THE UNIVERSITY OF SYDNEY

COVER PAGE

TITLE: EDA Analysis for Lifetime Value Prediction (LTV)

SID:530002502

Table of Contents

1.0 Executive briefing page 1

1.1 Background

The funding of UNICEF Australia programs for children depends on voluntary donations from supporters. The research examined 192,000 donors who made at least one donation to analyze their behavior patterns from their initial contribution through the first three months. The first 90 days following the first donation period shows the most significant signs about future supporter contribution levels throughout the next two years which represents the LTV horizon.

Business challenge and key questions; The team needs to understand which early donor behaviours indicate future high-value donors because fundraising budgets remain limited. The analysis addressed four essential questions.

- Which early behaviours point to higher 24-month value?
- Which acquisition channels and first products bring in higher value donors?
- Which audience groups (e.g., age, MOSAIC segments) deliver more value?
- What practical actions in the first 90 days are most likely to lift long-term value?

1.2 Key findings

1. Early monthly giving is the strongest signal. Donors who start Regular Giving (RG) in the first 90 days are worth several times more over two years. First product "RG – Global Parent" has the highest LTV; one-off cash and Inspired Gifts are far lower.
2. More early activity predicts higher value. Multiple gifts in the first 90 days consistently lift LTV. Welcome streams should accelerate a second gift, not just deliver thanks.
3. First channel matters. Telemarketing, Face-to-Face, and Door-to-Door deliver higher LTV than Web or Unsolicited. Web donors need rapid phone follow-up.
4. Older supporters give more. LTV rises with age, while "Unknown" age donors underperform—pointing to data gaps.
5. The analysis reveals that MOSAIC and seasonality produce smaller effects than channel and product variables but should be used for detailed optimization.
6. The ability to contact donors through both phone and email leads to higher donation amounts because it allows for follow-up communications and regular giving program enrollment.

1.3 Recommendations

1. Make monthly giving the default ask. Present monthly options on thank-you pages and welcome emails; follow with a call within 7–14 days for single-gift joiners to discuss monthly giving.
2. Design for a quick second gift. Introduce a 30–45 day "impact update + soft ask," plus an emergency or program story, to move one-gift donors to two-plus gifts in the 90-day window.
3. Prioritise higher-value acquisition. Sustain Telemarketing/F2F/D2D for quality; for Web, add phone capture and same-week call-backs to lift LTV.
4. Strengthen data at sign-up. Capture age, phone, and email at first contact with a simple permission prompt to enable the follow-up steps above.
5. Use Mosaic and season to fine-tune. Apply Mosaic to adjust messaging and calling strategy after the right channel and monthly-giving offer are set.

1.4 Areas for further investigation

- A/B testing of monthly-first flows and call timing and second-gift content will be conducted.
- Contact capture: Pilot incentives or design tweaks to lift phone/email capture.
- Mosaic should only be used for audience depth analysis after proving its ability to deliver incremental results.
- The predictive model will evaluate donor behavior during the initial stages to enable large-scale journey routing.

2.0 Problem description

The funding operations of UNICEF Australia depend on donations from individual supporters to support their programs. The commercial challenge requires organizations to determine donor contact levels and regular giving (RG) nurturing strategies for supporters who make their first donation as soon as possible. The organization faces limited financial resources and contact preference restrictions while dealing with donors who show different levels of lifetime value because most supporters stop giving shortly after their first donation, but a few donors generate most of the revenue. The business requires an evidence-based method to predict donor value over two years through initial 90-day indicators which include gift patterns and values and acquisition methods and contact information and demographic data and MOSAIC segmentation based on postcode. The ability to predict donor value in the early stages will determine the best approach for channel distribution and regular giving invitations and stewardship programs and ask amounts and reveal missing data that limits targeting efforts. The organization tracks its success through improved regular giving conversion rates and retention of high-value segments and reduced fundraising expenses and minimized unnecessary contact efforts which enable more children to receive UNICEF program support. Prior research shows that donor commitment and the quality of early relationship experiences are strong antecedents of loyalty and long-run giving, so identifying high-potential supporters quickly is critical (Sargeant & Woodliffe, 2007).

3.0 Data pre-processing

3.1 Data sources and structure

The dataset from UNICEF contains 1,780,140 donation records spanning from 2014 to 2023 which include donor profiling information about dates and amounts and product and channel descriptors and demographic data and contact availability. The dataset received additional postcode-level MOSAIC attributes (Group, Type) from USYD_ART.csv through a 4-digit postcode standardization process. The data diagnostics show that 93.55% of the records match correctly while maintaining the original number of rows but 5.84% of postcodes are invalid and GiftAmount has no missing values and only two entries have non-positive values.

```
AFTER CLEANING: key stats
  - postcode_missing_%: 5.84
  - giftamount_na: 0
  - giftamount_le0: 2
  - rows: 1780140
  - mosaic_match_%: 93.55
```

	value
postcode_missing_%	5.84
giftamount_na	0.00
giftamount_le0	2.00
rows	1,780,140.00
mosaic_match_%	93.55

Business relevance. With 93.55% MOSAIC coverage, geodemographic analyses are representative for most donors; the remaining ~6.45% form a natural “unknown” segment that should be retained to avoid overstating MOSAIC-linked effects in later EDA.

3.2 Cleaning and standardisation

Identifiers and duplicates. The process involved duplicate removal of full rows which were detected and the system maintained the first GiftDate entry from duplicate Gift_ID records. The total number of rows in the above table remained constant because the duplicate removal process did not affect the overall count.

The dates and chronological order of events; The system processed donation dates by placing the day first while min/max coverage analysis verified the time period for cohort and seasonal studies. The system kept

all 1,436 rows that had GiftDate values before First_GiftDate for complete data retention. The 90-day window for Donor-level features will start from the actual date of their first GiftDate.

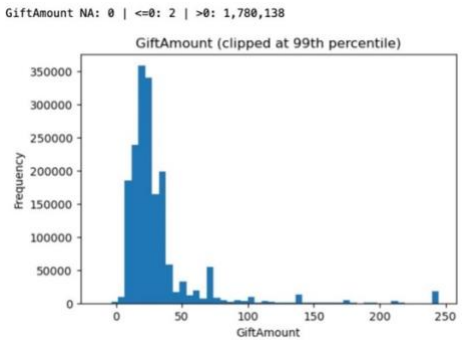
```
GiftDate earlier than First_GiftDate (count): 1436

Date coverage (min/max):
```

	First_GiftDate	First_CashDate	First_RecurringDate	GiftDate
min	2014-01-01	2014-01-01	2014-01-15	2014-01-01
max	2022-12-09	2025-12-06	2025-07-30	2024-12-12

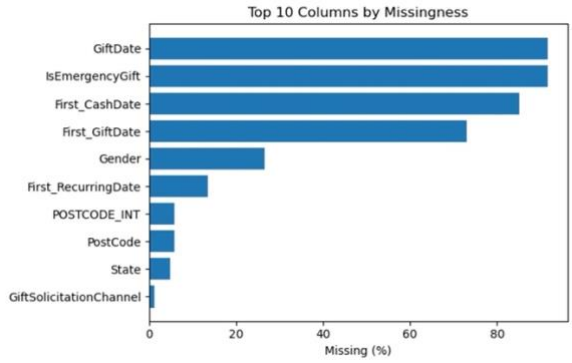
.

Monetary fields; *GiftAmount* is complete (0 missing) with only 2 non-positive entries. For display, values are clipped at the 99th percentile, though raw data remain intact.



Categorical normalisation; The data underwent standardization through whitespace removal and sentinel value positioning. The variables Age_Bucket, Gender, Have_Email/Phone, ProductType_Group and AppealSeason (C3/C11) maintained their "Unknown" values to avoid artificial data insertion and enable analysis of non-disclosure behavior.

Missing values; The system generated a complete missingness table which displayed missing data proportions in order and included a brief Top N chart for reporting purposes. The fields which pertain to Regular Giving show high rates of structural missing data because they include First_RecurringDate but monetary core fields remain complete. The presentation uses the Top N view but all variables with any missing data appear in the complete table to prevent any variable from being ignored.



Business relevance. "Unknown" segments in contact availability signal data capture opportunities to improve engagement and lifetime value, while structural missingness in Regular Giving dates is informative for segmentation, not imputation.

