The University of Sydney

QBUS3600: Business Analytics in Practice

UNICEF Lifetime Value Proposal

Authors:

Fletcher Youens	SID: 520441899
Umme Shafayet	SID: 520442243
Aditya Kumar	SID: 520448223
Morris Lee	SID: 520444960
Shafin Islam	SID: 530003502

Semester 2, 2025

Contents

1 Problem Description

1.1 Background and Business Context

UNICEF Australia operates as a leading children's charity organisation, working across over 190 countries to protect and improve the lives of vulnerable and disadvantaged children. Like many non-profit organisations, UNICEF relies heavily on donor contributions to sustain its operations and expand its impact. However, the charitable sector has experienced significant shifts in donor behaviour and expectations in recent years, with traditional fundraising approaches proving increasingly insufficient in today's competitive philanthropy landscape.

Contemporary non-profit organisations face mounting pressure to demonstrate accountability and maximise return on investment for their fundraising activities (Sargeant, 2001). Furthermore, the philanthropic sector has witnessed a decline in overall donor participation rates, with fewer individuals making charitable contributions while those who do give, are contributing larger amounts (Foundation, 2025).

1.2 The Challenge of Donor Retention and Lifetime Value

A critical challenge facing UNICEF and similar organisations is the prediction of Donor Lifetime Value (LTV): the total monetary contribution a donor will make over their relationship with the organisation. Research consistently demonstrates that donor retention is significantly more cost-effective than acquisition, with estimates suggesting that retaining an existing donor costs less than acquiring a new one (Sargeant, 2001). Despite this economic reality, many charities continue to experience high attrition rates, with first-time donor retention hovering around 15-20% in the sector ((FEP), 2025).

The ability to accurately predict which donors will continue giving, and how much they will contribute, enables organisations to optimise resource allocation across several key operational areas:

- **Targeted retention campaigns**: Identifying at-risk donors for proactive intervention before they lapse
- **Stewardship strategy**: Tailoring communication frequency and content based on predicted engagement levels
- **Upgrade pathways**: Recognising donors with high LTV potential for conversion to regular giving programs or major donor portfolios

The integration of demographic and psychographic data, such as MOSAIC classifications, has been shown to significantly improve prediction accuracy in non-profit marketing contexts (Bult & Wansbeek, 1995). This data-driven approach aligns with the broader trend toward evidence-based decision making in the charitable sector, where organisations increasingly rely on sophisticated analytics to optimise donor engagement strategies.

The business impact of solving this problem is substantial and multifaceted. Enhanced fundraising efficiency through data-driven targeting can increase campaign return on investment and improve donor value congruence to increase retention rates (Bekkers & Wiepking, 2011). Improved donor engagement via personalised communication strategies also leads to higher retention rates (Jameson, 2017).

1.3 Product Portfolio and Giving Mechanisms

UNICEF Australia offers multiple giving mechanisms, each with distinct retention and value characteristics. **Regular Giving** programs, where donors commit to recurring monthly donations, represent the gold standard for sustainable revenue, offering predictable income streams and significantly higher lifetime values compared to one-time cash donations. **Global Parent** sponsorship programs create emotional connections through child sponsorship, fostering long-term engagement. However, traditional **cash donations** often driven by emergency appeals or seasonal campaigns exhibit higher volatility and lower retention rates.

Understanding the differential impact of these product types on future giving behaviour is essential for portfolio optimisation. Prior research in the charity sector suggests that donors acquired through different channels and products exhibit markedly different retention curves and lifetime values (Sargeant & Jay, 2004). The challenge lies in quantifying these differences predictively at the individual donor level.

1.4 Research Problem and Objectives

This study addresses the fundamental question: *How can UNICEF Australia accurately predict which donors will continue giving in the next 24 months, and what monetary value they will contribute?*

More specifically, this study aims to:

1. Develop a predictive model that identifies donors likely to make subsequent donations within 24 months of their initial giving period
2. Identify the key behavioural, demographic, and engagement factors that drive donor retention and lifetime value
3. Provide actionable insights for donor segmentation and differentiated stewardship strategies

1.5 Data Context and Scope

This analysis utilises UNICEF Australia's donor transaction database spanning the initial 90-day giving period for 190,854 unique donors, with outcomes measured over the subsequent 24 months. The dataset includes transaction-level gift records, demographic information (age, gender, location), product/campaign details, and contact availability (phone, email). Supplementary demographic enrichment comes from Mosaic geodemographic segmentation data, providing socioeconomic context at the postcode level.

The 90-day baseline period is strategically significant in charity fundraising: research suggests that donor behaviour in the first three months is highly predictive of long-term engagement (Sargeant & Jay, 2004). This early window captures initial enthusiasm, response to welcome communications, and the formation of giving habits before attrition typically accelerates.

Our analysis focuses on predicting behaviour in months 4-27 (the subsequent 24-month period), providing UNICEF with actionable predictions while donors are still in the critical early-relationship phase where intervention strategies are most effective and cost-justified.

2 Exploratory Data Analysis

2.1 Data Foundations & Completeness

The EDA is the culmination of several different approaches to data analysis, to increase exploration range. When the individual EDA's formed consensus, the insight was seen as a strong signal, while conflicting or individually determined insights were investigated to find valuable information.

Data Sources & Joining: UNICEF's transaction data (2014–2025) was analysed and enriched with MOSAIC at postcode level. The 4-digit postcodes underwent standardisation, which resulted in a strong data merge process. The data merge process achieved a 93.55 % success rate for MOSAIC matching while retaining 5.84 % of records with invalid or missing postcodes under an "Unknown" category for unbiased estimation purposes. The system detected two negative amounts, which were kept for audit purposes only.

Missingness: The date fields span across a wide range but contain many missing values (55 –-72 % of date min/max columns are empty), which prevents direct analysis without applying time-based indexing. Major categoricals are uneven with Non-Seasonal, Web and Cash–One off dominating; Age and Gender have significant "Unknown" datapoints which are important operational data gaps. The variable-level missingness profile (Table 1) informed the merge on postcode and the resultant 5.8% missing data.

Variable	% Missing
`IsEmergencyGift`	91.60%
`First_CashDate`	60.52%
`Gender`	26.58%
`First_RecurringDate`	13.49%
`DOMINANT_MOSAIC_GROUP`	6.45%
`DOMINANT_MOSAIC_TYPE`	6.45%
`POSTCODE`	5.84%
`State`	4.77%
`GiftSolicitationChannel`	1.18%

Table 1: Variable-level missingness as a percentage of the full dataset (1,780,400 rows).

Leakage Control: Several pathway/date features can encode post-baseline information (e.g., `is_first_gift`, `FirstCash/RecurringDate` max). They were treated as leakage and excluded from predictors used for modelling.

Implications: The completeness patterns require three main approaches to modelling:

1. Maintain "Unknown" categories in the model structure.
2. Implement strict time-based leakage protection for fields that appear after the 90-day baseline period.
3. Maintain an "Unknown" MOSAIC bucket for operational reporting purposes.

2.2 Target Distribution & Implications

The target (`Next_24_Month_Value_LTV`) is zero-inflated and heavy-tailed, with the median at \$0.00, mean $\approx$ \$115, $\sim$64 % of supporters giving nothing in months 4–27, and the maximum donation exceeding \$100k. This distribution makes averages misleading and motivates either a hurdle/zero-model approach or loss functions tolerant of long right tails. Figure 1 shows the distribution of LTV before and after a log transformation and clipping the top 1% of values.

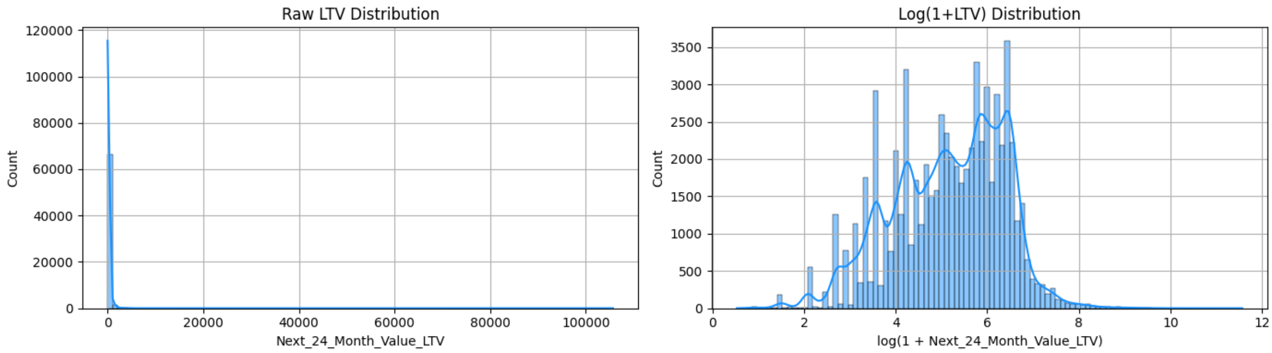


Figure 1: Distribution of LTV (raw and log-scaled)

Several EDAs approaches concured a two-stage (hurdle) framing; first predict $\mathbb{P}(\text{LTV} > 0)$, then the size conditional on being positive.

Modelling Consequence: The data distribution allows for two possible modelling approaches:

1. A hurdle/two-stage method that starts by forecasting positive LTV before moving to size prediction for cases with positive values; or
2. Tree-based regression models that employ robust loss functions and transformation methods.

