Stakeholder Report: Customer Retention & Revenue Forecasting Insights

Executive Summary

This report outlines our recent analysis of customer behaviour and predicted sales performance using Cohort analysis. The study is aimed at assessing customer retention patterns and evaluating how they influence our overall objective—**maximizing profitability through increased and predictable sales revenue**.

1. Objectives

- **Customer Retention Analysis**: Understand how frequently and recently customers make purchases, and the average revenue per customer.
- **Sales Revenue Forecasting**: Predict future purchasing behaviour and revenue using robust empirical approach.
- **Customer Segmentation**: Identify customer groups based on their buying behaviour to enhance marketing and retention strategies.

2. Methodology Overview

Using our transaction data[1], the following steps were performed:

- **Data Cleaning & Preparation**: We Removed invalid, Cancelled and or incomplete records; computed the Total revenue[2].
- **RFM Analysis**: Derived key customer metrics — Recency, Frequency, Monetary Value, Average Quantity, Average Unit Price, and Reorder Days.

- **Predictive Modelling**:

  To better understand customer behaviour and forecast revenue, we trained two predictive models—**Random Forest** and **XGBoost**—using key customer features like Recency, Frequency, and Monetary value (RFM), along with average quantity, unit price, and reorder patterns (see table table 1).

  Table 1: Evaluation Metrics

Model	MAE (Mean Absolute Error)	RMSE (Root Mean Squared Error)
Random Forest	2,103.83	61,537.00
X**GBoost**	1,693.91	5,606.55

[1] See Appendix

[2] `TotalPrice = Quantity × UnitPrice`

The **Mean Absolute Error (MAE)** tells us the average difference between the actual and predicted revenue values per customer. XGBoost's lower MAE (1,693.91) compared to Random Forest (2,103.83) indicates more accurate individual revenue predictions. **Root Mean Squared Error (RMSE)** penalizes larger errors more heavily. XGBoost again outperforms with a much lower RMSE of 5,606.55 vs. Random Forest's 61,537.00, showing better overall prediction stability and fewer large deviations. Based on both MAE and RMSE, **XGBoost** provides more reliable forecasts of our customer monetary value and is the preferred model for estimating our future revenue contributions and guiding marketing or retention efforts.

Figure 1: Customer segmenting based on RFM

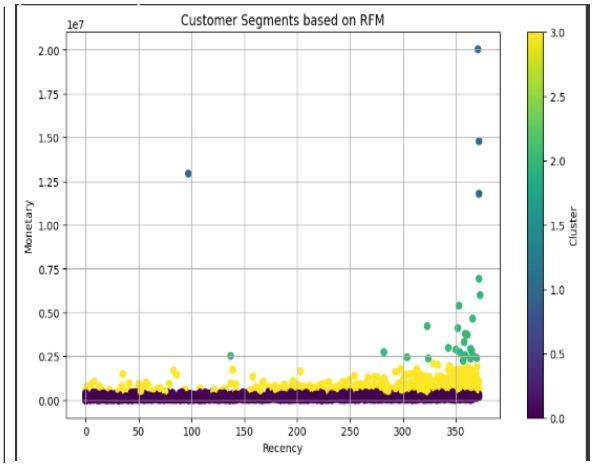

Figure 2: Customer segmenting based on RFM - Monetary

BG/NBD and **Gamma-Gamma** models were used to estimate future purchases and expected average revenue.

- **Customer Segmentation**: We Clustered customers into 4 segments using KMeans based on RFM features (see figure 1 and 2). **Cluster 0**: High frequency, low monetary — Frequent shoppers with low spend. **Cluster 1**: Low frequency, high monetary — Occasional big spenders. **Cluster 2**: High recency — at risk of churn, needs re-engagement. **Cluster 3**: High monetary and frequency — Ideal for VIP treatment.

- **Forecasting**: We Projected expected revenue over the next 30 days and exported top-performing customers for targeted campaigns.

3. Key Insights

Customer Behaviour

- A significant portion of customers exhibit **infrequent but high-value transactions**.
- Average reorder time varies widely, suggesting opportunities to encourage more consistent engagement.

Revenue Forecasting

Predicted 30-Day Revenue: based on our customer behaviour, we retrospectively expect total revenue of 102,896,379.42. Realistically, we have a Reliable expectation of incoming cash flow between 92,967,857.39 and 112,824,901.45 (see table 2 snippet below).

Top 10% customers expected to generate **the majority of revenue**, validating the 80/20 Pareto principle. These high-value customers were identified and exported for potential **exclusive deals or loyalty programs**.

Table 2: Top 10% Spenders

Customer	Monetary value	Predicted purchases	Predicted monetary	Expected revenue
14096.0	807006.890625	4.359137967168007	824859.4121308203	3595675.980995341
17841.0	179060.50837837838	8.700810908633688	179620.8937225191	1562847.4315194266
13521.0	376872.935	3.2124137855272155	455791.423834139	1464190.653249866
14911.0	112365.3610687023	10.266022175644162	112663.194709943	1156602.8552711909
12748.0	104278.21353982302	8.85733203533471	104598.77642066631	926466.0932475807
16360.0	324025.302	1.57954097732107	348136.8245591619	549896.3801056325
14606.0	77753.84113636364	6.900817951571337	78061.04499874469	538685.0606457554
15993.0	212881.6175	2.3065111483568663	233058.49451949674	537552.0158284869
15427.0	551410.31	0.6160983923583382	843508.6326877249	519684.3125392873
14456.0	339034.39857142855	1.4134908872663174	356679.517793609	504163.24807581067

From a different perspective, we could focus on spenders who are likely to make a purchase within the next 30 days. Table 3 below provides a snippet of their budget from past behaviour. Therefore, within the next 30days, they are expected to spend 6,418,255.07 the customer behaviour. The major catch is the fact that we have a churn rate of 6.30%. Therefore adopting a conversion strategy applied on eligible customers, we expect new revenue of 10,756,247.5 if 93.70% of slower customers convert. This translates to a change in revenue to the tone of 4,337,992.43. these expectations offers benchmark support for inventory planning and marketing budgets.

Table 3: 30 - Days Expected Revenue snippet

Customer	Average Quantity	Average Unit Price	Average Reorder Days	Cluster	Projected Revenue
15069.0	51	53.23	0.0	0	2759.67
15135.0	37	75.09	0.0	0	2836.85
15130.0	53	43.27	0.0	0	2311.20
15127.0	43	47.82	0.0	0	2077.58
15185.0	56	60.23	0.0	0	3393.19
15180.0	53	66.84	0.0	0	3564.97
15178.0	41	35.24	0.0	0	1456.58
16611.0	47	38.03	0.0	0	1787.50
15083.0	40	60.37	0.0	0	2451.02
15076.0	47	47.06	0.0	0	2233.25

4. Strategic Recommendations

- **Retention Focus**: Launch targeted reactivation campaigns for customers in Cluster 2 using time-sensitive offers.
- **Revenue Growth**: Leverage insights from Cluster 1 and 3 for **loyalty programs and personalized upselling**.
- **Inventory Planning**: Align procurement with predicted revenue inflows to reduce overstock and out-of-stock situations.
- **Marketing ROI**: Use the top 10% list to tailor **high-ROI campaigns** with minimal waste.

5. Conclusion

The integration of statistical forecasting, customer segmentation and predictive modelling provides a comprehensive view of how customer behaviour drives revenue. This enables **data-driven decisions** to maximize profit and foster long-term customer relationships.