import pandas as pd

from fuzzywuzzy import fuzz

from joblib import Parallel, delayed

# Chargez votre DataFrame avec une colonne de noms

# Remplacez "votre\_dataframe.csv" par le nom de votre fichier CSV ou ajustez selon votre source de données

df = pd.read\_csv("votre\_dataframe.csv")

# Définissez une fonction pour calculer le score de similarité fuzzy entre deux noms

def calculate\_fuzzy\_score(nom1, nom2):

return fuzz.ratio(nom1, nom2)

# Utilisez Parallel de joblib pour calculer les scores en parallèle

# N spécifie le nombre de processeurs à utiliser (ajustez selon votre système)

N = 4 # Vous pouvez ajuster cette valeur en fonction de votre matériel

scores = Parallel(n\_jobs=N)(delayed(calculate\_fuzzy\_score)(nom1, nom2) for nom1 in df["Nom"] for nom2 in df["Nom"])

# Remplacez les scores dans le DataFrame

df["Score"] = scores

# Filtrer les résultats pour ne conserver que les scores > 80

df = df[df["Score"] > 80]

# Affichez le DataFrame résultant

print(df)

………

import pandas as pd

from fuzzywuzzy import fuzz

from multiprocessing import Pool

# Chargez votre DataFrame avec une colonne de noms

# Remplacez "votre\_dataframe.csv" par le nom de votre fichier CSV ou ajustez selon votre source de données

df = pd.read\_csv("votre\_dataframe.csv")

# Définissez une fonction pour calculer le score de similarité fuzzy entre deux noms

def calculate\_fuzzy\_score(noms):

nom1, nom2 = noms

return nom1, nom2, fuzz.ratio(nom1, nom2)

# Utilisez la bibliothèque multiprocessing pour paralléliser le calcul des scores

# N spécifie le nombre de processus à utiliser (ajustez selon votre système)

N = 4 # Vous pouvez ajuster cette valeur en fonction de votre matériel

pool = Pool(N)

nom\_pairs = [(nom1, nom2) for nom1 in df["Nom"] for nom2 in df["Nom"]]

scores = pool.map(calculate\_fuzzy\_score, nom\_pairs)

pool.close()

pool.join()

# Créez un DataFrame à partir des résultats

df\_results = pd.DataFrame(scores, columns=["Nom1", "Nom2", "Score"])

# Filtrer les résultats pour ne conserver que les scores > 80

df\_filtered = df\_results[df\_results["Score"] > 80]

# Affichez le DataFrame résultant

print(df\_filtered)