



# Introduction

---

Non-Profit Organizations (NPOs) play a vital role in social, educational, and economic development, but some, especially in underdeveloped countries like Haiti, rely heavily on gut instinct and traditional methods to make decisions, which leads to poor donor retention, ineffective communication, weak donor interaction, and unsuccessful campaigns. Since NPOs rely heavily on fundraising, understanding donor behavior is a must, as one of the major challenges NPOs face is maintaining donor engagement and loyalty over time.

The primary objective of this project is to showcase the importance of using data to make informed decisions in NPOs. For that purpose, the project aims to analyze donor behavior and categorize donors for a random and fictitious organization. The insights gained throughout this analysis will foster the use of data and machine learning in decision-making, which can therefore help NPOs design more effective fundraising and communication strategies, ultimately enhancing their ability to sustain and grow their initiatives.

To achieve this, we use the [Donor Data](#) from Kaggle, which comprises six tables detailing donors, donations, campaigns, project results, and engagement activities. These data provide a comprehensive view of donor interactions and contributions, forming the basis for our analysis.

## Methodology

---

### Exploratory data Analysis

- data preparation
- business understanding
- data understanding

### Major questions

These questions deepen our grasp of the business and data context, steering the analysis toward the NPOs' core concerns and enabling a thorough exploration of donor dynamics.

This part is crucial, as it serves as a compass that guides our analysis and helps address stakeholders' concerns or uncover relevant information.

### Modeling/ Model Evaluation

### Recommendations

## Contact information

the contact information of the two data scientist who work on this project is provided.

NB: The scientist colaborated equally on this project.

- name
- email
- phone number
- LinkedIn

## Exploratory data Analysis

---

### Data preparation

There are 7 tables in the dataset; one of them is a duplicate, so we combined the 6 unique tables. Some columns are reformatted .Some are dropped when not necessary for the analysis and the NULL and duplicate values were removed.

### Business understanding

In today's fast-evolving fundraising landscape, nonprofit organisations face growing pressure to engage supporters meaningfully, demonstrate results, and build trust through transparency. Traditional, instinct-driven approaches no longer meet modern donor expectations

Adopting a data-informed approach is not just a technical upgrade—it's a cultural shift that aligns mission, message, and measurement. With the right mindset and existing tools, even small organisations can use data to boost engagement and impact.

Delivering measurable change increasingly depends on how well nonprofits engage and retain supporters. As donor expectations rise and attention spans shrink, sustaining long-term relationships becomes a major challenge.[McKinsey Global Institute](#) says that data-driven organisations are 23 times more likely to acquire customers, 6 times as likely to retain customers and 19 times as likely to be profitable, showing the importance of including data in decision making.

Our objectives focus on helping NPOs reduce donor churn, strengthen their relationships with donors, and improve alignment between fundraising and impact, with data driven analysis. We aim to help them engage donors and inspire new strategies for attraction and retention.

### Data Understanding

A synthetic dataset was created to reflect the operational characteristics of a nonprofit organisation, focused on youth and community welfare. It includes relational tables for donors, donations, campaigns, impact outcomes, and engagement activities. These tables are structured to support a complete engagement analysis from start to finish.

