Sheet 2019

Dataset Overview:

The dataset contains information about individuals who have participated in a blood donation campaign. The dataset includes a variety of demographic, health-related, and logistical details that are useful for analyzing and optimizing blood donation campaigns.

Key Columns/Attributes:

1. **Demographic Information**:
    - **Date de remplissage de la fiche**: The date when the donor registration form was filled. This could be useful for analyzing trends over time or campaign effectiveness.
    - **Date de naissance**: The birthdate of the individual, which can be used to calculate the age of the donor.
    - **Niveau d'etude**: The educational level of the individual. Categories include options like "Pas Précisé" (Not specified), "Universitaire" (University), "Secondaire" (Secondary), etc.
    - **Genre**: Gender of the individual (e.g., "Homme" for Male, "Femme" for Female).
    - **Taille**: Height of the individual (in meters or other units).
    - **Poids**: Weight of the individual (in kilograms or other units).
    - **Situation Matrimoniale (SM)**: Marital status of the individual (e.g., "Célibataire" for Single, "Marié(e)" for Married).
    - **Profession**: The profession or occupation of the individual (e.g., "Secrétaire comptable" for Accounting Secretary, "Chaudronnier" for Boilermaker).
2. **Health Conditions and Eligibility**:
    - **Raison de non-eligibilité totale**: Reasons why a donor is not eligible for donation. This includes a variety of health-related conditions.
    - **[Porteur(HIV,hbs,hcv)]**: Indicates whether the donor is a carrier of any blood-borne infections, such as HIV, Hepatitis B (HBS), or Hepatitis C (HCV).
    - **[Opéré]**: Indicates whether the donor has undergone any surgeries.
    - **[Drepanocytaire]**: Indicates whether the donor has sickle cell disease (Drepanocytosis).
    - **[Diabétique]**: Indicates whether the donor is diabetic.
    - **[Hypertendus]**: Indicates whether the donor has hypertension.
    - **[Asthmatiques]**: Indicates whether the donor has asthma.
    - **[Cardiaque]**: Indicates whether the donor has any heart-related conditions.
    - **[Tatoué]**: Indicates whether the donor has tattoos.
    - **[Scarifié]**: Indicates whether the donor has scars (perhaps related to other health conditions).
3. **Geographical Information**:

    - **Arrondissement de résidence**: The district or area where the donor resides (e.g., Douala 1, Douala 5).
    - **Quartier de Résidence**: The specific neighborhood or residential area of the donor (e.g., "Logbaba", "Bepanda").
4. **Other Variables**:
    - **Si autres raison préciser**: This column may contain additional feedback or reasons for non-eligibility. It could be useful for sentiment analysis to identify common reasons for not donating.
5. **Missing Values**:
    - Many of the columns, such as **Taille**, **Poids**, or health conditions, may have missing values. This will need to be handled during data preprocessing.

Data Format:

- **Date Columns**: The dates are presented in the format like "DD/MM/YYYY", but some may have more complex formats or inconsistent entries.
- **Categorical Columns**: The data includes multiple categorical columns that might require encoding for use in machine learning models (e.g., education level, profession, marital status).
- **Boolean/Flag Columns**: Health conditions (e.g., HIV carrier, diabetic, hypertensive) are typically marked with specific flags or indicators, such as "Oui" (Yes) or "Non" (No).

Potential Applications of the Dataset:

- **Blood Donation Eligibility Prediction**: Using demographic and health-related information to predict whether an individual is eligible to donate blood.
- **Geographic Distribution**: Mapping the geographical areas with higher or lower donor participation to optimize campaign locations.
- **Health Condition Analysis**: Investigating how certain health conditions impact blood donation eligibility.
- **Campaign Effectiveness**: Analyzing which factors (e.g., time of year, gender, education) correlate with higher donation rates.
- **Donor Profiling**: Using clustering techniques to identify ideal donor profiles for targeted outreach.

This dataset offers a comprehensive set of variables that can be analyzed for improving blood donation campaigns. It provides demographic, medical, and geographical data that can be used to derive valuable insights into donor behavior, eligibility, and campaign strategies.

