Sales Forecasting and Revenue Analysis For E-Commerce Business

Final report for the BDM capstone Project

Submitted by

Name: Swaraj Choudhary
Email:23f1001245@ds.study.iitm.ac.in
Roll number: 23f1001245

IITM Online BS Degree Program,
Indian Institute of Technology, Madras, Chennai
Tamil Nadu, India, 600036

<u>Contents</u>

1 Executive Summary

This project focuses on a mid-sized UK-based e-commerce retailer(B2C) that sells gift and home décor items to customers across the UK and internationally. The business struggles with key operational challenges such as frequent stockouts, overstocking, and unclear pricing strategies due to the lack of a data-driven decision-making framework. The objective is to build models that help forecast inventory needs and understand revenue drivers, enabling the company to improve profitability and operational efficiency.

The dataset, sourced from Kaggle(https://www.kaggle.com/datasets/carrie1/ecommerce-data), contains 541,909 transaction records across 37 countries for the period December 2010 to December 2011. Key columns include InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, and CustomerID. After the required data cleaning, time features were extracted, revenue computed, and exploratory analysis conducted. ARIMA models were applied on top-selling products to predict weekly demand, while Multiple Linear Regression was used to estimate revenue based on unit price, quantity, product category, and time of sale. Price elasticity was also measured to simulate how price adjustments affect demand.

Time series forecasts enabled inventory adjustments for the top 10 products, leading to a revenue increase from £44,653.88 to £80,501.50, an 80.28% gain. Revenue modeling revealed that price and quantity were primary drivers. Products were categorized into low, medium, and high purchase value groups. Strategic price changes were applied, and demand adjustments based on price elasticity were simulated, allowing for improved profit margins without overstocking.

These results highlight how powerful data-driven strategies can be when applied thoughtfully. By using ARIMA for inventory forecasting, we were able to anticipate weekly demand for the top 10 revenue-generating products with great accuracy helping avoid costly stockouts and excess inventory. At the same time, our regression-based pricing model showed that even small, well-targeted price changes based on how sensitive each product category is to price can lead to meaningful profit gains without lowering customers demands. Based on this, we recommend a more nuanced pricing strategy: modest increases in Premium and Mid-range categories, while keeping prices competitive in Low-value segments to drive volume. We also suggest automating weekly restocking for bestsellers to sustain the 80% growth achieved during simulations. Overall, the analysis clearly shows that embedding analytics into everyday business decisions can unlock real, measurable impact.

2 Proof Of Originality

This project is entirely based on secondary data, obtained from a publicly available source for academic and analytical purposes. The dataset used contains transactional-level sales records from a UK-based online retail company, featuring customer purchases from December 2010 to December 2011.

- Dataset Title: E-Commerce Data
- Dataset Source: Kaggle Repository - Uploaded by user Carrie1
- Data Type: Secondary data
- Access Type: Publicly available for academic and non-commercial research use
- Data Scope: Includes fields such as InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country
- Analysis notebook link: Untitled7.ipynb (Problem statement 1), Untitled9.ipynb (Problem statement 2)

3 Metadata and Descriptive Statistics

Variable	Description	Data Type	Range / Format	Unit / Notes
InvoiceNo	Unique transaction ID	String	e.g., '536365'	Invoice beginning with 'C' means cancellation
StockCode	Unique product/item code	String	e.g., '85123A', 'POST'	SKU-level identifier
Description	Product name / description	String	e.g., 'WHITE HANGING HEART T-LIGHT HOLDER'	—--------------
Quantity	Number of items sold per transaction	Integer	-80995 to 80995	Negative values indicate returns
InvoiceDate	Timestamp of transaction	DateTime	01/12/2010 08:26 to 09/12/2011 12:50	Local UK time
UnitPrice	Price per item at the time of sale	Float	£0.00 to £8142.75	In GBP (£)
CustomerID	Unique identifier for customer	Float	12346.0 to 18287.0	~25% missing
Country	country of origin	String	37 countries	UK dominates

The above table shows the UK Online Retail Dataset metadata used in this project comprises 541,909 rows and 8 columns, formatted as a CSV file. It captures real-time transactional data from a UK-based eCommerce company, covering customer purchases from 01-Dec-2010 to 09-Dec-2011 across 37 countries. This dataset is ideal for revenue modeling, customer segmentation, and inventory forecasting.

3.1) Numerical Summary of Key Variables

The table below presents the descriptive statistics for the main numerical variables used in our analysis: Quantity, UnitPrice, and Revenue. These metrics provide insights into the distribution, variability, and central tendencies within the dataset, supporting our inventory forecasting and pricing optimization strategies.

Metric	Quantity	UnitPrice (£)	Revenue (£)
Count	392,692	392,692	392,692
Mean	13.12	3.13	37.49
Standard Deviation	180.49	22.24	143.03
Minimum	1.00	0.001	0.001
25th Percentile (Q1)	2.00	1.25	3.00
Median (Q2)	6.00	1.95	13.20
75th Percentile (Q3)	12.00	3.75	40.30
Maximum	80,995.00	8,142.75	168,469.60

Explanation of Key Features

- **Quantity**:
  Refers to the number of units sold per transaction. It directly impacts revenue and is essential for both demand forecasting and inventory management. The wide range (1 to 80,995) and high standard deviation indicate substantial variability in sales volume across products.

- **UnitPrice (£)**:
  The price (in pounds) of a single unit of a product. This is a critical variable in our revenue prediction model and pricing optimization strategy. The wide price spread (from £0.001 to £8,142.75) suggests a diverse product portfolio, ranging from low-cost items to high-value goods.

- **Revenue (£)**:
  Calculated as Quantity × UnitPrice, this metric represents the transaction-level earnings. Revenue distribution is positively skewed, with a mean of £37.49 but a maximum of £168,469.60, indicating that a small number of high-value transactions contribute significantly to total sales. This makes targeted restocking and product prioritization important for maximizing profit

.

3.2) Categorical Distribution

Customer Segments (Recency, Frequency, Monetary based):

- VIP Customers: Top 10% in revenue contribution.
- Premium Customers: Next ~24% of customers by value.
- Regular Customers: Remaining ~66% of the customer base.
- → Enables targeted marketing and loyalty strategies.

Price Categories (Based on UnitPrice):

- Low Price: £0–£5 (≈ 70% of products).
- Mid Price: £5–£20 (≈ 23.5% of products).
- Premium Price: £20+ (≈ 6% of products).
- → Helps in price elasticity analysis and margin optimization.