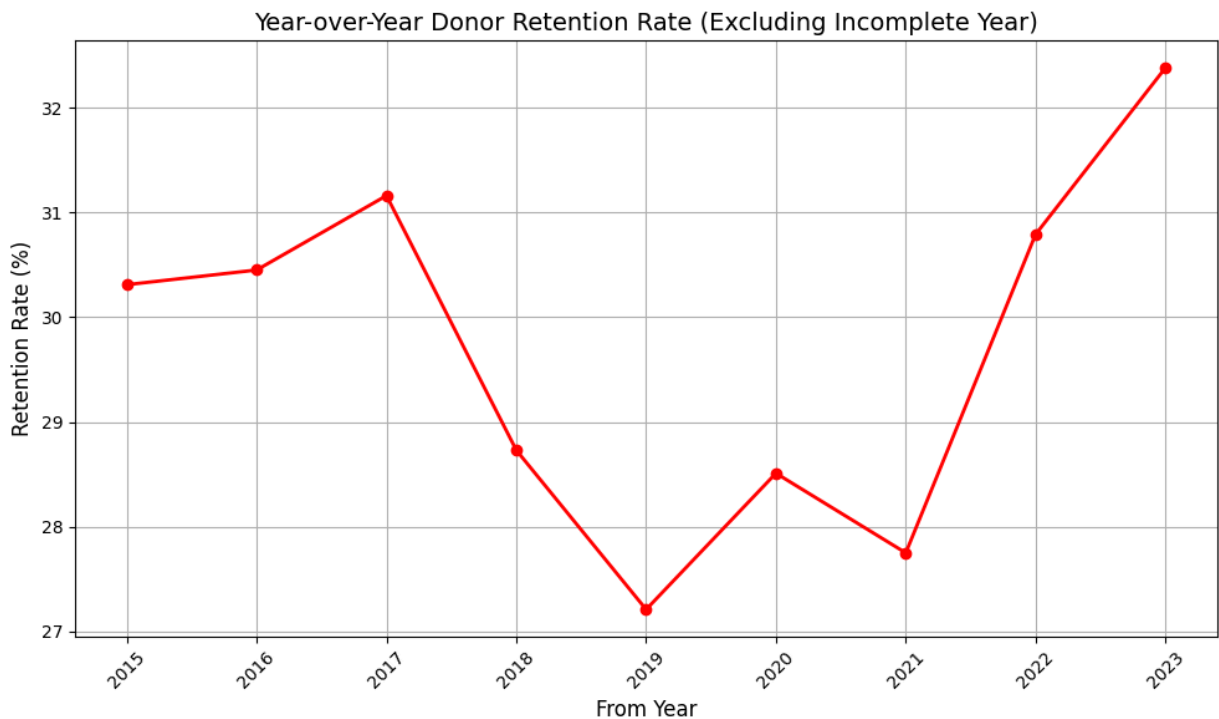
Although the data is fictitious, it reflects realistic patterns based on [sector](#) / [research](#). These patterns include donor attrition, variations in donation behavior, and inconsistent campaign performance (Virtuous, 2024). The goal is to encourage the use of data in nonprofit environments and to provide the NPO community with a clear example of what a data-informed approach can accomplish. This simulation allows for open analysis while maintaining privacy and ethical standards.

The structure of the dataset makes it possible to demonstrate how nonprofit engagement can be improved through analytics without relying on sensitive real-world data.

## Major questions

this section is a follow up of the two previous sections. it builds our business and data understanding and allow us to leverage the data to get meaningful and pertinent information. We will also tell the purpose of each Newfound information, and by the same occasion foster a data informed approached in NPOs.

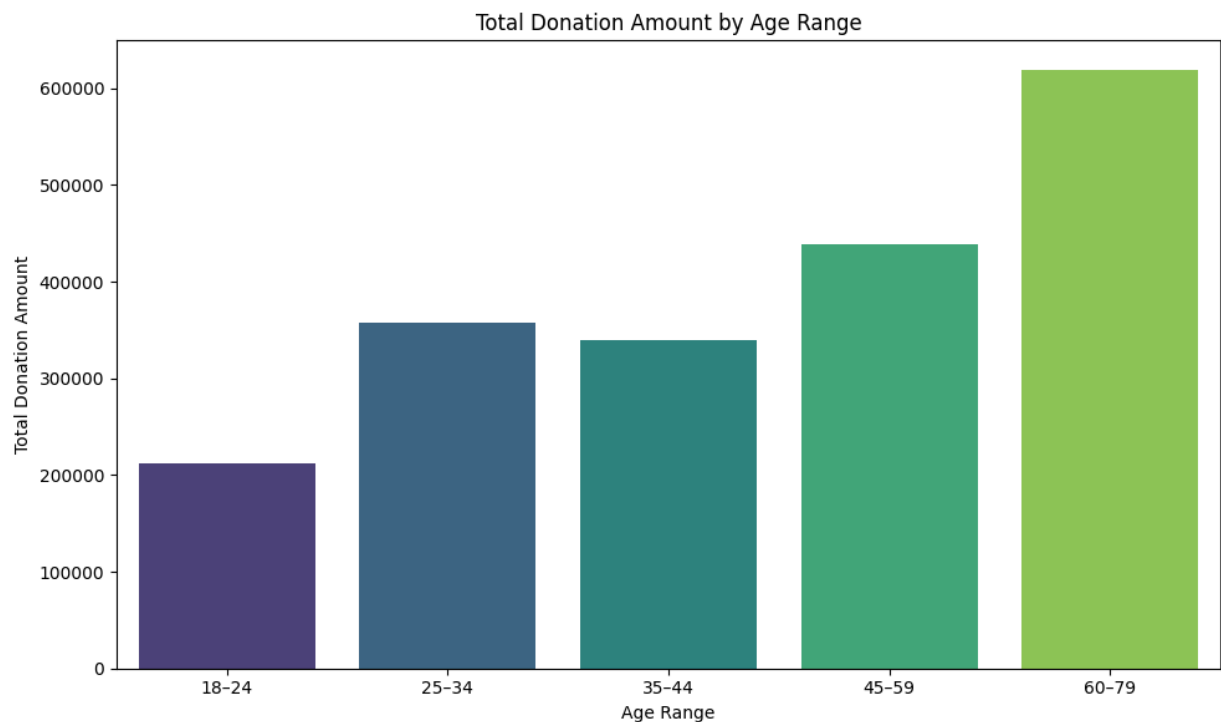
### How has donor retention evolved over time ?



