Customer Behaviour Analysis and Profitability Optimization for DAT UK Ltd

Prepared by:

CLBS Consultancy LTD
(33142417)

Client:

DAT UK Ltd.

Report Date: 2025-01-19

Table of Contents

1. Introduction

DAT UK Ltd, an internet merchant specializing in giftware products, is facing challenges in understanding customer behaviour, optimizing product performance, and increasing profitability. In this context, advanced data analytics helps to find out key insights of overall business activity and make evidence based future strategies. This report presents a comprehensive analysis of DAT UK Ltd's customer transactions, using different data analytical tools such as RFM (Recency, Frequency, Monetary) analysis, customer segmentation, and predictive modelling. Through data preparation, exploratory data analysis (EDA), and modelling, this study identifies key trends in customer purchasing behaviour, seasonal revenue patterns, and product performance. Additionally, clustering techniques are applied to categorize customers into different segments according to their purchasing behaviour.

2. Objectives of the Analysis:

The objectives of this analysis are as follows:

- Analyse Customer Behaviour: Identify actionable patterns and trends in customer purchasing.
- Segment Customers: Use RFM-based analysis for tailored customer engagement.
- Identify Product Performance: Highlight top-performing and underperforming products.
- Develop Predictive Models: Create a model to forecast revenue based on key factors.
- Formulate Recommendations: Provide strategies to increase profitability.

3. Data Preparation

Data preparation remained a critical aspect of this analysis. The initial dataset comprised two CSV files, which were merged to create a unified dataset. The combined file contains 1,067,371 rows and 9 columns. Data preparation was performed using Python in Jupyter Notebook Environment, and python libraries such as Pandas, NumPy were used to manipulate the complex data.

Key steps in data preparation:

1. Removing Duplicates: A total of 34,217 duplicate rows were identified and removed.
2. Handling Missing Values: Missing values in each column were identified. Missing values in the column Price and Quantity were imputed with median values. As Customer ID column was with significant number of missing values, it was filled with ‘Unknown’. Similarly, Description column was also with high number of missing values, which was filled with ‘Not Known’.
3. Eliminating Anomalies: Rows with negative values in Quantity and Price were removed.
4. Data Transformation: The column InvoiceDate was split into Year, Month, Day, and Time columns and new column ‘Revenue’ was created by multiplying Quantity and Price.

The final dataset comprise 1010548 rows and 12 columns, ready for analysis.

