import os

import pandas as pd

def preprocess\_csv\_files(input\_dir, output\_dir):

for filename in os.listdir(input\_dir):

if filename.endswith('.csv'):

size\_filename = len(filename)

filename\_short = filename[:size\_filename - 4].replace(".", "\_")

path\_main = os.path.join(input\_dir, filename)

path\_avg = os.path.join(output\_dir, f"{filename\_short}\_avg.csv")

print(f"Processing {filename}...")

# Read the original dataset

dataset = pd.read\_csv(path\_main)

# Convert timestamp to datetime and extract year, month, day, hour

dataset["datetime"] = pd.to\_datetime(dataset["timestamp"])

dataset["year"] = dataset["datetime"].dt.year

dataset["month"] = dataset["datetime"].dt.month

dataset["day"] = dataset["datetime"].dt.day

dataset["hour"] = dataset["datetime"].dt.hour

# Calculate hourly mean

datasethourlymean = dataset.groupby(["year", "month", "day", "hour"]).value.mean()

# Save the processed dataset

datasethourlymean.to\_csv(path\_avg, index=True, header=True)

datasetavg = pd.read\_csv(path\_avg)

datasetavg.rename(columns={'value': filename\_short}, inplace=True)

datasetavg.to\_csv(path\_avg, index=False)

if \_\_name\_\_ == "\_\_main\_\_":

input\_dir = '../DataPreprocessing/data/raw\_data'

output\_dir = '../DataPreprocessing/data/processed\_data'

if not os.path.exists(output\_dir):

os.makedirs(output\_dir)

preprocess\_csv\_files(input\_dir, output\_dir)

print("Preprocessing completed.")