insights to enhance customer engagement and optimise revenue strategies.

**Cluster Analysis:**

1. **Cluster 0 (Low-Value Customers):**  
   This segment, This segment, comprising the largest number of customers, exhibits low average sales (£1,601.83), transactions (83 orders), and quantities purchased (937 items). Customers in this cluster demonstrate consistent, low-value purchasing behaviour with infrequent transactions. The tight clustering in visualisation confirms their homogeneity in purchasing patterns (Figure 4).
2. **Cluster 1 (High-Value Customers):**  
   Representing the smallest but most lucrative segment, this cluster averages £225,721.65 in sales, 712 transactions, and 103,007 items purchased. These customers exhibit high-value, regular purchasing patterns. Visualisations show a dispersed cluster, highlighting varying purchasing behaviours within this group (Figure 4). This segment generates the highest revenue, with sales 141 times greater than Cluster 0, despite its smaller size (Figure 5).
3. **Cluster 2 (Medium-Value/Wholesale Customers):**  
   This group is characterised by the highest transaction frequency, averaging 2,395 orders, £72,682.47 in sales, and 47,137 items purchased. Notable outliers include customers with over 7,000 transactions, indicative of wholesale buying patterns. While engagement is high, their total revenue is significantly lower than that of Cluster 1, suggesting bulk purchasing at lower margins (Figure 5).

**Key Insights:**

1. **Revenue Distribution:**  
   Cluster 1 contributes the highest revenue, underscoring the importance of high-value customers. Cluster 2 demonstrates the impact of wholesale buyers on transaction volume, while Cluster 0 comprises casual, low-frequency buyers.
2. **Purchase Behaviour:**  
   High engagement in Cluster 2 does not correlate directly with revenue, as bulk buyers often benefit from lower prices. Cluster 1 combines high transaction volumes and sales value, making it a critical revenue driver.
3. **Order Patterns:**  
   Cluster 0’s customers favour small, consistent purchases, while Cluster 2 shows a clear preference for bulk buying. Cluster 1 represents high value purchasing behaviour with significant potential for strategic engagement.

**Strategic Recommendations:**

1. **High-Value Customer Strategy (Cluster 1):**  
   To retain and expand this profitable segment, introduce a VIP programme offering premium benefits such as personalised product recommendations, priority service, and exclusive early access to new products. Marketing efforts should also target similar high-value businesses for acquisition.
2. **Wholesale Customer Development (Cluster 2):**  
   Implement volume-based pricing tiers and create a dedicated B2B portal for bulk orders. Automated reordering systems and personalised account management can enhance loyalty, while wholesale-specific marketing materials can attract new business customers.
3. **Low-Value Customer Activation (Cluster 0):**  
   Focus on increasing engagement and purchase frequency through targeted email campaigns, first-time bulk purchase incentives, and an accessible loyalty programme. Efforts should aim to upgrade these customers to higher-value segments.
4. **Operational Improvements:**  
   Introduce segment-specific shipping solutions, inventory management, and tailored marketing campaigns. These strategies ensure operational efficiency while addressing the distinct needs of each customer segment.
5. **Data Strategy:**  
   Monitor migration patterns between segments and track segment-specific KPIs. Implement predictive analytics to identify potential high-value customers and develop early-warning systems for churn. Establish satisfaction metrics tailored to each segment to ensure continuous improvement.

**Conclusion:**  
The clustering analysis provides valuable insights into customer segmentation for this UK-based online retailer. Strategic initiatives tailored to each cluster can drive engagement, optimise operational efficiency, and maximise revenue potential.

Question 3: Exploring Correlations Between Variables

A red and blue squares with white text

Description automatically generated

Figure 6: Correlation heatmap showing relationships between sales metrics

A graph showing the number of countries/regions

Description automatically generated

Figure 7: Geographical distribution of total sales by country

A graph with many colored dots

Description automatically generated

Figure 8: Comparing variables between purchase frequency and total sales

**Analysis of Factors Influencing Customer Purchases**

The analysis aims to explore the relationship between customer demographics, particularly geographical location and purchasing frequency, and their purchasing behaviour. This investigation provides insights into what drives sales for a UK-based online retailer specialising in unique all-occasion gifts.

**Correlation Analysis**

A moderate correlation of 0.42 was observed between total sales and transaction metrics, indicating that while there is some relationship, other factors significantly influence sales (Figure 6). Notably, a perfect correlation (1.0) was found between transaction count and purchase frequency, suggesting that frequent transactions naturally reflect repeated purchasing activity. However, purchase frequency alone is not strongly predictive of total sales value, highlighting varied basket sizes among customer segments.

