MIS 779

Decision Analytics in Practice

Final Report

Data-Driven Strategies for Enhancing Child Sponsorship Retention

Team 10

Table of Contents:

Executive Summary:

World Vision Australia, a leading humanitarian organization since 1966, plays a crucial role in child well-being and community development across over 100 countries. Despite its extensive reach and impact, the organization is currently facing significant challenges related to donor retention and segmentation, which threaten the sustainability of its child sponsorship programs.

We conducted a comprehensive analysis, leveraging the power of both descriptive and predictive modelling techniques. This analysis revealed a clear picture of the challenges and, more importantly, opportunities for improvement.

Business Problems:

1. **Declining Child Sponsorship Retention:** Retention rates for child sponsorships have declined by 48% over the past three years, posing a serious risk to vital child development programs.
2. **Ineffective Donor Segmentation:** The existing segmentation model fails to meet donor needs. High churn persists across segments, and our most generous donors in the "Null" segment remain unsegmented due to significant data gaps.

Key Insights:

1. **Churn Trends:** Over the past three years, 48% of child sponsorships ended, with the highest churn occurring in 2023. This trend necessitates immediate action to stabilize sponsorship levels.
2. **Communication Impact:** There is a clear correlation between communication volume and retention. Periods of increased communication corresponded with lower churn rates, highlighting the need for consistent donor engagement.
3. **Segment-Specific Issues:** High churn rates in Segments 3 and 5, despite significant communication efforts, indicate that the current segmentation model is not addressing donor needs effectively. Additionally, the "Null" segment, which includes the highest-paying donors, is not segmented, leading to missed engagement opportunities.

Modelling Insights:

1. **Churn Prediction Accuracy:** Our logistic regression model accurately identified at-risk donors with an accuracy of 89.31%, enabling proactive retention efforts.
2. **New Clustering Model:** We developed a new segmentation model based on donor behaviour, identifying five distinct clusters. This model allows for more personalized and effective communication strategies.

Recommendations:

1. **Enhanced Segmentation:** Implement the new clustering model, that leverages donor communication and donation patterns, eliminating the need for demographic information, to ensure personalized communication strategies for all donor segments, particularly high-value donors.
2. **Consistent Communication:** Maintain regular and meaningful interactions with donors, especially during critical periods, to prevent churn. Tailor communication content and delivery methods based on donor preferences and behaviours.
3. **Proactive Retention Efforts:** Use the predictive model to identify at-risk donors and implement targeted retention strategies.

The Path Forward: A Brighter Future for Child Sponsorship

By implementing these data-driven recommendations, World Vision Australia can significantly improve child sponsorship retention rates, enhance the effectiveness of the segmentation model, and build stronger, more sustainable relationships with their donors. This will empower us to achieve our strategic goals and continue supporting vital child development programs that transform lives around the world.

Introduction & Approach:

Business Context and Strategy:

World Vision Australia, a prominent figure in global humanitarian efforts since 1966, empowers child well-being and community development across over 100 countries [World Vision Australia website, About Us]. Their mission has inspired positive change, impacting nearly 229,000 children [World Vision Australia website, About Us]. Cultivating strong donor relationships is a cornerstone of their strategy. Initiatives fostering deeper connections and engagement align with their objective of reducing donor churn and bolstering retention rates through impactful communication [World Vision Australia website, Our Work].

Business Problems:

Challenge 1: Declining Child Sponsorship Retention: Retention rates for child sponsorships have seen a concerning 48% decline over the past three years, jeopardizing vital child development programs.

Challenge 2: Ineffective Donor Segmentation: The current segmentation model's effectiveness in fostering engagement and retention is unclear. Significant data gaps hinder tailoring communication strategies, even existing segments exhibit high churn rates. Notably, the 'Null' segment, despite being the highest-paying donors, is not segmented at all, highlighting a critical oversight in engaging these valuable contributors.

Approach:

We propose a comprehensive data analysis framework utilizing descriptive and predictive components. Our primary focus is understanding sponsor behavior patterns, identifying key factors contributing to child sponsorship churn, and developing actionable recommendations to improve retention rates and segmentation accuracy.

The analysis delves into two key areas:

Communication & Donation Patterns: We examine the frequency and channels used for communication with donors, and the channels preferred by donors for contributions, particularly regarding child sponsorship programs. Analyzing trends in communication volume and engagement will help identify potential weaknesses and assess their impact on retention.

Donor Segmentation: We assess the effectiveness of the current segmentation model. Are specific segments experiencing higher churn rates? Are communication strategies tailored to different donor needs and preferences? Investigating segmentation effectiveness will determine if improvements are necessary and the most beneficial course of action.

Through this data-driven approach, we aim to gain valuable insights into child sponsorship retention and segmentation effectiveness. These insights will then be used to formulate actionable recommendations for refining communication strategies, enhancing donor engagement, and potentially improving the segmentation model. Ultimately, this will contribute to World Vision Australia's strategic objective of strengthening donor relationships and bolstering child sponsorship retention rates.

Assumptions:

- The donor data for the period 2021-2023 reflects typical trends and behavior patterns. This timeframe captures relevant events that may have influenced donor interactions.
- The definition of "churn status" accurately captures the loss of a child sponsorship based on the fulfillment end date.
- Communication disruptions, such as emergency appeals, directly affect churn rates.
- Donor interaction and donation history data provide sufficient information for predictive modeling and segmentation. While demographic data gaps exist, the focus on behavioral data offers valuable insights.
- Significant external events, like global crises, measurably affect donor engagement and retention.
- Missing values ('NULL') in the "Lewers_Segment" field indicate limited donor demographic information, preventing clear assignment to pre-defined segments by World Vision Australia (WVA).
- The average of missing values ('NULL') in "Lewers_Segment" is higher compared to other segments.
- The "Lewers_Segment" column was excluded from the cluster analysis to create new, independent clusters. This allows for an unbiased analysis tailored to specific criteria and data characteristics, rather than relying on pre-existing segments from WVA.