3.3 External enrichment and join process.

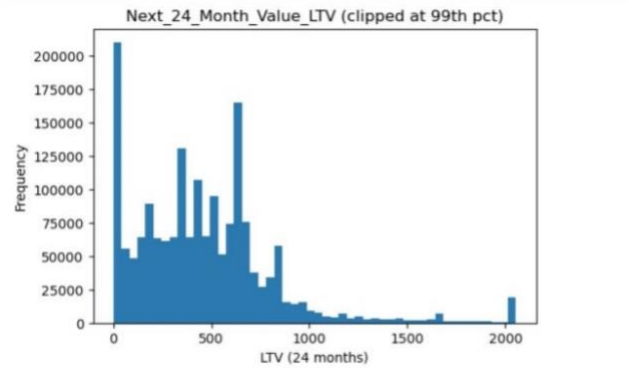
Postcode standardisation; The POSTCODE_INT numeric key was generated from PostCode data through validation of 0–9999 ranges which resulted in 5.84% of postcodes being invalid or missing thus determining the maximum potential for enrichment.

MOSAIC join; The POSTCODE_INT column in USYD_ART.csv received DOMINANT_MOSAIC_GROUP/TYPE data through a left join operation. The analysis shows that 93.55% of the records match and all rows were kept in the dataset. The “Mosaic Australia 2024 – Grand Index” workbook provided readable names for MOSAIC codes which were verified through inspection and count verification.

Business relevance; postcodes are classified as “Unknown MOSAIC” to maintain unbiased results. The high coverage rate of 93% allows for geodemographic profiling through enrichment but the remaining unmatched.

3.4 Transformations for analysis readiness

Type optimisation; The Initial_90_Days_LTV, Pre_RG_Donation. ConvertedTo_RG and ConvertedTo_RG_Within_6M binary indicators received Int8 nullable type conversion because they contained only 0 and 1 values. The LTV target and all amount values received float32 data type conversion for memory optimization purposes and descriptors were transformed into categories. The POSTCODE_INT key enables consistent data enrichment operations and MOSAIC name fields enhance stakeholder understanding through code-to-name transformations. The visualization displays histogram with 99th percentile limits for better understanding of imbalanced data points but maintain the original data values. The combination of specific dtypes and labels helps users avoid misinterpretation while allowing exact segmentation and maintains "Unknown" as a category to track operational data gaps.



4.0 Descriptive statistics

4.1 Scope and sample

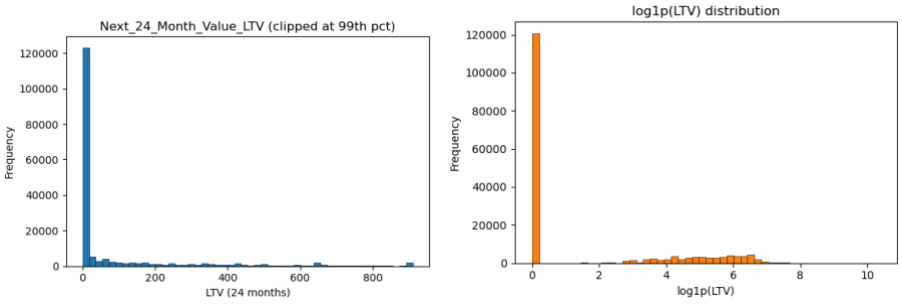
The analysis examines donors who show complete activity during their first 90 days because this period matches the prediction model used to calculate Lifetime Value (LTV). The donor level table shows 191,795 distinct donors with no missing target information while the leakage safe EDA subset includes 174,070 donors who received the Initial_90_Days_LTV = 1 flag.

LTV_24m	count	mean	std	min	25%	50%	75%	max
174,070.00	68,324.00	1.03	0.23	1.00	1.00	1.00	1.00	13.00
93.54	68,324.00	68.26	198.55	-0.07	21.00	35.00	70.00	14,000.00
281.77	68,324.00	65.89	186.10	-0.07	21.00	35.00	68.25	11,200.00
0.00	68,324.00	65.86	186.10	-0.07	21.00	35.00	68.19	11,200.00
0.00	68,324.00	66.52	187.62	-0.07	21.00	35.00	69.30	11,200.00
52.50	68,324.00	0.52	0.52	0.00	0.00	1.00	1.00	10.00
343.00	68,324.00	0.58	0.55	0.00	0.00	1.00	1.00	10.00
586.67	68,324.00	0.21	0.41	0.00	0.00	0.00	0.00	4.00
910.43	1,185.00	41.29	25.98	1.00	29.00	34.00	62.00	90.00
31,645.84	68,324.00	65.88	186.45	-0.07	21.00	35.00	68.60	11,200.00

4.2 Response variable: Next-24-Month LTV

Shape and scale; The distribution of LTV shows strong right-skewed characteristics with most values concentrated at zero. The EDA subset containing 174,070 observations shows that the median value equals 0.00 while the 75th percentile reaches 52.50 and the 90th percentile reaches 343.00 and the 95th percentile reaches 586.67 and the 99th percentile reaches 910.43 and the maximum value reaches 31,645.84 and the mean value equals 93.54 with a standard deviation of 281.77. The statistics demonstrate that most of the total value originates from a limited number of donors.

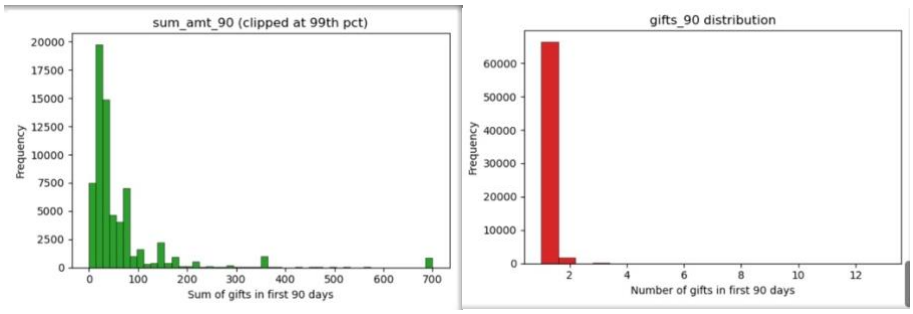
Implication for analysis; The analysis requires robust summary measures such as median and percentiles instead of averages because of the extreme distribution skewness and zero-value concentration. The raw and log histograms demonstrate these characteristics.



4.3 Early 90-day behaviour (predictor distributions)

Availability; The first 90-day behavioral data exists for 68,324 donors while 1,185 donors made their second donation between 29 to 62 days after their first gift with an average time of 41.29 days to their second donation.

Frequency and amount; The majority of donors give only one donation during their first 90 days of support (gifts_90 mean = 1.03; 25–75%: 1–1). The total value of donations during the first 90 days tends to be small (sum_amt_90 median = 35.00; IQR 21.00–70.00). The histograms display the expected right-skewed distribution with distinct intervals.



4.4 Numeric relationships with LTV (linear and non-linear)

Linear/monotonic associations (correlations and slopes).

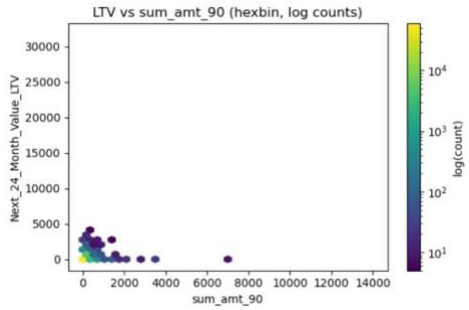
- Regular-giving activity inside 90 days is the strongest signal: `rg_gifts_90` shows Pearson r = 0.28, Spearman ρ = 0.48, and a significant slope of +217.16 LTV per additional RG gift (all $p < 0.001$).
- Frequency matters: `gifts_90` r = 0.17, slope +239.36 ($p < 0.001$).
- Amount is weaker than frequency: `sum_amt_90` r = 0.11, slope +0.17 ($p < 0.001$).
- Emergency/Cash emphasis is negatively associated with LTV: `emergency_gifts_90` r = −0.13 (slope −79.47), `cash_gifts_90` r = −0.04 (slope −24.59), both $p < 0.001$.
- Timing to second gift is weakly negative (`days_to_second_gift_90` r = −0.07, p = 0.01), consistent with faster repeat supporting higher value.

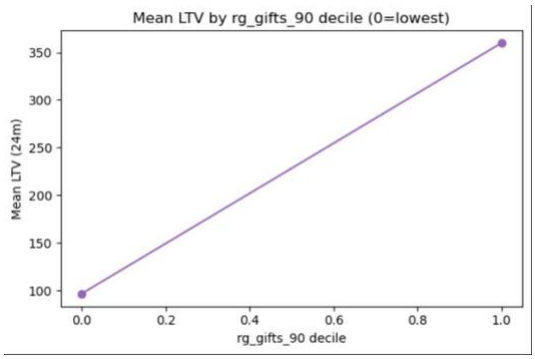
These results are summarised in the correlation/slope table and confirmed by the hexbin plot for large-N scatter (to mitigate over-plotting).

