

En utilisant la base de données et sa fiche descriptive, vous êtes invités à réaliser les tâches suivantes :

Tâche 1 : Importation des données

Importer la base de données de votre projet.

Tâche 2 : Pré-traitement des données

Préparation des données en étudiant :

1. Valeurs aberrantes :

- Analyser toutes les variables de la base de données et détecter les valeurs aberrantes.
- Imputer les valeurs aberrantes

2. Valeurs manquantes :

- Étudier le taux des valeurs manquantes.
- Proposer une méthode pour l'imputation des valeurs manquantes (justifier votre choix).

Tâche 3 : Analyse univariée :

1. Étudier la normalité de chacune des variables quantitatives de la base de données.
2. Étudier les modalités de chacune des variables qualitatives.

Tâche 4 : Analyse bivariée :

Étudier la corrélation entre les variables (deux à deux) de la base de données et valider les résultats avec les tests d'hypothèse appropriés.

Tâche 5 : Régression linéaire :

1. Régresser la variable cible quantitative (voir la FD de votre projet) de la base de données en fonctions des autres.
2. Proposer une stratégie détaillée pour améliorer la performance du modèle de régression. Justifier le choix des métriques et des test pour comparer les différents modèles.
3. Réduire la dimension de la base de données en utilisant l'analyse en composantes principales (Réf 1) et régresser la variable cible quantitative en fonction des variables de la nouvelle base de données.

Tâche 6 : Régression linéaire généralisée :

1. Effectuer une étude bibliographique sur les modèles de régression linéaire généralisée (Réf 2).
2. Choisir le meilleur modèle linéaire généralisé pour régresser la variable cible quantitative de votre projet.

Livrables :

Vous êtes invités à préparer :

1. **Un scrip R** : le script doit être fonctionnel et exécutable, contenant toutes les tâches demandées.
2. **Une présentation** : Une présentation descriptive du travail demandé. Un maximum de 30 diapositives.

Références :

Réf 1 : Jolliffe Ian T. and Cadima Jorge 2016 Principal component analysis : a review and recent developments Phil. Trans. R. Soc. A.3742015020220150202 <http://doi.org/10.1098/rsta.2015.0202>

Réf 2 : https://en.wikipedia.org/wiki/Generalized_linear_model