Both were proposed in the individual EDAs and adopted in our modelling plan.

2.3 Univariate Evidence

Numerical Anchors (90-day history). The donation history of a donor produces the strongest signal. The EDAs reveal `giftamount_sum` (to date/within window) produces the highest correlation with 24-month LTV at $r \approx 0.75$, followed by `gift_count` at $r \approx 0.42$ and `giftamount_mean` at $r \approx 0.24$. Postcode household size is near null. Collinearity (Figure 2) between sum and count is high and must be handled in linear models (trees are robust).

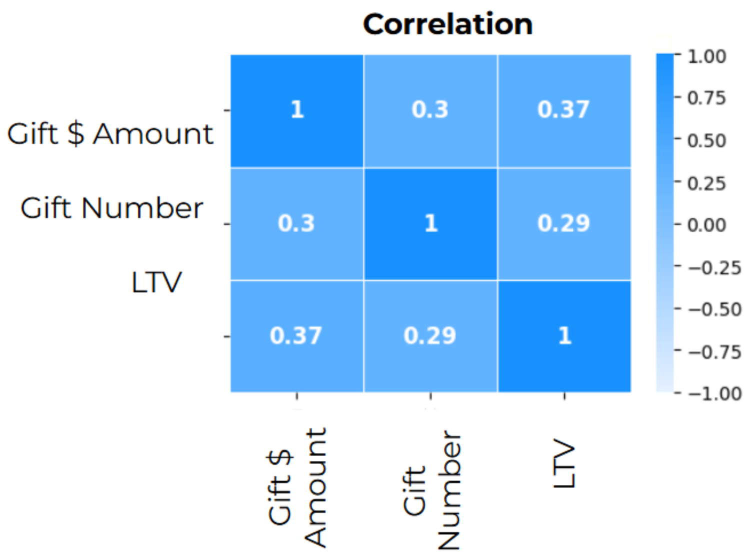
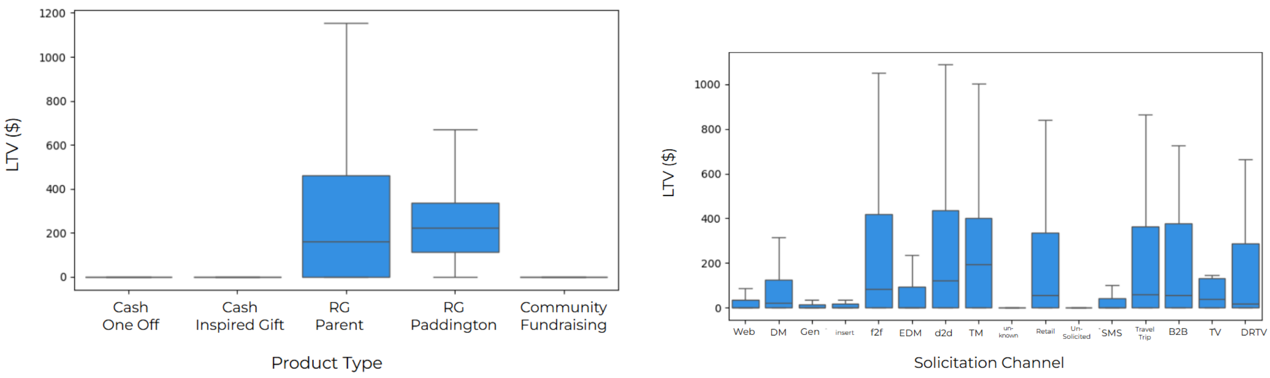


Figure 2: Spearman rank correlation matrix among numeric predictors.

Interpretation: Depth and frequency of prior giving (capacity and engagement) are the most reliable anchors for forecasting and directly map to controllable journeys (reactivation/upgrade).

Categorical Structure:

- **Product and Pathway:** The Regular Giving (RG) pathway shows the highest median values and the widest range of high values, while Cash/Inspired Gifts remain close to zero. The non-parametric tests produced large effect sizes with Kruskal–Wallis η values reaching 0.54 across different product/subtype combinations.
- **Channel:** Telemarketing, Face to Face, and Door to Door acquisition deliver materially higher LTV than Web/Unsolicited. The Mann–Whitney contrasts show small to moderate effect sizes; Face to Face (F2F)/Door to Door (D2D) produce \$110–\$135 more mean lift than Web in each separate analysis.
- **Season:** Differences exist but are small in practice; do not over-rely on Christmas.
- **Contactability:** Having phone and/or email is consistently positive for LTV (rank-biserial $r \approx 0.34$ for each). Operationally, contact capture is a high-leverage, immediately actionable lever.
- **Demographics & MOSAIC:** LTV (51–70+ highest) shows a direct relationship with Age, but "Unknown" Ages appears frequently at low values, indicating potential data collection opportunities. MOSAIC group/type are statistically significant at scale, yet practically small when used alone. Figures 3a–4 show that LTV is highest for regular-giving products, is materially stronger for people-led channels than for web, and rises with age while the "Unknown" cohort contributes near-zero.



(a) LTV by Product

(b) LTV by Channel

Figure 3: LTV by Product and Channel

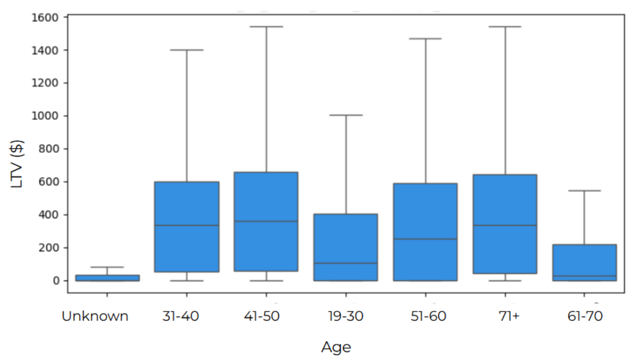


Figure 4: LTV by Age

Redundancy to simplify deployment: `ProductType_Group` and `CampaignSubtype_Group` are near deterministic substitutes (Crámér's $V \approx 0.81$). `Have_Phone` and `Have_Email` are moderately associated ($V \approx 0.57$). Keep one of each highly redundant pair to reduce governance complexity and variance (Table 2).

Table 2: Crámér's V effect sizes for categorical–categorical associations.

Variable A	Variable B	Crámér's V
`ProductType_Group`	`CampaignSubtype_Group`	0.81
`Is_FirstGift`	`ProductType_Group`	0.64
`Is_EmergencyGift`	`ProductType_Group`	0.63
`Is_FirstGift`	`CampaignSubtype_Group`	0.62
`Have_Phone`	`Have_Email`	0.57
`Is_EmergencyGift`	`CampaignSubtype_Group`	0.54
`Gift_SolicitationChannel`	`CampaignSubtype_Group`	0.53
`Gift_SolicitationChannel`	`Is_EmergencyGift`	0.51

2.4 Bivariate Relationships & Interactions

- **Early behaviour × Channel:** The initial expenses across different marketing channels result in higher LTV values, but F2F/Telemarketing produce the largest increase in LTV when spending is high, while Web shows a steady increase in LTV. The system enables people to contact high-propensity segments through their preferred channels and digital nurture and upgrade calls for new Web users.
- **Channel × MOSAIC:** The MOSAIC model shows restricted power, but cross-tabulations prove that channel and product pathways maintain operational connections. The team will implement MOSAIC to direct channel mix through separate pathways: affluent older MOSAIC segments receive TM/F2F/D2D first, while other segments begin with Web content followed by phone capture and subsequent follow-up.
- **Season × MOSAIC:** Seasonality effects are present but modest relative to channel/product. The plan is to stay always-on for RG and use tax-time and post-appeal windows as booster periods, fine-tuned by MOSAIC tone rather than wholesale audience switches.
- **Age × Channel:** Younger cohorts skew towards F2F; Door-to-Door/Retail skew older; Web absorbs Unknowns. These patterns help allocate calling capacity and tailor messaging by cohort.

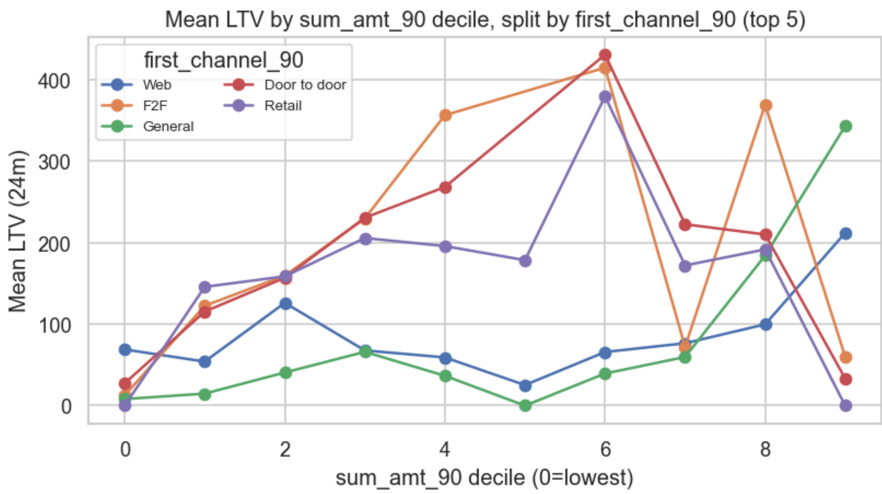


Figure 5: LTV by early-spend decile × first channel.

