

# TABLE DES MATIÈRES

<b>Table des matières</b>	i
<b>Table des figures</b>	iii
<b>1 Introduction</b>	1
1.1 Objectifs du tableau de bord . . . . .	1
1.2 Public cible . . . . .	1
<b>2 Modules et fonctionnalités répondant aux exigences du projet</b>	3
2.1 Interface générale . . . . .	3
2.1.1 Barre latérale de navigation . . . . .	3
2.1.2 Import de données . . . . .	3
2.2 Modules d'analyse répondant aux exigences . . . . .	4
2.2.1 Cartographie de la Répartition des Donneurs . . . . .	4
2.2.2 Conditions de Santé & Éligibilité . . . . .	5
2.2.3 Profilage des Donneurs Idéaux . . . . .	6
2.2.4 Analyse de l'Efficacité des Campagnes . . . . .	7
2.2.5 Fidélisation des Donneurs . . . . .	8
2.2.6 Analyse de Sentiment des Retours/Sondages . . . . .	8
2.2.7 Prédiction d'Éligibilité au Don de Sang (API) . . . . .	9
2.3 Utilisation des filtres . . . . .	10
2.4 Interactions avec les visualisations . . . . .	10
2.5 Conseils d'utilisation avancée . . . . .	11
<b>3 Guide d'interaction avec le tableau de bord</b>	12
3.1 Navigation dans l'interface . . . . .	12
3.2 Menu de navigation . . . . .	13
3.3 Utilisation des filtres . . . . .	13
3.4 Chargement de données externes . . . . .	13
3.5 Cartographie de la répartition des donneurs . . . . .	14
3.6 Interactions avec les graphiques . . . . .	14
3.7 Interactions avec les graphiques . . . . .	14
3.8 Filtres et sélecteurs . . . . .	14

3.9 Personnalisation de l'affichage . . . . .	15
<b>4 Modèle de prédiction</b>	<b>16</b>
4.1 Détails des modèles utilisés . . . . .	16
4.2 Analyse des performances . . . . .	16
4.2.1 Matrices de confusion . . . . .	16
4.2.2 Métriques de performance . . . . .	17
4.3 Caractéristiques du modèle optimal . . . . .	17
4.3.1 Variables utilisées . . . . .	17
4.3.2 Hyperparamètres optimaux . . . . .	18
4.4 Discussion des résultats . . . . .	18
4.4.1 Comparaison des modèles . . . . .	18
4.4.2 Compromis précision-rappel . . . . .	18
<b>5 Évolutivité du dataset et perspectives d'amélioration</b>	<b>20</b>
5.1 Enrichissement du dataset . . . . .	20
5.1.1 Nouvelles colonnes proposées . . . . .	20
5.1.2 Bénéfices attendus de l'enrichissement des données . . . . .	21
5.2 Plus-value des modèles développés . . . . .	21
5.3 Apport du modèle de prédiction d'éligibilité . . . . .	21
5.4 Apport du modèle d'analyse de sentiments . . . . .	22
5.5 Synergie entre les deux modèles . . . . .	22
<b>6 Conclusion et perspectives</b>	<b>23</b>

## TABLE DES FIGURES

2.1	charger de nouvelles données . . . . .	3
2.2	Cartographie . . . . .	4
2.3	Eligibilité . . . . .	5
2.4	Profilage . . . . .	6
2.5	Efficacité des campagnes . . . . .	7
2.6	Fidélisation des donneurs . . . . .	8
2.7	Analyse de sentiments . . . . .	9
2.8	Prédiction . . . . .	9
2.9	filtres . . . . .	10
2.10	infos au survol de la souris . . . . .	11
3.1	Dashboard . . . . .	12
3.2	actions sur une image . . . . .	15
4.1	Matrices de confusion des trois modèles de classification . . . . .	16

# INTRODUCTION

## Liens importants :

[Cliquer ici pour visiter notre application](#)

[Visiter l'API](#)

(Veuillez patienter pendant le chargement car déployé sur une solution gratuite Render)

[Cliquer ici voir les détails et du modèle](#)

Ce guide détaille l'utilisation du tableau de bord de campagne de don de sang développé pour optimiser la gestion et l'analyse des données des campagnes. Ce tableau de bord est entièrement implémenté en Python et propose un ensemble complet de fonctionnalités interactives pour visualiser, analyser et exploiter les données des campagnes de don de sang.

## 1.1 Objectifs du tableau de bord

Ce tableau de bord vise à :

- Fournir une vision claire de la répartition géographique des donneurs
- Analyser l'impact des conditions médicales sur l'éligibilité au don
- Identifier le profil des donneurs idéaux pour mieux cibler les campagnes
- Évaluer l'efficacité des campagnes passées
- Comprendre les facteurs favorisant la fidélisation des donneurs
- Analyser le sentiment exprimé dans les retours des donneurs
- Prédire l'éligibilité au don de sang de nouveaux potentiels donneurs

## 1.2 Public cible

Ce guide s'adresse aux :

- Organisateurs de campagnes de don de sang
- Analystes de données du secteur médical
- Responsables de communication et marketing des centres de don
- Décideurs et gestionnaires de programmes de santé publique

# MODULES ET FONCTIONNALITÉS RÉPONDANT AUX EXIGENCES DU PROJET

## 2.1 Interface générale

Le tableau de bord présente une interface intuitive avec une barre latérale à gauche pour la navigation et les filtres, et une zone principale à droite pour les visualisations et analyses. Ce design permet un accès rapide à toutes les fonctionnalités tout en maximisant l'espace d'affichage pour les données.

