

# Processus de Nettoyage des Données

Ce document explique en détail chaque étape du processus de nettoyage des données pour garantir leur cohérence et leur précision. Il couvre les principales fonctions utilisées dans un projet de nettoyage de données en Python pour préparer les données efficacement. Voici une description détaillée de chaque étape :

## 1. Identification du type de vente :

Fonction : `changer_louer_vendre` Objectif : Déterminer si le bien est à 'vendre' ou à 'louer' en analysant les mots-clés dans le titre ou en fonction de la valeur d'une colonne. Methodologie : La fonction convertit le titre en minuscules pour une correspondance insensible à la casse, puis vérifie la présence de mots spécifiques.

## 2. Nettoyage de la colonne prix :

Fonction : `nettoyer_prix` Objectif : Éliminer les espaces et caractères inutiles dans la colonne des prix et convertir ces valeurs en nombres. Methodologie : La fonction remplace les espaces et caractères non pertinents et convertit le prix en valeur numérique.

## 3. Normalisation des noms de villes :

Fonction : `changer_nom_ville` Objectif : Assurer l'uniformité des noms de villes en appliquant un mapping. Methodologie : Utilise un dictionnaire pour remplacer les noms de villes avec leur équivalent dans le mapping.

## 4. Analyse statistique et asymétrie :

Fonction : `analyse_statistique_et_asymetrie` Objectif : Analyser la distribution des données en calculant des statistiques descriptives comme la moyenne, médiane, écart-type et asymétrie. Methodologie : La fonction calcule les valeurs statistiques et affiche un histogramme pour visualiser la distribution.

## 5. Remplacement des valeurs nulles :

Fonctions : `replace`, `remplacer_valeur_null` Objectif : Remplacer les valeurs manquantes en fonction de la distribution des données (moyenne, médiane, mode). Methodologie : La fonction détermine l'asymétrie de la distribution et utilise des statistiques appropriées pour remplacer les valeurs nulles.

## 6. Analyse des valeurs catégoriques :

Fonction : `analyse_categorique` Objectif : Étudier la distribution des catégories dans une colonne et visualiser la répartition. Methodologie : Compte les occurrences de chaque catégorie et génère un graphique à barres pour afficher la fréquence de chaque catégorie.

## 7. Détection des valeurs aberrantes :

Methodologies : - Avec Z-score (`detecter_et_plot_zscore`): Calcule le Z-score pour identifier les valeurs situées à plus de 3 écarts-types. - Par fréquence (`detecter_aberrantes_par_frequence_nombre`): Identifie les valeurs aberrantes selon une fréquence minimale. - Avec l'IQR (`detecter_avec_IQR`): Détermine les valeurs aberrantes en calculant les bornes en fonction de l'IQR.