**Geographical Distribution**

The United Kingdom emerges as the dominant market, contributing approximately £7 million in sales, far surpassing the Netherlands, which ranks second at around £0.4 million (Figure 7). Beyond the top five countries, sales drop significantly, indicating a long tail of smaller markets spread across 30+ countries. This reflects the retailer's strong home market advantage but limited penetration internationally.

**Purchase Patterns**

UK customers display the highest variability in purchase behaviour, with transaction frequencies ranging from 0 to over 8,000. Clusters of purchasing behaviour are evident:

1. Low frequency/low value, representing most transactions.
2. Medium frequency/medium value.
3. High frequency/high value, where a few outliers contribute disproportionately to sales (Figure 8).

Interestingly, several high-value UK customers generate more than £200,000 in revenue, yet their purchase frequencies vary widely. This underscores the importance of basket size and product mix over transaction frequency alone.

**Key Insights**

1. **Market Concentration**  
   Over 80% of the retailer’s revenue is derived from the UK, underscoring a strong domestic focus. However, the limited contribution from international markets suggests an opportunity for global expansion, particularly in high-potential regions like the Netherlands and Ireland.
2. **Customer Behaviour**  
   Transaction frequency does not consistently correlate with high sales value, pointing to significant variations in basket size and product preferences across customer segments.
3. **Business Model Implications**  
   The retailer operates with a mix of retail and wholesale customers, showing strong domestic performance but revealing untapped potential in international markets.

**Strategic Recommendations**

1. **Market Development**  
   To address the market concentration, prioritise international expansion by targeting high-potential regions such as the Netherlands, Ireland, and Germany. Develop market-specific product offerings and implement localisation strategies, including pricing and tailored marketing campaigns.
2. **Customer Segmentation Strategy**  
   Introduce tiered service levels based on customer purchase patterns. Target specific segments with personalised promotions and consider launching a VIP programme for high-value customers. Reactivation campaigns can be designed to engage low-frequency buyers.
3. **Operational Optimisation**  
   To enhance efficiency, establish regional distribution centres and optimise inventory for varying market demands. Develop international shipping solutions and offer market-specific customer service to improve the purchasing experience.
4. **Product Strategy**  
   Conduct an in-depth analysis of successful products in each market to create targeted product bundles and wholesale-specific lines. Adjust pricing to reflect local market dynamics and maximise appeal.
5. **Technology and Analytics**  
   Leverage predictive analytics to model customer behaviour and build customer lifetime value models. Implement real-time sales monitoring and automate market-specific reporting to enable data-driven decision-making.

Question 4:Identifying the Most Popular Products

A screen shot of a computer

Description automatically generated

Figure 9: Table of top 10 products by revenue

A screen shot of a computer

Description automatically generated

Figure 10: Table of top 10 products by quantity sold

A screen shot of a computer

Description automatically generated

Figure 11: Table of characteristics of top-selling products

A graph of blue and white bars

Description automatically generated

Figure 12: Bar chart of top 10 products by revenue

A green bar graph with white text

Description automatically generated

Figure 13: Bar chart of top 10 products by sales volume

**Analysis: Identifying the Most Popular Products**

**Product Performance Analysis**

The analysis revealed key insights into the best-performing products in terms of both revenue generation and sales volume. These insights provide a comprehensive understanding of the characteristics that contribute to the success of top-selling items.

**Top Revenue Generators**

The **Paper Craft Little Birdie** emerged as the highest revenue-generating product, contributing £168,469 in total revenue. This product also recorded the largest sales volume, with 80,995 units sold, highlighting its strong appeal within the craft and hobby market (Figure 9 and 12). Its moderate price point of £2.08 makes it accessible while generating significant revenue through high sales volumes.  
The **Regency Cakestand**, a premium product priced at £12.48, generated £142,592 in revenue despite its lower sales volume of 12,402 units. This indicates its value as a high-margin item, likely appealing to customers seeking premium offerings in the party or event category.  
The **White Hanging Heart T-Light Holder** contributed £100,448 to total revenue, with a mid-range price point of £2.89 and 36,725 units sold. This product’s consistent sales performance reflects its popularity in the home decor category.

