

```

1  """
2  Script : preparation_donnees.py
3  Auteur : CISSE Ibrahim
4  Objectif : Nettoyer, préparer et standardiser un jeu de données client pour analyse
   ou modélisation.
5
6  Étapes réalisées :
7  1. Chargement des données brutes depuis un fichier CSV
8  2. Nettoyage :
9     - Suppression des doublons
10    - Suppression des valeurs manquantes
11    - Filtrage des lignes incohérentes (revenu <= 0)
12  3. Standardisation :
13     - Mise en minuscules des noms
14     - Uniformisation des adresses (strip, remplacement)
15     - Formatage des dates
16  4. Création de variables dérivées :
17     - Âge à partir de la date de naissance
18     - Catégorie de revenu
19  5. Export du jeu de données nettoyé vers un nouveau fichier CSV
20  """
21
22  import pandas as pd
23  from datetime import datetime
24
25  # 1. Chargement
26  df = pd.read_csv("clients_bruts.csv")
27
28  # 2. Nettoyage
29  df.drop_duplicates(inplace=True)
30  df.dropna(inplace=True)
31  df = df[df["revenu"] > 0]
32
33  # 3. Standardisation
34  df["nom"] = df["nom"].str.lower().str.strip()
35  df["adresse"] = df["adresse"].str.replace(" av ", " avenue ").str.strip()
36  df["date_naissance"] = pd.to_datetime(df["date_naissance"], errors="coerce")
37
38  # 4. Variables dérivées
39  today = pd.Timestamp("today")
40  df["age"] = (today - df["date_naissance"]).dt.days // 365
41  df["categorie_revenu"] = pd.cut(df["revenu"],
42     bins=[0, 20000, 50000, 100000, float("inf")],
43     labels=["faible", "moyen", "élevé", "très élevé"]
44 )
45
46  # 5. Export
47  df.to_csv("clients_nettoyes.csv", index=False)
48

```