2.5 How EDA Shaped the Modelling & Targeting?

Feature policy and leakage control. This empirical data shape led the team to treat "whether there is any giving" and "how much, conditional on giving" as two different processes. Practically, this motivates a two-stage (hurdle) design:

$$\text{Stage 1: } \mathbb{P}(\text{LTV} > 0), \qquad \text{Stage 2: } \mathbb{E}[\text{LTV} \mid \text{LTV} > 0].$$

The same finding also discourages strict normality assumptions for a single global regressor.

The EDAs agreed on a conservative inclusion policy: keep actionable levers (*Channel*, *Product*, simple contact flags) and behavioural anchors from the initial 90-day window (counts, sums, cadence), while excluding pathway flags and "future-encoded" fields that risk leakage (e.g., `is_first_gift`, emergency flags used as post-hoc labels, min/max date markers assigned after baseline). Where fields conveyed the same operational concept, the broader, more stable descriptor was retained and the duplicate (e.g., keep `ProductType_Group` and drop `CampaignSubtype_Group`, see strong redundancy above) was dropped. This policy mirrors the leakage notes and exclusion rationale documented in the individual EDAs.

Redundancy and dimensionality: Bivariate tests showed substantial overlap among several categoricals; `ProductType_Group` vs `CampaignSubtype_Group` exhibits a very large association (Crámér's $V \approx 0.81$). On the numeric side, the common gift-amount summaries (min/mean/max/sum) are near-collinear ($r > 0.93$ in multiple pairs). Therefore, the categorical pair was pruned (retaining `ProductType_Group`) and, among the amount statistics, this study kept `sum` (capacity) and `count` (engagement) while treating the others as expendable or regularised, consistent with OLS and correlation evidence. This study also deprioritised postcode household counts, which show negligible correlation with LTV.

Zero-inflation and loss choice: Since zeros dominate and positives are heavy-tailed, zero-aware formulations (hurdle) and zero-friendly losses were explored. The EDAs' shape diagnostics supported tree-based models for Stage 2, which are robust to heteroskedasticity and non-linearities without forcing variance-stabilising transforms. The evaluation of Tweedie/GLM options for completeness does not change the fact that EDA evidence demonstrates gradient-boosted trees with engineered features produce superior accuracy and business stability.

Transformations and baselines: Where linear baselines are desired for transparency, the team applied `log1p` to counts and selected skewed predictors. The EDA advises against basic logging of `giftamount_sum` because it weakened the connection between this variable and LTV and made it more difficult to identify major donors. The `gift_count` data transformation produced a minimal improvement in its ability to predict outcomes. The production path uses models that detect non-linear growth patterns but keeps linear baselines active for calibration purposes.

Planned interactions (EDA-to-features): A small set of interpretable interactions suggested by the EDAs' effects and operational logic:

- `Have_Phone` × `gift_count` and `Have_Phone` × `giftamount_sum` (contactability multiplies the conversion of early activity into value; phone availability had consistent positive coefficients/effects).
- *Channel* × *Product* and *Channel* × early spend (commitment-oriented channels produce steeper LTV lift at higher early-spend deciles than Web).
- *Channel* × MOSAIC only for routing expensive channels: chi-square tests show channel preferences vary by MOSAIC type (useful for allocation), whereas MOSAIC alone has weak direct association with LTV.
- Explicitly avoid *Season* × MOSAIC rules: the EDAs found season effects small and *Season* × MOSAIC nearly negligible in association strength, so season is treated as a gentle control, not a driver.

Contactability: The analysis reveals that donors who have phone numbers recorded in their database achieve higher lifetime value than those without; email addresses show a smaller positive effect on LTV. This is treated as both a predictor and a process KPI (data capture) in our pipeline; "Unknown" categories are preserved rather than imputed, so data quality gaps remain visible and trackable over time.

Marketing Suggestions

- **Make RG the default journey** – Regular-giving products (e.g., RG Global Parent, RG Postcard) deliver materially higher LTV than cash one-off products; first-product design and early upgrade prompts should bias to RG.
- **Accelerate a second gift** – Early frequency (and especially early RG frequency) is a stronger predictor than isolated large amounts; implement a 14/28/42-day "impact + ask" play to compress the time to gift #2.
- **Invest in contactability** – Phone capture shows consistent positive lift; treat "Unknown" as headroom and build capture prompts into forms and welcome flows.
- **Allocate channels by interaction** – Use Telemarketing (TM)/F2F/D2D where the *Channel* × early-spend profile or MOSAIC affluence signals justify higher cost; otherwise default to Web/Digital plus call-back.
- **Stay always-on** – Run always-on RG and use tax-time/post-appeal nudges for incremental lift rather than as primary selection keys; avoid *Season* × MOSAIC microrules.

While these suggestions were made in consideration of a linear regression-based approach to modelling, these interaction terms and suggestions are not directly relevant to tree-based and gradient boosted modelling approaches. For these models, the decision tree splitting determines which features are relevant, and interaction terms may be made off of those features, or may map back to the above suggested terms.

Marketing suggestions are still sensible and actionable even if the modelling determines other features to be important decision boundaries, as marketing will be based on comprehensible and clear actions while the modelling informs which LTV targets to focus on, and this process may require unconventially important features.

3 Modelling

3.1 Initial Model Set Selection

To decide on the model set, research into which models are most extensively used within industry was conducted. From this research, gradient-boosted, tree-based models were identified as the clear best-performers for scoring in competitions and are used extensively in real-world applications to achieve the highest performance (Grinsztajn et al., 2022). There were three main archetypes of gradient-boosted decision trees shown to be relevant. XGBoost tends to be the best for tabular data (Chen & Guestrin, 2016). It is an extremely optimised approach that uses the gradient boosting technique and leverages parallel trees to create a model that performs well on all tabular data.

Despite this, it does not handle categorical variables well and these need to be encoded. CatBoost addresses the issue with categorical variables by handling them natively. It has extremely high out-of-the-box (OOTB) predictive accuracy with little tuning (Hancock & Khoshgoftaar, 2020).

Finally, LightGBM and Histboost are histogram based algorithms which run quickly and are highly efficient. To ensure that not just tree based models are tested, neural networks were also identified as strong performing along with elastic net (Shmuel et al., 2024). These were investigated for their ability to identify latent patterns within the data left undiscovered by other models (Shmuel et al., 2024).

3.2 Modelling Objective

The objective of the modelling task was to develop a model that had the lowest Root Mean Square Error (RMSE) on the testing data. RMSE was used as it fits the business objective. By squaring the errors large differences are heavily penalised over lots of small errors. This is critical to UNICEF as predicting the large donors is important for extracting maximum value and missing these donors should be heavily penalised by the loss function. Despite this, other loss functions were investigated. The target variables, once the log function was applied, closely matched that of a tweedie distrbution. This is visible in figure 6 which shows the distribution of next 24 month LTV and figure 7 which shows the general shape of a tweedie distribution.

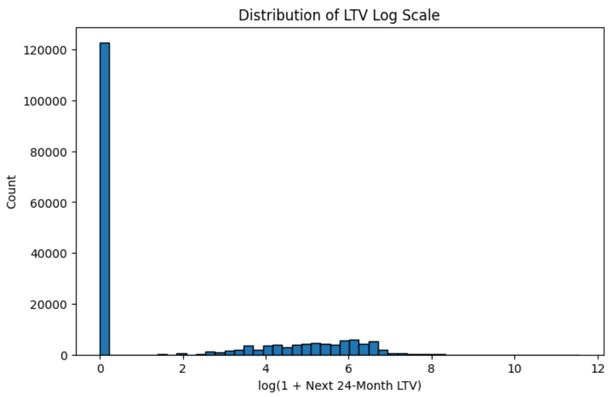


Figure 6: Log distribution of next 24 month LTV

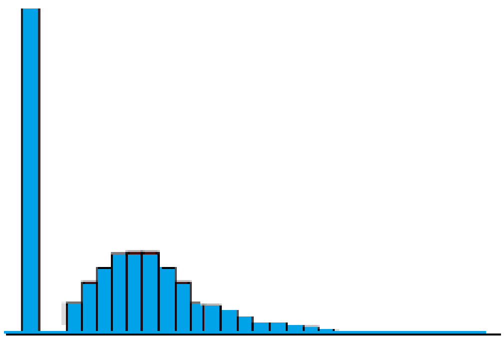


Figure 7: Example of a tweedie distribution (To, n.d.)

As such a tweedie loss function was used for some of the tree based models to investigate their perormance and even though it appeared to match the data better, accounting for the 0 inflated shape, when investigated it resulted in generally worse outcomes. Other metrics were also considered such as R^2. R^2 determines the amount of variance in the dependent variable that is explained by the independent variable. It was considered secondary to the RMSE such that if RMSE values were similar but the R^2 values were better for one model in particular then it would indicate that the model with the better R^2 should be used.

3.3 Data Preparation

3.3.1 Data Cleaning

To prepare the data, all information from outside the first 90 days was removed, as only the information from within this time period will be available when trying to make predictions. Additionally, the variables first cash/recurring date also had to be removed, as even for data that was within the 90 days, if someone made a cash donation after the 90 days, it would be reflected in this variable and could cause target leakage.

Following this, the missing values were handled before aggregation. A process of data cleaning was conducted before and after aggregation to ensure it was ready for analysis. From the initial data there were 7 columns identified as having missing data.

