**—> Convert Halfhourly dataset -**

Blockwise, compiled energy used by a single house as per block

halfhourly\_block\_path = "/content/drive/MyDrive/archive/halfhourly\_dataset/halfhourly\_dataset/"

block\_dataset\_csv = []

temp = pd.DataFrame(columns=['tstp', 'energy'])

for i in os.listdir(halfhourly\_block\_path):

f = os.path.join(halfhourly\_block\_path,i)

block\_file = pd.read\_csv(f)

l1 = block\_file['tstp'].unique()

energy\_dict = {}

for k in l1:

energy\_dict[k] = 0

# Replace 'Null' with '0' and convert to float

block\_file['energy(kWh/hh)'] = block\_file['energy(kWh/hh)'].replace('Null', '0').astype(float)

# Now iterate over the DataFrame and update the energy\_dict

for index, row in block\_file.iterrows():

energy\_dict[row['tstp']] += row['energy(kWh/hh)']

file\_name = i[:len(i)-4]

# name\_dict = {f'{file\_name}': energy\_dict}

with open(f'/content/drive/MyDrive/Colab Notebooks/block\_json/{file\_name}.json', 'w') as f:

json\_val = json.dumps(energy\_dict)

f.write(json\_val)

**—-> Convert halfhourly to hourly dataset -**

import csv

file = open(f"/content/drive/MyDrive/Colab Notebooks/block\_json/block\_0.json")

data = json.load(file)

updated\_data = {}

temp\_key = 0

temp\_val = 0

for key, value in data.items():

# print(key.split()[0])

if key.split()[1][3] == '3':

temp\_key = key

temp\_val = value

# print(f"Temp\_key = {temp\_key}, Temp\_val = {temp\_val}")

else:

updated\_data.setdefault(key.split()[0], []).append([key.split()[1], value + temp\_val])

# print(f"Key: {key}, Value: {value}, Temp\_key = {temp\_key}, Temp\_val = {temp\_val}")

with open('/content/drive/MyDrive/Colab Notebooks/block\_0\_final.csv', 'w', newline='') as csvfile:

# Create a CSV writer object

writer = csv.writer(csvfile)

# Write the header row (assuming all sub-lists have the same length)

header = ["Date"] + ["Time"]+["Energy"]

writer.writerow(header)

# Write each key and its corresponding sub-lists as rows

for key, value\_list in updated\_data.items():

for sub\_list in value\_list:

writer.writerow([key] + sub\_list)

print("CSV file created successfully!")

**—->Compile energy used by all blocks along with hourly weather conditions -**

file = pd.read\_csv('/content/drive/MyDrive/Colab Notebooks/block\_0\_final.csv')

for i in range(1,112):

file\_json = open(f"/content/drive/MyDrive/Colab Notebooks/block\_json/block\_{i}.json")

data = json.load(file\_json)

updated\_data = {}

temp\_key = 0

temp\_val = 0

for key, value in data.items():

if key.split()[1][3] == '3':

temp\_key = key

temp\_val = value

else:

updated\_data.setdefault(key.split()[0], []).append([key.split()[1], value + temp\_val])

for key, value\_list in updated\_data.items():

for sublist in value\_list:

try:

value\_to\_update = float(file.loc[(file.Date == key) & (file.Time == sublist[0])]['Energy'].iloc[0]) + sublist[1]

file.loc[(file.Date == key) & (file.Time == sublist[0]), 'Energy'] = value\_to\_update

except:

continue

print(f"File block\_{i} done.")

file.to\_csv('/content/drive/MyDrive/Colab Notebooks/full\_and\_final.csv', index = False)

final\_csv = pd.read\_csv('/content/drive/MyDrive/Colab Notebooks/full\_and\_final.csv')

final\_csv

**—-> Getting final dataset ready and organized -**

final\_csv['Date'] = pd.to\_datetime(final\_csv['Date'])

final\_csv['Time'] = pd.to\_datetime(final\_csv['Time'])

final\_csv

import pandas as pd

weather\_data = pd.read\_csv('/content/drive/MyDrive/archive/weather\_hourly\_darksky.csv')

x = weather\_data['time'].str.split(expand=True)

weather\_data = weather\_data.drop(['time'], axis=1)

weather\_data.insert(3, 'Date', x[0], False)

weather\_data.insert(4, 'Time', x[1], True)

weather\_data['Date'] = pd.to\_datetime(weather\_data['Date'])

weather\_data['Time'] = pd.to\_datetime(weather\_data['Time']) # Assuming Time format is HH:MM:SS

# weather\_data

merged\_data = pd.merge(left=weather\_data, right=final\_csv[['Date', 'Time', 'Energy']],how='left', on=['Date', 'Time'])

selected\_columns = list(weather\_data.columns) + ['Energy']

merged\_data = merged\_data[selected\_columns]

merged\_data.to\_csv('/content/drive/MyDrive/Colab Notebooks/Final\_dataset.csv', index = False)