CHURN ANALYSIS AND PREDICTION FOR FOODCORP

MACHINE LEARNING AND PREDICTIVE ANALYTICS

20703562
SPRING 2024-2025
BUSI4373 UNUK

Executive Summary

This project delivers a scalable, data-driven churn prediction system tailored to FoodCorp's weekly customer retention strategy. The primary goal was to proactively identify currently active customers at high risk of churn—empowering marketing teams to prioritize timely outreach and optimize resource allocation. To align with FoodCorp's operational objectives, churn was defined conditionally: customers were labelled as churned only if they were active in the past 30 days but made no purchases in the following 30-day period. This definition ensures the model targets customers while they are still recoverable, enabling interventions before disengagement becomes irreversible.

This report presents a complete end-to-end churn prediction framework for FoodCorps, grounded in behavioural analytics and machine learning. The objective was to proactively identify customers at risk of churning, enabling timely and targeted retention strategies. Churn was operationally defined in alignment with FoodCorp's business logic: customers active in the past 30 days but inactive in the subsequent 30-day window were labelled as churned.(Figure 1). This conditional definition improved model relevance by focusing on recoverable customers rather than those long inactive.

Using a 150-day rolling behavioral history segmented into five 30-day input windows, the model forecasts churn in the next month. Multiple machine learning algorithms were evaluated using temporally consistent validation. Random Forest emerged as the top performer, achieving an F1 score of 76% and an AUC of 87%, effectively balancing precision (limiting false positives) and recall (capturing true churners) . Feature importance analysis, SHAP and permutation methods revealed that churn is most strongly driven by behavioral irregularities such as visit gaps, frequency variance, and declining engagement, rather than spending volume. As shown in Figure 2, which displays churn counts, these insights support early churn detection—often before revenue loss occurs.

Key insights from SHAP analysis revealed that churn is driven more by irregular visit patterns (mean_gap, frequency_variance, average_visit) than by monetary metrics like spend — allowing early detection of disengagement before revenue drop-offs occur. The model also maintained an effective balance between recall and precision, avoiding excessive false positives while capturing most at-risk customers.

To support interpretability and deployment, a hybrid feature schema was adopted, combining lagged and aggregated metrics that generalise across weekly cycles. The entire pipeline is modular and fully automatable, enabling seamless weekly re-training and scoring across FoodCorp's store network with minimal manual intervention. Therefore, this project delivers a deployment-ready churn prediction solution that is technically robust, behaviorally insightful, and aligned with FoodCorps's real-world retention operations — equipping the business with the tools to proactively protect customer value at scale.

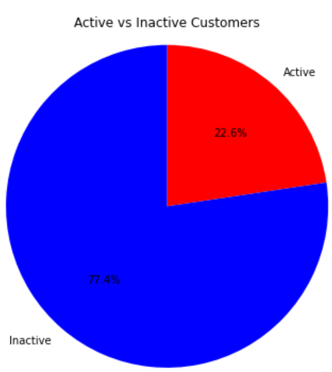


***Figure 1**. Distribution of Active vs Inactive Customers*

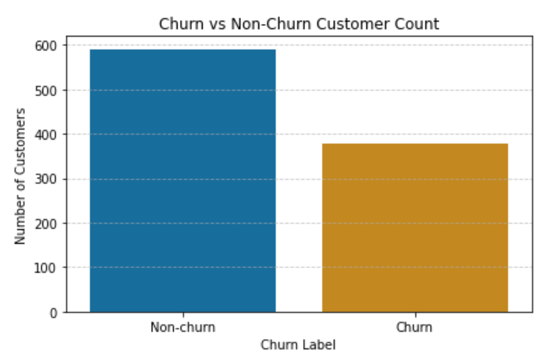


***Figure 2.** Churn vs Non-Churn Customer Counts Predicted*

Current Churn Levels:

To establish a suitable churn definition, this project begins by reviewing the consultancy report commissioned by FoodCorps. The consulting firm proposed using a global churn definition based on customer inactivity, where a customer is considered churned if they have not returned after a fixed number of days. To support this, the report presented two key analyses: (1) the distribution of median days between visits for each customer and (2) the proportion of customers that would be classified as churned under various inactivity thresholds. These visualizations help determine a threshold that reflects meaningful disengagement without prematurely flagging regular but infrequent shoppers.