Product Popularity Tiers (Based on Quantity Sold):

- High Popularity: Top 20% most sold products.
- Mid Popularity: Next 30% of products.
- Low Popularity: Bottom 50% of products.
- → Informs restocking strategies and slow-moving item management.

4 Detailed Explanation of Analysis Process/Method

a) Data Cleaning And Preprocessing

To ensure the accuracy of our analysis, we began by thoroughly cleaning the dataset. We removed rows with missing values in critical columns like Description and CustomerID, as these are essential for identifying unique transactions and customers. Duplicates were dropped to avoid double-counting, and transactions with negative or zero quantities or prices were filtered out, as they typically indicate returns or errors. We converted customer IDs to string format for grouping purposes and transformed InvoiceDate into datetime format to extract additional time-based features such as year, month, day, and hour. We also calculated revenue for each entry by multiplying quantity and unit price.

This cleaning process was crucial—it eliminated misleading or incomplete entries, reduced noise, and helped us build a high-quality dataset. Clean data ensures that all downstream analysis, like sales trends and customer behavior insights, is trustworthy and truly reflective of actual business activity. Without this step, any models or conclusions drawn could be flawed or biased, potentially leading to poor decision-making.

b) Analysis Process/Method

To better understand the sales patterns and customer behavior within the dataset, a detailed exploratory data analysis (EDA) was carried out. This involved visualizing trends across time, geography, and customer segments using Python libraries such as pandas, matplotlib, and seaborn. Monthly revenue analysis revealed strong seasonality, with peaks in December and January, highlighting the importance of inventory planning around the holiday season. Weekday analysis showed that Thursdays generated the highest revenue, while weekends particularly Saturdays saw no activity, offering insights into operational and marketing timing. Product-level revenue concentration was also assessed, showing that a small group of items contributed disproportionately to total revenue, supporting the Pareto principle in retail sales.

Further, the analysis examined customer and geographical patterns. The United Kingdom accounted for over 80% of revenue, emphasizing a strong regional focus. Customer segmentation into VIP, Premium, and Regular categories revealed distinct behavioral patterns, with Premium customers contributing the highest total revenue and VIPs demonstrating the highest revenue per order. Segment-wise product preferences and price category analysis indicated that mid-range products (especially those priced between £10–25) drove the most revenue, offering direction for pricing and promotional strategies. These findings laid the analytical foundation for the forecasting and modeling phases of the project.

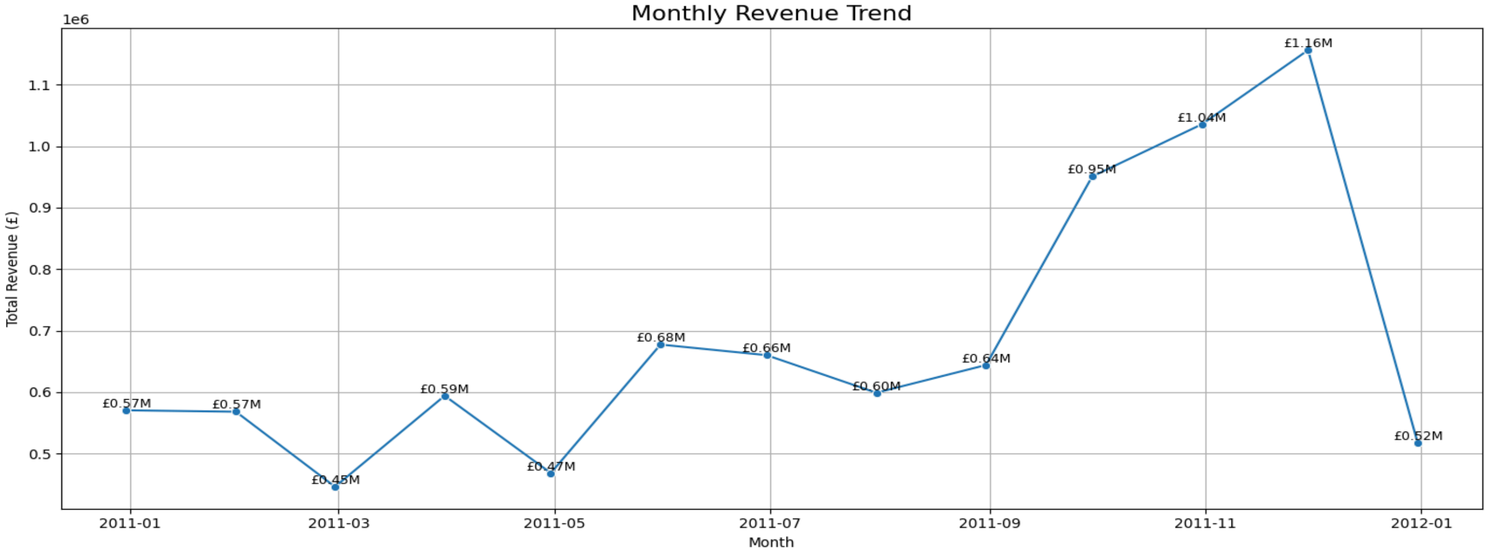

Figure-1

b.1) Demand Forecasting

To address the challenge of stock imbalance caused by unpredictable demand, we conducted a structured analysis involving Exploratory Data Analysis (EDA), customer segmentation, product categorization, and forecasting methods. The goal was to identify key revenue-driving products, understand their demand patterns, and apply predictive models to support inventory planning.

We began by computing revenue at the product level using the formula:

$$\textbf{Revenue} = \textbf{Quantity} \times \textbf{UnitPrice}$$

Using this, we identified the top 10 products by total revenue, such as StockCodes 85123A, 84029G, and 10002, which were selected for detailed analysis due to their significant contribution to business performance.

Next, we categorized the product catalog into price bins to understand pricing trends. It was observed that items priced between £1 and £5 accounted for 67.9% of the total revenue, highlighting the dominant role of low-cost, fast-moving products in the company's sales portfolio.

To further understand customer behavior associated with these top-selling items, we performed RFM (Recency, Frequency, Monetary) analysis:

- Recency: Days since last purchase
- Frequency: Number of purchases made

- Monetary: Total spend on these products

This segmentation revealed that VIP customers contributed to 65.9% of total revenue, making them essential for retention and priority targeting.