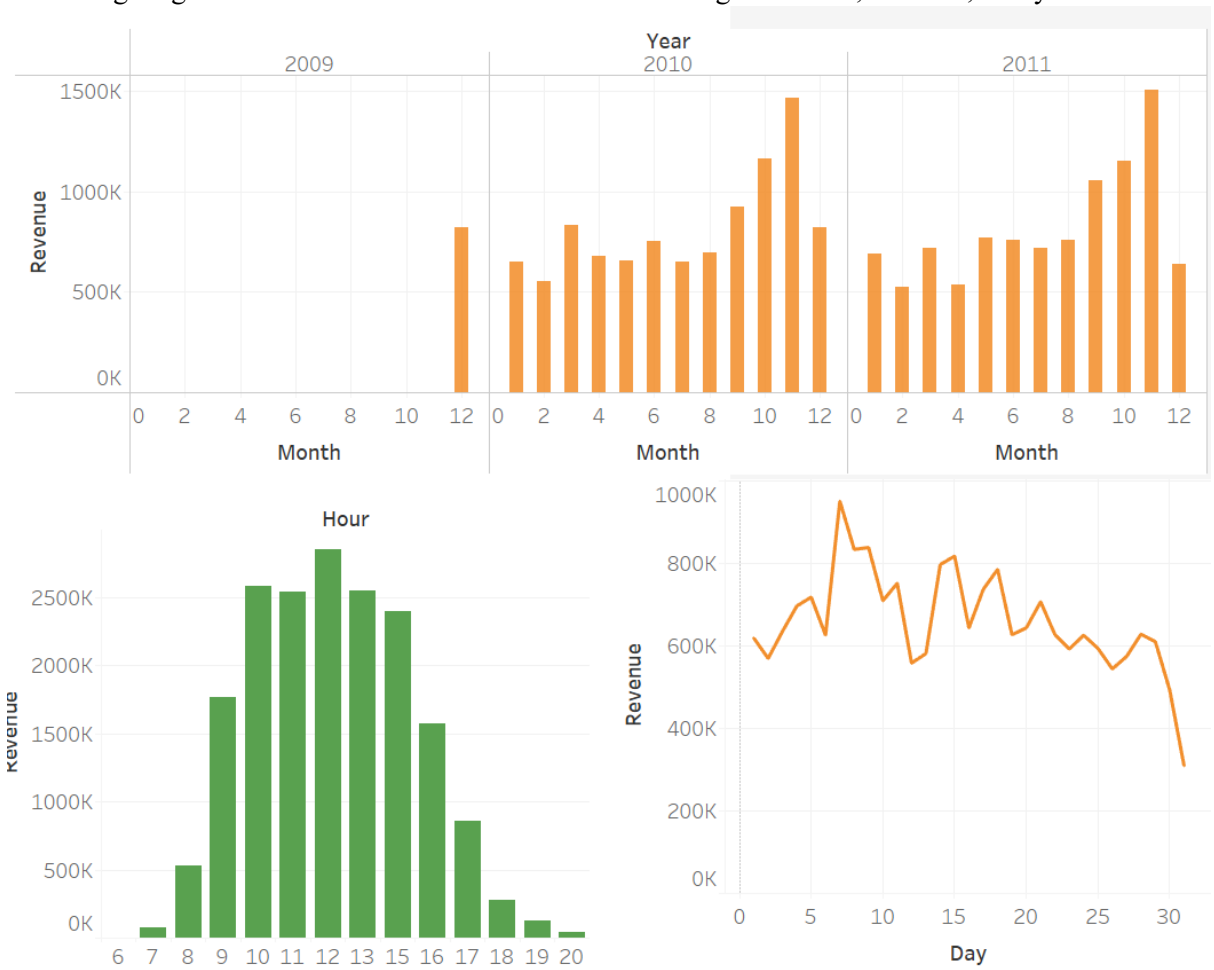
4. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was also performed using Python using Jupyter Notebook Environment, and libraries such as Matplotlib, Seaborn Were used for visual representation of the data. EDA provided valuable insights into DAT UK Ltd's customer and sales data. Key results of the EDA are discussed below.

4.1 Revenue Trends:

As revenue generation was the key factor in optimizing profitability, its trend was analysed. Following trend and patterns were recognized from Analysis.

Following figure shows the revenue trend according to Year, Month, Day and Time:



4.1.1 Seasonal trend of Revenue Generation:

Revenue shows significant seasonal trend with low level of revenue generation from December to August and starts to increase revenue generation from September, it becomes high in November in each year. Peak value of revenue is in November 2011 at £1503866.78 whereas low level of revenue is in February 2011 at £522545.560.

4.1.2 Hourly and Daily Patterns of Revenue generation:

Revenue generation varies significantly by hour, where beginning and end of the day with low level of revenue generation. This pattern shows that revenue generation gradually increases from morning hour and reaches peak at 12 noon, again starts to decline after that time. Overall, revenue generation remains

at significantly low level before 9 am and after 5 pm. Similarly, daily revenue generation pattern in a month was analysed. Last day of the month shows low level of revenue generation which gradually increases with the starting day of the month and reaches to the peak on 7th day of the month from where it starts to decrease gradually.

4.2 Country-Wise Analysis:

5882 customers with distinct customer ID are from 43 different countries. UK have dominant number of customers (above 90 % customers are from UK) followed by Germany and France. The UK contributes to the highest revenue, followed by EIRE and Netherlands. Surprisingly, revenue per customer data is different, EIRE remains in top followed by Singapore and Netherlands.

4.3 Distribution of Quantity Price and Revenue

4.3.1 Distribution of Quantity

Most of the Quantity values are concentrated near 0, with very few higher values (extreme right on the x-axis), indicating that most transactions involve small quantities, but a few involve very large quantities. This indicates that bulk purchases are rare but exist in the dataset.

4.3.2 Distribution of Price

Most of the Price values are very small (near 0), with a few much higher values (to the far right). The distribution is again highly skewed, which reflects the dataset is dominated by low-priced items, which likely drive the bulk of transactions and high-priced items are infrequent but may contribute significantly to revenue.

4.3.3 Distribution of Revenue

Like the previous plots, Revenue has a highly skewed distribution. Most revenues are small, the small revenue values correspond to low-quantity and low-priced transactions.