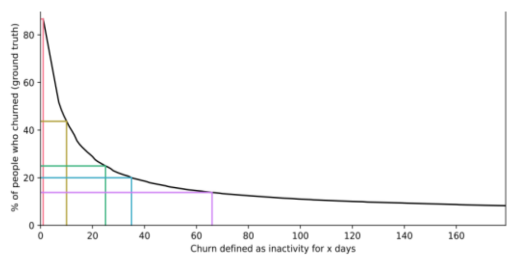


***Figure 3**. Target Class Proportions vs. Churn Definition Threshold*

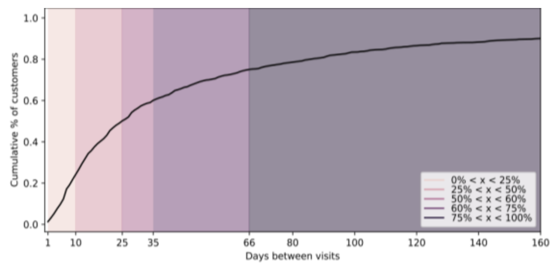


***Figure 4.** Cumulative Distribution of Days Between Customer Visits*

Figure 3 shows the cumulative distribution of customers by median days between purchases. About 50% return within 25–30 days, and 75% within 60–66 days. The curve flattens beyond 60 days, indicating most customers return quickly, with only a minority showing long gaps. This supports using a 30-day threshold to detect deviations from typical behavior while allowing timely intervention. Figure 4 displays the proportion of customers classified as churned at various inactivity thresholds. The curve drops steeply from 0 to 30 days, then flattens, showing most churnable customers are captured by day 30. Beyond this point, precision gains are minor, and the churn class shrinks considerably—potentially harming model training.

In alignment with FoodCorps global churn strategy, where churn is defined as a customer exceeding a fixed period of inactivity (β days) since their last purchase, this project operationalized churn labelling to reflect both that conceptual rule and the company's stated requirement to only consider currently active customers. However, the original definition from the consulting firm—classifying churn solely based on prolonged inactivity proposed a key limitation for predictive modelling: it included many customers who were already dormant and likely beyond recovery, making real-time interventions less effective.

To overcome this, a refined, conditional churn definition was adopted: customers were labelled as churned only if they made at least one purchase in the 30-day (β) input window prior to the reference date but none in the 30-day output window following it. This implementation ensures temporal practical relevance, allowing the model to focus on predicting imminent churn among recently active users, a critical shift that enhances the model's operational utility for FoodCorps weekly intervention framework. Using the conditional definition, 967 active customers were identified at the prediction point, of which 406 were labelled as churned, resulting in a churn rate of 41.99%.

Technical Report:

This section details the churn prediction system developed for FoodCorps, designed to run weekly and identify active customers at risk of disengagement. The pipeline combines SQL-based feature engineering with Python-based modelling and evaluation, using temporal validation to ensure reliability.

Feature Engineering

The churn prediction model is designed for weekly deployment, which requires a temporally consistent structure for both input and output windows. As shown in Figure 5, customer activity was segmented into five non-overlapping 30-day input windows (F5 to F1), covering a total of 150 days prior to the reference date. Each window captures lagged behavioral signals, allowing the model to learn both recent and historical trends. Churn labels were generated using a subsequent 30-day output window, with customers flagged as churned if they were active during the final input window (F1) but had no transactions during the output period. This setup reflects a real-world scenario in which predictions are made for engaged customers and intervention is still feasible.