- Entries with an "Ongoing Financial" commitment type and a fulfillment end date are considered churned donors.
- Fields like eDM1, eDM2, eDM3, eDM4, and eDM5 are categorized as "Email."
- Fields like DM, DM1, and DM2 are categorized as "Letter."
- Communication mediums not explicitly classified as Email, Letter, SMS, or Telephone are categorized as "None."
- Under "Echo_Outbound_Communication_Category," only "Sponsorship and other regular contributions" directly relates to child sponsorship, while other categories do not.

Data Analysis:

Initially, we had three separate datasets: donor data, interaction data, and donation and pledge data, each unlinked. We later integrated these datasets using the unique donor identifier as the primary key. This process allowed us to filter and retain only the linkable entries, enhancing the coherence and utility of our data for analysis.

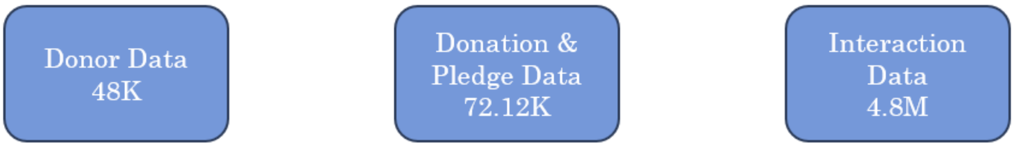


Figure 1 - Initial Datasets

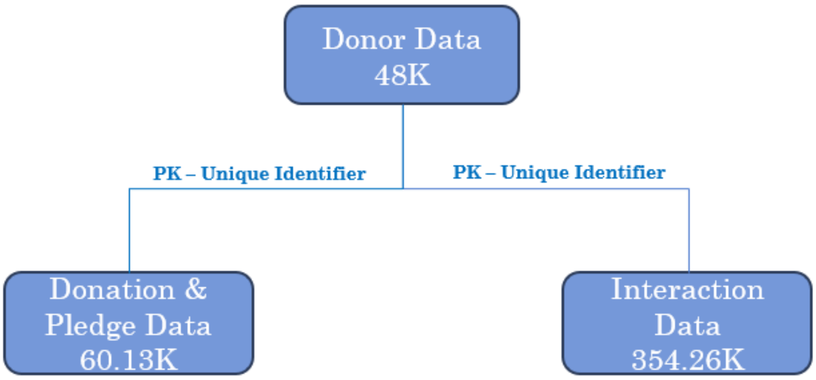


Figure 2 - Datasets after tracing through Primary Key (PK)

Descriptive Analysis:

This analysis delves into key trends within child sponsorship donor data, focusing on two critical challenges: declining retention rates and potentially ineffective donor segmentation. By gaining a deeper understanding of these issues, we can develop strategies to strengthen donor relationships and increase program sustainability.

Data Preparation: Building a Strong Foundation

The initial data preparation focused on understanding overall donor engagement trends. We filtered the interaction and donation data to include only active donors from 2021 to 2023. This selection was based on unique donor identifiers within the dataset. Additionally, a "Churn Status" variable was created to classify donors based on their "Fulfillment End Date." Finally, a "Mode Attribute" was generated to identify the communication channels (e.g., email, phone) used by donors for interaction.

Escalating Churn and Annual Trends

One critical concern is the rise of the churning of child sponsorships. Over the past three years (2021-2023), nearly 48% of sponsorships have ended, translating to a significant loss of over 11,650 sponsorships (See Figure 3). This steady increase, with the highest number of churned sponsors (6,800) occurring in 2023, necessitates immediate action.

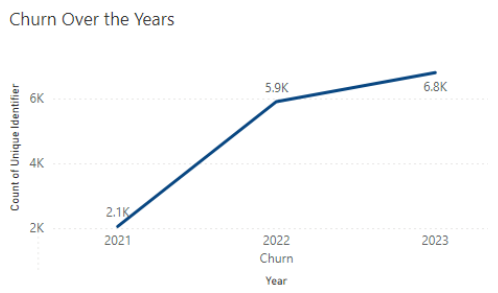


Figure 3 - Overall Churn Rate (2021-2023)

Digging Deeper: Churn by Donor Segment

While all segments exhibit churn, a closer look reveals concerningly high rates in Segments 3 (42%) and 5 (40.88%), despite receiving the most communication (See Figure 4). Conversely, Segment 6 exhibits a lower churn rate (39.54%) with less frequent communication. This suggests the current segmentation model may not be effectively targeting donor needs. A bar chart visualizing churn rates by segment would be insightful.

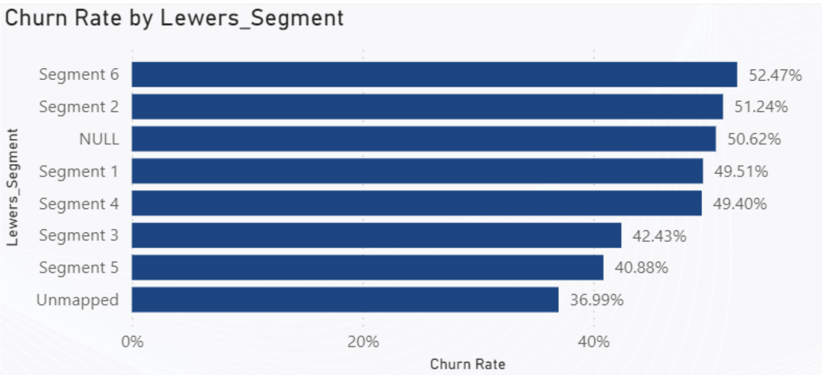


Figure 4 – Churn – Donor Segment wise

Communication Patterns and the Churn Correlation

Communication volume appears to be linked to churn rates. In 2022, when communication was highest (56,452), churn remained relatively low (See Figure 5). However, a significant decrease in communication (35,907 in 2023) coincided with a rise in churn – 11.33% (See Figure 5 & Figure 3).

