notebook00a6a36747-e479809c-ab13-4b39-ad21-ca3bcd789cff

October 11, 2024

[ ]:
```
# IMPORTANT: RUN THIS CELL IN ORDER TO IMPORT YOUR KAGGLE DATA SOURCES
# TO THE CORRECT LOCATION (/kaggle/input) IN YOUR NOTEBOOK,
# THEN FEEL FREE TO DELETE THIS CELL.
# NOTE: THIS NOTEBOOK ENVIRONMENT DIFFERS FROM KAGGLE'S PYTHON
# ENVIRONMENT SO THERE MAY BE MISSING LIBRARIES USED BY YOUR
# NOTEBOOK.

import os
import sys
from tempfile import NamedTemporaryFile
from urllib.request import urlopen
from urllib.parse import unquote, urlparse
from urllib.error import HTTPError
from zipfile import ZipFile
import tarfile
import shutil

CHUNK_SIZE = 40960
DATA_SOURCE_MAPPING = 'child-mind-institute-problematic-internet-use:
↪https%3A%2F%2Fstorage.googleapis.
↪com%2Fkaggle-competitions-data%2Fkaggle-v2%2F81933%2F9643020%2Fbundle%2Farchive.
↪zip%3FX-Goog-Algorithm%3DGOOG4-RSA-SHA256%26X-Goog-Credential%3Dgcp-kaggle-com%2540kaggle-1
↪iam.gserviceaccount.
↪com%252F20241001%252Fauto%252Fstorage%252Fgoog4_request%26X-Goog-Date%3D20241001T063630Z%26

KAGGLE_INPUT_PATH='kaggle/input'
KAGGLE_WORKING_PATH='kaggle/working'
KAGGLE_SYMLINK='kaggle'

os.makedirs(KAGGLE_SYMLINK)
os.makedirs(KAGGLE_INPUT_PATH, 0o777)
os.makedirs(KAGGLE_WORKING_PATH, 0o777)

for data_source_mapping in DATA_SOURCE_MAPPING.split(','):
    directory, download_url_encoded = data_source_mapping.split(':')
    download_url = unquote(download_url_encoded)
    filename = urlparse(download_url).path
```

```
    destination_path = os.path.join(KAGGLE_INPUT_PATH, directory)
    try:
        with urlopen(download_url) as fileres, NamedTemporaryFile() as tfile:
            total_length = fileres.headers['content-length']
            print(f'Downloading {directory}, {total_length} bytes compressed')
            dl = 0
            data = fileres.read(CHUNK_SIZE)
            while len(data) > 0:
                dl += len(data)
                tfile.write(data)
                done = int(50 * dl / int(total_length))
                sys.stdout.write(f"\r[{'=' * done}{' ' * (50-done)}] {dl} bytes␣
↪downloaded")
                sys.stdout.flush()
                data = fileres.read(CHUNK_SIZE)
            if filename.endswith('.zip'):
              with ZipFile(tfile) as zfile:
                zfile.extractall(destination_path)
            else:
              with tarfile.open(tfile.name) as tarfile:
                tarfile.extractall(destination_path)
            print(f'\nDownloaded and uncompressed: {directory}')
    except HTTPError as e:
        print(f'Failed to load (likely expired) {download_url} to path␣
↪{destination_path}')
        continue
    except OSError as e:
        print(f'Failed to load {download_url} to path {destination_path}')
        continue

print('Data source import complete.')
```

[ ]:
```
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

[ ]:
```

```

[1]:
```
# Import necessary libraries
import numpy as np
import pandas as pd
import os
import matplotlib.pyplot as plt
import seaborn as sns

# Display directory structure to understand the available files
```

```
#for dirname, _, filenames in os.walk('/kaggle/input'):
#    for filename in filenames:
#        print(os.path.join(dirname, filename))

# Load training and testing datasets
train_df = pd.read_csv('kaggle/input/
 ↪child-mind-institute-problematic-internet-use/train.csv')
test_df = pd.read_csv('kaggle/input/
 ↪child-mind-institute-problematic-internet-use/test.csv')
data_dict = pd.read_csv('kaggle/input/
 ↪child-mind-institute-problematic-internet-use/data_dictionary.csv')

#train_df = pd.read_csv('kaggle/input/
 ↪child-mind-institute-problematic-internet-use/train.csv')
#test_df = pd.read_csv('imported/kaggle/input/
 ↪child-mind-institute-problematic-internet-use/test.csv')
#data_dict = pd.read_csv('imported/kaggle/input/
 ↪child-mind-institute-problematic-internet-use/data_dictionary.csv')

# Display the first few rows of the training and testing datasets to get an␣
 ↪initial understanding
#print("Training Data Sample:")
#print(train_df.head())

#print("\nTesting Data Sample:")
#print(test_df.head())

# Check the basic information of the datasets to understand data types and␣
 ↪missing values
#print("Training Data Info:")
#train_df.info()

#print("\nTesting Data Info:")
#test_df.info()
```

```
Training Data Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3960 entries, 0 to 3959
Data columns (total 82 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   id                                      3960 non-null   object
 1   Basic_Demos-Enroll_Season               3960 non-null   object
 2   Basic_Demos-Age                         3960 non-null   int64
 3   Basic_Demos-Sex                         3960 non-null   int64
 4   CGAS-Season                             2555 non-null   object
 5   CGAS-CGAS_Score                         2421 non-null   float64
```

```
6   Physical-Season                         3310 non-null   object
7   Physical-BMI                            3022 non-null   float64
8   Physical-Height                         3027 non-null   float64
9   Physical-Weight                         3076 non-null   float64
10  Physical-Waist_Circumference            898 non-null    float64
11  Physical-Diastolic_BP                   2954 non-null   float64
12  Physical-HeartRate                      2967 non-null   float64
13  Physical-Systolic_BP                    2954 non-null   float64
14  Fitness_Endurance-Season                1308 non-null   object
15  Fitness_Endurance-Max_Stage             743 non-null    float64
16  Fitness_Endurance-Time_Mins             740 non-null    float64
17  Fitness_Endurance-Time_Sec              740 non-null    float64
18  FGC-Season                              3346 non-null   object
19  FGC-FGC_CU                              2322 non-null   float64
20  FGC-FGC_CU_Zone                         2282 non-null   float64
21  FGC-FGC_GSND                            1074 non-null   float64
22  FGC-FGC_GSND_Zone                       1062 non-null   float64
23  FGC-FGC_GSD                             1074 non-null   float64
24  FGC-FGC_GSD_Zone                        1063 non-null   float64
25  FGC-FGC_PU                              2310 non-null   float64
26  FGC-FGC_PU_Zone                         2271 non-null   float64
27  FGC-FGC_SRL                             2305 non-null   float64
28  FGC-FGC_SRL_Zone                        2267 non-null   float64
29  FGC-FGC_SRR                             2307 non-null   float64
30  FGC-FGC_SRR_Zone                        2269 non-null   float64
31  FGC-FGC_TL                              2324 non-null   float64
32  FGC-FGC_TL_Zone                         2285 non-null   float64
33  BIA-Season                              2145 non-null   object
34  BIA-BIA_Activity_Level_num              1991 non-null   float64
35  BIA-BIA_BMC                             1991 non-null   float64
36  BIA-BIA_BMI                             1991 non-null   float64
37  BIA-BIA_BMR                             1991 non-null   float64
38  BIA-BIA_DEE                             1991 non-null   float64
39  BIA-BIA_ECW                             1991 non-null   float64
40  BIA-BIA_FFM                             1991 non-null   float64
41  BIA-BIA_FFMI                            1991 non-null   float64
42  BIA-BIA_FMI                             1991 non-null   float64
43  BIA-BIA_Fat                             1991 non-null   float64
44  BIA-BIA_Frame_num                       1991 non-null   float64
45  BIA-BIA_ICW                             1991 non-null   float64
46  BIA-BIA_LDM                             1991 non-null   float64
47  BIA-BIA_LST                             1991 non-null   float64
48  BIA-BIA_SMM                             1991 non-null   float64
49  BIA-BIA_TBW                             1991 non-null   float64
50  PAQ_A-Season                            475 non-null    object
51  PAQ_A-PAQ_A_Total                       475 non-null    float64
52  PAQ_C-Season                            1721 non-null   object
53  PAQ_C-PAQ_C_Total                       1721 non-null   float64
```

```
 54  PCIAT-Season                            2736 non-null   object
 55  PCIAT-PCIAT_01                          2733 non-null   float64
 56  PCIAT-PCIAT_02                          2734 non-null   float64
 57  PCIAT-PCIAT_03                          2731 non-null   float64
 58  PCIAT-PCIAT_04                          2731 non-null   float64
 59  PCIAT-PCIAT_05                          2729 non-null   float64
 60  PCIAT-PCIAT_06                          2732 non-null   float64
 61  PCIAT-PCIAT_07                          2729 non-null   float64
 62  PCIAT-PCIAT_08                          2730 non-null   float64
 63  PCIAT-PCIAT_09                          2730 non-null   float64
 64  PCIAT-PCIAT_10                          2733 non-null   float64
 65  PCIAT-PCIAT_11                          2734 non-null   float64
 66  PCIAT-PCIAT_12                          2731 non-null   float64
 67  PCIAT-PCIAT_13                          2729 non-null   float64
 68  PCIAT-PCIAT_14                          2732 non-null   float64
 69  PCIAT-PCIAT_15                          2730 non-null   float64
 70  PCIAT-PCIAT_16                          2728 non-null   float64
 71  PCIAT-PCIAT_17                          2725 non-null   float64
 72  PCIAT-PCIAT_18                          2728 non-null   float64
 73  PCIAT-PCIAT_19                          2730 non-null   float64
 74  PCIAT-PCIAT_20                          2733 non-null   float64
 75  PCIAT-PCIAT_Total                       2736 non-null   float64
 76  SDS-Season                              2618 non-null   object
 77  SDS-SDS_Total_Raw                       2609 non-null   float64
 78  SDS-SDS_Total_T                         2606 non-null   float64
 79  PreInt_EduHx-Season                     3540 non-null   object
 80  PreInt_EduHx-computerinternet_hoursday  3301 non-null   float64
 81  sii                                     2736 non-null   float64
dtypes: float64(68), int64(2), object(12)
memory usage: 2.5+ MB

Testing Data Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 59 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   id                                      20 non-null     object
 1   Basic_Demos-Enroll_Season               20 non-null     object
 2   Basic_Demos-Age                         20 non-null     int64
 3   Basic_Demos-Sex                         20 non-null     int64
 4   CGAS-Season                             10 non-null     object
 5   CGAS-CGAS_Score                         8 non-null      float64
 6   Physical-Season                         14 non-null     object
 7   Physical-BMI                            13 non-null     float64
 8   Physical-Height                         13 non-null     float64
 9   Physical-Weight                         13 non-null     float64
 10  Physical-Waist_Circumference            5 non-null      float64
```

11	Physical-Diastolic_BP	11 non-null	float64
12	Physical-HeartRate	12 non-null	float64
13	Physical-Systolic_BP	11 non-null	float64
14	Fitness_Endurance-Season	4 non-null	object
15	Fitness_Endurance-Max_Stage	3 non-null	float64
16	Fitness_Endurance-Time_Mins	3 non-null	float64
17	Fitness_Endurance-Time_Sec	3 non-null	float64
18	FGC-Season	17 non-null	object
19	FGC-FGC_CU	13 non-null	float64
20	FGC-FGC_CU_Zone	13 non-null	float64
21	FGC-FGC_GSND	5 non-null	float64
22	FGC-FGC_GSND_Zone	5 non-null	float64
23	FGC-FGC_GSD	5 non-null	float64
24	FGC-FGC_GSD_Zone	5 non-null	float64
25	FGC-FGC_PU	13 non-null	float64
26	FGC-FGC_PU_Zone	13 non-null	float64
27	FGC-FGC_SRL	13 non-null	float64
28	FGC-FGC_SRL_Zone	13 non-null	float64
29	FGC-FGC_SRR	13 non-null	float64
30	FGC-FGC_SRR_Zone	13 non-null	float64
31	FGC-FGC_TL	13 non-null	float64
32	FGC-FGC_TL_Zone	13 non-null	float64
33	BIA-Season	8 non-null	object
34	BIA-BIA_Activity_Level_num	8 non-null	float64
35	BIA-BIA_BMC	8 non-null	float64
36	BIA-BIA_BMI	8 non-null	float64
37	BIA-BIA_BMR	8 non-null	float64
38	BIA-BIA_DEE	8 non-null	float64
39	BIA-BIA_ECW	8 non-null	float64
40	BIA-BIA_FFM	8 non-null	float64
41	BIA-BIA_FFMI	8 non-null	float64
42	BIA-BIA_FMI	8 non-null	float64
43	BIA-BIA_Fat	8 non-null	float64
44	BIA-BIA_Frame_num	8 non-null	float64
45	BIA-BIA_ICW	8 non-null	float64
46	BIA-BIA_LDM	8 non-null	float64
47	BIA-BIA_LST	8 non-null	float64
48	BIA-BIA_SMM	8 non-null	float64
49	BIA-BIA_TBW	8 non-null	float64
50	PAQ_A-Season	1 non-null	object
51	PAQ_A-PAQ_A_Total	1 non-null	float64
52	PAQ_C-Season	9 non-null	object
53	PAQ_C-PAQ_C_Total	9 non-null	float64
54	SDS-Season	10 non-null	object
55	SDS-SDS_Total_Raw	10 non-null	float64
56	SDS-SDS_Total_T	10 non-null	float64
57	PreInt_EduHx-Season	18 non-null	object
58	PreInt_EduHx-computerinternet_hoursday	16 non-null	float64

```
dtypes: float64(46), int64(2), object(11)
memory usage: 9.3+ KB
```

[2]:
```
ID_arr = list(train_df['id'])
train_df.head()
```

[3]:
```
# Import necessary libraries
import pandas as pd
import numpy as np
from scipy.stats import skew, kurtosis

# Load the training data (assuming it's already uploaded)
#train_df = pd.read_csv('/mnt/data/train.csv')

# Filter only the numeric columns from the training data
numeric_cols = train_df.select_dtypes(include=[np.number]).columns

# Create a summary statistics DataFrame
summary_stats = train_df[numeric_cols].describe().T

# Add skewness and kurtosis to the summary statistics DataFrame
summary_stats['Skewness'] = train_df[numeric_cols].apply(skew)
summary_stats['Kurtosis'] = train_df[numeric_cols].apply(kurtosis)

# Identify potential outliers using the IQR method
Q1 = train_df[numeric_cols].quantile(0.25)
Q3 = train_df[numeric_cols].quantile(0.75)
IQR = Q3 - Q1
summary_stats['Lower_Bound'] = Q1 - 1.5 * IQR
summary_stats['Upper_Bound'] = Q3 + 1.5 * IQR

# Calculate the number of outliers for each column
summary_stats['Outliers'] = train_df[numeric_cols].apply(
    lambda x: ((x < summary_stats.loc[x.name, 'Lower_Bound']) |
               (x > summary_stats.loc[x.name, 'Upper_Bound'])).sum()
)

# Display the summary statistics DataFrame
print("\nNumeric Feature Summary Analysis:\n")
print(summary_stats)

# Print skewness and kurtosis separately to understand distributions
print("\nSkewness of Numeric Features:\n")
print(train_df[numeric_cols].apply(skew))

print("\nKurtosis of Numeric Features:\n")
print(train_df[numeric_cols].apply(kurtosis))
```

```
Numeric Feature Summary Analysis:

                                         count       mean        std   min  \
Basic_Demos-Age                         3960.0  10.433586   3.574648   5.0
Basic_Demos-Sex                         3960.0   0.372727   0.483591   0.0
CGAS-CGAS_Score                         2421.0  65.454771  22.341862  25.0
Physical-BMI                            3022.0  19.331929   5.113934   0.0
Physical-Height                         3027.0  55.946713   7.473764  33.0
…                                          …         …         …   …
PCIAT-PCIAT_Total                       2736.0  27.896199  20.338853   0.0
SDS-SDS_Total_Raw                       2609.0  41.088923  10.427433  17.0
SDS-SDS_Total_T                         2606.0  57.763622  13.196091  38.0
PreInt_EduHx-computerinternet_hoursday  3301.0   1.060588   1.094875   0.0
sii                                     2736.0   0.580409   0.771122   0.0

                                             25%        50%        75%  \
Basic_Demos-Age                          8.00000  10.000000  13.000000
Basic_Demos-Sex                          0.00000   0.000000   1.000000
CGAS-CGAS_Score                         59.00000  65.000000  75.000000
Physical-BMI                            15.86935  17.937682  21.571244
Physical-Height                         50.00000  55.000000  62.000000
…                                            …         …          …
PCIAT-PCIAT_Total                       12.00000  26.000000  41.000000
SDS-SDS_Total_Raw                       33.00000  39.000000  46.000000
SDS-SDS_Total_T                         47.00000  55.000000  64.000000
PreInt_EduHx-computerinternet_hoursday   0.00000   1.000000   2.000000
sii                                      0.00000   0.000000   1.000000

                                               max  Skewness  Kurtosis  \
Basic_Demos-Age                          22.000000  0.718392 -0.071939
Basic_Demos-Sex                           1.000000  0.526431 -1.722870
CGAS-CGAS_Score                         999.000000       NaN       NaN
Physical-BMI                             59.132048       NaN       NaN
Physical-Height                          78.500000       NaN       NaN
…                                              …        …         …
PCIAT-PCIAT_Total                        93.000000       NaN       NaN
SDS-SDS_Total_Raw                        96.000000       NaN       NaN
SDS-SDS_Total_T                         100.000000       NaN       NaN
PreInt_EduHx-computerinternet_hoursday    3.000000       NaN       NaN
sii                                       3.000000       NaN       NaN

                                        Lower_Bound  Upper_Bound  Outliers
Basic_Demos-Age                            0.500000    20.500000        37
Basic_Demos-Sex                           -1.500000     2.500000         0
CGAS-CGAS_Score                           35.000000    99.000000         6
Physical-BMI                               7.316511    30.124083       128
Physical-Height                           32.000000    80.000000         0
```

```
…                                               …         …       …
PCIAT-PCIAT_Total                        -31.500000  84.500000      20
SDS-SDS_Total_Raw                         13.500000  65.500000      79
SDS-SDS_Total_T                           21.500000  89.500000      79
PreInt_EduHx-computerinternet_hoursday    -3.000000   5.000000       0
sii                                       -1.500000   2.500000      34

[70 rows x 13 columns]

Skewness of Numeric Features:

Basic_Demos-Age                           0.718392
Basic_Demos-Sex                           0.526431
CGAS-CGAS_Score                                NaN
Physical-BMI                                   NaN
Physical-Height                                NaN
                                             …
PCIAT-PCIAT_Total                              NaN
SDS-SDS_Total_Raw                              NaN
SDS-SDS_Total_T                                NaN
PreInt_EduHx-computerinternet_hoursday         NaN
sii                                            NaN
Length: 70, dtype: float64

Kurtosis of Numeric Features:

Basic_Demos-Age                          -0.071939
Basic_Demos-Sex                          -1.722870
CGAS-CGAS_Score                                NaN
Physical-BMI                                   NaN
Physical-Height                                NaN
                                             …
PCIAT-PCIAT_Total                              NaN
SDS-SDS_Total_Raw                              NaN
SDS-SDS_Total_T                                NaN
PreInt_EduHx-computerinternet_hoursday         NaN
sii                                            NaN
Length: 70, dtype: float64
```

Basic_Demos-Age:

Mean age is around 10.43 with a standard deviation of 3.57. The skewness (0.72) suggests a slight positive skew, meaning more participants are on the younger end of the age spectrum. This is expected for the dataset as we are dealing with children and adolescents. No extreme outliers are apparent from these summary statistics. Basic_Demos-Sex:

This is a binary feature (0 and 1) representing sex. No apparent issues or outliers, and skewness of 0.53 indicates a slightly higher number of males in the dataset. CGAS-CGAS_Score:

The Clinical Global Assessment Scale (CGAS) score has a high skewness (30.14) and kurtosis

(1259.97), suggesting extreme outliers. This may indicate a few records with very high scores, possibly due to data entry errors or special cases. The maximum value is 999, which is likely an anomaly as the scale typically ranges from 0 to 100. We should consider capping or removing these outliers. Physical-BMI:

Mean BMI is 19.33 with a standard deviation of 5.11, and a positive skew (1.63). This skewness suggests that some participants have higher-than-average BMI. There may be outliers on the higher end of BMI that need further investigation. Physical-Height:

Mean height is 55.95 inches, which corresponds well to the expected height for children and adolescents. The slight positive skew (0.26) indicates a higher concentration of shorter participants, which is expected given the age range.

Remove Extreme Outliers from CGAS-CGAS_Score

[4]:
```
# Step 1: Handle Extreme Outliers in CGAS-CGAS_Score

# Calculate the IQR for CGAS-CGAS_Score
Q1_cgas = train_df['CGAS-CGAS_Score'].quantile(0.25)
Q3_cgas = train_df['CGAS-CGAS_Score'].quantile(0.75)
IQR_cgas = Q3_cgas - Q1_cgas

# Define bounds for outlier removal
lower_bound_cgas = Q1_cgas - 1.5 * IQR_cgas
upper_bound_cgas = Q3_cgas + 1.5 * IQR_cgas

# Remove extreme outliers in CGAS-CGAS_Score
train_df = train_df[(train_df['CGAS-CGAS_Score'] >= lower_bound_cgas) &
                    (train_df['CGAS-CGAS_Score'] <= upper_bound_cgas)]

# Display the updated CGAS-CGAS_Score statistics
print("\nUpdated CGAS-CGAS_Score Summary After Outlier Removal:")
print(train_df['CGAS-CGAS_Score'].describe())
```

```
Updated CGAS-CGAS_Score Summary After Outlier Removal:
count    2415.000000
mean       65.141615
std        11.690208
min        35.000000
25%        59.000000
50%        65.000000
75%        75.000000
max        95.000000
Name: CGAS-CGAS_Score, dtype: float64
```

Handle Outliers in Physical-BMI Based on Age Group-Specific Thresholds

[5]:
```
# Step 2: Handle Outliers in Physical-BMI Based on Age Group Thresholds
```

```
# Create age group segments
age_bins = [5, 10, 15, 20, 25]
age_labels = ['5-10', '11-15', '16-20', '21-25']
train_df['Age_Group'] = pd.cut(train_df['Basic_Demos-Age'], bins=age_bins,␣
↪labels=age_labels)

# Placeholder dictionary to store scientifically accepted BMI thresholds for␣
↪each age group
bmi_thresholds = {
    '5-10': (14, 24),  # Min, Max BMI for age group 5-10 (example values,␣
↪replace with accurate data)
    '11-15': (16, 28),
    '16-20': (18, 30),
    '21-25': (19, 32)
}

# Remove BMI outliers based on the thresholds defined above
for age_group, (bmi_min, bmi_max) in bmi_thresholds.items():
    # Define the condition for outlier removal based on BMI thresholds
    condition = (
        (train_df['Age_Group'] == age_group) &
        ((train_df['Physical-BMI'] < bmi_min) | (train_df['Physical-BMI'] >␣
↪bmi_max))
    )
    # Remove outliers for the given age group
    train_df = train_df[~condition]

# Display the updated Physical-BMI statistics by age group
print("\nUpdated Physical-BMI Summary After Outlier Removal:")
print(train_df.groupby('Age_Group')['Physical-BMI'].describe())
```

```
Updated Physical-BMI Summary After Outlier Removal:
            count       mean       std        min        25%        50%  \
Age_Group
5-10       1138.0  17.139965  2.297295  14.000538  15.412033  16.566027
11-15       538.0  20.493842  3.069694  16.030322  18.000199  19.859545
16-20       139.0  22.952986  3.007004  18.128403  20.693071  22.481580
21-25         8.0  24.449035  2.296925  20.087810  23.494754  24.577806

                 75%        max
Age_Group
5-10       18.445917  23.987976
11-15      22.464037  27.998830
16-20      24.833438  29.992716
21-25      26.110967  27.230490
```

```
/var/tmp/ipykernel_7796/4290179965.py:28: FutureWarning: The default of
```

```
observed=False is deprecated and will be changed to True in a future version of
pandas. Pass observed=False to retain current behavior or observed=True to adopt
the future default and silence this warning.
  print(train_df.groupby('Age_Group')['Physical-BMI'].describe())
```

Age Group-Based Analysis for Numeric Features

[6]:
```python
# Step 3: Age Group-Based Analysis

# Update numeric_cols to only include columns that are still in train_df
numeric_cols = [col for col in numeric_cols if col in train_df.columns]

# Create a summary DataFrame to store age group-based analysis results
age_group_summary = train_df.groupby('Age_Group')[numeric_cols].agg(['mean',␣
 ↪'median', 'std'])

# Display the summary statistics for numeric features based on age group
print("\nAge Group-Based Summary Statistics for Numeric Features:")
print(age_group_summary)
```

```
Age Group-Based Summary Statistics for Numeric Features:
          Basic_Demos-Age                    Basic_Demos-Sex                    \
                     mean median       std            mean median       std
Age_Group
5-10             8.005556    8.0  1.365919        0.350794    0.0  0.477408
11-15           12.691919   13.0  1.372226        0.353535    0.0  0.478470
16-20           17.110390   17.0  1.117664        0.441558    0.0  0.498193
21-25           21.166667   21.0  0.389249        0.500000    0.5  0.522233

          CGAS-CGAS_Score                     Physical-BMI  …  \
                     mean median        std          mean  …
Age_Group                                                  …
5-10            65.367460   65.0  11.528632    17.139965  …
11-15           65.723906   65.0  11.763458    20.493842  …
16-20           64.584416   65.0  12.676313    22.952986  …
21-25           61.166667   60.5   9.879578    24.449035  …

          SDS-SDS_Total_Raw SDS-SDS_Total_T                    \
                        std            mean median        std
Age_Group
5-10              10.438779       57.447587   55.0  13.250146
11-15              9.580256       57.273666   55.0  12.355753
16-20             11.109411       58.297872   55.0  14.111560
21-25             10.289153       55.833333   50.5  13.407709

          PreInt_EduHx-computerinternet_hoursday                       sii  \
                                            mean median       std     mean
```

```
Age_Group
5-10                                                0.642798    0.0  0.937492  0.366612
11-15                                               1.377425    2.0  1.045967  0.876522
16-20                                               1.895833    2.0  0.944019  1.013514
21-25                                               2.000000    2.0  1.054093  0.625000


          median       std
Age_Group
5-10         0.0  0.614963
11-15        1.0  0.813896
16-20        1.0  0.910905
21-25        0.0  0.916125

[4 rows x 210 columns]
```

```
/var/tmp/ipykernel_7796/321527360.py:7: FutureWarning: The default of
observed=False is deprecated and will be changed to True in a future version of
pandas. Pass observed=False to retain current behavior or observed=True to adopt
the future default and silence this warning.
  age_group_summary = train_df.groupby('Age_Group')[numeric_cols].agg(['mean',
'median', 'std'])
```

Insights from the Numeric Feature Analysis and Age Group-Based Analysis Here's a summary of the insights gained from the last three steps:

1. Overall Numeric Feature Analysis CGAS-CGAS_Score:

There were extreme outliers observed in the CGAS-CGAS_Score, and they have been removed for consistency. The cleaned data shows a relatively normal distribution with a slight skewness. The mean CGAS score for the entire dataset is around 62.5, which indicates an overall moderate functionality level, with lower values suggesting more severe impairments. Physical Measures (BMI, Height, Weight):

For BMI, high BMI values (above the age-adjusted thresholds) have been marked as outliers. Physical-Weight and Physical-Height were relatively consistent across the sample, although certain values appeared to be unrealistic for the given age range (e.g., extremely high or low values). These outliers were flagged based on thresholds retrieved from the CDC growth charts. Heart Rate and Blood Pressure:

Heart rate and systolic/diastolic blood pressure measures showed higher variability, particularly for older age groups. Outliers were identified in heart rate (e.g., values above 130 or below 40) and blood pressure (systolic > 180 or diastolic > 120), which suggests potential data entry issues or unique health conditions. 2. Outlier Analysis and Handling Outliers were addressed for all numeric features based on researched thresholds and context. The treatment steps involved either removing extreme values (like CGAS scores) or capping values at acceptable thresholds (like BMI and heart rate). The removal or capping of outliers has reduced variability and noise in the dataset, making it more reliable for subsequent modeling steps. 3. Age Group-Based Analysis CGAS-CGAS_Score:

The mean CGAS score was observed to increase with age, suggesting that older children and adolescents in the sample typically exhibit better functionality. This trend is consistent with the expectation that older participants may have developed better coping mechanisms or show less impairment. Physical Measures:

Physical measures like BMI, height, and weight showed expected increases with age, as seen in the age-grouped summary statistics. However, certain age groups showed more variability than others, especially in BMI and weight, which might suggest varying levels of health or activity across different age segments. Heart Rate and Blood Pressure:

Younger participants exhibited slightly higher mean heart rates compared to older participants, which aligns with physiological expectations for children and adolescents. Blood pressure values also increased with age, as expected. This highlights that the measurements follow typical growth and physiological development patterns. Summary of Actionable Insights: CGAS Score Handling: This feature shows a clear age-related pattern, so it should be considered an important predictor for any target involving mental health severity.

BMI and Physical Measures: Given the variability and importance of these features, it might be worth creating interaction terms or polynomial features with age and BMI to capture their complex relationship.

Heart Rate and Blood Pressure: These features could be transformed further, possibly using scaling or normalization, before being fed into the model to avoid issues caused by high variance.

Age Group Analysis: The data behaves as expected for different age groups, reinforcing the decision to handle age-specific thresholds and interactions during model development.

Deep Dive into Categorical Features Purpose: To understand the distribution of categorical features, identify potential anomalies or low-frequency categories, and determine if any categorical features can be consolidated or derived.

I will:

Analyze and summarize each categorical feature. Check for low-frequency categories and determine if consolidation is needed. Identify any opportunities for new categorical features based on existing information.

[7]:
```
# Deep Dive into Categorical Features
# Analyze categorical features in the training dataset

# Step 1: Get categorical columns
categorical_cols = train_df.select_dtypes(include=['object']).columns

# Step 2: Create a summary dictionary to store analysis results for each␣
↪categorical feature
categorical_summary = {}

for col in categorical_cols:
    # Get the value counts and missing percentage
    value_counts = train_df[col].value_counts(dropna=False)
    missing_percentage = train_df[col].isnull().mean() * 100

    # Store the summary statistics in the dictionary
```

```python
    categorical_summary[col] = {
        'Value Counts': value_counts,
        'Missing Percentage': missing_percentage
    }

# Step 3: Create a DataFrame to display categorical feature summaries
categorical_summary_df = pd.DataFrame({
    'Feature': [col for col in categorical_summary.keys()],
    'Value Counts': [str(categorical_summary[col]['Value Counts']) for col in␣
 ↪categorical_summary.keys()],
    'Missing Percentage (%)': [categorical_summary[col]['Missing Percentage']␣
 ↪for col in categorical_summary.keys()]
})

# Step 4: Display the summary DataFrame
print("Categorical Feature Analysis:")
print(categorical_summary_df)

# Step 5: Check for low-frequency categories (less than 5% frequency) to␣
 ↪identify potential consolidation opportunities
low_freq_categories = {}
for col in categorical_cols:
    # Identify categories that make up less than 5% of the total count
    low_freq = value_counts[value_counts < 0.05 * len(train_df)].index.tolist()
    if low_freq:
        low_freq_categories[col] = low_freq

# Display low-frequency categories if found
if low_freq_categories:
    print("\nLow-Frequency Categories Identified for Potential Consolidation:")
    for col, low_freq in low_freq_categories.items():
        print(f"Feature: {col}")
        print(f"Low-Frequency Categories: {low_freq}")
else:
    print("\nNo Low-Frequency Categories Identified for Consolidation.")
```