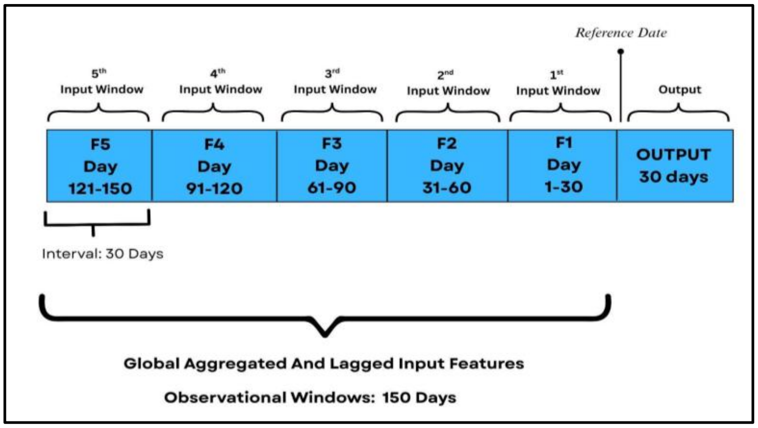


***Figure 5**. Windowing Strategy used for temporal feature construction*

A comprehensive feature set was engineered to capture key aspects of customer behavior, as detailed in Table 1 by functional categories. Temporal variables for each input window included total spend (**f1_spend to f5_spend**), quantity (**f1_qty to f5_qty**), and visit frequency (**f1_frequency to f5_frequency**). Aggregated indicators—such as **average_spend, unit_cost, and basket_value**—reflected customer value and consistency. Additional attributes included store location (**store_code_0 to store_code_3**), product diversity (**num_pro**), and purchase regularity (**mean_gap**). Outcome variables (is_active, churned) and transaction timing (max_purchase_day, min_purchase_day_output) supported labeling. This layered feature design enabled the model to learn temporal behavior patterns while supporting profiling.

GLOBAL AGGREGATE INPUT	
customer_id	Unique customer identifier
ref_day	Reference date for building the feature snapshot
mean_gap	Average number of days between a customer's purchases (recency/regularity measure)
Store_code_0 - Store_code_3	One-hot encoded store location (where purchases occurred)
active_day	Number of distinct days the customer shopped in the last 5 periods
basket_value	Average basket value (total value per shopping day) over last 5 periods
unit_cost	Cost per item = total spend / total quantity over last 5 periods
num_pro	Number of unique products purchased in last 5 periods
average_visit	Avg number of days per period the customer made a purchase (visits per period)
average_spend	Avg total spend per period (sum of spend over 5 periods divided by 5)

OUTPUT FEATURES	
max_purchase_day_f1_window	Most recent purchase day before ref_day in the last window
min_purchase_day_output_window	First purchase day in the output window (used to detect churn)
is_active	Flag: did the customer make purchases before ref_day?
Churned	Label: customer made purchases before ref_day but none after, in output window

WINDOW AGGREGATE INPUT	
f1_qty– f5_qty	Total quantity bought in each window
f1_spend – f5_spend	Total spend in each of the last 5 windows
f1_frequency– f5_frequency	Count of unique shopping days in each window

Table 1. Features generation approach

Feature selection

Following the initial feature generation a two-stage feature selection process was carried out on 929-sample training set of active customers identified,to reduce redundancy, enhance interpretability, and improve model performance. The first stage involved **analyzing correlations** between features and the output, churn label.