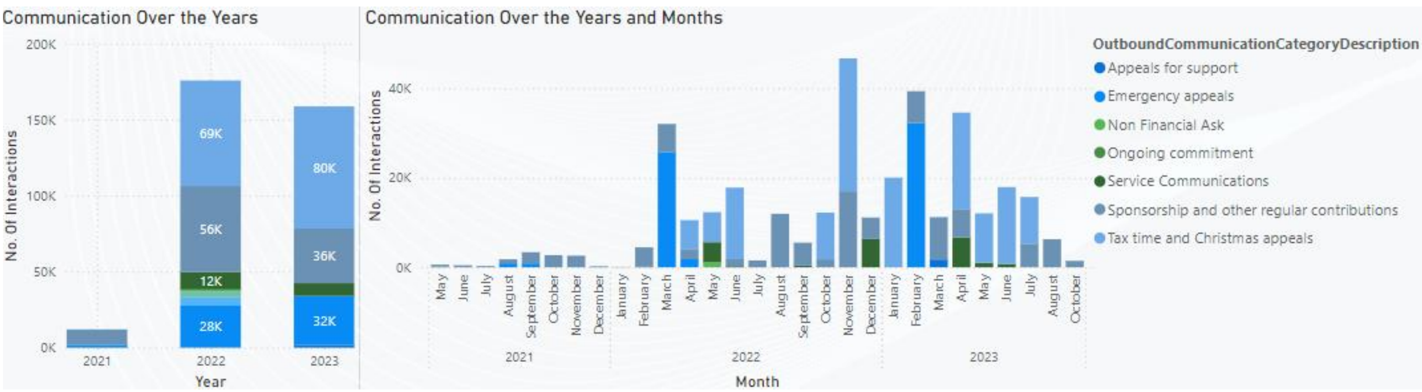


Figure 5 - Communication Over the Period

Impact of Communication Disruptions and Channel Shifts

Communication disruptions also seem to influence churn. A decline in communication from April to June 2022 (Interactions related to Child sponsorship – April = 19.97%, May = 0%, June = 6.73%) correlated with a rise in churn(11.33%). Similarly, during the 2023 emergency appeals, lower sponsorship interactions resulted in higher churn the following month.

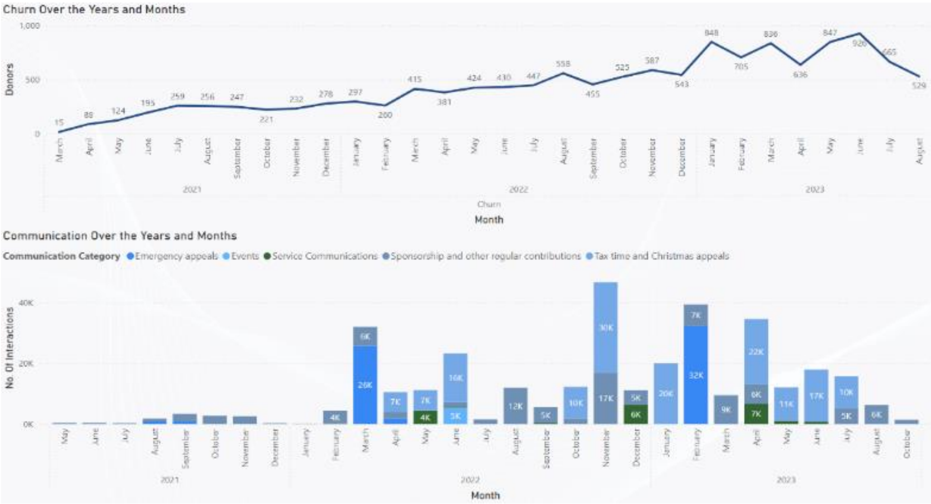


Figure 6 - Communications vs Churn over the years

Another factor to consider is the recent shift to a letter-based communication strategy in 2023. (Interaction via Email in 2022 =92.8%, emails in 2023 = 86.25%)

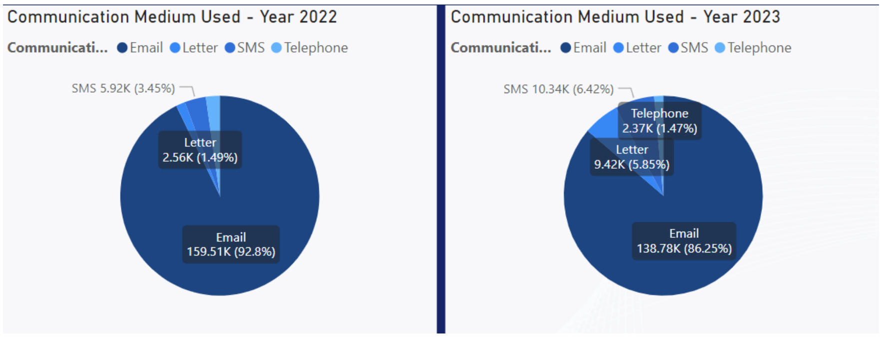


Figure 7 - Communication Medium Used - 2022 vs. 2023

Unearthing Potential: The "Null" Segment

The "Null" segment, with the highest total donations ($9 million), presents an intriguing opportunity for increased engagement. Despite limited demographic data, this segment exhibits a high average donation amount ($823.96) and a significant number of donations (38,130).

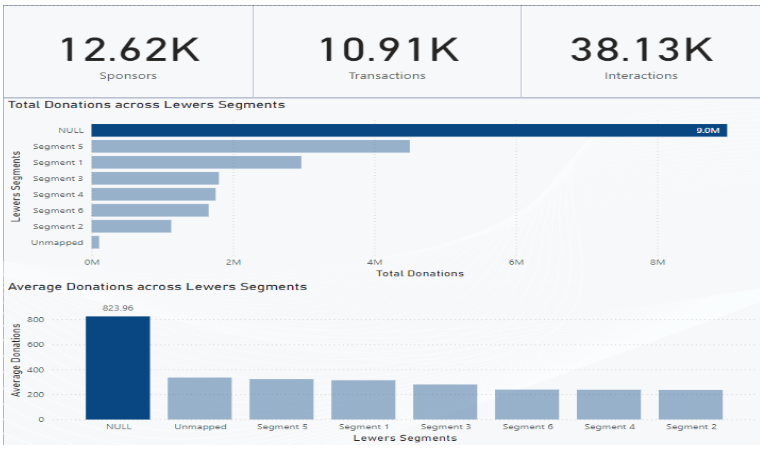


Figure 8 - Donors in Segments

New Clustering Model for More Effective Segmentation:

To address the ineffectiveness of the current segmentation model, a new model was developed based on factors like tenure, churn status, donations, and communication volume. This model identified five distinct donor clusters, each with potentially unique characteristics and communication needs (See Figure 9). Utilizing these clusters for targeted communication strategies can potentially improve engagement and retention.