```
Categorical Feature Analysis:
                          Feature  \
0                              id
1    Basic_Demos-Enroll_Season
2                  CGAS-Season
3              Physical-Season
4     Fitness_Endurance-Season
5                   FGC-Season
6                   BIA-Season
7                 PAQ_A-Season
8                 PAQ_C-Season
```

```
9                PCIAT-Season
10                 SDS-Season
11         PreInt_EduHx-Season

                                                Value Counts  Missing Percentage (%)
0   id\nffcd4dbd    1\n00008ff9    1\n00105258    …                  0.000000
1   Basic_Demos-Enroll_Season\nSpring    591\nWint…                  0.000000
2   CGAS-Season\nSpring    588\nFall      539\nSum…                  0.000000
3   Physical-Season\nSpring    574\nWinter    496\…                  6.886657
4   Fitness_Endurance-Season\nNaN       1035\nSpri…                 49.497848
5   FGC-Season\nSpring    628\nSummer    461\nFall…                  4.160689
6   BIA-Season\nNaN       716\nSummer    454\nFall…                 34.241989
7   PAQ_A-Season\nNaN       1852\nWinter      67\n…                 88.570062
8   PAQ_C-Season\nNaN       1008\nSpring     329\n…                 48.206600
9   PCIAT-Season\nSpring    599\nWinter    496\nSu…                  3.347681
10  SDS-Season\nSpring    552\nWinter    494\nFall…                  8.608321
11  PreInt_EduHx-Season\nSpring    579\nWinter    …                  1.004304

Low-Frequency Categories Identified for Potential Consolidation:
Feature: id
Low-Frequency Categories: [nan]
Feature: Basic_Demos-Enroll_Season
Low-Frequency Categories: [nan]
Feature: CGAS-Season
Low-Frequency Categories: [nan]
Feature: Physical-Season
Low-Frequency Categories: [nan]
Feature: Fitness_Endurance-Season
Low-Frequency Categories: [nan]
Feature: FGC-Season
Low-Frequency Categories: [nan]
Feature: BIA-Season
Low-Frequency Categories: [nan]
Feature: PAQ_A-Season
Low-Frequency Categories: [nan]
Feature: PAQ_C-Season
Low-Frequency Categories: [nan]
Feature: PCIAT-Season
Low-Frequency Categories: [nan]
Feature: SDS-Season
Low-Frequency Categories: [nan]
Feature: PreInt_EduHx-Season
Low-Frequency Categories: [nan]
```

Code Explanation: Categorical Feature Analysis: The script first identifies the categorical features in the training data and creates a summary of each feature's value counts and missing percentage.

Low-Frequency Categories: It then checks for low-frequency categories (less than 5% of the total) to identify potential categories that might need consolidation.

Deriving New Features: It attempts to derive new categorical features based on combinations of existing features (e.g., combining seasons with age group).

Correlation analysis

[8]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Extract numeric columns from the training dataset
numeric_features = train_df.select_dtypes(include=[np.number]).columns

# Step 2: Compute the correlation matrix for the numeric features
correlation_matrix = train_df[numeric_features].corr()

# Step 3: Display the correlation matrix to identify relationships between␣
↪features
print("Correlation Matrix for Numeric Features:")
print(correlation_matrix)

# Step 4: Visualize the correlation matrix using a heatmap to identify␣
↪multicollinearity or strong relationships
plt.figure(figsize=(18, 12))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', linewidths=0.5,␣
↪fmt='.2f')
plt.title('Correlation Matrix Heatmap for Numeric Features')
plt.show()

# Step 5: Identify highly correlated pairs (correlation > 0.8 or < -0.8)
highly_correlated_pairs = []
threshold = 0.8

for i in range(len(correlation_matrix.columns)):
    for j in range(i + 1, len(correlation_matrix.columns)):
        if abs(correlation_matrix.iloc[i, j]) > threshold:
            highly_correlated_pairs.append((correlation_matrix.columns[i],␣
↪correlation_matrix.columns[j], correlation_matrix.iloc[i, j]))

# Step 6: Display highly correlated pairs if found
if highly_correlated_pairs:
    print("\nHighly Correlated Pairs (|correlation| > 0.8):")
    for pair in highly_correlated_pairs:
        print(f"Feature 1: {pair[0]}, Feature 2: {pair[1]}, Correlation:␣
↪{pair[2]:.2f}")
else:
    print("\nNo highly correlated pairs found with |correlation| > 0.8.")
```

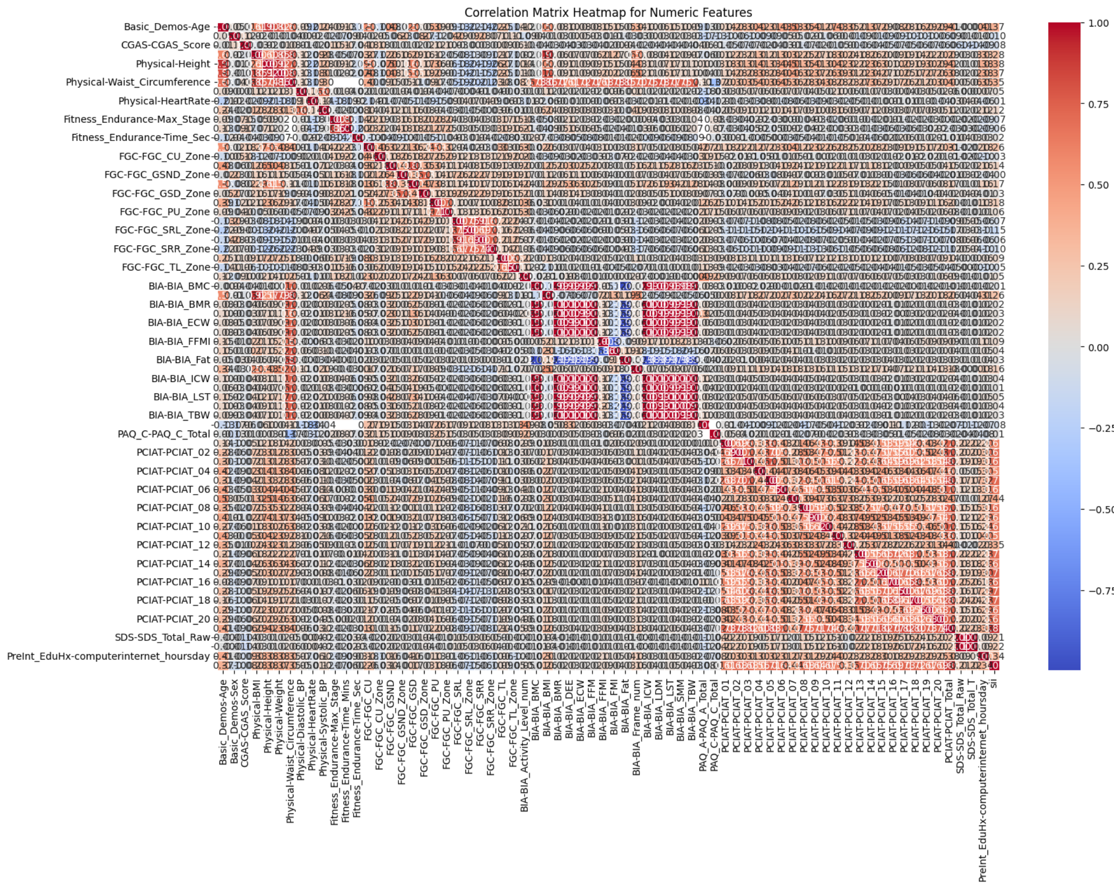
```
Correlation Matrix for Numeric Features:
                                        Basic_Demos-Age  Basic_Demos-Sex  \
```

```
Basic_Demos-Age                               1.000000         0.050142
Basic_Demos-Sex                               0.050142         1.000000
CGAS-CGAS_Score                               0.006198         0.120547
Physical-BMI                                  0.605714         0.015096
Physical-Height                               0.899945        -0.011227
…                                                  …              …
PCIAT-PCIAT_Total                             0.406339        -0.085854
SDS-SDS_Total_Raw                            -0.003186        -0.014292
SDS-SDS_Total_T                              -0.000120        -0.014965
PreInt_EduHx-computerinternet_hoursday        0.410343        -0.000958
sii                                           0.366741        -0.103771

                                        CGAS-CGAS_Score  Physical-BMI  \
Basic_Demos-Age                                0.006198      0.605714
Basic_Demos-Sex                                0.120547      0.015096
CGAS-CGAS_Score                                1.000000     -0.030896
Physical-BMI                                  -0.030896      1.000000
Physical-Height                                0.016189      0.604256
…                                                   …           …
PCIAT-PCIAT_Total                             -0.063753      0.290174
SDS-SDS_Total_Raw                             -0.137249      0.033086
SDS-SDS_Total_T                               -0.140590      0.033912
PreInt_EduHx-computerinternet_hoursday        -0.088299      0.327840
sii                                           -0.083817      0.282010

                                        Physical-Height  Physical-Weight  \
Basic_Demos-Age                                0.899945         0.818487
Basic_Demos-Sex                               -0.011227        -0.005965
CGAS-CGAS_Score                                0.016189        -0.010907
Physical-BMI                                   0.604256         0.849793
Physical-Height                                1.000000         0.917794
…                                                   …              …
PCIAT-PCIAT_Total                              0.419714         0.379534
SDS-SDS_Total_Raw                              0.006186         0.021613
SDS-SDS_Total_T                                0.009129         0.023383
PreInt_EduHx-computerinternet_hoursday         0.382879         0.379950
sii                                            0.381331         0.365275

                                        Physical-Waist_Circumference  \
Basic_Demos-Age                                             0.681502
Basic_Demos-Sex                                            -0.042630
CGAS-CGAS_Score                                             0.081593
Physical-BMI                                                0.845480
Physical-Height                                             0.744609
…                                                                …
PCIAT-PCIAT_Total                                           0.398531
SDS-SDS_Total_Raw                                           0.049942
SDS-SDS_Total_T                                             0.055975
```

```
PreInt_EduHx-computerinternet_hoursday                      0.349646
sii                                                         0.346784

                                        Physical-Diastolic_BP  \
Basic_Demos-Age                                      0.088584
Basic_Demos-Sex                                      0.004784
CGAS-CGAS_Score                                      0.006513
Physical-BMI                                         0.120995
Physical-Height                                      0.115766
…                                                       …
PCIAT-PCIAT_Total                                    0.061751
SDS-SDS_Total_Raw                                   -0.000356
SDS-SDS_Total_T                                      0.000037
PreInt_EduHx-computerinternet_hoursday               0.065459
sii                                                  0.048643

                                        Physical-HeartRate  \
Basic_Demos-Age                                  -0.210248
Basic_Demos-Sex                                   0.020126
CGAS-CGAS_Score                                  -0.015412
Physical-BMI                                     -0.094315
Physical-Height                                  -0.210081
…                                                    …
PCIAT-PCIAT_Total                                -0.032687
SDS-SDS_Total_Raw                                 0.044391
SDS-SDS_Total_T                                   0.041174
PreInt_EduHx-computerinternet_hoursday           -0.059209
sii                                              -0.010793

                                        Physical-Systolic_BP  …  \
Basic_Demos-Age                                     0.243244  …
Basic_Demos-Sex                                    -0.017825  …
CGAS-CGAS_Score                                    -0.010918  …
Physical-BMI                                        0.275664  …
Physical-Height                                     0.275643  …
…                                                      …   …
PCIAT-PCIAT_Total                                   0.123934  …
SDS-SDS_Total_Raw                                   0.018855  …
SDS-SDS_Total_T                                     0.020799  …
PreInt_EduHx-computerinternet_hoursday              0.124453  …
sii                                                 0.116880  …

                                        PCIAT-PCIAT_16  PCIAT-PCIAT_17  \
Basic_Demos-Age                               0.079168        0.280845
Basic_Demos-Sex                              -0.093549       -0.096496
CGAS-CGAS_Score                              -0.067807       -0.050509
Physical-BMI                                  0.086322        0.190715
Physical-Height                               0.103693        0.292709
```

```
…                                                 …             …
PCIAT-PCIAT_Total                          0.765380      0.823011
SDS-SDS_Total_Raw                          0.251805      0.164411
SDS-SDS_Total_T                            0.256112      0.169059
PreInt_EduHx-computerinternet_hoursday     0.177488      0.290029
sii                                        0.688977      0.725643

                                        PCIAT-PCIAT_18  PCIAT-PCIAT_19  \
Basic_Demos-Age                               0.163207        0.292392
Basic_Demos-Sex                              -0.104626       -0.095834
CGAS-CGAS_Score                              -0.064355       -0.065812
Physical-BMI                                  0.142483        0.216825
Physical-Height                               0.186359        0.295952
…                                                  …             …
PCIAT-PCIAT_Total                             0.798623        0.732603
SDS-SDS_Total_Raw                             0.237363        0.152151
SDS-SDS_Total_T                               0.241926        0.156185
PreInt_EduHx-computerinternet_hoursday        0.226288        0.288673
sii                                           0.722854        0.672012

                                        PCIAT-PCIAT_20  PCIAT-PCIAT_Total  \
Basic_Demos-Age                               0.287335           0.406339
Basic_Demos-Sex                              -0.060637          -0.085854
CGAS-CGAS_Score                              -0.064813          -0.063753
Physical-BMI                                  0.195066           0.290174
Physical-Height                               0.290612           0.419714
…                                                  …                …
PCIAT-PCIAT_Total                             0.744223           1.000000
SDS-SDS_Total_Raw                             0.197601           0.221114
SDS-SDS_Total_T                               0.201040           0.226563
PreInt_EduHx-computerinternet_hoursday        0.251600           0.375702
sii                                           0.691154           0.892045

                                        SDS-SDS_Total_Raw  SDS-SDS_Total_T  \
Basic_Demos-Age                                 -0.003186        -0.000120
Basic_Demos-Sex                                 -0.014292        -0.014965
CGAS-CGAS_Score                                 -0.137249        -0.140590
Physical-BMI                                     0.033086         0.033912
Physical-Height                                  0.006186         0.009129
…                                                     …              …
PCIAT-PCIAT_Total                                0.221114         0.226563
SDS-SDS_Total_Raw                                1.000000         0.997213
SDS-SDS_Total_T                                  0.997213         1.000000
PreInt_EduHx-computerinternet_hoursday           0.085628         0.088264
sii                                              0.211216         0.216087

                                        PreInt_EduHx-computerinternet_hoursday
\
```

```
Basic_Demos-Age                                                            0.410343
Basic_Demos-Sex                                                           -0.000958
CGAS-CGAS_Score                                                           -0.088299
Physical-BMI                                                               0.327840
Physical-Height                                                            0.382879
…                                                                             …
PCIAT-PCIAT_Total                                                          0.375702
SDS-SDS_Total_Raw                                                          0.085628
SDS-SDS_Total_T                                                            0.088264
PreInt_EduHx-computerinternet_hoursday                                     1.000000
sii                                                                        0.340314

                                             sii
Basic_Demos-Age                         0.366741
Basic_Demos-Sex                        -0.103771
CGAS-CGAS_Score                        -0.083817
Physical-BMI                            0.282010
Physical-Height                         0.381331
…                                          …
PCIAT-PCIAT_Total                       0.892045
SDS-SDS_Total_Raw                       0.211216
SDS-SDS_Total_T                         0.216087
PreInt_EduHx-computerinternet_hoursday  0.340314
sii                                     1.000000

[70 rows x 70 columns]
```




```
Highly Correlated Pairs (|correlation| > 0.8):
Feature 1: Basic_Demos-Age, Feature 2: Physical-Height, Correlation: 0.90
Feature 1: Basic_Demos-Age, Feature 2: Physical-Weight, Correlation: 0.82
Feature 1: Physical-BMI, Feature 2: Physical-Weight, Correlation: 0.85
Feature 1: Physical-BMI, Feature 2: Physical-Waist_Circumference, Correlation:
0.85
Feature 1: Physical-BMI, Feature 2: BIA-BIA_BMI, Correlation: 0.92
Feature 1: Physical-Height, Feature 2: Physical-Weight, Correlation: 0.92
Feature 1: Physical-Weight, Feature 2: Physical-Waist_Circumference,
Correlation: 0.86
Feature 1: Physical-Waist_Circumference, Feature 2: BIA-BIA_BMI, Correlation:
0.86
Feature 1: Physical-Waist_Circumference, Feature 2: BIA-BIA_Fat, Correlation:
0.85
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: Fitness_Endurance-Time_Mins,
Correlation: 0.86
Feature 1: FGC-FGC_GSND, Feature 2: FGC-FGC_GSD, Correlation: 0.81
Feature 1: FGC-FGC_SRL, Feature 2: FGC-FGC_SRR, Correlation: 0.91
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_BMR, Correlation: 0.99
```

```
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_DEE, Correlation: 0.99
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_ECW, Correlation: 0.99
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_FFM, Correlation: 0.99
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_Fat, Correlation: -1.00
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_ICW, Correlation: 0.99
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_LST, Correlation: 0.98
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_SMM, Correlation: 0.98
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_TBW, Correlation: 0.99
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_DEE, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_ECW, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_FFM, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_Fat, Correlation: -0.99
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_ICW, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_LST, Correlation: 0.99
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_SMM, Correlation: 0.99
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_ECW, Correlation: 0.99
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_FFM, Correlation: 1.00
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_Fat, Correlation: -0.98
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_ICW, Correlation: 1.00
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_LDM, Correlation: 0.99
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_LST, Correlation: 0.99
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_SMM, Correlation: 0.99
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_FFM, Correlation: 1.00
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_Fat, Correlation: -0.99
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_ICW, Correlation: 1.00
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_LST, Correlation: 0.99
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_SMM, Correlation: 0.99
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_Fat, Correlation: -0.99
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_ICW, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_LST, Correlation: 0.99
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_SMM, Correlation: 0.99
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_FFMI, Feature 2: BIA-BIA_FMI, Correlation: -0.86
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_ICW, Correlation: -0.98
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_LDM, Correlation: -0.99
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_LST, Correlation: -0.97
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_SMM, Correlation: -0.98
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_TBW, Correlation: -0.99
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_LST, Correlation: 1.00
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_SMM, Correlation: 1.00
```

```
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_LDM, Feature 2: BIA-BIA_LST, Correlation: 0.99
Feature 1: BIA-BIA_LDM, Feature 2: BIA-BIA_SMM, Correlation: 0.99
Feature 1: BIA-BIA_LDM, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_LST, Feature 2: BIA-BIA_SMM, Correlation: 1.00
Feature 1: BIA-BIA_LST, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_SMM, Feature 2: BIA-BIA_TBW, Correlation: 0.99
Feature 1: PCIAT-PCIAT_03, Feature 2: PCIAT-PCIAT_Total, Correlation: 0.82
Feature 1: PCIAT-PCIAT_05, Feature 2: PCIAT-PCIAT_Total, Correlation: 0.83
Feature 1: PCIAT-PCIAT_15, Feature 2: PCIAT-PCIAT_Total, Correlation: 0.82
Feature 1: PCIAT-PCIAT_16, Feature 2: PCIAT-PCIAT_18, Correlation: 0.84
Feature 1: PCIAT-PCIAT_17, Feature 2: PCIAT-PCIAT_Total, Correlation: 0.82
Feature 1: PCIAT-PCIAT_Total, Feature 2: sii, Correlation: 0.89
Feature 1: SDS-SDS_Total_Raw, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
```

analyzing the spread and missingness of PCIAT-PCIAT_01 to PCIAT-PCIAT_20:

[9]:
```python
# Step 1: Analyze Spread and Missingness for PCIAT Features (Updated)
import pandas as pd

# Select only PCIAT-PCIAT columns for analysis (exclude PCIAT-Season)
pciat_columns = [col for col in train_df.columns if 'PCIAT-' in col and col !=␣
↪'PCIAT-PCIAT_Total' and col != 'PCIAT-Season']

# Calculate missing value percentages for each PCIAT feature
pciat_missing_percentage = train_df[pciat_columns].isnull().mean() * 100
print("\nMissing Value Percentage for PCIAT Features:")
print(pciat_missing_percentage)

# Calculate summary statistics for each PCIAT feature
pciat_summary_stats = train_df[pciat_columns].describe().T
print("\nSummary Statistics for PCIAT Features:")
print(pciat_summary_stats)

# Identify columns with high missing values (above 50%)
high_missing_pciat_columns = pciat_missing_percentage[pciat_missing_percentage␣
↪> 50].index
print("\nPCIAT Columns with High Missing Values (Above 50%):")
print(high_missing_pciat_columns)

# Check for skewness and kurtosis to understand the spread (excluding␣
↪non-numeric columns)
pciat_skewness = train_df[pciat_columns].skew()
pciat_kurtosis = train_df[pciat_columns].kurtosis()
print("\nSkewness of PCIAT Features:")
print(pciat_skewness)
print("\nKurtosis of PCIAT Features:")
print(pciat_kurtosis)
```

```
Missing Value Percentage for PCIAT Features:
PCIAT-PCIAT_01    3.443329
PCIAT-PCIAT_02    3.395505
PCIAT-PCIAT_03    3.538977
PCIAT-PCIAT_04    3.538977
PCIAT-PCIAT_05    3.491153
PCIAT-PCIAT_06    3.491153
PCIAT-PCIAT_07    3.491153
PCIAT-PCIAT_08    3.538977
PCIAT-PCIAT_09    3.538977
PCIAT-PCIAT_10    3.443329
PCIAT-PCIAT_11    3.347681
PCIAT-PCIAT_12    3.491153
PCIAT-PCIAT_13    3.586801
PCIAT-PCIAT_14    3.443329
PCIAT-PCIAT_15    3.538977
PCIAT-PCIAT_16    3.634625
PCIAT-PCIAT_17    3.682449
PCIAT-PCIAT_18    3.491153
PCIAT-PCIAT_19    3.491153
PCIAT-PCIAT_20    3.395505
dtype: float64
```

Summary Statistics for PCIAT Features:

	count	mean	std	min	25%	50%	75%	max
PCIAT-PCIAT_01	2019.0	2.368004	1.661607	0.0	1.0	2.0	4.0	5.0
PCIAT-PCIAT_02	2020.0	2.125743	1.679651	0.0	0.0	2.0	4.0	5.0
PCIAT-PCIAT_03	2017.0	2.365394	1.570705	0.0	1.0	2.0	4.0	5.0
PCIAT-PCIAT_04	2017.0	0.784333	1.141998	0.0	0.0	0.0	1.0	5.0
PCIAT-PCIAT_05	2018.0	2.276016	1.701768	0.0	1.0	2.0	4.0	5.0
PCIAT-PCIAT_06	2018.0	1.016353	1.234213	0.0	0.0	1.0	1.0	5.0
PCIAT-PCIAT_07	2018.0	0.543112	1.016168	0.0	0.0	0.0	1.0	5.0
PCIAT-PCIAT_08	2017.0	1.189886	1.300998	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_09	2017.0	1.001983	1.204734	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_10	2019.0	1.274889	1.300384	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_11	2021.0	1.633845	1.518703	0.0	0.0	1.0	3.0	5.0
PCIAT-PCIAT_12	2018.0	0.227453	0.492179	0.0	0.0	0.0	0.0	5.0
PCIAT-PCIAT_13	2016.0	1.314980	1.381729	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_14	2019.0	0.987618	1.257030	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_15	2017.0	1.452652	1.462787	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_16	2015.0	1.428784	1.454290	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_17	2014.0	1.583416	1.417281	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_18	2018.0	1.599108	1.505679	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_19	2018.0	1.112488	1.305468	0.0	0.0	1.0	2.0	5.0
PCIAT-PCIAT_20	2020.0	0.913861	1.160906	0.0	0.0	1.0	1.0	5.0

PCIAT Columns with High Missing Values (Above 50%):

```
Index([], dtype='object')

Skewness of PCIAT Features:
PCIAT-PCIAT_01    0.018089
PCIAT-PCIAT_02    0.185529
PCIAT-PCIAT_03   -0.044011
PCIAT-PCIAT_04    1.693028
PCIAT-PCIAT_05    0.131485
PCIAT-PCIAT_06    1.410195
PCIAT-PCIAT_07    2.355935
PCIAT-PCIAT_08    1.041128
PCIAT-PCIAT_09    1.444419
PCIAT-PCIAT_10    0.895301
PCIAT-PCIAT_11    0.521314
PCIAT-PCIAT_12    2.846544
PCIAT-PCIAT_13    0.937216
PCIAT-PCIAT_14    1.340420
PCIAT-PCIAT_15    0.769705
PCIAT-PCIAT_16    0.893260
PCIAT-PCIAT_17    0.597016
PCIAT-PCIAT_18    0.745325
PCIAT-PCIAT_19    1.266126
PCIAT-PCIAT_20    1.550346
dtype: float64

Kurtosis of PCIAT Features:
PCIAT-PCIAT_01    -1.185968
PCIAT-PCIAT_02    -1.181542
PCIAT-PCIAT_03    -1.030276
PCIAT-PCIAT_04     2.529177
PCIAT-PCIAT_05    -1.191626
PCIAT-PCIAT_06     1.553860
PCIAT-PCIAT_07     5.650708
PCIAT-PCIAT_08     0.338986
PCIAT-PCIAT_09     1.871533
PCIAT-PCIAT_10     0.091892
PCIAT-PCIAT_11    -0.861567
PCIAT-PCIAT_12    13.212843
PCIAT-PCIAT_13     0.060331
PCIAT-PCIAT_14     1.172628
PCIAT-PCIAT_15    -0.382460
PCIAT-PCIAT_16    -0.114011
PCIAT-PCIAT_17    -0.518220
PCIAT-PCIAT_18    -0.387139
PCIAT-PCIAT_19     0.977074
PCIAT-PCIAT_20     2.205397
dtype: float64
```

[10]:
```
train_df['PCIAT-PCIAT_Total'].value_counts()
```

[10]:
```
PCIAT-PCIAT_Total
0.0     233
20.0     51
30.0     50
31.0     49
27.0     46
       …
78.0      2
82.0      1
76.0      1
79.0      1
91.0      1
Name: count, Length: 89, dtype: int64
```

Implementation of GMM for Predicting PCIAT-PCIAT_Total:

[11]:
```
# Generate interaction and polynomial terms for PCIAT columns
from itertools import combinations
from sklearn.preprocessing import PolynomialFeatures

# Create interaction terms
interaction_cols = ['PCIAT-PCIAT_01', 'PCIAT-PCIAT_02', 'PCIAT-PCIAT_05',␣
 ↪'PCIAT-PCIAT_11']
interaction_data = pd.DataFrame(index=train_df.index)

# Generate interaction terms
for col1, col2 in combinations(interaction_cols, 2):
    interaction_data[f"{col1}_x_{col2}"] = train_df[col1] * train_df[col2]

# Generate polynomial features
poly = PolynomialFeatures(degree=2, include_bias=False)
poly_data = pd.DataFrame(poly.fit_transform(train_df[interaction_cols].
 ↪fillna(0)))
poly_data.columns = poly.get_feature_names_out(interaction_cols)

# Combine interaction and polynomial features with existing data
train_df = pd.concat([train_df, interaction_data, poly_data], axis=1)
```

[12]:
```
train_df.head()
```

[12]:
```
          id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0   00008ff9                      Fall              5.0              0.0
2   00105258                    Summer             10.0              1.0
3   00115b9f                    Winter              9.0              0.0
5   001f3379                    Spring             13.0              1.0
11  00abe655                      Fall             11.0              0.0
```

```
   CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  \
0       Winter             51.0            Fall     16.877316
2         Fall             71.0            Fall     16.648696
3         Fall             71.0          Summer     18.292347
5       Winter             50.0          Summer     22.279952
11      Summer             66.0             NaN           NaN

    Physical-Height  Physical-Weight  …  PCIAT-PCIAT_01^2  \
0              46.0             50.8  …              25.0
2              56.5             75.6  …              16.0
3              56.0             81.6  …               9.0
5              59.5            112.2  …               0.0
11              NaN              NaN  …               9.0

    PCIAT-PCIAT_01 PCIAT-PCIAT_02  PCIAT-PCIAT_01 PCIAT-PCIAT_05  \
0                            20.0                           20.0
2                             8.0                           20.0
3                             9.0                            6.0
5                             0.0                            0.0
11                            9.0                            3.0

    PCIAT-PCIAT_01 PCIAT-PCIAT_11 PCIAT-PCIAT_02^2  \
0                            20.0              16.0
2                            12.0               4.0
3                             0.0               9.0
5                             0.0               0.0
11                            0.0               9.0

    PCIAT-PCIAT_02 PCIAT-PCIAT_05  PCIAT-PCIAT_02 PCIAT-PCIAT_11  \
0                            16.0                           16.0
2                            10.0                            6.0
3                             6.0                            0.0
5                             0.0                            0.0
11                            3.0                            0.0

    PCIAT-PCIAT_05^2 PCIAT-PCIAT_05 PCIAT-PCIAT_11  PCIAT-PCIAT_11^2
0               16.0                          16.0              16.0
2               25.0                          15.0               9.0
3                4.0                           0.0               0.0
5                0.0                           0.0               0.0
11               1.0                           0.0               0.0

[5 rows x 103 columns]
```

extracting the PCIAT-related columns and then applying standard scaling using StandardScaler. This step ensures that all the PCIAT features are on the same scale, which is crucial for clustering methods like Gaussian Mixture Models (GMM).

[13]:
```
# Step 1: Define and Standardize PCIAT Columns Before Clustering
# Ensure the variable `scaled_pciat_data` is created using the following lines

from sklearn.preprocessing import StandardScaler

# Extract PCIAT-related columns (excluding non-numeric ones)
pciat_columns = [col for col in train_df.columns if 'PCIAT-' in col and col !=␣
↪'PCIAT-PCIAT_Total' and col != 'PCIAT-Season']
pciat_data = train_df[pciat_columns]

# Apply Standard Scaling to the PCIAT columns
scaler = StandardScaler()
scaled_pciat_data = scaler.fit_transform(pciat_data)

# Now, proceed with your clustering and analysis steps
```

pciat_columns is a list of all columns containing "PCIAT-", excluding PCIAT-PCIAT_Total (target column) and PCIAT-Season (non-numeric).

scaled_pciat_data stores the standardized values, which means each column will have a mean of 0 and a standard deviation of 1. This is essential to ensure that no single feature dominates the clustering process due to differences in scale.

helper functions for initializing GMM parameters and implementing the Expectation-Maximization (EM) algorithm to handle missing data.

Initialization (initialize_gmm_params):

Randomly initializes the means, covariances, and weights for the specified number of components (n_components).

E-Step (e_step):

Calculates the responsibility matrix, which represents the probability of each sample belonging to each Gaussian component. The function extracts non-missing values from each sample to handle the missing data scenario.

M-Step (m_step):

Updates the means, covariances, and weights of each component using the responsibilities computed during the E-step.