Comment: This line graph illustrates the percentage of donors retained from one year to the next over a ten-year period. While retention fluctuates around 30%, occasional dips highlight the importance of targeted re-engagement strategies and sustained communication with donor segments.

Any NPO will want this information to position themselves regarding donor retention. By knowing how they retain donors, they can determine whether or not to foster interaction with donors, build more campaigns or any other inspired-data decision that could improve donor retention.

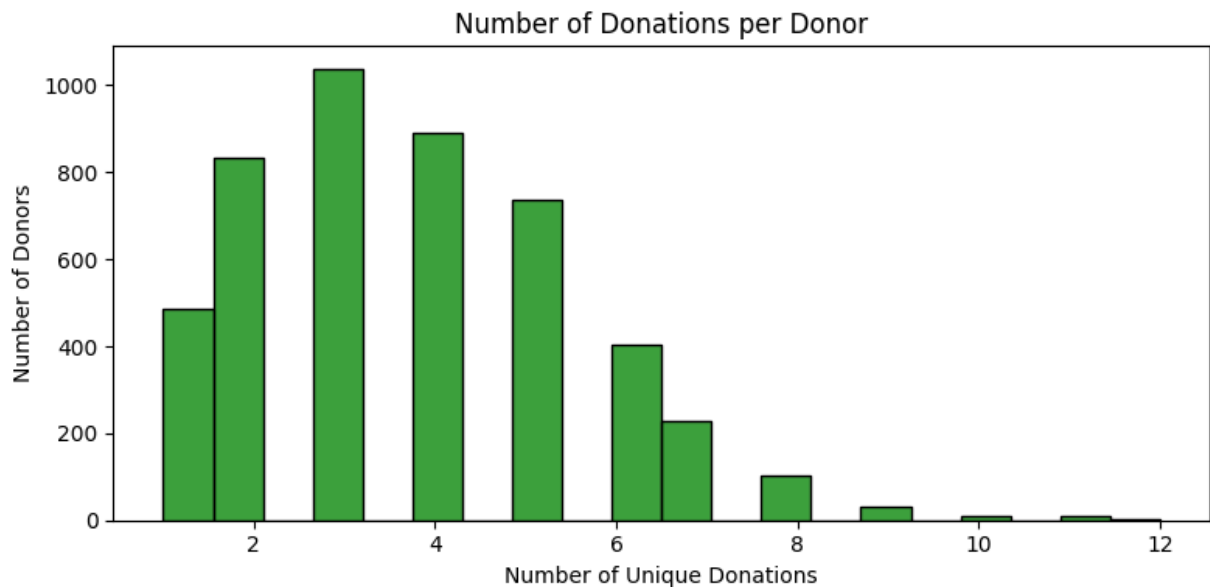
### What is the donation frequency per age range ?