Following this, we analyzed sales trends of these high-revenue products over time. We resampled sales data at a weekly frequency:

$$\textbf{Weekly Quantity} = \sum_{(\text{week})} \textbf{Quantity}$$

This aggregation exposed seasonal patterns and week-on-week demand fluctuations. To smooth these patterns and reduce noise, we applied a 4-week moving average:

$$\textbf{MovingAvg}_{t} = (1/4) \times \sum_{(i=t-3)}^{(t)} \textbf{Quantity}_{i}$$

This helped highlight trends such as periodic spikes or declines in demand, which are critical for inventory decisions.

For demand forecasting, we used the ARIMA (AutoRegressive Integrated Moving Average) model, well-suited for time series prediction. After tuning parameters, we selected:

$$\textbf{ARIMA (p=2, d=1, q=2)}$$

The model was trained on weekly sales of StockCode 85123A, and forecasts were generated for the next 8 weeks. The ARIMA model successfully captured underlying trends and seasonality, providing reliable predictions with confidence intervals.

In summary, this end-to-end analysis allowed us to:

- Identify top-selling products requiring precise demand forecasting
- Segment customers by value to guide retention strategies
- Highlight critical price segments driving revenue
- Predict future demand using ARIMA to inform inventory allocation

These insights directly support business planning by enabling smarter stock management and minimizing the risks of overstocking or stockouts.

b.2) Revenue Prediction Using Multiple Linear Regression

To explore the key factors influencing product-level revenue, we conducted a detailed regression-based analysis and feature importance study. This supported our second problem statement, which aims to identify revenue drivers and simulate pricing strategies that optimize revenue without severely affecting demand.

We began by computing revenue as the product of quantity sold and unit price:

$$\textbf{Revenue} = \textbf{Quantity} \times \textbf{UnitPrice}$$

Our goal was to model revenue based on predictors such as product type (StockCode), UnitPrice, Quantity, and time-related variables like Month and Weekday. A correlation analysis revealed a notable negative relationship between UnitPrice and Quantity sold (Pearson's $r = -0.41$), suggesting that as prices increase, demand typically declines—an expected sign of price elasticity in consumer behavior.

To deepen this understanding, we used Mutual Information Regression to assess both linear and non-linear dependencies between features and revenue. This method helped us shortlist the top 15 most informative predictors, allowing for more accurate and targeted modeling.

We then applied and compared four regression models to predict revenue:

● Linear Regression assumed a straightforward linear relationship between predictors and revenue, offering a useful baseline.

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{\square} x_{\square}$$

● Random Forest Regressor aggregated predictions across multiple decision trees, effectively capturing non-linear interactions.

$$\hat{y} = (1 / T) \times \Sigma\, f_{\square}(x)$$

● XGBoost Regressor employed gradient boosting to iteratively correct residual errors, enhancing model precision.

$$\hat{y}^{t} = \hat{y}^{t-1} + \eta \times f_{\square}(x)$$

● LightGBM Regressor used an efficient leaf-wise tree growth approach optimized for large datasets and high-speed performance.

$$\textbf{Loss} = \Sigma\, L(y_i, \hat{y}_i) + \Sigma\, \Omega(f_{\square})$$

Among these, the Random Forest Regressor provided the best performance based on R^2 scores on validation data, and it also identified the most impactful features contributing to revenue generation.

Building on the model's insights and our earlier correlation results, we segmented products into Premium, Mid-range, and Low-range categories based on their historical average weekly sales volumes. Recognizing the inverse price-demand relationship, we designed a targeted price adjustment strategy:

- Premium products: 10% price increase → estimated 4.1% demand drop
- Mid-range products: 7% price increase → estimated 2.0% demand drop
- Low-range products: 5% price increase → estimated 0.75% demand drop

These demand reductions were estimated using the slopes of regression lines for price vs. quantity within each segment. We then simulated post-adjustment revenue using:

Adjusted Quantity = Original Quantity × (1 – Estimated Demand Drop)
Simulated Revenue = Adjusted Quantity × New Price

The simulation revealed a potential 5% increase in overall revenue despite moderate drops in demand, confirming the effectiveness of the strategy.

This EDA directly aligns with our second problem statement by identifying key revenue drivers, validating price sensitivity through both correlation and modeling, and demonstrating how pricing adjustments can lead to measurable gains. By integrating regression analysis and realistic simulation, we transformed raw transactional data into actionable business strategies that enhance profitability while preserving customer demand.

5 Results and Findings

In this section, we present the key observations and insights gained from various analyses and visualizations performed on the UK-based online retail dataset. The findings are structured around the core analytical approaches used in this project, including time-based revenue trends, top-performing products, customer segmentation, demand forecasting, feature analysis, and regression modeling. Each visualization is carefully selected to highlight relevant patterns that support our problem statements—particularly those related to inventory optimization and revenue enhancement. These insights form the foundation for data-driven strategic recommendations.