This new model offers a promising solution to the challenge of ineffective donor segmentation, allowing for more personalized communication strategies. Crucially, this new model relies solely on behavioral and engagement data and does not incorporate any demographic information about the donors.

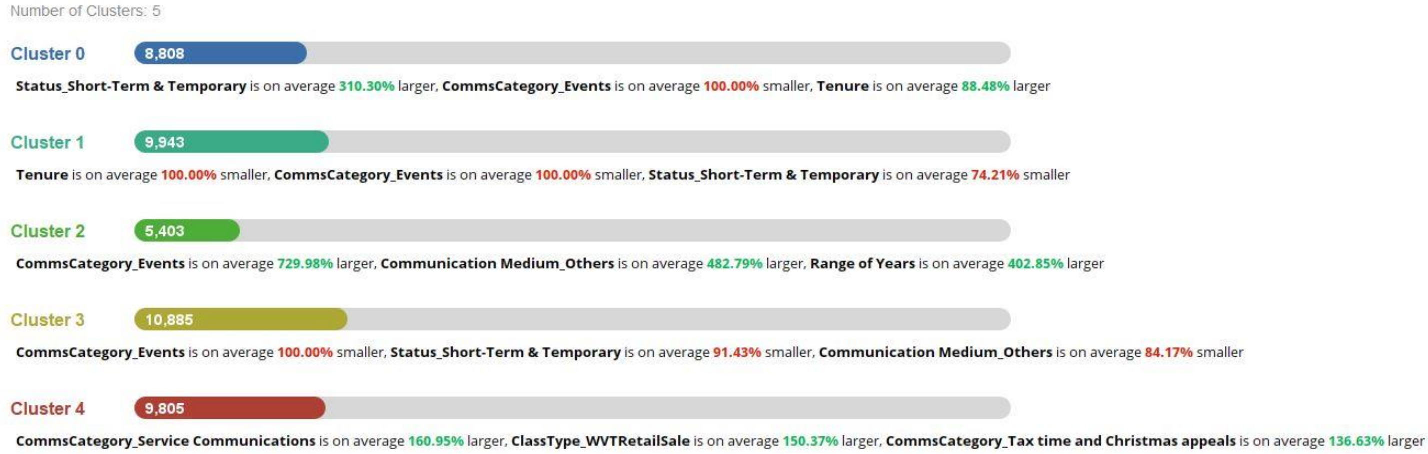
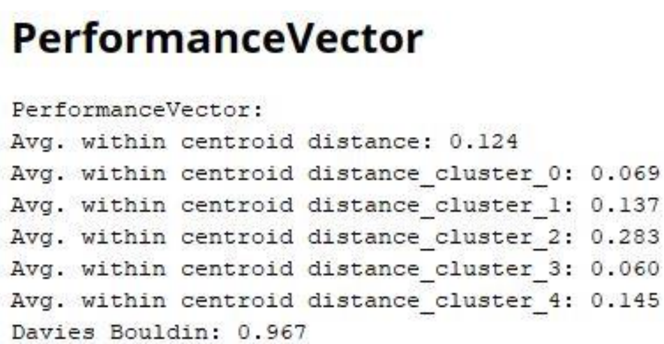


Figure 9 - Donor Clusters

The clusters comprised the following numbers of donors:

Cluster	Number of Donors
Cluster 0	8808 donors
Cluster 1	9943 donors
Cluster 2	5403 donors
Cluster 3	10885 donors
Cluster 4	9805 donors

We employed the Davies-Bouldin (DBI) index to evaluate its performance. In this analysis, the DBI score obtained was 0.967. While a lower score is generally desirable, interpreting DBI scores can be subjective and depends on the specific dataset and clustering algorithm used. The PerformanceVector is obtained as follows:


PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.124
Avg. within centroid distance_cluster_0: 0.069
Avg. within centroid distance_cluster_1: 0.137
Avg. within centroid distance_cluster_2: 0.283
Avg. within centroid distance_cluster_3: 0.060
Avg. within centroid distance_cluster_4: 0.145
Davies Bouldin: 0.967
```

Figure 10 - Results of PerformanceVector in RapidMiner

The centroid chart generated was as follows:

The centroid chart illustrates the distribution of key features across different clusters. The spikes in communication mediums and categories, such as email and emergency appeals, highlight the distinct preferences and behaviors within each cluster. This differentiation aids in tailoring communication strategies to better align with the specific needs and engagement patterns of each donor segment

Further analysis, such as exploring the characteristics of each cluster, is recommended to fully understand the potential benefits of this new segmentation approach. However, the initial results suggest this model holds promise for improving donor engagement and retention through targeted communication strategies.

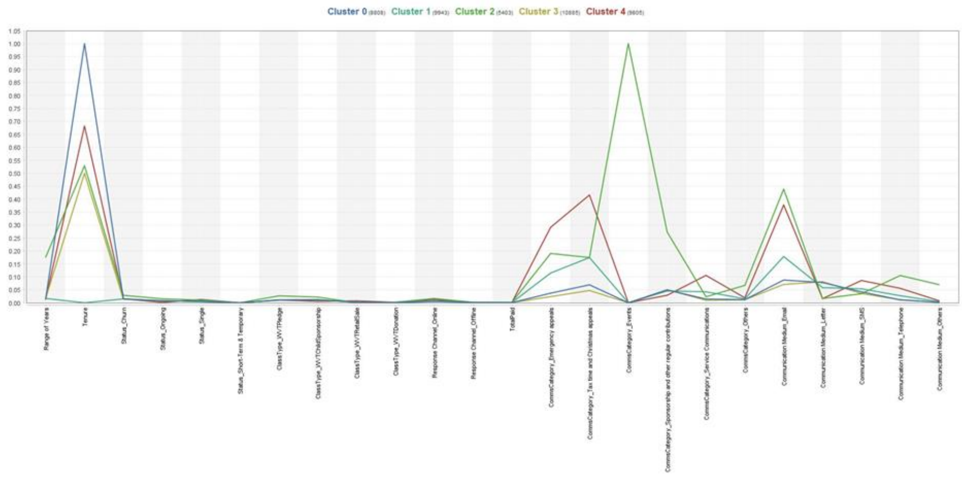


Figure 11 - Centroid Chart though RapidMIner

Predictive Analysis:

To gain a deeper understanding of donor churn within our child sponsorship program, we developed a predictive model. This model analyzes donor data to forecast churn rates and identify individuals at high risk of discontinuing their sponsorship.

Data Preparation

The foundation of the model lies in a meticulously prepared dataset. We constructed a new dataset specifically focused on donors actively participating in child sponsorship programs. This dataset aggregates donor interactions and donation history, providing a comprehensive view of their engagement with our organization. Furthermore, we implemented rigorous filtering and normalization techniques to ensure data integrity and consistency.

Attributes	Range of Years	Tenure	CommsCategory_Child S...	CommsCategory_Other Purposes	Communication Medium_Email	Communication Medium_Others	Response Channel_Online	Response Channel_Offline	TotalPaid	Lewers_Seg...
Range of Years	1	0.512	0.743	0.252	0.608	0.302	0.260	-0.169	0.621	-0.148
Tenure	0.512	1	0.519	0.639	0.743	0.353	0.186	-0.143	0.353	-0.162
CommsCategory_Child Sponsorship	0.743	0.519	1	0.236	0.706	0.566	0.284	-0.194	0.602	-0.211
CommsCategory_Other Purposes	0.252	0.639	0.236	1	0.834	0.333	0.080	-0.063	0.139	-0.172
Communication Medium_Email	0.608	0.743	0.706	0.834	1	0.402	0.205	-0.149	0.431	-0.229
Communication Medium_Others	0.302	0.353	0.566	0.333	0.402	1	0.183	-0.115	0.300	-0.174
Response Channel_Online	0.260	0.186	0.284	0.080	0.205	0.183	1	-0.525	0.357	-0.056
Response Channel_Offline	-0.169	-0.143	-0.194	-0.063	-0.149	-0.115	-0.525	1	0.184	0.059
TotalPaid	0.621	0.353	0.602	0.139	0.431	0.300	0.357	0.184	1	-0.154
Lewers_Segment	-0.148	-0.162	-0.211	-0.172	-0.229	-0.174	-0.056	0.059	-0.154	1

Figure 12 - Correlation Matrix

The correlation matrix clearly depicts how the predictors selected affect the churn of child sponsorships.

Model Selection and Training

Considering the binary nature of churn prediction (churn vs. retain), we opted for a logistic regression model. This approach is well-suited for this task due to its effectiveness in binary classification. The model was trained (80% split) and tested (20% split) on a carefully partitioned dataset, ensuring a comprehensive evaluation of its predictive capabilities.

