

Analysis Report

1. Data Cleaning Steps

a. Handling Missing Values

- **Description:** Rows with missing Description values were removed.
- **CustomerID:** Missing values in CustomerID were filled with 'Unknown', and the column type was set to object.

b. Removing Duplicates

- Duplicates in the dataset were identified and removed to ensure data integrity.

c. Inconsistent Data

- Standardized the Description column by stripping whitespace and converting text to lowercase.
- Removed rows with negative or zero values in Quantity and UnitPrice to eliminate invalid data.

d. Derived Columns

- Created a Revenue column by multiplying Quantity and UnitPrice.
 - Converted InvoiceDate to datetime format and extracted the YearMonth for trend analysis.
 - Identified first purchase date for each customer to flag returning customers.
-

2. Key Findings from the Analysis

a. Top-Performing Products

- The top products by revenue include items such as White Hanging Heart T-Light Holder and Metal Lantern.

b. Top Regions

- The majority of sales come from the **United Kingdom**, followed by **Netherlands**, **EIRE**, **Germany**, and **France**.

c. Seasonal Trends

- Sales exhibit clear seasonal patterns with peaks around the holiday season (November and December).
- December 2011 sales appear lower, likely due to incomplete data.

d. Customer Segmentation

- High-value customers were identified based on revenue contribution.

- Retention analysis showed a significant portion of sales coming from returning customers.

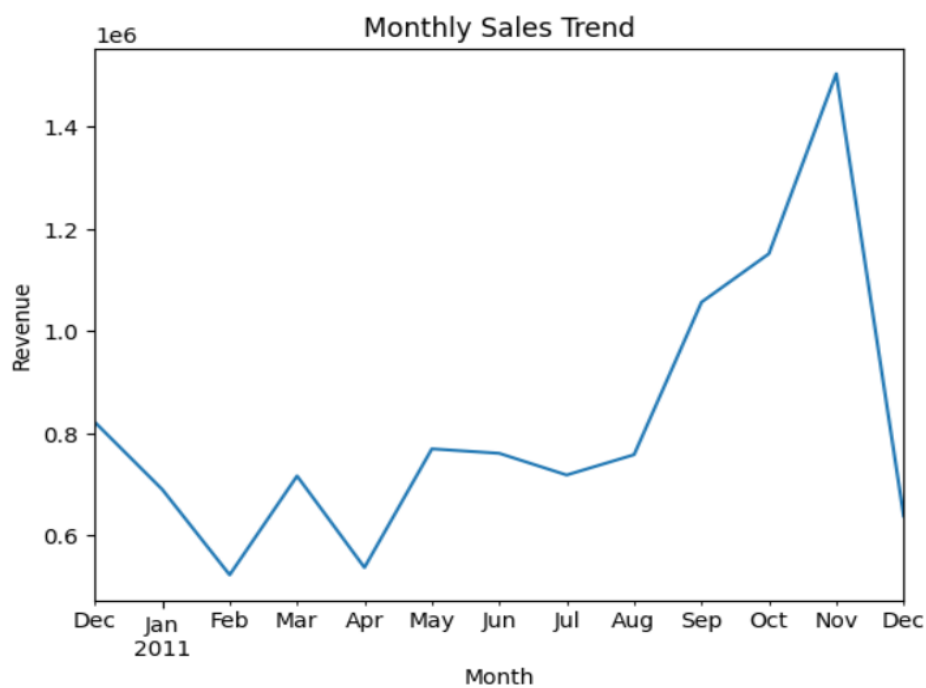
e. Anomalies

- Seasonal decomposition highlighted anomalies such as unexpected drops in sales during mid-year.
-