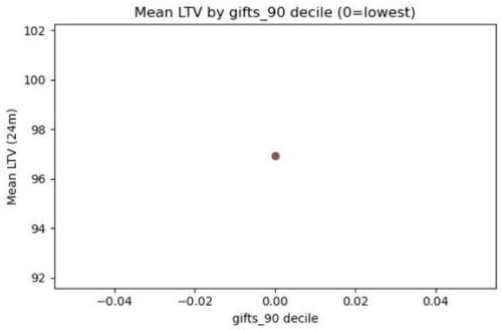
Non linear patterns (deciles); The decile profiles of rg_gifts_90 and gifts_90 demonstrates minimal change across middle deciles but a significant increase in the top decile which reveals a steady increase in values throughout the distribution. The sum_amt_90 data shows threshold behaviour because of spending patterns among early high spenders.

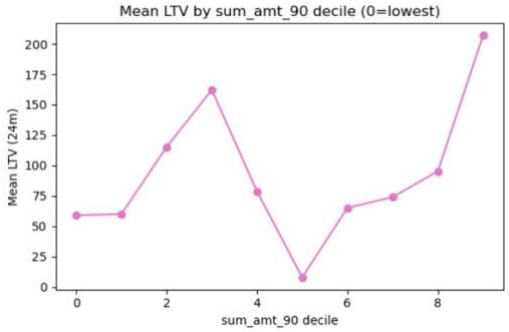
Fig4.4:

feature	n	pearson_r	pearson_p	spearman_rho	spearman_p	linreg_slope	linreg_p
rg_gifts_90	68,324.00	0.28	0.00	0.48	0.00	217.16	0.00
gifts_90	68,324.00	0.17	0.00	0.11	0.00	239.36	0.00
sum_amt_90	68,324.00	0.11	0.00	-0.05	0.00	0.17	0.00
max_amt_90	68,324.00	0.09	0.00	-0.06	0.00	0.15	0.00
mean_amt_90	68,324.00	0.08	0.00	-0.07	0.00	0.14	0.00
first_gift_amount_90	68,324.00	0.08	0.00	-0.07	0.00	0.13	0.00
median_amt_90	68,324.00	0.08	0.00	-0.07	0.00	0.13	0.00
cash_gifts_90	68,324.00	-0.04	0.00	-0.16	0.00	-24.59	0.00
days_to_second_gift_90	1,185.00	-0.07	0.01	-0.04	0.20	-3.80	0.01
emergency_gifts_90	68,324.00	-0.13	0.00	-0.26	0.00	-79.47	0.00









4.5 Categorical drivers (group summaries and tests)

Product at first gift (90 days); The first product group of Regular Giving (RG) donors results in higher LTV values than Cash – One off and Inspired Gifts donors. The LTV mean values for RG – Global Parent donors reach 271.16 while RG – Paddington donors reach 230.40 but Cash – One off donors reach 64.55 and Inspired Gifts donors reach 37.41. The Kruskal–Wallis test shows a substantial practical effect ($H = 17{,}224.88$, $\eta^2 \approx 0.252$).

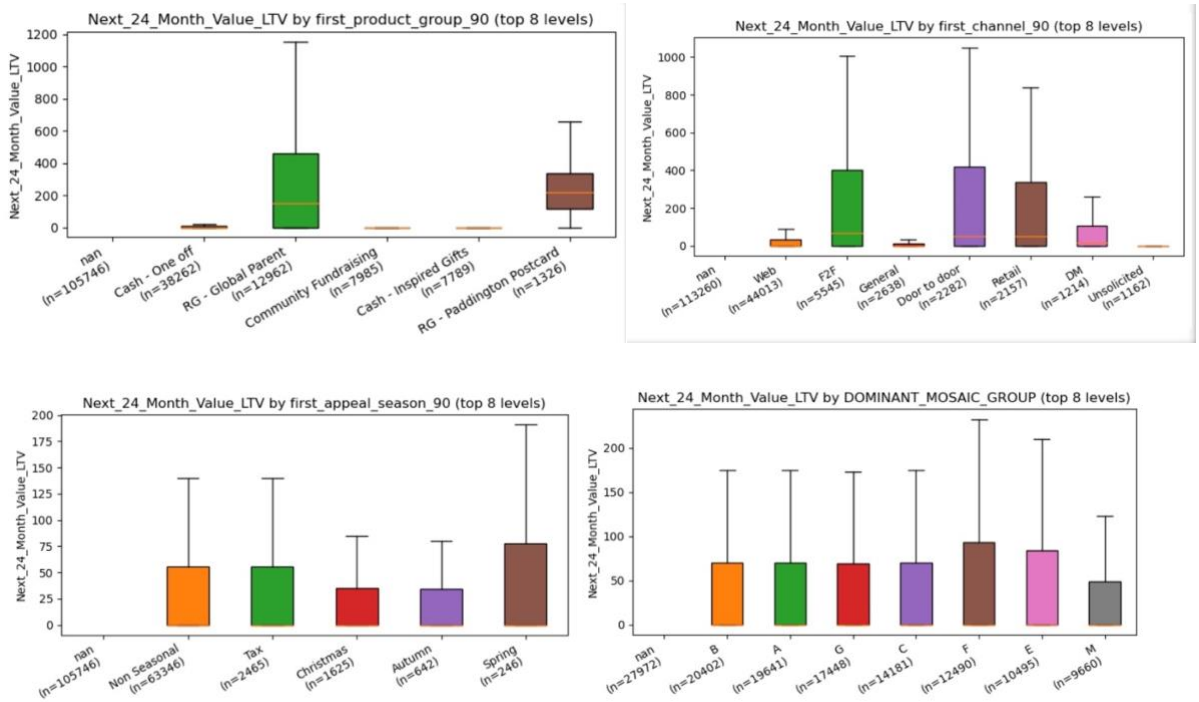
Solicitation channel at first gift (90 days); Channels associated with direct recruitment perform strongly: Telemarketing mean = 231.18, Door-to-Door = 219.58, F2F = 211.67, Retail = 192.05; Web has a lower mean (85.41), and Unsolicited is lowest (8.58). The overall effect is moderate ($H = 4{,}647.32$, $\eta^2 \approx 0.077$).

Appeal season of first gift (90 days); Differences are statistically detectable (H = 16.25, p ≈ 0.0027) but negligible in magnitude ($\eta^2 \approx 0.00$); practically minor relative to product and channel.

Age; LTV rises with age; means are 356.94 (61–70), 340.87 (51–60), 338.53 (71+), then 292.49 (41–50), 204.37 (31–40), 137.07 (19–30). The effect is large (H = 29,291.51, $\eta^2 \approx 0.168$). The "Unknown" segment is substantial and low-value (mean 59.37), highlighting a data-capture gap with likely business relevance.

MOSAIC group (postcode level); The large H value of 265.53 indicates statistical significance but the small effect size of $\eta^2 \approx 0.002$ and zero medians across groups show that the model does not effectively discriminate between groups at this level. The model maintains its utility for both MOSAIC Type level analysis and when used in combination with other models.

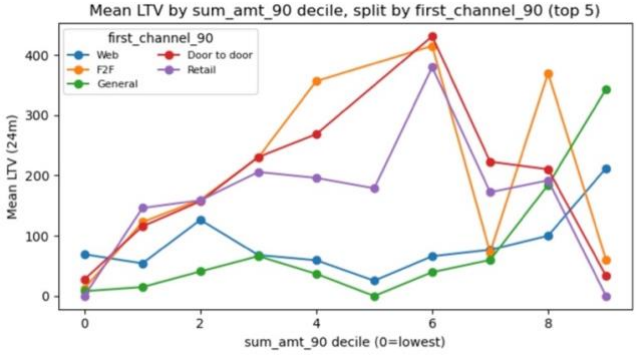
Fig.4.5.



4.6 Interaction: early spend × first channel

The early spend deciles of Mean LTV show growth across all channels but F2F and Telemarketing produce more significant increases at the highest deciles than Web does. The results confirm previous findings because channels that focus on frequent donations and early donor retention produce substantial long-term value.

Fig 4.6:



4.7 Key observations and implications for UNICEF

Value concentration; The top donors who give zero in the following 24 months represent at least half of all donors yet they generate most of the value. The program metrics and dashboards need to show percentile-based views and top decile growth instead of using averages as the primary measurement.