4.3.4 General Insights

All three distributions are heavily skewed to the right, meaning that a small number of extreme values exist in the dataset. Most transactions involve low quantities, low prices, and low revenues. The bulk purchases or high-priced items can disproportionately affect overall metrics like average revenue or price.

5. Customer Segmentation

Customer segmentation was performed using RFM (Recency, Frequency, Monetary) analysis combined with K-means clustering. Customers were grouped into meaningful different 4 segments based on RFM analysis. Characteristics of different segments are discussed below.

5.1 Cluster Insights

From the plot and cluster summary, the clusters can be divided as follows:

- Cluster 3 (Yellow Cluster): This cluster belongs to customer with Low recency (recent customers), high frequency, and high monetary value. This cluster includes the most valuable and loyal

customers. are likely wholesale buyers or corporate clients. Although his cluster contains only 4 customers, it contributes significant to revenue.

- Cluster 2 (Green Cluster): In this cluster, the customers have Moderate recency, Moderate frequency, and High monetary value. These customers have decent engagement, but lower monetary contributions compared to Cluster 3. In the dataset this customer contains 38 customers.
- Cluster 0 (Purple Cluster): In this cluster the customers have Moderate recency, Low frequency, and Low monetary value. This is the largest cluster which consists of 3840 customers. These customers who provide steady but not exceptional revenue. They are moderately active and represent a significant growth opportunity. These customers are new or occasional buyers.
- Cluster 1 (Blue Cluster): In this cluster customers have high recency (very inactive customers), low frequency, and low monetary value. These customers are dormant or churned.

5.2 Cluster Identification

Based on the above clustering customers segments can be identified as follows.

High-Value Loyal Customers (Cluster 3): Frequent and recent buyers contributing the highest revenue.

Occasional Spenders (Cluster 2) : Moderate recency, frequency, and spending

Potential High Value Customers (Cluster 0): Recent buyers with moderate frequency and spending and represents highest number of customers.

Dormant Customers (Cluster 1): High recency (last purchase made a long time ago), low frequency, and spending

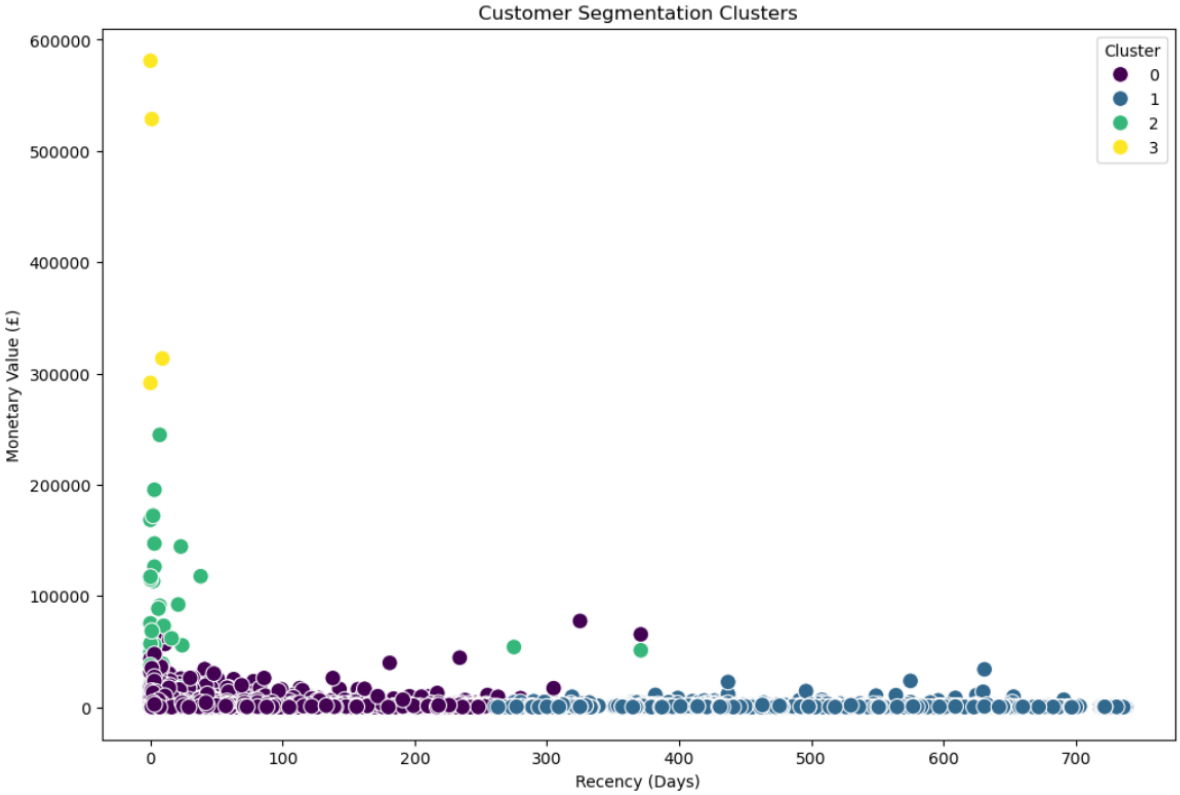
5.3 Key Observation:

A small group of customers contributes significantly to the revenue (high monetary value), indicating that a minority of customers drive most of the profit. A large group of customers shows high recency and minimal spending, which might indicate disengagement or lost customers. There are significant number of recent customers who could be converted into loyal, high-value customers with the right strategies.

5.4 Recommended Strategies for each Cluster:

- Retain High-Value Customers (Cluster 3): Company should implement loyalty programs, offer exclusive discounts, and prioritize customer service for this group.
- Engage Occasional Spenders (Cluster 2): Company should use targeted promotions and personalized recommendations to increase their purchasing frequency and spending.
- Nurture Potential High Value Customers (Cluster 0): Focus on onboarding strategies, product education, and incentives to encourage repeat purchases. As highest number of customers are in this cluster, it may be the major source of revenue generation in the future.
- Reactivate Dormant Customers (Cluster 1): Company should Send time-sensitive reactivation campaigns with attractive discounts or offers.

Graphical representation of customer segmentation is shown below:



6. Predictive Modelling:

Hyperparameter-tuned XGBoost Regressor was used for the predictive modelling in this analysis. Results from predictive modelling are below.

6.1 Key performance metrics:

RMSE (Root Mean Squared Error): RMSE measures the average deviation between predicted and actual revenue values. Relatively lower RMSE (44.64) indicates better model performance.

R^2 Score: The R^2 score explains the proportion of variance in revenue accounted for by the model. A value of 0.526 means the model explains 52.6% of the variance in revenue.

6.2 Feature Importance

The feature importance graph highlights the relative contributions of the features to the model's predictions:

Quantity (73.80% Importance): This is the most critical feature driving revenue. It reflects that the size of purchases (quantities sold) is the primary determinant of revenue.

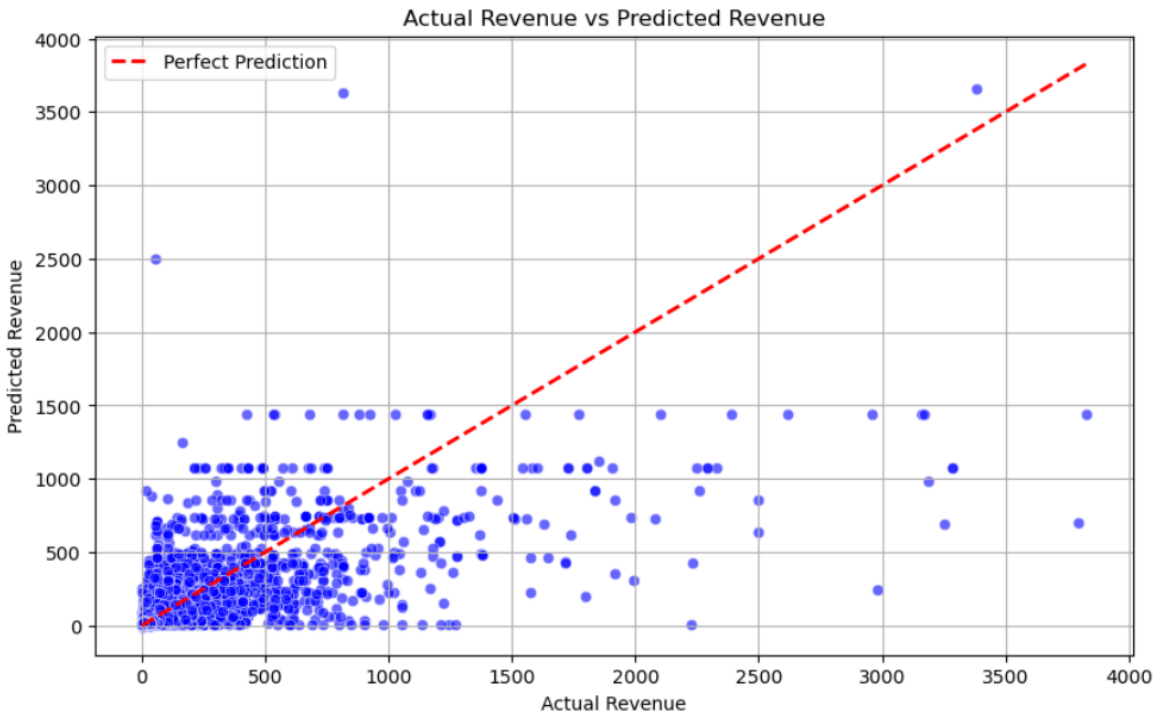
Recency (25.34% Importance): This measures how recently a customer made a purchase. It significantly influences revenue, showing that recent buyers are more likely to contribute to revenue.