**Sales Volume Leaders**

When considering total quantity sold, the **Ceramic Storage Jar** led the list with 77,916 units sold, followed by **WWII Gliders**, which recorded 54,415 units sold (Figure 10 and 13). Both products are low-priced items (ranging between £0.29 and £2.89), which aligns with the trend of high sales volumes for budget-friendly items. These results suggest that affordability is a critical factor in driving unit sales.

**Product Characteristics Analysis**

A deeper analysis of the product characteristics revealed distinct patterns across price and category distributions:

* **Price Stratification**: Products fall into three main tiers:
  + Budget (<£1): Items like gliders, cake cases, and popcorn holders dominate this range, appealing to customers seeking affordable options.
  + Mid-range (£1–£3): Products such as bird ornaments and storage jars attract a broader audience due to their balance of affordability and perceived value.
  + Premium (>£10): Items like the Regency Cakestand and product manuals cater to niche markets seeking higher-end products.
* **Category Distribution**: Products span various categories, reflecting a diverse customer base:
  + Home Decor: Items like t-light holders and storage jars are popular in this category, indicating consistent demand for decorative and functional products.
  + Party/Events: Products such as cake stands and bunting suggest strong sales driven by celebrations and special occasions.
  + Crafts/Hobbies: Paper craft kits and paint sets demonstrate the appeal of creative pursuits.
  + Seasonal: Items such as rabbit lights and ornaments highlight the significance of seasonality in driving sales.

**Strategic Recommendations**

To capitalise on these insights, the following recommendations are proposed:

1. **Product Development**  
   Expanding the craft and hobby line to introduce new designs and products can build on the success of items like Paper Craft Little Birdie. Premium home decor offerings could also attract higher-margin customers. Creating product bundles that combine complementary items across categories, particularly within the £2–£3 price range, may boost sales further.
2. **Inventory Management**  
   Optimising stock levels for top-performing products, especially during peak seasons, is critical. Implementing seasonal forecasting and category-specific reorder points can ensure product availability while minimising excess inventory. Balancing premium products with volume-driven items can help maximise profitability.
3. **Marketing Strategy**  
   High-margin items like the Regency Cakestand should be featured prominently in promotional materials. Craft and hobby-focused campaigns could tap into the strong demand for creative products. Bundling complementary items, such as storage jars with t-light holders, and creating category-specific promotions would encourage cross-selling.
4. **Pricing Strategy**  
   Maintaining a multi-tier pricing structure ensures accessibility across different customer segments. Margins on best-selling products can be optimised through strategic pricing adjustments. Offering volume discounts for craft items could incentivise bulk purchases, while premium pricing for unique products maintains brand differentiation.
5. **Category Management**  
   Expanding successful categories like home decor and crafts ensures sustained growth. Underperforming items should be phased out, and seasonal product rotations can keep the inventory fresh and relevant. Focusing on complementary product lines, such as party supplies paired with home decor, can further enhance customer appeal.

Question 5: Customer Churn Analysis

A screenshot of a computer

Description automatically generated

Figure 14: Confusion matrix for churn prediction model

A graph of a bar

Description automatically generated with medium confidence

Figure 15: Feature importance bar plot

**Analysis of Customer Churn**

To assess signs of customer churn, a classification model was developed using customer transaction history to differentiate between repeat buyers and one-time purchasers. The model's objective was to predict potential churn and identify key behavioural factors influencing customer retention.

**Model Performance and Findings**

The classification model achieved a perfect accuracy score of 1.0, meaning it effectively distinguished active customers from churned ones (Figure 14). The confusion matrix revealed that the model correctly identified 854 active customers (true positives) and 448 churned customers (true negatives), with no false positives or false negatives. The total customer base consisted of 1,302 customers, with 65.6% classified as active and 34.4% as churned. This distribution highlights a significant portion of customers at risk of churn, underlining the importance of proactive retention strategies.

**Feature Importance**

The model's feature importance analysis shed light on key predictors of customer churn (Figure 15):

1. **Frequency**: The most influential predictor, with a 0.75 importance score. This indicates that purchase frequency is critical to customer retention and serves as a primary indicator of churn risk.
2. **Monetary Value**: Contributed with a 0.15 importance score. While spending levels are important, they are less predictive of churn compared to frequency. High spending alone does not necessarily guarantee loyalty.
3. **Recency**: The least impactful feature, with a 0.10 importance score. Recent purchases were not strongly indicative of whether a customer would churn or remain active.

**Customer Behaviour Insights**

The analysis revealed clear behavioural distinctions between active and churned customers. Transaction frequency emerged as the key determinant of customer retention, overshadowing monetary value and recency. Interestingly, the findings showed that approximately one-third of customers are at risk of churn, underscoring the need for targeted intervention strategies.