The first 90 days of regular giving activity together with RG signals create the most significant impact on LTV. The LTV depends most on the number of gifts and the regular giving activity during the first 90 days rather than the size of the first donation. The success of LTV depends on two essential factors which are the quick conversion of RG donors and the fast acceleration of their second donation.

Channels differ materially; Telemarketing, Door-to-Door and F2F donors exhibit higher LTV than Web/Unsolicited; prioritising RG-oriented recruitment channels, or applying targeted follow-ups for Web entrants, can lift long-run value.

The age distribution shows a clear pattern while MOSAIC segments demonstrate minimal differences between each other. The value of older donors exceeds that of other groups while MOSAIC Group segments fail to create meaningful distinctions at this level of analysis which indicates that geodemographic applications should evaluate Type segments or their interactions with channel and 90-day behaviour.

Data capture matters; Large “Unknown” segments (e.g., Age) coincide with lower LTV; improving capture of basic descriptors (and contact channels) is likely to pay off via better engagement.

5.0 Exploring potential relationships

The analysis in this section examines the relationship between early window behavior during the first 90 days and basic descriptors with the 24-month lifetime value (LTV). The analysis of 191,795 donors who have Initial_90_Days_LTV = 1 is conducted on the donor level table that was previously assembled. The analysis employs robust distribution-aware statistics and visualizations because both the predictor variables and the LTV target variable show strong skewness and contain zero values.

5.1 Numerical–numerical relationships

Approach

Pairwise associations between LTV and the 90-day behavioural features were assessed using:

- Pearson r (linear association) and Spearman ρ (rank/monotone association),
- Simple OLS slope estimates with p-values (for interpretability in original units), and
- Non-parametric trend checks via decile profiles (mean LTV by predictor decile) to visualise non-linearities.

All metrics are summarised in Fig. 5.1. The underlying table and top findings show where features are sorted by Pearson r.

Fig.5.1:

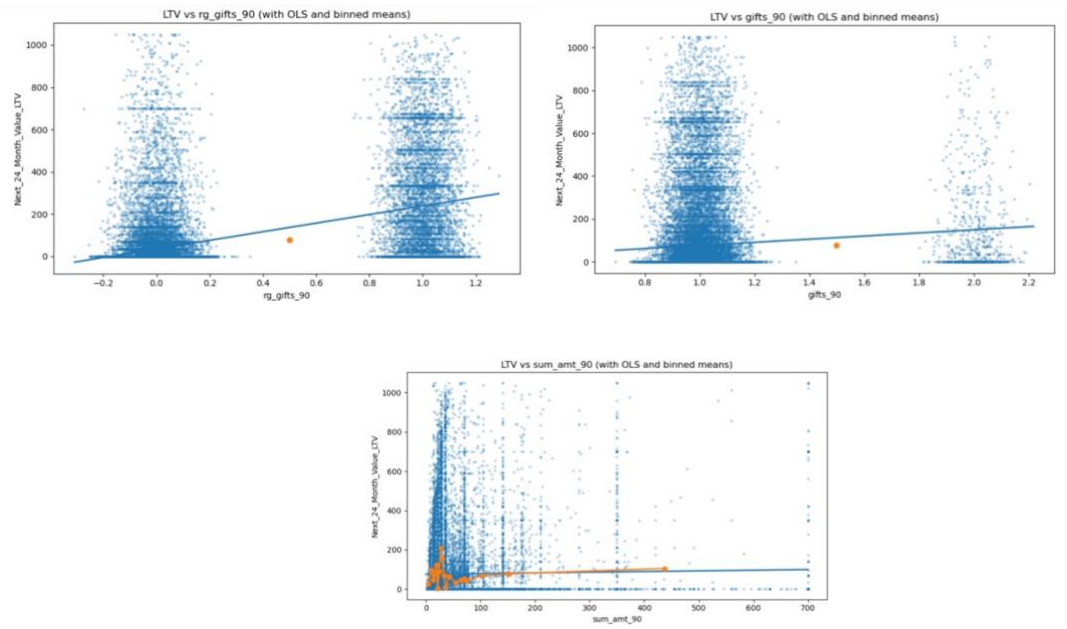
feature	n	pearson_r	pearson_p	spearman_rho	spearman_p	ols_beta	ols_p
rg_gifts_90	68,324.00	0.28	0.00	0.48	0.00	217.16	0.00
gifts_90	68,324.00	0.17	0.00	0.11	0.00	239.36	0.00
days_to_second_gift_90	1,185.00	-0.07	0.01	-0.04	0.20	-3.80	0.01
sum_amt_90	68,324.00	0.11	0.00	-0.05	0.00	0.17	0.00
max_amt_90	68,324.00	0.09	0.00	-0.06	0.00	0.15	0.00
mean_amt_90	68,324.00	0.08	0.00	-0.07	0.00	0.14	0.00
median_amt_90	68,324.00	0.08	0.00	-0.07	0.00	0.13	0.00
first_gift_amount_90	68,324.00	0.08	0.00	-0.07	0.00	0.13	0.00
cash_gifts_90	68,324.00	-0.04	0.00	-0.16	0.00	-24.59	0.00
emergency_gifts_90	68,324.00	-0.13	0.00	-0.26	0.00	-79.47	0.00

Key findings

RG activity dominates predictive signal.

- rg_gifts_90 shows the highest correlation with LTV because it reaches Pearson r ≈ 0.28 and Spearman ρ ≈ 0.48 and OLS slope ≈ +217 AUD per additional RG gift with all p < 0.001. The rank correlation value exceeds the linear correlation value which shows that the relationship between variables follows a non-linear pattern that increases in value as donors start regular giving. The analysis includes Fig. 5.1 and Fig. 5.2a (scatter + OLS) and the decile profile (Fig4.4).

Fig. 5.2(a-c)



More gifts help; amount is secondary.

- The number of gifts displayed in the window (gifts_90) shows a positive relationship with LTV (Pearson r ≈ 0.17, p < 0.001; Fig. 4.2). The average LTV shows a general upward trend with increasing frequency of donations although most donors give only once (zero heavy differences appear across the lower deciles).
- The Pearson r value of 0.11 indicates a weak linear relationship between sum_amt_90 and the total amount given in the window (p < 0.001). The decile curve shows a non-monotonic pattern because higher LTV values do not always appear in intermediate deciles which supports the conclusion that early frequency and RG conversion generate more long-term value than early cash amounts. See Fig. 5.2c/4.4.

Product mix within 90 days matters.

- cash_gifts_90 and emergency_gifts_90 correlate negatively with LTV (Pearson r ≈ −0.04 and −0.13, respectively; p < 0.001), while rg_gifts_90 is positive. This pattern suggests cash-only supporters and emergency responders are less likely to generate high 24-month value unless converted to RG. See Fig. 5.1.

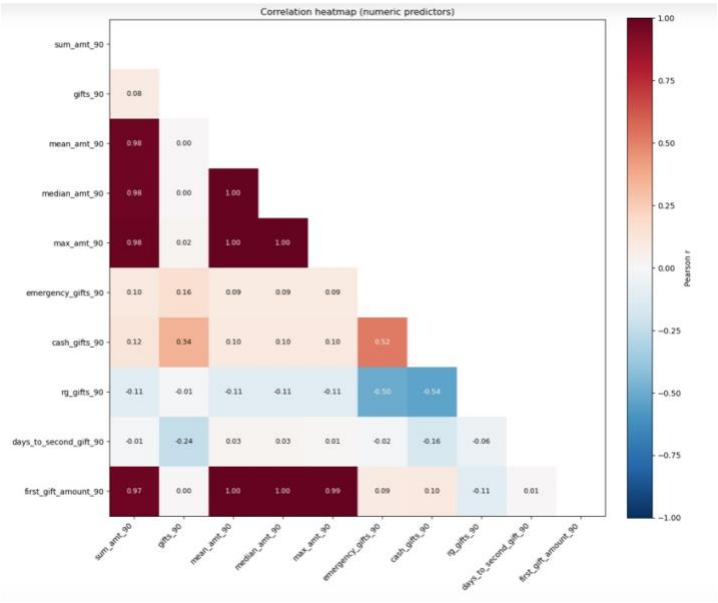
Donation cadence signal is weak but directional.

- days_to_second_gift_90 has a small negative association (Pearson r ≈ −0.07, p ≈ 0.01): shorter time to the second gift is modestly associated with higher LTV. See the scatter/OLS panel and the summary table.

Visual evidence and non-linearity

- Correlation heatmap. The numeric predictors group into three distinct categories which include (i) gift amount statistics and (ii) gift type counts and (iii) cadence. The amount summaries show high correlation (r≈0.97–1.00) because donors tend to make single donations every 90 days. The correlation between cash donations and emergency one-off gifts reaches +0.52 while early regular giving shows a negative relationship of -0.54. The timing of second-gift donations shows no significant relationship with other variables. The model should use one amount proxy while behavior types should be modeled through nonlinear/share-based features.