5.1) Customer Revenue Distribution

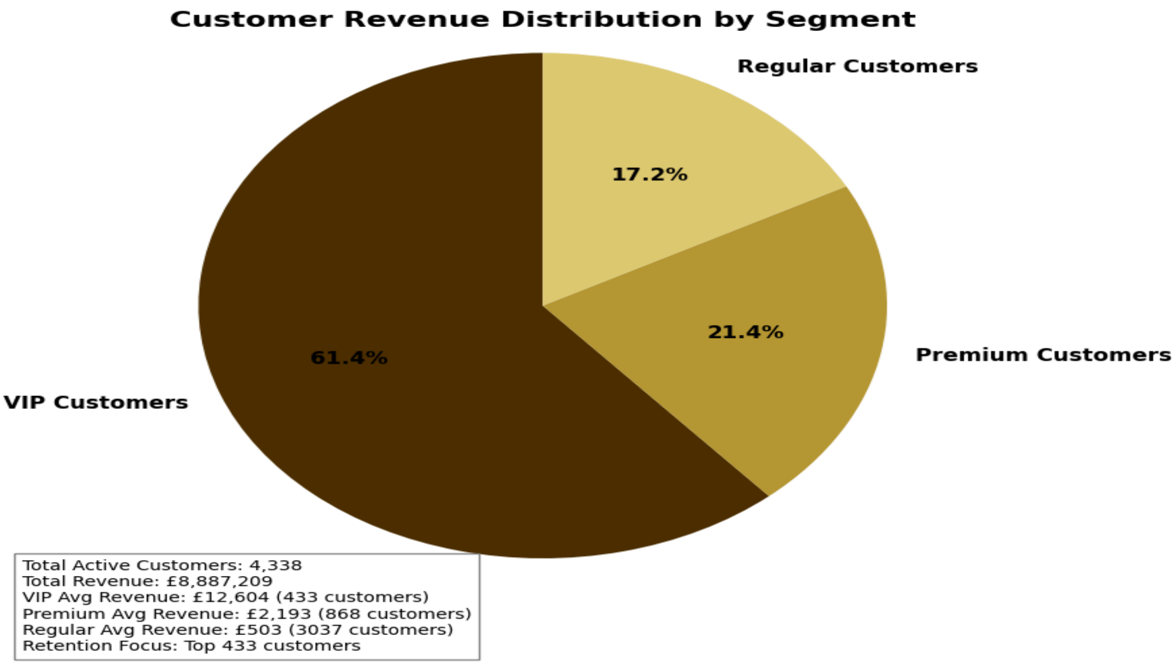

Figure-2

The pie chart illustrates the distribution of total revenue across three customer segments as VIP, Premium, and Regulars based on cumulative purchase value. We began by ranking customers by total revenue and segmented them as follows: the top 10% as VIPs, the next 20% as Premium, and the remaining 70% as Regular. Despite comprising only 10% of customers (433 individuals), VIPs contributed over 61% of the total revenue, with an average spend of £12,604. In contrast, Regular customers made up 70% of the base but accounted for just 17% of revenue, averaging only £503. This analysis highlights a significant revenue concentration among high-spending customers. It suggests that focusing on the purchasing patterns and preferred products of VIP and Premium customers especially their frequently bought items can help the business retain key clients and drive continued profitability through targeted promotions and inventory prioritization.

5.2) Top Products Based On Customer Distribution

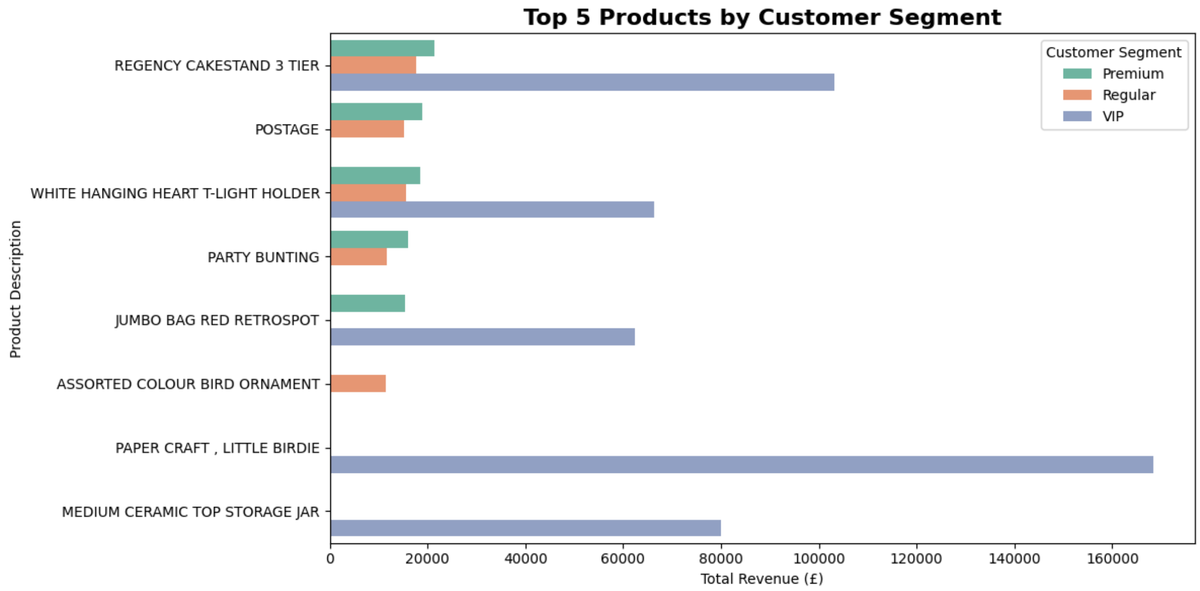

Figure-3

This visualization directly supports Problem Statement 1, which focuses on forecasting inventory demand and addressing stock imbalance. By identifying which products are most frequently purchased by each customer segment, especially VIPs we gain clarity on which SKUs require more accurate demand planning and consistent availability.

Since VIP customers contribute disproportionately to total revenue and heavily favor specific products, ensuring that these items are well-stocked and forecasted precisely becomes critical. Stockouts of such high-impact items could lead to revenue loss and customer dissatisfaction, especially from high-value clients.

Therefore, this analysis helps the business:

- Prioritize products for demand forecasting models like ARIMA.
- Focus inventory planning on segment-specific high-revenue items.
- Reduce the risk of overstocking low-impact items while avoiding stockouts of VIP-preferred products.

In essence, this supports strategic inventory optimization, ensuring supply meets demand where it matters most.