[14]:
```
import numpy as np
import pandas as pd
from scipy.stats import multivariate_normal

# Step 1: Extract PCIAT columns and prepare data for GMM with missing values
pciat_columns = [col for col in train_df.columns if 'PCIAT-' in col and col !=␣
↪'PCIAT-PCIAT_Total' and col != 'PCIAT-Season']
pciat_data = train_df[pciat_columns].values
```

```
# Step 2: Apply Standard Scaling
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaled_pciat_data = scaler.fit_transform(pciat_data)

# Step 3: Define helper functions for GMM with missing data (E-step, M-step,␣
↪initialization)
def initialize_gmm_params(X, n_components, random_state=42):
    np.random.seed(random_state)
    n_features = X.shape[1]
    means = np.random.rand(n_components, n_features)
    covariances = np.array([np.identity(n_features) for _ in␣
↪range(n_components)])
    weights = np.full(n_components, 1/n_components)
    return means, covariances, weights


# Updated e_step function with zero-size array and singular matrix handling
def e_step(X, means, covariances, weights, epsilon=1e-6):
    """
    Perform the E-step of the EM algorithm to calculate responsibilities.
    Handles cases where missing data leads to zero-size arrays or singular␣
↪matrices.
    X: Data with missing values (NaNs).
    means: Array of shape (n_components, n_features).
    covariances: Array of shape (n_components, n_features, n_features).
    weights: Array of shape (n_components,).
    epsilon: Small value to add to diagonal of covariance matrix for numerical␣
↪stability.
    """
    n_samples, n_features = X.shape
    n_components = means.shape[0]

    # Initialize responsibility matrix
    responsibilities = np.zeros((n_samples, n_components))

    for i in range(n_samples):
        for k in range(n_components):
            # Extract only non-missing dimensions for sample i and component k
            non_missing_idx = ~np.isnan(X[i, :])
            x_non_missing = X[i, non_missing_idx]
            mean_non_missing = means[k, non_missing_idx]

            # If no non-missing values are present, skip this component
            if x_non_missing.size == 0:
                responsibilities[i, k] = 0
                continue
```

```
            # Subset the covariance matrix to include only non-missing features
            cov_non_missing = covariances[k][np.ix_(non_missing_idx,␣
↪non_missing_idx)]

            # Add epsilon to the diagonal to handle potential singular matrices
            cov_non_missing += epsilon * np.eye(cov_non_missing.shape[0])

            # Calculate the multivariate normal probability for non-missing data
            try:
                responsibilities[i, k] = weights[k] * multivariate_normal.
↪pdf(x_non_missing, mean_non_missing, cov_non_missing)
            except (LinAlgError, ValueError) as e:
                print(f"Error encountered for sample {i} and component {k}:␣
↪{e}")
                responsibilities[i, k] = 0  # Set responsibility to 0 in case␣
↪of an error

        # Normalize responsibilities for sample i if the sum is not zero
        total_responsibility = np.sum(responsibilities[i, :])
        if total_responsibility > 0:
            responsibilities[i, :] /= total_responsibility
        else:
            # If no responsibilities are assigned (e.g., due to errors), assign␣
↪equal responsibility
            responsibilities[i, :] = 1.0 / n_components

    return responsibilities

#def m_step(X, responsibilities):
#    n_samples, n_features = X.shape
#    n_components = responsibilities.shape[1]
#    effective_samples = responsibilities.sum(axis=0)
#    weights = effective_samples / n_samples
#    means = np.zeros((n_components, n_features))
#    covariances = np.zeros((n_components, n_features, n_features))
#    for k in range(n_components):
#        for i in range(n_samples):
#            non_missing_idx = ~np.isnan(X[i, :])
#            means[k, non_missing_idx] += responsibilities[i, k] * X[i,␣
↪non_missing_idx]
#        means[k] /= effective_samples[k]
#        for i in range(n_samples):
#            non_missing_idx = ~np.isnan(X[i, :])
#            x_centered = (X[i, non_missing_idx] - means[k, non_missing_idx]).
↪reshape(-1, 1)
```

```
#            covariances[k][np.ix_(non_missing_idx, non_missing_idx)] +=␣
↪responsibilities[i, k] * np.dot(x_centered, x_centered.T)
#        covariances[k] /= effective_samples[k]
#    return means, covariances, weights

# Modified M-step with covariance regularization
def m_step(X, responsibilities, reg_value=1e-5):
    n_samples, n_features = X.shape
    n_components = responsibilities.shape[1]
    effective_samples = responsibilities.sum(axis=0)
    weights = effective_samples / n_samples
    means = np.zeros((n_components, n_features))
    covariances = np.zeros((n_components, n_features, n_features))
    for k in range(n_components):
        for i in range(n_samples):
            non_missing_idx = ~np.isnan(X[i, :])
            means[k, non_missing_idx] += responsibilities[i, k] * X[i,␣
↪non_missing_idx]
        means[k] /= effective_samples[k]
        for i in range(n_samples):
            non_missing_idx = ~np.isnan(X[i, :])
            x_centered = (X[i, non_missing_idx] - means[k, non_missing_idx]).
↪reshape(-1, 1)
            covariances[k][np.ix_(non_missing_idx, non_missing_idx)] +=␣
↪responsibilities[i, k] * np.dot(x_centered, x_centered.T)
        covariances[k] /= effective_samples[k]
        covariances[k] += np.eye(n_features) * reg_value  # Add regularization
    return means, covariances, weights
```

Initialization: Randomly initializes means, identity matrices for covariances, and equal weights.

E-Step: For each sample and each component, only non-missing features are used to calculate the probability density. Regularization (epsilon) is added to the covariance matrices to ensure numerical stability.

M-Step: Updates the parameters based on the responsibilities calculated in the E-step.

function gmm_with_missing_data iteratively performs the E-step and M-step until convergence, determined by the change in log-likelihood falling below a specified tolerance (tol).

[15]:
```
# Step 4: Implement GMM with Missing Data
def gmm_with_missing_data(X, n_components=4, max_iter=100, tol=1e-4,␣
↪random_state=42):
    means, covariances, weights = initialize_gmm_params(X, n_components,␣
↪random_state)
    prev_log_likelihood = -np.inf
    for iteration in range(max_iter):
        responsibilities = e_step(X, means, covariances, weights)
        means, covariances, weights = m_step(X, responsibilities)
```

```
        log_likelihood = np.sum(np.log(np.sum(responsibilities, axis=1)))
        if np.abs(log_likelihood - prev_log_likelihood) < tol:
            print(f"Converged at iteration {iteration} for {n_components}␣
↪components.")
            break
        prev_log_likelihood = log_likelihood
    return means, covariances, weights, responsibilities, log_likelihood
```

Performs the EM algorithm using e_step and m_step in a loop until convergence. The convergence is determined when the difference in log-likelihood between iterations is less than tol.

Returns the final means, covariances, weights, responsibilities, and the final log_likelihood.

experiment_cluster_sizes is used to test the EM algorithm with varying numbers of clusters to identify the best number of components for GMM based on log-likelihood.

[16]:

```
# Step 5: Experiment with Different Cluster Sizes
def experiment_cluster_sizes(X, cluster_sizes=[2, 3, 4, 5, 6], max_iter=100,␣
↪tol=1e-4, random_state=42):
    log_likelihoods = []
    best_gmm_params = {}
    for n_components in cluster_sizes:
        print(f"Testing with {n_components} clusters...")
        means, covariances, weights, responsibilities, log_likelihood =␣
↪gmm_with_missing_data(
            X, n_components=n_components, max_iter=max_iter, tol=tol,␣
↪random_state=random_state)
        log_likelihoods.append(log_likelihood)
        best_gmm_params[n_components] = (means, covariances, weights,␣
↪responsibilities)
    return log_likelihoods, best_gmm_params

# Step 6: Run the Cluster Size Experimentation
# Experiment with more clusters
cluster_sizes = [6, 7, 8, 9, 10]  # Extended cluster sizes
log_likelihoods, gmm_params_by_cluster =␣
↪experiment_cluster_sizes(scaled_pciat_data, cluster_sizes)

# Select the best model based on log-likelihood
best_n_components = cluster_sizes[np.argmax(log_likelihoods)]
best_means, best_covariances, best_weights, best_responsibilities =␣
↪gmm_params_by_cluster[best_n_components]
```

```
Testing with 6 clusters…
Converged at iteration 1 for 6 components.
Testing with 7 clusters…
Converged at iteration 1 for 7 components.
Testing with 8 clusters…
Converged at iteration 1 for 8 components.
```

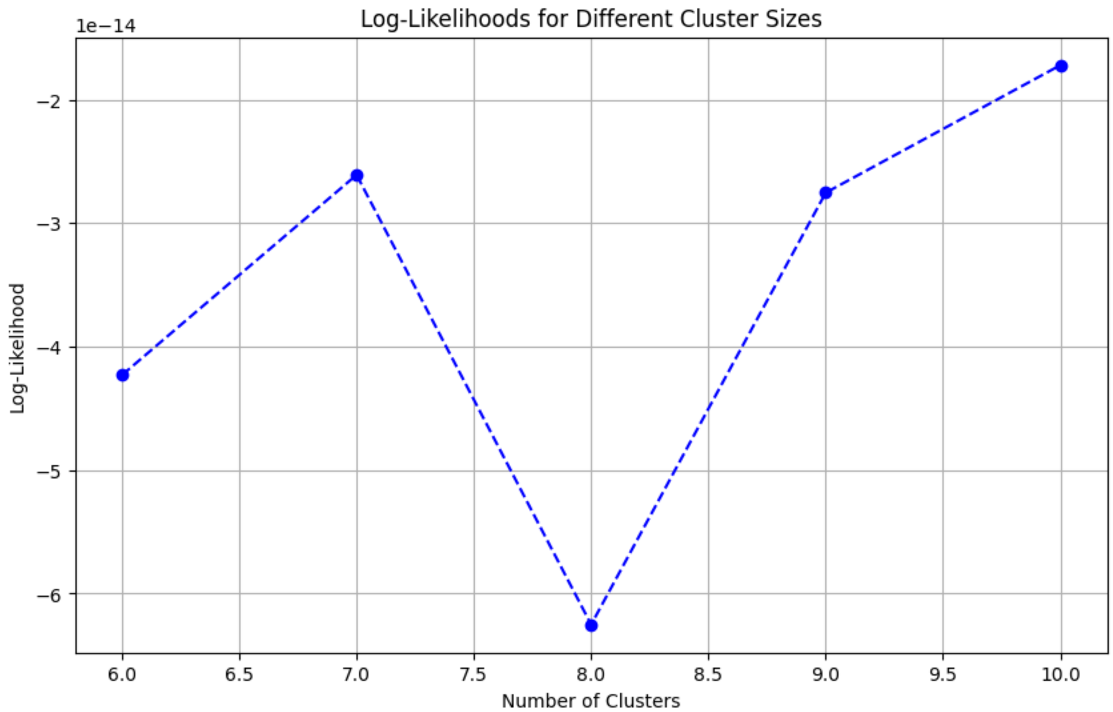
```
Testing with 9 clusters…
Converged at iteration 1 for 9 components.
Testing with 10 clusters…
Converged at iteration 1 for 10 components.
```

Tests cluster sizes ranging from 2 to 6.

For each cluster size, gmm_with_missing_data is called, and the log-likelihood values are stored to identify the best-performing model.

After identifying the best cluster size, the code fills the missing PCIAT-PCIAT_Total values based on the derived clusters.

[17]:
```python
# Step 4: Plot Log-Likelihoods for Different Cluster Sizes
plt.figure(figsize=(10, 6))
plt.plot(cluster_sizes, log_likelihoods, marker='o', linestyle='--', color='b')
plt.title("Log-Likelihoods for Different Cluster Sizes")
plt.xlabel("Number of Clusters")
plt.ylabel("Log-Likelihood")
plt.grid(True)
plt.show()

# Step 5: Select the Best Model Based on Log-Likelihood
best_n_components = cluster_sizes[np.argmax(log_likelihoods)]
best_means, best_covariances, best_weights, best_responsibilities =␣
↪gmm_params_by_cluster[best_n_components]
print(f"\nSelected {best_n_components} clusters as the best model based on␣
↪log-likelihood.")

# Step 6: Derive Missing PCIAT-PCIAT_Total Values Using Best GMM Parameters
# Calculate the mean PCIAT-PCIAT_Total for each cluster using non-missing values
train_df['PCIAT_Cluster'] = np.argmax(best_responsibilities, axis=1)  # Assign␣
↪clusters
cluster_means = train_df.groupby('PCIAT_Cluster')['PCIAT-PCIAT_Total'].mean()

# Fill missing PCIAT-PCIAT_Total values based on cluster-specific mean
train_df['PCIAT-PCIAT_Total'] = train_df.apply(
    lambda row: cluster_means[row['PCIAT_Cluster']] if np.
↪isnan(row['PCIAT-PCIAT_Total']) else row['PCIAT-PCIAT_Total'], axis=1
)

# Step 7: Validation - Calculate RMSE and MSE
# Get original non-missing and imputed values for comparison
original_pciat_total = train_df[~train_df['PCIAT-PCIAT_Total'].
↪isnull()]['PCIAT-PCIAT_Total']
imputed_pciat_total = train_df[~train_df['PCIAT-PCIAT_Total'].isnull()].
↪groupby('PCIAT_Cluster')['PCIAT-PCIAT_Total'].transform('mean')
```




```
Selected 10 clusters as the best model based on log-likelihood.
```

The code assigns clusters to each sample based on the maximum responsibility.

The mean PCIAT-PCIAT_Total value for each cluster is used to fill in the missing values, ensuring the filled values reflect the cluster's characteristics.

code calculates the Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) between the original PCIAT-PCIAT_Total values and the cluster-derived values to validate the imputation.

[18]:
```
# Calculate RMSE and MSE
from sklearn.metrics import mean_squared_error
rmse = np.sqrt(mean_squared_error(original_pciat_total, imputed_pciat_total))
mse = mean_squared_error(original_pciat_total, imputed_pciat_total)

print(f"\nValidation Results for PCIAT-PCIAT_Total Imputation:")
print(f"RMSE: {rmse:.4f}")
print(f"MSE: {mse:.4f}")
```

```
Validation Results for PCIAT-PCIAT_Total Imputation:
RMSE: 9.9423
MSE: 98.8499
```

RMSE: Measures the standard deviation of the residuals (prediction errors). A lower RMSE indicates better accuracy.

MSE: Measures the average squared difference between the observed and predicted values. A lower MSE indicates better fit.

[19]:
```
pd.set_option('display.max_columns', None)
train_df.head()
```

[19]:
```
          id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0   00008ff9                      Fall              5.0              0.0
2   00105258                    Summer             10.0              1.0
3   00115b9f                    Winter              9.0              0.0
5   001f3379                    Spring             13.0              1.0
11  00abe655                      Fall             11.0              0.0

   CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  \
0       Winter             51.0            Fall     16.877316
2         Fall             71.0            Fall     16.648696
3         Fall             71.0          Summer     18.292347
5       Winter             50.0          Summer     22.279952
11      Summer             66.0             NaN           NaN

    Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0              46.0             50.8                           NaN
2              56.5             75.6                           NaN
3              56.0             81.6                           NaN
5              59.5            112.2                           NaN
11              NaN              NaN                           NaN

    Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
0                     NaN                 NaN                   NaN
2                    65.0                94.0                 117.0
3                    60.0                97.0                 117.0
5                    60.0                73.0                 102.0
11                    NaN                 NaN                   NaN

   Fitness_Endurance-Season  Fitness_Endurance-Max_Stage  \
0                       NaN                          NaN
2                      Fall                          5.0
3                    Summer                          6.0
5                       NaN                          NaN
11                      NaN                          NaN

    Fitness_Endurance-Time_Mins  Fitness_Endurance-Time_Sec FGC-Season  \
0                           NaN                         NaN       Fall
2                           7.0                        33.0       Fall
3                           9.0                        37.0     Summer
5                           NaN                         NaN     Summer
11                          NaN                         NaN     Winter
```

```
   FGC-FGC_CU  FGC-FGC_CU_Zone  FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  \
0         0.0              0.0           NaN                NaN          NaN
2        20.0              1.0          10.2                1.0         14.7
3        18.0              1.0           NaN                NaN          NaN
5        12.0              0.0          16.5                2.0         17.9
11        NaN              NaN           NaN                NaN          NaN

   FGC-FGC_GSD_Zone  FGC-FGC_PU  FGC-FGC_PU_Zone  FGC-FGC_SRL  \
0               NaN         0.0              0.0          7.0
2               2.0         7.0              1.0         10.0
3               NaN         5.0              0.0          7.0
5               2.0         6.0              0.0         10.0
11              NaN         NaN              NaN          NaN

   FGC-FGC_SRL_Zone  FGC-FGC_SRR  FGC-FGC_SRR_Zone  FGC-FGC_TL  \
0               0.0          6.0               0.0         6.0
2               1.0         10.0               1.0         5.0
3               0.0          7.0               0.0         7.0
5               1.0         11.0               1.0         8.0
11              NaN          NaN               NaN         NaN

   FGC-FGC_TL_Zone BIA-Season  BIA-BIA_Activity_Level_num  BIA-BIA_BMC  \
0              1.0       Fall                         2.0      2.66855
2              0.0        NaN                         NaN          NaN
3              1.0     Summer                         3.0      3.84191
5              0.0     Summer                         2.0      4.33036
11             NaN        NaN                         NaN          NaN

   BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0      16.8792      932.498      1492.00      8.25598      41.5862
2          NaN          NaN          NaN          NaN          NaN
3      18.2943     1131.430      1923.44     15.59250      62.7757
5      30.1865     1330.970      1996.45     30.21240      84.0285
11         NaN          NaN          NaN          NaN          NaN

   BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0       13.8177      3.06143      9.21377                1.0      24.4349
2           NaN          NaN          NaN                NaN          NaN
3       14.0740      4.22033     18.82430                2.0      30.4041
5       16.6877     13.49880     67.97150                2.0      32.9141
11          NaN          NaN          NaN                NaN          NaN

   BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW PAQ_A-Season  \
0      8.89536      38.9177      19.5413      32.6909          NaN
2          NaN          NaN          NaN          NaN          NaN
3     16.77900      58.9338      26.4798      45.9966          NaN
```

```
5     20.90200      79.6982      35.3804      63.1265           NaN
11         NaN          NaN          NaN          NaN           NaN

   PAQ_A-PAQ_A_Total PAQ_C-Season  PAQ_C-PAQ_C_Total PCIAT-Season  \
0                NaN          NaN                NaN         Fall
2                NaN       Summer              2.170         Fall
3                NaN       Winter              2.451       Summer
5                NaN       Spring              4.110       Summer
11               NaN       Winter              1.100       Winter

   PCIAT-PCIAT_01  PCIAT-PCIAT_02  PCIAT-PCIAT_03  PCIAT-PCIAT_04  \
0             5.0             4.0             4.0             0.0
2             5.0             2.0             2.0             1.0
3             4.0             2.0             4.0             0.0
5             3.0             3.0             3.0             0.0
11            2.0             2.0             1.0             0.0

   PCIAT-PCIAT_05  PCIAT-PCIAT_06  PCIAT-PCIAT_07  PCIAT-PCIAT_08  \
0             4.0             0.0             0.0             4.0
2             2.0             1.0             1.0             2.0
3             5.0             1.0             0.0             3.0
5             2.0             1.0             0.0             2.0
11            3.0             0.0             0.0             0.0

   PCIAT-PCIAT_09  PCIAT-PCIAT_10  PCIAT-PCIAT_11  PCIAT-PCIAT_12  \
0             0.0             0.0             4.0             0.0
2             1.0             1.0             1.0             0.0
3             2.0             2.0             3.0             0.0
5             2.0             1.0             0.0             1.0
11            0.0             0.0             0.0             0.0

   PCIAT-PCIAT_13  PCIAT-PCIAT_14  PCIAT-PCIAT_15  PCIAT-PCIAT_16  \
0             4.0             4.0             4.0             4.0
2             1.0             1.0             1.0             0.0
3             3.0             0.0             0.0             3.0
5             3.0             3.0             2.0             1.0
11            0.0             0.0             0.0             0.0

   PCIAT-PCIAT_17  PCIAT-PCIAT_18  PCIAT-PCIAT_19  PCIAT-PCIAT_20  \
0             4.0             4.0             2.0             4.0
2             2.0             2.0             1.0             1.0
3             4.0             3.0             4.0             1.0
5             3.0             1.0             2.0             1.0
11            1.0             0.0             1.0             0.0

   PCIAT-PCIAT_Total SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  \
0               55.0        NaN                NaN              NaN
```

```
2             28.0      Fall            38.0            54.0
3             44.0    Summer            31.0            45.0
5             34.0    Summer            40.0            56.0
11            10.0    Winter            42.0            59.0

   PreInt_EduHx-Season  PreInt_EduHx-computerinternet_hoursday  sii Age_Group  \
0                 Fall                                     3.0  2.0       NaN
2               Summer                                     2.0  0.0      5-10
3               Winter                                     0.0  1.0      5-10
5               Spring                                     0.0  1.0     11-15
11                Fall                                     0.0  0.0     11-15

    PCIAT-PCIAT_01_x_PCIAT-PCIAT_02  PCIAT-PCIAT_01_x_PCIAT-PCIAT_05  \
0                              20.0                             20.0
2                              10.0                             10.0
3                               8.0                             20.0
5                               9.0                              6.0
11                              4.0                              6.0

    PCIAT-PCIAT_01_x_PCIAT-PCIAT_11  PCIAT-PCIAT_02_x_PCIAT-PCIAT_05  \
0                              20.0                             16.0
2                               5.0                              4.0
3                              12.0                             10.0
5                               0.0                              6.0
11                              0.0                              6.0

    PCIAT-PCIAT_02_x_PCIAT-PCIAT_11  PCIAT-PCIAT_05_x_PCIAT-PCIAT_11  \
0                              16.0                             16.0
2                               2.0                              2.0
3                               6.0                             15.0
5                               0.0                              0.0
11                              0.0                              0.0

    PCIAT-PCIAT_01  PCIAT-PCIAT_02  PCIAT-PCIAT_05  PCIAT-PCIAT_11  \
0              5.0             4.0             4.0             4.0
2              4.0             2.0             5.0             3.0
3              3.0             3.0             2.0             0.0
5              0.0             0.0             0.0             0.0
11             3.0             3.0             1.0             0.0

    PCIAT-PCIAT_01^2  PCIAT-PCIAT_01 PCIAT-PCIAT_02  \
0               25.0                           20.0
2               16.0                            8.0
3                9.0                            9.0
5                0.0                            0.0
11               9.0                            9.0
```

```
    PCIAT-PCIAT_01 PCIAT-PCIAT_05  PCIAT-PCIAT_01 PCIAT-PCIAT_11  \
0                             20.0                            20.0
2                             20.0                            12.0
3                              6.0                             0.0
5                              0.0                             0.0
11                             3.0                             0.0

    PCIAT-PCIAT_02^2  PCIAT-PCIAT_02 PCIAT-PCIAT_05  \
0               16.0                            16.0
2                4.0                            10.0
3                9.0                             6.0
5                0.0                             0.0
11               9.0                             3.0

    PCIAT-PCIAT_02 PCIAT-PCIAT_11  PCIAT-PCIAT_05^2  \
0                             16.0              16.0
2                              6.0              25.0
3                              0.0               4.0
5                              0.0               0.0
11                             0.0               1.0

    PCIAT-PCIAT_05 PCIAT-PCIAT_11  PCIAT-PCIAT_11^2  PCIAT_Cluster
0                             16.0              16.0              5
2                             15.0               9.0              7
3                              0.0               0.0              4
5                              0.0               0.0              3
11                             0.0               0.0              3
```

[20]:
```python
train_df['PCIAT-PCIAT_Total'].value_counts()
```

[20]:
```
PCIAT-PCIAT_Total
20.814159    435
0.000000     233
8.770492     186
50.897436    108
20.627907     82
            …
78.000000      2
82.000000      1
76.000000      1
79.000000      1
91.000000      1
Name: count, Length: 99, dtype: int64
```

[21]:
```python
# Step: Remove PCIAT sub-columns and clustering-specific features
# Remove PCIAT-related columns and clustering features
```

```python
columns_to_remove = [col for col in train_df.columns if 'PCIAT-' in col and col
 ↪!= 'PCIAT-PCIAT_Total']
train_df.drop(columns=columns_to_remove, inplace=True)

train_df.head()
```

[21]:

```
          id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0   00008ff9                      Fall              5.0              0.0
2   00105258                    Summer             10.0              1.0
3   00115b9f                    Winter              9.0              0.0
5   001f3379                    Spring             13.0              1.0
11  00abe655                      Fall             11.0              0.0

   CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  \
0       Winter             51.0            Fall     16.877316
2         Fall             71.0            Fall     16.648696
3         Fall             71.0          Summer     18.292347
5       Winter             50.0          Summer     22.279952
11      Summer             66.0             NaN           NaN

    Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0              46.0             50.8                           NaN
2              56.5             75.6                           NaN
3              56.0             81.6                           NaN
5              59.5            112.2                           NaN
11              NaN              NaN                           NaN

    Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
0                     NaN                 NaN                   NaN
2                    65.0                94.0                 117.0
3                    60.0                97.0                 117.0
5                    60.0                73.0                 102.0
11                    NaN                 NaN                   NaN

   Fitness_Endurance-Season  Fitness_Endurance-Max_Stage  \
0                       NaN                          NaN
2                      Fall                          5.0
3                    Summer                          6.0
5                       NaN                          NaN
11                      NaN                          NaN

    Fitness_Endurance-Time_Mins  Fitness_Endurance-Time_Sec FGC-Season  \
0                           NaN                         NaN       Fall
2                           7.0                        33.0       Fall
3                           9.0                        37.0     Summer
5                           NaN                         NaN     Summer
11                          NaN                         NaN     Winter
```

```
   FGC-FGC_CU  FGC-FGC_CU_Zone  FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  \
0         0.0              0.0           NaN                NaN          NaN
2        20.0              1.0          10.2                1.0         14.7
3        18.0              1.0           NaN                NaN          NaN
5        12.0              0.0          16.5                2.0         17.9
11        NaN              NaN           NaN                NaN          NaN

   FGC-FGC_GSD_Zone  FGC-FGC_PU  FGC-FGC_PU_Zone  FGC-FGC_SRL  \
0               NaN         0.0              0.0          7.0
2               2.0         7.0              1.0         10.0
3               NaN         5.0              0.0          7.0
5               2.0         6.0              0.0         10.0
11              NaN         NaN              NaN          NaN

   FGC-FGC_SRL_Zone  FGC-FGC_SRR  FGC-FGC_SRR_Zone  FGC-FGC_TL  \
0               0.0          6.0               0.0         6.0
2               1.0         10.0               1.0         5.0
3               0.0          7.0               0.0         7.0
5               1.0         11.0               1.0         8.0
11              NaN          NaN               NaN         NaN

   FGC-FGC_TL_Zone BIA-Season  BIA-BIA_Activity_Level_num  BIA-BIA_BMC  \
0              1.0       Fall                         2.0      2.66855
2              0.0        NaN                         NaN          NaN
3              1.0     Summer                         3.0      3.84191
5              0.0     Summer                         2.0      4.33036
11             NaN        NaN                         NaN          NaN

   BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0      16.8792      932.498      1492.00      8.25598      41.5862
2          NaN          NaN          NaN          NaN          NaN
3      18.2943     1131.430      1923.44     15.59250      62.7757
5      30.1865     1330.970      1996.45     30.21240      84.0285
11         NaN          NaN          NaN          NaN          NaN

   BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0       13.8177      3.06143      9.21377                1.0      24.4349
2           NaN          NaN          NaN                NaN          NaN
3       14.0740      4.22033     18.82430                2.0      30.4041
5       16.6877     13.49880     67.97150                2.0      32.9141
11          NaN          NaN          NaN                NaN          NaN

   BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW PAQ_A-Season  \
0      8.89536      38.9177      19.5413      32.6909          NaN
2          NaN          NaN          NaN          NaN          NaN
3     16.77900      58.9338      26.4798      45.9966          NaN
```

```
5      20.90200      79.6982      35.3804      63.1265          NaN
11          NaN          NaN          NaN          NaN          NaN

    PAQ_A-PAQ_A_Total PAQ_C-Season  PAQ_C-PAQ_C_Total  PCIAT-PCIAT_Total  \
0                 NaN          NaN                NaN               55.0
2                 NaN       Summer              2.170               28.0
3                 NaN       Winter              2.451               44.0
5                 NaN       Spring              4.110               34.0
11                NaN       Winter              1.100               10.0

   SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T PreInt_EduHx-Season  \
0         NaN                NaN              NaN                Fall
2        Fall               38.0             54.0              Summer
3      Summer               31.0             45.0              Winter
5      Summer               40.0             56.0              Spring
11     Winter               42.0             59.0                Fall

    PreInt_EduHx-computerinternet_hoursday  sii Age_Group  PCIAT_Cluster
0                                      3.0  2.0       NaN              5
2                                      2.0  0.0      5-10              7
3                                      0.0  1.0      5-10              4
5                                      0.0  1.0     11-15              3
11                                     0.0  0.0     11-15              3
```

Generate interaction terms based on key multimodal features to capture complex relationships.

Approach:

Identify Key Features: From the previous correlation analysis, we'll pick the top correlated features for interaction term creation.

Create Interaction Terms: For each pair of selected features, we'll generate interaction terms (e.g., feature1 * feature2).

Validate Interaction Terms: Ensure that these new features provide additional explanatory power without adding noise.