Frequency (0.85% Importance): This has minimal importance, suggesting that repeat purchases (frequency) do not significantly influence revenue compared to transaction size or recency.

6.3 Key findings:

The tuned XGBoost model shows moderate predictive capabilities with R Squired 0.52 and RMSE of 44.64. The high level of feature importance of Quantity and Recency reflects the need to focus on increasing transaction sizes and most recent buyers to increase revenue. Although Frequency has a smaller effect, it remains important feature to increase profitability. Moderate level of R Squired indicates that nearly half of the variance in revenue is still unexplained.

Scatter plots of actual vs. predicted revenue to evaluate the model's predictive accuracy is given below.



7. Recommendations

Based on the analysis, the following strategies are recommended:

- Optimize pricing strategies: Company should adjust prices for high-demand products, such as *World war 2 gliders asstd designs* and *White hanging heart t-light holder,* to maximize profitability while maintaining competitiveness.
- Use of customer segmentation: In this report customers are segmented in 4 group according to the purchasing behaviour. Company should focus on the strategies mentioned in section 5.4 (*Recommended Strategies for each cluster)* of this report. However, retaining high value customer and focusing high potential customers are key strategies.
- Use of seasonal trends: Availability of stock should be maintained in the peak seasonal period (e.g., September, October, November).
- Follow hourly and daily pattern of revenue generation: It needs to be insured that online transaction system and manpower are up to date at high revenue generation day and time. Maintenance and modification in the system needs to be done using pattern of revenue generation, probably at the end of the month or within 2 or 3 days of beginning of the month.
- Expansion of business to different region: About 90% of customers are concentrated in UK. To expand profitability, it is recommended to expand business in the countries with high value customer such as AIRE and Netherland

- Increase revenue through bulk sales: Company should focus on bulk purchases offering attractive discount and other marketing campaign.

8. Implementation Plan:

Implementation plan for execution of recommendation in different period is given below.

8.1 Short-Term Plan (0-3 months):

- Launch promotional and marketing campaigns for top-performing products.
- Design and roll out loyalty and reactivation programs.
- Make comprehensive implementation plan to execute the recommendations of this report.

8.2 Medium-Term Plan (3-12 months):

- Develop marketing strategies focusing on potential high-value customers.
- Optimize pricing strategies based on product performance and demand.
- Start marketing campaigns in the countries outside UK, where revenue customers is high.

8.3 Long-Term Plan (12+ months):

- Regularly monitor customer segments and review strategies using data analytics tools.
- Using updated data continuously improve the predictive model to enhance forecasting.

9. Conclusion

This report provides a detailed analysis of DAT UK Ltd's customer data to understand customer behaviour and increase profitability. Analysis shows that bulk purchases and recent customer engagement are the main drivers of revenue. Seasonal and hourly patterns show when revenue is highest, emphasizing the need for better inventory management and operational planning. Additionally, Customer segmentation identified four distinct groups of customers, each requiring specific strategies. Predictive modelling showed moderate accuracy, with Quantity and Recency being the most important features in increase revenue. By implementing the recommended strategies, including customer-focused campaigns, optimized pricing, and regional expansion, DAT UK Ltd can increase profitability significantly.