

# Rapport de Récupération des Données de Population Canine

## 1. Introduction et Problématique

Ce rapport documente le processus de traitement des données manquantes (NaN) pour la population canine en France pour l'année 2020. Le jeu de données initial présentait 10 957 communes sans données, principalement dues au secret statistique ou à des défauts de collecte locale.

## 2. Méthodologie de Récupération Hybride

Pour maximiser la précision, une approche en deux étapes a été adoptée :

### Étape A : Récupération Temporelle

Utilisation des données historiques de la même commune (période 2013-2019). La valeur la plus récente disponible a été utilisée pour combler 2020.

### Étape B : Imputation Spatiale (KNN)

Pour les communes sans historique, nous avons utilisé l'algorithme des K-Plus Proches Voisins (K=5). La valeur manquante est estimée par la moyenne de la population canine de ses 5 voisins géographiques.

## 3. Preuve de Fiabilité Statistique

La validité de l'utilisation des données historiques est confirmée par :

- Une corrélation de 0.978 entre 2019 et 2020.
- Une variation médiane de seulement 7 chiens par commune entre ces deux années.

Ces indicateurs démontrent une très forte stabilité temporelle de la population canine.

## 4. Résultats de l'Opération

Nombre total de communes : 36594

NaNs initiaux : 10957

NaNs après étape temporelle : 6294

NaNs finaux : 78

Taux de récupération global : 99.29%

# Rapport de Récupération des Données de Population Canine

