

# Projet : Prédiction de l'espérance de vie à l'aide de la régression linéaire simple

## Objectif

Votre tâche consiste à prédire l'espérance de vie en utilisant une régression linéaire simple. Vous devez identifier et choisir la meilleure variable indépendante parmi celles fournies pour prédire l'espérance de vie. Vous utiliserez soit la méthode de validation croisée, soit la méthode de division train-test, ainsi que les bibliothèques Scikit-learn, Pandas, Seaborn et Matplotlib. Vous devrez également visualiser les données et vérifier les valeurs aberrantes (outliers) en utilisant des statistiques univariées et vérifier la relation entre les variables en utilisant des statistiques bivariées.

## Variables disponibles

### Étapes du projet

#### 1. Exploration des données :

- Charger les données.
- Comprendre la distribution des variables et leur relation potentielle avec l'espérance de vie.

#### 2. Traitement des valeurs aberrantes :

- Utiliser des statistiques univariées pour identifier les valeurs aberrantes dans les variables.
- Traitement des valeurs aberrantes.

#### 3. Analyse des relations bivariées :

- Utiliser des statistiques bivariées pour examiner la relation entre l'espérance de vie et les autres variables.

#### 4. Choix de la variable indépendante :

- Explorer les corrélations entre l'espérance de vie et les autres variables pour identifier la variable la plus prometteuse.
- Justifier votre choix de la variable indépendante.

#### 5. Préparation des données :

- Diviser les données en ensembles d'entraînement et de test ou préparer pour la validation croisée.

## 6. Construction du modèle :

- Utiliser une régression linéaire simple pour modéliser la relation entre la variable choisie et l'espérance de vie.
- Entraîner le modèle sur l'ensemble d'entraînement ou en utilisant la validation croisée.

## 7. Évaluation du modèle :

- Prédire les valeurs de l'espérance de vie sur l'ensemble de test ou en utilisant la validation croisée.
- Évaluer les performances du modèle en utilisant des métriques telles que l'erreur quadratique moyenne (MSE).

## 8. Visualisation des résultats :

- Tracer les résultats de la prédiction par rapport aux valeurs réelles.
- Visualiser la ligne de régression.

## 9. Interprétation des résultats :

- Analyser les coefficients de régression.
- Discuter des implications des résultats obtenus.

## Consignes supplémentaires

- Documentez votre processus et justifiez vos choix à chaque étape.
- Utilisez des visualisations pour illustrer vos résultats.
- Préparez un rapport final décrivant votre approche, vos résultats et vos conclusions.