

# Rapport d'Analyse du Projet de Tableau de Bord pour Donneurs de Sang Réalisé par l'équipe SENDA CONSULTING

---

## Rapport d'Analyse du Projet de Tableau de Bord pour Donneurs de Sang

Lien github : <https://github.com/vic1er03/Don-de-sang.git>

### Résumé

Ce rapport présente l'analyse d'un projet développé dans le cadre d'un concours de data visualisation portant sur l'analyse des données de donneurs de sang. Le projet comprend un tableau de bord interactif développé avec Streamlit et un modèle prédictif d'éligibilité au don de sang. L'objectif principal était de créer un outil permettant aux organisateurs de campagnes de don de sang d'optimiser leurs stratégies en se basant sur des insights tirés des données.

Après une analyse approfondie, nous concluons que le projet répond avec succès à la majorité des objectifs fixés par le concours, obtenant un score global estimé à . Les principales forces du projet sont l'interactivité du tableau de bord, la pertinence des visualisations et la performance du modèle prédictif.

### Introduction

Le don de sang est une composante essentielle des systèmes de santé modernes, permettant de sauver des vies lors d'interventions chirurgicales, d'accidents ou pour les patients atteints de maladies chroniques. Optimiser les campagnes de don de sang représente donc un enjeu majeur de santé publique.

Ce projet s'inscrit dans le cadre d'un concours visant à développer un tableau de bord interactif pour analyser les données de donneurs de sang et prédire l'éligibilité au don. L'objectif est de fournir aux organisateurs de campagnes des outils d'analyse performants pour améliorer l'efficacité de leurs actions.

### Méthodologie d'Analyse

Notre analyse s'est basée sur l'examen des fichiers suivants :

- Scripts Python (`train\_model.py`, `dashboard\_final.py`)
- Notebook Jupyter (`analyse\_donneurs\_sang.ipynb`)
- Fichiers de modèle (`eligibility\_model.pkl`, `label\_encoders.pkl`)

- Documents de description (`Challenge\_Proposal.pdf`, `Dataset Description.pdf`)
- Jeu de données (`Challenge dataset.xlsx`)

Nous avons abordé le projet selon les critères définis dans le document du concours :

1. Fonctionnalité
2. Utilisabilité
3. Insights sur les données
4. Innovation
5. Qualité du code

## Analyse des Objectifs du Concours

### Objectifs Principaux

Le concours définissait quatre objectifs principaux avec 7 objectifs spécifiques :

1. Création d'un tableau de bord interactif pour l'analyse des données de donneurs de sang
2. Développement de visualisations innovantes permettant d'explorer les données
3. Implémentation d'un modèle prédictif pour déterminer l'éligibilité au don (fonctionnalité bonus)
4. Documentation complète du projet

### Fonctionnalités Requises

Les fonctionnalités requises incluaient :

- Visualisations interactives (démographiques, géographiques, facteurs de santé, temporelles)
- Filtres et interface utilisateur intuitive
- Modèle prédictif intégré au tableau de bord (bonus)
- Documentation complète

## Évaluation de l'Atteinte des Objectifs

### 1. Tableau de Bord Interactif

**\*\*Objectif atteint ☒\*\***

Le fichier `dashboard\_final.py` implémente un tableau de bord Streamlit complet avec :

- Navigation par barre latérale permettant d'accéder à différentes sections d'analyse
- Filtres interactifs pour explorer les données selon différents critères
- Interface utilisateur claire et professionnelle
- Multiples visualisations couvrant tous les aspects requis

Le tableau de bord est structuré en plusieurs sections :

- Aperçu des données
- Distribution géographique des donneurs
- Conditions de santé et éligibilité
- Profils des donneurs
- Analyse des campagnes
- Fidélisation des donneurs
- Analyse de sentiment
- Prédiction d'éligibilité

## 2. Visualisations Innovantes

**\*\*Objectif atteint ☒\*\***

Le tableau de bord inclut diverses visualisations couvrant tous les aspects requis :

**Analyse démographique :** Le projet propose des visualisations détaillées des profils des donneurs, incluant l'âge, le sexe, la profession et le niveau d'éducation. Ces visualisations permettent d'identifier les segments de population les plus susceptibles de donner leur sang.

**Analyse géographique :** Des cartes interactives (``create_map``, ``create_geo_map``) montrent la distribution géographique des donneurs, permettant d'identifier les zones à forte ou faible participation et d'adapter les stratégies de campagne en conséquence.

**Analyse des facteurs de santé :** Des graphiques détaillés (``create_health_eligibility_chart``) illustrent l'impact des différentes conditions de santé sur l'éligibilité au don, aidant à comprendre les principales raisons d'inéligibilité.

**Analyses avancées :** Le projet va au-delà des exigences de base en incluant :

- Clustering des donneurs (``create_donor_clustering``) pour identifier des segments de population

- Analyse des campagnes (``create_campaign_analysis``) pour évaluer l'efficacité des différentes stratégies
- Analyse de la fidélisation des donateurs (``create_donor_retention_analysis``) pour comprendre les facteurs influençant les dons répétés
- Analyse de sentiment (``create_sentiment_analysis``) pour traiter les retours textuels des donateurs

### 3. Modèle Prédicatif d'Éligibilité (Bonus)

Objectif atteint ☒

Le projet inclut un modèle prédictif performant pour déterminer l'éligibilité au don de sang :

- Plusieurs algorithmes ont été testés (RandomForest, SVM, KNN, LogisticRegression)
- RandomForest a été sélectionné comme meilleur modèle avec un score F1 de 0.85
- Le modèle est intégré au tableau de bord pour des prédictions en temps réel
- Le prétraitement des données est robuste, avec gestion des valeurs manquantes et encodage approprié des variables catégorielles

Le script ``train_model.py`` montre une approche méthodique pour la sélection et l'entraînement du modèle, avec utilisation de validation croisée et optimisation des hyperparamètres via GridSearchCV.

### Points d'Attention Particuliers

Le projet a bien pris en compte les points d'attention mentionnés dans le document du concours :

Confidentialité des données : Les données semblent être correctement anonymisées, avec des identifiants uniques pour les donateurs.

Déséquilibre des données : Le modèle prend en compte les classes déséquilibrées avec l'utilisation du score F1 pondéré comme métrique d'évaluation.

Complexité: Le tableau de bord maintient un bon équilibre entre fonctionnalités avancées et clarté, évitant la surcharge d'information.

Prédiction en temps réel : Le modèle est intégré au tableau de bord pour des prédictions instantanées, permettant une utilisation pratique lors des campagnes.

## Conclusion

Le projet répond avec succès à la majorité des objectifs fixés par le concours. Le tableau de bord est complet, interactif et fournit des insights précieux pour les campagnes de don de sang. Le modèle prédictif d'éligibilité est bien implémenté et performant.

Notre projet représente une contribution significative à l'optimisation des campagnes de don de sang, offrant aux organisateurs des outils d'analyse puissants pour améliorer l'efficacité de leurs actions et, in fine, augmenter le nombre de dons de sang disponibles pour les patients qui en ont besoin.