Fig. 5.3:



- The scatter panels with transformation show zero inflation and long tails but log transformation helps to compress the spread and reveal the monotone trends. The four-panel plot in Appendix A shows the relationship between sum_amt_90 and LTV.
- Decile profiles (Fig. 5.2a–c) show a near-monotone lift in average LTV across rg_gifts_90 and gifts_90 deciles, while sum_amt_90 shows curvature, reinforcing the need for non-linear modelling or transforms/interactions later. The decile figures and points are visible.

To quantify scatter-plot associations, simple OLS was run on log-transformed variables with HC3 standard errors.

- The model shows that log(LTV) is positively related to log1p(rg_gifts_90) with a coefficient of 4.25 (95% CI 4.18–4.32) and p<0.001 and Adj. R^2 = 0.24 and N=68,324. The analysis demonstrates that higher RG frequency during the first 90 days leads to significant growth in two-year value. The use of log1p transformation on counts and log transformation of LTV produces an elasticity value that approximates the coefficient.
- log(LTV) ~ log1p(sum_amt_90): Coefficient –0.005 (95% CI –0.025–0.015), p=0.64, Adj. R^2≈0.00, N=68,324. No significant linear association; result aligns with weak correlations and flat binned means.

--- LINEAR OLS SUMMARY: y_log ~ log1p(rg_gifts_90) ---
OLS Regression Results

Dep. Variable:	y_log	R-squared:	0.240
Model:	OLS	Adj. R-squared:	0.240
Method:	Least Squares	F-statistic:	1.565e+04
Date:	Fri, 12 Sep 2025	Prob (F-statistic):	0.00
Time:	11:54:05	Log-Likelihood:	-1.4894e+05
No. Observations:	68324	AIC:	2.979e+05
Df Residuals:	68322	BIC:	2.979e+05
Df Model:	1		
Covariance Type:	HC3		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	0.9894	0.009	115.521	0.000	0.973	1.006
x	4.2487	0.034	125.118	0.000	4.182	4.315

Omnibus:	6371.754	Durbin-Watson:	1.824
Prob(Omnibus):	0.000	Jarque-Bera (JB):	8292.895
Skew:	0.830	Prob(JB):	0.00
Kurtosis:	3.393	Cond. No.	3.62

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

OLS Regression Results

Dep. Variable:	y_log	R-squared:	0.000
Model:	OLS	Adj. R-squared:	-0.000
Method:	Least Squares	F-statistic:	0.2176
Date:	Fri, 12 Sep 2025	Prob (F-statistic):	0.641
Time:	11:54:05	Log-Likelihood:	-1.5830e+05
No. Observations:	68324	AIC:	3.166e+05
Df Residuals:	68322	BIC:	3.166e+05
Df Model:	1		
Covariance Type:	HC3		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	1.6280	0.038	43.071	0.000	1.554	1.702
x	-0.0048	0.010	-0.466	0.641	-0.025	0.015

Omnibus:	11882.606	Durbin-Watson:	1.552
Prob(Omnibus):	0.000	Jarque-Bera (JB):	13082.050
Skew:	1.013	Prob(JB):	0.00
Kurtosis:	2.298	Cond. No.	16.1

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

The Correlation and OLS checks demonstrate that early spend does not create a linear connection with LTV. The hexbin (Appendix B) demonstrates that donors tend to group at x values below $50 with LTV values near 0 while most donors achieve higher LTV values when their sum_amt_90 reaches higher levels. The striping pattern reveals how price and count restrictions affect the data while the fan-shaped distribution of variance shows that the error remains unstable. The data patterns confirm the need for log1p/decile transformations and the evaluation of channel/product interaction effects.

Business interpretation; Early regular giving is the strongest lever: even a few RG gifts in 90 days drive disproportionate value. Early frequency outweighs amount unless tied to RG, supporting RG conversion and second-gift acceleration.

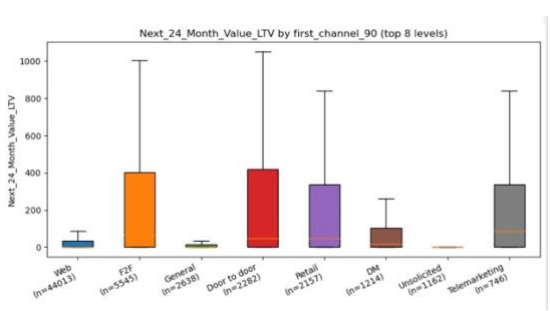
5.2 Categorical–numerical relationships

The Kruskal–Wallis tests were applied to evaluate LTV distribution differences between categories because LTV data follows a non-normal distribution with numerous zero values. The test results appear directly under each figure in the following section. The Mann–Whitney U tests (Appendix C) with Holm correction were performed for top categories to calculate rank biserial r effect sizes.

First solicitation channel (within 90 days)

- Boxplots (Fig. 5.4a) show substantial separation across channels: Telemarketing, Door-to-Door/F2F, and Retail display much higher medians and broader upper tails than Web, General, Unsolicited. The Kruskal–Wallis test is highly significant ($p < 0.001$), confirming heterogeneous LTV distributions by channel.
- Mann–Whitney U tests show channel differences (Appendix C): *F2F* and *Door-to-door* outperform *Web* by ~A$110–A$135 in mean LTV (Holm-adjusted $p < 0.001$, $r \approx 0.22–0.39$, small–moderate).

Fig.5.4a.



first_channel_90: H=4647.32, p=0, eta²=0.077

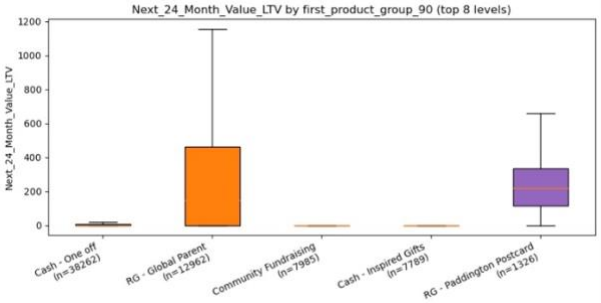
count	mean	median	kruskal_H	p_value	eta2
746.00	231.18	86.80	4,647.32	0.00	0.08
2,282.00	219.58	49.00	4,647.32	0.00	0.08
5,545.00	211.67	68.60	4,647.32	0.00	0.08
2,157.00	192.05	49.00	4,647.32	0.00	0.08
155.00	154.13	0.00	4,647.32	0.00	0.08
1,214.00	113.18	17.50	4,647.32	0.00	0.08
598.00	106.66	0.00	4,647.32	0.00	0.08
2,638.00	100.70	0.00	4,647.32	0.00	0.08
125.00	88.36	0.00	4,647.32	0.00	0.08
44,013.00	85.41	0.00	4,647.32	0.00	0.08

- Implication: channel choice at first contact is not neutral. Channels designed to secure commitments (telemarketing/F2F) are more likely to seed long-term value than passive online acquisition, even controlling for the same 90-day window.

First product group in the window

- The boxplots in Fig. 5.4b demonstrate that RG Global Parent donors achieve higher median values than Cash One off and Inspired Gifts. The Kruskal–Wallis test produces a highly significant result at $p < 0.001$.
- The Mann–Whitney U tests in Appendix C demonstrate that RG – Global Parent produces a substantially higher 24-month LTV than Cash – One off and Cash – Inspired Gifts (Holm adjusted $p < 0.001$, rank biserial $r \approx 0.5–0.9$, large).

Fig.5.4b.



first_product_group_90: H=17224.88, p=0, eta²=0.252

count	mean	median	kruskal_H	p_value	eta2
12,962.00	271.16	147.00	17,224.88	0.00	0.25
1,326.00	230.40	219.71	17,224.88	0.00	0.25
38,262.00	64.55	0.00	17,224.88	0.00	0.25
7,789.00	37.41	0.00	17,224.88	0.00	0.25
7,985.00	5.02	0.00	17,224.88	0.00	0.25

- **Implication:** securing a **regular giving product** within 90 days is a strong marker for high LTV, aligning with the numeric results in §4.1.