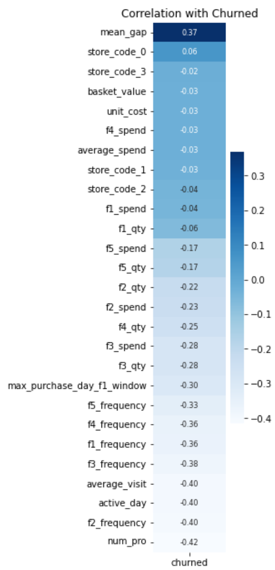


Figure 6. Correlation between input and output feature

As shown in Figure 6, several of the raw temporal features, particularly **frequency-based** variables such as **f1_frequency, f2_frequency, and f3_frequency**, showed strong negative correlations with churn, indicating that more frequent recent visits were predictive of retention. In contrast, **mean_gap**, representing average time between visits, had a strong positive correlation (0.37), suggesting that irregular or infrequent shoppers were more likely to churn. This analysis also revealed high multicollinearity across sequential lag features (f1_ to f5_), supporting the need for dimensionality reduction.

To address redundancy and improve interpretability, **several raw lag features were aggregated** into higher-level behavioural indicators. These included **average_qty** to summarize purchase volume, **spend_variance and frequency_variance** to capture behavioural fluctuations, and **days_since_last_purchase** to reflect short-term recency. Redundant features such as the full set of f*_spend, f*_qty, along with one-hot encoded store indicators were removed to streamline the feature space.

In the second stage, **multiple feature importance methods** were applied, including **Mutual information (filter), Random Forest importances (embedded), and permutation-based** methods across holdout, train/test split, and 5-fold cross-validation.

As shown in Figure 3, features such as m**ean_gap, frequency_variance, average_visit, and num_**pro consistently ranked highest—capturing behavioural regularity, diversity, and early disengagement. The final modelling set included mean_gap and days_since_last_purchase (recency); frequency_variance and spend_variance (stability); lagged Frequencies, average_visit, average_qty, and average_spend (engagement); and num_pro (breadth). This refined set balanced predictive power, interpretability, and operational relevance (Figure 7).

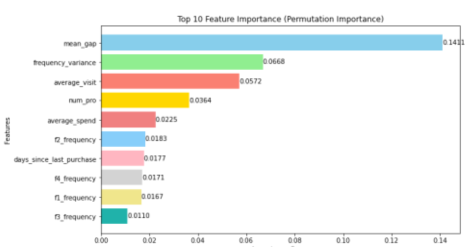


Figure 7. Top 10 Feature Importance Scores (Permutation Method)

GLOBALLY AGGREGATED FEATURES	mean gap , frequency variance , average visit, num pro , average spend,days since last purchase
WINDOWED LAGGED INPUT FEATURES	f1_frequency , f2_frequency , f3_frequency, f4_frequency, f5_frequency

Table 2. Final Feature sets

Based on these findings, the final feature set, combining globally aggregated and windowed lagged inputs (Table 2) set for balanced predictive power, interpretability, and efficiency for weekly deployment.

Model Evaluation Strategy

To develop a churn prediction system aligned with FoodCorps's weekly deployment cycle, a temporally consistent rolling evaluation framework was implemented. As illustrated in Figure 8, the evaluation was structured in two main stages: model tuning and validation, followed by final testing on a temporally held-out dataset.

During the tuning phase, a **rolling window approach** was used to simulate real-world conditions where only past data is available at prediction time. Customer behavior was captured using **five consecutive 30-day input windows** (totaling 150 days), followed by a **30-day output** window for churn labelling. This setup was repeated across five rolling reference dates, ending on **12-01-2022**, to form multiple training and validation folds. In each fold, models were trained on earlier data and validated on the next chronological window, ensuring temporal structure and preventing data leakage. **Feature scaling** was also handled with care—lagged and aggregated variables were scaled using statistics derived exclusively from the training data, balanced class rates meant no resampling was needed.

