

## **Rapport de Nettoyage des Données pour PIMENTO\_CASES.xlsx et PIMENTO\_PROGRAMS.xlsx**

### **Objectif**

L'objectif de ce rapport est de nettoyer les ensembles de données provenant de PIMENTO\_CASES.xlsx et PIMENTO\_PROGRAMS.xlsx en abordant les problèmes liés aux entrées non valides, aux valeurs manquantes, aux observations anormales, à la création de nouvelles variables et à la réduction de la dimension. De plus, nous allons créer des ensembles de données dérivés agrégés au niveau du consulat/ambassade et de la région, ainsi que des dictionnaires de données correspondants.

## **1. Processus de Nettoyage des Données**

### **1.1 Vérification des Heures de Travail des Employés**

- **Tâche** : S'assurer qu'aucun employé ne travaille trop d'heures en enfilade. Pour cela, il est nécessaire de calculer le total des heures travaillées en enfilade en additionnant les valeurs des colonnes de temps associées à chaque employé.

### **1.2 Suppression des Colonnes Non Souhaitées**

- **Tâche** : Éliminer les colonnes spécifiques (par exemple, les colonnes AJ à BL) pour les lignes 122579 à 122953, car elles contiennent des valeurs excessives.

### **1.3 Gestion des Valeurs Manquantes**

- **Tâche** : Supprimer les lignes avec des valeurs manquantes ou les remplacer par la médiane des colonnes correspondantes.

### **1.4 Lignes Dupliquées**

- **Tâche** : Supprimer les entrées en double de l'ensemble de données pour garantir l'unicité des enregistrements.

### **1.5 Identification des Typos et des Entrées Invalides**

- **Tâche** : Vérifier les erreurs typographiques dans les régions et les missions, et identifier les entrées invalides.

### **1.6 Observations Anormales**

- **Tâche** : Identifier les valeurs aberrantes dans les colonnes numériques, en utilisant des méthodes statistiques pour détecter les observations qui s'éloignent significativement de la moyenne.

## 2. Agrégation des Ensembles de Données Dérivés

- Après avoir nettoyé les ensembles de données, nous avons agrégé les données en fonction de la région, du titre de mission et de l'identifiant de l'employé. Cette agrégation permettra de résumer les données en fonction des dimensions clés et d'obtenir des insights sur les opérations.

## 3. Création de nouvelles variables

### 3.1. Taux de Désastre (Disaster Rate)

- **Description :** Le taux de désastre est défini comme le pourcentage de désastres sur le nombre total de missions. Cette variable fournit une mesure claire de la fréquence des désastres par rapport aux opérations menées. Un taux élevé peut indiquer une concentration de désastres dans certaines régions ou missions, ce qui peut nécessiter une attention particulière en termes de ressources et de planification.

### 3.2. Temps Moyen de Désastre (Avg Disaster Time)

- **Description :** Le temps moyen de désastre indique le temps moyen consacré à chaque désastre. Cela permet d'évaluer l'efficacité des interventions en termes de temps de réponse. Une moyenne élevée pourrait signaler des problèmes dans les processus d'intervention, tandis qu'une moyenne basse peut indiquer une gestion efficace des crises. Cette variable aide également à comprendre l'allocation des ressources et les efforts nécessaires pour faire face aux désastres.

### 3.3. Taux de Décès (Death Rate)

- **Description :** Le taux de décès représente le pourcentage de décès par rapport au nombre total d'interventions. Cette variable est cruciale pour évaluer l'impact des missions sur la vie humaine et comprendre les risques associés aux interventions. Un taux de décès élevé peut signaler la nécessité d'améliorations dans les protocoles d'intervention et d'assistance, en mettant l'accent sur la réduction des pertes humaines.

### 3.4. Fréquence des Communications d'Assistance (Assistance Comm Frequency)

- **Description :** La fréquence des communications d'assistance indique le nombre total de communications d'assistance par mission. Cette variable mesure l'intensité des communications durant les interventions. Une fréquence élevée peut refléter une bonne coordination et une communication efficace entre les équipes sur le terrain, ce qui est essentiel pour le succès des opérations. Cette variable peut également aider à identifier des lacunes dans la communication qui pourraient affecter les résultats des missions.

### **3.5. Taux d'Intervention Légale (Legal Intervention Rate)**

- **Description :** Le taux d'intervention légale est le pourcentage d'interventions légales par rapport aux opérations totales. Cela permet d'évaluer l'importance des interventions juridiques dans le cadre des missions. Un taux élevé peut indiquer que les problèmes juridiques sont fréquents dans certaines missions, ce qui pourrait nécessiter des ajustements dans la formation ou les ressources disponibles pour le personnel.

### **3.6. Taux d'Enlèvement d'Enfants (Child Abduction Rate)**

- **Description :** Le taux d'enlèvement d'enfants représente le pourcentage de cas d'enlèvements d'enfants par rapport au total des interventions. Cette variable est essentielle pour sensibiliser à la problématique des enlèvements d'enfants dans les régions concernées. Un taux élevé peut signaler la nécessité d'une attention accrue et de programmes spécifiques pour protéger les enfants et répondre à ces incidents.

| Variable                  | Type de Données | Description  |
|---------------------------|-----------------|--|
| Disaster Rate             | Numérique       | Pourcentage de désastres par rapport au nombre total de missions. Représente la fréquence des désastres.                                     |
| Avg Disaster Time         | Numérique       | Temps moyen (en heures ou minutes) consacré à chaque désastre. Permet d'évaluer l'efficacité des interventions.                              |
| Death Rate                | Numérique       | Pourcentage de décès par rapport au nombre total d'interventions. Évalue l'impact des missions sur la vie humaine.                           |
| Assistance Comm Frequency | Numérique       | Nombre total de communications d'assistance par mission. Mesure l'intensité des communications durant les interventions.                     |
| Legal Intervention Rate   | Numérique       | Pourcentage d'interventions légales par rapport aux opérations totales. Évalue l'importance des interventions juridiques.                    |
| Child Abduction Rate      | Numérique       | Pourcentage de cas d'enlèvements d'enfants par rapport au total des interventions. Sensibilise à la problématique des enlèvements d'enfants. |