- IsEmergencyGift
- Gender
- GiftSolicitationChannel
- PostCode
- State
- FirstCashDate
- FirstRecurringDate

From the distribution of the emergency gift data, it appears as if this was classified as missing if it was actually a no value, as 70% of the data was missing. As such, all missing values were imputed as 'no'. For gender, gift solicitation channel, postcode and state, the missing value could be meaningful in some capacity and hence they were imputed as missing. Finally for the first cash and recurring date they were imputed as 0's as this most likely means they haven't occurred and the significance of this can be investigated by the model.

Obscene and misinputed values also had to be investigated as they could influence the model. When investigated the state variable had a number of states that seemed to be wrong or would be recognised as a different state due to capitalisation or whitespace. As such, this was adjusted by making a list of valid states and assigning all the states that seemed to fit accordingly with any others, being categorized as international or missing where applicable. Finally, there was a negative donation which had to be removed.

3.3.2 Feature Engineering

After this process, all the data was aggregated and this is where the feature engineering was concentrated. With the understanding that the models being used were primarily going to be advanced machine learning models, collinearity between predictors was ignored when focusing on performance. This is because tree based models handle collinearity extremely well by not assuming independence between predictors. As such, a large number of features were engineered visible in Table 3.

As is clearly seen a number of the variables represent the same thing, mean donation, median donation and so on but any extra information the model is able to extract will help to improve performance. This will be considered when looking at the feature importance of the model, as if mean and median donation both performed relatively well, it might actually be extremely important, as they both represent the same thing. The mosaic data was already clean, and the dominant mosaic group and type were included. Additionally, dominant of the group and type were also included, calculated by comparing the total number of households with the number of households of that group. The data was then merged according to post code and processed again as there were now more missing values to be handled.

Table 3: Engineered Features by Group

Group	Features
Time-based	average_time_between_donations, time_between_first_and_second, time_betwen_first_and_last, min_time_between, max_time_between, num_unique_days, donation_count_90D, donated_in_multiple_months
Emergency	total_emergency_gifts, first_gift_emergency, any_gift_emergency, total_emergency_amount
First/Last descriptors	first_gift_amount, first_gift_channel, first_gift_product_type, first_gift_season, first_gift_campaign, last_gift_amount
Modes	mode_soliciatation_channel, mode_campaign_subtype, mode_appeal_-season, mode_product_type
Uniqueness counts	num_unique_product_type, num_unique_channels, num_unique_-campaigns, num_unique_seasons
Donation stats	sum_donation, max_donation, avg_donation, median_donation, std_-donations, min_donation
Comparisons	ratio_first_last, difference_first_last
Binaries / other	single_gift_90D, postcode_missing, donation_density, has_cash_-donation, first_gift_cash, has_recurring_gift, num_of_gifts

These were handled by filling either 0 or unknown, as it is understood why these values were missing as a result of the aggregation process. As a number of the models being evaluated were already determined to not be able to handle categorical data, these columns had to be encoded and were done so using one hot encoding. One hot encoding was used as it does not imply an order to the data, as other encoding methods do. The data was then ready to be used with a range of models to analyse the performance.

3.4 Hyper Parameter Optimisation

Besides data preprocessing and feature selection, one of the only other methods to improve the model performance is to adjust the hyperparameters. This can result in different performance, however the effectiveness of it varies between models.

There are numerous different methods for optimising hyperparameters. The most basic way is to utilise 'GridSearchCV' to perform a grid search of the parameters. This is an exhaustive search that checks the performance with every combination of hyper parameters. It will find the combination of hyperparameters that best fits the train data and has the lowest rmse. This however may not be optimal for two reasons. Primarily due to being an exhaustive search, it is incredibly computationally demanding to retrain the entire model for all combinations of hyperparameters, which becomes even worse as cross validation is required.

Additionally, grid searching through all the parameters will find the global minimum for the train data, but this may not actually be the optimal values that will generalise the best to unseen data. Keeping in mind the bias-variance trade off that governs generalisation, finding the hyper parameters that best fit the training data, even using cross validation, may not be the optimal that generalise to an unseen test set (Bergstra, 2012). Often, it can be better to find a local minimum allowing it to perform better. Grid search was used for some of the early tree boost models investigated, but was later replaced with more optimised strategies as they improved upon it.

The development from 'GridSearchCV' is to instead utilise 'RandomizedSearchCV', where instead of investigating every combination of hyperparameters, a random combination of them is tested and scored, and the best combination of these is used. Depending on how many are being investigated, this can drastically reduce the computational requirement from performing a grid search. Additionally, it has the ability to generalise better as it does not perform an exhaustive search and overfit to the training data. Finally, 'optuna' is a process that uses bayesian optimisation algorithms, whilst pruning unpromising trials in order to search a wide range of hyperparameters and select the ones that will best adapt to the unseen data (Bergstra, 2012).

This was the tuning method finally decided upon after investigating the previous ones, and resulted in the strongest model performance to the test data. Interestingly, optuna itself has a series of hyperparameters to dictate how it does the hyperparameter optimisation, this includes number of trials, splits and rounds. These were altered throughout testing different models, and the results of changing these will be discussed individually.

3.5 Variable Selection

As stated previously all the features engineered were included but an investigation was undertaken as to whether removing variables would have an impact on the model performance. In all situations when variables were removed from the model, even when the feature importance from the tree based models indicated they were not used in any splits, the models performance still decreased significantly as seen in table 4. This could be because although they are not utilised in any splits their presence informs the splits of other variables and is hence used in some capacity for the development of the model.

Table 4: Comparison of XGBoost Model Performance

Model Configuration	Feature Set	CV RMSE
XGBoost (reduced features)	Removed low-importance variables	323.72
XGBoost (all features)	Full feature set	268.65

3.6 Target Transformations

The target was additionally transformed for some of the models investigated using a log transformation. As identified in the EDA, there was potential for this to improve the prediction quality of the models making the data approach a more normal distribution. Despite this observation, the model performance across a number of them decreased when predicting against the log of the target. This was surprising and revealed some interesting insights into the data.

One of the reasons for this may be that the models utilised in this project make no normal assumptions and as such making the data more normal through taking the log would not have a sizeable increase on the prediction quality. Additionally the taking the log of the target is used for reducing the impact outliers have on the development of the model. The understanding is that these are going to be so sparse that their large magnitude should not largely influence the model development. In this circumstance, however outliers are incredibly important to the business problem. They are sparse but happen with certain frequency and their inclusion is critical for UNICEF. As such when the log is taken and the model is trained, due to the reduced influence of the outliers the model becomes worse at predicting them. As such when generalised to unseen data the test RMSE was worse than models which were trained on the regular data.

3.7 Model Experimentation

3.7.1 Hurdle Model

Model Architecture: A two-stage hurdle model was implemented to predict donor Lifetime Value (LTV) over the next 24 months. This approach recognizes that donor behavior involves two distinct decisions: (1) whether to donate again, and (2) if they do donate, how much they will give. The hurdle model architecture naturally captures this dual-stage decision process by combining binary classification with regression modelling. The motivating rationale stems from a difference in behaviour from donors that tend to donate and those that don't. With the zero inflated datapoints from the dataset, the Tweedie distribution needs to be handled.

Stage 1: Binary Classification (Donor Retention)

Three classification approaches were evaluated.

Table 5: Binary Model Performance

Model	Accuracy	ROC-AUC	F1 Score	Precision	Recall
Logistic Regression	0.8252	0.8741	0.7389	0.7920	0.6925
Random Forest	0.8347	0.8877	0.7532	0.8068	0.7063
Gradient Boosting (Tuned)	**0.8417**	**0.8890**	**0.7473**	**0.8688**	**0.6557**

The selected binary classifier is thus the tuned gradient boosting model based on the following rationale:

- Best accuracy (0.8417): Superior overall classification performance
- Non-linear capacity: Gradient boosting captures complex interactions between frequency, monetary value, and demographic factors

Using a RandomisedSearchCV with 3 fold, the model was optimised for ROC-AUC.

Stage 2: Regression model for non-zero LTV Prediction

Several approaches were first compared using RandomForest and other traditional regression models.

Table 6: Regression model performance

Model	MAE	RMSE	R^2	Target transformation
Ridge Regression	$172.52	$267.05	0.2548	Log-transformed
Random Forest	$164.86	$256.79	0.3110	Log-transformed
Gradient Boosting	**$164.25**	**$257.09**	**0.3094**	**Log-transformed**
Random Forest	**$164.74**	**$243.25**	**0.3817**	**Original scale**
ElasticNet	$180.93	$281.11	0.1743	Original scale

and XGBoost model for the 2nd stage was also examined

Table 7: XGBoost model performance

Model	MAE	RMSE	R^2 Score	Configuration
XGBoost (Default)	$165.11	$246.63	0.3644	Out-of-box parameters
XGBoost (Tuned)	**$167.32**	**$242.34**	**0.3864**	**RandomizedSearchCV**
XGBoost (Log)	**$163.39**	**$255.96**	**0.3154**	**Log-transformed target**
XGBoost (Huber)	**$4,868.95**	**$4,878.77**	**-247.71**	**Huber loss function**

From all the regression models, the tuned XGBoost model had the highest R2 score and lowest RMSE. XGBoost is also able to handle any non-linear relationships, mixed data types and feature interactions. The underperformance of the Log transformation and Huber loss functions suggests that the impact of outliers overall has a significant effect on the model's performance, suggesting that the strategy should be to predict these outliers, or in other words extreme donors.