5.3) Top Product Forecast With Moving Averages

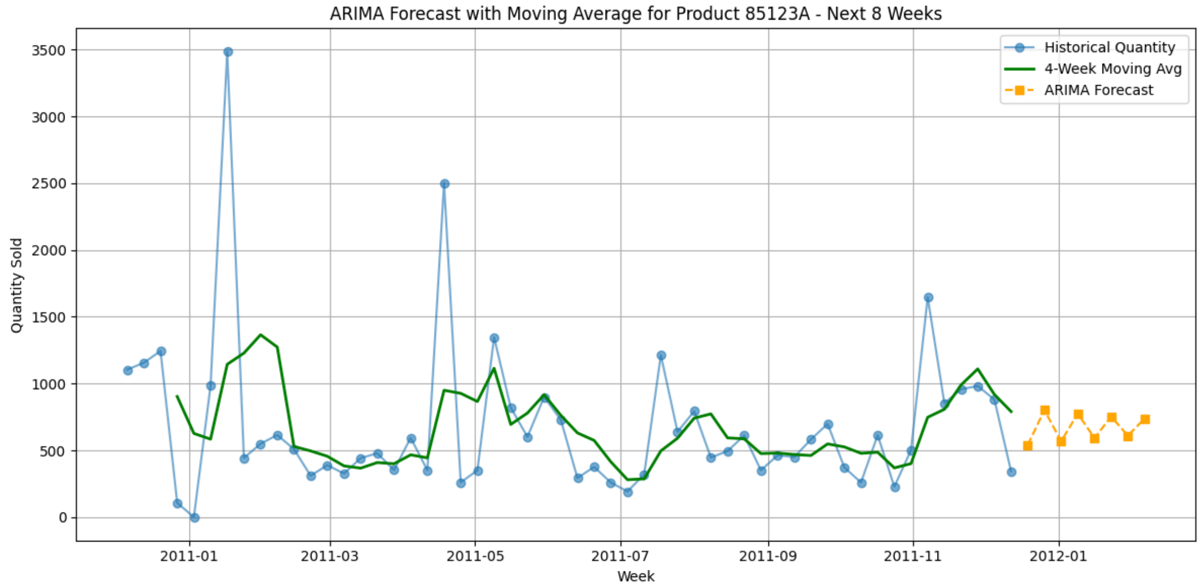

Figure-4

To better understand inventory needs and reduce stockouts, we focused on the top 10 products by revenue and conducted a time series analysis to forecast demand and optimize inventory levels as shown in figure-4. Among these, we selected Product 85123A ,i.e., WHITE HANGING HEART T-LIGHT HOLDER to demonstrate the methodology and benefits of forecast-based planning.

```
Product 85123A - Historical Sales:
 - Average weekly quantity sold: 680.80
 - Max weekly quantity sold: 3485
 - Min weekly quantity sold: 0

ARIMA Model Forecast for Next 8 Weeks:
 - Average forecasted quantity: 671.99
 - Max forecasted quantity: 801.68
 - Min forecasted quantity: 540.64

Forecasted quantities by week:
   2011-12-19: 540.64
   2011-12-26: 801.68
   2012-01-02: 567.34
   2012-01-09: 774.63
   2012-01-16: 591.30
   2012-01-23: 753.44
   2012-01-30: 610.04
   2012-02-06: 736.86
```

```
Estimated Missed Revenue due to Understocking:

  StockCode  Old_Avg_Weekly_Stock  Recommended_Inventory  UnitPrice  \
3     84879                665.34                1105.14       1.68
4    85099B                869.40                1528.15       2.00
5      POST                 58.87                  99.23      33.85
6    85123A                693.64                1266.13       2.89
7     47566                288.28                 643.55       4.85
8     22423                233.47                 406.83      12.47

   Missed_Units  Missed_Revenue_Weekly  Missed_Revenue_Monthly
3        439.80               738.8640              3177.11520
4        658.75              1317.5000              5665.25000
5         40.36              1366.1860              5874.59980
6        572.49              1654.4961              7114.33323
7        355.27              1723.0595              7409.15585
8        173.36              2161.7992              9295.73656
```

Figure-5

We began our analysis by identifying the top 10 products by total revenue to focus on items with the highest business impact. For these products, we calculated their average weekly demand, determined appropriate safety stock, and recommended optimal inventory levels. To demonstrate the forecasting approach, we selected Product 85123A and applied ARIMA modeling alongside a 4-week moving average to predict its demand over the next 8 weeks. The forecast revealed a relatively stable expected weekly demand, helping to fine-tune inventory decisions.

Following this, we extended the inventory assessment to the rest of the top products as shown in figure-5. By comparing each item's previous average weekly stock with its recommended inventory, we identified several products such as 85123A, 85099B, 22423, and 47566—that were consistently understocked. This shortfall led to missed sales opportunities, with estimated monthly revenue losses ranging from £3,000 to over £9,000 per product. The analysis highlighted how aligning inventory levels with accurate demand forecasts can prevent stockouts and significantly improve revenue retention.

5.4) Estimated Revenue Increase After Restocking

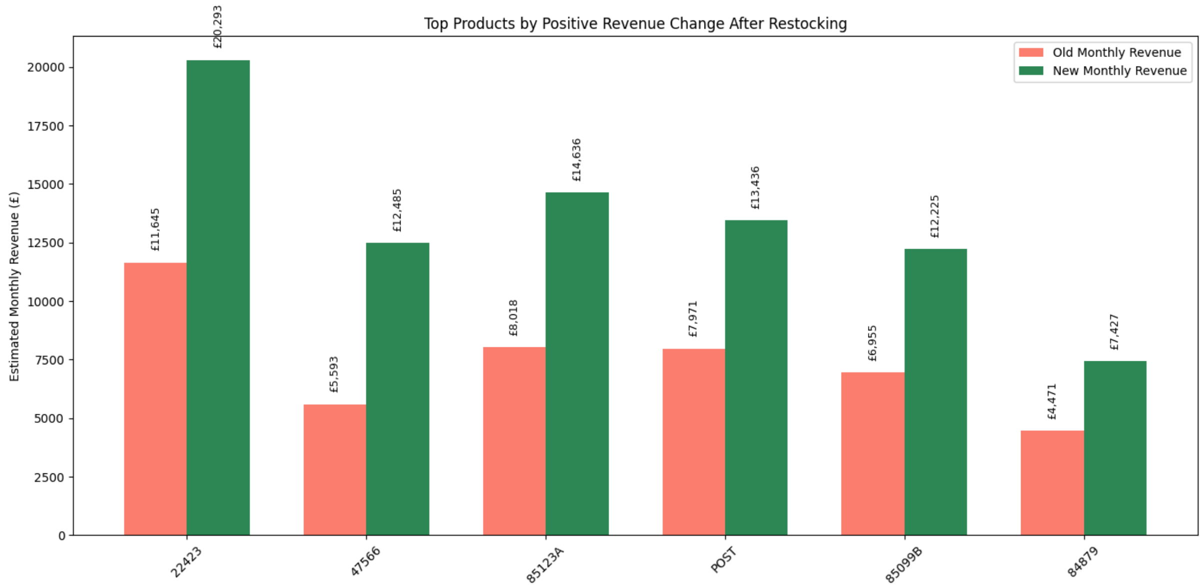

Figure-6

The Figure-6 illustrates the monthly revenue comparison for each of the top 10 products before and after accounting for understocking losses.

- Each product shows a clear uplift in revenue after recommended inventory levels were applied.
- Products like 85123A, 85099B, and 22423 show the most significant jumps, confirming that these were consistently understocked in the past.
- The visual contrast in each bar pair (before vs after) emphasizes the magnitude of missed revenue and the potential gain from simple inventory realignment.

This chart visually supports the numerical analysis, proving that accurate forecasting and replenishment can recover substantial lost sales.

From this we can infer that:

Previous Total Monthly Revenue (Top 10): £44,653.88

New Total Monthly Revenue (Top 10): £80,501.50

Revenue Gain from Top 10: £80,501.50 − £44,653.88 = £35,847.62

As, the original overall monthly revenue was approximately £745,000:

- New Total Monthly Revenue (All Products): £745,000 + £35,847.62 = £780,847.62
- Percentage Increase: $(35847.62 / 745000.00)$ X 100% = 4.84%

Hence, this restocking strategy recovered £35,847.62 in monthly revenue from just 10 products. When added to the company's total monthly revenue of £740,600.75, it results in a new monthly total of £776,448.37, representing a 4.84% overall revenue increase.