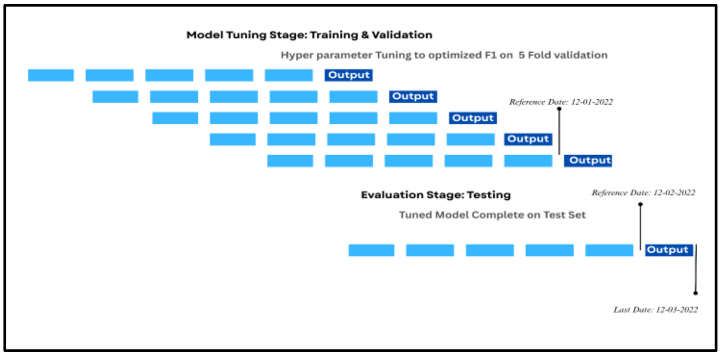


***Figure 8**. Model evaluation strategy*

A diverse set of machine learning models was considered, including **Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, and Linear Support Vector Classifier (SVC). A Dummy Classifier** was included as a baseline, predicting the most frequent class (non-churn), offering a lower-bound benchmark (F1 = 0%, AUC = 50%). Each model was integrated into a pipeline and tuned using **GridSearchCV** across model-specific parameter grids, with **F1 score** as the primary metric to balance false positives and false negatives—both critical for FoodCorp's retention strategy.

In the final evaluation, the best-tuned model was tested on a held-out dataset using a reference date of **12-02-2022**, with predictions extending to **12-03-2022**. This test set was isolated from prior tuning to assess generalization under deployment-like conditions. The same five 30-day input windows generated features, with a 30-day prediction horizon. This ensured evaluation was temporally realistic and operationally representative, confirming the model's readiness for live deployment.

Model Evaluation Results

Models were evaluated on a rolling validation set using 150 days of behavioural input and a 30-day churn prediction window. Metrics such as F1 score, AUC, precision, and recall were computed to assess performance, with F1 used for model selection.

	Parameters	Accuracy	Precision	Recall	F1	AUC
Random Forest	max_depth=20, min_samples_leaf=1, n_estimators=50	77%	76%	77%	76%	85%
Decision Tree	max_depth=10, min_samples_leaf=2	74%	70%	62%	65%	75%
Logistic Regression	C=0.01, class_weight=None, penalty='l2'	76%	69%	76%	72%	83%
SVC	C=0.01, class_weight=None	76%	68%	78%	72%	83%
KNN	n_neighbors=7, weights='uniform'	76%	74%	65%	69%	83%
Dummy Classifier	None	58%	0%	0%	0%	50%

***Table 3.** Performance Comparison of Classifiers Across Key Metrics*

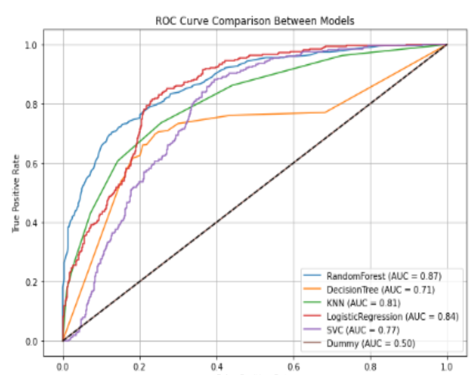


Figure 9. *ROC Curve Comparisons*

Table 3 shows comparison of six classifiers across five metrics on the validation set**: Accuracy, Precision, Recall, F1 Score, and AUC**. Random Forest outperformed all with an F1 score of 76% and the highest AUC of 85%, balancing recall and precision effectively. Logistic Regression, SVC, and KNN showed decent AUCs (83%) but lower F1 due to weaker precision or recall.

The ROC curves in fig 9. visualize each model's ability to discriminate between churners and non-churners across all thresholds. The Random Forest curve lies farthest above the diagonal, confirming superior performance with the highest AUC (0.87). Logistic Regression and KNN followed with AUCs of 0.84 and 0.81, while Decision Tree showed minimal separation from the random baseline (AUC = 0.71), reflecting weaker generalization. The Dummy Classifier traced the diagonal line, reinforcing its role as a non-informative benchmark.