3. Visualizations

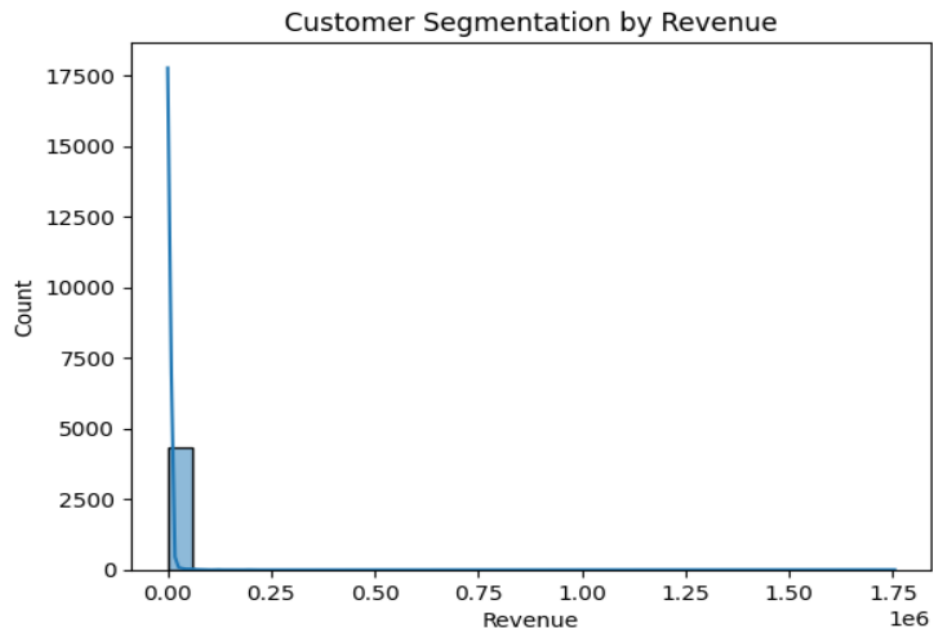
a. Monthly Sales Trends

Caption: Monthly sales trends show seasonal spikes, particularly around the year-end.



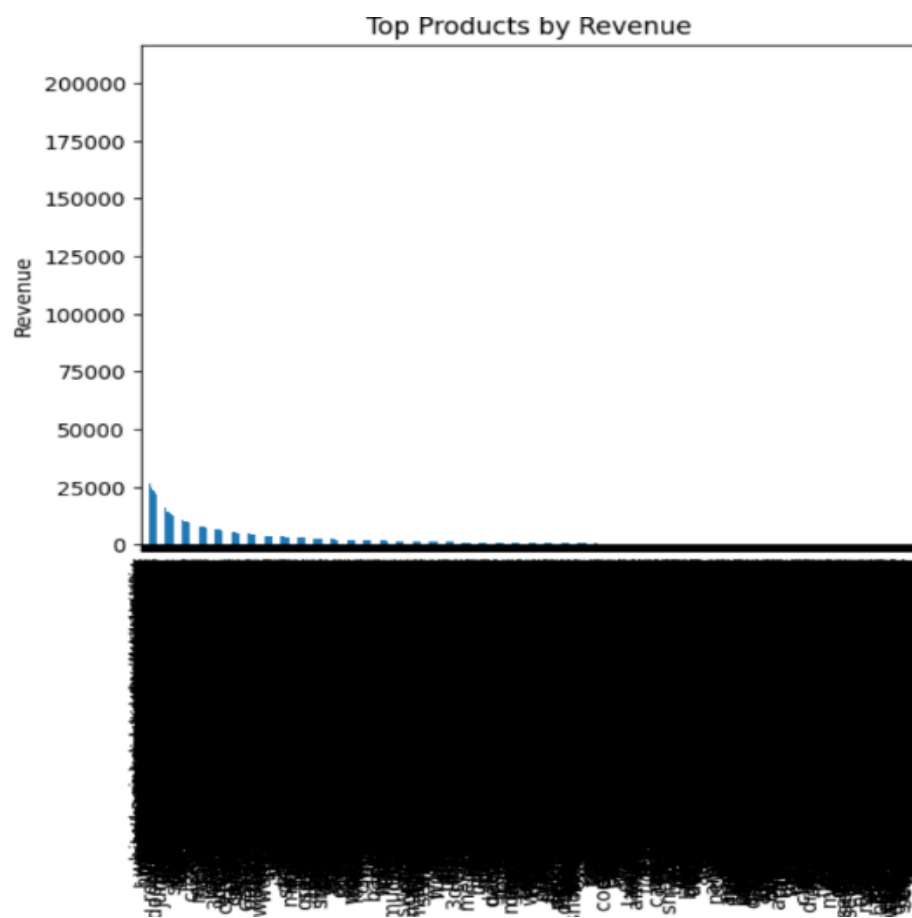
b. Customer Segmentation by Revenue

Caption: Distribution of customers by their total revenue contribution.



c. Top Products by Revenue

Caption: Top-selling products based on revenue.



4. Actionable Recommendations

a. Increase Sales During Low-Performing Months

Insight: Low-sales months are observed during mid-year.

- **Recommendations:**
 - Implement targeted promotional campaigns (e.g., discounts, bundles).
 - Boost marketing efforts in underperforming regions.

b. Retain High-Value Customers

Insight: High-value customers contribute significantly to revenue.

- **Recommendations:**
 - Launch loyalty programs with tiered rewards.
 - Use personalized marketing (e.g., product recommendations).

c. Product Recommendations Based on Behavior

Insight: Frequently purchased or paired products can reveal customer preferences.

- **Recommendations:**
 - Recommend complementary products through cross-selling strategies.
 - Suggest premium versions or add-ons through upselling at checkout.
-

5. Forecasting

ARIMA Model Forecast

- The ARIMA model (order = (1,1,1)) was applied to predict next month's revenue.
- **Forecasted value:** 144456.77796049142.

Evaluation

- **Mean Squared Error (MSE):** 47845149707.35.
-

6. Conclusion

The analysis highlights seasonal trends, top-performing products, and key customer segments. By implementing the provided recommendations, the business can enhance sales and improve customer retention.