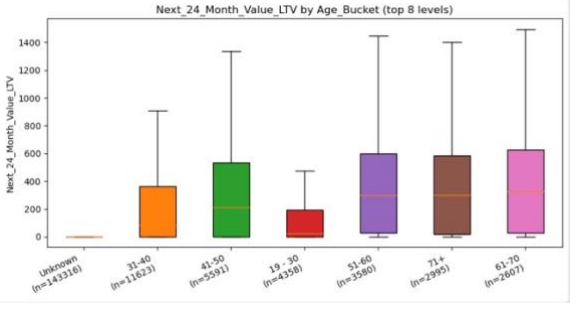
Appeal season at first gift

- The boxplots (Fig 4.5) show that seasonal patterns have a small impact on the data compared to channel and product variations ($p < 0.01$ but with minimal practical differences in median values).
- The results indicate that promotional events like Christmas and Tax holidays increase sales, but they do not automatically result in higher customer lifetime value unless the customer makes a return purchase.

Demographics and contactability

- Age_Bucket: Older donors (51–70+) exhibit materially higher medians than younger or Unknown (see Fig. 5.4c with Kruskal–Wallis $p < 0.001$).

Fig.5.4c.



Age_Bucket: H=29291.51, p=0, eta²=0.168

count	mean	median	kruskal_H	p_value	eta2
2,607.00	356.94	327.60	29,291.51	0.00	0.17
3,580.00	340.87	300.30	29,291.51	0.00	0.17
2,995.00	338.53	301.00	29,291.51	0.00	0.17
5,591.00	292.49	210.00	29,291.51	0.00	0.17
11,623.00	204.37	81.90	29,291.51	0.00	0.17
4,358.00	137.07	21.00	29,291.51	0.00	0.17
143,316.00	59.37	0.00	29,291.51	0.00	0.17

- The statistical analysis shows that gender differences are either non-significant or extremely small (Kruskal–Wallis near null).
- Have_Phone / Have_Email; The two flags show positive relationships with LTV medians ($p < 0.001$) because contactability seems to be the main factor rather than causality.

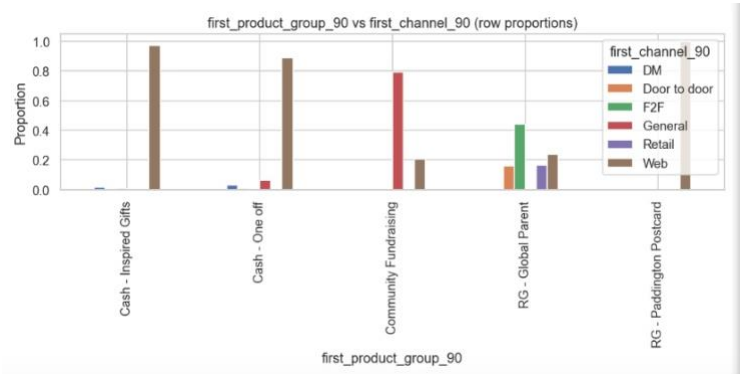
- The State variable shows minimal yet statistically significant differences ($p < 0.05$) which do not reach operational importance when compared to channel and product factors.
- The Mann–Whitney U tests demonstrate that donors between 51 and 70 years old and those 70 and older receive higher LTV than donors between 19 and 30 years old (Holm adjusted $p < 0.001$, $r \approx 0.30$–0.62, moderate–large).

MOSAIC enrichment

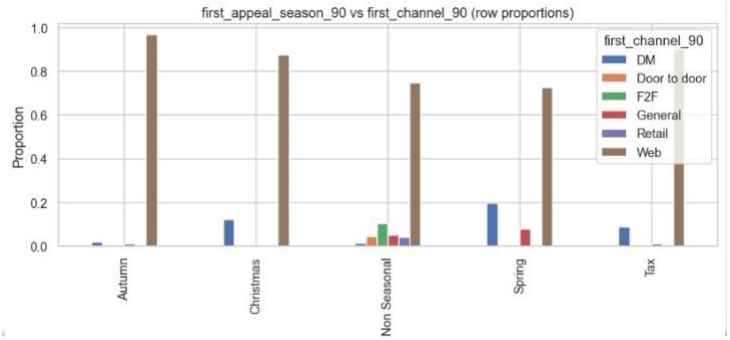
- DOMINANT_MOSAIC_GROUP / TYPE: While statistically significant across many levels (large N), boxplots show limited practical separation and the effect is small compared with solicitation/channel effects (see Fig. 4.5)
- Implication: MOSAIC is useful for fine segmentation and to refine targeting within channel/product, but not a primary driver of LTV by itself.

5.3 Categorical–categorical relationships (assortment/segmentation) page 13

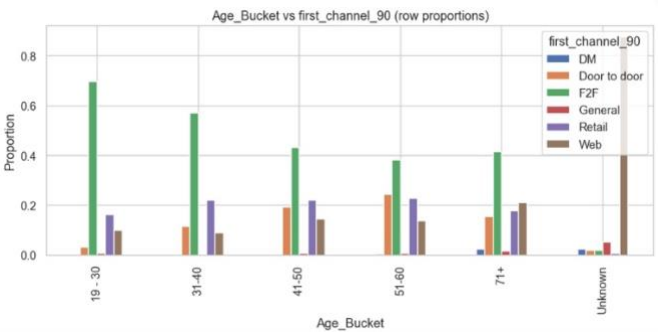
Channel × Product; The analysis through cross tabulations shows that regular giving products mainly appear in channels which focus on commitment-based solicitations (Telemarketing and Door to Door/F2F) yet Cash – One off donations tend to occur through Web/DM channels. The chi square test in Appendix D shows a significant relationship between the variables at $p < 0.001$. The medium range Cramér's V value shows substantial practical dependence between the variables because of the study sample size.



Channel × Appeal season; The expected seasonal appeals such as Tax and Christmas appear during DM/Web periods according to chi square tests which show strong statistical connections ($p < 0.001$). The joint determination of season selection and communication channels means that observed LTV improvements from channels cannot be explained by seasonal factors alone.



Age bucket × acquisition channel; Younger donors between 19 and 30 years old and 31 to 40 years old prefer F2F while Door to Door/Retail becomes more popular with increasing age but Web acquisitions dominate Unknown age donors who were captured digitally. The test statistic printed (Appendix D and Appendix E) shows $\chi^2 \approx 37{,}400$ (df = 25), $p < 0.001$, Cramér's $V \approx 0.36$, i.e., a moderate–strong association.



5.4 Interactions and non-linear effects

The top initiation channels received mean LTV by sum_amt_90 decile analysis to diagnose interactions (Fig. 4.6). The lines show different growth patterns based on channel selection because Telemarketing and Door to Door channels produce greater increases in LTV at higher early sum values yet Web shows consistent results across all deciles. The results show that high early sums combined with commitment-oriented first contact methods produce significant increases in incremental LTV.

The decile profiles in §4 demonstrate curvature for sum_amt_90 while rg_gifts_90 and gifts_90 show near monotone uplifts (non linearity). The practical evidence supports three modelling requirements: (i) the use of non linear terms or transformed amount variables (e.g., log, splines) and (ii) the inclusion of explicit frequency features and (iii) the addition of interaction terms between variables (e.g., channel × product and channel × amount) in subsequent modelling stages.

5.5 Simple regression cross-checks

Simple OLS lines were fit on scatter plots (blue regression lines with binned means). For rg_gifts_90, slope ≈ +217 AUD per gift ($p<0.001$), matching decile lifts (Fig. 5.2a). gifts_90 slope is smaller; sum_amt_90 weak, better handled with rank/nonlinear summaries. Diagnostics derive from regression summary columns (Pearson/Spearman/slopes).

6. Linking the EDA to business decisions

The analysis of statistically important connections from EDA leads to fundraising strategies for UNICEF Australia to achieve its LTV goal of forecasting and expanding donations during the 24 months following the first 90-day period.

6.1 What the distributions imply for strategy

- LTV is highly skewed; the median donor contributes $0, with a long right-tail of higher-value givers. A small minority accounts for most revenue, while typical donors lapse after 90 days. Identifying and nurturing high-yield prospects is the most impactful strategy.
- Analysis universe**:** 191,795 donors with targets; 174,070 meet the "Initial_90_Days_LTV=1" eligibility rule. Forecasting and resourcing should anchor to this reachable base.

6.2 Behaviour in the first 90 days that predicts—and can be influenced

1. • Early Regular Giving is the strongest predictor**;** Higher RG counts in the first 90 days show the largest positive association with 24-month value; converting to monthly giving early is the highest-leverage action.
   Action: Route new donors with any RG activity to upgrade journeys (welcome → upgrade call/EDM in 7–14 days). For cash-only donors making a second gift, trigger an immediate RG ask.
2. • Frequency over size**;** More gifts in the first window predict higher LTV; one-off cash/emergency gifts are neutral to negative once frequency is controlled.
   Action: Optimise journeys for multiple touches (thank-you → impact story → second ask) instead of

single high-value appeals. Create distinct tracks for emergency givers—celebrate impact, then test RG after a cooling period.

