Customer Transaction Analysis Report

Executive Summary

This report presents an analysis of a customer transaction dataset from an online retail business. The goal was to clean and explore the data, identify meaningful patterns, and build a basic predictive model to classify purchasing behavior. The analysis revealed several insights into customer trends, transaction behavior across countries and months, and relationships between key variables. A Random Forest model was used to predict purchases and achieved 100% accuracy on the available dataset. While the model's performance is strong, further validation is recommended to ensure its generalizability.

Key Charts and Insights

1. Distribution of Purchase Amount

A histogram of total purchase values showed that most transactions are of relatively low value, with a few high-value purchases. This indicates the presence of a small group of high-spending customers, which could be strategically important for targeted marketing.

2. Top 10 Countries by Number of Transactions

The United Kingdom accounts for the majority of transactions, followed by countries like Germany and France. These markets may warrant additional focus in terms of marketing, logistics, and customer service.

3. Monthly Sales Trends

Sales volume fluctuates across months, with certain periods showing clear peaks. These trends can be used for more accurate demand forecasting and seasonal marketing campaigns.

4. Correlation Heatmap

The heatmap showed strong correlations between variables like quantity and total price. Understanding these relationships can help improve pricing strategies and inventory planning.

5. Quantity vs. Total Price

A scatter plot analysis showed that purchases involving higher quantities typically result in higher total prices. This insight supports the idea of offering bulk discounts or bundling products to encourage larger orders.

Model Performance

- **Model Used:** Random Forest Classifier
- **Accuracy:** 100%
- **Evaluation Metrics:**
    - Precision, recall, and F1-score were all perfect for both classes.
    - The confusion matrix confirmed that there were no misclassifications.

Important Note:
While the model performed perfectly on the current dataset, this may indicate data leakage or an overly simplistic classification task. All records were labeled as a purchase, so redefining the prediction goal (for example, predicting high vs. low spenders) could provide more meaningful results.

Recommendations

- **Focus on High-Spending Customers:** A small group of customers contributes disproportionately to total revenue. These customers should be prioritized for loyalty programs or personalized offers.
- **Plan Around Seasonal Peaks:** Sales are higher during certain months, so inventory and marketing strategies should be aligned accordingly.
- **Prioritize Top-Performing Countries:** Countries with high transaction volumes, such as the UK and Germany, should receive focused business development efforts.
- **Leverage Bulk Purchase Behavior:** Since total price increases with quantity, offering bundle deals or discounts for higher quantities could encourage larger purchases.
- **Reconsider Target Variable for Prediction:** Because the current model predicts purchases that already occurred, consider creating a new target variable (such as likelihood to repurchase or predict high-value orders) for future modeling efforts.