Comment: Donors who are older are likely to donate more than young individuals in the range of 25 to 34 years old. The results match [real world data](#).

Any NPO would want this information to know which age range to target the most and which one to sensitize to donate more, whether it's through campaign messages or any other means.

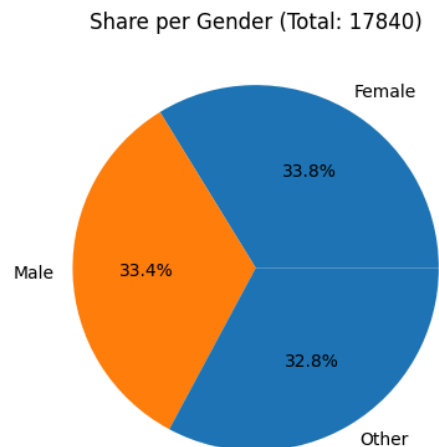
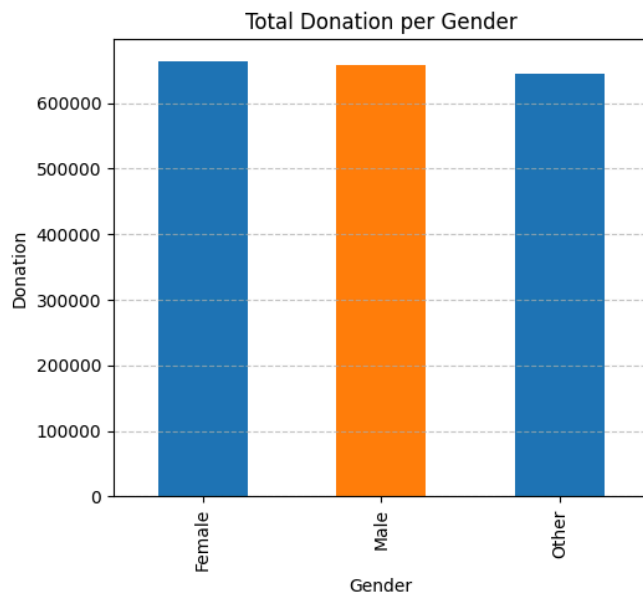
## How many time donors are likely to donate ?



Comment: Most donors donate around 2 to 4 times. We can observe that the Number of Donors decreases for higher number of donations, showing moderate donor retention.

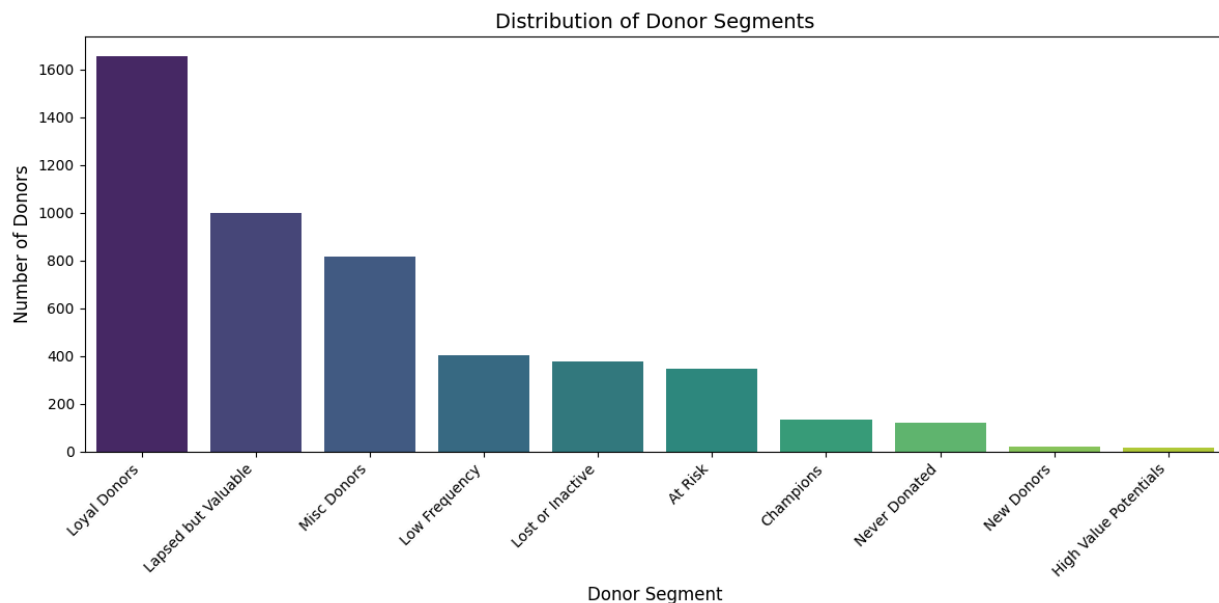
This type of information is very useful because some NPOs might want to set their new objectives to higher number of donations per donor.