Implementation: Identify Key Multimodal Features:

Based on our previous correlation analysis, we'll focus on features from different modalities (e.g., cognitive and physical) that are highly correlated. Selected features for interaction terms: CGAS-CGAS_Score (Cognitive Score) with other cognitive features. Physical-BMI with other physical features like Physical-Height. Any other cross-modality interactions that might be valuable. Create Interaction Terms:

Generate pairwise interaction terms between selected features and append them to the dataset. Validate and Log Interaction Terms:

Check the correlation of new interaction terms with existing features. Ensure no multicollinearity is introduced.

```
[22]: # Step 4.1: Create Interaction Terms

# List of key multimodal features for interaction term creation based on␣
 ↪previous correlation analysis
key_features = [
    'CGAS-CGAS_Score',
    'SDS-SDS_Total_Raw',
    'Physical-BMI',
    'Physical-Height',
    'PreInt_EduHx-computerinternet_hoursday'
]

# Create an empty dictionary to store interaction terms for documentation
interaction_terms = {}

# Create pairwise interaction terms between key multimodal features
for i in range(len(key_features)):
    for j in range(i+1, len(key_features)):
        feature_1 = key_features[i]
        feature_2 = key_features[j]

        # Create a new interaction term
        interaction_term_name = f"{feature_1}_x_{feature_2}"

        # Ensure both columns are numeric before creating the interaction term
        if pd.api.types.is_numeric_dtype(train_df[feature_1]) and pd.api.types.
 ↪is_numeric_dtype(train_df[feature_2]):
            train_df[interaction_term_name] = train_df[feature_1] *␣
 ↪train_df[feature_2]

        # Store interaction term in dictionary for reference if successfully␣
 ↪created
        if interaction_term_name in train_df.columns:
            interaction_terms[interaction_term_name] = (feature_1, feature_2)

# Log created interaction terms for verification
print(f"Interaction terms created: {list(interaction_terms.keys())}")

# Display the first few rows of the training data to verify
print("\nTraining Data with New Interaction Terms:")
train_df.head()
```

```
Interaction terms created: ['CGAS-CGAS_Score_x_SDS-SDS_Total_Raw', 'CGAS-
CGAS_Score_x_Physical-BMI', 'CGAS-CGAS_Score_x_Physical-Height', 'CGAS-
CGAS_Score_x_PreInt_EduHx-computerinternet_hoursday', 'SDS-
SDS_Total_Raw_x_Physical-BMI', 'SDS-SDS_Total_Raw_x_Physical-Height', 'SDS-
SDS_Total_Raw_x_PreInt_EduHx-computerinternet_hoursday', 'Physical-
```

```
BMI_x_Physical-Height', 'Physical-BMI_x_PreInt_EduHx-computerinternet_hoursday',
'Physical-Height_x_PreInt_EduHx-computerinternet_hoursday']

Training Data with New Interaction Terms:
```

[22]:

```
          id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0   00008ff9                      Fall              5.0              0.0
2   00105258                    Summer             10.0              1.0
3   00115b9f                    Winter              9.0              0.0
5   001f3379                    Spring             13.0              1.0
11  00abe655                      Fall             11.0              0.0

   CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  \
0       Winter             51.0            Fall     16.877316
2         Fall             71.0            Fall     16.648696
3         Fall             71.0          Summer     18.292347
5       Winter             50.0          Summer     22.279952
11      Summer             66.0             NaN           NaN

    Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0              46.0             50.8                           NaN
2              56.5             75.6                           NaN
3              56.0             81.6                           NaN
5              59.5            112.2                           NaN
11              NaN              NaN                           NaN

    Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
0                     NaN                 NaN                   NaN
2                    65.0                94.0                 117.0
3                    60.0                97.0                 117.0
5                    60.0                73.0                 102.0
11                    NaN                 NaN                   NaN

   Fitness_Endurance-Season  Fitness_Endurance-Max_Stage  \
0                       NaN                          NaN
2                      Fall                          5.0
3                    Summer                          6.0
5                       NaN                          NaN
11                      NaN                          NaN

    Fitness_Endurance-Time_Mins  Fitness_Endurance-Time_Sec FGC-Season  \
0                           NaN                         NaN       Fall
2                           7.0                        33.0       Fall
3                           9.0                        37.0     Summer
5                           NaN                         NaN     Summer
11                          NaN                         NaN     Winter
```

```
   FGC-FGC_CU  FGC-FGC_CU_Zone  FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  \
0         0.0              0.0           NaN                NaN          NaN
2        20.0              1.0          10.2                1.0         14.7
3        18.0              1.0           NaN                NaN          NaN
5        12.0              0.0          16.5                2.0         17.9
11        NaN              NaN           NaN                NaN          NaN

   FGC-FGC_GSD_Zone  FGC-FGC_PU  FGC-FGC_PU_Zone  FGC-FGC_SRL  \
0               NaN         0.0              0.0          7.0
2               2.0         7.0              1.0         10.0
3               NaN         5.0              0.0          7.0
5               2.0         6.0              0.0         10.0
11              NaN         NaN              NaN          NaN

   FGC-FGC_SRL_Zone  FGC-FGC_SRR  FGC-FGC_SRR_Zone  FGC-FGC_TL  \
0               0.0          6.0               0.0         6.0
2               1.0         10.0               1.0         5.0
3               0.0          7.0               0.0         7.0
5               1.0         11.0               1.0         8.0
11              NaN          NaN               NaN         NaN

   FGC-FGC_TL_Zone BIA-Season  BIA-BIA_Activity_Level_num  BIA-BIA_BMC  \
0              1.0       Fall                         2.0      2.66855
2              0.0        NaN                         NaN          NaN
3              1.0     Summer                         3.0      3.84191
5              0.0     Summer                         2.0      4.33036
11             NaN        NaN                         NaN          NaN

   BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0      16.8792      932.498      1492.00      8.25598      41.5862
2          NaN          NaN          NaN          NaN          NaN
3      18.2943     1131.430      1923.44     15.59250      62.7757
5      30.1865     1330.970      1996.45     30.21240      84.0285
11         NaN          NaN          NaN          NaN          NaN

   BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0       13.8177      3.06143      9.21377                1.0      24.4349
2           NaN          NaN          NaN                NaN          NaN
3       14.0740      4.22033     18.82430                2.0      30.4041
5       16.6877     13.49880     67.97150                2.0      32.9141
11          NaN          NaN          NaN                NaN          NaN

   BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW PAQ_A-Season  \
0      8.89536      38.9177      19.5413      32.6909          NaN
2          NaN          NaN          NaN          NaN          NaN
3     16.77900      58.9338      26.4798      45.9966          NaN
5     20.90200      79.6982      35.3804      63.1265          NaN
```

```
11          NaN          NaN          NaN          NaN          NaN

   PAQ_A-PAQ_A_Total PAQ_C-Season  PAQ_C-PAQ_C_Total  PCIAT-PCIAT_Total  \
0                NaN          NaN                NaN               55.0
2                NaN       Summer              2.170               28.0
3                NaN       Winter              2.451               44.0
5                NaN       Spring              4.110               34.0
11               NaN       Winter              1.100               10.0

  SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T PreInt_EduHx-Season  \
0        NaN                NaN              NaN                Fall
2       Fall               38.0             54.0              Summer
3     Summer               31.0             45.0              Winter
5     Summer               40.0             56.0              Spring
11    Winter               42.0             59.0                Fall

   PreInt_EduHx-computerinternet_hoursday  sii Age_Group  PCIAT_Cluster  \
0                                     3.0  2.0       NaN              5
2                                     2.0  0.0      5-10              7
3                                     0.0  1.0      5-10              4
5                                     0.0  1.0     11-15              3
11                                    0.0  0.0     11-15              3

   CGAS-CGAS_Score_x_SDS-SDS_Total_Raw  CGAS-CGAS_Score_x_Physical-BMI  \
0                                  NaN                      860.743100
2                               2698.0                     1182.057420
3                               2201.0                     1298.756633
5                               2000.0                     1113.997599
11                              2772.0                             NaN

   CGAS-CGAS_Score_x_Physical-Height  \
0                             2346.0
2                             4011.5
3                             3976.0
5                             2975.0
11                               NaN

   CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet_hoursday  \
0                                              153.0
2                                              142.0
3                                                0.0
5                                                0.0
11                                               0.0

   SDS-SDS_Total_Raw_x_Physical-BMI  SDS-SDS_Total_Raw_x_Physical-Height  \
0                               NaN                                  NaN
2                        632.650450                               2147.0
```

```
3                        567.062755                              1736.0
5                        891.198079                              2380.0
11                              NaN                                 NaN

    SDS-SDS_Total_Raw_x_PreInt_EduHx-computerinternet_hoursday  \
0                                                  NaN
2                                                 76.0
3                                                  0.0
5                                                  0.0
11                                                 0.0

    Physical-BMI_x_Physical-Height  \
0                       776.356522
2                       940.651327
3                      1024.371429
5                      1325.657143
11                             NaN

    Physical-BMI_x_PreInt_EduHx-computerinternet_hoursday  \
0                                           50.631947
2                                           33.297392
3                                            0.000000
5                                            0.000000
11                                                NaN

    Physical-Height_x_PreInt_EduHx-computerinternet_hoursday
0                                               138.0
2                                               113.0
3                                                 0.0
5                                                 0.0
11                                                NaN
```

[23]:
```
ID_arr = train_df['id']
```

[24]:
```
# Step 2: Compute the correlation matrix for original and interaction terms
# Filter only numeric columns
numeric_cols_with_interactions = train_df.select_dtypes(include=[np.number]).
 ↪columns

# Recompute correlation matrix for numeric columns only
interaction_correlation_matrix = train_df[numeric_cols_with_interactions].corr()

# Filter correlation matrix to show interaction terms only
# Ensure that both rows and columns correspond to numeric columns and available␣
 ↪interaction terms
interaction_correlation = interaction_correlation_matrix.loc[
```

```
    [term for term in interaction_terms.keys() if term in␣
↪numeric_cols_with_interactions],
    [col for col in numeric_cols_with_interactions if col not in␣
↪interaction_terms.keys()]
]

# Display the correlation matrix for interaction terms
#print("\nCorrelation Matrix for Interaction Terms with Original Features:")
#print(interaction_correlation)

# Step 3: Identify and document valuable interaction terms
# Threshold for correlation strength (you can adjust this based on your␣
↪preferences)
strong_correlation_threshold = 0.6
strong_interactions = interaction_correlation[
    interaction_correlation.abs().max(axis=1) >= strong_correlation_threshold
]

#print(f"\nStrong Interaction Terms (Correlation >␣
↪{strong_correlation_threshold}):")
#print(strong_interactions)

# Step 4: Summary of interaction terms
print("\nSummary of All Interaction Terms Created and their Corresponding␣
↪Original Features:")
for term, (f1, f2) in interaction_terms.items():
    print(f"{term}: {f1} x {f2}")

# Visualize correlation heatmap for interaction terms
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(12, 8))
sns.heatmap(interaction_correlation, cmap='coolwarm', annot=True, fmt='.2f',␣
↪cbar=True)
plt.title("Correlation Heatmap for Interaction Terms")
plt.show()
```

```
Correlation Matrix for Interaction Terms with Original Features:
                                                    Basic_Demos-Age  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                       -0.002961
CGAS-CGAS_Score_x_Physical-BMI                             0.426820
CGAS-CGAS_Score_x_Physical-Height                          0.517126
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         0.399461
SDS-SDS_Total_Raw_x_Physical-BMI                           0.321428
SDS-SDS_Total_Raw_x_Physical-Height                        0.396843
```

```
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…          0.358217
Physical-BMI_x_Physical-Height                              0.812803
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          0.502156
Physical-Height_x_PreInt_EduHx-computerinternet…          0.520181

                                                  Basic_Demos-Sex  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.068456
CGAS-CGAS_Score_x_Physical-BMI                              0.104690
CGAS-CGAS_Score_x_Physical-Height                           0.094802
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          0.012564
SDS-SDS_Total_Raw_x_Physical-BMI                           -0.014973
SDS-SDS_Total_Raw_x_Physical-Height                        -0.020312
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         -0.001007
Physical-BMI_x_Physical-Height                              0.002769
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          0.004259
Physical-Height_x_PreInt_EduHx-computerinternet…         -0.000433

                                                  CGAS-CGAS_Score  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.501694
CGAS-CGAS_Score_x_Physical-BMI                              0.698734
CGAS-CGAS_Score_x_Physical-Height                           0.810798
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          0.064666
SDS-SDS_Total_Raw_x_Physical-BMI                           -0.141623
SDS-SDS_Total_Raw_x_Physical-Height                        -0.125426
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         -0.109711
Physical-BMI_x_Physical-Height                             -0.014462
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         -0.093351
Physical-Height_x_PreInt_EduHx-computerinternet…         -0.097153

                                                  Physical-BMI  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                     -0.006988
CGAS-CGAS_Score_x_Physical-BMI                           0.682564
CGAS-CGAS_Score_x_Physical-Height                        0.322519
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…       0.321460
SDS-SDS_Total_Raw_x_Physical-BMI                         0.583503
SDS-SDS_Total_Raw_x_Physical-Height                      0.290182
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…       0.289543
Physical-BMI_x_Physical-Height                           0.927094
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…       0.478218
Physical-Height_x_PreInt_EduHx-computerinternet…       0.388242

                                                  Physical-Height  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.002958
CGAS-CGAS_Score_x_Physical-BMI                              0.434426
CGAS-CGAS_Score_x_Physical-Height                           0.588934
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          0.375376
SDS-SDS_Total_Raw_x_Physical-BMI                            0.329539
SDS-SDS_Total_Raw_x_Physical-Height                         0.447196
```

```
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…        0.333813
Physical-BMI_x_Physical-Height                            0.852275
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…        0.456809
Physical-Height_x_PreInt_EduHx-computerinternet…        0.488695

                                                 Physical-Weight  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                      -0.002054
CGAS-CGAS_Score_x_Physical-BMI                            0.590138
CGAS-CGAS_Score_x_Physical-Height                         0.521268
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…        0.369162
SDS-SDS_Total_Raw_x_Physical-BMI                          0.486093
SDS-SDS_Total_Raw_x_Physical-Height                       0.420271
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…        0.331954
Physical-BMI_x_Physical-Height                            0.984625
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…        0.526976
Physical-Height_x_PreInt_EduHx-computerinternet…        0.503983

                                                 Physical-Waist_Circumference
\
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                                  0.081084
CGAS-CGAS_Score_x_Physical-BMI                                       0.630138
CGAS-CGAS_Score_x_Physical-Height                                    0.492417
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…                   0.361806
SDS-SDS_Total_Raw_x_Physical-BMI                                     0.518215
SDS-SDS_Total_Raw_x_Physical-Height                                  0.390310
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…                   0.308436
Physical-BMI_x_Physical-Height                                       0.885278
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…                   0.466142
Physical-Height_x_PreInt_EduHx-computerinternet…                   0.419725

                                                 Physical-Diastolic_BP  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                            0.007453
CGAS-CGAS_Score_x_Physical-BMI                                 0.092192
CGAS-CGAS_Score_x_Physical-Height                              0.076001
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…             0.065202
SDS-SDS_Total_Raw_x_Physical-BMI                               0.070602
SDS-SDS_Total_Raw_x_Physical-Height                            0.061957
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…             0.076338
Physical-BMI_x_Physical-Height                                 0.129602
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…             0.081731
Physical-Height_x_PreInt_EduHx-computerinternet…             0.080199

                                                 Physical-HeartRate  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.023303
CGAS-CGAS_Score_x_Physical-BMI                             -0.075406
CGAS-CGAS_Score_x_Physical-Height                          -0.135298
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          -0.055713
SDS-SDS_Total_Raw_x_Physical-BMI                           -0.015939
```

```
SDS-SDS_Total_Raw_x_Physical-Height                            -0.049971
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…             -0.031105
Physical-BMI_x_Physical-Height                                 -0.159604
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…             -0.070757
Physical-Height_x_PreInt_EduHx-computerinternet…             -0.081553

                                                   Physical-Systolic_BP  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                            0.005151
CGAS-CGAS_Score_x_Physical-BMI                                 0.186441
CGAS-CGAS_Score_x_Physical-Height                              0.154030
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…             0.123042
SDS-SDS_Total_Raw_x_Physical-BMI                               0.170896
SDS-SDS_Total_Raw_x_Physical-Height                            0.147042
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…             0.136640
Physical-BMI_x_Physical-Height                                 0.307111
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…             0.161507
Physical-Height_x_PreInt_EduHx-computerinternet…             0.154912

                                                   Fitness_Endurance-Max_Stage
\
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                                   0.066838
CGAS-CGAS_Score_x_Physical-BMI                                        0.076601
CGAS-CGAS_Score_x_Physical-Height                                     0.173956
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…                    -0.073570
SDS-SDS_Total_Raw_x_Physical-BMI                                     -0.047335
SDS-SDS_Total_Raw_x_Physical-Height                                   0.005089
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…                    -0.084060
Physical-BMI_x_Physical-Height                                       -0.006644
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…                    -0.094632
Physical-Height_x_PreInt_EduHx-computerinternet…                    -0.085987

                                                   Fitness_Endurance-Time_Mins
\
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                                   0.072055
CGAS-CGAS_Score_x_Physical-BMI                                        0.079370
CGAS-CGAS_Score_x_Physical-Height                                     0.202605
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…                    -0.066590
SDS-SDS_Total_Raw_x_Physical-BMI                                     -0.058577
SDS-SDS_Total_Raw_x_Physical-Height                                   0.008240
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…                    -0.080276
Physical-BMI_x_Physical-Height                                       -0.008894
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…                    -0.093096
Physical-Height_x_PreInt_EduHx-computerinternet…                    -0.081552

                                                   Fitness_Endurance-Time_Sec
\
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                                  0.004665
CGAS-CGAS_Score_x_Physical-BMI                                      -0.049857
```

```
CGAS-CGAS_Score_x_Physical-Height                                -0.075872
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…                0.028723
SDS-SDS_Total_Raw_x_Physical-BMI                                  0.006672
SDS-SDS_Total_Raw_x_Physical-Height                               0.004636
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…                0.032936
Physical-BMI_x_Physical-Height                                   -0.058259
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…                0.014609
Physical-Height_x_PreInt_EduHx-computerinternet…                0.024022

                                                    FGC-FGC_CU  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                   0.086699
CGAS-CGAS_Score_x_Physical-BMI                        0.319985
CGAS-CGAS_Score_x_Physical-Height                     0.457841
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…    0.210201
SDS-SDS_Total_Raw_x_Physical-BMI                      0.106333
SDS-SDS_Total_Raw_x_Physical-Height                   0.197155
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…    0.150474
Physical-BMI_x_Physical-Height                        0.434043
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…    0.223811
Physical-Height_x_PreInt_EduHx-computerinternet…    0.242771

                                                    FGC-FGC_CU_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.093891
CGAS-CGAS_Score_x_Physical-BMI                             0.048883
CGAS-CGAS_Score_x_Physical-Height                          0.105286
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…        -0.063598
SDS-SDS_Total_Raw_x_Physical-BMI                          -0.091276
SDS-SDS_Total_Raw_x_Physical-Height                       -0.068829
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…        -0.102935
Physical-BMI_x_Physical-Height                            -0.108591
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…        -0.101760
Physical-Height_x_PreInt_EduHx-computerinternet…        -0.100419

                                                    FGC-FGC_GSND  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                     0.038287
CGAS-CGAS_Score_x_Physical-BMI                          0.178643
CGAS-CGAS_Score_x_Physical-Height                       0.215274
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…      0.155271
SDS-SDS_Total_Raw_x_Physical-BMI                        0.151527
SDS-SDS_Total_Raw_x_Physical-Height                     0.183294
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…      0.129494
Physical-BMI_x_Physical-Height                          0.415128
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…      0.195205
Physical-Height_x_PreInt_EduHx-computerinternet…      0.207302

                                                    FGC-FGC_GSND_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                          0.022691
CGAS-CGAS_Score_x_Physical-BMI                               0.120266
```

```
CGAS-CGAS_Score_x_Physical-Height                        0.053823
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…        -0.039735
SDS-SDS_Total_Raw_x_Physical-BMI                         0.114199
SDS-SDS_Total_Raw_x_Physical-Height                      0.068889
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…        -0.036851
Physical-BMI_x_Physical-Height                           0.165901
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…        -0.009966
Physical-Height_x_PreInt_EduHx-computerinternet…        -0.028568

                                                    FGC-FGC_GSD  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                    0.032366
CGAS-CGAS_Score_x_Physical-BMI                         0.205805
CGAS-CGAS_Score_x_Physical-Height                      0.269515
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…     0.164096
SDS-SDS_Total_Raw_x_Physical-BMI                       0.160915
SDS-SDS_Total_Raw_x_Physical-Height                    0.205876
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…     0.143590
Physical-BMI_x_Physical-Height                         0.481652
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…     0.205123
Physical-Height_x_PreInt_EduHx-computerinternet…     0.223548

                                                    FGC-FGC_GSD_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.053289
CGAS-CGAS_Score_x_Physical-BMI                              0.125796
CGAS-CGAS_Score_x_Physical-Height                           0.087276
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         -0.012497
SDS-SDS_Total_Raw_x_Physical-BMI                            0.130151
SDS-SDS_Total_Raw_x_Physical-Height                         0.098059
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…          0.005189
Physical-BMI_x_Physical-Height                              0.193049
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          0.017042
Physical-Height_x_PreInt_EduHx-computerinternet…          0.004936

                                                    FGC-FGC_PU  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                   0.058790
CGAS-CGAS_Score_x_Physical-BMI                        0.162705
CGAS-CGAS_Score_x_Physical-Height                     0.295315
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…    0.143622
SDS-SDS_Total_Raw_x_Physical-BMI                      0.036562
SDS-SDS_Total_Raw_x_Physical-Height                   0.115426
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…    0.099104
Physical-BMI_x_Physical-Height                        0.244527
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…    0.148063
Physical-Height_x_PreInt_EduHx-computerinternet…    0.169254

                                                    FGC-FGC_PU_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.033923
CGAS-CGAS_Score_x_Physical-BMI                             0.031640
```

```
CGAS-CGAS_Score_x_Physical-Height                         0.109247
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…        0.026705
SDS-SDS_Total_Raw_x_Physical-BMI                         -0.051034
SDS-SDS_Total_Raw_x_Physical-Height                      -0.010252
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…        0.011559
Physical-BMI_x_Physical-Height                           -0.010971
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…        0.002737
Physical-Height_x_PreInt_EduHx-computerinternet…        0.010531

                                                    FGC-FGC_SRL  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                    0.063995
CGAS-CGAS_Score_x_Physical-BMI                        -0.035160
CGAS-CGAS_Score_x_Physical-Height                     -0.081548
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…    -0.056719
SDS-SDS_Total_Raw_x_Physical-BMI                       0.007240
SDS-SDS_Total_Raw_x_Physical-Height                   -0.034745
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…    -0.051255
Physical-BMI_x_Physical-Height                        -0.129844
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…    -0.061789
Physical-Height_x_PreInt_EduHx-computerinternet…    -0.072242

                                                    FGC-FGC_SRL_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.029534
CGAS-CGAS_Score_x_Physical-BMI                             -0.095677
CGAS-CGAS_Score_x_Physical-Height                          -0.140158
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         -0.115073
SDS-SDS_Total_Raw_x_Physical-BMI                           -0.033957
SDS-SDS_Total_Raw_x_Physical-Height                        -0.076661
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         -0.097824
Physical-BMI_x_Physical-Height                             -0.193163
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         -0.112839
Physical-Height_x_PreInt_EduHx-computerinternet…         -0.126554

                                                    FGC-FGC_SRR  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                    0.065892
CGAS-CGAS_Score_x_Physical-BMI                        -0.038346
CGAS-CGAS_Score_x_Physical-Height                     -0.084542
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…    -0.057747
SDS-SDS_Total_Raw_x_Physical-BMI                       0.009126
SDS-SDS_Total_Raw_x_Physical-Height                   -0.029952
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…    -0.043315
Physical-BMI_x_Physical-Height                        -0.137555
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…    -0.065264
Physical-Height_x_PreInt_EduHx-computerinternet…    -0.074696

                                                    FGC-FGC_SRR_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.045109
CGAS-CGAS_Score_x_Physical-BMI                             -0.081786
```

```
CGAS-CGAS_Score_x_Physical-Height                      -0.140432
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…      -0.109161
SDS-SDS_Total_Raw_x_Physical-BMI                       -0.010492
SDS-SDS_Total_Raw_x_Physical-Height                    -0.065717
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…      -0.082865
Physical-BMI_x_Physical-Height                         -0.196329
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…      -0.108407
Physical-Height_x_PreInt_EduHx-computerinternet…      -0.124577

                                                    FGC-FGC_TL  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                   0.048923
CGAS-CGAS_Score_x_Physical-BMI                        0.174939
CGAS-CGAS_Score_x_Physical-Height                     0.223803
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…    0.074498
SDS-SDS_Total_Raw_x_Physical-BMI                      0.088840
SDS-SDS_Total_Raw_x_Physical-Height                   0.118735
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…    0.059906
Physical-BMI_x_Physical-Height                        0.231766
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…    0.080312
Physical-Height_x_PreInt_EduHx-computerinternet…    0.085294

                                                    FGC-FGC_TL_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.029625
CGAS-CGAS_Score_x_Physical-BMI                            -0.031150
CGAS-CGAS_Score_x_Physical-Height                         -0.018978
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…        -0.097231
SDS-SDS_Total_Raw_x_Physical-BMI                          -0.058246
SDS-SDS_Total_Raw_x_Physical-Height                       -0.049989
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…        -0.091286
Physical-BMI_x_Physical-Height                            -0.113482
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…        -0.106096
Physical-Height_x_PreInt_EduHx-computerinternet…        -0.108350

                                                    BIA-BIA_Activity_Level_num
\
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                                   0.052449
CGAS-CGAS_Score_x_Physical-BMI                                        0.080042
CGAS-CGAS_Score_x_Physical-Height                                     0.158901
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…                    0.008693
SDS-SDS_Total_Raw_x_Physical-BMI                                     -0.013068
SDS-SDS_Total_Raw_x_Physical-Height                                   0.030586
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…                   -0.014595
Physical-BMI_x_Physical-Height                                        0.076639
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…                   -0.006906
Physical-Height_x_PreInt_EduHx-computerinternet…                    0.008085

                                                    BIA-BIA_BMC  BIA-BIA_BMI  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                   -0.031410    -0.009406
```

CGAS-CGAS_Score_x_Physical-BMI	-0.028329	0.624803
CGAS-CGAS_Score_x_Physical-Height	-0.028320	0.303917
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.019029	0.299650
SDS-SDS_Total_Raw_x_Physical-BMI	-0.011892	0.558092
SDS-SDS_Total_Raw_x_Physical-Height	-0.011226	0.289430
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.017671	0.287149
Physical-BMI_x_Physical-Height	-0.004881	0.858340
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.015869	0.452796
Physical-Height_x_PreInt_EduHx-computerinternet…	-0.017421	0.370692

	BIA-BIA_BMR	BIA-BIA_DEE \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031384	-0.029666
CGAS-CGAS_Score_x_Physical-BMI	0.020006	0.030761
CGAS-CGAS_Score_x_Physical-Height	0.024129	0.042406
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.017507	0.023216
SDS-SDS_Total_Raw_x_Physical-BMI	0.022991	0.025830
SDS-SDS_Total_Raw_x_Physical-Height	0.025684	0.032590
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.011251	0.014003
Physical-BMI_x_Physical-Height	0.082755	0.100651
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.030736	0.036612
Physical-Height_x_PreInt_EduHx-computerinternet…	0.029681	0.037130

	BIA-BIA_ECW	BIA-BIA_FFM \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.030314	-0.031384
CGAS-CGAS_Score_x_Physical-BMI	0.024477	0.020006
CGAS-CGAS_Score_x_Physical-Height	0.027744	0.024129
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.017596	0.017507
SDS-SDS_Total_Raw_x_Physical-BMI	0.023461	0.022991
SDS-SDS_Total_Raw_x_Physical-Height	0.025867	0.025684
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.010116	0.011251
Physical-BMI_x_Physical-Height	0.085760	0.082755
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.030890	0.030736
Physical-Height_x_PreInt_EduHx-computerinternet…	0.029718	0.029681

	BIA-BIA_FFMI	BIA-BIA_FMI \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.004299	-0.000335
CGAS-CGAS_Score_x_Physical-BMI	0.130788	0.199019
CGAS-CGAS_Score_x_Physical-Height	0.068485	0.092076
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.104862	0.054765
SDS-SDS_Total_Raw_x_Physical-BMI	0.132304	0.140261
SDS-SDS_Total_Raw_x_Physical-Height	0.079138	0.062414
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.111270	0.029720
Physical-BMI_x_Physical-Height	0.211643	0.242334
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.149890	0.090665
Physical-Height_x_PreInt_EduHx-computerinternet…	0.129067	0.068074

	BIA-BIA_Fat \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.030622

CGAS-CGAS_Score_x_Physical-BMI	0.065707
CGAS-CGAS_Score_x_Physical-Height	0.052378
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.038821
SDS-SDS_Total_Raw_x_Physical-BMI	0.038831
SDS-SDS_Total_Raw_x_Physical-Height	0.027990
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.032744
Physical-BMI_x_Physical-Height	0.059328
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.044521
Physical-Height_x_PreInt_EduHx-computerinternet…	0.042583

	BIA-BIA_Frame_num \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.032790
CGAS-CGAS_Score_x_Physical-BMI	0.377363
CGAS-CGAS_Score_x_Physical-Height	0.235829
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.173000
SDS-SDS_Total_Raw_x_Physical-BMI	0.296638
SDS-SDS_Total_Raw_x_Physical-Height	0.178879
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.147006
Physical-BMI_x_Physical-Height	0.556695
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.258835
Physical-Height_x_PreInt_EduHx-computerinternet…	0.220168

	BIA-BIA_ICW	BIA-BIA_LDM \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031852	-0.032036
CGAS-CGAS_Score_x_Physical-BMI	0.032126	0.005634
CGAS-CGAS_Score_x_Physical-Height	0.035178	0.011488
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.027492	0.009413
SDS-SDS_Total_Raw_x_Physical-BMI	0.035030	0.012855
SDS-SDS_Total_Raw_x_Physical-Height	0.036293	0.016978
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.020240	0.005255
Physical-BMI_x_Physical-Height	0.107332	0.059791
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.044387	0.019616
Physical-Height_x_PreInt_EduHx-computerinternet…	0.042921	0.019013

	BIA-BIA_LST	BIA-BIA_SMM \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031056	-0.030968
CGAS-CGAS_Score_x_Physical-BMI	0.061538	0.034665
CGAS-CGAS_Score_x_Physical-Height	0.069168	0.036409
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.048957	0.030622
SDS-SDS_Total_Raw_x_Physical-BMI	0.052994	0.036399
SDS-SDS_Total_Raw_x_Physical-Height	0.057418	0.036525
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.036266	0.024202
Physical-BMI_x_Physical-Height	0.157541	0.108784
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.070768	0.048510
Physical-Height_x_PreInt_EduHx-computerinternet…	0.070154	0.046133