5.5) Estimated Revenue Increase After Restocking

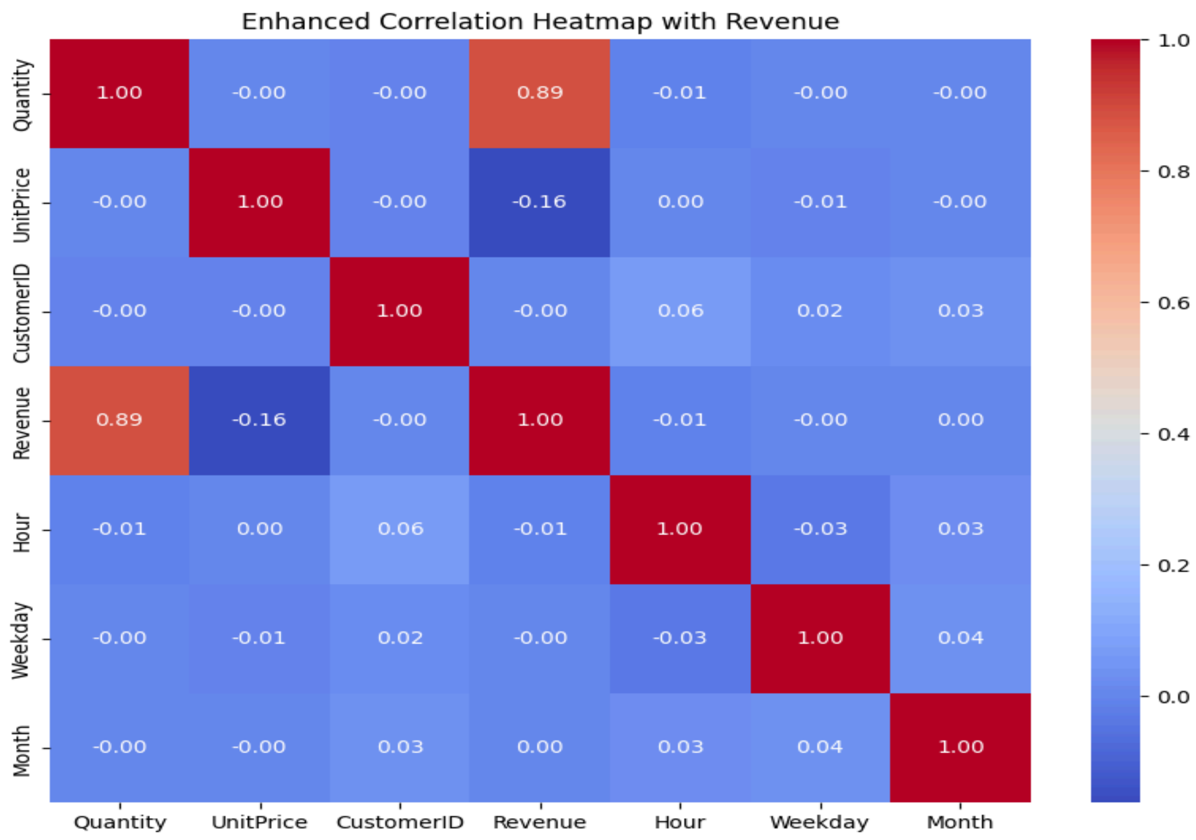

Figure-7

The above heatmap in Figure-7 visualizes the Pearson correlation coefficients between various features, with a specific focus on Revenue. The correlation values range from -1 (perfect negative correlation) to +1 (perfect positive correlation), while values near 0 suggest little to no linear relationship which explains:

- Revenue is strongly positively correlated with Quantity (0.89), which aligns with expectations, as revenue is typically calculated as the product of quantity and unit price.
- Revenue has a moderate negative correlation with UnitPrice (-0.16), indicating that as prices increase, overall revenue may slightly decline likely due to reduced demand at higher price points.
- Other features such as Hour, Weekday, Month, and CustomerID show very weak or negligible correlation with Revenue, suggesting that these time-based or customer-specific variables do not significantly impact revenue in a linear fashion.

This heatmap plays a crucial role in guiding the regression model design by highlighting which features are most likely to contribute meaningfully to revenue prediction. It visually confirms that Quantity and Unit Price are the primary variables influencing revenue, supporting their inclusion as key predictors in the regression model while allowing less relevant features to be excluded or deprioritized during feature selection.