Random Forest was selected as the final optimized model for deployment due to its consistent superiority across both classification and ranking metrics throughout the validation process. It demonstrated stable performance across folds, balanced error trade-offs, and strong generalization to future data. These optimized hyperparameters achieved best balance between predictive performance and model complexity, making Random Forest a robust and scalable solution for FoodCorps's weekly churn prediction pipeline.

Figure 10 shows the precision-recall trade-off across classification thresholds. A **0.5** threshold offers balance, with precision at 0.78 and recall at 0.70. Lower thresholds improve recall but over-target, while higher ones boost precision but miss churners. Given FoodCorps' goal of early, efficient intervention, 0.5 was chosen to support actionable retention without overreaching.

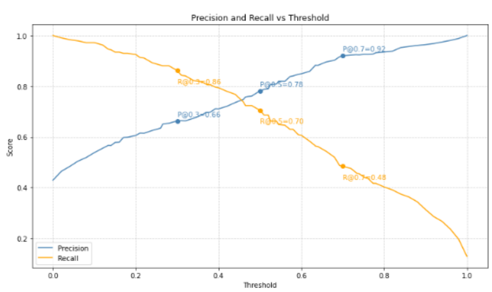


Figure 10. *Precision and Recall vs. Classification Threshold*

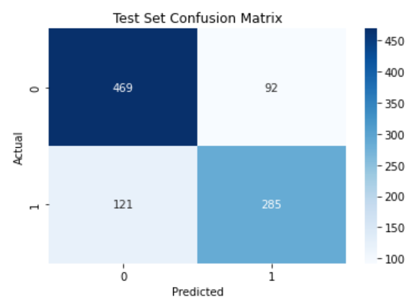


Figure 11. *Confusion Matrix on Held-Out Test Set*

After training on the combined training and validation sets, the final Random Forest model was evaluated on the held-out test set to assess generalization. It achieved an **AUC of 0.87**, with **balanced precision (0.76**) and **recall (0.78)** with **78% accuracy**, the model effectively identifies at-risk customers without excessive false positives, making it suitable for deployment in FoodCorps's weekly churn pipeline. The test set predicted 590 non-churners (62%) and 377 churners (38%).

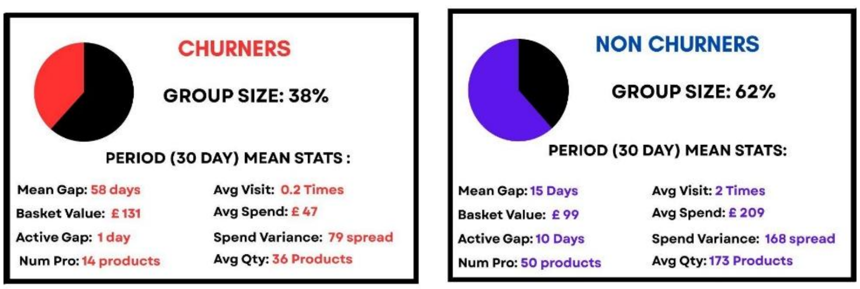
Insight Report

Marketing Summary

While profiling actual churners helps distinguish groups, analysing model predictions on recent data reveals deeper insights into key features from SHAP and Summary Statistics of the groups. As shown in Table 4**, non-churners** make up 62% of the base, while **churners** account for 38%—a sizable at-risk segment, emphasizing the need for strong retention strategies. A key behavioural difference lies in **visit frequency**: non-churners visit about twice per month, while churners average just 0.2 visits. This indicates that churners often disengage early, without forming habits. Timely reminders or follow-up prompts after the first visit could help increase engagement and reduce early-stage drop-off.