3. • Time to second gift matters**;** Shorter days-to-second gift consistently link to higher LTV. *Action:* Introduce a "Second-Gift within 45 days" KPI; automate prompts at 14/28/42 days.
4. • Amount patterns are non-linear**;** LTV does not rise linearly with amount deciles; mid-deciles sometimes underperform, and the top decile is not always strongest once RG is controlled. *Action:* Build combined segments (e.g., {≥2 gifts & any RG} vs. {1 gift & $X spend}); avoid simplistic thresholds based on amount.

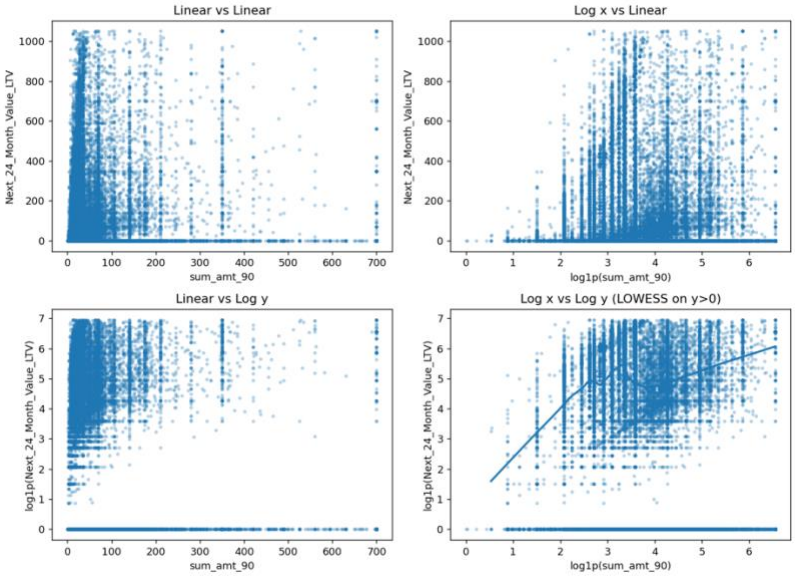
6.3 Channel, product, cohort and MOSAIC insights that guide budget

1. **First contact channel drives future value.**
   - Telemarketing, Door-to-door and F2F yield substantially higher LTV than Web and Unsolicited, with significant differences.
   - **Action:** Acquire or upgrade via people-led channels when donors show "RG readiness" signals. For Web-acquired donors, budget for a TM upgrade sweep in weeks 4–6.
2. **Product at first gift sets trajectory.**
   - RG – Global Parent and RG – Paddington Postcard yield 3–4× the mean LTV of Cash – One-off or Inspired Gifts.
   - **Action:** Bias media and landing experiences toward RG first-product journeys. For cash first-product donors, pre-commit an RG upgrade touch within 14 days.
3. **Seasonality is modest.**
   - First-appeal season matters, but effects are small; Non-Seasonal and Tax outperform Christmas slightly.
   - **Action:** Maintain always-on RG offers and avoid over-reliance on Christmas.
4. **Age matters; yet "Unknown" dominates.**
   - The analysis shows that donors between 51 and 70 years old have higher lifetime value but the system lacks sufficient age information for most donors. The "Unknown" category represents the biggest group, yet it contains donors with the lowest and highest values which indicates a problem with data collection.
   - **Action:** The solution includes two steps: First we should implement simple age indicators through birth year ranges and create a flag for unknown ages and then direct donors without age information to standard streams while testing senior-oriented ads in phone-based channels.
5. **MOSAIC enriches but is secondary.**
   - The results show statistical significance, but the effect size remains small because the distributions share common areas which makes MOSAIC less effective as a standalone tool.
   - **Action:** The MOSAIC tool should be used for creative tone and channel mix optimization in high-yield behavioural segments.
6. **Interactions are actionable.**
   - Channel effects vary by amount decile: F2F/Telemarketing outperform Web in higher deciles; gaps compress at lower bands.
   - **Action:** Optimise ROI using interaction-aware targeting rather than single-factor rules.

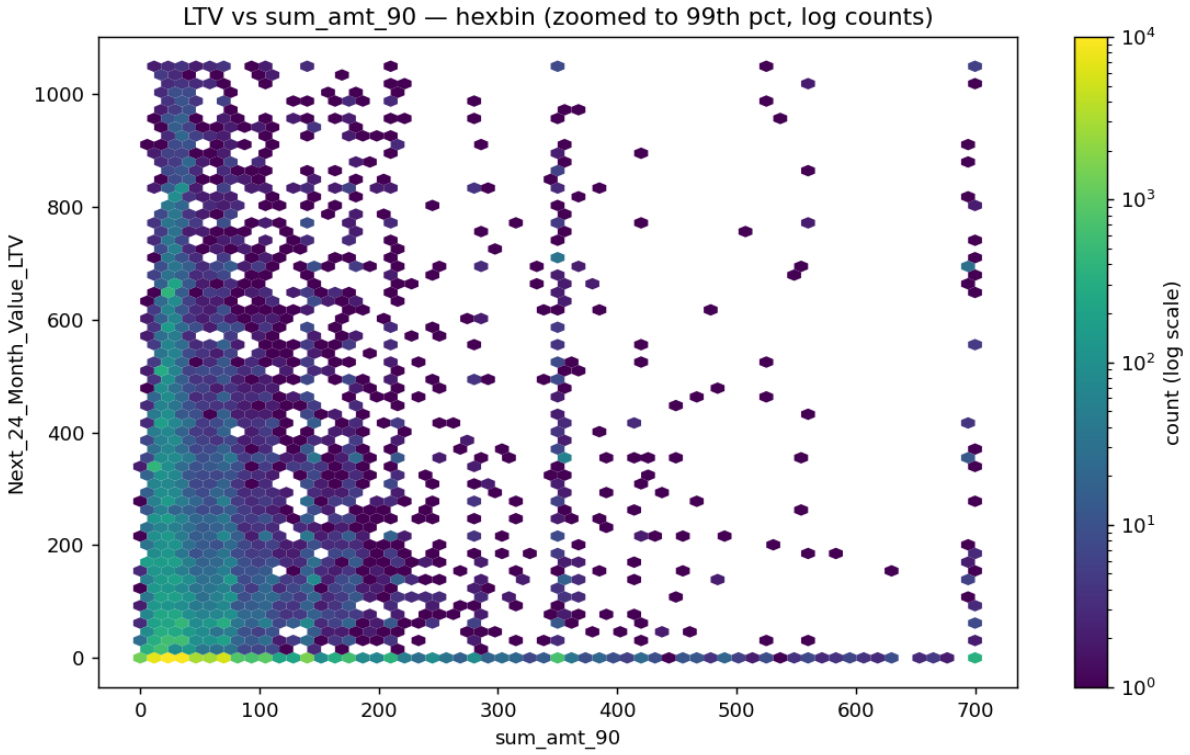
6.4 Prioritised recommendations (testable next steps)

1. **Make "RG within 90 days" the north-star outcome;** Fund *TM/F2F upgrade programs* for (a) any RG signal, (b) ≥2 gifts, or (c) fast second gift. Track LTV lift vs. control.
2. **Shift web journeys from one-off to monthly-first;** Default to monthly on donation forms; keep one-off as secondary. Use post-donation upgrade prompts for cash first-product cohorts.
3. **Implement a 45-day second-gift play.**; Structured touch plan at day 14/28/42 (story + ask). Objective: increase the proportion of donors with a second gift inside the window.
4. **Allocate channels by interaction segments;** High **sum_amt_90** deciles and multi-gift donors → TM/F2F upgrade; low deciles → lower-cost digital nurture first. Validate with split-tests using the interaction pattern.
5. **Data capture and enrichment;** Reduce "Unknown" **Age_Bucket** with light prompts; keep MOSAIC in the feature set for personalisation rather than primary selection. Monitor the postcode-match rate (currently ~93.6%) to maintain enrichment coverage.