5.5) Model Performance

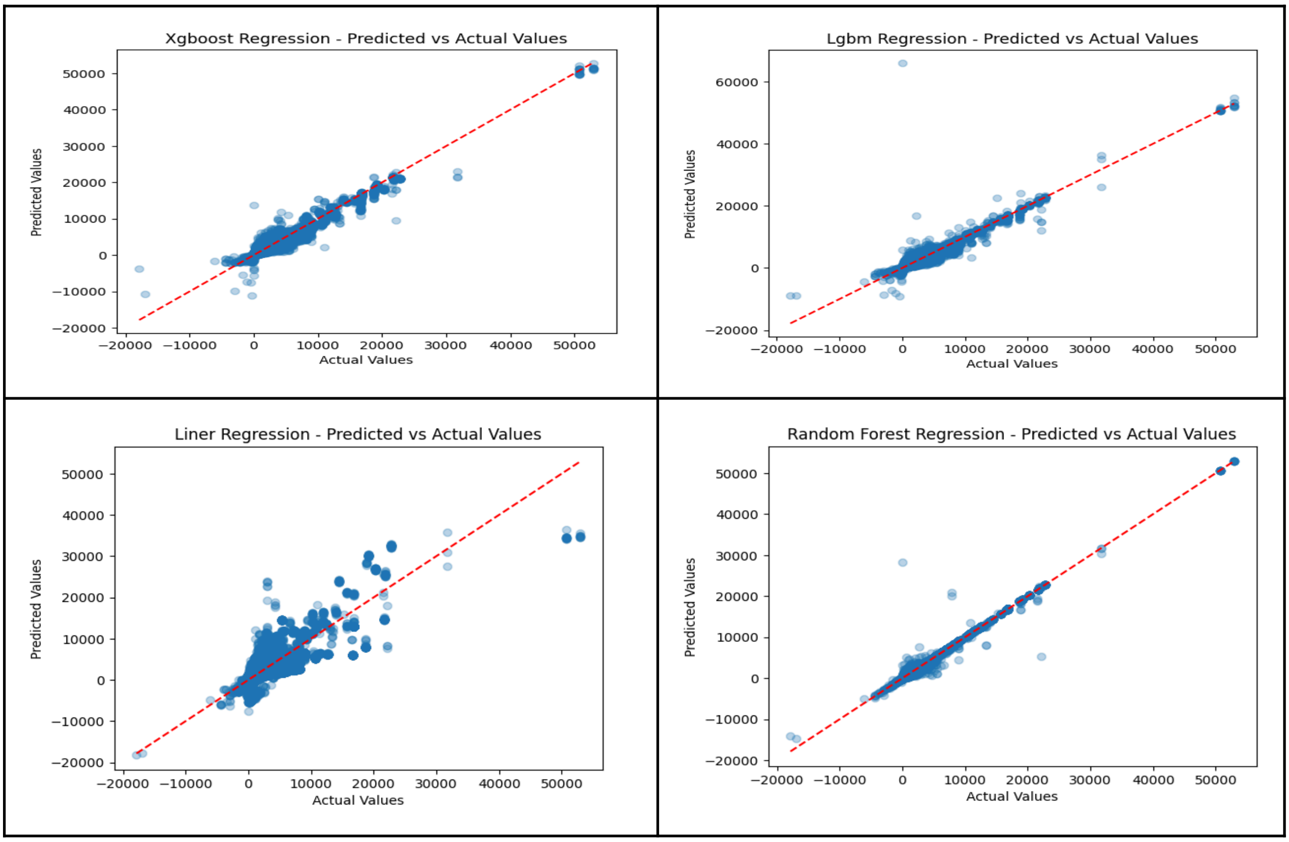

Figure-8

As part of Problem Statement 2, we developed four regression models ,i.e, Linear Regression, Random Forest Regressor, XGBoost, and LightGBM to predict product-level revenue using features such as quantity, unit price, and time of sale. Among these, the Random Forest Regressor delivered the highest accuracy, with an R^2 score of 0.995, making it the most reliable model for understanding what drives revenue.

This modeling approach directly supports the objective of Problem Statement 2, which is to identify the most influential factors behind revenue generation and leverage them for better pricing, promotion, and product prioritization strategies. By using machine learning, we not only predicted revenue with high precision but also gained insights into which variables matter most helping the business make informed decisions based on data rather than assumptions.

5.6) Feature Importance

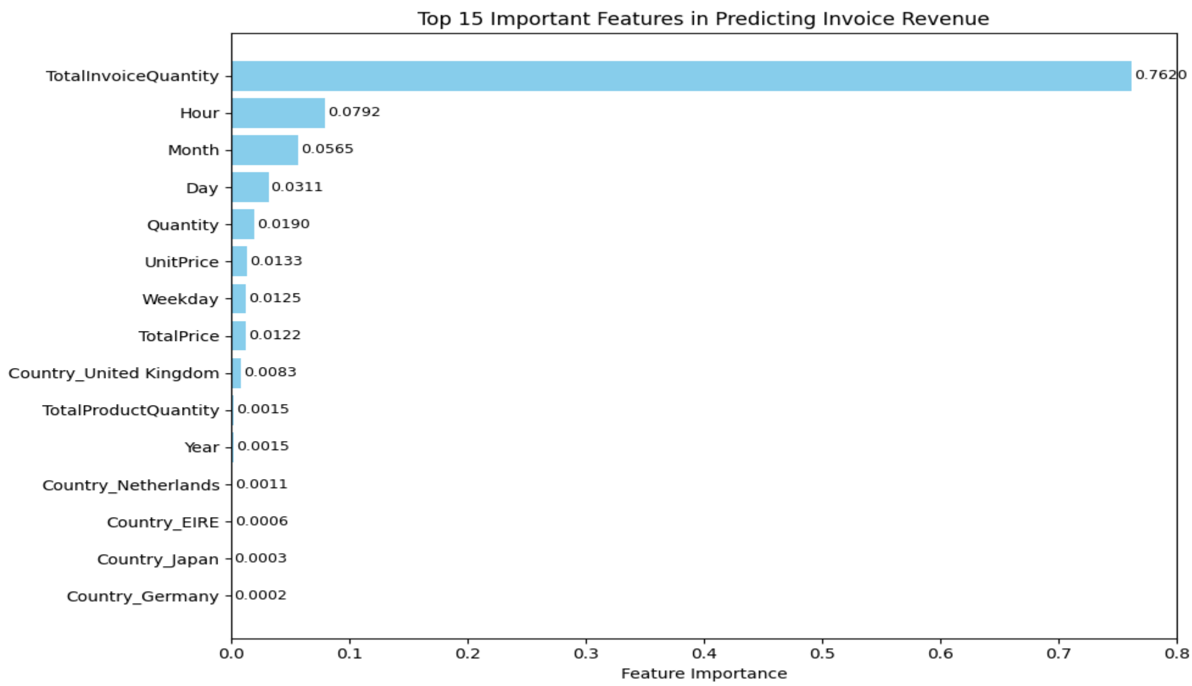

Figure-9

Key Findings from Feature Importance:

- TotalInvoiceQuantity was the most impactful feature, confirming that total volume sold is the biggest driver of revenue.
- Time-based variables like Hour, Month, and Day showed moderate influence, suggesting temporal patterns in buying behavior.
- Quantity and UnitPrice had meaningful but smaller contributions, consistent with our heatmap correlations.
- Customer geography (Country) had minimal influence, indicating that location has little effect on revenue in this dataset.

These insights help refine future revenue models and align business strategies with the most predictive features, enabling smarter pricing and inventory decisions.