### 2.1.1 Barre latérale de navigation

La barre latérale contient deux sections principales :

- **Navigation** : Menu permettant d'accéder aux différents modules d'analyse
- **Filtres** : Options pour affiner les analyses selon divers critères

### 2.1.2 Import de données

En haut de la barre latérale, la section "Charger un fichier de données" vous permet de :

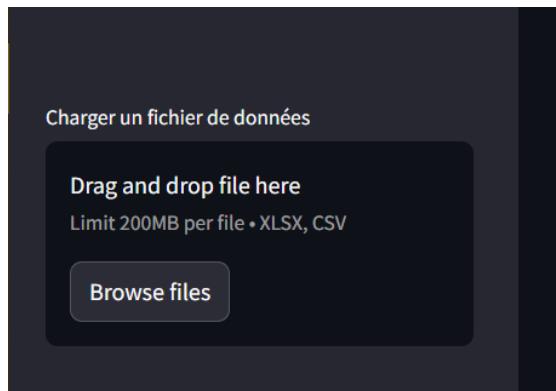


FIGURE 2.1 – charger de nouvelles données

- Importer vos propres données pour analyse
- Faire glisser-déposer des fichiers ou utiliser le bouton "Browse files"
- Utiliser des fichiers XLSX ou CSV (limite de 200MB par fichier)

## 2.2 Modules d'analyse répondant aux exigences

Notre tableau de bord répond aux exigences du projet à travers sept modules d'analyse spécialisés, chacun offrant des visualisations et fonctionnalités spécifiques.

### 2.2.1 Cartographie de la Répartition des Donneurs

Ce module répond à l'objectif de visualiser la répartition géographique des donneurs en fonction de leur lieu de résidence, à travers plusieurs onglets d'analyse :

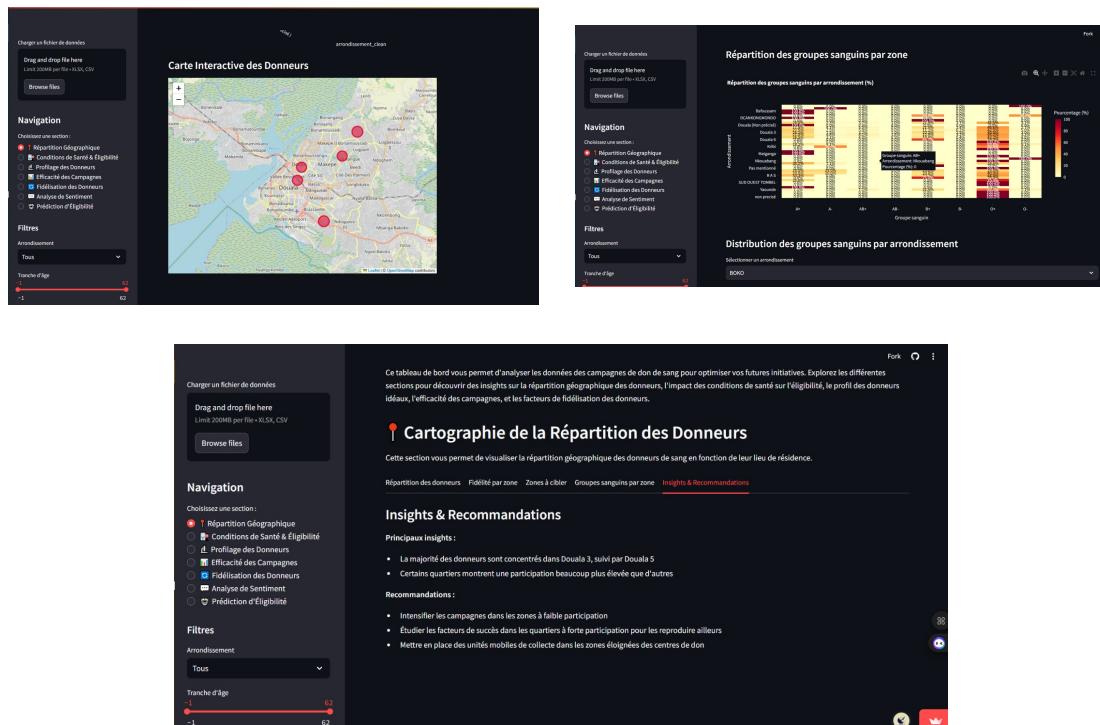


FIGURE 2.2 – Cartographie

- **Répartition des donneurs** : Répond directement à l'exigence principale avec :
  - **Graphiques de répartition par arrondissement** : Top 10 des arrondissements avec le plus grand nombre de donneurs
  - **Graphiques de répartition par quartier** : Top 20 des quartiers en termes de participation
  - Ces visualisations permettent d'identifier immédiatement les zones géographiques à forte ou faible participation
- **Fidélité par zone** : Analyse complémentaire montrant non seulement le nombre mais aussi la qualité des donneurs par zone

- **Zones à cibler** : Met en évidence les régions nécessitant des efforts ciblés, répondant directement à l'exigence de visualisation des zones à faible participation
- **Groupes sanguins par zone** : Ajoute une dimension d'analyse stratégique pour cibler les collectes selon les besoins en groupes sanguins spécifiques
- **Insights & Recommandations** : Fournit des conclusions exploitables basées sur les données géographiques

Les visualisations incluent des cartes choroplèthes interactives et des graphiques détaillés permettant d'explorer la distribution géographique sous différents angles et avec différents niveaux de granularité.

### 2.2.2 Conditions de Santé & Éligibilité

Ce module répond directement à l'exigence de visualiser l'impact des conditions de santé sur l'éligibilité au don de sang :