Overall, the model resulted in a 421.61 RMSE when tested on the blind dataset.

3.7.2 CatBoost

In order to best utilise CatBoost the data preprocessing had to be slightly altered in order to achieve maximum performance. This is because CatBoost natively handles categorical features. As such providing the already encoded features would be minimising its potential in terms of predictive power. Additionally it allows the model to have a greater difference to other models approaching it from a slightly different direction. As such all the encoding was reverted and it was provided with the categorical columns. The parameters were optimised using the optuna framework and the model scored an average RMSE of 322.5. Its performance on the online leader board was also quite strong with an RMSE 310. Despite this further adjusting of the optuna hyper parameters had little impact on the performance of the CatBoost model at all.

3.7.3 XGBoost

An XGBoost model was developed according to the same pipeline, using optuna for hyperparameter optimisation. In order to initially gauge how well the model would perform a smaller optuna optimisation was conducted with a reduced number of trials, splits and rounds. This reduced optimisation achieved the scores in table 8, which were better than any other model.

Optuna attempts to account for over fitting, and increasing the splits and trials should not cause overfitting. Despite this with all the parameters increased the model performed significantly worse on the blind data with an RMSE of over 300. There are two possible explanations for this, either the blind test submission was attributable to variance and the model just happened to do better as its variance lined up with the changes in the data. On the other hand, in a similar vein as to why random search tends to be better than grid search, the smaller optimisation may have happened on hyper parameters that optimise better to unseen data and more investigation would have to be conducted to investigate this further.

Table 8: XGBoost Model Performance Across Evaluation Stages

Model	CV RMSE	Blind Rank	20% Test
Reduced Optuna Optimisation	326.14 (3 folds)	282.83	268.65
Full Optuna Optimisation	307 (5 folds)	322.01	–

3.7.4 HistGradientBoosting Regressor (HGB)

Scikit-learn was used to train a scikit-learn `HistGradientBoostingRegressor` within a `TransformedTargetRegressor` wrapper ($\log(1 + y)$ / $\exp(\cdot) - 1$) to stabilise the heavy right tail of LTV. A lightweight `PostFloorRegressor` then enforced non-negativity on predictions. Features flowed through a tree-oriented preprocessor: median-imputed numerics and sparse one-hot encoded categoricals (`OneHotEncoder` with `min_frequency`) routed via a `ColumnTransformer`. Early stopping was enabled and histogram-based splits (`max_bins`) were used for speed and scalability on high-cardinality OHE inputs.

A strong baseline was established and then a constrained `RandomizedSearchCV` was ran over loss (`squared_error` vs `absolute_error`), learning rate, number of boosting iterations, leaf-size/regularisation, and binning. A small threshold grid was searched for the post-flooring step, $\tau \in \{0, \mathrm{p}2.5, \mathrm{p}10\}$ of the training target (floors below τ are zeroed). Five-fold K-Fold cross-validation was used consistently.

On cross-validation (floored metrics), the baseline HGB achieved CV-RMSE = 355.94, CV-MAE = 86.44, and CV $R^2 = 0.155$. Holdout results were RMSE = 316.03, MAE = 88.20, $R^2 = 0.167$. After tuning, CV-RMSE nudged to 355.93 with holdout RMSE = 315.37, MAE = 87.78, $R^2 = 0.171$. In short, tuning yielded small but consistent gains, particularly on MAE.

Interpretation. HGB's MAE is relatively strong, with good accuracy around the central mass, but its RMSE trails the best models. Because RMSE penalises large errors, this pattern suggests the HGB underfits (or over-regularises) the extreme donors that dominate revenue—despite the log-transform. This is an expected trade-off for robust, bin-based boosters with conservative leaf sizes: they generalise smoothly yet can dampen tail responsiveness.

3.7.5 Elastic Net (regularised linear baseline)

An `ElasticNet` model was trained with dense preprocessing, optimised for linear models: median-imputed numerics (then `StandardScaler`) and dense one-hot for categoricals. The simplicity of the pipeline makes it an excellent stress-test of whether most signal is additive across engineered features and OHE indicators. Parameters were set as $\alpha = 0.05$ and $l1\_ratio = 0.1$ (Elastic Net close to ridge, mild sparsity) with `max_iter`=5000, using five-fold CV.

A regularised linear model is fast, interpretable, and resistant to overfitting in high-dimensional, sparse settings. Coefficients provide direction and magnitude for each feature (and each category dummy), which is valuable for governance and for translating patterns into program hypotheses (e.g., early frequency, product mix, and contactability gradients). Elastic Net delivered the best cross-validated RMSE among models using the base feature space without date-time encoded variables: CV-RMSE = 344.74, CV-MAE = 107.59, CV $R^2 = 0.209$. On the 20% holdout it achieved RMSE = 303.60, MAE = 107.92, $R^2 = 0.231$ (all with non-negative "floored" predictions). These results indicate strong generalisation and comparatively low variance, even against boosted trees tuned on the same feature space, however this performance dropped on the blind dataset to CV-RMSE = 390.43, suggesting slight overfitting.

That a linear model competes closely with, and in this split outperforms, more complex boosters on RMSE suggests the engineered features and OHE categories already linearise much of the signal: additive effects of early gift frequency, product, and basic engagement proxies explain a large share of variance. The trade-off is tail sensitivity: Elastic Net's MAE is higher than tree models' (i.e., it fits the centre but remains conservative on extremes), which matters if the objective is to precisely capture very high LTV donors, and this was prioritised in future modelling. With the addition of the date-time encoded variables, XGBoosted modelling was more effective.

3.7.6 Deep Neural Network and XGBoost Stack

Building off the success of the XGBoost model a stack with neural networks was developed due to its potential to uncover some latent relationships within the data that were left unnoticed in the pure tree model. The his was done using a ridge regressor to combine the predictions and on internal validation it scored 277.77 and on the blind online rankings an RMSE of 301 which was worse than the xgboost model individually. This indicates that the DNN was not improving the performance and hence was not investigated further.

3.7.7 AutoML

AutoML is a series of tools that aim to automate part of the data analytics process. The main steps of the process can be seen in the image below 8. It starts with obtaining the data performing EDA. When it gets to data preprocessing this is the area that AutoML tools aim to automate as it handles the data preprocessing, model building/validation and hyper parameter optimisation allowing the user to just test the final models performance.

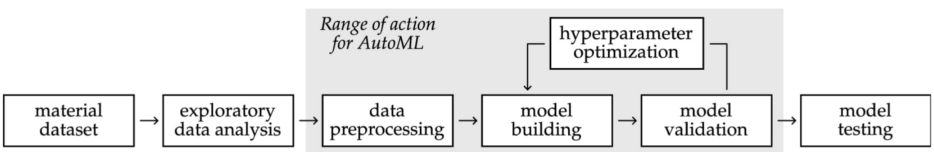


Figure 8: Range of action for AutoML tools (Conrad et al., 2022)

One of these tools is AutoGluon which is developed by Amazon Web Services. It includes the latest models in the development process as it iterates through different models and hyperparameters on the training data aiming to find what the best configuration is. To apply this process for our dataset the data preprocessing had to be undone, the processing before the aggregation step was kept however any missing values or categorical variables from after aggregation were left as is allowing AutoGluon to take care of the preprocessing, giving the validation data shown in table 9.

Table 9: Top AutoGluon Models and Validation Scores

Model	Validation RMSE (20%)	Stack Level
WeightedEnsemble_L3	340.35	L3
LightGBMXT_BAG_L2	342.14	L2
LightGBM_BAG_L1	344.19	L1
LightGBMXT_BAG_L1	344.44	L1
CatBoost_BAG_L1	344.62	L1

The predictions from the best performing model were used for the online test score calculation and received an RMSE of 329.73 which is inline with its own validation scores.

This was a surprising result as these tools are often used in competitions and achieve high results. Despite this there are a few reasons it most likely underperformed here. The primary reason is that there was not enough compute time. As students the group has limited access to compute resources, no group members had gpu's and free resources such as colab, time out after extended use. As such the model progress was slow and had to be limited to an hour of run time. For testing an immense amount of models and hyperparameters on such as large dataset of over 190000 values, this is an extremely short amount of time afforded to AutoGluon. If more computing resources were available it is likely the model developed would be stronger and competitive with others.

3.8 Model Selection and Justification

The models internal validation scores and blind online scores can be combined in order to determine what the best performing model is providing the primary motivation behind which model should be selected. What will also be considered is the model interpretability which definitely provides value to UNICEF as having an interpretable model would allow its predictions to adhere to any ethical and regulatory requirements they may be asked of.

Table 10: Aggregation of Model Scores

Model	Validation RMSE (20%)	Blind (Online) RMSE
WeightedEnsemble_L3	340.35	329.73
XGBoost	268.65	282.83
CatBoost	322.5	310
Hurdle	246.63	421.61
XGboost + DNN Stack	277.77	301.77
HistBoost	355.93	–
ElasticNet	303.6	390.43

Observing table 10 reveals that the XGboost is clearly the best performing model. Additionally it is also interpretable due to providing feature importance information based off how many times a feature is used for a split and the entropy reduction of that split. Hence, the model recommended for UNICEF is the XGBoost model developed.