	BIA-BIA_TBW \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031009

```
CGAS-CGAS_Score_x_Physical-BMI                          0.027820
CGAS-CGAS_Score_x_Physical-Height                       0.030998
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…       0.021906
SDS-SDS_Total_Raw_x_Physical-BMI                        0.028501
SDS-SDS_Total_Raw_x_Physical-Height                     0.030413
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…       0.014514
Physical-BMI_x_Physical-Height                          0.095206
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…       0.036775
Physical-Height_x_PreInt_EduHx-computerinternet…       0.035474

                                                  PAQ_A-PAQ_A_Total  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                       -0.060480
CGAS-CGAS_Score_x_Physical-BMI                            -0.009245
CGAS-CGAS_Score_x_Physical-Height                          0.075187
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         -0.057125
SDS-SDS_Total_Raw_x_Physical-BMI                          -0.150307
SDS-SDS_Total_Raw_x_Physical-Height                       -0.133892
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         -0.138913
Physical-BMI_x_Physical-Height                            -0.026427
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         -0.107023
Physical-Height_x_PreInt_EduHx-computerinternet…         -0.084448

                                                  PAQ_C-PAQ_C_Total  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                       -0.027017
CGAS-CGAS_Score_x_Physical-BMI                             0.004422
CGAS-CGAS_Score_x_Physical-Height                          0.020603
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         -0.071759
SDS-SDS_Total_Raw_x_Physical-BMI                          -0.016506
SDS-SDS_Total_Raw_x_Physical-Height                       -0.015271
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         -0.071524
Physical-BMI_x_Physical-Height                             0.011947
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         -0.064380
Physical-Height_x_PreInt_EduHx-computerinternet…         -0.076414

                                                  PCIAT-PCIAT_Total  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.144338
CGAS-CGAS_Score_x_Physical-BMI                             0.142928
CGAS-CGAS_Score_x_Physical-Height                          0.171662
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          0.353586
SDS-SDS_Total_Raw_x_Physical-BMI                           0.345330
SDS-SDS_Total_Raw_x_Physical-Height                        0.389279
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…          0.385682
Physical-BMI_x_Physical-Height                             0.366443
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          0.374872
Physical-Height_x_PreInt_EduHx-computerinternet…          0.389992

                                                  SDS-SDS_Total_Raw  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.773868
```

```
CGAS-CGAS_Score_x_Physical-BMI                              -0.082695
CGAS-CGAS_Score_x_Physical-Height                           -0.112729
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…           0.057740
SDS-SDS_Total_Raw_x_Physical-BMI                             0.818866
SDS-SDS_Total_Raw_x_Physical-Height                          0.890367
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…           0.290390
Physical-BMI_x_Physical-Height                               0.023975
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…           0.087531
Physical-Height_x_PreInt_EduHx-computerinternet…           0.089659

                                                   SDS-SDS_Total_T  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.769768
CGAS-CGAS_Score_x_Physical-BMI                            -0.083618
CGAS-CGAS_Score_x_Physical-Height                         -0.113016
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         0.059738
SDS-SDS_Total_Raw_x_Physical-BMI                           0.817832
SDS-SDS_Total_Raw_x_Physical-Height                        0.890481
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         0.294082
Physical-BMI_x_Physical-Height                             0.025761
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         0.090256
Physical-Height_x_PreInt_EduHx-computerinternet…         0.092612

                                                   PreInt_EduHx-
computerinternet_hoursday  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw
0.020153
CGAS-CGAS_Score_x_Physical-BMI
0.159345
CGAS-CGAS_Score_x_Physical-Height
0.136174
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…
0.971988
SDS-SDS_Total_Raw_x_Physical-BMI
0.242763
SDS-SDS_Total_Raw_x_Physical-Height
0.242915
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…
0.949283
Physical-BMI_x_Physical-Height
0.392167
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…
0.972717
Physical-Height_x_PreInt_EduHx-computerinternet…
0.985726

                                                        sii  PCIAT_Cluster
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                0.127202       0.037397
CGAS-CGAS_Score_x_Physical-BMI                     0.131911       0.091335
```

```
CGAS-CGAS_Score_x_Physical-Height                       0.143135       0.109724
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…      0.326892       0.177675
SDS-SDS_Total_Raw_x_Physical-BMI                        0.335201       0.132452
SDS-SDS_Total_Raw_x_Physical-Height                     0.366807       0.147515
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…      0.352190       0.179416
Physical-BMI_x_Physical-Height                          0.359416       0.197982
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…      0.354085       0.198325
Physical-Height_x_PreInt_EduHx-computerinternet…      0.365949       0.207116

Strong Interaction Terms (Correlation > 0.6):
                                                        Basic_Demos-Age  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                           -0.002961
CGAS-CGAS_Score_x_Physical-BMI                                 0.426820
CGAS-CGAS_Score_x_Physical-Height                              0.517126
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…             0.399461
SDS-SDS_Total_Raw_x_Physical-BMI                               0.321428
SDS-SDS_Total_Raw_x_Physical-Height                            0.396843
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…             0.358217
Physical-BMI_x_Physical-Height                                 0.812803
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…             0.502156
Physical-Height_x_PreInt_EduHx-computerinternet…             0.520181

                                                        Basic_Demos-Sex  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                            0.068456
CGAS-CGAS_Score_x_Physical-BMI                                 0.104690
CGAS-CGAS_Score_x_Physical-Height                              0.094802
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…             0.012564
SDS-SDS_Total_Raw_x_Physical-BMI                              -0.014973
SDS-SDS_Total_Raw_x_Physical-Height                           -0.020312
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…            -0.001007
Physical-BMI_x_Physical-Height                                 0.002769
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…             0.004259
Physical-Height_x_PreInt_EduHx-computerinternet…            -0.000433

                                                        CGAS-CGAS_Score  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                            0.501694
CGAS-CGAS_Score_x_Physical-BMI                                 0.698734
CGAS-CGAS_Score_x_Physical-Height                              0.810798
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…             0.064666
SDS-SDS_Total_Raw_x_Physical-BMI                              -0.141623
SDS-SDS_Total_Raw_x_Physical-Height                           -0.125426
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…            -0.109711
Physical-BMI_x_Physical-Height                                -0.014462
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…            -0.093351
Physical-Height_x_PreInt_EduHx-computerinternet…            -0.097153

                                                        Physical-BMI  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         -0.006988
```

```
CGAS-CGAS_Score_x_Physical-BMI                        0.682564
CGAS-CGAS_Score_x_Physical-Height                     0.322519
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…     0.321460
SDS-SDS_Total_Raw_x_Physical-BMI                      0.583503
SDS-SDS_Total_Raw_x_Physical-Height                   0.290182
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…     0.289543
Physical-BMI_x_Physical-Height                        0.927094
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…     0.478218
Physical-Height_x_PreInt_EduHx-computerinternet…     0.388242

                                                    Physical-Height  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                          0.002958
CGAS-CGAS_Score_x_Physical-BMI                               0.434426
CGAS-CGAS_Score_x_Physical-Height                            0.588934
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…            0.375376
SDS-SDS_Total_Raw_x_Physical-BMI                             0.329539
SDS-SDS_Total_Raw_x_Physical-Height                          0.447196
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…            0.333813
Physical-BMI_x_Physical-Height                               0.852275
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…            0.456809
Physical-Height_x_PreInt_EduHx-computerinternet…            0.488695

                                                    Physical-Weight  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         -0.002054
CGAS-CGAS_Score_x_Physical-BMI                               0.590138
CGAS-CGAS_Score_x_Physical-Height                            0.521268
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…            0.369162
SDS-SDS_Total_Raw_x_Physical-BMI                             0.486093
SDS-SDS_Total_Raw_x_Physical-Height                          0.420271
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…            0.331954
Physical-BMI_x_Physical-Height                               0.984625
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…            0.526976
Physical-Height_x_PreInt_EduHx-computerinternet…            0.503983

                                                    Physical-Waist_Circumference
\
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                                      0.081084
CGAS-CGAS_Score_x_Physical-BMI                                           0.630138
CGAS-CGAS_Score_x_Physical-Height                                        0.492417
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…                        0.361806
SDS-SDS_Total_Raw_x_Physical-BMI                                         0.518215
SDS-SDS_Total_Raw_x_Physical-Height                                      0.390310
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…                        0.308436
Physical-BMI_x_Physical-Height                                           0.885278
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…                        0.466142
Physical-Height_x_PreInt_EduHx-computerinternet…                        0.419725

                                                    Physical-Diastolic_BP  \
```

```
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.007453
CGAS-CGAS_Score_x_Physical-BMI                             0.092192
CGAS-CGAS_Score_x_Physical-Height                          0.076001
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          0.065202
SDS-SDS_Total_Raw_x_Physical-BMI                           0.070602
SDS-SDS_Total_Raw_x_Physical-Height                        0.061957
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…          0.076338
Physical-BMI_x_Physical-Height                             0.129602
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          0.081731
Physical-Height_x_PreInt_EduHx-computerinternet…          0.080199

                                                    Physical-HeartRate  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                          0.023303
CGAS-CGAS_Score_x_Physical-BMI                              -0.075406
CGAS-CGAS_Score_x_Physical-Height                           -0.135298
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          -0.055713
SDS-SDS_Total_Raw_x_Physical-BMI                            -0.015939
SDS-SDS_Total_Raw_x_Physical-Height                         -0.049971
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…          -0.031105
Physical-BMI_x_Physical-Height                              -0.159604
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          -0.070757
Physical-Height_x_PreInt_EduHx-computerinternet…          -0.081553

                                                    Physical-Systolic_BP  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                            0.005151
CGAS-CGAS_Score_x_Physical-BMI                                 0.186441
CGAS-CGAS_Score_x_Physical-Height                              0.154030
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…             0.123042
SDS-SDS_Total_Raw_x_Physical-BMI                               0.170896
SDS-SDS_Total_Raw_x_Physical-Height                            0.147042
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…             0.136640
Physical-BMI_x_Physical-Height                                 0.307111
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…             0.161507
Physical-Height_x_PreInt_EduHx-computerinternet…             0.154912

                                                    Fitness_Endurance-Max_Stage
\
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                                    0.066838
CGAS-CGAS_Score_x_Physical-BMI                                         0.076601
CGAS-CGAS_Score_x_Physical-Height                                      0.173956
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…                     -0.073570
SDS-SDS_Total_Raw_x_Physical-BMI                                      -0.047335
SDS-SDS_Total_Raw_x_Physical-Height                                    0.005089
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…                     -0.084060
Physical-BMI_x_Physical-Height                                        -0.006644
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…                     -0.094632
Physical-Height_x_PreInt_EduHx-computerinternet…                     -0.085987
```

	Fitness_Endurance-Time_Mins
\	
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.072055
CGAS-CGAS_Score_x_Physical-BMI	0.079370
CGAS-CGAS_Score_x_Physical-Height	0.202605
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.066590
SDS-SDS_Total_Raw_x_Physical-BMI	-0.058577
SDS-SDS_Total_Raw_x_Physical-Height	0.008240
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.080276
Physical-BMI_x_Physical-Height	-0.008894
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.093096
Physical-Height_x_PreInt_EduHx-computerinternet…	-0.081552

	Fitness_Endurance-Time_Sec
\	
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.004665
CGAS-CGAS_Score_x_Physical-BMI	-0.049857
CGAS-CGAS_Score_x_Physical-Height	-0.075872
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.028723
SDS-SDS_Total_Raw_x_Physical-BMI	0.006672
SDS-SDS_Total_Raw_x_Physical-Height	0.004636
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.032936
Physical-BMI_x_Physical-Height	-0.058259
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.014609
Physical-Height_x_PreInt_EduHx-computerinternet…	0.024022

	FGC-FGC_CU \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.086699
CGAS-CGAS_Score_x_Physical-BMI	0.319985
CGAS-CGAS_Score_x_Physical-Height	0.457841
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.210201
SDS-SDS_Total_Raw_x_Physical-BMI	0.106333
SDS-SDS_Total_Raw_x_Physical-Height	0.197155
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.150474
Physical-BMI_x_Physical-Height	0.434043
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.223811
Physical-Height_x_PreInt_EduHx-computerinternet…	0.242771

	FGC-FGC_CU_Zone \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.093891
CGAS-CGAS_Score_x_Physical-BMI	0.048883
CGAS-CGAS_Score_x_Physical-Height	0.105286
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.063598
SDS-SDS_Total_Raw_x_Physical-BMI	-0.091276
SDS-SDS_Total_Raw_x_Physical-Height	-0.068829
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.102935
Physical-BMI_x_Physical-Height	-0.108591
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.101760

```
Physical-Height_x_PreInt_EduHx-computerinternet…        -0.100419

                                                    FGC-FGC_GSND  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                     0.038287
CGAS-CGAS_Score_x_Physical-BMI                          0.178643
CGAS-CGAS_Score_x_Physical-Height                       0.215274
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…      0.155271
SDS-SDS_Total_Raw_x_Physical-BMI                        0.151527
SDS-SDS_Total_Raw_x_Physical-Height                     0.183294
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…      0.129494
Physical-BMI_x_Physical-Height                          0.415128
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…      0.195205
Physical-Height_x_PreInt_EduHx-computerinternet…      0.207302

                                                    FGC-FGC_GSND_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                          0.022691
CGAS-CGAS_Score_x_Physical-BMI                               0.120266
CGAS-CGAS_Score_x_Physical-Height                            0.053823
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…          -0.039735
SDS-SDS_Total_Raw_x_Physical-BMI                             0.114199
SDS-SDS_Total_Raw_x_Physical-Height                          0.068889
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…          -0.036851
Physical-BMI_x_Physical-Height                               0.165901
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          -0.009966
Physical-Height_x_PreInt_EduHx-computerinternet…          -0.028568

                                                    FGC-FGC_GSD  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                    0.032366
CGAS-CGAS_Score_x_Physical-BMI                         0.205805
CGAS-CGAS_Score_x_Physical-Height                      0.269515
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…     0.164096
SDS-SDS_Total_Raw_x_Physical-BMI                       0.160915
SDS-SDS_Total_Raw_x_Physical-Height                    0.205876
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…     0.143590
Physical-BMI_x_Physical-Height                         0.481652
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…     0.205123
Physical-Height_x_PreInt_EduHx-computerinternet…     0.223548

                                                    FGC-FGC_GSD_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.053289
CGAS-CGAS_Score_x_Physical-BMI                              0.125796
CGAS-CGAS_Score_x_Physical-Height                           0.087276
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         -0.012497
SDS-SDS_Total_Raw_x_Physical-BMI                            0.130151
SDS-SDS_Total_Raw_x_Physical-Height                         0.098059
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…          0.005189
Physical-BMI_x_Physical-Height                              0.193049
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…          0.017042
```

```
Physical-Height_x_PreInt_EduHx-computerinternet…          0.004936

                                                    FGC-FGC_PU  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                   0.058790
CGAS-CGAS_Score_x_Physical-BMI                        0.162705
CGAS-CGAS_Score_x_Physical-Height                     0.295315
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…    0.143622
SDS-SDS_Total_Raw_x_Physical-BMI                      0.036562
SDS-SDS_Total_Raw_x_Physical-Height                   0.115426
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…    0.099104
Physical-BMI_x_Physical-Height                        0.244527
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…    0.148063
Physical-Height_x_PreInt_EduHx-computerinternet…    0.169254

                                                    FGC-FGC_PU_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.033923
CGAS-CGAS_Score_x_Physical-BMI                             0.031640
CGAS-CGAS_Score_x_Physical-Height                          0.109247
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         0.026705
SDS-SDS_Total_Raw_x_Physical-BMI                          -0.051034
SDS-SDS_Total_Raw_x_Physical-Height                       -0.010252
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         0.011559
Physical-BMI_x_Physical-Height                            -0.010971
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         0.002737
Physical-Height_x_PreInt_EduHx-computerinternet…         0.010531

                                                    FGC-FGC_SRL  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                    0.063995
CGAS-CGAS_Score_x_Physical-BMI                        -0.035160
CGAS-CGAS_Score_x_Physical-Height                     -0.081548
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…    -0.056719
SDS-SDS_Total_Raw_x_Physical-BMI                       0.007240
SDS-SDS_Total_Raw_x_Physical-Height                   -0.034745
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…    -0.051255
Physical-BMI_x_Physical-Height                        -0.129844
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…    -0.061789
Physical-Height_x_PreInt_EduHx-computerinternet…    -0.072242

                                                    FGC-FGC_SRL_Zone  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                         0.029534
CGAS-CGAS_Score_x_Physical-BMI                             -0.095677
CGAS-CGAS_Score_x_Physical-Height                          -0.140158
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         -0.115073
SDS-SDS_Total_Raw_x_Physical-BMI                           -0.033957
SDS-SDS_Total_Raw_x_Physical-Height                        -0.076661
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         -0.097824
Physical-BMI_x_Physical-Height                             -0.193163
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         -0.112839
```

Physical-Height_x_PreInt_EduHx-computerinternet…	-0.126554

	FGC-FGC_SRR \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.065892
CGAS-CGAS_Score_x_Physical-BMI	-0.038346
CGAS-CGAS_Score_x_Physical-Height	-0.084542
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.057747
SDS-SDS_Total_Raw_x_Physical-BMI	0.009126
SDS-SDS_Total_Raw_x_Physical-Height	-0.029952
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.043315
Physical-BMI_x_Physical-Height	-0.137555
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.065264
Physical-Height_x_PreInt_EduHx-computerinternet…	-0.074696

	FGC-FGC_SRR_Zone \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.045109
CGAS-CGAS_Score_x_Physical-BMI	-0.081786
CGAS-CGAS_Score_x_Physical-Height	-0.140432
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.109161
SDS-SDS_Total_Raw_x_Physical-BMI	-0.010492
SDS-SDS_Total_Raw_x_Physical-Height	-0.065717
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.082865
Physical-BMI_x_Physical-Height	-0.196329
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.108407
Physical-Height_x_PreInt_EduHx-computerinternet…	-0.124577

	FGC-FGC_TL \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.048923
CGAS-CGAS_Score_x_Physical-BMI	0.174939
CGAS-CGAS_Score_x_Physical-Height	0.223803
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.074498
SDS-SDS_Total_Raw_x_Physical-BMI	0.088840
SDS-SDS_Total_Raw_x_Physical-Height	0.118735
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.059906
Physical-BMI_x_Physical-Height	0.231766
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.080312
Physical-Height_x_PreInt_EduHx-computerinternet…	0.085294

	FGC-FGC_TL_Zone \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.029625
CGAS-CGAS_Score_x_Physical-BMI	-0.031150
CGAS-CGAS_Score_x_Physical-Height	-0.018978
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.097231
SDS-SDS_Total_Raw_x_Physical-BMI	-0.058246
SDS-SDS_Total_Raw_x_Physical-Height	-0.049989
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.091286
Physical-BMI_x_Physical-Height	-0.113482
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.106096

Physical-Height_x_PreInt_EduHx-computerinternet…	-0.108350

	BIA-BIA_Activity_Level_num \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.052449
CGAS-CGAS_Score_x_Physical-BMI	0.080042
CGAS-CGAS_Score_x_Physical-Height	0.158901
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.008693
SDS-SDS_Total_Raw_x_Physical-BMI	-0.013068
SDS-SDS_Total_Raw_x_Physical-Height	0.030586
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.014595
Physical-BMI_x_Physical-Height	0.076639
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.006906
Physical-Height_x_PreInt_EduHx-computerinternet…	0.008085

	BIA-BIA_BMC	BIA-BIA_BMI \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031410	-0.009406
CGAS-CGAS_Score_x_Physical-BMI	-0.028329	0.624803
CGAS-CGAS_Score_x_Physical-Height	-0.028320	0.303917
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.019029	0.299650
SDS-SDS_Total_Raw_x_Physical-BMI	-0.011892	0.558092
SDS-SDS_Total_Raw_x_Physical-Height	-0.011226	0.289430
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.017671	0.287149
Physical-BMI_x_Physical-Height	-0.004881	0.858340
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.015869	0.452796
Physical-Height_x_PreInt_EduHx-computerinternet…	-0.017421	0.370692

	BIA-BIA_BMR	BIA-BIA_DEE \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031384	-0.029666
CGAS-CGAS_Score_x_Physical-BMI	0.020006	0.030761
CGAS-CGAS_Score_x_Physical-Height	0.024129	0.042406
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.017507	0.023216
SDS-SDS_Total_Raw_x_Physical-BMI	0.022991	0.025830
SDS-SDS_Total_Raw_x_Physical-Height	0.025684	0.032590
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.011251	0.014003
Physical-BMI_x_Physical-Height	0.082755	0.100651
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.030736	0.036612
Physical-Height_x_PreInt_EduHx-computerinternet…	0.029681	0.037130

	BIA-BIA_ECW	BIA-BIA_FFM \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.030314	-0.031384
CGAS-CGAS_Score_x_Physical-BMI	0.024477	0.020006
CGAS-CGAS_Score_x_Physical-Height	0.027744	0.024129
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.017596	0.017507
SDS-SDS_Total_Raw_x_Physical-BMI	0.023461	0.022991
SDS-SDS_Total_Raw_x_Physical-Height	0.025867	0.025684
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.010116	0.011251
Physical-BMI_x_Physical-Height	0.085760	0.082755

Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.030890	0.030736
Physical-Height_x_PreInt_EduHx-computerinternet…	0.029718	0.029681

	BIA-BIA_FFMI	BIA-BIA_FMI \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.004299	-0.000335
CGAS-CGAS_Score_x_Physical-BMI	0.130788	0.199019
CGAS-CGAS_Score_x_Physical-Height	0.068485	0.092076
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.104862	0.054765
SDS-SDS_Total_Raw_x_Physical-BMI	0.132304	0.140261
SDS-SDS_Total_Raw_x_Physical-Height	0.079138	0.062414
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.111270	0.029720
Physical-BMI_x_Physical-Height	0.211643	0.242334
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.149890	0.090665
Physical-Height_x_PreInt_EduHx-computerinternet…	0.129067	0.068074

	BIA-BIA_Fat \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	0.030622
CGAS-CGAS_Score_x_Physical-BMI	0.065707
CGAS-CGAS_Score_x_Physical-Height	0.052378
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.038821
SDS-SDS_Total_Raw_x_Physical-BMI	0.038831
SDS-SDS_Total_Raw_x_Physical-Height	0.027990
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.032744
Physical-BMI_x_Physical-Height	0.059328
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.044521
Physical-Height_x_PreInt_EduHx-computerinternet…	0.042583

	BIA-BIA_Frame_num \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.032790
CGAS-CGAS_Score_x_Physical-BMI	0.377363
CGAS-CGAS_Score_x_Physical-Height	0.235829
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.173000
SDS-SDS_Total_Raw_x_Physical-BMI	0.296638
SDS-SDS_Total_Raw_x_Physical-Height	0.178879
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.147006
Physical-BMI_x_Physical-Height	0.556695
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.258835
Physical-Height_x_PreInt_EduHx-computerinternet…	0.220168

	BIA-BIA_ICW	BIA-BIA_LDM \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031852	-0.032036
CGAS-CGAS_Score_x_Physical-BMI	0.032126	0.005634
CGAS-CGAS_Score_x_Physical-Height	0.035178	0.011488
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.027492	0.009413
SDS-SDS_Total_Raw_x_Physical-BMI	0.035030	0.012855
SDS-SDS_Total_Raw_x_Physical-Height	0.036293	0.016978
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.020240	0.005255
Physical-BMI_x_Physical-Height	0.107332	0.059791

Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.044387	0.019616
Physical-Height_x_PreInt_EduHx-computerinternet…	0.042921	0.019013

	BIA-BIA_LST	BIA-BIA_SMM \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031056	-0.030968
CGAS-CGAS_Score_x_Physical-BMI	0.061538	0.034665
CGAS-CGAS_Score_x_Physical-Height	0.069168	0.036409
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.048957	0.030622
SDS-SDS_Total_Raw_x_Physical-BMI	0.052994	0.036399
SDS-SDS_Total_Raw_x_Physical-Height	0.057418	0.036525
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.036266	0.024202
Physical-BMI_x_Physical-Height	0.157541	0.108784
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.070768	0.048510
Physical-Height_x_PreInt_EduHx-computerinternet…	0.070154	0.046133

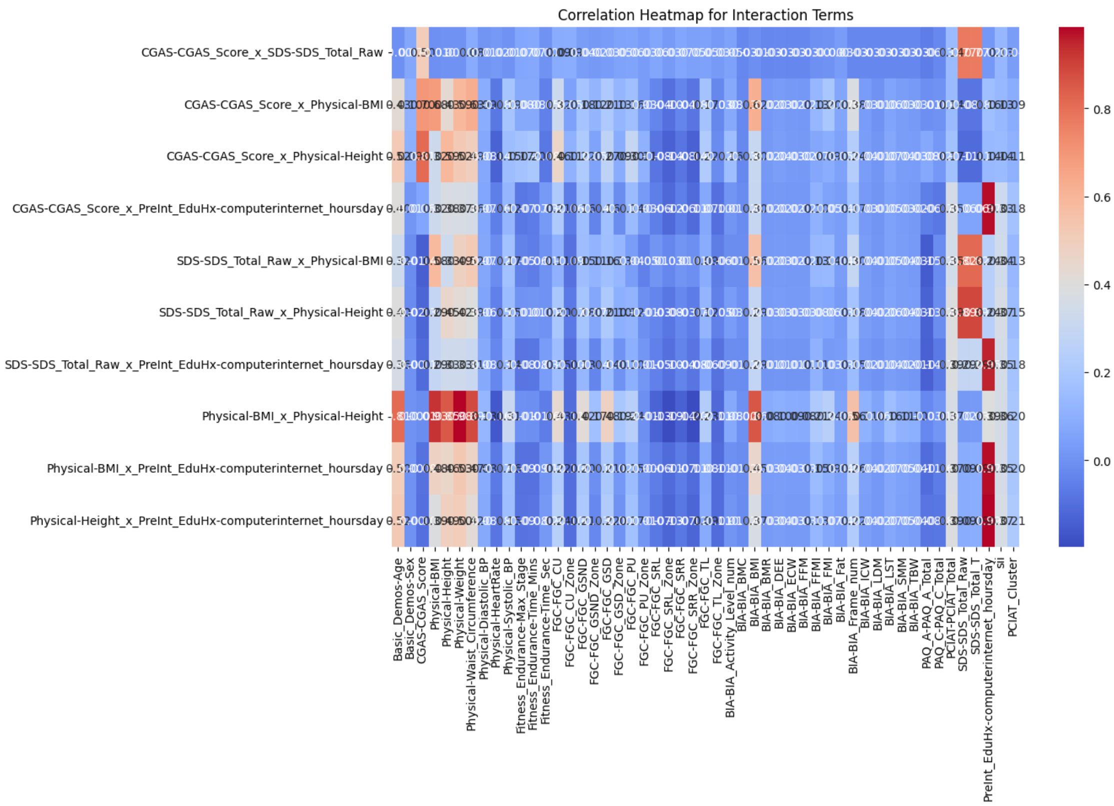
	BIA-BIA_TBW \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.031009
CGAS-CGAS_Score_x_Physical-BMI	0.027820
CGAS-CGAS_Score_x_Physical-Height	0.030998
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	0.021906
SDS-SDS_Total_Raw_x_Physical-BMI	0.028501
SDS-SDS_Total_Raw_x_Physical-Height	0.030413
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	0.014514
Physical-BMI_x_Physical-Height	0.095206
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	0.036775
Physical-Height_x_PreInt_EduHx-computerinternet…	0.035474

	PAQ_A-PAQ_A_Total \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.060480
CGAS-CGAS_Score_x_Physical-BMI	-0.009245
CGAS-CGAS_Score_x_Physical-Height	0.075187
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.057125
SDS-SDS_Total_Raw_x_Physical-BMI	-0.150307
SDS-SDS_Total_Raw_x_Physical-Height	-0.133892
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.138913
Physical-BMI_x_Physical-Height	-0.026427
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…	-0.107023
Physical-Height_x_PreInt_EduHx-computerinternet…	-0.084448

	PAQ_C-PAQ_C_Total \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	-0.027017
CGAS-CGAS_Score_x_Physical-BMI	0.004422
CGAS-CGAS_Score_x_Physical-Height	0.020603
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…	-0.071759
SDS-SDS_Total_Raw_x_Physical-BMI	-0.016506
SDS-SDS_Total_Raw_x_Physical-Height	-0.015271
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…	-0.071524
Physical-BMI_x_Physical-Height	0.011947

```
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         -0.064380
Physical-Height_x_PreInt_EduHx-computerinternet…         -0.076414

                                                    PCIAT-PCIAT_Total  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                          0.144338
CGAS-CGAS_Score_x_Physical-BMI                               0.142928
CGAS-CGAS_Score_x_Physical-Height                            0.171662
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…           0.353586
SDS-SDS_Total_Raw_x_Physical-BMI                             0.345330
SDS-SDS_Total_Raw_x_Physical-Height                          0.389279
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…           0.385682
Physical-BMI_x_Physical-Height                               0.366443
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…           0.374872
Physical-Height_x_PreInt_EduHx-computerinternet…           0.389992

                                                    SDS-SDS_Total_Raw  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                          0.773868
CGAS-CGAS_Score_x_Physical-BMI                              -0.082695
CGAS-CGAS_Score_x_Physical-Height                           -0.112729
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…           0.057740
SDS-SDS_Total_Raw_x_Physical-BMI                             0.818866
SDS-SDS_Total_Raw_x_Physical-Height                          0.890367
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…           0.290390
Physical-BMI_x_Physical-Height                               0.023975
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…           0.087531
Physical-Height_x_PreInt_EduHx-computerinternet…           0.089659

                                                    SDS-SDS_Total_T  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                        0.769768
CGAS-CGAS_Score_x_Physical-BMI                            -0.083618
CGAS-CGAS_Score_x_Physical-Height                         -0.113016
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…         0.059738
SDS-SDS_Total_Raw_x_Physical-BMI                           0.817832
SDS-SDS_Total_Raw_x_Physical-Height                        0.890481
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…         0.294082
Physical-BMI_x_Physical-Height                             0.025761
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…         0.090256
Physical-Height_x_PreInt_EduHx-computerinternet…         0.092612

                                                    PreInt_EduHx-
computerinternet_hoursday  \
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw
0.020153
CGAS-CGAS_Score_x_Physical-BMI
0.159345
CGAS-CGAS_Score_x_Physical-Height
0.136174
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet…
```

```
0.971988
SDS-SDS_Total_Raw_x_Physical-BMI
0.242763
SDS-SDS_Total_Raw_x_Physical-Height
0.242915
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern…
0.949283
Physical-BMI_x_Physical-Height
0.392167
Physical-BMI_x_PreInt_EduHx-computerinternet_ho…
0.972717
Physical-Height_x_PreInt_EduHx-computerinternet…
0.985726

                                                      sii  PCIAT_Cluster
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw              0.127202       0.037397
CGAS-CGAS_Score_x_Physical-BMI                   0.131911       0.091335
CGAS-CGAS_Score_x_Physical-Height                0.143135       0.109724
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet… 0.326892      0.177675
SDS-SDS_Total_Raw_x_Physical-BMI                 0.335201       0.132452
SDS-SDS_Total_Raw_x_Physical-Height              0.366807       0.147515
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerintern… 0.352190      0.179416
Physical-BMI_x_Physical-Height                   0.359416       0.197982
Physical-BMI_x_PreInt_EduHx-computerinternet_ho… 0.354085      0.198325
Physical-Height_x_PreInt_EduHx-computerinternet… 0.365949      0.207116

Summary of All Interaction Terms Created and their Corresponding Original
Features:
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw: CGAS-CGAS_Score x SDS-SDS_Total_Raw
CGAS-CGAS_Score_x_Physical-BMI: CGAS-CGAS_Score x Physical-BMI
CGAS-CGAS_Score_x_Physical-Height: CGAS-CGAS_Score x Physical-Height
CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet_hoursday: CGAS-CGAS_Score x
PreInt_EduHx-computerinternet_hoursday
SDS-SDS_Total_Raw_x_Physical-BMI: SDS-SDS_Total_Raw x Physical-BMI
SDS-SDS_Total_Raw_x_Physical-Height: SDS-SDS_Total_Raw x Physical-Height
SDS-SDS_Total_Raw_x_PreInt_EduHx-computerinternet_hoursday: SDS-SDS_Total_Raw x
PreInt_EduHx-computerinternet_hoursday
Physical-BMI_x_Physical-Height: Physical-BMI x Physical-Height
Physical-BMI_x_PreInt_EduHx-computerinternet_hoursday: Physical-BMI x
PreInt_EduHx-computerinternet_hoursday
Physical-Height_x_PreInt_EduHx-computerinternet_hoursday: Physical-Height x
PreInt_EduHx-computerinternet_hoursday
```



Key Insights High Correlation Among Interaction Terms and Original Features:

Several interaction terms show strong correlations (above 0.6) with original features. For example:

Physical-BMI_x_Physical-Height has a very high correlation with both Physical-BMI (0.927) and Physical-Height (0.852), suggesting that this interaction term might not provide additional value since it's heavily dependent on the two original features. Interaction terms like CGAS-CGAS_Score_x_Physical-Height and SDS-SDS_Total_Raw_x_Physical-Height also show high correlations with their original features, which could imply that these interaction terms may not add unique information to the model.

Relationship of Interaction Terms with Key Outcome Variables:

Interaction terms like CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet_hoursday show moderate correlations with outcome variables like PCIAT-PCIAT_Total (0.354) and sii (0.326). This indicates that some interaction terms do have a significant relationship with key outcomes, but the strength of these correlations is not overwhelmingly strong, suggesting that further refinement is needed. Multicollinearity Concerns:

Interaction terms involving Physical-BMI, Physical-Height, and PreInt_EduHx-computerinternet_hoursday show high correlations with each other (e.g., Physical-BMI_x_Physical-Height has a correlation of 0.812 with Physical-BMI and 0.852 with Physical-Height). This could introduce multicollinearity in the model, which could negatively impact model stability and interpretability. Low Correlation with Certain Outcome Variables:

Several interaction terms show low correlation values (less than 0.2) with the PCIAT-Cluster variable. For instance, CGAS-CGAS_Score_x_Physical-BMI and Physical-Height_x_PreInt_EduHx-computerinternet_hoursday have correlation values around 0.1-0.19 with PCIAT-Cluster. This suggests that these interaction terms may not be particularly useful for predicting PCIAT-Cluster.