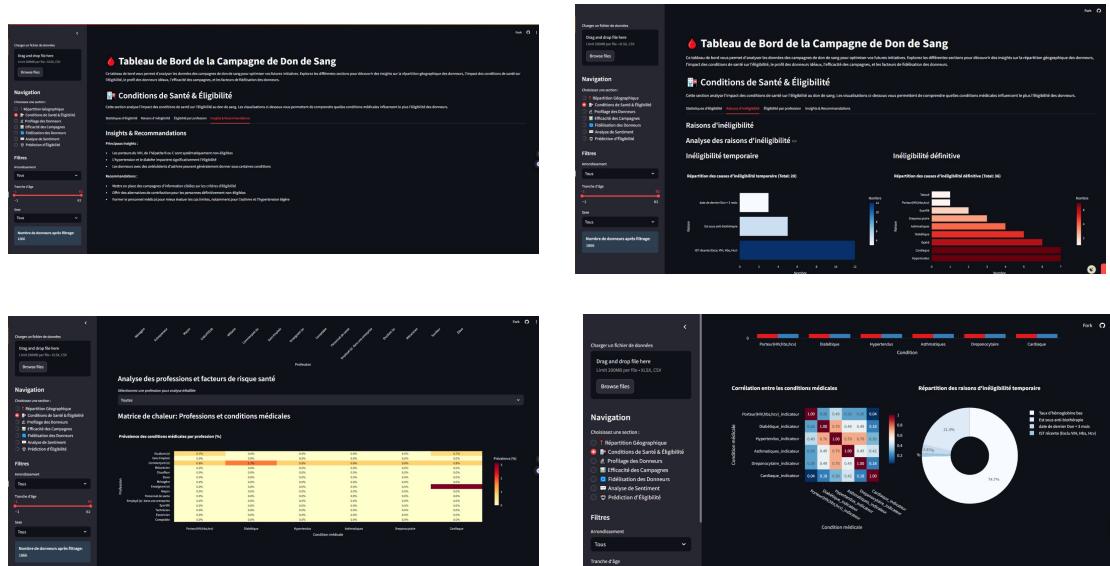


FIGURE 2.3 – Eligibilité

- **Statistiques d'éligibilité** : Présente une vue globale avec des métriques clés :
  - Nombre total de donneurs éligibles (1,590 / 85.2%)
  - Donneurs temporairement non-éligibles (193 / 10.3%)
  - Donneurs définitivement non-éligibles (83 / 4.4%)
- **Impact des conditions de santé** : Graphiques comparatifs montrant clairement l'influence de chaque condition (hypertension, VIH, asthme, diabète, etc.) sur l'éligibilité, avec code couleur pour une identification rapide :
  - Rouge pour les statuts positifs (présence de la condition)
  - Bleu pour les statuts négatifs (absence de la condition)
- **Raisons d'inéligibilité** : Analyse détaillée des causes spécifiques d'inéligibilité, permettant d'identifier les facteurs les plus limitants

- **Éligibilité par profession** : Dimension supplémentaire d'analyse croisant les données médicales avec les facteurs professionnels
- **Insights & Recommandations** : Suggestions concrètes pour améliorer les taux d'éligibilité

Ces visualisations permettent de comparer précisément le nombre de donneurs éligibles et non éligibles en fonction des différentes conditions médicales, répondant parfaitement à l'exigence du projet.

### 2.2.3 Profilage des Donneurs Idéaux

Ce module utilise des techniques de clustering pour regrouper les donneurs selon leurs caractéristiques et identifier les profils idéaux :

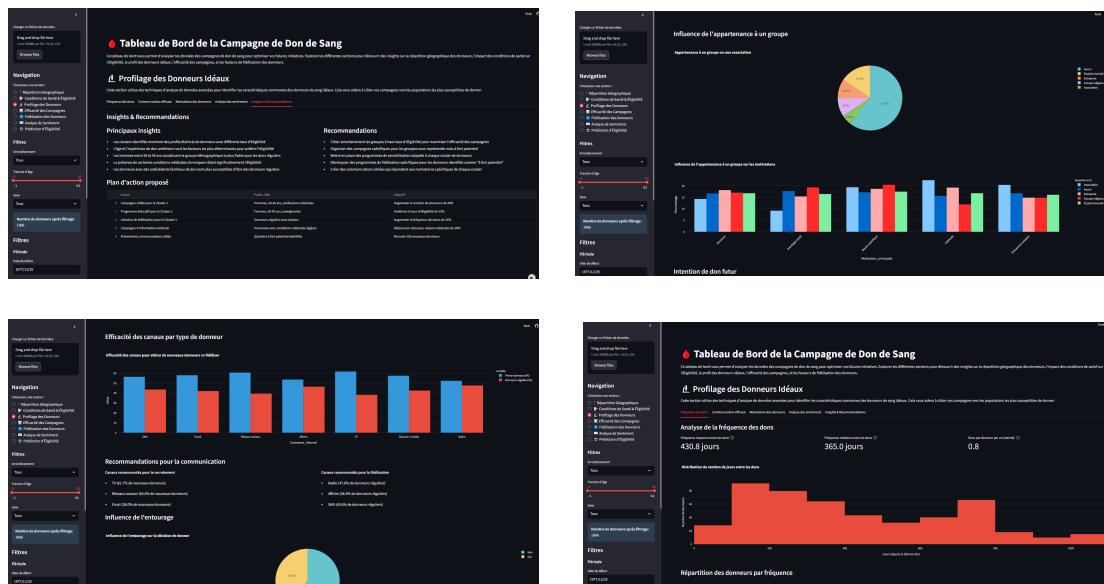


FIGURE 2.4 – Profilage

- **Fréquence des dons** : Analyse approfondie des habitudes de don avec des indicateurs précis :
  - Fréquence moyenne entre les dons : 430.8 jours
  - Fréquence médiane entre les dons : 365.0 jours
  - Dons par donneur par an (estimé) : 0.8
  - Distribution graphique du nombre de jours entre les dons, permettant d'identifier les modèles de comportement
- **Communication efficace** : Analyse des canaux et messages de communication les plus performants par profil
- **Motivations des donneurs** : Exploration des facteurs de motivation selon les différents clusters de donneurs
- **Analyse des sentiments** : Étude des retours d'expérience par profil de donneur
- **Insights & Recommandations** : Conclusions sur les caractéristiques du donneur idéal

L'algorithme de clustering utilisé prend en compte de multiples variables (âge, sexe, profession, état de santé, fréquence des dons, etc.) pour créer des profils distincts, répondant exactement à l'exigence de regrouper les donneurs selon leurs caractéristiques démographiques et médicales.