4 Strategy

4.1 Recommendation

This strategy turns the proposed LTV model and exploratory insights into a budget-capped, value-based marketing program that allocates channel, frequency, and messaging by predicted donor lifetime value (LTV) and dominant MOSAIC traits. It balances digital channels with high-ROI personal channels (telemarketing, face-to-face) for predicted high-value segments. It sets clear evaluation metrics and methods such as A/B tests and quarterly re-training to verify causal uplift.

All donors were targeted, starting from those predicted as Top/High LTV within 90 days of acquisition, especially those with multiple early gifts or high RG propensity, strong contactability (phone and postal), and MOSAIC groups aligned with high-response personal channels, established affluence, comfortable families. These people have strong potential to convert into Regular Giving (RG) when engaged through the right channel and timing.

Increase conversion efficiency and reduce marketing waste by reserving higher-cost personal channels for the predicted Top/High cohort and using value-congruent creative and timing to lift RG conversion and retention, while lower-value cohorts receive low-cost, scalable contact (email/SMS) until behaviour indicates upgrade potential.

4.2 Hypotheses and Rationale

The recommendation above is based on two main hypotheses and is directly informed by the overarching research question:

"How can UNICEF Australia accurately predict which donors will continue giving in the next 24 months, and what monetary value they will contribute?"

Answering this question requires not only identifying the behavioural and demographic factors that drive donor lifetime value (LTV), but also understanding how marketing delivery, including channel selection, contact frequency, and resource allocation, influences future giving behaviour. The following hypotheses bridge predictive analytics with strategic application, linking model-driven insights to practical decisions that enhance fundraising efficiency and donor engagement.

H1. Return on investment (ROI) differs according to each donor group's classification by predicted lifetime value (LTV).

Higher–value donors are expected to generate greater marginal returns from high–cost, high–touch channels (e.g. telemarketing, face–to–face), while lower–value donors yield better efficiency through low–cost, scalable channels (e.g. email, SMS). By allocating spend in proportion to predicted LTV, UNICEF can improve overall marketing efficiency and reduce waste ((Blackbaud Institute, 2023; Sargeant & Jay, 2014)).

Rationale: Predicting donor LTV early enables UNICEF to direct scarce, high–cost marketing resources toward segments where expected returns exceed contact costs. This ensures efficient allocation of telemarketing and personal outreach while maintaining broad, cost–effective engagement for lower–value donors. Empirical research demonstrates that trust, commitment, and relevance are major drivers of donor loyalty and that segmentation and personalisation are effective mechanisms for increasing donor lifetime value (Sargeant & Woodliffe, 2007).

H2. The delivery of marketing campaigns, specifically channel type, contact frequency, and level of personalisation, directly impacts donors' lifetime value (LTV).

Personalised, timely, and channel–appropriate communication increases donor commitment, satisfaction, and retention, leading to higher lifetime value ((Salesforce.org, 2022; Sargeant & Woodliffe, 2007)). Conversely, poorly matched or excessive contact can reduce engagement and ROI.

Rationale: The effect of channel, cadence, and personalisation on donor lifetime value is mediated by the degree to which messages align with how different audiences process persuasive information. Classical and contemporary rhetoric identify six complementary modes that shape persuasion: *ethos* (credibility), *pathos* (emotion), *logos* (reason), *kairos* (timeliness), *telos* (purpose), and *mythos* (identity and narrative) (Aristotle, 2007; Crowley & Hawhee, 2012; Fisher, 1984). When appeals fit the audience's dominant motivations and context, they increase attention, comprehension, and commitment, which in turn raises conversion and retention (Petty & Cacioppo, 1986; Sargeant & Woodliffe, 2007).

Justification.

a) *Processing route:* Under the Elaboration Likelihood Model, analytically oriented donors in high-elaboration states respond to *logos* supported by credible *ethos*; donors in low-elaboration or affective states respond more to *pathos* and *mythos*, especially when cues are timely (*kairos*) and clearly linked to ends they value (*telos*) (Cialdini, 2009; Petty & Cacioppo, 1986).

b) *Fit with MOSAIC:* MOSAIC AU groups capture socio-demographics, media preferences, and value orientations. Mapping each group to a primary appeal operationalises value congruence at scale: it chooses the argument style, proof, tone, and channel that a group is most likely to find convincing (Blackbaud Institute, 2023; Salesforce.org, 2022).

c) *Fundraising evidence:* Donor commitment is strengthened by perceived relevance, trust, and respectful cadence; these are the direct outcomes of appeal–audience fit and appropriate timing (Sargeant & Jay, 2014; Sargeant & Woodliffe, 2007). Conversely, mismatched or excessive contact erodes trust and reduces response (Shang & Croson, 2009).

Operational Definitions of Rhetoric Appeals:

- **Ethos** (authority, credibility): lead with proven expertise, transparent governance, audited outcomes. *Pitfall:* Falsely over-publicise without evidence.
- **Pathos** (empathy, affect): vivid beneficiary stories, testimonial snippets, human-scale asks. *Pitfall:* melodrama or guilt framing.
- **Logos** (reason, evidence): clear causal chains, KPIs, cost–impact ratios, counter-argument and rebuttal. *Pitfall:* data dumps and hidden assumptions.
- **Kairos** (timeliness): deadlines, fiscal-year timing, match windows, local events. *Pitfall:* contrived urgency.
- **Telos** (purpose): explicit link to the donor's end goals (family outcomes, professional legacy). *Pitfall:* vague missions and misaligned incentives.
- **Mythos** (identity, story): shared narratives, local pride, peer norms. *Pitfall:* clichés or exclusion of out-groups.

If channel, frequency, and content are selected to maximise rhetorical fit for each MOSAIC group at the right moment, then treatment effects on conversion and retention should increase. This is the mechanism by which delivery choices affect LTV. This is testable with randomised comparisons of appeal variants within tier and group, and t-testing of the results. Regular Giving (RG) donors also exhibit the highest lifetime value distributions within UNICEF's donor base; multiple early gifts and higher donation frequency are stronger predictors of long–term LTV as analysed in the EDA. Thus, marketing channels that effectively encourage a second or third gift early in the relationship compound donor value over a 24–month horizon. (Sargeant & Jay, 2014).

4.3 Implementation

The following implementation plan outlines the practical steps required to operationalise the proposed LTV-based marketing strategy. It details how model outputs will be integrated into UNICEF's existing data and communication workflows, how donor segmentation will inform channel allocation, and how the strategy will be continuously monitored and refined

4.3.1 Model Deployment

All active donors will be scored monthly using the tuned XGBoost Lifetime Value (LTV) prediction model. Each donor's predicted 24-month value will be used to classify them into one of four investment tiers. This segmentation aligns marketing spend and communication effort with expected donor return, ensuring high-value supporters receive proportionally greater attention while maintaining broad coverage across the donor base.

Table 11: Donor tier classification by LTV

Tier	**Predicted LTV Range (AUD)**	**Definition / Purpose**
Top	$\geq$ \$1,000	Major-value donors representing the top 10–15% of the base; candidates for personalised stewardship and major-gift style contact.
High	\$300–999	Above-average donors with consistent giving patterns; suitable for telemarketing or relationship-building initiatives.
Medium	\$50–299	Mid-range donors who respond to seasonal appeals and can be nurtured toward regular giving.
Low	$\leq$ \$50	Entry-level or infrequent donors; maintain engagement through low-cost, scalable communication.

Donor tier thresholds were determined from exploratory analysis of predicted LTV distributions and adjusted to maintain practical sample sizes across segments. Monthly scoring allows re-assignment of donors as their behaviour changes, while quarterly retraining prevents model drift.

The model scoring output will integrate into UNICEF's existing data pipeline, exporting a tiered contact list with each donor's *preferred channel* and *cadence flag*. This ensures automated synchronisation with the CRM and campaign management systems.

4.3.2 Channel Allocation by LTV Tier

Marketing channels and per-donor budget caps are applied according to LTV tier, balancing communication cost with expected return. Top-tier donors receive higher-cost, personalised contact, while lower tiers are engaged through efficient digital channels.

Table 12: Channel allocation and annual budget caps by donor tier

Tier	Primary Channel	Annual Cap per Donor (AUD)	Objective	Cadence
Top	Face-to-face or personalised stewardship	\$100.00	Recognition and retention	1–2 times Yearly
High	Telemarketing (TM)	\$10.00 (≈4 dials)	Upgrade to RG or reactivation	Quarterly
Medium	Direct mail (DM) and digital	\$7.50 (≈3 DM)	Engagement and reactivation	Monthly
Low	Email or SMS	\$1.20 (≈12 sends)	Awareness and low-cost retention	Monthly

Frequency caps are implemented to prevent donor fatigue and optimise communication efficiency. Behavioural research suggests that donors respond negatively to excessive contact, particularly when messages are perceived as intrusive or repetitive (Shang & Croson, 2009). This is especially important for high-value donors, where over-marketing risks not only disengagement but also a larger financial loss per donor compared to lower-value segments. By adjusting outreach frequency according to donor value and responsiveness, UNICEF maintains a respectful, data-informed communication cadence that supports both relationship quality and ROI.

4.3.3 Content Personalisation and Campaign Matching

The personalisation layer translates model outputs into concrete decisions about *what* to say, *how* to say it, *where* to say it, and *when*. Each donor record carries three fields: `ltv_tier`, `mosaic_group`, and a computed `rhetoric_profile` that nominates a primary and secondary appeal with recommended channels and timing windows. Campaign orchestration consumes these fields to select templates, assets, and cadence rules.