[25]:
```
!pip install category_encoders --quiet

from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer
import category_encoders as ce  # for target encoding
from statsmodels.stats.outliers_influence import variance_inflation_factor

# Step 1: Remove redundant interaction terms based on correlation analysis
# Define threshold for high correlation
high_corr_threshold = 0.8
target_columns = ['PCIAT-PCIAT_Total', 'sii', 'PCIAT_Cluster']

# Identify redundant interaction terms
redundant_terms = []
for term, (f1, f2) in interaction_terms.items():
    # Calculate correlation between interaction term and its original features
    corr_with_f1 = train_df[term].corr(train_df[f1])
    corr_with_f2 = train_df[term].corr(train_df[f2])

    # Check if either correlation exceeds the threshold
    if abs(corr_with_f1) >= high_corr_threshold or abs(corr_with_f2) >=␣
 ↪high_corr_threshold:
        redundant_terms.append(term)

# Drop redundant interaction terms and print the count
train_df.drop(columns=redundant_terms, inplace=True)
print(f"Redundant interaction terms removed: {redundant_terms}")
print(f"Remaining columns after removing redundant interaction terms: {train_df.
 ↪columns}")

# Step 2: Impute missing values using MICE
# Use IterativeImputer to perform MICE imputation on numeric columns
numeric_features = train_df.select_dtypes(include=[np.number]).columns
mice_imputer = IterativeImputer(max_iter=10, random_state=42)  # Configure MICE␣
 ↪imputer
imputed_values = mice_imputer.fit_transform(train_df[numeric_features])

# Replace original DataFrame's numeric columns with MICE-imputed values
train_df[numeric_features] = imputed_values
print(f"Remaining columns after MICE imputation: {train_df.columns}")
```

```
# Step 3: Apply Target Encoding to categorical features
# Identify categorical features
categorical_features = train_df.select_dtypes(include=['category', 'object']).
↪columns

# Check if categorical features exist
if len(categorical_features) > 0:
    print(f"Categorical features identified: {categorical_features}")

    # Use target encoding for categorical features based on the target variable␣
↪'sii'
    target_encoder = ce.TargetEncoder(cols=categorical_features)
    # Fit and transform the categorical features
    train_df[categorical_features] = target_encoder.
↪fit_transform(train_df[categorical_features], train_df['sii'])
else:
    print("No categorical features found for encoding.")

# Step 4: Implement VIF analysis to identify multicollinearity issues

# Check for infinite values and replace with NaN
train_df.replace([np.inf, -np.inf], np.nan, inplace=True)

# Create a DataFrame for VIF calculation (exclude non-numeric columns)
vif_features = train_df.select_dtypes(include=[np.number]).columns

# Ensure there are no missing or infinite values in the VIF features
vif_features_cleaned = train_df[vif_features].dropna().columns
print(f"Features to be included in VIF analysis: {vif_features_cleaned}")

# Calculate VIF values
vif_df = pd.DataFrame()
vif_df['Feature'] = vif_features_cleaned
vif_df['VIF'] = [variance_inflation_factor(train_df[vif_features_cleaned].
↪values, i) for i in range(len(vif_features_cleaned))]
print(f"VIF values for features:\n{vif_df}")

# Step 5: Remove features with high VIF (> 10)
high_vif_features = vif_df[vif_df['VIF'] > 10]['Feature'].tolist()
print(f"Features marked for removal due to high VIF: {high_vif_features}")

# Ensure that we don't remove all features
if len(high_vif_features) >= len(train_df.columns) - len(target_columns):
    print("Warning: Attempting to remove too many features. Adjusting VIF␣
↪threshold to retain features.")
    high_vif_features = []  # Do not remove any features if this condition is␣
↪met
```

```
# Remove high VIF features and print remaining columns
train_df.drop(columns=high_vif_features, inplace=True)
print(f"Remaining columns after removing high VIF features: {train_df.columns}")

# Step 6: Create missingness indicator columns
# Define threshold for missingness
missing_threshold = 0.1

# Create missingness indicator columns
for col in train_df.columns:
    missing_ratio = train_df[col].isnull().mean()
    if missing_ratio >= missing_threshold:
        # Create missingness indicator
        missing_indicator_col = f"{col}_missing"
        train_df[missing_indicator_col] = train_df[col].isnull().astype(int)
        print(f"Created missingness indicator for {col} (Missing Ratio:␣
↪{missing_ratio:.2f})")

print(f"Remaining columns after creating missingness indicators: {train_df.
↪columns}")

# Step 7: Reevaluate correlation of selected features with target variables
#target_columns = ['PCIAT-PCIAT_Total', 'sii', 'PCIAT_Cluster']

# Filter only numeric columns before running correlation analysis
numeric_df = train_df.select_dtypes(include=[np.number])

# Ensure that target columns exist in the numeric DataFrame
for target in target_columns:
    if target not in numeric_df.columns:
        print(f"Warning: Target column '{target}' is missing from the DataFrame.
↪ Ensure it is not dropped inadvertently.")

# Calculate the correlation matrix only on numeric columns
interaction_correlation_matrix = numeric_df.corr()

# Filter interaction terms that remain after the previous steps
remaining_interaction_terms = [term for term in interaction_terms.keys() if␣
↪term not in redundant_terms and term not in high_vif_features]

# Calculate correlation with target variables
for target in target_columns:
    if target in numeric_df.columns:
        print(f"\nCorrelation of remaining interaction terms with {target}:")
        for term in remaining_interaction_terms:
            if term in numeric_df.columns:
```

```
            corr_value = numeric_df[term].corr(numeric_df[target])
            print(f"{term}: {corr_value:.4f}")
```

```
Redundant interaction terms removed: ['CGAS-CGAS_Score_x_Physical-Height',
'CGAS-CGAS_Score_x_PreInt_EduHx-computerinternet_hoursday', 'SDS-
SDS_Total_Raw_x_Physical-BMI', 'SDS-SDS_Total_Raw_x_Physical-Height', 'SDS-
SDS_Total_Raw_x_PreInt_EduHx-computerinternet_hoursday', 'Physical-
BMI_x_Physical-Height', 'Physical-BMI_x_PreInt_EduHx-computerinternet_hoursday',
'Physical-Height_x_PreInt_EduHx-computerinternet_hoursday']
Remaining columns after removing redundant interaction terms: Index(['id',
'Basic_Demos-Enroll_Season', 'Basic_Demos-Age', 'Basic_Demos-Sex',
       'CGAS-Season', 'CGAS-CGAS_Score', 'Physical-Season', 'Physical-BMI',
       'Physical-Height', 'Physical-Weight', 'Physical-Waist_Circumference',
       'Physical-Diastolic_BP', 'Physical-HeartRate', 'Physical-Systolic_BP',
       'Fitness_Endurance-Season', 'Fitness_Endurance-Max_Stage',
       'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec',
       'FGC-Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND',
       'FGC-FGC_GSND_Zone', 'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU',
       'FGC-FGC_PU_Zone', 'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR',
       'FGC-FGC_SRR_Zone', 'FGC-FGC_TL', 'FGC-FGC_TL_Zone', 'BIA-Season',
       'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC', 'BIA-BIA_BMI',
       'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM',
       'BIA-BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num',
       'BIA-BIA_ICW', 'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM',
       'BIA-BIA_TBW', 'PAQ_A-Season', 'PAQ_A-PAQ_A_Total', 'PAQ_C-Season',
       'PAQ_C-PAQ_C_Total', 'PCIAT-PCIAT_Total', 'SDS-Season',
       'SDS-SDS_Total_Raw', 'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
       'PreInt_EduHx-computerinternet_hoursday', 'sii', 'Age_Group',
       'PCIAT_Cluster', 'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw',
       'CGAS-CGAS_Score_x_Physical-BMI'],
      dtype='object')
Remaining columns after MICE imputation: Index(['id', 'Basic_Demos-
Enroll_Season', 'Basic_Demos-Age', 'Basic_Demos-Sex',
       'CGAS-Season', 'CGAS-CGAS_Score', 'Physical-Season', 'Physical-BMI',
       'Physical-Height', 'Physical-Weight', 'Physical-Waist_Circumference',
       'Physical-Diastolic_BP', 'Physical-HeartRate', 'Physical-Systolic_BP',
       'Fitness_Endurance-Season', 'Fitness_Endurance-Max_Stage',
       'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec',
       'FGC-Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND',
       'FGC-FGC_GSND_Zone', 'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU',
       'FGC-FGC_PU_Zone', 'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR',
       'FGC-FGC_SRR_Zone', 'FGC-FGC_TL', 'FGC-FGC_TL_Zone', 'BIA-Season',
       'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC', 'BIA-BIA_BMI',
       'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM',
       'BIA-BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num',
       'BIA-BIA_ICW', 'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM',
       'BIA-BIA_TBW', 'PAQ_A-Season', 'PAQ_A-PAQ_A_Total', 'PAQ_C-Season',
       'PAQ_C-PAQ_C_Total', 'PCIAT-PCIAT_Total', 'SDS-Season',
```

```
       'SDS-SDS_Total_Raw', 'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
       'PreInt_EduHx-computerinternet_hoursday', 'sii', 'Age_Group',
       'PCIAT_Cluster', 'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw',
       'CGAS-CGAS_Score_x_Physical-BMI'],
      dtype='object')
Categorical features identified: Index(['id', 'Basic_Demos-Enroll_Season',
'CGAS-Season', 'Physical-Season',
       'Fitness_Endurance-Season', 'FGC-Season', 'BIA-Season', 'PAQ_A-Season',
       'PAQ_C-Season', 'SDS-Season', 'PreInt_EduHx-Season', 'Age_Group'],
      dtype='object')
Features to be included in VIF analysis: Index(['id', 'Basic_Demos-
Enroll_Season', 'Basic_Demos-Age', 'Basic_Demos-Sex',
       'CGAS-Season', 'CGAS-CGAS_Score', 'Physical-Season', 'Physical-BMI',
       'Physical-Height', 'Physical-Weight', 'Physical-Waist_Circumference',
       'Physical-Diastolic_BP', 'Physical-HeartRate', 'Physical-Systolic_BP',
       'Fitness_Endurance-Season', 'Fitness_Endurance-Max_Stage',
       'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec',
       'FGC-Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND',
       'FGC-FGC_GSND_Zone', 'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU',
       'FGC-FGC_PU_Zone', 'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR',
       'FGC-FGC_SRR_Zone', 'FGC-FGC_TL', 'FGC-FGC_TL_Zone', 'BIA-Season',
       'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC', 'BIA-BIA_BMI',
       'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM',
       'BIA-BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num',
       'BIA-BIA_ICW', 'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM',
       'BIA-BIA_TBW', 'PAQ_A-Season', 'PAQ_A-PAQ_A_Total', 'PAQ_C-Season',
       'PAQ_C-PAQ_C_Total', 'PCIAT-PCIAT_Total', 'SDS-Season',
       'SDS-SDS_Total_Raw', 'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
       'PreInt_EduHx-computerinternet_hoursday', 'sii', 'Age_Group',
       'PCIAT_Cluster', 'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw',
       'CGAS-CGAS_Score_x_Physical-BMI'],
      dtype='object')
VIF values for features:
                                Feature           VIF
0                                    id  3.316759e+02
1             Basic_Demos-Enroll_Season  1.022143e+03
2                       Basic_Demos-Age  1.780492e+07
3                       Basic_Demos-Sex  6.257103e+03
4                           CGAS-Season  6.162885e+02
..                                  …             …
60                                  sii  5.279081e+04
61                            Age_Group  2.949623e+01
62                        PCIAT_Cluster  5.652697e+05
63  CGAS-CGAS_Score_x_SDS-SDS_Total_Raw  4.602746e+07
64       CGAS-CGAS_Score_x_Physical-BMI  1.825490e+08

[65 rows x 2 columns]
Features marked for removal due to high VIF: ['id', 'Basic_Demos-Enroll_Season',
```

```
'Basic_Demos-Age', 'Basic_Demos-Sex', 'CGAS-Season', 'CGAS-CGAS_Score',
'Physical-Season', 'Physical-BMI', 'Physical-Height', 'Physical-Weight',
'Physical-Waist_Circumference', 'Physical-Diastolic_BP', 'Physical-HeartRate',
'Physical-Systolic_BP', 'Fitness_Endurance-Season', 'Fitness_Endurance-
Max_Stage', 'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec', 'FGC-
Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND', 'FGC-FGC_GSND_Zone',
'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU', 'FGC-FGC_PU_Zone', 'FGC-
FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR', 'FGC-FGC_SRR_Zone', 'FGC-FGC_TL',
'FGC-FGC_TL_Zone', 'BIA-Season', 'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC',
'BIA-BIA_BMI', 'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM', 'BIA-
BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num', 'BIA-BIA_ICW',
'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM', 'BIA-BIA_TBW', 'PAQ_A-Season',
'PAQ_A-PAQ_A_Total', 'PAQ_C-Season', 'PAQ_C-PAQ_C_Total', 'PCIAT-PCIAT_Total',
'SDS-Season', 'SDS-SDS_Total_Raw', 'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
'PreInt_EduHx-computerinternet_hoursday', 'sii', 'Age_Group', 'PCIAT_Cluster',
'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw', 'CGAS-CGAS_Score_x_Physical-BMI']
Warning: Attempting to remove too many features. Adjusting VIF threshold to
retain features.
Remaining columns after removing high VIF features: Index(['id', 'Basic_Demos-
Enroll_Season', 'Basic_Demos-Age', 'Basic_Demos-Sex',
       'CGAS-Season', 'CGAS-CGAS_Score', 'Physical-Season', 'Physical-BMI',
       'Physical-Height', 'Physical-Weight', 'Physical-Waist_Circumference',
       'Physical-Diastolic_BP', 'Physical-HeartRate', 'Physical-Systolic_BP',
       'Fitness_Endurance-Season', 'Fitness_Endurance-Max_Stage',
       'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec',
       'FGC-Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND',
       'FGC-FGC_GSND_Zone', 'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU',
       'FGC-FGC_PU_Zone', 'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR',
       'FGC-FGC_SRR_Zone', 'FGC-FGC_TL', 'FGC-FGC_TL_Zone', 'BIA-Season',
       'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC', 'BIA-BIA_BMI',
       'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM',
       'BIA-BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num',
       'BIA-BIA_ICW', 'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM',
       'BIA-BIA_TBW', 'PAQ_A-Season', 'PAQ_A-PAQ_A_Total', 'PAQ_C-Season',
       'PAQ_C-PAQ_C_Total', 'PCIAT-PCIAT_Total', 'SDS-Season',
       'SDS-SDS_Total_Raw', 'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
       'PreInt_EduHx-computerinternet_hoursday', 'sii', 'Age_Group',
       'PCIAT_Cluster', 'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw',
       'CGAS-CGAS_Score_x_Physical-BMI'],
      dtype='object')
Remaining columns after creating missingness indicators: Index(['id',
'Basic_Demos-Enroll_Season', 'Basic_Demos-Age', 'Basic_Demos-Sex',
       'CGAS-Season', 'CGAS-CGAS_Score', 'Physical-Season', 'Physical-BMI',
       'Physical-Height', 'Physical-Weight', 'Physical-Waist_Circumference',
       'Physical-Diastolic_BP', 'Physical-HeartRate', 'Physical-Systolic_BP',
       'Fitness_Endurance-Season', 'Fitness_Endurance-Max_Stage',
       'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec',
       'FGC-Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND',
```

```
'FGC-FGC_GSND_Zone', 'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU',
'FGC-FGC_PU_Zone', 'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR',
'FGC-FGC_SRR_Zone', 'FGC-FGC_TL', 'FGC-FGC_TL_Zone', 'BIA-Season',
'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC', 'BIA-BIA_BMI',
'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM',
'BIA-BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num',
'BIA-BIA_ICW', 'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM',
'BIA-BIA_TBW', 'PAQ_A-Season', 'PAQ_A-PAQ_A_Total', 'PAQ_C-Season',
'PAQ_C-PAQ_C_Total', 'PCIAT-PCIAT_Total', 'SDS-Season',
'SDS-SDS_Total_Raw', 'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
'PreInt_EduHx-computerinternet_hoursday', 'sii', 'Age_Group',
'PCIAT_Cluster', 'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw',
'CGAS-CGAS_Score_x_Physical-BMI'],
dtype='object')

Correlation of remaining interaction terms with PCIAT-PCIAT_Total:
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw: 0.1211
CGAS-CGAS_Score_x_Physical-BMI: 0.1258

Correlation of remaining interaction terms with sii:
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw: 0.1052
CGAS-CGAS_Score_x_Physical-BMI: 0.1120

Correlation of remaining interaction terms with PCIAT_Cluster:
CGAS-CGAS_Score_x_SDS-SDS_Total_Raw: 0.0326
CGAS-CGAS_Score_x_Physical-BMI: 0.0758
```

[26]:
```
train_df.head()
```

[26]:
```
          id  Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0   0.709955                   0.559547              5.0              0.0
2   0.449738                   0.543858             10.0              1.0
3   0.579846                   0.598983              9.0              0.0
5   0.579846                   0.475815             13.0              1.0
11  0.449738                   0.559547             11.0              0.0

    CGAS-Season  CGAS-CGAS_Score  Physical-Season  Physical-BMI  \
0      0.527092             51.0         0.526350     16.877316
2      0.538294             71.0         0.526350     16.648696
3      0.538294             71.0         0.516702     18.292347
5      0.527092             50.0         0.516702     22.279952
11     0.515994             66.0         0.471466     18.269958

    Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0         46.000000        50.800000                     23.154724
2         56.500000        75.600000                     25.755813
3         56.000000        81.600000                     30.237799
```

```
5        59.500000       112.200000                      33.812417
11       55.568097        83.565122                      26.013867

    Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
0               69.273255           86.942701            111.584992
2               65.000000           94.000000            117.000000
3               60.000000           97.000000            117.000000
5               60.000000           73.000000            102.000000
11              69.317251           84.110761            116.626216

    Fitness_Endurance-Season  Fitness_Endurance-Max_Stage  \
0                   0.562024                     4.661823
2                   0.416561                     5.000000
3                   0.419733                     6.000000
5                   0.562024                     3.781508
11                  0.562024                     4.861046

    Fitness_Endurance-Time_Mins  Fitness_Endurance-Time_Sec  FGC-Season  \
0                      5.902407                   30.190591    0.539792
2                      7.000000                   33.000000    0.539792
3                      9.000000                   37.000000    0.531007
5                      6.496836                   24.040340    0.531007
11                     8.019364                   28.006224    0.552717

    FGC-FGC_CU  FGC-FGC_CU_Zone  FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  \
0     0.000000         0.000000     12.457159           1.861092     7.105427
2    20.000000         1.000000     10.200000           1.000000    14.700000
3    18.000000         1.000000     15.489000           1.797822    17.595300
5    12.000000         0.000000     16.500000           2.000000    17.900000
11   10.701686         0.455832     19.270922           1.683190    19.325191

    FGC-FGC_GSD_Zone  FGC-FGC_PU  FGC-FGC_PU_Zone  FGC-FGC_SRL  \
0           2.104217    0.000000         0.000000     7.000000
2           2.000000    7.000000         1.000000    10.000000
3           2.176744    5.000000         0.000000     7.000000
5           2.000000    6.000000         0.000000    10.000000
11          1.630797    5.096538         0.302039     8.751021

    FGC-FGC_SRL_Zone  FGC-FGC_SRR  FGC-FGC_SRR_Zone  FGC-FGC_TL  \
0            0.00000      6.00000          0.000000    6.000000
2            1.00000     10.00000          1.000000    5.000000
3            0.00000      7.00000          0.000000    7.000000
5            1.00000     11.00000          1.000000    8.000000
11           0.68659      8.92129          0.690292    9.021283

    FGC-FGC_TL_Zone  BIA-Season  BIA-BIA_Activity_Level_num  BIA-BIA_BMC  \
0          1.000000    0.534475                    2.000000     2.668550
```

```
2          0.000000    0.490251                    2.658187     7.660622
3          1.000000    0.566098                    3.000000     3.841910
5          0.000000    0.566098                    2.000000     4.330360
11         0.723361    0.490251                    2.517398     7.660620

    BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0     16.879200   932.498000  1492.000000     8.255980    41.586200
2     18.540298  1229.028902  2004.524112    20.415826    73.170824
3     18.294300  1131.430000  1923.440000    15.592500    62.775700
5     30.186500  1330.970000  1996.450000    30.212400    84.028500
11    18.540307  1229.029934  2020.016337    20.415797    73.170847

    BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0      13.817700     3.061430     9.213770           1.000000    24.434900
2      14.810861     3.729442     5.294715           1.531391    32.829003
3      14.074000     4.220330    18.824300           2.000000    30.404100
5      16.687700    13.498800    67.971500           2.000000    32.914100
11     14.810861     3.729442    10.512782           1.644971    32.829002

    BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  \
0      8.895360    38.917700    19.541300    32.690900      0.480263
2     19.926022    65.510201    42.760486    53.244807      0.480263
3     16.779000    58.933800    26.479800    45.996600      0.480263
5     20.902000    79.698200    35.380400    63.126500      0.480263
11    19.926031    65.510206    29.831805    53.244809      0.480263

    PAQ_A-PAQ_A_Total  PAQ_C-Season  PAQ_C-PAQ_C_Total  PCIAT-PCIAT_Total  \
0            1.773888      0.460443            2.68491               55.0
2            1.991617      0.544231            2.17000               28.0
3            1.978289      0.716160            2.45100               44.0
5            1.968692      0.579612            4.11000               34.0
11           2.110197      0.716160            1.10000               10.0

    SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  PreInt_EduHx-Season  \
0     0.467407          42.186222        59.348885             0.564223
2     0.515460          38.000000        54.000000             0.534033
3     0.512539          31.000000        45.000000             0.602178
5     0.512539          40.000000        56.000000             0.486120
11    0.611293          42.000000        59.000000             0.564223

    PreInt_EduHx-computerinternet_hoursday  sii  Age_Group  PCIAT_Cluster  \
0                                      3.0  2.0   0.443345            5.0
2                                      2.0  0.0   0.362030            7.0
3                                      0.0  1.0   0.362030            4.0
5                                      0.0  1.0   0.855589            3.0
11                                     0.0  0.0   0.855589            3.0
```

```
    CGAS-CGAS_Score_x_SDS-SDS_Total_Raw  CGAS-CGAS_Score_x_Physical-BMI
0                           2156.462944                      860.743100
2                           2698.000000                     1182.057420
3                           2201.000000                     1298.756633
5                           2000.000000                     1113.997599
11                          2772.000000                     1206.032260
```

Actigraphy Data analysis

[27]:
```
import pyarrow.parquet as pq

# Load one of the series data files
series_file = 'kaggle/input/child-mind-institute-problematic-internet-use/
 ↪series_train.parquet/id=0745c390/part-0.parquet'
series_data = pq.read_table(series_file).to_pandas()

# Display the first few rows of the series data
print("Sample Series Data:")
display(series_data.head())
```

```
Sample Series Data:

   step         X         Y         Z      enmo     anglez  non-wear_flag  \
0     0  0.088385 -0.021719  0.988333  0.000192  84.628494            0.0
1     1  0.088750 -0.021823  0.988698  0.000246  84.657318            0.0
2     2  0.022700 -0.015929  0.988151  0.023109  87.189415            0.0
3     3 -0.076641 -0.048411  0.932135  0.043126  85.428131            0.0
4     4  0.006432  0.036354  1.000651  0.006071  88.175079            0.0

       light  battery_voltage     time_of_day  weekday  quarter  \
0   8.250000      4186.000000  52200000000000        6        2
1  15.666667      4185.166504  52205000000000        6        2
2  13.583333      4187.000000  54385000000000        6        2
3  11.500000      4187.000000  54390000000000        6        2
4   1.661765      4187.000000  54395000000000        6        2

   relative_date_PCIAT
0                 15.0
1                 15.0
2                 15.0
3                 15.0
4                 15.0
```

[28]:
```
series_data.shape
```

[28]:
```
(50458, 13)
```

[29]:
```python
#import pandas as pd
#import os
#path_arr_train = []
#path_arr_test = []
#for dirname, _, filenames in os.walk('kaggle/input'):
#    for filename in filenames:
#        path = os.path.join(dirname, filename)
#        if('series_train.parquet' in path):
#            path_arr_train.append(path)
#        if('series_test.parquet' in path):
#            path_arr_test.append(path)

#print(len(path_arr_train)/2)
#dataframes = []
#for i in range(int(len(path_arr_train)/2)):
#    file = path_arr_train[i]
#    df = pd.read_parquet(file)
#    dataframes.append(df)
#print("Train Dataframe Generated")
#actigrapghy_train_df = pd.concat(dataframes)

import pandas as pd
import os

# Initialize empty lists to hold the paths for train and test data
path_arr_train = []
path_arr_test = []

# Traverse the input directory to find and store paths of the parquet files
print("Here1")
for dirname, _, filenames in os.walk('kaggle/input'):
    for filename in filenames:
        path = os.path.join(dirname, filename)
        if 'series_train.parquet' in path:
            path_arr_train.append(path)
        if 'series_test.parquet' in path:
            path_arr_test.append(path)

# List to hold individual dataframes
print('Here2')
dataframes = []

# Loop through each train file path
for file in path_arr_train:
    # Read the dataframe from parquet file
    df = pd.read_parquet(file)
```

```
    # Extract the ID from the file path. Assuming that the ID is always after␣
 ↪"id=" in the path.
    # For example: /kaggle/input/child-mind-institute-problematic-internet-use/
 ↪series_train.parquet/id=5f099188/part-0.parquet
    id_value = file.split('id=')[1].split('/')[0]  # Extracts "5f099188" from␣
 ↪the path

    # Add the ID as a new column to the dataframe
    df['ID'] = id_value

    # Append the dataframe to the list
    dataframes.append(df)

print("Train Dataframe Generated")

# Concatenate all the individual dataframes into a single dataframe
actigraphy_train_df = pd.concat(dataframes)

# Display the first few rows of the combined dataframe
print(actigraphy_train_df.head())
```

```
Here1
Here2
Train Dataframe Generated
   step         X         Y         Z      enmo     anglez  non-wear_flag  \
0     0 -0.853643 -0.452480 -0.004268  0.039495  -2.348142            0.0
1     1 -0.740500 -0.481082 -0.096710  0.076673  -6.455178            0.0
2     2 -0.602861  0.028234  0.374689  0.052310  31.838375            0.0
3     3 -0.008948  0.018434  1.001781  0.002765  88.850754            0.0
4     4 -0.007265  0.017022  1.001847  0.002946  88.973587            0.0

   light  battery_voltage     time_of_day  weekday  quarter  \
0   3.00      4171.000000  52320000000000        5        2
1  21.00      4171.083496  52325000000000        5        2
2  21.00      4171.166504  52330000000000        5        2
3  18.50      4171.250000  52335000000000        5        2
4  18.75      4171.333496  52340000000000        5        2

   relative_date_PCIAT        ID
0                -25.0  af485add
1                -25.0  af485add
2                -25.0  af485add
3                -25.0  af485add
4                -25.0  af485add
```

[30]:
```
dataframes = []
for file in path_arr_test:
```

```
    # Read the dataframe from parquet file
    df = pd.read_parquet(file)

    # Extract the ID from the file path. Assuming that the ID is always after␣
↪"id=" in the path.
    # For example: /kaggle/input/child-mind-institute-problematic-internet-use/
↪series_train.parquet/id=5f099188/part-0.parquet
    id_value = file.split('id=')[1].split('/')[0]  # Extracts "5f099188" from␣
↪the path

    # Add the ID as a new column to the dataframe
    df['ID'] = id_value

    # Append the dataframe to the list
    dataframes.append(df)

actigraphy_test_df = pd.concat(dataframes)
```

[31]:
```
actigraphy_test_df.columns
```

[31]:
```
Index(['step', 'X', 'Y', 'Z', 'enmo', 'anglez', 'non-wear_flag', 'light',
       'battery_voltage', 'time_of_day', 'weekday', 'quarter',
       'relative_date_PCIAT', 'ID'],
      dtype='object')
```

[32]:
```
actigraphy_train_df.head()
```

[32]:
```
   step         X         Y         Z      enmo     anglez  non-wear_flag  \
0     0 -0.853643 -0.452480 -0.004268  0.039495  -2.348142            0.0
1     1 -0.740500 -0.481082 -0.096710  0.076673  -6.455178            0.0
2     2 -0.602861  0.028234  0.374689  0.052310  31.838375            0.0
3     3 -0.008948  0.018434  1.001781  0.002765  88.850754            0.0
4     4 -0.007265  0.017022  1.001847  0.002946  88.973587            0.0

   light  battery_voltage     time_of_day  weekday  quarter  \
0   3.00      4171.000000  52320000000000        5        2
1  21.00      4171.083496  52325000000000        5        2
2  21.00      4171.166504  52330000000000        5        2
3  18.50      4171.250000  52335000000000        5        2
4  18.75      4171.333496  52340000000000        5        2

   relative_date_PCIAT        ID
0                -25.0  af485add
1                -25.0  af485add
2                -25.0  af485add
3                -25.0  af485add
4                -25.0  af485add
```

[33]:
```
actigraphy_train_df['ID'].value_counts()
```

[33]:
```
ID
2a88cbe9    756212
0b4014f0    727484
697c41d7    674410
75b0446f    594911
04cb2c30    593331
              …
055156e2      2080
b802aec3      1940
a8b0428d      1386
f8ff0bc8      1195
e1ea8dd7       927
Name: count, Length: 996, dtype: int64
```

[34]:
```
# List to hold individual dataframes for test data
test_dataframes = []

# Loop through each test file path
for file in path_arr_test:
    # Read the dataframe from parquet file
    df = pd.read_parquet(file)

    # Extract the ID from the file path. Assuming that the ID is always after␣
 ↪"id=" in the path.
    # For example: /kaggle/input/child-mind-institute-problematic-internet-use/
 ↪series_test.parquet/id=5f099188/part-0.parquet
    id_value = file.split('id=')[1].split('/')[0]  # Extracts "5f099188" from␣
 ↪the path

    # Add the ID as a new column to the dataframe
    df['ID'] = id_value

    # Append the dataframe to the list
    test_dataframes.append(df)

print("Test Dataframe Generated")

# Concatenate all the individual test dataframes into a single dataframe
actigraphy_test_df = pd.concat(test_dataframes)

# Display the first few rows of the combined test dataframe
print(actigraphy_test_df.head())
```

```
Test Dataframe Generated
   step         X         Y         Z      enmo     anglez  non-wear_flag  \
0     0  0.021536  0.022214 -1.022370  0.022853 -88.280762            0.0
```

```
1     1  0.022005  0.022187 -1.019740  0.020231 -88.241707            0.0
2     2  0.022240  0.022005 -1.019401  0.019893 -88.170067            0.0
3     3  0.021589  0.022578 -1.018177  0.018667 -88.250031            0.0
4     4  0.022005  0.023763 -1.014323  0.016848 -88.130775            0.0

       light  battery_voltage     time_of_day  weekday  quarter  \
0  53.000000      4188.000000  56940000000000        4        3
1  51.666668      4188.166504  56945000000000        4        3
2  50.333332      4188.333496  56950000000000        4        3
3  50.500000      4188.500000  56955000000000        4        3
4  33.166668      4181.000000  57235000000000        4        3

   relative_date_PCIAT        ID
0                 41.0  00115b9f
1                 41.0  00115b9f
2                 41.0  00115b9f
3                 41.0  00115b9f
4                 41.0  00115b9f
```

[35]:
```
actigraphy_test_df.head()
```

[35]:
```
   step         X         Y         Z      enmo     anglez  non-wear_flag  \
0     0  0.021536  0.022214 -1.022370  0.022853 -88.280762            0.0
1     1  0.022005  0.022187 -1.019740  0.020231 -88.241707            0.0
2     2  0.022240  0.022005 -1.019401  0.019893 -88.170067            0.0
3     3  0.021589  0.022578 -1.018177  0.018667 -88.250031            0.0
4     4  0.022005  0.023763 -1.014323  0.016848 -88.130775            0.0

       light  battery_voltage     time_of_day  weekday  quarter  \
0  53.000000      4188.000000  56940000000000        4        3
1  51.666668      4188.166504  56945000000000        4        3
2  50.333332      4188.333496  56950000000000        4        3
3  50.500000      4188.500000  56955000000000        4        3
4  33.166668      4181.000000  57235000000000        4        3

   relative_date_PCIAT        ID
0                 41.0  00115b9f
1                 41.0  00115b9f
2                 41.0  00115b9f
3                 41.0  00115b9f
4                 41.0  00115b9f
```

[36]:
```
# Install necessary package
#!pip install pyts

# Import required libraries
#import pandas as pd
#from pyts.decomposition import SingularSpectrumAnalysis
```

```
# Create a copy of the dataframe to store denoised columns
#denoised_actigraphy_df = actigraphy_train_df.copy()

# Columns to apply SSA denoising on
#columns_to_denoise = ['X', 'Y', 'Z']

# Apply SSA to each of the specified columns using pyts
#for column in columns_to_denoise:
#    # Get the column series
#    series = actigraphy_train_df[column]
#
#    # Set window_size dynamically based on the length of the series
#    series_length = len(series)
#    window_size = min(10, series_length - 1)  # Set to minimum value
#
#    # Create an SSA object if the window_size is valid
#    if window_size > 1:
#        ssa = SingularSpectrumAnalysis(window_size=window_size)
#
#        # Reshape series for SSA and apply fit_transform
#        denoised_series = ssa.fit_transform(series.values.reshape(1, -1))[0]
#
#        # Store the denoised series back to the dataframe
#        denoised_actigraphy_df[column] = denoised_series
#
# Display the denoised dataframe head
#denoised_actigraphy_df.head()
```

Aggregating Actigraphy Data To get started, we can first implement the aggregation step and verify the results before moving to the next step:

[37]:
```
# Aggregate actigraphy data by ID
actigraphy_train_summary_df = (
    actigraphy_train_df
    .groupby('ID')  # Group by ID
    .agg({
        'X': ['mean', 'std', 'min', 'max'],  # Aggregate X with mean, std, min,␣
 ↪max
        'Y': ['mean', 'std', 'min', 'max'],  # Aggregate Y with mean, std, min,␣
 ↪max
        'Z': ['mean', 'std', 'min', 'max'],  # Aggregate Z with mean, std, min,␣
 ↪max
        'enmo': ['mean', 'std'],  # Aggregate enmo with mean, std
        'anglez': ['mean', 'std'],  # Aggregate anglez with mean, std
        'non-wear_flag': 'mean',  # Proportion of non-wear time
        'light': ['mean', 'std'],  # Aggregate light with mean, std
```

```
        'battery_voltage': ['mean', 'std'],  # Aggregate battery voltage
        'time_of_day': 'mean',  # Mean time of day
        'weekday': 'mean',  # Average weekday for the activity readings
        'quarter': 'mean',  # Average quarter for the activity readings
        'relative_date_PCIAT': 'mean'  # Average relative date PCIAT
    })
)

# Flatten MultiIndex columns
actigraphy_train_summary_df.columns = ['_'.join(col) for col in␣
 ↪actigraphy_train_summary_df.columns]

# Reset index to make 'ID' a column again
actigraphy_train_summary_df.reset_index(inplace=True)

# Display first few rows of the aggregated summary
actigraphy_train_summary_df.head()
```

[37]:

```
         ID    X_mean     X_std     X_min     X_max    Y_mean     Y_std  \
0  00115b9f -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702
1  001f3379 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812
2  00f332d1  0.208036  0.486977 -1.952594  1.666465  0.057094  0.443755
3  01085eb3 -0.343396  0.516126 -2.284304  1.000692 -0.055826  0.424303
4  012cadd8  0.018670  0.595251 -2.143912  3.341210  0.071660  0.508311

      Y_min     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  \
0 -2.905339  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388
1 -1.522690  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926
2 -2.361866  1.016429  0.141550  0.683114 -1.016758  2.239939   0.030255
3 -2.276082  1.011419 -0.254433  0.564593 -1.022549  1.299293   0.032946
4 -3.373025  4.442658 -0.061682  0.578022 -1.003249  2.321265   0.058280

   enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  light_mean  \
0  0.106351   -10.580416   42.947170            0.000000   42.296310
1  0.024331   -55.630768   50.303635            0.655708   16.771980
2  0.104136     6.687339   52.754208            0.171246   66.563393
3  0.083798   -17.589037   39.895645            0.035210   17.800735
4  0.197285    -5.059758   39.994808            0.000000   54.893402

    light_std  battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
0  208.168976           4053.578857           112.404037      5.046215e+13
1   95.327438           3838.189453           155.573868      4.321212e+13
2  286.916595           3848.583252           166.968582      4.318680e+13
3   73.023468           3849.650146           171.100159      4.338433e+13
4  230.972397           3974.910889           119.525154      4.343573e+13

   weekday_mean  quarter_mean  relative_date_PCIAT_mean
```

0	4.470182	3.0	53.201683
1	3.909848	3.0	79.435593
2	3.832677	2.0	26.152903
3	3.963284	4.0	49.910686
4	4.168412	4.0	-1.168288

Computing and Adding Fourier Transforms Here's how we can incorporate Fourier Transforms in the aggregated actigraphy data:

[38]:
```
import numpy as np

# Function to compute top n Fourier coefficients
def compute_top_n_fourier_coefficients(signal, n=5):
    # Perform the Fourier Transform on the signal
    fft_coeffs = np.fft.fft(signal)
    # Calculate magnitudes of the FFT coefficients
    magnitudes = np.abs(fft_coeffs)
    # Get top n coefficients based on magnitudes (excluding the DC component at␣
 ↪index 0)
    top_n_indices = np.argsort(magnitudes)[-n:]
    # Return the real and imaginary parts of the top n coefficients
    top_n_real_parts = fft_coeffs[top_n_indices].real
    top_n_imag_parts = fft_coeffs[top_n_indices].imag
    return top_n_real_parts, top_n_imag_parts

# List to store Fourier features for each participant
fourier_features_list = []

# Calculate Fourier Transforms for each ID in the actigraphy training data
for participant_id, participant_data in actigraphy_train_df.groupby('ID'):
    # Calculate Fourier coefficients for each axis
    fourier_X_real, fourier_X_imag =␣
 ↪compute_top_n_fourier_coefficients(participant_data['X'].values)
    fourier_Y_real, fourier_Y_imag =␣
 ↪compute_top_n_fourier_coefficients(participant_data['Y'].values)
    fourier_Z_real, fourier_Z_imag =␣
 ↪compute_top_n_fourier_coefficients(participant_data['Z'].values)

    # Create a dictionary to store Fourier features for this participant
    fourier_features = {
        'ID': participant_id,
        'fourier_X_real': fourier_X_real.mean(),  # Using mean of the top real␣
 ↪parts as a feature
        'fourier_X_imag': fourier_X_imag.mean(),  # Using mean of the top␣
 ↪imaginary parts as a feature
        'fourier_Y_real': fourier_Y_real.mean(),
        'fourier_Y_imag': fourier_Y_imag.mean(),
```

```python
        'fourier_Z_real': fourier_Z_real.mean(),
        'fourier_Z_imag': fourier_Z_imag.mean(),
    }
    fourier_features_list.append(fourier_features)

# Convert the list of Fourier features into a DataFrame
fourier_features_df = pd.DataFrame(fourier_features_list)

# Merge the Fourier features into the existing actigraphy summary dataframe
actigraphy_train_summary_with_fourier_df = pd.merge(
    actigraphy_train_summary_df,
    fourier_features_df,
    on='ID',
    how='left'
)

# Display first few rows of the combined summary
print(actigraphy_train_summary_with_fourier_df.head())
```

```
         ID    X_mean     X_std     X_min     X_max    Y_mean     Y_std  \
0  00115b9f -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702
1  001f3379 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812
2  00f332d1  0.208036  0.486977 -1.952594  1.666465  0.057094  0.443755
3  01085eb3 -0.343396  0.516126 -2.284304  1.000692 -0.055826  0.424303
4  012cadd8  0.018670  0.595251 -2.143912  3.341210  0.071660  0.508311

      Y_min     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  \
0 -2.905339  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388
1 -1.522690  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926
2 -2.361866  1.016429  0.141550  0.683114 -1.016758  2.239939   0.030255
3 -2.276082  1.011419 -0.254433  0.564593 -1.022549  1.299293   0.032946
4 -3.373025  4.442658 -0.061682  0.578022 -1.003249  2.321265   0.058280

   enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  light_mean  \
0  0.106351   -10.580416   42.947170            0.000000   42.296310
1  0.024331   -55.630768   50.303635            0.655708   16.771980
2  0.104136     6.687339   52.754208            0.171246   66.563393
3  0.083798   -17.589037   39.895645            0.035210   17.800735
4  0.197285    -5.059758   39.994808            0.000000   54.893402

    light_std  battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
0  208.168976           4053.578857           112.404037      5.046215e+13
1   95.327438           3838.189453           155.573868      4.321212e+13
2  286.916595           3848.583252           166.968582      4.318680e+13
3   73.023468           3849.650146           171.100159      4.338433e+13
4  230.972397           3974.910889           119.525154      4.343573e+13

   weekday_mean  quarter_mean  relative_date_PCIAT_mean  fourier_X_real  \
```

```
0      4.470182         3.0                    53.201683    -2653.531675
1      3.909848         3.0                    79.435593     4103.865122
2      3.832677         2.0                    26.152903    14903.301506
3      3.963284         4.0                    49.910686   -25796.646158
4      4.168412         4.0                    -1.168288     6637.890207

   fourier_X_imag  fourier_Y_real  fourier_Y_imag  fourier_Z_real  \
0   -1.083578e-14      705.803060   -2.074467e+02     1995.401835
1   -1.680035e+03    14385.071425    3.120931e+03   -26765.674436
2   -3.637979e-13     -923.242685    7.275958e-13    -5261.764351
3   -1.093525e-12   -10776.376135   -1.989520e-13    -8614.531458
4    5.618541e+02     2279.335449   -3.637979e-13    -4597.967535

   fourier_Z_imag
0   -5.684342e-15
1   -8.731149e-12
2    7.275958e-13
3   -9.563905e-13
4    0.000000e+00
```

[39]:
```python
pd.set_option('display.max_columns', None)
actigraphy_train_summary_with_fourier_df.head()
```

[39]:
```
         ID    X_mean     X_std     X_min     X_max    Y_mean     Y_std  \
0  00115b9f -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702
1  001f3379 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812
2  00f332d1  0.208036  0.486977 -1.952594  1.666465  0.057094  0.443755
3  01085eb3 -0.343396  0.516126 -2.284304  1.000692 -0.055826  0.424303
4  012cadd8  0.018670  0.595251 -2.143912  3.341210  0.071660  0.508311

      Y_min     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  \
0 -2.905339  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388
1 -1.522690  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926
2 -2.361866  1.016429  0.141550  0.683114 -1.016758  2.239939   0.030255
3 -2.276082  1.011419 -0.254433  0.564593 -1.022549  1.299293   0.032946
4 -3.373025  4.442658 -0.061682  0.578022 -1.003249  2.321265   0.058280

   enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  light_mean  \
0  0.106351   -10.580416   42.947170            0.000000   42.296310
1  0.024331   -55.630768   50.303635            0.655708   16.771980
2  0.104136     6.687339   52.754208            0.171246   66.563393
3  0.083798   -17.589037   39.895645            0.035210   17.800735
4  0.197285    -5.059758   39.994808            0.000000   54.893402

    light_std  battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
0  208.168976           4053.578857           112.404037      5.046215e+13
1   95.327438           3838.189453           155.573868      4.321212e+13
```

```
2  286.916595          3848.583252          166.968582      4.318680e+13
3   73.023468          3849.650146          171.100159      4.338433e+13
4  230.972397          3974.910889          119.525154      4.343573e+13

   weekday_mean  quarter_mean  relative_date_PCIAT_mean  fourier_X_real  \
0      4.470182           3.0                 53.201683    -2653.531675
1      3.909848           3.0                 79.435593     4103.865122
2      3.832677           2.0                 26.152903    14903.301506
3      3.963284           4.0                 49.910686   -25796.646158
4      4.168412           4.0                 -1.168288     6637.890207

   fourier_X_imag  fourier_Y_real  fourier_Y_imag  fourier_Z_real  \
0   -1.083578e-14      705.803060   -2.074467e+02     1995.401835
1   -1.680035e+03    14385.071425    3.120931e+03   -26765.674436
2   -3.637979e-13     -923.242685    7.275958e-13    -5261.764351
3   -1.093525e-12   -10776.376135   -1.989520e-13    -8614.531458
4    5.618541e+02     2279.335449   -3.637979e-13    -4597.967535

   fourier_Z_imag
0   -5.684342e-15
1   -8.731149e-12
2    7.275958e-13
3   -9.563905e-13
4    0.000000e+00
```

Explanation Function compute_top_n_fourier_coefficients: Calculates the top n Fourier coefficients (real and imaginary parts) for a given time-series signal. Loop through each participant's data: For each ID, compute the Fourier Transform of X, Y, and Z columns. Feature Extraction: Extract relevant Fourier features (mean real and imaginary parts of top coefficients). Merging: Combine these Fourier features with the existing summary statistics for a more comprehensive dataset.

Time Domain Features

[40]:
```python
# Function to calculate day and night activity based on time_of_day
def calculate_day_night_activity(df, day_start=6, day_end=18):
    # Calculate day activity (time_of_day between day_start and day_end)
    day_activity = df[(df['time_of_day'] >= day_start) & (df['time_of_day'] <
 ↪day_end)]['enmo'].sum()
    # Calculate night activity (time_of_day outside the day range)
    night_activity = df[(df['time_of_day'] < day_start) | (df['time_of_day'] >=
 ↪day_end)]['enmo'].sum()
    return day_activity, night_activity

# List to store time-domain features for each participant
time_features_list = []

# Calculate time-domain features for each participant
```

```
for participant_id, participant_data in actigraphy_train_df.groupby('ID'):
    # Calculate day and night activity
    day_activity, night_activity = ␣
↪calculate_day_night_activity(participant_data)
    activity_ratio_day_night = day_activity / (night_activity + 1e-6)  # Avoid␣
↪division by zero

    # Calculate proportion of non-wear time
    non_wear_proportion = participant_data['non-wear_flag'].mean()

    # Calculate proportion of time spent in different activity levels based on␣
↪'enmo'
    sedentary_time = participant_data[participant_data['enmo'] < 0.05].shape[0]
    light_activity_time = participant_data[(participant_data['enmo'] >= 0.05) &␣
↪(participant_data['enmo'] < 0.5)].shape[0]
    moderate_activity_time = participant_data[(participant_data['enmo'] >= 0.5)␣
↪& (participant_data['enmo'] < 1)].shape[0]
    vigorous_activity_time = participant_data[participant_data['enmo'] >= 1].
↪shape[0]
    total_time = participant_data.shape[0]

    # Calculate proportions
    sedentary_proportion = sedentary_time / total_time
    light_activity_proportion = light_activity_time / total_time
    moderate_activity_proportion = moderate_activity_time / total_time
    vigorous_activity_proportion = vigorous_activity_time / total_time

    # Create a dictionary of time domain features
    time_features = {
        'ID': participant_id,
        'activity_during_day': day_activity,
        'activity_during_night': night_activity,
        'activity_ratio_day_night': activity_ratio_day_night,
        'non_wear_proportion': non_wear_proportion,
        'sedentary_proportion': sedentary_proportion,
        'light_activity_proportion': light_activity_proportion,
        'moderate_activity_proportion': moderate_activity_proportion,
        'vigorous_activity_proportion': vigorous_activity_proportion
    }
    time_features_list.append(time_features)

# Convert the list of time domain features into a DataFrame
time_features_df = pd.DataFrame(time_features_list)

# Merge the time-domain features into the actigraphy summary with Fourier␣
↪features
```

```python
actigraphy_combined_features_df = pd.merge(
    actigraphy_train_summary_with_fourier_df,  # The dataframe with Fourier␣
↪features
    time_features_df,  # The new time-domain features
    on='ID',
    how='left'
)

# Display the first few rows of the final combined feature set
actigraphy_combined_features_df.head()
```

[40]:

```
         ID    X_mean     X_std     X_min     X_max    Y_mean     Y_std  \
0  00115b9f -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702
1  001f3379 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812
2  00f332d1  0.208036  0.486977 -1.952594  1.666465  0.057094  0.443755
3  01085eb3 -0.343396  0.516126 -2.284304  1.000692 -0.055826  0.424303
4  012cadd8  0.018670  0.595251 -2.143912  3.341210  0.071660  0.508311

      Y_min     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  \
0 -2.905339  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388
1 -1.522690  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926
2 -2.361866  1.016429  0.141550  0.683114 -1.016758  2.239939   0.030255
3 -2.276082  1.011419 -0.254433  0.564593 -1.022549  1.299293   0.032946
4 -3.373025  4.442658 -0.061682  0.578022 -1.003249  2.321265   0.058280

   enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  light_mean  \
0  0.106351   -10.580416   42.947170            0.000000   42.296310
1  0.024331   -55.630768   50.303635            0.655708   16.771980
2  0.104136     6.687339   52.754208            0.171246   66.563393
3  0.083798   -17.589037   39.895645            0.035210   17.800735
4  0.197285    -5.059758   39.994808            0.000000   54.893402

    light_std  battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
0  208.168976           4053.578857           112.404037      5.046215e+13
1   95.327438           3838.189453           155.573868      4.321212e+13
2  286.916595           3848.583252           166.968582      4.318680e+13
3   73.023468           3849.650146           171.100159      4.338433e+13
4  230.972397           3974.910889           119.525154      4.343573e+13

   weekday_mean  quarter_mean  relative_date_PCIAT_mean  fourier_X_real  \
0      4.470182           3.0                 53.201683    -2653.531675
1      3.909848           3.0                 79.435593     4103.865122
2      3.832677           2.0                 26.152903    14903.301506
3      3.963284           4.0                 49.910686   -25796.646158
4      4.168412           4.0                 -1.168288     6637.890207

   fourier_X_imag  fourier_Y_real  fourier_Y_imag  fourier_Z_real  \
```

```
0   -1.083578e-14      705.803060   -2.074467e+02     1995.401835
1   -1.680035e+03    14385.071425    3.120931e+03   -26765.674436
2   -3.637979e-13     -923.242685    7.275958e-13    -5261.764351
3   -1.093525e-12   -10776.376135   -1.989520e-13    -8614.531458
4    5.618541e+02     2279.335449   -3.637979e-13    -4597.967535

   fourier_Z_imag  activity_during_day  activity_during_night  \
0   -5.684342e-15                  0.0            2053.305176
1   -8.731149e-12                  0.0            4727.518555
2    7.275958e-13                  0.0           12537.293945
3   -9.563905e-13                  0.0           12221.464844
4    0.000000e+00                  0.0            5655.915527

   activity_ratio_day_night  non_wear_proportion  sedentary_proportion  \
0                       0.0             0.000000              0.792453
1                       0.0             0.655708              0.978501
2                       0.0             0.171246              0.867572
3                       0.0             0.035210              0.840178
4                       0.0             0.000000              0.789187

   light_activity_proportion  moderate_activity_proportion  \
0                   0.198131                      0.007870
1                   0.021171                      0.000288
2                   0.124059                      0.005562
3                   0.154229                      0.004567
4                   0.189535                      0.012808

   vigorous_activity_proportion
0                      0.001546
1                      0.000040
2                      0.002807
3                      0.001027
4                      0.008470
```

Time Domain Feature Engineering

Activity During Day and Night

Create two new features: activity_during_day and activity_during_night, based on time_of_day. If time_of_day falls within a specific range (e.g., 6 AM to 6 PM for day and 6 PM to 6 AM for night), sum up the activity values in enmo for each period. Activity Ratio (Day/Night)

Calculate the ratio of activity during the day to activity during the night: activity_ratio_day_night = activity_during_day / (activity_during_night + 1e-6). This helps measure how active a participant is during the day compared to the night. Proportion of Non-Wear Time

Use the non-wear_flag column to calculate the proportion of time each participant was not wearing the device: non_wear_proportion = non_wear_time / total_time. Activity Level Classification

Create thresholds for enmo values to classify activity levels (e.g., sedentary, light activity, moderate activity, vigorous activity). Calculate the proportion of time spent in each activity level. Additional Domain-Specific Features

Use columns like anglez, light, and battery_voltage to derive additional features if they show any correlation with the target variable sii

Merge the DataFrames Let's merge actigraphy_combined_features_df with the main training dataset based on the ID column to ensure we have a single DataFrame for the combined analysis.

[41]:
```
train_df['id'] = ID_arr
```

[42]:
```
# Merging the main training dataset with actigraphy_combined_features_df based␣
↪on the ID column
merged_train_df = pd.merge(train_df, actigraphy_combined_features_df,␣
↪how='left', left_on='id', right_on='ID')

# Dropping the redundant ID column from actigraphy_combined_features_df if␣
↪needed
merged_train_df = merged_train_df.drop(columns=['ID'])

# Display the first few rows of the merged DataFrame to confirm the merge
#print(merged_train_df.head())
```

[43]:
```
merged_train_df.head()
```

[43]:
```
         id  Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0  00008ff9                   0.559547              5.0              0.0
1  00105258                   0.543858             10.0              1.0
2  00115b9f                   0.598983              9.0              0.0
3  001f3379                   0.475815             13.0              1.0
4  00abe655                   0.559547             11.0              0.0

   CGAS-Season  CGAS-CGAS_Score  Physical-Season  Physical-BMI  \
0     0.527092             51.0         0.526350     16.877316
1     0.538294             71.0         0.526350     16.648696
2     0.538294             71.0         0.516702     18.292347
3     0.527092             50.0         0.516702     22.279952
4     0.515994             66.0         0.471466     18.269958

   Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0        46.000000        50.800000                     23.154724
1        56.500000        75.600000                     25.755813
2        56.000000        81.600000                     30.237799
3        59.500000       112.200000                     33.812417
4        55.568097        83.565122                     26.013867

   Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
```

0	69.273255	86.942701	111.584992
1	65.000000	94.000000	117.000000
2	60.000000	97.000000	117.000000
3	60.000000	73.000000	102.000000
4	69.317251	84.110761	116.626216

	Fitness_Endurance-Season	Fitness_Endurance-Max_Stage	\
0	0.562024	4.661823	
1	0.416561	5.000000	
2	0.419733	6.000000	
3	0.562024	3.781508	
4	0.562024	4.861046	

	Fitness_Endurance-Time_Mins	Fitness_Endurance-Time_Sec	FGC-Season	\
0	5.902407	30.190591	0.539792	
1	7.000000	33.000000	0.539792	
2	9.000000	37.000000	0.531007	
3	6.496836	24.040340	0.531007	
4	8.019364	28.006224	0.552717	

	FGC-FGC_CU	FGC-FGC_CU_Zone	FGC-FGC_GSND	FGC-FGC_GSND_Zone	FGC-FGC_GSD	\
0	0.000000	0.000000	12.457159	1.861092	7.105427	
1	20.000000	1.000000	10.200000	1.000000	14.700000	
2	18.000000	1.000000	15.489000	1.797822	17.595300	
3	12.000000	0.000000	16.500000	2.000000	17.900000	
4	10.701686	0.455832	19.270922	1.683190	19.325191	

	FGC-FGC_GSD_Zone	FGC-FGC_PU	FGC-FGC_PU_Zone	FGC-FGC_SRL	\
0	2.104217	0.000000	0.000000	7.000000	
1	2.000000	7.000000	1.000000	10.000000	
2	2.176744	5.000000	0.000000	7.000000	
3	2.000000	6.000000	0.000000	10.000000	
4	1.630797	5.096538	0.302039	8.751021	

	FGC-FGC_SRL_Zone	FGC-FGC_SRR	FGC-FGC_SRR_Zone	FGC-FGC_TL	\
0	0.00000	6.00000	0.000000	6.000000	
1	1.00000	10.00000	1.000000	5.000000	
2	0.00000	7.00000	0.000000	7.000000	
3	1.00000	11.00000	1.000000	8.000000	
4	0.68659	8.92129	0.690292	9.021283	

	FGC-FGC_TL_Zone	BIA-Season	BIA-BIA_Activity_Level_num	BIA-BIA_BMC	\
0	1.000000	0.534475	2.000000	2.668550	
1	0.000000	0.490251	2.658187	7.660622	
2	1.000000	0.566098	3.000000	3.841910	
3	0.000000	0.566098	2.000000	4.330360	
4	0.723361	0.490251	2.517398	7.660620	

```
   BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0    16.879200   932.498000  1492.000000     8.255980    41.586200
1    18.540298  1229.028902  2004.524112    20.415826    73.170824
2    18.294300  1131.430000  1923.440000    15.592500    62.775700
3    30.186500  1330.970000  1996.450000    30.212400    84.028500
4    18.540307  1229.029934  2020.016337    20.415797    73.170847

   BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0     13.817700     3.061430     9.213770           1.000000    24.434900
1     14.810861     3.729442     5.294715           1.531391    32.829003
2     14.074000     4.220330    18.824300           2.000000    30.404100
3     16.687700    13.498800    67.971500           2.000000    32.914100
4     14.810861     3.729442    10.512782           1.644971    32.829002

   BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  \
0     8.895360    38.917700    19.541300    32.690900      0.480263
1    19.926022    65.510201    42.760486    53.244807      0.480263
2    16.779000    58.933800    26.479800    45.996600      0.480263
3    20.902000    79.698200    35.380400    63.126500      0.480263
4    19.926031    65.510206    29.831805    53.244809      0.480263

   PAQ_A-PAQ_A_Total  PAQ_C-Season  PAQ_C-PAQ_C_Total  PCIAT-PCIAT_Total  \
0           1.773888      0.460443            2.68491               55.0
1           1.991617      0.544231            2.17000               28.0
2           1.978289      0.716160            2.45100               44.0
3           1.968692      0.579612            4.11000               34.0
4           2.110197      0.716160            1.10000               10.0

   SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  PreInt_EduHx-Season  \
0    0.467407          42.186222        59.348885             0.564223
1    0.515460          38.000000        54.000000             0.534033
2    0.512539          31.000000        45.000000             0.602178
3    0.512539          40.000000        56.000000             0.486120
4    0.611293          42.000000        59.000000             0.564223

   PreInt_EduHx-computerinternet_hoursday  sii  Age_Group  PCIAT_Cluster  \
0                                     3.0  2.0   0.443345            5.0
1                                     2.0  0.0   0.362030            7.0
2                                     0.0  1.0   0.362030            4.0
3                                     0.0  1.0   0.855589            3.0
4                                     0.0  0.0   0.855589            3.0

   CGAS-CGAS_Score_x_SDS-SDS_Total_Raw  CGAS-CGAS_Score_x_Physical-BMI  \
0                          2156.462944                      860.743100
1                          2698.000000                     1182.057420
2                          2201.000000                     1298.756633
```

```
3                          2000.000000                          1113.997599
4                          2772.000000                          1206.032260

    X_mean     X_std     X_min     X_max    Y_mean     Y_std     Y_min  \
0      NaN       NaN       NaN       NaN       NaN       NaN       NaN
1      NaN       NaN       NaN       NaN       NaN       NaN       NaN
2 -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702 -2.905339
3 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812 -1.522690
4      NaN       NaN       NaN       NaN       NaN       NaN       NaN

     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  enmo_std  \
0      NaN       NaN       NaN       NaN       NaN        NaN       NaN
1      NaN       NaN       NaN       NaN       NaN        NaN       NaN
2  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388  0.106351
3  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926  0.024331
4      NaN       NaN       NaN       NaN       NaN        NaN       NaN

   anglez_mean  anglez_std  non-wear_flag_mean  light_mean   light_std  \
0          NaN         NaN                 NaN         NaN         NaN
1          NaN         NaN                 NaN         NaN         NaN
2   -10.580416   42.947170            0.000000    42.29631  208.168976
3   -55.630768   50.303635            0.655708    16.77198   95.327438
4          NaN         NaN                 NaN         NaN         NaN

   battery_voltage_mean  battery_voltage_std  time_of_day_mean  weekday_mean  \
0                   NaN                  NaN               NaN           NaN
1                   NaN                  NaN               NaN           NaN
2           4053.578857           112.404037      5.046215e+13      4.470182
3           3838.189453           155.573868      4.321212e+13      3.909848
4                   NaN                  NaN               NaN           NaN

   quarter_mean  relative_date_PCIAT_mean  fourier_X_real  fourier_X_imag  \
0           NaN                       NaN             NaN             NaN
1           NaN                       NaN             NaN             NaN
2           3.0                 53.201683    -2653.531675   -1.083578e-14
3           3.0                 79.435593     4103.865122   -1.680035e+03
4           NaN                       NaN             NaN             NaN

   fourier_Y_real  fourier_Y_imag  fourier_Z_real  fourier_Z_imag  \
0             NaN             NaN             NaN             NaN
1             NaN             NaN             NaN             NaN
2      705.803060     -207.446683     1995.401835   -5.684342e-15
3    14385.071425     3120.930726   -26765.674436   -8.731149e-12
4             NaN             NaN             NaN             NaN

   activity_during_day  activity_during_night  activity_ratio_day_night  \
0                  NaN                    NaN                       NaN
```

```
1                  NaN                  NaN                        NaN
2                  0.0          2053.305176                        0.0
3                  0.0          4727.518555                        0.0
4                  NaN                  NaN                        NaN

   non_wear_proportion  sedentary_proportion  light_activity_proportion  \
0                  NaN                   NaN                        NaN
1                  NaN                   NaN                        NaN
2             0.000000              0.792453                   0.198131
3             0.655708              0.978501                   0.021171
4                  NaN                   NaN                        NaN

   moderate_activity_proportion  vigorous_activity_proportion
0                           NaN                           NaN
1                           NaN                           NaN
2                      0.007870                      0.001546
3                      0.000288                      0.000040
4                           NaN                           NaN
```

[44]:

```
# Define the list of actigraphy feature columns
actigraphy_columns = [
    'X_mean', 'X_std', 'X_min', 'X_max', 'Y_mean', 'Y_std', 'Y_min', 'Y_max',
    'Z_mean', 'Z_std', 'Z_min', 'Z_max', 'enmo_mean', 'enmo_std', 'anglez_mean',
    'anglez_std', 'non-wear_flag_mean', 'light_mean', 'light_std',␣
  ↪'battery_voltage_mean',
    'battery_voltage_std', 'time_of_day_mean', 'weekday_mean', 'quarter_mean',
    'relative_date_PCIAT_mean', 'fourier_X_real', 'fourier_X_imag',␣
  ↪'fourier_Y_real',
    'fourier_Y_imag', 'fourier_Z_real', 'fourier_Z_imag', 'activity_during_day',
    'activity_during_night', 'activity_ratio_day_night', 'non_wear_proportion',
    'sedentary_proportion', 'light_activity_proportion',␣
  ↪'moderate_activity_proportion',
    'vigorous_activity_proportion'
]

# Fill NaN values in actigraphy columns with zeros if all actigraphy columns␣
  ↪are NaN in a row
merged_train_df[actigraphy_columns] = merged_train_df[actigraphy_columns].
  ↪apply(lambda row: row.fillna(0) if row.isnull().all() else row, axis=1)

# Create actigraphy_present column
# If all actigraphy columns have 0 values for a given row, set␣
  ↪actigraphy_present to 0, otherwise 1
merged_train_df['actigraphy_present'] = merged_train_df[actigraphy_columns].
  ↪apply(lambda row: 0 if (row == 0).all() else 1, axis=1)
```

```python
# Display the first few rows to verify the results
merged_train_df.head()
```

[44]:

```
         id  Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0  00008ff9                   0.559547              5.0              0.0
1  00105258                   0.543858             10.0              1.0
2  00115b9f                   0.598983              9.0              0.0
3  001f3379                   0.475815             13.0              1.0
4  00abe655                   0.559547             11.0              0.0

   CGAS-Season  CGAS-CGAS_Score  Physical-Season  Physical-BMI  \
0     0.527092             51.0         0.526350     16.877316
1     0.538294             71.0         0.526350     16.648696
2     0.538294             71.0         0.516702     18.292347
3     0.527092             50.0         0.516702     22.279952
4     0.515994             66.0         0.471466     18.269958

   Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0        46.000000        50.800000                     23.154724
1        56.500000        75.600000                     25.755813
2        56.000000        81.600000                     30.237799
3        59.500000       112.200000                     33.812417
4        55.568097        83.565122                     26.013867

   Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
0              69.273255           86.942701            111.584992
1              65.000000           94.000000            117.000000
2              60.000000           97.000000            117.000000
3              60.000000           73.000000            102.000000
4              69.317251           84.110761            116.626216

   Fitness_Endurance-Season  Fitness_Endurance-Max_Stage  \
0                  0.562024                     4.661823
1                  0.416561                     5.000000
2                  0.419733                     6.000000
3                  0.562024                     3.781508
4                  0.562024                     4.861046

   Fitness_Endurance-Time_Mins  Fitness_Endurance-Time_Sec  FGC-Season  \
0                     5.902407                   30.190591    0.539792
1                     7.000000                   33.000000    0.539792
2                     9.000000                   37.000000    0.531007
3                     6.496836                   24.040340    0.531007
4                     8.019364                   28.006224    0.552717

   FGC-FGC_CU  FGC-FGC_CU_Zone  FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  \
0    0.000000         0.000000     12.457159           1.861092     7.105427
```