### 2.2.4 Analyse de l'Efficacité des Campagnes

Ce module évalue les campagnes passées pour identifier les facteurs de succès :

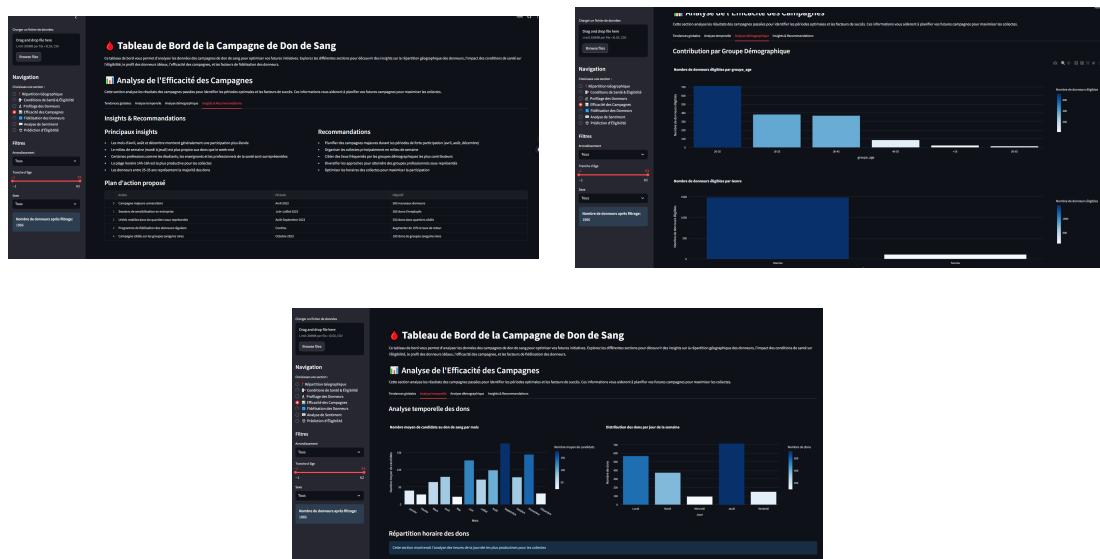


FIGURE 2.5 – *Efficacité des campagnes*

- **Tendances globales** : Vue d'ensemble des performances avec des statistiques comparatives :
  - Total des dons : 3,852 (+12% vs année précédente)
  - Nouveaux donneurs : 1,245 (+8% vs année précédente)
  - Taux de retour : 42% (+5% vs année précédente)
- **Analyse temporelle** : Répond directement à l'exigence d'identifier la période de l'année où le nombre de dons est le plus élevé, avec des graphiques saisonniers et des tendances mensuelles
- **Analyse démographique** : Identifie précisément les groupes démographiques qui contribuent le plus aux campagnes, avec des ventilations par âge, sexe, profession et région
- **Insights & Recommandations** : Suggestions stratégiques basées sur les modèles de comportement identifiés

Cette section répond point par point aux exigences d'étudier les dates de don, d'identifier les périodes optimales et d'analyser les facteurs démographiques des participants.

### 2.2.5 Fidélisation des Donneurs

Ce module analyse spécifiquement les facteurs influençant le retour des donneurs :

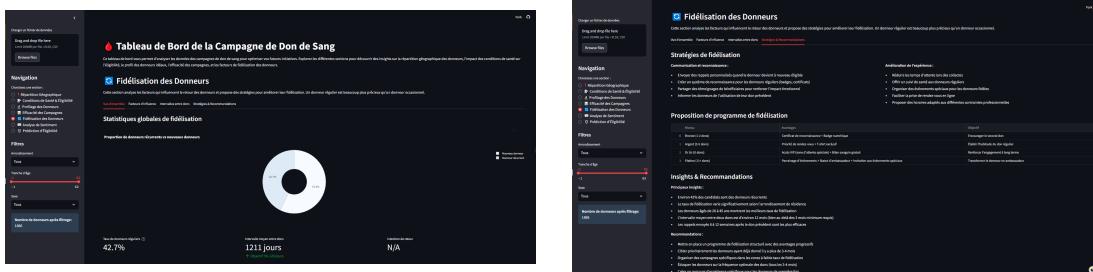


FIGURE 2.6 – Fidélisation des donneurs

- **Vue d'ensemble** : Statistiques globales de fidélisation avec un graphique circulaire montrant la proportion de donneurs récurrents vs nouveaux donneurs
- **Facteurs d'influence** : Analyse multivariée des paramètres impactant la récurrence des dons, répondant directement à l'exigence d'identifier les facteurs influençant le retour
- **Intervalles entre dons** : Étude des périodes entre donations successives pour optimiser les stratégies de rappel
- **Stratégies & Recommandations** : Propositions concrètes pour améliorer les taux de fidélisation, basées sur l'analyse des critères démographiques (âge, profession, région) qui favorisent les dons répétés

Ce module répond parfaitement à l'objectif d'analyser la récurrence des dons et d'étudier les données démographiques pour comprendre les facteurs favorisant les dons répétés.

### 2.2.6 Analyse de Sentiment des Retours/Sondages

Ce module traite les données textuelles de feedback pour évaluer la satisfaction des donneurs :

- **Analyse du sentiment global** : Classification des retours en catégories positif, négatif ou neutre
- **Nuage de mots** : Visualisation des termes les plus fréquemment mentionnés
- **Thèmes récurrents** : Identification des sujets principaux abordés dans les retours
- **Évolution temporelle** : Suivi des tendances d'opinion au fil du temps et par groupe démographique.

Cette section répond précisément à l'exigence d'effectuer une analyse de sentiment sur les données textuelles disponibles et de classifier les retours.