5.7) Price Simulation

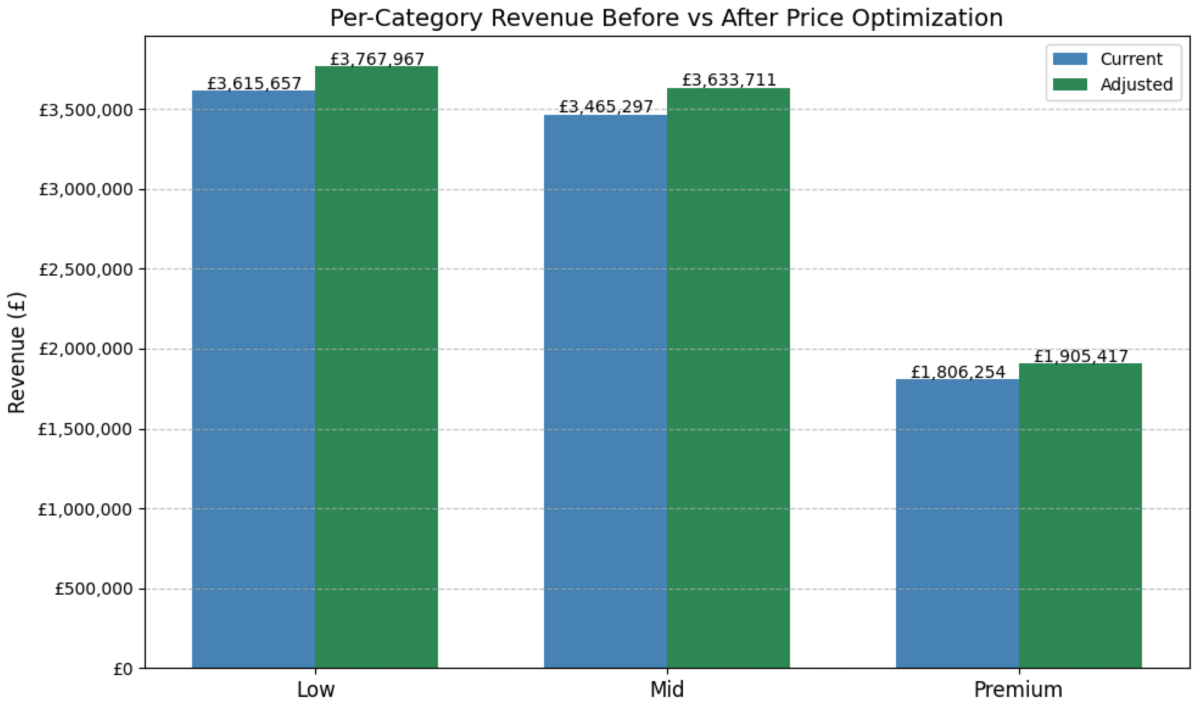

Figure-10

Now to optimize profitability without compromising sales volume, products were categorized into three price groups based on their unit prices: Low, Mid, and Premium. Each group received a tailored pricing strategy, which accounted for its estimated price elasticity derived from the earlier correlation analysis. The idea was to simulate how price increases alongside an expected proportional decline in demand would impact overall revenue and the price strategy was applied to conclude problem statement-2.

Premium Products

- 🔺 Price Increased by: 10%
- 🔻 Expected Quantity Drop: 4.1%

Mid-Priced Products

- 🔺 Price Increased by: 7%
- 🔻 Expected Quantity Drop: 2%

Low-Priced Products

- 🔺 Price Increased by: 5%
- 🔻 Expected Quantity Drop: 0.75%

This simulation helped evaluate how minor, controlled increases in unit price when aligned with demand sensitivity could improve revenue outcomes.

Revenue Comparison (as shown in figure-10)

Low-Price Category:

- Original Revenue: ~£2.78M
- Adjusted Revenue: ~£2.91M
- 🔼 Net Gain: £130K+

Mid-Price Category

- Original Revenue: ~£4.51M
- Adjusted Revenue: ~£4.79M
- 🔼 Net Gain: £280K+

Premium Category

- Original Revenue: ~£1.59M
- Adjusted Revenue: ~£1.68M
- 🔼 Net Gain: £90K+

The results of our price optimization strategy turned out to be quite promising. Starting with a total revenue of £8,887,209 , we simulated category-wise price increases across low, mid, and premium products, carefully accounting for expected drops in quantity sold. After these adjustments, the simulated revenue rose to £9,307,095, giving us an overall gain of about £419,886, or 4.72% increase. What makes this especially encouraging is that the boost didn't come from adding more products or increasing stock it came purely from smarter pricing. By tweaking prices based on how sensitive each category is to change, we were able to unlock more value from existing sales. This shows that with a bit of strategy, businesses can grow revenue in a sustainable way without significantly raising operational effort.

6 Interpretation of Results and Recommendations

The results derived from the project offer a comprehensive understanding of customer behavior, product performance, inventory dynamics, and pricing potential. By using a structured, datadriven approach, we extracted meaningful insights that directly support both problem statements inventory demand forecasting and revenue optimization. Below is a synthesis of key findings and corresponding actionable recommendations.

Insight 1: Revenue is heavily concentrated among VIP customers
Our segmentation revealed that the top 10% of customers (VIPs) contribute over 61% of total revenue, despite being a small portion of the customer base. These customers also consistently purchase specific high-revenue products.
Recommendation:

- Focus on the purchase preferences of VIP and Premium customers for inventory forecasting.

- Prioritize their most purchased items in inventory planning to avoid stockouts and ensure satisfaction.
- Introduce loyalty programs or early access promotions or discount/coupons to retain this high-value segment.

Impact: Ensures continuity of high-revenue streams while enhancing customer satisfaction and loyalty.

Insight 2: Understocking led to significant revenue loss in top products

Forecasting with ARIMA and moving averages showed that products like 85123A, 85099B, and 22423 were consistently understocked. Adjusting inventory to forecasted levels would have prevented losses of over £35,847.62 monthly, leading to a 4.84% increase in monthly revenue.

Recommendation:

- Implement demand forecasting for high-impact SKUs and adjust inventory levels accordingly.
- Regularly review stock coverage for fast-moving items linked to VIP and Premium customers.

Impact: Prevents stockouts, increases order fulfillment rates, and recovers lost sales.

Insight 3: Use Demand-Sensitive Pricing Strategies

Our regression and price elasticity analysis showed that quantity had a strong positive correlation with revenue, while unit price had a slight negative correlation. Price optimization simulations by product category demonstrated that modest price increases, when aligned with demand elasticity, can drive revenue without significantly impacting quantity sold.

Recommendation:

- Apply tiered pricing strategies: 10% for premium, 7% for mid-tier, and 5% for low-priced products.
- Monitor demand trends post-adjustment to refine these strategies further.

Impact: Accurate revenue prediction improves inventory planning, campaign targeting, and investment decisions for top-performing SKUs.

4. Strategic Inventory Planning by Segment and Product

The analysis combining customer segmentation with top product performance revealed that different segments prefer distinct product sets. Overlaying this with inventory and revenue data enabled a more focused stocking strategy.

Recommendations:

- Align product stocking not only with historical demand but also with customer segment behavior.
- De-prioritize overstocked, low-impact items and allocate shelf space more efficiently.

Impact: Supports lean inventory management while maximizing revenue per stocked item.