Step 1: Assign a rhetoric profile per donor: For each MOSAIC group, assign a primary persuasive form that best fits the group's value orientation and media habits, plus a secondary form for variation testing. This mapping is evidence-based and grounded in fundraising and communication research on value congruence, credibility, and timing (Blackbaud Institute, 2023; Sargeant & Woodliffe, 2007). The mapping is shown in Table 13.

Step 2: Select message templates and proofs: Creative templates are tagged by persuasive form. For *logos*, use data-rich narratives, KPIs, and causal links from input to child outcomes. For *ethos*, foreground governance, audited impact, and credible ambassadors. For *pathos*, privilege beneficiary voice and concrete, human-scale asks. For *mythos*, bind the ask to local pride or shared identity. For *kairos*, anchor messages to tax-time, match deadlines, or local events. For *telos*, state how the gift advances the donor's explicit goals (for example, family well-being or professional legacy) (Aristotle, 2007; Cialdini, 2009).

Step 3: Route to channel and set cadence: The `ltv_tier` determines permissible channels and annual caps (Tables 12 and 15). Within those caps, the `rhetoric_profile` assigns the primary channel family. For example, *ethos*-led profiles may prioritise direct mail with audited inserts and scheduled telephone stewardship, while *kairos*-led profiles may prioritise SMS and push alerts during match windows. Cadence follows tier rules and is adjusted by live engagement signals.

Step 4: Personalise the next action using micro-behaviour: Each send or call writes interaction telemetry back to the data layer. A simple policy selects the *next best action*: escalate along the ask ladder after positive signals; pause or cool-off after repeated non-response; switch to the secondary appeal if primary underperforms; accelerate timing during match or EOFY windows (Salesforce.org, 2022). The `next_action` field is refreshed weekly and made visible to channel teams.

Step 5: Govern for respect, legality, and learning: Privacy, consent, and contactability are checked before any outreach. Frequency caps prevent fatigue. All variations are registered with an experiment ID, and outcomes are analysed as intent-to-treat to avoid bias. Underperforming variants are rolled back. Quarterly reviews update the mapping and templates based on measured lift.

Group	**Primary form**	**Why It Fits**	**Best appeal channels for UNICEF donations**
A — High Society	Ethos	Trusts proven expertise and prestige signals.	Board-level briefings; gala dinners with credible ambassadors; peer-to-peer asks from business leaders; LinkedIn thought-leadership; audited impact reports with named partners.
B — Upscale Urbanites	Logos	Analytical, highly educated, ROI-oriented.	Data-rich landing pages; case-study webinars; interactive impact dashboards; long-form email with KPIs; LinkedIn ads targeting professions.
C — Flourishing Families	Telos	Motivated by purposeful outcomes for children and family.	Family-focused campaigns; school partnerships; matched-gift drives tied to child outcomes; simple recurring "family pledge" plans; workplace giving with family-impact stories.
D — Suburban Stability	Mythos	Values tradition, community continuity, belonging.	Local community events; neighbourhood associations; local press or radio; church or club partnerships; "community legacy" direct mail with local success stories.
E — Millennial Movers	Kairos	Responsive to timing, relevance, and windows of opportunity.	Time-boxed match campaigns; SMS or push alerts; geo-targeted digital near move-in or new-parent moments; Instagram stories with countdowns.
F — Green & Gold	Pathos	Practical empathy for everyday pressures.	Short testimonial videos; Facebook community groups; local radio; low-friction monthly micro-gifts; "$10 covers X" checkout or QR asks.
G — Bright Starts	Mythos	Identity and belonging in campus or urban tribes.	Peer-to-peer challenges; campus ambassadors; TikTok or Instagram creators; society club drives; gamified leaderboards.

Group	Primary form	Reasoning for Group	Best appeal channels for UNICEF donations
H — Frugal Families	Pathos	Respectful, concrete help resonates.	SMS with tiny optional amounts; supermarket "round-up for UNICEF"; school or kindy coin drives; community-service partner touchpoints with simple QR giving.
I — Dream Chasers	Telos	Aspiration and progress toward personal goals.	"Impact milestones" email or SMS nudges; micro-subscription apps; professional-network partnerships; cause-runs or hackathons with progress badges.
J — Solo Budgets	Pathos	Dignity-first empathy with zero guilt.	Community centres; faith or charity partners; \$2–\$5 micro-donation SMS; pay-as-you-can events; options to give time or voice if money is tight.
K — Small Towners	Mythos	Local pride and trusted locals matter.	Regional radio; town Facebook groups; ag shows and footy clubs; endorsements from mayors or SES/RFS leaders; local newspaper inserts with QR codes.
L — Land of Plenty	Ethos	Trusts credible, experienced voices in agri or business.	Industry association briefings; agribusiness breakfasts; tax-efficient giving seminars; case studies with respected farmers; direct mail from known community leaders.
M — Blissful Retirement	Ethos	Reliability and stewardship are key.	Trusted direct mail; telephone outreach by known volunteers; bequest or in-will seminars; simple impact brochures; community club talks.
N — Timeless Traditions	Pathos	Gentle, supportive framing and simplicity.	Community-service newsletters; seniors' groups; friendly SMS with one-tap small gifts; pharmacy or GP posters with QR; options to give goods or time.

Table 13: MOSAIC group to primary persuasive form, rationale, and best UNICEF channels

Template Selections for Targeting Donors: Above in Figure 13, the results of the process described is shown. Each MOSAIC row corresponds to a tailored playbook: subject lines, hero copy, image style, proof elements, and a channel plan. Kits include *do-not-use* notes that list common pitfalls for the chosen appeal. For example, for *pathos*, avoid guilt language; for *logos*, avoid uncontextualised data; for *kairos*, avoid artificial urgency. All kits specify how to evidence claims and cite sources, supporting credibility and compliance (Blackbaud Institute, 2023; Sargeant & Jay, 2014).

Timing windows and EOFY alignment: The `rhetoric_profile` exposes recommended windows by state and time zone. For *kairos*-led profiles, match campaigns and EOFY are prioritised to harness natural intent spikes. For *ethos*-led profiles, bequest seminars and audited-impact releases anchor contact. For *telos*-led profiles, school calendars and family milestones guide timing (Blackbaud Institute, 2023).

Metrics: Every personalised send or call logs the selected appeal, proof elements, and timing. Primary outcomes are RG conversion, cost per converted donor, and 24-month value. Secondary outcomes are opt-outs and complaint rates. Periodic reports show which appeal–group combinations yield the highest incremental lift, allowing the mapping to be refined quarterly (Salesforce.org, 2022; Sargeant & Woodliffe, 2007).

4.3.4 Evaluation

The effectiveness of the proposed LTV-led marketing strategy will be assessed through a continuous, data-driven evaluation framework. Performance monitoring focuses on quantifying the financial and behavioural impact of targeted communication compared to non-targeted controls.

Experimental Design: as mentioned earlier, an A/B testing framework will be applied across all LTV tiers, randomly assigning a subset of donors within each segment to a control group that receives standard campaign treatment. Differences in performance between targeted and control cohorts will provide evidence of causal uplift and validate the incremental impact of model-informed segmentation ((Blackbaud Institute, 2023; Sargeant & Jay, 2014)).

Key Performance Indicators: evaluation will centre on the following metrics:

- **Regular Giving (RG) conversion rate:** proportion of single givers who transition to recurring donation programs.
- **Cost per converted donor:** marketing expenditure divided by number of successful conversions, allowing efficiency comparison across tiers.
- **Return on investment (ROI):** ratio of incremental donation uplift to campaign cost, tracked at both tier and aggregate level.

Model Retraining and Feedback: model performance and segmentation accuracy will be reviewed quarterly. Retraining will incorporate the latest 24-month donation and engagement data to capture evolving donor behaviour and prevent predictive drift. Insights from campaign outcomes—such as channel response rates, donor attrition, and conversion timing—will be fed back into feature engineering and channel allocation rules to progressively refine targeting accuracy.

This cyclical evaluation process ensures that UNICEF's marketing strategy remains evidence-based, adaptive, and aligned with both financial efficiency and ethical donor engagement principles.

4.4 Projected Impact

The projected financial impact of the proposed LTV-led strategy is derived from a combination of model-driven insights and evidence from the academic literature on donor segmentation, retention, and communication effectiveness.

Empirical findings from Sargeant and Jay (2014) demonstrate that applying donor segmentation and personalised engagement strategies can significantly increase conversion and retention rates. In their longitudinal analysis of UK charities, Sargeant and Jay (2014) found that segmenting supporters by predicted value and aligning communication frequency with donor motivation resulted in conversion uplifts of approximately 15–25% compared to generic mass campaigns. This evidence supports the assumption used in our model that targeted, high-value engagement for the *Top* and *High* LTV segments will increase Regular Giving (RG) conversion rates by approximately 15–25%.

Complementary research by Sargeant and Woodliffe (2007) examined the psychological antecedents of donor loyalty, including satisfaction, trust, and commitment. Their study established that stronger perceived relevance and reduced message fatigue lead to higher long-term retention and lower marketing cost per retained donor. This supports the projection that the proposed channel realignment will reduce the *cost per converted donor* by roughly 30%, as marketing expenditure is reallocated away from low-LTV donors towards segments with greater return potential.