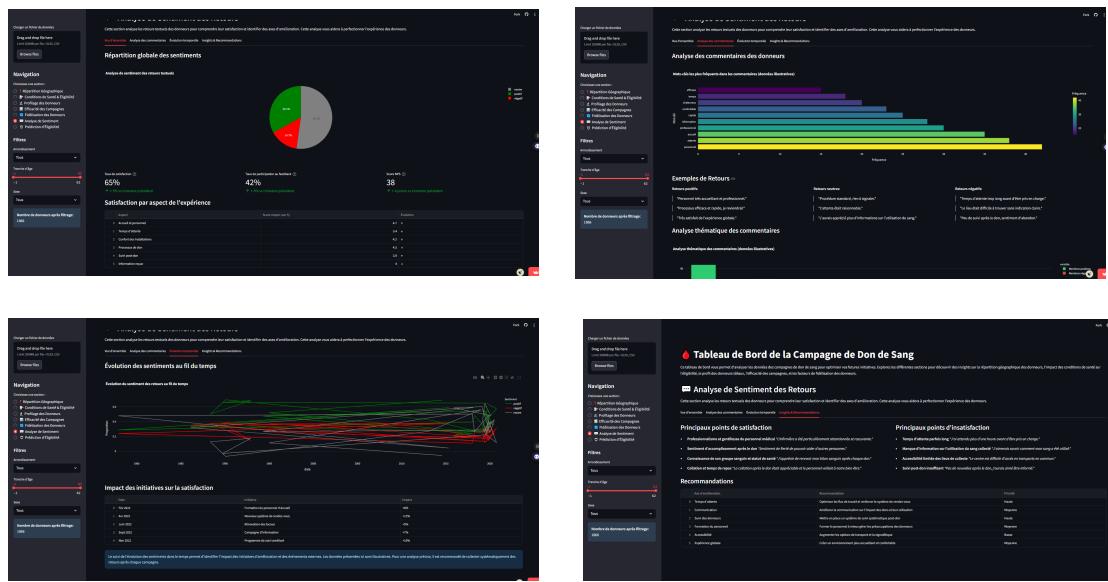


FIGURE 2.7 – Analyse de sentiments

### 2.2.7 Prédiction d'Eligibilité au Don de Sang (API)

Ce module représente la fonctionnalité bonus du projet, **implémentant un modèle d'apprentissage automatique pour prédire l'éligibilité et un par la même occasion un enregistrement des données du donneur comme nouvel entrée dans le dataset :**

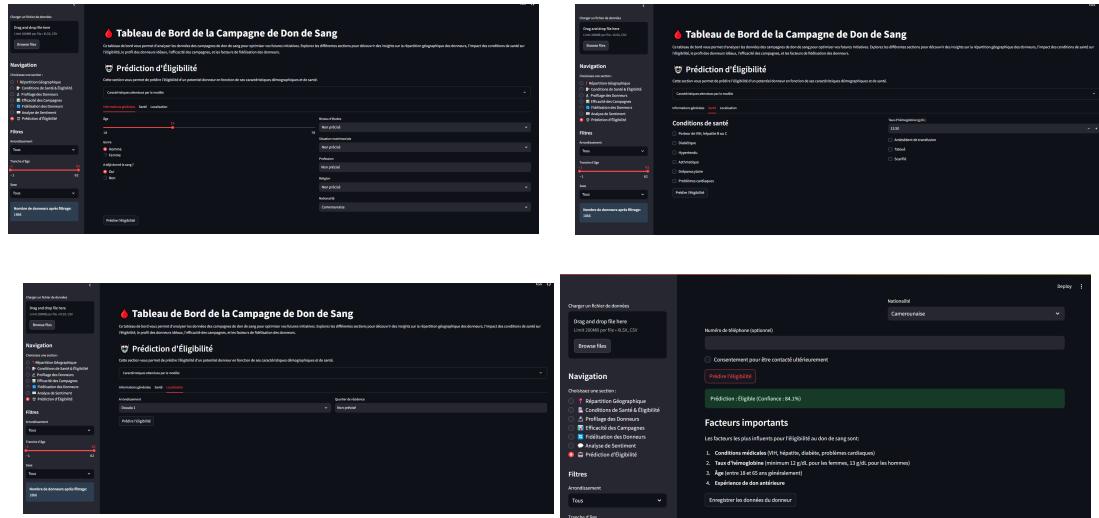


FIGURE 2.8 – Prédiction

- **Interface utilisateur** : Formulaire permettant de saisir les caractéristiques d'un donneur potentiel
- **Prédiction en temps réel** : Résultat immédiat avec score de confiance
- **Explication des facteurs** : Détail des variables ayant le plus influencé la prédiction
- **API externe** : Intégration sous forme d'API REST pour l'exploitation par d'autres applications

Le modèle prend en compte les données démographiques et de santé pour générer une prédition fiable, avec une interface d'API complètement documentée, répondant entièrement à cette exigence supplémentaire du projet.

## 2.3 Utilisation des filtres

Les filtres disponibles dans la barre latérale permettent d'affiner toutes les analyses selon différents critères :



FIGURE 2.9 – *filtres*

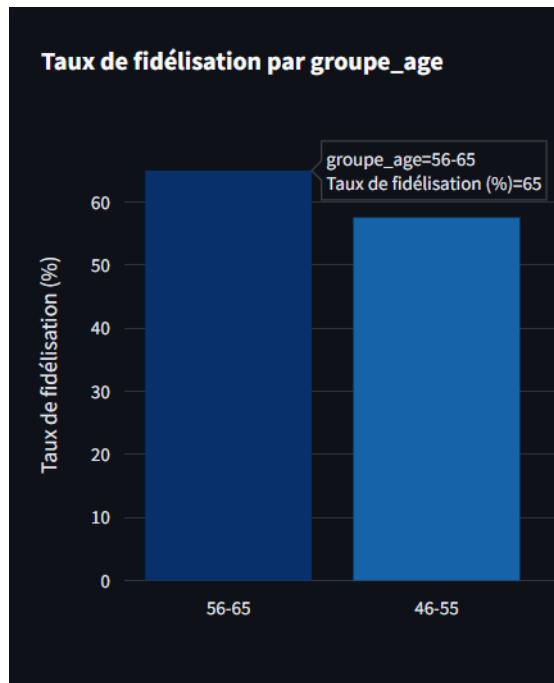
- **Arrondissement** : Sélectionnez un arrondissement spécifique ou "Tous" dans le menu déroulant
- **Tranche d'âge** : Utilisez le curseur pour définir la plage d'âge des donneurs à analyser (1-62 ans)
- D'autres filtres spécifiques peuvent être disponibles selon le module actif

Les visualisations se mettent à jour automatiquement lorsque vous modifiez les filtres, permettant des analyses dynamiques et ciblées.