## What is the donation amount repartition per gender?



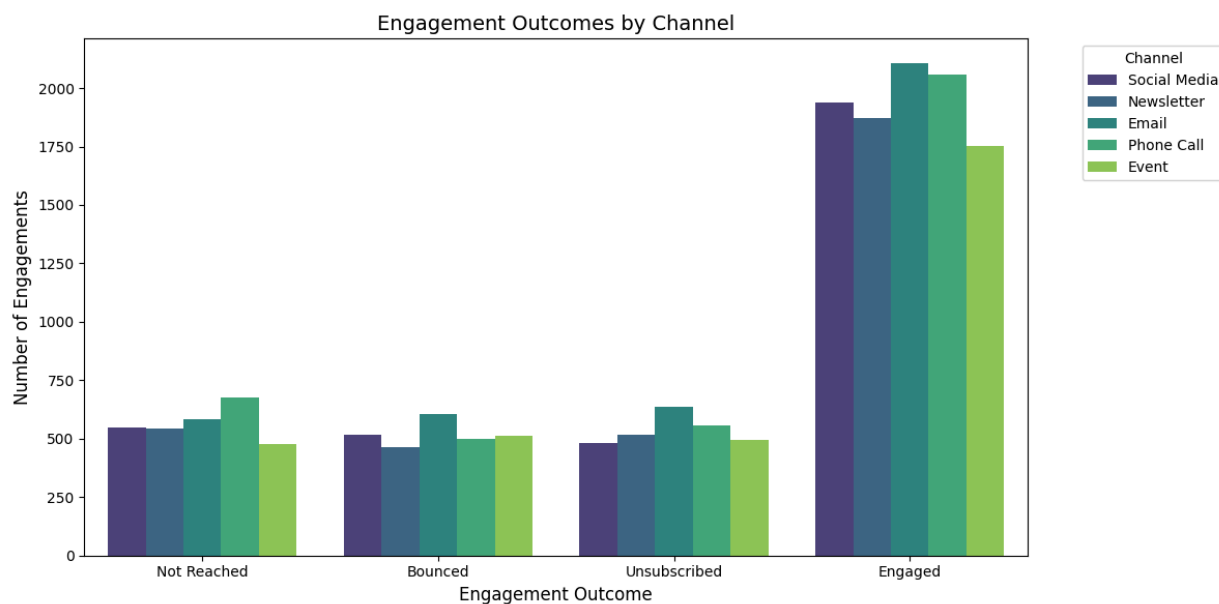
Comment: For this fictitious data set, gender is equally distributed, but in some regions, NPOs might observe some imbalance and, therefore, make it their objective to understand why and do something about it.

## How can the different type of donors be classified ?



Comments: Not every donor contributes the same. Therefore, donors have to be interacted with differently. Some donors may receive messages of recognition and heartfelt remerciements, while others may need to be re-engaged toward donating again. Knowing the different donor segments is crucial information for NPOs.

## what channel results to the best engagement outcome?



Comment: Channels seem to have approximately the same engagement outcome in the context of this analysis. However, some NPOs might have a certain channel that leads to better results. For example, in the context of Haiti, some NPOs may want to use social media instead of email, since it is more widely used.

## Modeling

=====

=====

We will develop a concrete and practical plan, along with an ML pipeline, that aligns with the objective of reducing churn and unsubscriptions, strengthening donor relationships, and improving alignment between collection and impact — with clear, actionable ML deliverables.