Sheet Volontaire

The **"Volontaire"** sheet provides another dataset of individuals related to the blood donation campaign. This dataset contains additional information for each donor, including their **ID**, **age**, and **health conditions**. Below are the key attributes from this sheet:

Key Columns/Attributes in the "Volontaire" Sheet:

1. **Identification and Demographic Information**:
    - **ID**: A unique identifier for each donor (e.g., "DONOR_1842").
    - **Age**: The age of the donor.
    - **Horodateur**: Timestamp (likely when the data was recorded).
    - **Niveau_d'etude**: Educational level (e.g., "Universitaire", "Secondaire", "Aucun" for none).
    - **Genre_**: Gender of the donor (e.g., "Homme", "Femme").
    - **Taille_**: Height of the donor (missing for some records).
    - **Poids**: Weight of the donor (missing for some records).
    - **Situation_Matrimoniale_(SM)**: Marital status (e.g., "Célibataire" for Single, "Marié(e)" for Married).
    - **Profession_**: The profession or occupation of the individual (e.g., "Etudiant (e)" for Student, "Militaire" for Military).
2. **Health Conditions and Eligibility**:
    - **Raison_de_non-eligibilité_totale__[Porteur(HIV,hbs,hcv)]**: Indicates whether the donor carries HIV, Hepatitis B (HBS), or Hepatitis C (HCV).
    - **Raison_de_non-eligibilité_totale__[Opéré]**: Indicates whether the donor has undergone surgery.
    - **Raison_de_non-eligibilité_totale__[Drepanocytaire]**: Indicates whether the donor has sickle cell disease (Drepanocytosis).
    - **Raison_de_non-eligibilité_totale__[Diabétique]**: Indicates whether the donor is diabetic.
    - **Raison_de_non-eligibilité_totale__[Hypertendus]**: Indicates whether the donor has hypertension.
    - **Raison_de_non-eligibilité_totale__[Asthmatiques]**: Indicates whether the donor has asthma.
    - **Raison_de_non-eligibilité_totale__[Cardiaque]**: Indicates whether the donor has heart-related conditions.
    - **Raison_de_non-eligibilité_totale__[Tatoué]**: Indicates whether the donor has tattoos.
    - **Raison_de_non-eligibilité_totale__[Scarifié]**: Indicates whether the donor has scars.
3. **Geographical Information**:
    - **Arrondissement_de_résidence_**: The district where the donor resides (e.g., Douala 1, Douala 3).
    - **Quartier_de_Résidence_**: The specific neighborhood or area where the donor lives (e.g., "Logbaba").
4. **Other Variables**:
    - **Si_autres_raison_préciser_**: This column appears to capture additional feedback or reasons for non-eligibility. It may contain textual data, such as "Aucune information" (No information) or other reasons provided by the donor.

Observations:

- The dataset has many missing values, especially for attributes like height, weight, and health conditions. Preprocessing will be required to handle these missing values effectively.
- There are several health-related flags that help indicate whether a donor is eligible to donate blood based on specific medical conditions.
- **Geographical data** is also provided, which can be valuable for analyzing the distribution of blood donors across different areas.
- The **"Si_autres_raison_préciser_"** column can be used for performing sentiment analysis on textual feedback if it contains useful donor comments.

Usefulness of the Data:

- **Eligibility Analysis**: The health-related flags are critical for analyzing donor eligibility and understanding how conditions like HIV, hypertension, or diabetes affect eligibility.
- **Geographical Mapping**: The residential information can be used to analyze the geographical distribution of donors and target specific regions.
- **Sentiment Analysis**: The textual feedback can provide insight into donor experiences and potential reasons for non-participation, which is useful for improving campaigns.
- **Donor Profiling**: The demographic data (age, education, profession, etc.) can help profile ideal donors for targeted campaigns.

This dataset, combined with the one you shared earlier, offers a rich set of attributes that will help generate meaningful insights for improving blood donation campaigns.