Model Performance

The model achieved a remarkable overall accuracy of **89.31%**, indicating its ability to accurately distinguish between donors likely to churn and those who will remain engaged. This is a strong result, demonstrating the model's potential to effectively predict churn behavior. The confusion matrix obtained is as follows:

	true Ongoing	true Churn	class precision
pred. Ongoing	2295	342	87.03%
pred. Churn	155	1857	92.30%
class recall	93.67%	84.45%	

Figure 13 - Confusion matrix

True positives (TP): 2295 - These represent donors the model correctly identified as at risk of churn, who ultimately did churn.

True negatives (TN): 1857 - These represent donors the model correctly predicted would not churn, and they remained active sponsors.

False positives (FP): 155 - These represent donors the model incorrectly flagged as at risk of churn, but they did not churn.

False negatives (FN): 342 - These represent donors the model incorrectly predicted would remain active, but they ultimately churned.

The model's performance is further characterized by the following metrics:

Measures Evaluated with	Value
Precision	92.30%
Recall	84.45%
F-measure	88.20%
Sensitivity (True Positive Rate)	84.45%
Specificity (True Negative Rate)	93.67%
Kappa Statistic	0.785

The AUC Curve obtained is as follows:

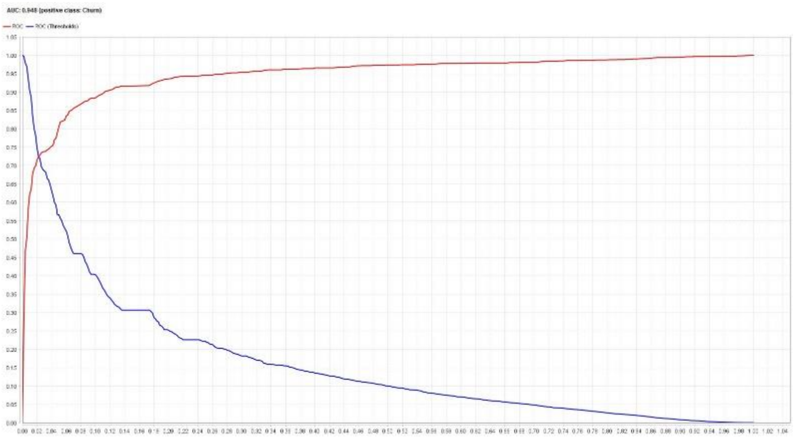


Figure 14 - AUC Curve for Churn Analysis

The ROC curve below illustrates the performance of our logistic regression model in predicting donor churn, with an AUC of 0.948. This high AUC indicates excellent model accuracy, enabling effective identification of high-risk donors.

Interpretations:

Our descriptive and predictive analyses provide critical insights directly addressing the business problems identified:

Challenge 1: Declining Child Sponsorship Retention

Descriptive Insights:

1. **Churn Rates**: The descriptive analysis highlighted a concerning 48% decline in retention rates over the past three years. This trend jeopardises vital child development programs, underscoring the need for immediate action to stabilise sponsorship levels.

2. **Communication Patterns**: The correlation between communication volume and churn rates suggests that more consistent and targeted communication can help reduce churn. Periods of high communication saw lower churn rates, indicating the importance of maintaining regular donor engagement.

Predictive Insights:

1. **Churn Prediction**: The predictive model's high accuracy in identifying at-risk donors (89.31%) allows us to proactively address the declining retention rates. By focusing retention efforts on these high-risk donors, we can mitigate the impact of churn on child sponsorship programs.

2. **Targeted Strategies**: The model's ability to predict churn enables us to tailor our communication and engagement strategies more effectively, ensuring that donors receive the support and information they need to stay engaged.

Challenge 2: Ineffective Donor Segmentation

Descriptive Insights:

1. **High Churn in Certain Segments**: The high churn rates in Segments 3 and 5, despite significant communication efforts, indicate that the current segmentation model is not effectively addressing donor needs. This points to a need for a more refined segmentation approach.

2. **The "Null" Segment**: The analysis revealed that the highest-paying donors, grouped in the "Null" segment, are not segmented at all. This critical oversight means that these valuable contributors are not receiving targeted engagement, potentially leading to higher churn rates among our most significant donors.