Appendix A: four-panel plot for `sum_amt_90` vs LTV



Appendix B: Hexbin Plot (LTV vs sum_amt_90)



Appendix C: pairwise MWU with Holm correction & rank-biserial effect size page 17

Pairwise MWU — Product (top 5)

A	B	n_A	n_B	mean_A	mean_B	diff_mean	U	p_raw	r_rank_biserial	p_holm
Cash - One off	RG - Global Parent	38,262.00	12,962.00	64.55	271.16	-206.60	122,466,073.50	0.00	0.51	0.00
Cash - One off	Community Fundraising	38,262.00	7,985.00	64.55	5.02	59.53	187,126,248.50	0.00	-0.22	0.00
Cash - One off	RG - Paddington Postcard	38,262.00	1,326.00	64.55	230.40	-165.85	6,123,059.00	0.00	0.76	0.00
RG - Global Parent	Community Fundraising	12,962.00	7,985.00	271.16	5.02	266.13	86,482,283.50	0.00	-0.67	0.00
RG - Global Parent	Cash - Inspired Gifts	12,962.00	7,789.00	271.16	37.41	233.74	78,716,384.50	0.00	-0.56	0.00
Community Fundraising	RG - Paddington Postcard	7,985.00	1,326.00	5.02	230.40	-225.38	419,244.50	0.00	0.92	0.00
Cash - Inspired Gifts	RG - Paddington Postcard	7,789.00	1,326.00	37.41	230.40	-192.99	933,081.50	0.00	0.82	0.00
Community Fundraising	Cash - Inspired Gifts	7,985.00	7,789.00	5.02	37.41	-32.39	25,981,554.00	0.00	0.16	0.00
Cash - One off	Cash - Inspired Gifts	38,262.00	7,789.00	64.55	37.41	27.14	158,354,767.00	0.00	-0.06	0.00
RG - Global Parent	RG - Paddington Postcard	12,962.00	1,326.00	271.16	230.40	40.75	7,727,116.50	0.00	0.10	0.00

Pairwise MWU — Channel (top 6)

A	B	n_A	n_B	mean_A	mean_B	diff_mean	U	p_raw	r_rank_biserial	p_holm
Web	F2F	44,013.00	5,545.00	85.41	211.67	-126.26	79,266,104.00	0.00	0.35	0.00
F2F	General	5,545.00	2,638.00	211.67	100.70	110.98	10,156,717.50	0.00	-0.39	0.00
Web	Door to door	44,013.00	2,282.00	85.41	219.58	-134.17	34,827,588.50	0.00	0.31	0.00
Web	Retail	44,013.00	2,157.00	85.41	192.05	-106.64	33,265,163.00	0.00	0.30	0.00
General	Door to door	2,638.00	2,282.00	100.70	219.58	-118.89	1,980,422.00	0.00	0.34	0.00
General	Retail	2,638.00	2,157.00	100.70	192.05	-91.35	1,882,447.50	0.00	0.34	0.00
Web	DM	44,013.00	1,214.00	85.41	113.18	-27.77	20,708,029.50	0.00	0.22	0.00
General	DM	2,638.00	1,214.00	100.70	113.18	-12.49	1,158,801.50	0.00	0.28	0.00
F2F	DM	5,545.00	1,214.00	211.67	113.18	98.49	3,998,799.50	0.00	-0.19	0.00
Door to door	DM	2,282.00	1,214.00	219.58	113.18	106.40	1,589,173.50	0.00	-0.15	0.00

Pairwise MWU — Age (top 6)

A	B	n_A	n_B	mean_A	mean_B	diff_mean	U	p_raw	r_rank_biserial	p_holm
Unknown	31-40	143,316.00	11,623.00	59.37	204.37	-144.99	473,730,309.00	0.00	0.43	0.00
Unknown	41-50	143,316.00	5,591.00	59.37	292.49	-233.11	178,270,967.00	0.00	0.56	0.00
Unknown	19 - 30	143,316.00	4,358.00	59.37	137.07	-77.70	217,662,063.50	0.00	0.30	0.00
Unknown	51-60	143,316.00	3,580.00	59.37	340.87	-281.49	97,904,916.50	0.00	0.62	0.00
Unknown	71+	143,316.00	2,995.00	59.37	338.53	-279.16	83,835,003.50	0.00	0.61	0.00
19 - 30	51-60	4,358.00	3,580.00	137.07	340.87	-203.79	4,674,469.00	0.00	0.40	0.00
19 - 30	71+	4,358.00	2,995.00	137.07	338.53	-201.46	3,987,728.50	0.00	0.39	0.00
41-50	19 - 30	5,591.00	4,358.00	292.49	137.07	155.41	16,056,855.00	0.00	-0.32	0.00
31-40	51-60	11,623.00	3,580.00	204.37	340.87	-136.50	15,395,075.00	0.00	0.26	0.00
31-40	71+	11,623.00	2,995.00	204.37	338.53	-134.17	13,102,096.00	0.00	0.25	0.00

Appendix D: chi-square + Cramér's V

first_product_group_90 × first_channel_90: χ^2=43511.4, dof=20, p=0, Cramér's V=0.434

first_channel_90	DM	Door to door	F2F	General	Retail	Web
first_product_group_90						
Cash - Inspired Gifts	139	0	66	17	0	7467
Cash - One off	1075	301	61	2355	150	32251
Community Fundraising	0	0	0	266	0	70
RG - Global Parent	0	1981	5418	0	2007	2914
RG - Paddington Postcard	0	0	0	0	0	1311

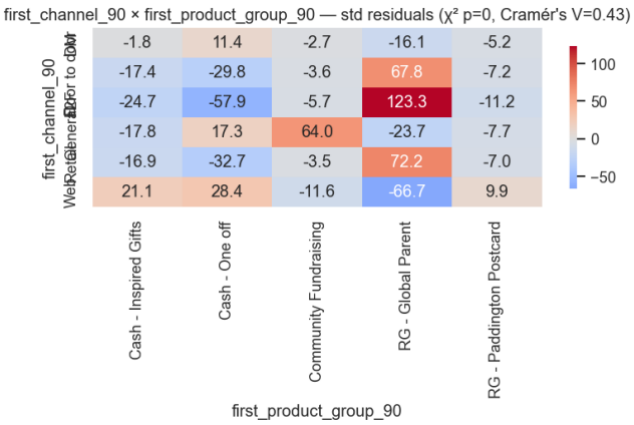
Age_Bucket × first_channel_90: χ^2=37400.1, dof=25, p=0, Cramér's V=0.362

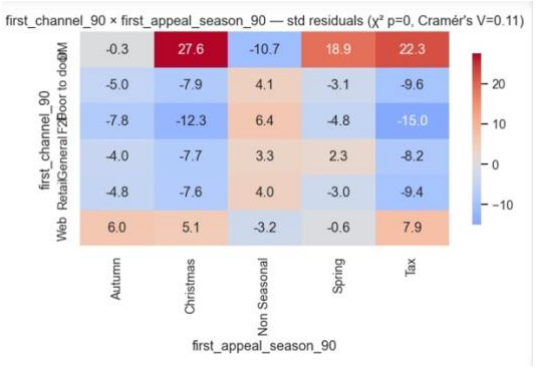
first_channel_90	DM	Door to door	F2F	General	Retail	Web
Age_Bucket						
19 - 30	0	41	867	9	202	123
31-40	2	357	1763	8	680	281
41-50	2	335	746	13	383	250
51-60	4	265	416	6	248	151
71+	26	162	432	17	187	220
Unknown	1175	941	986	2576	319	42848

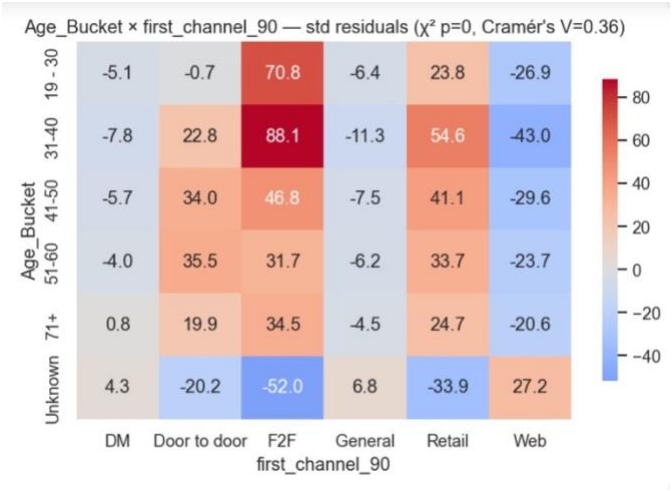
first_appeal_season_90 × first_channel_90: χ^2=2926.9, dof=20, p=0, Cramér's V=0.112

first_channel_90	DM	Door to door	F2F	General	Retail	Web
first_appeal_season_90						
Autumn	12	0	0	7	0	609
Christmas	191	0	0	6	0	1370
Non Seasonal	757	2282	5545	2584	2157	39737
Spring	48	0	0	19	0	178
Tax	206	0	0	22	0	2119

Appendix E:







8.0 References

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**