## Model 1

**purpose of model:** predict donor loyalty

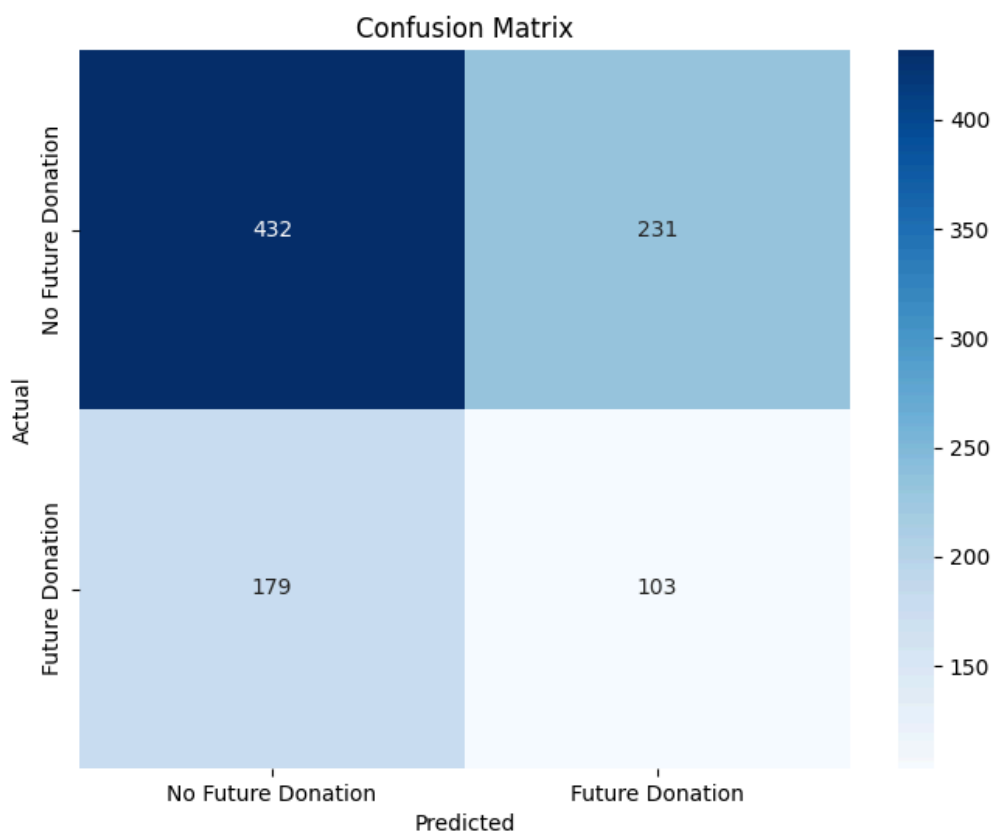
**steps taken:**

to build the model we generated new features.we splitted the data into two parts: one for the train set and one for the test set. the train set wil contain all historical data that are three years before 2025. we engineered Additional Behavioral Variables

**Chosen model:**

The model used is XGBoost Classifier, selected for its ability to handle non-linear relationships and imbalanced datasets.