**Strategic Recommendations**

Based on these findings, several strategic measures can be implemented to reduce churn and improve customer retention:

1. **Frequency-Based Interventions**
   * Introduce an early warning system to flag customers with declining purchase frequency.
   * Develop reactivation campaigns targeting customers who drop below critical frequency thresholds.
   * Implement automated engagement programs that encourage consistent purchasing.
   * Create loyalty tiers that reward customers based on their purchase frequency.
2. **Customer Retention Initiatives**
   * Focus retention efforts on maintaining or increasing purchase frequency.
   * Launch subscription or repeat-order programs to ensure regular transactions.
   * Establish reminder systems for replenishable or frequently purchased products.
   * Offer frequency-based rewards to incentivise recurring purchases.
3. **Churn Prevention Strategies**
   * Monitor transaction patterns to identify early indicators of churn.
   * Implement intervention strategies, such as personalised offers or discounts, at key churn-risk points.
   * Develop win-back campaigns targeting customers who have recently churned.
   * Utilise predictive alerts to proactively address potential churn cases.
4. **Data Enhancement**
   * Expand data collection to include additional behavioural metrics, such as browsing activity or time spent on site.
   * Analyse purchase patterns for specific product categories to identify trends.
   * Investigate seasonal variations in customer behaviour to better anticipate churn.
   * Introduce customer satisfaction tracking to gather feedback and address concerns proactively.
5. **Business Process Optimisation**
   * Automate the monitoring of churn prediction to ensure timely responses.
   * Establish a dedicated retention team focused on customer engagement and loyalty.
   * Develop a churn-risk scoring system to prioritise interventions effectively.
   * Implement real-time triggers to deploy retention strategies at critical moments.

**Conclusion**

This analysis highlights the critical role of purchase frequency in retaining customers and reducing churn. While monetary value and recency offer some insights, they are less predictive of customer behaviour compared to frequency patterns. The proposed strategies aim to strengthen customer engagement and improve retention outcomes, ultimately enhancing business sustainability and profitability. This approach also demonstrates the potential for strong customer segmentation, enabling data-driven decisions to maximise value and mitigate churn risks.

Question 6: Price-Sales Relationship

A graph with numbers and lines

Description automatically generated

Figure 16: Scatterplot of price vs. sales volume

Correlation Coefficient: -0.04, P-value: 9.32e-03

A graph with blue dots

Description automatically generated

Figure 17: Boxplot showing price distribution across sales categories

**Analysis of the Relationship Between Product Prices and Sales Volume**

This analysis explores the relationship between product pricing and sales volume, aiming to determine whether price changes impact sales patterns. The dataset, comprising transactions from a UK-based online retailer specialising in unique, all-occasion gifts, offers insights into price sensitivity and purchasing behaviours over a year-long period.

**Price Sensitivity**

The correlation analysis reveals a weak negative relationship between product prices and sales volume, with a correlation coefficient of -0.04 (Figure 16). Although weak, this relationship is statistically significant (p-value: 9.32e-3), indicating that as prices increase, sales volumes tend to decrease slightly. Most products fall below the £100 price range, and the highest sales volumes, typically between 60,000 and 80,000 units, are concentrated in the £0-£20 range.

**Price Distribution Patterns**

The price distribution is heavily skewed towards lower price ranges, with a dense cluster of products priced under £50. Products priced above £100 are sparsely distributed, with a notable outlier around £700 that exhibits minimal sales volume. Additionally, sales volumes drop sharply for products priced above £50, emphasising customer preference for budget-friendly items.

**Price Variability and Its Impact**

Products in the lower price range exhibit the highest sales volatility, with a standard deviation clustering below 100 (Figure 16 and 17). Some outliers have deviations of 300 and 600, indicating significant variability in sales performance for specific items. Conversely, products in the premium price range show more stable sales volumes, reflecting consistent but low demand.

**Key Insights**

1. **Price Sensitivity**  
   Products in the £0-£20 range, particularly those priced between £2 and £5, drive the highest sales volumes (Figure 17). Premium products, while maintaining steady sales, exhibit considerably lower volumes.
2. **Price Stability**  
   Price variations have a more pronounced impact on budget items, while premium products demonstrate stability. Seasonal or promotional factors also appear to influence the volatility in the low-price segment.