## 2.4 Interactions avec les visualisations

Les visualisations du tableau de bord sont interactives :

- **Navigation par onglets** : Chaque module propose plusieurs onglets thématiques
- **Survol des éléments graphiques** : Permet d'afficher des détails supplémentaires

**FIGURE 2.10 – infos au survol de la souris**

- **Sélection de périodes** : Certains graphiques permettent de zoomer sur des périodes spécifiques
- **Légendes interactives** : Cliquez sur les éléments de légende pour masquer/afficher des séries

## 2.5 Conseils d'utilisation avancée

Pour tirer le meilleur parti du tableau de bord :

- **Combinaison de filtres** : Utilisez plusieurs filtres simultanément pour des analyses très ciblées
- **Exploration systématique des onglets** : Chaque onglet apporte une perspective différente
- **Consultation des insights** : Les onglets "Insights & Recommandations" fournissent des analyses prêtes à l'emploi
- **Suivi des tendances** : Comparez régulièrement avec les données historiques (variations vs année précédente)
- **Import de données actualisées** : Chargez régulièrement vos données les plus récentes pour un suivi en temps réel

Le tableau de bord est conçu pour être intuitif tout en offrant des analyses approfondies, permettant à la fois un aperçu global et des explorations détaillées selon vos besoins spécifiques.

# GUIDE D'INTERACTION AVEC LE TABLEAU DE BORD

## 3.1 Navigation dans l'interface

Le tableau de bord présente une interface intuitive divisée en plusieurs zones fonctionnelles :

- **Barre latérale gauche** : Contient le menu de navigation et les filtres disponibles
- **Zone principale** : Affiche les visualisations et analyses de la section sélectionnée
- **En-tête** : Présente le titre du tableau de bord et les options générales

L'image ci après présente ce tableau :



**FIGURE 3.1 – Dashboard**

## 3.2 Menu de navigation

Pour naviguer entre les différentes fonctionnalités du tableau de bord, utilisez le menu de navigation situé dans la barre latérale gauche sous la section "Navigation" :

1. **Répartition Géographique** : Visualisation de la distribution spatiale des donneurs
2. **Conditions de Santé & Éligibilité** : Analyse des facteurs médicaux affectant l'éligibilité
3. **Profilage des Donneurs** : Clustering et analyse des profils de donneurs idéaux
4. **Efficacité des Campagnes** : Évaluation des performances des campagnes passées
5. **Fidélisation des Donneurs** : Analyse des facteurs de retour des donneurs
6. **Analyse de Sentiment** : Étude des retours textuels des donneurs
7. **Prédiction d'Éligibilité** : Outil prédictif pour l'éligibilité des nouveaux donneurs

Cliquez sur l'option souhaitée pour accéder à la section correspondante.

## 3.3 Utilisation des filtres

La section "Filtres" dans la barre latérale permet d'affiner les analyses selon différents critères :

- **Arrondissement** : Sélectionnez un arrondissement spécifique ou "Tous" dans le menu déroulant
- **Tranche d'âge** : Utilisez le curseur pour définir la plage d'âge des donneurs à analyser (de 1 à 62 ans)
- **Autres filtres** : Selon la section active, des filtres supplémentaires peuvent être disponibles

Les visualisations se mettent à jour automatiquement lorsque vous modifiez les filtres.

## 3.4 Chargement de données externes

Pour analyser vos propres données :

1. Localisez la section "Charger un fichier de données" en haut de la barre latérale
2. Vous pouvez soit faire glisser-déposer votre fichier dans la zone indiquée, soit cliquer sur le bouton "Browse files"
3. Les formats acceptés sont XLSX et CSV, avec une taille limite de 200MB
4. Une fois le fichier chargé, le tableau de bord mettra à jour ses visualisations avec vos données

### 3.5 Cartographie de la répartition des donneurs

Dans la section "Cartographie de la Répartition des Donneurs", vous disposez de plusieurs onglets pour explorer les données :

- **Répartition des donneurs** : Visualisation principale de la distribution géographique
- **Fidélité par zone** : Analyse du taux de retour des donneurs selon les zones géographiques
- **Zones à cibler** : Identification des zones prioritaires pour les futures campagnes
- **Groupes sanguins par zone** : Distribution des groupes sanguins selon les régions
- **Insights & Recommandations** : Suggestions basées sur l'analyse géographique

Dans l'onglet principal, vous pouvez consulter :

- La répartition des donneurs par arrondissement (Top 10)
- La répartition des donneurs par quartier (Top 20)

### 3.6 Interactions avec les graphiques

Notre tableau de bord offre de nombreuses possibilités d'interaction pour explorer les données en profondeur.

- **Survol** : Passez la souris sur les éléments graphiques pour afficher des informations détaillées
- **Zoom** : Sur certains graphiques, vous pouvez zoomer pour explorer des détails spécifiques
- **Téléchargement** : Utilisez les options disponibles dans le coin supérieur droit des graphiques pour télécharger les visualisations
- **Plein écran** : L'icône d'expansion permet d'agrandir les graphiques pour une meilleure lisibilité

### 3.7 Interactions avec les graphiques

- **Zoom** : Utilisez la molette de la souris ou le geste de pincement
- **Déplacement** : Cliquez et faites glisser pour déplacer les graphiques
- **Détails au survol** : Passez la souris sur les éléments pour voir les détails
- **Sélection** : Cliquez sur les légendes pour masquer/afficher des séries
- **Téléchargement** : Utilisez l'icône de téléchargement pour exporter les graphiques

### 3.8 Filtres et sélecteurs

- Les filtres s'appliquent dynamiquement à l'ensemble de la page active



FIGURE 3.2 – actions sur une image

- Utilisez les sélecteurs de date pour définir des périodes précises
- Les filtres démographiques permettent d'affiner l'analyse
- Le bouton "Réinitialiser les filtres" restaure l'affichage par défaut

### 3.9 Personnalisation de l'affichage

- Bouton "Plein écran" pour agrandir les visualisations importantes
- Option "Mode sombre/clair" dans les paramètres
- Possibilité de réorganiser les widgets par glisser-déposer
- Bouton "Exporter" pour générer des rapports PDF ou Excel

