import os

import pandas as pd

import hashlib

def get\_all\_csv\_files(root\_folder):

csv\_files = []

for root, dirs, files in os.walk(root\_folder):

for file in files:

if file.endswith('.csv'):

csv\_files.append(os.path.join(root, file))

return csv\_files

def hash\_file\_content(file\_path):

with open(file\_path, 'rb') as f:

return hashlib.md5(f.read()).hexdigest()

def load\_and\_tag\_csv(file\_path):

try:

df = pd.read\_csv(file\_path)

df['Source\_File'] = os.path.basename(file\_path)

return df

except Exception as e:

print(f"Error reading {file\_path}: {e}")

return None

def merge\_csvs\_from\_folders(folders):

seen\_hashes = set()

merged\_df = pd.DataFrame()

for folder in folders:

csv\_files = get\_all\_csv\_files(folder)

for file in csv\_files:

file\_hash = hash\_file\_content(file)

if file\_hash not in seen\_hashes:

df = load\_and\_tag\_csv(file)

if df is not None:

merged\_df = pd.concat([merged\_df, df], ignore\_index=True)

seen\_hashes.add(file\_hash)

else:

print(f"Skipped duplicate file: {file}")

return merged\_df

# Paths to your two main folders

folder1 = r"C:\Users\vigne\Downloads\2025-20250717T074716Z-1-001\15-07-2025"

folder2 = r"C:\Users\vigne\Downloads\2025-20250715T093935Z-1-001\Web Scrapping - 07-07-2025"

# Merge process

final\_df = merge\_csvs\_from\_folders([folder1, folder2])

# Save to Excel with a new name and custom sheet name

output\_file\_name = "Final\_Merged\_File.xlsx"

output\_path = os.path.join(r"C:\Users\vigne\Downloads", output\_file\_name)

final\_df.to\_excel(output\_path, index=False, sheet\_name="MergedData")

print(f" Merged Excel file created: {output\_path}")