**Recommendations**

1. **Pricing Strategy**
   * Focus on optimising the £2-£5 price range to capitalise on high-volume sales.
   * Retain premium products to maintain margins, despite their lower sales volumes.
   * Implement dynamic pricing for items below £20 to maximise revenue during peak demand.
   * Introduce volume-based discounts to boost sales of budget-friendly items.
2. **Inventory Management**
   * Maintain higher stock levels for low-priced, high-demand items.
   * Use a just-in-time inventory strategy for premium products to reduce holding costs.
   * Keep buffer stock for items exhibiting high sales volatility.
3. **Product Development**
   * Prioritise new product launches in the optimal price range (£2-£5).
   * Develop bundles combining low-priced and premium items to appeal to diverse customer segments.
   * Create premium versions of best-selling products to expand the premium product portfolio.
4. **Sales Optimisation**
   * Leverage strategic discounting for high-volume, low-price items.
   * Position premium products with clear value propositions to justify their pricing.
   * Adopt category-specific pricing strategies to align with customer purchasing behaviour.
   * Adjust prices seasonally to capture demand fluctuations effectively.

**Conclusion and Recommendations**

**Main Insights and Findings**

This e-commerce data analysis provided several key insights into customer purchasing behaviour, product performance, and sales patterns. The analysis revealed that:

1. **Seasonal Trends and Sales Volumes**
   * Peak sales periods coincide with holiday seasons, indicating strong seasonal demand for the retailer's all-occasion gift products.
   * Sales performance is consistent across weeks, with notable spikes around the Christmas period.
2. **Customer Segmentation and Profitability**
   * Customers can be categorised into distinct groups based on purchasing behaviour. The most profitable segment comprises loyal customers with frequent and high-value transactions, primarily wholesalers.
   * Retention strategies targeting high-value customers could significantly enhance revenue.
3. **Product Performance**
   * Low-priced items (£2-£5 range) contribute the highest sales volumes, whereas premium products generate steady but limited sales, maintaining profitability through higher margins.
   * Bestselling products are characterised by their affordability and practical appeal, making them attractive to both retail and wholesale customers.
4. **Pricing and Sales Volume Relationship**
   * A weak negative correlation between price and sales volume underscores customer price sensitivity. Products in the budget range exhibit high sales volatility, whereas premium products show stable demand.
5. **Geographical Insights**
   * The United Kingdom remains the primary revenue source, with noticeable contributions from European markets. Expanding into underperforming regions could unlock additional growth.

**Positives and Negatives of This Project**

**Positives:**

* The dataset provided a comprehensive view of transaction details, enabling a deep dive into sales trends, customer behaviour, and product performance.
* The use of clustering, correlation analysis, and time series methods allowed for robust and actionable insights.
* The project showcased a variety of analytical techniques, making it a strong portfolio piece for data analyst and data scientist roles.

**Negatives:**

* The dataset lacked specific promotional and marketing data, which could have provided insights into campaign effectiveness.
* Missing customer demographic details limited the depth of customer segmentation.
* Outliers in the data required extensive pre-processing, which, while manageable, could introduce bias if not handled carefully.

**Recommendations and Next Steps**

1. **Optimise Pricing Strategies**
   * Maintain a focus on products priced in the £2-£5 range to drive volume sales.
   * Experiment with dynamic pricing for high-demand periods and implement discounts for budget-friendly products to attract more customers.
   * Introduce bundled pricing strategies to combine low- and high-priced products, increasing average transaction value.
2. **Enhance Customer Retention**
   * Develop loyalty programmes for high-value customers, particularly wholesalers, to ensure repeat purchases.
   * Personalise offers and incentives for frequent buyers using segmentation insights.
3. **Expand Product Development**
   * Focus on introducing new products in the optimal price range (£2-£5) to capture customer demand.
   * Consider launching premium variants of bestselling items to appeal to high-margin segments.
4. **Regional Growth Strategy**
   * Expand marketing efforts in underperforming regions to diversify revenue streams.
   * Tailor product offerings to the preferences of international markets, based on regional sales data.
5. **Data Enrichment and Future Analysis**
   * Incorporate additional data, such as marketing campaigns and customer demographics, to enhance future analyses.
   * Use advanced machine learning techniques, such as predictive modelling, to forecast sales trends more accurately.

This project highlights the importance of leveraging data-driven strategies to optimise pricing, target high-value customers, and enhance product offerings. Implementing these recommendations will position the retailer to capture new opportunities, improve operational efficiency, and sustain long-term growth in the competitive e-commerce market.