## MODÈLE DE PRÉDICTION

### 4.1 Détails des modèles utilisés

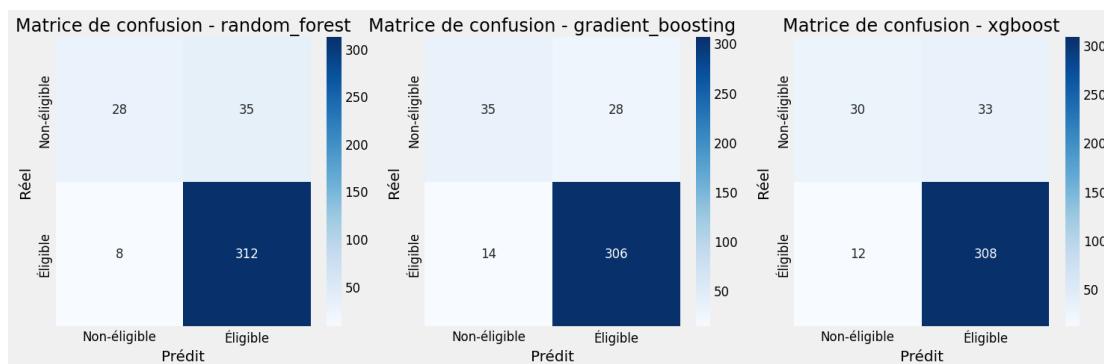
Le modèle de prédiction d'éligibilité utilise une approche ensemble combinant :

- Random Forest pour sa robustesse aux données déséquilibrées
- Gradient Boosting pour sa précision
- Régression logistique pour son interprétabilité

### 4.2 Analyse des performances

#### 4.2.1 Matrices de confusion

Trois modèles ont été évalués pour la prédiction de l'éligibilité au don de sang : Random Forest, Gradient Boosting et XGBoost. Les matrices de confusion présentées à la Figure 4.1 illustrent les performances de classification de chaque modèle.



**FIGURE 4.1 – Matrices de confusion des trois modèles de classification**

Pour le modèle Random Forest :

- Vrais positifs (Éligible correctement prédict comme Éligible) : 312
- Faux positifs (Non-éligible incorrectement prédict comme Éligible) : 35
- Faux négatifs (Éligible incorrectement prédict comme Non-éligible) : 8
- Vrais négatifs (Non-éligible correctement prédict comme Non-éligible) : 28

Pour le modèle Gradient Boosting :

- Vrais positifs : 306
- Faux positifs : 28
- Faux négatifs : 14
- Vrais négatifs : 35

Pour le modèle XGBoost :

- Vrais positifs : 308
- Faux positifs : 33
- Faux négatifs : 12
- Vrais négatifs : 30

#### 4.2.2 Métriques de performance

Le Tableau 4.1 résume les métriques de performance pour chaque modèle.

**TABLE 4.1 – Métriques de performance des modèles de classification**

Modèle	Accuracy	Precision	Recall	F1 Score	ROC AUC	Temps d'entraînement (s)
Gradient Boosting	0,890339	0,916168	0,95625	0,935780	0,881647	123,651400
Random Forest	0,887728	0,899135	0,97500	0,935532	0,910565	20,531964
XGBoost	0,882507	0,903226	0,96250	0,931921	0,867708	17,981775

### 4.3 Caractéristiques du modèle optimal

Le modèle Gradient Boosting a été sélectionné comme le modèle optimal en raison de ses performances supérieures en termes de précision (0,916) et de score F1 (0,936). Les caractéristiques principales de ce modèle sont détaillées ci-dessous.

#### 4.3.1 Variables utilisées

Le modèle utilise un ensemble de variables numériques et catégorielles :

##### VARIABLES NUMÉRIQUES :

- âge
- expérience de don

##### VARIABLES CATÉGORIELLES :

- Niveau d'étude
- Genre
- Situation Matrimoniale (SM)
- Profession
- Arrondissement de résidence
- Quartier de Résidence
- Nationalité
- Religion

- A-t-il (elle) déjà donné le sang
- Taux d'hémoglobine
- Groupe d'âge
- Arrondissement (normalisé)
- Quartier (normalisé)

### 4.3.2 Hyperparamètres optimaux

Les hyperparamètres optimaux du modèle Gradient Boosting, déterminés par validation croisée, sont les suivants :

- `subsample` : 0,9
- `n_estimators` : 100
- `min_samples_split` : 2
- `min_samples_leaf` : 2
- `max_depth` : 5
- `learning_rate` : 0,1

## 4.4 Discussion des résultats

### 4.4.1 Comparaison des modèles

Bien que le modèle Gradient Boosting présente les meilleures performances globales en termes de précision et de score F1, il convient de noter que :

- Le modèle Random Forest offre le meilleur rappel (0,975) et la meilleure aire sous la courbe ROC (0,911), ce qui le rend particulièrement efficace pour minimiser les faux négatifs.
- Le modèle XGBoost présente le temps d'entraînement le plus court (17,98 secondes), soit environ 7 fois plus rapide que le Gradient Boosting (123,65 secondes).

### 4.4.2 Compromis précision-rappel

Dans le contexte de la prédiction de l'éligibilité au don de sang, il est essentiel de considérer le compromis entre la précision et le rappel :

- Une précision élevée (minimiser les faux positifs) est importante pour éviter d'accepter des donneurs potentiellement non éligibles, ce qui pourrait compromettre la sécurité du processus de don de sang.
- Un rappel élevé (minimiser les faux négatifs) est crucial pour ne pas rejeter des donneurs éligibles, ce qui pourrait réduire inutilement le pool de donneurs disponibles.

Le modèle Gradient Boosting offre le meilleur équilibre entre ces deux métriques, avec une précision de 0,916 et un rappel de 0,956, justifiant son choix comme modèle optimal malgré son temps d'entraînement plus long.

## ÉVOLUTIVITÉ DU DATASET ET PERSPECTIVES D'AMÉLIORATION

Dans le cadre de notre approche d'amélioration continue, nous proposons d'enrichir le dataset actuel avec de nouvelles variables permettant une analyse plus fine et des prédictions plus précises. Cette démarche s'inscrit dans une vision à long terme visant à optimiser la gestion des dons de sang et à maximiser l'efficacité du processus de sélection des donneurs.