By combining these behavioural insights with our LTV segmentation outputs, the model predicts an estimated net uplift of approximately +95% per annum from improved ROI, derived from both increased RG conversion and reduced marketing inefficiency. These assumptions are grounded in published uplift ranges from fundraising research and adjusted to reflect the distribution of predicted LTV within UNICEF Australia's donor base.

Segment	Conversion Uplift	Cost Change	ROI Impact
Top	+20%	+10% spend	+40% uplift
High	+20%	+10% spend	+40% uplift
Medium	+10%	Neutral	+15% uplift
Low	Stable	–30% spend	Budget reallocated

Table 14: Projected impact by donor segment

4.5 Estimated Costs and Budget

Channel budget Per-donor caps and volumes imply the following spend and expected return envelope:

Channel	Yearly Cost per Donor (AUD)	Donors	Est. Spend (AUD)	Est. LTV Return per Donor (AUD)
Face-to-face	$100.00	145	$14,500	$2,421.83
Telemarketing (4 dials)	$40.00	2,891	$115,640	$513.72
Direct Mail (3 waves)	$7.50	5,619	$42,143	$148.97
Email/SMS (12 sends)	$1.20	12,523	$15,028	$15.34
Totals		**21,178**	**$187,310**	*See impact envelope*

Table 15: Estimated costs and budget allocation by donor segment

The per–channel unit costs are taken directly from the UNICEF stakeholder Q&A brief, and the expected returns are generated from our two–stage LTV model. In line with established fundraising evidence, conservative uplift ranges are applied and prioritise high social interaction channels for high propensity segments while maintaining low cost digital coverage for scale. The table should therefore be read as a *supplemental* allocation layered on top of business–as–usual (BAU) activity: an incremental envelope of $187,310 directed to 21,178 treated donors with projected 24-month returns of $2.65–$3.05m for that cohort (i.e., an incremental revenue multiple of roughly 14–16× on spend). The logic is intentionally practical: higher unit-cost channels (F2F/TM) are restricted to donor segments where early behaviour and contactability indicate strong lift, while Direct Mail and Email/SMS provide broad, lower-cost reinforcement to compress time-to-second-gift and encourage regular-giving.

The yearly cost per donor is constructed from empirically observed contact frequencies and feasible timing windows for each channel, using impact pleas in December and June, 28-48 day impact pleas for early frequency, as well as additional quarterly DM waves. The inclination for donors to donate in these periods within the framework of the financial cycle is shown below in Figure 9.

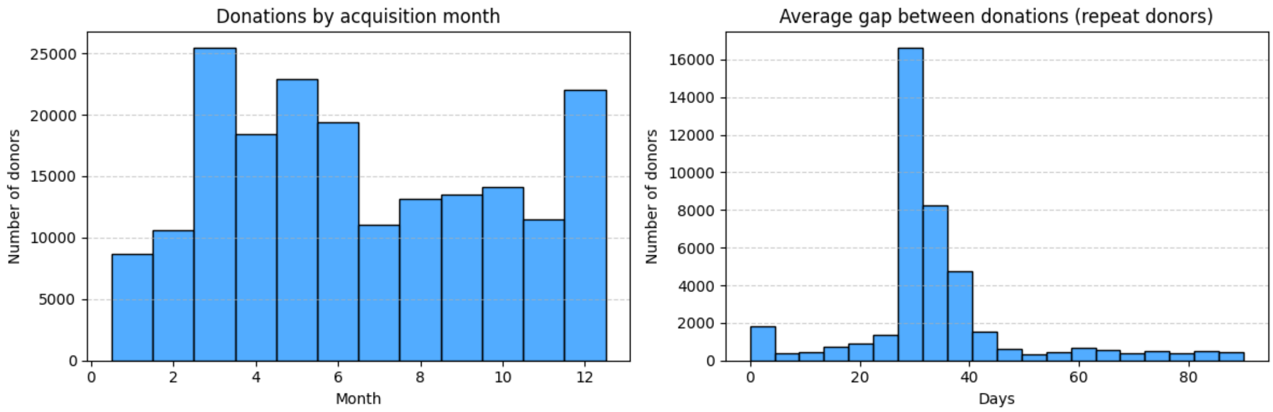


Figure 9: Seasonal Shifts and Return Giving Windows

These frequencies are not fixed tariffs; they are starting estimates to be tuned by UNICEF through test-and-learn through A/B tests on DM cadence, and contact capture prompts in web flows. In production, the cadence should be governed by measured marginal ROI at the segment–channel level, with clear *stop* rules to suppress continuing marketing after conversion or nonresponsiveness to contain costs and avoid fatigue for potential donors and UNICEF resources.

Contextually, the envelope is modest relative to UNICEF Australia's overall finances. In 2023, UNICEF Australia reported total revenue of **$64.3m** and total expenditure of **$62.7m**, including **$16.27m** in fundraising costs (Australian Committee for UNICEF Limited, 2023). The proposed incremental spend of **$187,310** is therefore about **1.2%** of 2023 fundraising costs and roughly **0.3%** of total expenditure. This is well within reason for an optimisation scheme as suggested, with room for targeted testing and potential scale-up. This brief reallocates a small, risk-contained budget toward segments and channels where the model shows the highest probability of positive LTV.

References

Aristotle. (2007). *On rhetoric: A theory of civic discourse* (G. A. Kennedy, Trans.; 2nd). Oxford University Press.

Australian Committee for UNICEF Limited. (2023). *Annual financial report – 31 december 2023* (Statement of Comprehensive Income). Australian Committee for UNICEF Limited. Sydney, Australia.

Bekkers, R., & Wiepking, P. (2011). A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving. *Nonprofit and voluntary sector quarterly*, *40*(5), 924–973.

Bergstra, J. (2012, July). Grid search is a bad hyper-parameter optimization algorithm [Available online].

Blackbaud Institute. (2023). *Charitable giving report 2023: How nonprofits raised money in a changing world*. Blackbaud Inc. Charleston, SC.

Bult, J. R., & Wansbeek, T. (1995). Optimal selection for direct mail. *Marketing science (Providence, R.I.)*, *14*(4), 378–394.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Cialdini, R. B. (2009). *Influence: Science and practice* (5th). Pearson.

Conrad, F., Mälzer, M., Schwarzenberger, M., Wiemer, H., & Ihlenfeldt, S. (2022). Benchmarking automl for regression tasks on small tabular data in materials design. *Scientific Reports*, *12*, 19350. https://doi.org/10.1038/s41598-022-23327-1

Crowley, S., & Hawhee, D. (2012). *Ancient rhetorics for contemporary students* (5th). Pearson.

(FEP), F. E. P. (2025, April). *Quarterly fundraising report™: Year-to-date nonprofit sector trends, q1 2025 (jan 1, 2025 – mar 31, 2025)* (tech. rep.) (Report No. Q1 2025). Fundraising Effectiveness Project. https://publications.fepreports.org/

Fisher, W. R. (1984). Narration as a human communication paradigm: The case of public moral argument. *Communication Monographs*, *51*(1), 1–22. https://doi.org/10.1080/03637758409390180

Foundation, C. A. (2025). Uk giving report 2025. https://www.cafonline.org/insights/research/uk-giving-report

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in Neural Information Processing Systems 35 (NeurIPS 2022) – Datasets & Benchmarks Track*, —. https://proceedings.neurips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf

Hancock, J. T., & Khoshgoftaar, T. M. (2020). Catboost for big data: An interdisciplinary review. *Journal of Big Data*, *7*(1), 94. https://doi.org/10.1186/s40537-020-00369-8

Jameson, T. (2017). *Communication strategies as drivers of nonprofit donor retention* [Doctoral dissertation, Walden University]. https://scholarworks.waldenu.edu/dissertations/3893/

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer.

Salesforce.org. (2022). *Nonprofit trends report 2022*. Salesforce Inc. San Francisco, CA.

Sargeant, A. (2001). Relationship fundraising: How to keep donors loyal. *Nonprofit management & leadership*, *12*(2), 177–192.

Sargeant, A., & Jay, E. (2004). Reasons for lapse: The case of face-to-face donors. *International journal of nonprofit and voluntary sector marketing*, *9*(2), 171–182.

Sargeant, A., & Jay, E. (2014). *Fundraising management: Analysis, planning and practice* (3rd). Routledge.

Sargeant, A., & Woodliffe, L. (2007). Building donor loyalty: The antecedents and role of commitment in the context of charity giving. *Journal of Nonprofit & Public Sector Marketing*, *18*(2), 47–68. https://doi.org/10.1300/J054v18n02_03

Shang, J., & Croson, R. (2009). A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods. *Economic Journal*, *119*(540), 1422–1439. https://doi.org/10.1111/j.1468-0297.2009.02267.x

Shmuel, A., Glickman, O., & Lazebnik, T. (2024). A comprehensive benchmark of machine and deep learning across diverse tabular datasets [Available at https://arxiv.org/abs/2408.14817]. *arXiv preprint arXiv:2408.14817*.

To, S. H. (n.d.). Tweedie distribution: Definition and examples [Accessed: 2025-10-27]. %5Curl%7Bhttps://www.statisticshowto.com/tweedie-distribution/%7D

5 AI Usage Acknowledgement

Various Large Language Models (LLMs) were utilised throughout the study to assist with debugging, generate plots, and support general problem-solving and documentation tasks. These models served as auxiliary tools to streamline development, verify results, and enhance the overall efficiency of the workflow.