### Model Evaluation

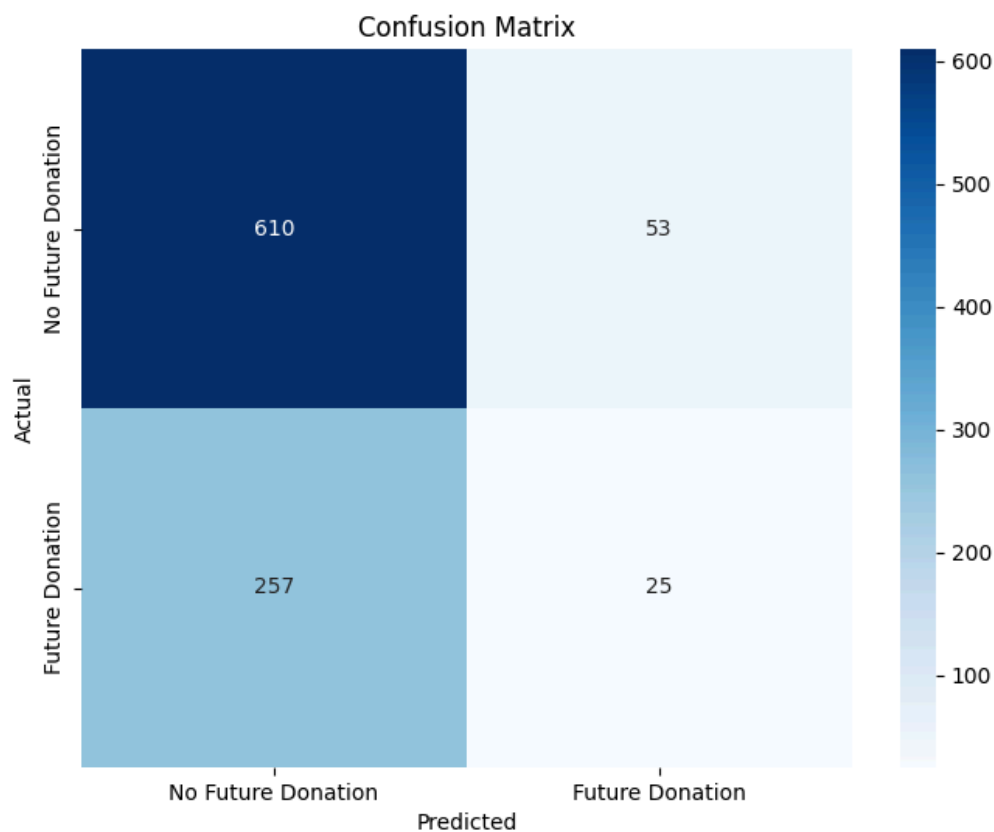


**interpretation**



The model correctly predicts donor loyalty about 57% of the time, slightly better than random (50%), but not yet reliable for decision-making. It has limited ability to distinguish between loyal and non-loyal donors (0.5 = random). It needs improvement. The model correctly identifies many non-loyal donors (432), but struggles with loyal donors (103 correctly, 179 missed)

## Model 2



### Interpretation

The model correctly predicts about 67% of loyal donors. That's higher than the XGBoost model ( $\approx 0.57$ ). We can therefore say that this one does better overall at predicting loyal donors

However, accuracy alone is misleading when classes are imbalanced (as is the case here project.)

## Model 3

**Model purpose :** this model is build to try to predict future donation amount

**Model performance:** RMSE  $\approx 165.5$  On average, the model's predictions for future donation amounts deviate by about 165 units from the actual values. Since donations vary widely, this represents a relatively large error, indicating that the predictions are not very precise.

$R^2 \approx -0.10$  An  $R^2$  value below 0 indicates that the model performs worse than a simple baseline that predicts the mean donation for all donors. This suggests that the Random Forest model fails to capture meaningful patterns from the current features for predicting future donation amounts.

## Recommendations

---

The use of data plays a crucial role in decision-making. Data can be used for inferential analysis, predictive analysis, etc... data allows organisation to make better and more informed decision overall.

## Contact information 1

---

Bellow: The Contact Information Of the Two Data scientist in charge of the

- First Name: Haender Michael
- Last Name: Jean Louis
- Email: [michaelhaenderjeanlouis@gmail.com](mailto:michaelhaenderjeanlouis@gmail.com)
- Phone Number: +509 41 75 0264
- WhatsApp : +509 33905060
- LinkedIn: [https://www.linkedin.com/in/michael-haender-jean-louis-4b7320316?utm\\_source=share&utm\\_campaign=share\\_via&utm\\_content=profile&utm\\_medium=ios\\_app](https://www.linkedin.com/in/michael-haender-jean-louis-4b7320316?utm_source=share&utm_campaign=share_via&utm_content=profile&utm_medium=ios_app)

## Contact information 2

---

- First Name : John Widno
- Last Name : DORCY
- Email : [dorcyjohnwidno97@gmail.com](mailto:dorcyjohnwidno97@gmail.com)
- Phone : +509 38 05 8388
- LinkedIn : <https://www.linkedin.com/in/john-widno-dorcy-19399a216/>

For further inquiries, feedback, or collaboration on this analysis, feel free to reach out. . .