Modelling Insights:

1. **New Clustering Model**: The new segmentation model, based on behavioural and engagement data, identifies five distinct donor clusters. This model allows for more personalised communication strategies that can better align with the specific needs and preferences of each donor segment, including the high-value "Null" segment.

2. **Effective Targeting**: By utilising the new clusters for targeted communication, we can enhance donor engagement and retention, addressing the ineffectiveness of the current segmentation model. This approach ensures that even high-paying donors receive the tailored communication necessary to keep them engaged and reduce churn.

Strategic Implications

To effectively tackle the identified business challenges:

1. **Enhanced Segmentation**: Implement the new clustering model to ensure that all donor segments, including high-value donors, receive personalised communication strategies. This targeted approach is crucial for improving engagement and retention rates.

2. **Consistent Communication**: Maintain regular and meaningful interactions with donors, particularly during key periods, to prevent churn. The positive correlation between communication volume and retention highlights the importance of this strategy.

3. **Proactive Retention Efforts**: Utilise the predictive model to identify at-risk donors and implement targeted retention strategies. By focusing on these high-risk individuals, we can significantly mitigate the decline in sponsorship retention.

In conclusion, the insights from both the descriptive and predictive analyses provide actionable strategies to address our business problems. By refining our segmentation model and enhancing our communication efforts, we can improve donor retention and foster stronger, more sustainable relationships with our contributors.

Recommendations:

This section outlines a series of data-driven recommendations to address World Vision Australia's child sponsorship retention challenges, informed by the insights gleaned from our descriptive and predictive analyses. These recommendations prioritize actionable strategies aligned with a business analytics perspective.

1. Communication Optimization:

Frequency and Channel Reassessment:

Comprehensive Analysis: Perform a detailed analysis to determine the optimal frequency and channels of communication that correlate with lower churn rates. Adjust communication strategies to increase engagement through preferred channels and frequencies.

Personalization: Utilize the predictive model to tailor communication content and delivery channels for different donor groups, focusing on their specific needs and preferences.

Content Personalization:

Segmentation-Based Personalization: Leverage donor segmentation data to customize communication messages for different segments, ensuring that content is relevant and engaging for each group.

Performance Monitoring: Regularly monitor key performance metrics such as open rates and click-through rates to optimize communication strategies continually.

2. Segmentation Enhancement:

Data Enrichment:

Targeted Data Collection: Implement strategies to gather missing demographic data, which will help enrich donor profiles and improve segmentation accuracy. Consider using incentives or simplified data collection methods to encourage donors to provide this information.

Model Refinement:

Segmentation Model Review: Assess and refine the current segmentation model by incorporating interaction styles, behavior history, and churn rates. Address the high percentage of donors in the "null" category by integrating more data points such as communication preferences and donation history.

Machine Learning Integration: Explore advanced analytics and machine learning algorithms to identify new patterns within the donor data, leading to more granular and effective segmentation.

3. Predictive Modelling for Churn Prevention:

Proactive Identification and Mitigation:

High-Risk Donor Identification: Use the logistic regression model to proactively identify donors at high risk of churn. Develop targeted communication and retention strategies for these donors to address their specific concerns and reduce churn.

Continuous Improvement: Regularly update and refine the predictive model to improve its accuracy and effectiveness in identifying at-risk donors.

4. Strategic Engagement Initiatives:

Consistent and Targeted Communication:

Regular Engagement: Ensure that communication with donors is consistent and meaningful, particularly during key periods identified in the analysis. Maintain high engagement levels to prevent donor churn.

Tailored Strategies: Implement targeted communication strategies based on the new clustering model, ensuring that each donor segment receives personalized and relevant content.

Feedback and Iteration:

Donor Feedback: Conduct regular surveys and track donor engagement to gather feedback on their preferences and experiences. Use this feedback to identify areas for improvement and optimize future engagement efforts.

Data-Driven Iteration: Establish a cycle of continuous refinement by utilizing ongoing data analysis, performance evaluation, and donor feedback to refine communication strategies and segmentation models.

5. Leveraging the "Null" Segment:

Engagement Focus: Develop targeted engagement strategies for the "null" segment, which includes the highest-paying donors. Use personalized communication to engage these valuable contributors effectively and reduce their churn rates.

6. Implementing Advanced Analytics:

Behavioral Insights: Use the new clustering model to gain deeper insights into donor behaviors and preferences. Tailor communication and engagement strategies to align with these insights, ensuring higher retention rates.

Predictive Analytics: Continuously improve predictive models to better identify at-risk donors and develop proactive strategies to retain them.

By implementing these recommendations, World Vision Australia can enhance donor retention, improve segmentation accuracy, and build stronger, more sustainable relationships with their donors. This data-driven approach will enable the organization to achieve its strategic objectives and support child development programs more effectively.

Appendices:

The appendix has relevant screenshots for the analysis.

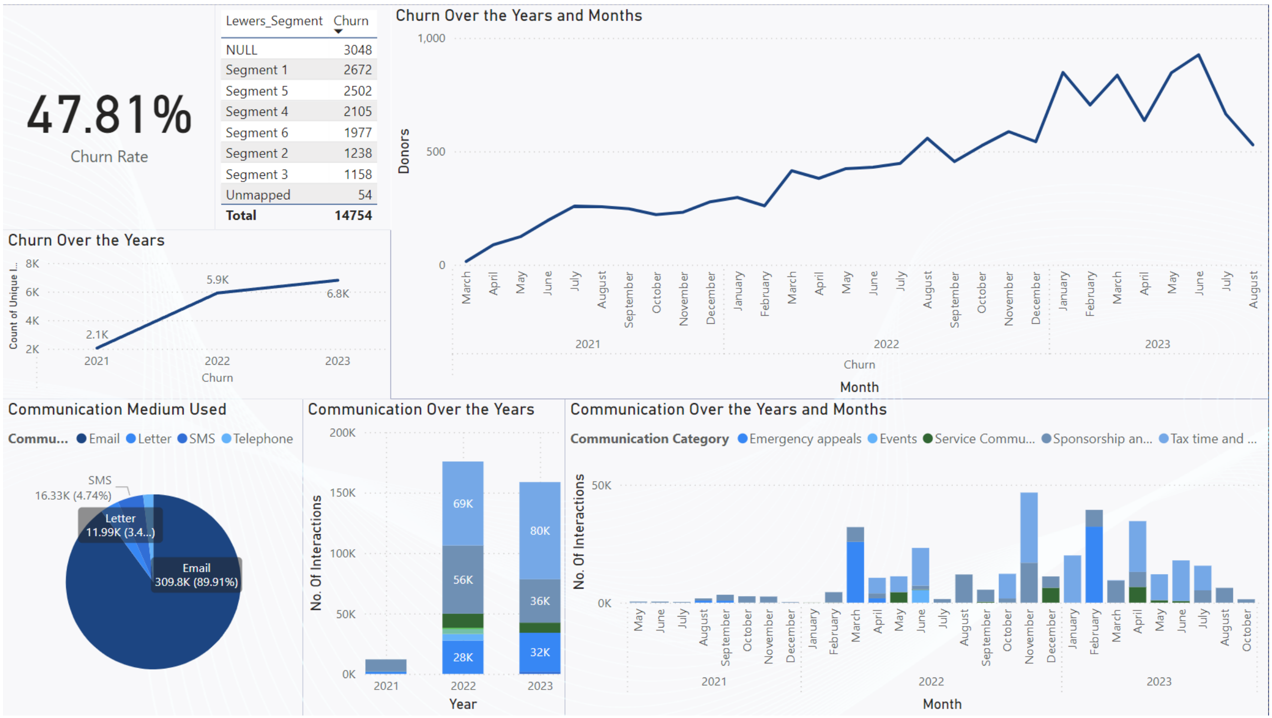


Figure 12 - Dashboard 1: Churn Analysis

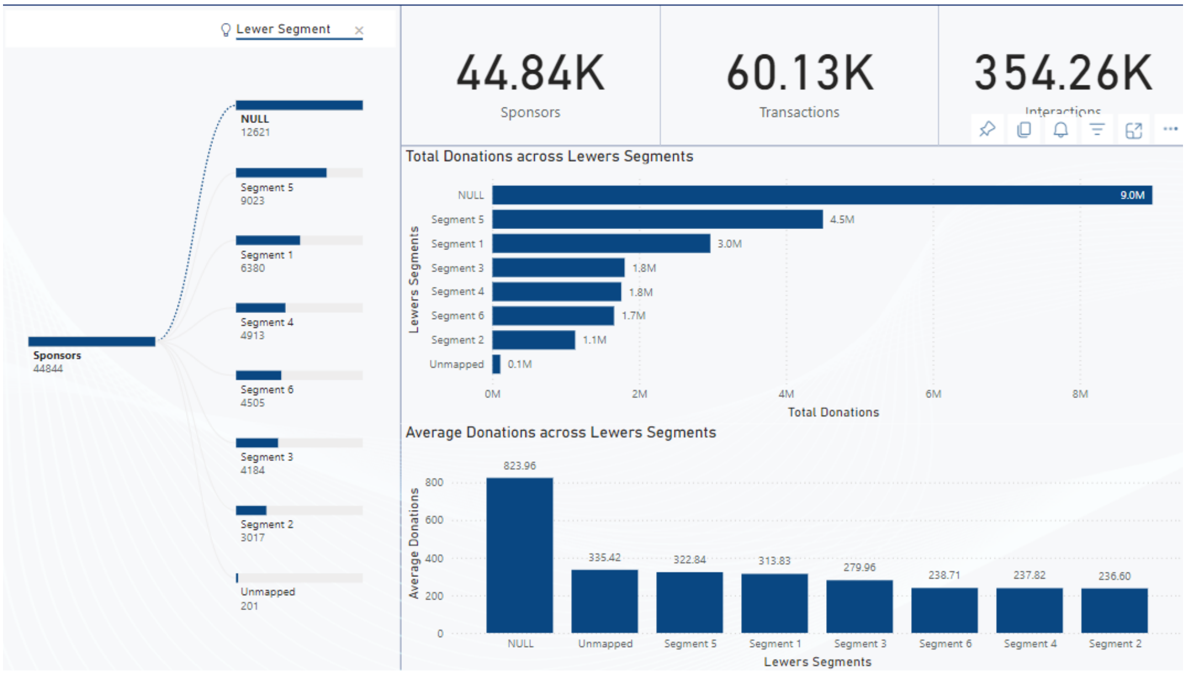


Figure 13 - Dashboard 2: Clustering

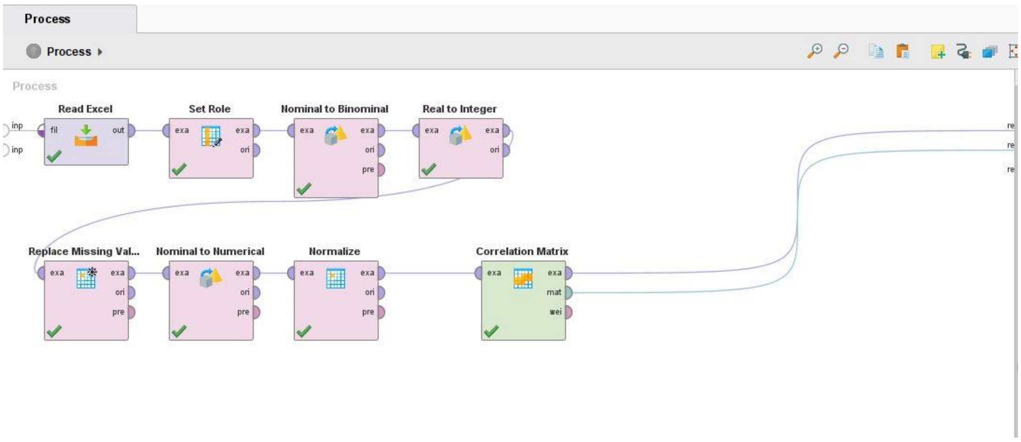


Figure 14 - RapidMiner process for Correlation Matrix

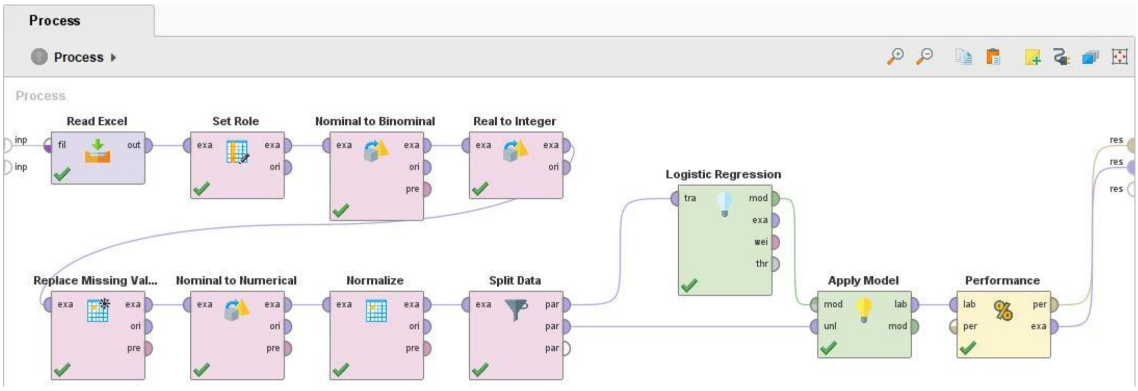


Figure 15 - RapidMiner Process for Churn Analysis

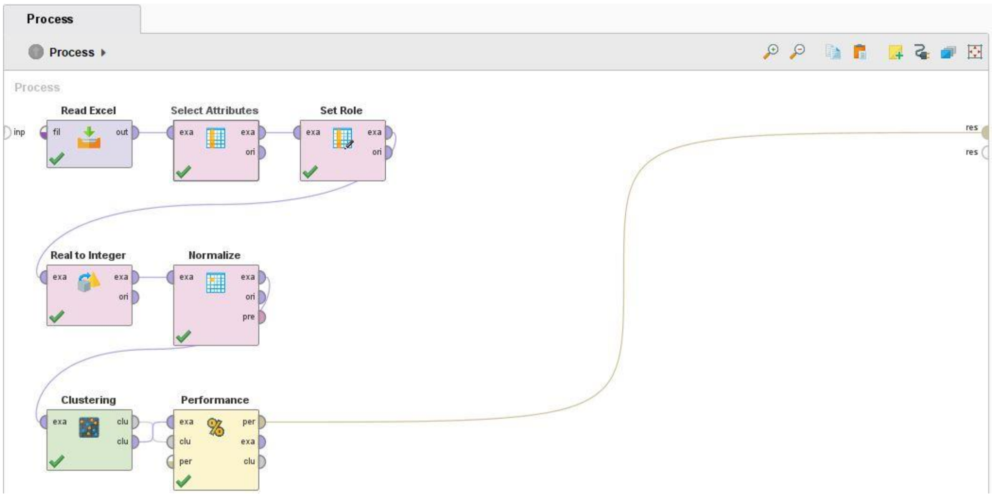


Figure 1615 - RapidMiner Process for Clustering

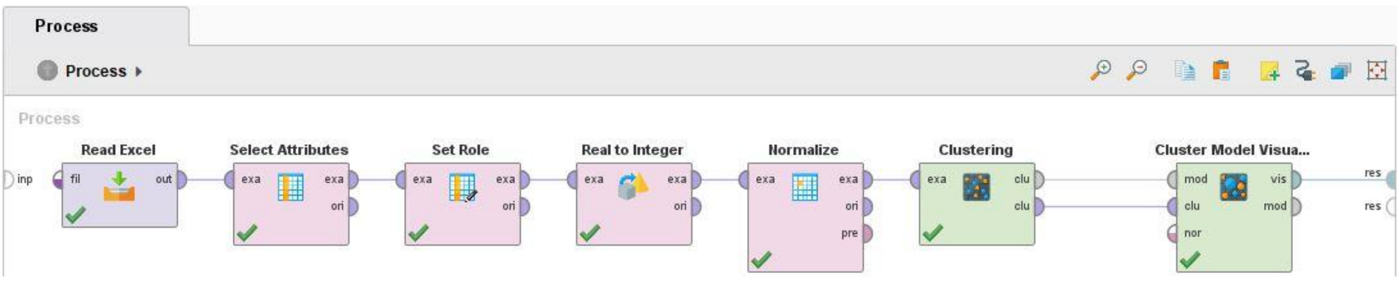


Figure 16 - RapidMiner Process for Visusaling Clusters