Basket composition also varies considerably. While churners have a **higher basket value** (£131 vs. £99), **their average spending per month** is drastically lower (£47 vs. £209). This suggests they make fewer but more expensive purchases likely non-essential or one-off items and then disengage. Promoting everyday products alongside premium items may increase repeat visits and basket breadth. Churners demonstrate significantly less **product diversity**, purchasing just 14 unique products, compared to 50 for non-churners. This limited exposure to the store's full offering likely reduces their connection to the brand. Introducing bundles, sampling strategies, or personalized product recommendations could help expand their basket and build store familiarity.

In terms of consistency, the **mean purchase gap and active gap** (i.e., days between visits) show that churners shop much less regularly. Their average gap is 58 days, while non-churners average 15 days. Churners are active for only 1 day per month, compared to 10 days for retained customers. This sporadic pattern suggests that churners never fully go onboard, making them ideal targets for early reactivation campaigns or first-time purchase follow-ups. Finally, **spend variance**, which reflects fluctuation in monthly spending, is also lower among churners (79 vs. 168). This further supports the pattern that churners tend to be one-time, high-spend buyers rather than ongoing participants in varied shopping cycles. Encouraging more consistent behavior through subscription models or loyalty programs may increase their long-term value.



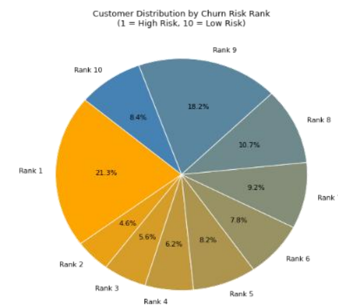


***Table 4**. Summary Statistics of Churners vs non-Churners (30-Day Means)*

***Figure 12.** Customer Distribution by Churn Risk Rank*

To support resource allocation, a priority ranking based on predicted churn probability was added to each customer profile. At the **highest risk (Ranks 1–3),** customers make very few visits, have long gaps between purchases, and buy a small range of products — often one-time, high-spend shoppers likely testing the store. These customers (**making up ~31% of the base**) would benefit from immediate re-engagement strategies such as personalized welcome offers, reminders, or follow-up promotions to encourage a second visit. **Mid-ranked customers (Ranks 4–6)** show increasing engagement with more product variety and shopping days they're on the path to loyalty and could respond well to nudges like loyalty points, bundle suggestions, or cross-selling. Finally**, low-risk customers (Ranks 7–10),** who shop frequently and consistently, represent the most loyal segment ideal for retention through rewards, early-access perks, or exclusive benefits.

Technical Summary:

The first step in distinguishing churners from non-churners was using SHAP values to identify key features and their impact direction on churn likelihood.

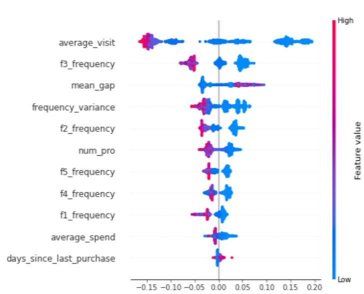


Figure 13. SHAP Summary Plot

SHAP analysis was conducted on the final Random Forest model to interpret the influence (Fig.13) of each feature on churn predictions, it highlighted the key behavioural drivers that the Random Forest model used to predict customer churn. The most influential feature is **average_visit**, where lower visit frequency is strongly associated with a higher likelihood of churn. This is followed by mid-range frequency features such as **f3_frequency** and **f2_frequency**, which also indicate that reduced and inconsistent engagement contributes significantly to churn risk. Additionally, **mean_gap**, which measures the average number of days between purchases, shows a clear positive correlation with churn — longer gaps between visits increase the model's predicted churn probability. High **frequency_variance** further supports this, suggesting that customers with irregular visit patterns are more prone to disengage. In contrast, **features like days_since_last_purchase** and **average_spend** had minimal impact, implying that the model places greater emphasis on behavioural consistency than on spending levels. Overall, the model captures churn risk primarily through indicators of declining or unstable engagement, rather than monetary factors.