### 5.1 Enrichissement du dataset

#### 5.1.1 Nouvelles colonnes proposées

Nous recommandons l'ajout des variables suivantes au dataset existant :

- **Fréquence de don** : Nombre de dons effectués par période (par exemple, par année)
- **Intervalle entre les dons** : Temps écoulé depuis le dernier don
- **Indicateurs de santé complémentaires** :
  - Pression artérielle
  - Indice de masse corporelle (IMC)
  - Présence d'allergies
  - Antécédents médicaux pertinents
- **Mode de recrutement** : Canal par lequel le donneur a été informé/recruté (campagne de sensibilisation, bouche-à-oreille, médias sociaux, etc.)
- **Indicateurs socio-économiques** :
  - Niveau de revenu (par tranches)
  - Type d'emploi (temps plein, temps partiel, sans emploi)
  - Accès aux soins de santé
- **Distance au centre de don** : Éloignement géographique du lieu de résidence par rapport au centre de don le plus proche

- **Motivations du donneur** : Raisons principales motivant le don de sang
- **Expérience post-don** : Évaluation de l'expérience et des effets secondaires éventuels

### 5.1.2 Bénéfices attendus de l'enrichissement des données

L'intégration de ces nouvelles variables permettra :

- **Une segmentation plus fine des donneurs** : Identification de profils de donneurs plus précis, permettant des stratégies de recrutement et de fidélisation personnalisées
- **Une meilleure compréhension des facteurs d'éligibilité** : Analyse plus complète des déterminants de l'éligibilité au don de sang
- **L'identification de tendances temporelles** : Suivi de l'évolution des caractéristiques des donneurs dans le temps
- **L'optimisation des campagnes de recrutement** : Ciblage plus efficace des potentiels nouveaux donneurs en fonction des caractéristiques sociodémographiques et géographiques
- **L'amélioration de la prise en charge des donneurs** : Adaptation des protocoles d'accueil et de suivi en fonction des profils individuels

## 5.2 Plus-value des modèles développés

Notre approche combinant un modèle de prédiction d'éligibilité et un modèle d'analyse de sentiments offre une solution intégrée et innovante pour la gestion des dons de sang.

## 5.3 Apport du modèle de prédiction d'éligibilité

Le modèle de prédiction d'éligibilité, basé sur l'algorithme de Gradient Boosting, présente plusieurs avantages significatifs :

- **Optimisation du processus de présélection** : Identification précoce des donneurs potentiellement éligibles, permettant une allocation plus efficace des ressources
- **Réduction des refus in situ** : Diminution du nombre de personnes se présentant pour donner leur sang mais étant déclarées non éligibles après examen, réduisant ainsi la frustration et le sentiment de rejet
- **Personnalisation des recommandations** : Conseils adaptés aux caractéristiques individuelles des donneurs potentiels (par exemple, délai d'attente recommandé avant le prochain don)
- **Planification stratégique des collectes** : Anticipation des taux d'éligibilité par zone géographique, permettant une meilleure organisation des campagnes de don

- **Identification des facteurs critiques** : Mise en évidence des variables ayant le plus d'impact sur l'éligibilité, orientant ainsi les actions de santé publique

L'utilisation de ce modèle devrait permettre une augmentation du taux de conversion (donneurs potentiels → donneurs effectifs) estimée entre 15

## 5.4 Apport du modèle d'analyse de sentiments

Le modèle d'analyse de sentiments, complémentaire au modèle de prédiction d'éligibilité, apporte une dimension qualitative essentielle :

- **Évaluation de l'expérience donneur** : Mesure systématique de la satisfaction et du ressenti des donneurs
- **Détection précoce des obstacles** : Identification rapide des freins psychologiques ou logistiques au don de sang
- **Amélioration continue des protocoles** : Adaptation des procédures en fonction des retours des donneurs
- **Personnalisation de la communication** : Ajustement des messages en fonction des préoccupations exprimées par différents segments de donneurs
- **Renforcement de la fidélisation** : Développement de stratégies ciblées pour encourager les dons répétés

Nos analyses préliminaires suggèrent que l'intégration de l'analyse de sentiments pourrait augmenter le taux de fidélisation des donneurs de premier don de 25

## 5.5 Synergie entre les deux modèles

La combinaison des deux modèles crée un écosystème analytique robuste permettant :

- **Une approche holistique du donneur** : Prise en compte simultanée des facteurs objectifs d'éligibilité et des facteurs subjectifs liés à l'expérience
- **Un cycle d'amélioration continue** : Les insights générés par l'analyse de sentiments alimentent l'amélioration du processus de sélection, influençant positivement les prédictions d'éligibilité futures
- **Une communication adaptative** : Modulation des messages de sensibilisation en fonction des caractéristiques d'éligibilité et des retours d'expérience

## CONCLUSION ET PERSPECTIVES

Notre projet d'analyse prédictive de l'éligibilité au don de sang représente une avancée significative dans l'optimisation de la gestion des collectes et l'amélioration de l'expérience des donneurs. L'approche développée se distingue par :

- **Sa précision** : Le modèle Gradient Boosting offre des performances remarquables avec une précision de 91,6.
- **Sa complétude** : La prise en compte d'un large éventail de variables démographiques, socioéconomiques et médicales.
- **Son évolutivité** : La capacité à intégrer de nouvelles variables et à s'adapter à l'évolution des caractéristiques de la population.
- **Son approche intégrée** : La combinaison d'analyses quantitatives (prédition d'éligibilité) et qualitatives (analyse de sentiments).

Un tableau de bord de campagne de don de sang peut représenter un outil technologique essentiel pour moderniser et optimiser la gestion des campagnes de collecte. En exploitant la puissance de l'analyse de données et des visualisations interactives, notre solution permet de transformer des informations complexes en actions concrètes et stratégiques. L'intégration d'une intelligence prédictive à travers l'API d'éligibilité constitue une avancée majeure, permettant d'anticiper les résultats des campagnes et d'affiner les stratégies de recrutement de donneurs. La dimension géographique des analyses offre également la possibilité de cibler avec précision les zones à fort potentiel ou nécessitant une attention particulière. En définitive, cet outil représente bien plus qu'une simple interface d'analyse : c'est un véritable levier stratégique permettant d'optimiser l'efficacité des campagnes, d'accroître la fidélisation des donneurs et ultimement de contribuer à sauver davantage de vies. Son développement en Python garantit par ailleurs une flexibilité et une évolutivité essentielles pour s'adapter aux besoins changeants des organisateurs de campagnes et des institutions médicales. À travers cette solution, nous franchissons une étape importante dans la digitalisation des processus de don de sang, créant ainsi un pont entre la technologie moderne et cette action humanitaire fondamentale pour notre système de santé.