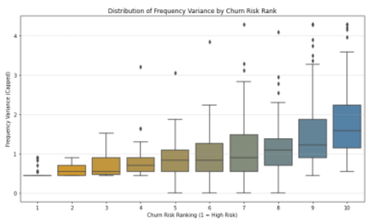


Figure 14. *Distribution of Frequency Variance by Churn Risk Rank*

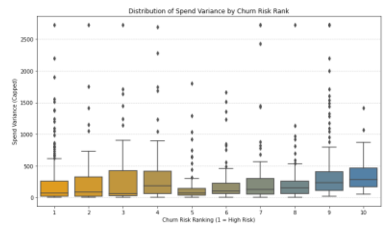


Figure 15. *Distribution of Spend Variance by Churn Risk Rank*

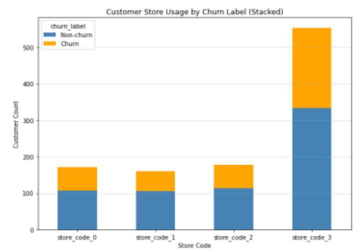


Figure 16. *Customer Store Usage by Churn Label (Stacked)*

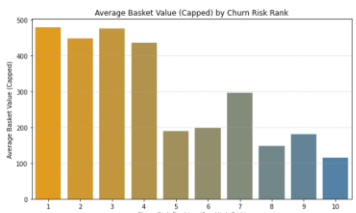


Figure 17*. Average Basket Value (Capped****)*** *by Churn Risk Rank*

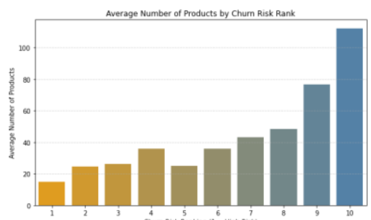


Figure 18*. Average Number of Products by Churn Risk Rank*

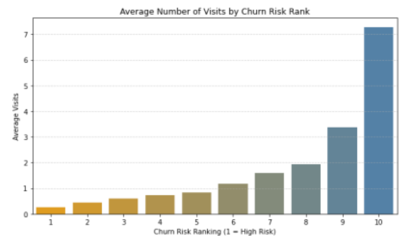


Figure 19*. Average Number of Visits by Churn Risk Rank*

The **model-driven pen portraits** based on churn probability rankings (Ranks 1–10) reveal distinct behavioural segments informed by SHAP logic. **Rank 1** customers, over 20% of the base are mostly first-time users with minimal engagement and low product diversity but high unit costs, representing high-risk trialists. **Ranks 2–4** show early disengagement with inconsistent activity and widening purchase gaps. **Ranks 5–7** include shallow regulars with moderate frequency and limited product depth, indicating fragile loyalty. In contrast, **Ranks 8–10** reflect stable, high-frequency users with diverse purchases and low churn risk. SHAP analysis confirms the model prioritizes recency, regularity, and visit patterns—features like average_visit, mean_gap, and frequency_variance drive predictions, signalling behavioural decline rather than spend as the key churn trigger. These segments support precision targeting and tailored interventions. Figure 14,15,16 17, 18, 19, respectively illustrate key statistics of the 10 groups.

Random Forest was chosen for its strong performance and business alignment, achieving the highest F1 score (76%) and AUC (85%). It effectively captured non-linear patterns, distinguished trialists from disengaging regulars, and avoided overfitting using rolling validation. SHAP results emphasized frequency-based features, reinforcing behavioural focus. The model balances recall and precision, minimizing false positives while identifying at-risk users. Emphasizing temporal signals over spending, it supports proactive retention and scales through a modular, weekly-updating pipeline.

Feature importance analysis across Random Forest, permutation, and mutual information highlighted **mean gap, frequency variance, average visit, and number of products** as key churn predictors, emphasizing behavioural patterns over monetary value. This supported the use of features that generalize across rolling windows. The hybrid schema, blending aggregated and lagged metrics, enabled early churn detection by tracking shifts in visit regularity.