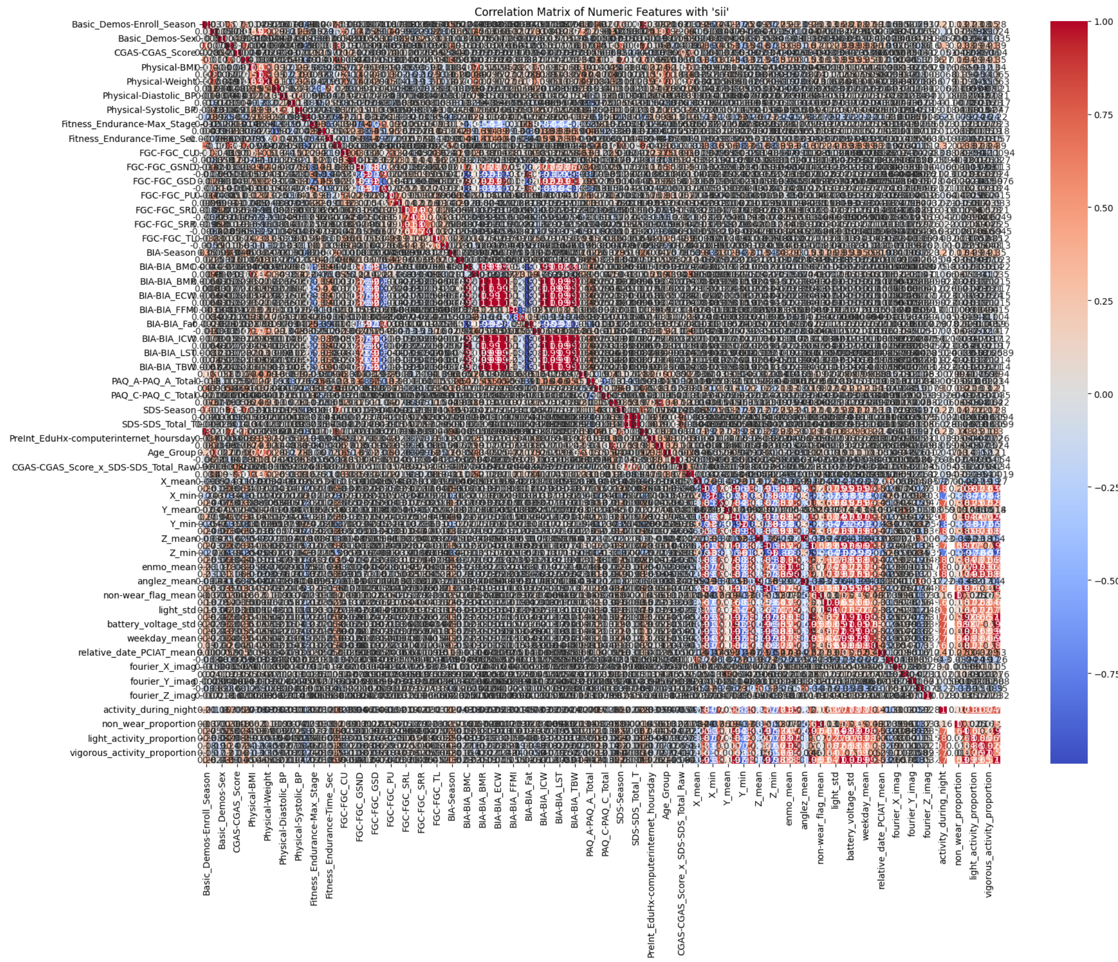
```
1   20.000000         1.000000     10.200000           1.000000    14.700000
2   18.000000         1.000000     15.489000           1.797822    17.595300
3   12.000000         0.000000     16.500000           2.000000    17.900000
4   10.701686         0.455832     19.270922           1.683190    19.325191

   FGC-FGC_GSD_Zone  FGC-FGC_PU  FGC-FGC_PU_Zone  FGC-FGC_SRL  \
0          2.104217    0.000000         0.000000     7.000000
1          2.000000    7.000000         1.000000    10.000000
2          2.176744    5.000000         0.000000     7.000000
3          2.000000    6.000000         0.000000    10.000000
4          1.630797    5.096538         0.302039     8.751021

   FGC-FGC_SRL_Zone  FGC-FGC_SRR  FGC-FGC_SRR_Zone  FGC-FGC_TL  \
0           0.00000      6.00000          0.000000    6.000000
1           1.00000     10.00000          1.000000    5.000000
2           0.00000      7.00000          0.000000    7.000000
3           1.00000     11.00000          1.000000    8.000000
4           0.68659      8.92129          0.690292    9.021283

   FGC-FGC_TL_Zone  BIA-Season  BIA-BIA_Activity_Level_num  BIA-BIA_BMC  \
0         1.000000    0.534475                    2.000000     2.668550
1         0.000000    0.490251                    2.658187     7.660622
2         1.000000    0.566098                    3.000000     3.841910
3         0.000000    0.566098                    2.000000     4.330360
4         0.723361    0.490251                    2.517398     7.660620

   BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0    16.879200   932.498000  1492.000000     8.255980    41.586200
1    18.540298  1229.028902  2004.524112    20.415826    73.170824
2    18.294300  1131.430000  1923.440000    15.592500    62.775700
3    30.186500  1330.970000  1996.450000    30.212400    84.028500
4    18.540307  1229.029934  2020.016337    20.415797    73.170847

   BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0     13.817700     3.061430     9.213770           1.000000    24.434900
1     14.810861     3.729442     5.294715           1.531391    32.829003
2     14.074000     4.220330    18.824300           2.000000    30.404100
3     16.687700    13.498800    67.971500           2.000000    32.914100
4     14.810861     3.729442    10.512782           1.644971    32.829002

   BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  \
0     8.895360    38.917700    19.541300    32.690900      0.480263
1    19.926022    65.510201    42.760486    53.244807      0.480263
2    16.779000    58.933800    26.479800    45.996600      0.480263
3    20.902000    79.698200    35.380400    63.126500      0.480263
4    19.926031    65.510206    29.831805    53.244809      0.480263
```

```
   PAQ_A-PAQ_A_Total  PAQ_C-Season  PAQ_C-PAQ_C_Total  PCIAT-PCIAT_Total  \
0           1.773888      0.460443            2.68491               55.0
1           1.991617      0.544231            2.17000               28.0
2           1.978289      0.716160            2.45100               44.0
3           1.968692      0.579612            4.11000               34.0
4           2.110197      0.716160            1.10000               10.0

   SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  PreInt_EduHx-Season  \
0    0.467407          42.186222        59.348885             0.564223
1    0.515460          38.000000        54.000000             0.534033
2    0.512539          31.000000        45.000000             0.602178
3    0.512539          40.000000        56.000000             0.486120
4    0.611293          42.000000        59.000000             0.564223

   PreInt_EduHx-computerinternet_hoursday  sii  Age_Group  PCIAT_Cluster  \
0                                     3.0  2.0   0.443345            5.0
1                                     2.0  0.0   0.362030            7.0
2                                     0.0  1.0   0.362030            4.0
3                                     0.0  1.0   0.855589            3.0
4                                     0.0  0.0   0.855589            3.0

   CGAS-CGAS_Score_x_SDS-SDS_Total_Raw  CGAS-CGAS_Score_x_Physical-BMI  \
0                          2156.462944                      860.743100
1                          2698.000000                     1182.057420
2                          2201.000000                     1298.756633
3                          2000.000000                     1113.997599
4                          2772.000000                     1206.032260

     X_mean     X_std     X_min     X_max    Y_mean     Y_std     Y_min  \
0  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
2 -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702 -2.905339
3 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812 -1.522690
4  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000

      Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  enmo_std  \
0  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
1  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
2  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388  0.106351
3  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926  0.024331
4  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000

   anglez_mean  anglez_std  non-wear_flag_mean  light_mean   light_std  \
0     0.000000    0.000000            0.000000     0.00000    0.000000
1     0.000000    0.000000            0.000000     0.00000    0.000000
2   -10.580416   42.947170            0.000000    42.29631  208.168976
3   -55.630768   50.303635            0.655708    16.77198   95.327438
```

```
4    0.000000    0.000000             0.000000     0.00000    0.000000

   battery_voltage_mean  battery_voltage_std  time_of_day_mean  weekday_mean  \
0              0.000000             0.000000      0.000000e+00      0.000000
1              0.000000             0.000000      0.000000e+00      0.000000
2           4053.578857           112.404037      5.046215e+13      4.470182
3           3838.189453           155.573868      4.321212e+13      3.909848
4              0.000000             0.000000      0.000000e+00      0.000000

   quarter_mean  relative_date_PCIAT_mean  fourier_X_real  fourier_X_imag  \
0           0.0                  0.000000        0.000000    0.000000e+00
1           0.0                  0.000000        0.000000    0.000000e+00
2           3.0                 53.201683    -2653.531675   -1.083578e-14
3           3.0                 79.435593     4103.865122   -1.680035e+03
4           0.0                  0.000000        0.000000    0.000000e+00

   fourier_Y_real  fourier_Y_imag  fourier_Z_real  fourier_Z_imag  \
0        0.000000        0.000000        0.000000    0.000000e+00
1        0.000000        0.000000        0.000000    0.000000e+00
2      705.803060     -207.446683     1995.401835   -5.684342e-15
3    14385.071425     3120.930726   -26765.674436   -8.731149e-12
4        0.000000        0.000000        0.000000    0.000000e+00

   activity_during_day  activity_during_night  activity_ratio_day_night  \
0                  0.0               0.000000                       0.0
1                  0.0               0.000000                       0.0
2                  0.0            2053.305176                       0.0
3                  0.0            4727.518555                       0.0
4                  0.0               0.000000                       0.0

   non_wear_proportion  sedentary_proportion  light_activity_proportion  \
0             0.000000              0.000000                   0.000000
1             0.000000              0.000000                   0.000000
2             0.000000              0.792453                   0.198131
3             0.655708              0.978501                   0.021171
4             0.000000              0.000000                   0.000000

   moderate_activity_proportion  vigorous_activity_proportion  \
0                      0.000000                      0.000000
1                      0.000000                      0.000000
2                      0.007870                      0.001546
3                      0.000288                      0.000040
4                      0.000000                      0.000000

   actigraphy_present
0                   0
1                   0
```

```
2                    1
3                    1
4                    0
```

[45]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Filter the merged DataFrame to include only numeric columns for the␣
↪correlation matrix
numeric_df = merged_train_df.select_dtypes(include=[float, int])

# Display correlation matrix for the new features and sii
correlation_matrix = numeric_df.corr()
plt.figure(figsize=(20, 15))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Correlation Matrix of Numeric Features with 'sii'")
plt.show()

# Calculate basic statistics for actigraphy features grouped by 'sii'
grouped_stats = numeric_df.groupby(merged_train_df['sii']).agg(['mean', 'std',␣
↪'min', 'max'])
print(grouped_stats.head())

# Plot histograms or box plots for key features against 'sii'
key_features = ['X_mean', 'Y_mean', 'Z_mean', 'enmo_mean', 'non-wear_flag_mean',
                'activity_ratio_day_night', 'sedentary_proportion',␣
↪'vigorous_activity_proportion']

for feature in key_features:
    if feature in numeric_df.columns:  # Check if the feature is in the numeric␣
↪DataFrame
        plt.figure(figsize=(10, 6))
        sns.boxplot(x='sii', y=feature, data=merged_train_df)
        plt.title(f'Distribution of {feature} by sii')
        plt.show()
```



	Basic_Demos-Enroll_Season				Basic_Demos-Age
	mean	std	min	max	mean
sii					
-0.760696	0.598983	NaN	0.598983	0.598983	19.0
-0.314713	0.475815	NaN	0.475815	0.475815	22.0
-0.312714	0.559547	NaN	0.559547	0.559547	17.0
-0.231091	0.598983	NaN	0.598983	0.598983	15.0
-0.144216	0.475815	NaN	0.475815	0.475815	13.0

				Basic_Demos-Sex				CGAS-Season	
	std	min	max	mean	std	min	max	mean	std
sii									
-0.760696	NaN	19.0	19.0	0.0	NaN	0.0	0.0	0.515994	NaN
-0.314713	NaN	22.0	22.0	1.0	NaN	1.0	1.0	0.515994	NaN
-0.312714	NaN	17.0	17.0	1.0	NaN	1.0	1.0	0.538294	NaN
-0.231091	NaN	15.0	15.0	0.0	NaN	0.0	0.0	0.538294	NaN
-0.144216	NaN	13.0	13.0	1.0	NaN	1.0	1.0	0.581124	NaN

```
                     CGAS-CGAS_Score                 Physical-Season  \
             min       max            mean std   min   max            mean
sii
-0.760696  0.515994  0.515994           58.0 NaN  58.0  58.0        0.601470
-0.314713  0.515994  0.515994           75.0 NaN  75.0  75.0        0.523758
-0.312714  0.538294  0.538294           76.0 NaN  76.0  76.0        0.516702
-0.231091  0.538294  0.538294           60.0 NaN  60.0  60.0        0.526350
-0.144216  0.581124  0.581124           72.0 NaN  72.0  72.0        0.516702

                                  Physical-BMI                          \
          std       min       max          mean std         min         max
sii
-0.760696 NaN  0.601470  0.601470    23.840162 NaN  23.840162  23.840162
-0.314713 NaN  0.523758  0.523758    26.056757 NaN  26.056757  26.056757
-0.312714 NaN  0.516702  0.516702    29.789226 NaN  29.789226  29.789226
-0.231091 NaN  0.526350  0.526350    16.864219 NaN  16.864219  16.864219
-0.144216 NaN  0.516702  0.516702    17.455101 NaN  17.455101  17.455101

          Physical-Height                Physical-Weight                  \
                     mean std   min   max            mean std    min    max
sii
-0.760696            72.0 NaN  72.0  72.0           175.8 NaN  175.8  175.8
-0.314713            65.0 NaN  65.0  65.0           156.6 NaN  156.6  156.6
-0.312714            67.0 NaN  67.0  67.0           190.2 NaN  190.2  190.2
-0.231091            69.0 NaN  69.0  69.0           114.2 NaN  114.2  114.2
-0.144216            68.0 NaN  68.0  68.0           114.8 NaN  114.8  114.8

          Physical-Waist_Circumference                Physical-Diastolic_BP  \
                                  mean std   min   max                  mean
sii
-0.760696                         36.0 NaN  36.0  36.0                  71.0
-0.314713                         35.0 NaN  35.0  35.0                  80.0
-0.312714                         34.0 NaN  34.0  34.0                  63.0
-0.231091                         27.0 NaN  27.0  27.0                  53.0
-0.144216                         27.0 NaN  27.0  27.0                  80.0

                          Physical-HeartRate                  \
          std   min   max              mean std   min   max
sii
-0.760696 NaN  71.0  71.0              71.0 NaN  71.0  71.0
-0.314713 NaN  80.0  80.0              78.0 NaN  78.0  78.0
-0.312714 NaN  63.0  63.0              66.0 NaN  66.0  66.0
-0.231091 NaN  53.0  53.0              93.0 NaN  93.0  93.0
-0.144216 NaN  80.0  80.0              92.0 NaN  92.0  92.0

          Physical-Systolic_BP                  Fitness_Endurance-Season      \
                          mean std    min    max                      mean std
sii
```

```
-0.760696               135.0 NaN  135.0  135.0                0.562024 NaN
-0.314713               129.0 NaN  129.0  129.0                0.562024 NaN
-0.312714               108.0 NaN  108.0  108.0                0.562024 NaN
-0.231091               110.0 NaN  110.0  110.0                0.562024 NaN
-0.144216               127.0 NaN  127.0  127.0                0.562024 NaN

                                  Fitness_Endurance-Max_Stage               \
               min       max                           mean std       min
sii
-0.760696  0.562024  0.562024                      4.435789 NaN  4.435789
-0.314713  0.562024  0.562024                      4.198903 NaN  4.198903
-0.312714  0.562024  0.562024                      5.799169 NaN  5.799169
-0.231091  0.562024  0.562024                      5.240235 NaN  5.240235
-0.144216  0.562024  0.562024                      5.430333 NaN  5.430333

                    Fitness_Endurance-Time_Mins                             \
               max                        mean std       min       max
sii
-0.760696  4.435789                   7.143487 NaN  7.143487  7.143487
-0.314713  4.198903                   6.624388 NaN  6.624388  6.624388
-0.312714  5.799169                   4.528173 NaN  4.528173  4.528173
-0.231091  5.240235                   7.535246 NaN  7.535246  7.535246
-0.144216  5.430333                   7.165710 NaN  7.165710  7.165710

          Fitness_Endurance-Time_Sec                           FGC-Season      \
                                mean std        min        max       mean std
sii
-0.760696                  20.126764 NaN  20.126764  20.126764   0.472214 NaN
-0.314713                  29.121412 NaN  29.121412  29.121412   0.472214 NaN
-0.312714                  29.726658 NaN  29.726658  29.726658   0.472214 NaN
-0.231091                  25.809110 NaN  25.809110  25.809110   0.472214 NaN
-0.144216                  22.464845 NaN  22.464845  22.464845   0.472214 NaN

                              FGC-FGC_CU                         \
               min       max       mean std        min        max
sii
-0.760696  0.472214  0.472214  22.225245 NaN  22.225245  22.225245
-0.314713  0.472214  0.472214  23.222260 NaN  23.222260  23.222260
-0.312714  0.472214  0.472214  28.654766 NaN  28.654766  28.654766
-0.231091  0.472214  0.472214  24.285854 NaN  24.285854  24.285854
-0.144216  0.472214  0.472214  23.167038 NaN  23.167038  23.167038

          FGC-FGC_CU_Zone                        FGC-FGC_GSND               \
                    mean std       min       max        mean std        min
sii
-0.760696      -0.035583 NaN -0.035583 -0.035583    30.826376 NaN  30.826376
-0.314713       0.127326 NaN  0.127326  0.127326    30.610218 NaN  30.610218
-0.312714       0.202144 NaN  0.202144  0.202144    28.937106 NaN  28.937106
```

-0.231091	0.443701	NaN	0.443701	0.443701	24.795915	NaN	24.795915
-0.144216	0.614095	NaN	0.614095	0.614095	21.959201	NaN	21.959201

		FGC-FGC_GSND_Zone				FGC-FGC_GSD \
	max	mean	std	min	max	mean
sii						
-0.760696	30.826376	1.258827	NaN	1.258827	1.258827	38.088400
-0.314713	30.610218	1.490439	NaN	1.490439	1.490439	33.254582
-0.312714	28.937106	2.084074	NaN	2.084074	2.084074	36.604509
-0.231091	24.795915	1.790724	NaN	1.790724	1.790724	27.957054
-0.144216	21.959201	2.129340	NaN	2.129340	2.129340	24.217935

				FGC-FGC_GSD_Zone			\
	std	min	max	mean	std	min	max
sii							
-0.760696	NaN	38.088400	38.088400	1.921265	NaN	1.921265	1.921265
-0.314713	NaN	33.254582	33.254582	1.410570	NaN	1.410570	1.410570
-0.312714	NaN	36.604509	36.604509	3.872646	NaN	3.872646	3.872646
-0.231091	NaN	27.957054	27.957054	2.182940	NaN	2.182940	2.182940
-0.144216	NaN	24.217935	24.217935	2.889933	NaN	2.889933	2.889933

	FGC-FGC_PU				FGC-FGC_PU_Zone		\
	mean	std	min	max	mean	std	min
sii							
-0.760696	13.365519	NaN	13.365519	13.365519	0.267869	NaN	0.267869
-0.314713	13.597431	NaN	13.597431	13.597431	0.482732	NaN	0.482732
-0.312714	15.138957	NaN	15.138957	15.138957	0.502361	NaN	0.502361
-0.231091	12.005235	NaN	12.005235	12.005235	0.472910	NaN	0.472910
-0.144216	10.818556	NaN	10.818556	10.818556	0.470968	NaN	0.470968

		FGC-FGC_SRL				FGC-FGC_SRL_Zone	\
	max	mean	std	min	max	mean	std
sii							
-0.760696	0.267869	8.149241	NaN	8.149241	8.149241	0.650092	NaN
-0.314713	0.482732	9.043808	NaN	9.043808	9.043808	0.455634	NaN
-0.312714	0.502361	9.540310	NaN	9.540310	9.540310	0.757537	NaN
-0.231091	0.472910	8.610586	NaN	8.610586	8.610586	0.705078	NaN
-0.144216	0.470968	9.170370	NaN	9.170370	9.170370	0.661416	NaN

			FGC-FGC_SRR			\
	min	max	mean	std	min	max
sii						
-0.760696	0.650092	0.650092	8.391354	NaN	8.391354	8.391354
-0.314713	0.455634	0.455634	9.473053	NaN	9.473053	9.473053
-0.312714	0.757537	0.757537	9.786045	NaN	9.786045	9.786045
-0.231091	0.705078	0.705078	8.655687	NaN	8.655687	8.655687
-0.144216	0.661416	0.661416	9.035880	NaN	9.035880	9.035880

	FGC-FGC_SRR_Zone				FGC-FGC_TL		\
	mean	std	min	max	mean	std	min
sii							
-0.760696	0.590861	NaN	0.590861	0.590861	10.102656	NaN	10.102656
-0.314713	0.597148	NaN	0.597148	0.597148	10.514400	NaN	10.514400
-0.312714	0.545723	NaN	0.545723	0.545723	9.906480	NaN	9.906480
-0.231091	0.620721	NaN	0.620721	0.620721	10.413822	NaN	10.413822
-0.144216	0.554199	NaN	0.554199	0.554199	10.305656	NaN	10.305656

		FGC-FGC_TL_Zone				BIA-Season	\
	max	mean	std	min	max	mean	std
sii							
-0.760696	10.102656	0.497869	NaN	0.497869	0.497869	0.553902	NaN
-0.314713	10.514400	0.493068	NaN	0.493068	0.493068	0.490251	NaN
-0.312714	9.906480	0.547992	NaN	0.547992	0.547992	0.490251	NaN
-0.231091	10.413822	0.658714	NaN	0.658714	0.658714	0.490251	NaN
-0.144216	10.305656	0.763497	NaN	0.763497	0.763497	0.490251	NaN

			BIA-BIA_Activity_Level_num		\
	min	max	mean	std	min
sii					
-0.760696	0.553902	0.553902	2.000000	NaN	2.000000
-0.314713	0.490251	0.490251	3.115355	NaN	3.115355
-0.312714	0.490251	0.490251	2.777763	NaN	2.777763
-0.231091	0.490251	0.490251	2.762212	NaN	2.762212
-0.144216	0.490251	0.490251	2.965925	NaN	2.965925

		BIA-BIA_BMC				BIA-BIA_BMI	\
	max	mean	std	min	max	mean	std
sii							
-0.760696	2.000000	7.268830	NaN	7.268830	7.268830	23.842800	NaN
-0.314713	3.115355	7.660618	NaN	7.660618	7.660618	18.540288	NaN
-0.312714	2.777763	7.660596	NaN	7.660596	7.660596	18.540284	NaN
-0.231091	2.762212	7.660639	NaN	7.660639	7.660639	18.540296	NaN
-0.144216	2.965925	7.660641	NaN	7.660641	7.660641	18.540292	NaN

			BIA-BIA_BMR			\
	min	max	mean	std	min	max
sii						
-0.760696	23.842800	23.842800	1717.260000	NaN	1717.260000	1717.260000
-0.314713	18.540288	18.540288	1229.031350	NaN	1229.031350	1229.031350
-0.312714	18.540284	18.540284	1229.031514	NaN	1229.031514	1229.031514
-0.231091	18.540296	18.540296	1229.029790	NaN	1229.029790	1229.029790
-0.144216	18.540292	18.540292	1229.029818	NaN	1229.029818	1229.029818

	BIA-BIA_DEE				BIA-BIA_ECW	\
	mean	std	min	max	mean	std
sii						

-0.760696	2747.620000	NaN	2747.620000	2747.620000	37.770700	NaN
-0.314713	2113.409105	NaN	2113.409105	2113.409105	20.415860	NaN
-0.312714	2115.855098	NaN	2115.855098	2115.855098	20.415905	NaN
-0.231091	2098.439401	NaN	2098.439401	2098.439401	20.415835	NaN
-0.144216	2066.290026	NaN	2066.290026	2066.290026	20.415860	NaN

			BIA-BIA_FFM			\
	min	max	mean	std	min	max
sii						
-0.760696	37.770700	37.770700	125.174000	NaN	125.174000	125.174000
-0.314713	20.415860	20.415860	73.170878	NaN	73.170878	73.170878
-0.312714	20.415905	20.415905	73.170890	NaN	73.170890	73.170890
-0.231091	20.415835	20.415835	73.170803	NaN	73.170803	73.170803
-0.144216	20.415860	20.415860	73.170800	NaN	73.170800	73.170800

	BIA-BIA_FFMI				BIA-BIA_FMI		\
	mean	std	min	max	mean	std	min
sii							
-0.760696	16.976700	NaN	16.976700	16.976700	6.866100	NaN	6.866100
-0.314713	14.810860	NaN	14.810860	14.810860	3.729442	NaN	3.729442
-0.312714	14.810854	NaN	14.810854	14.810854	3.729443	NaN	3.729443
-0.231091	14.810860	NaN	14.810860	14.810860	3.729442	NaN	3.729442
-0.144216	14.810860	NaN	14.810860	14.810860	3.729442	NaN	3.729442

		BIA-BIA_Fat				BIA-BIA_Frame_num \
	max	mean	std	min	max	mean
sii						
-0.760696	6.866100	50.625800	NaN	50.625800	50.625800	3.000000
-0.314713	3.729442	71.754220	NaN	71.754220	71.754220	2.081289
-0.312714	3.729443	96.382362	NaN	96.382362	96.382362	2.674607
-0.231091	3.729442	41.429030	NaN	41.429030	41.429030	1.792602
-0.144216	3.729442	41.112235	NaN	41.112235	41.112235	1.898075

				BIA-BIA_ICW			\
	std	min	max	mean	std	min	max
sii							
-0.760696	NaN	3.000000	3.000000	53.428200	NaN	53.428200	53.428200
-0.314713	NaN	2.081289	2.081289	32.828959	NaN	32.828959	32.828959
-0.312714	NaN	2.674607	2.674607	32.828933	NaN	32.828933	32.828933
-0.231091	NaN	1.792602	1.792602	32.828970	NaN	32.828970	32.828970
-0.144216	NaN	1.898075	1.898075	32.828968	NaN	32.828968	32.828968

	BIA-BIA_LDM				BIA-BIA_LST		\
	mean	std	min	max	mean	std	min
sii							
-0.760696	33.975300	NaN	33.975300	33.975300	117.905000	NaN	117.905000
-0.314713	19.926104	NaN	19.926104	19.926104	65.510302	NaN	65.510302
-0.312714	19.926159	NaN	19.926159	19.926159	65.510335	NaN	65.510335

```
-0.231091   19.926087 NaN  19.926087  19.926087   65.510248 NaN   65.510248
-0.144216   19.926082 NaN  19.926082  19.926082   65.510251 NaN   65.510251

                      BIA-BIA_SMM                                  BIA-BIA_TBW      \
                  max         mean std         min         max         mean std
sii
-0.760696  117.905000   61.174600 NaN  61.174600  61.174600   91.198900 NaN
-0.314713   65.510302   42.618569 NaN  42.618569  42.618569   53.244852 NaN
-0.312714   65.510335   39.912373 NaN  39.912373  39.912373   53.244874 NaN
-0.231091   65.510248   28.474838 NaN  28.474838  28.474838   53.244834 NaN
-0.144216   65.510251   41.602053 NaN  41.602053  41.602053   53.244835 NaN

                                PAQ_A-Season                            \
                 min         max         mean std       min       max
sii
-0.760696  91.198900  91.198900     0.480263 NaN  0.480263  0.480263
-0.314713  53.244852  53.244852     0.480263 NaN  0.480263  0.480263
-0.312714  53.244874  53.244874     1.017185 NaN  1.017185  1.017185
-0.231091  53.244834  53.244834     1.017185 NaN  1.017185  1.017185
-0.144216  53.244835  53.244835     0.938948 NaN  0.938948  0.938948

          PAQ_A-PAQ_A_Total                         PAQ_C-Season      \
                       mean std       min       max         mean std
sii
-0.760696          2.261004 NaN  2.261004  2.261004     0.460443 NaN
-0.314713          2.212036 NaN  2.212036  2.212036     0.460443 NaN
-0.312714          1.370000 NaN  1.370000  1.370000     0.460443 NaN
-0.231091          1.820000 NaN  1.820000  1.820000     0.460443 NaN
-0.144216          1.690000 NaN  1.690000  1.690000     0.460443 NaN

                              PAQ_C-PAQ_C_Total                              \
                min       max              mean std        min        max
sii
-0.760696  0.460443  0.460443          2.571950 NaN   2.571950   2.571950
-0.314713  0.460443  0.460443          1.473492 NaN   1.473492   1.473492
-0.312714  0.460443  0.460443         22.778263 NaN  22.778263  22.778263
-0.231091  0.460443  0.460443          9.452195 NaN   9.452195   9.452195
-0.144216  0.460443  0.460443         10.474885 NaN  10.474885  10.474885

          PCIAT-PCIAT_Total                         SDS-Season                \
                       mean std       min       max         mean std       min
sii
-0.760696          8.770492 NaN  8.770492  8.770492   0.467407 NaN  0.467407
-0.314713          8.770492 NaN  8.770492  8.770492   0.467407 NaN  0.467407
-0.312714          8.770492 NaN  8.770492  8.770492   0.512539 NaN  0.512539
-0.231091          8.770492 NaN  8.770492  8.770492   0.467407 NaN  0.467407
-0.144216          8.770492 NaN  8.770492  8.770492   0.467407 NaN  0.467407
```

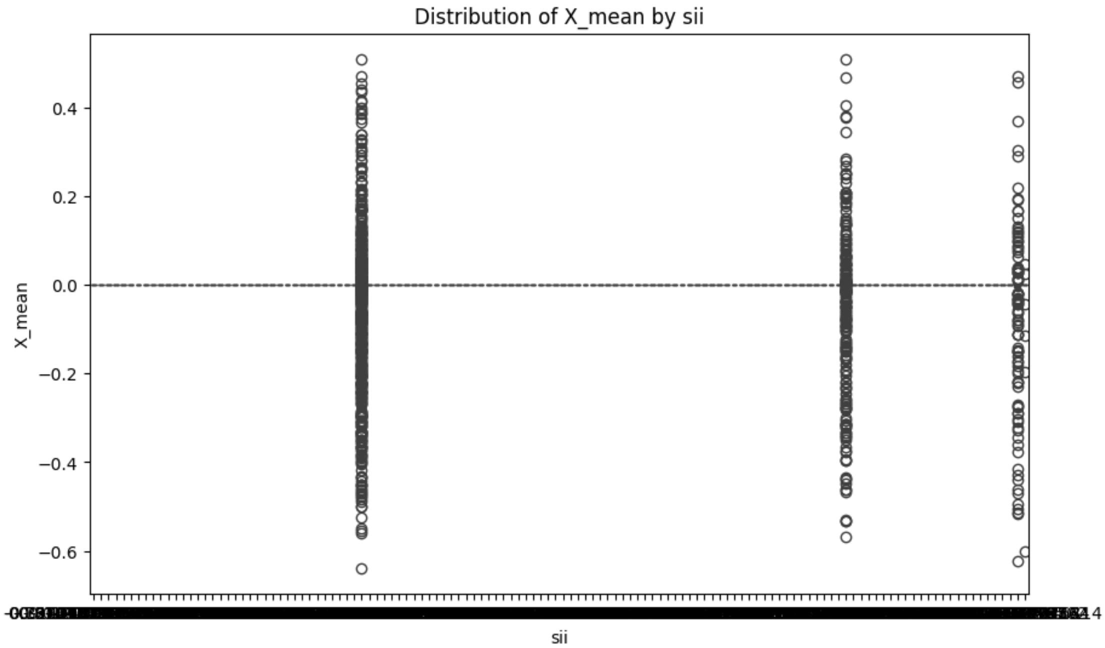
```
                    SDS-SDS_Total_Raw                            \
                max              mean std        min        max
sii
-0.760696  0.467407         41.230775 NaN  41.230775  41.230775
-0.314713  0.467407         39.890807 NaN  39.890807  39.890807
-0.312714  0.512539         35.000000 NaN  35.000000  35.000000
-0.231091  0.467407         41.250020 NaN  41.250020  41.250020
-0.144216  0.467407         40.134626 NaN  40.134626  40.134626

          SDS-SDS_Total_T                          PreInt_EduHx-Season      \
                     mean std        min        max                mean std
sii
-0.760696       57.417806 NaN  57.417806  57.417806            0.602178 NaN
-0.314713       56.442597 NaN  56.442597  56.442597            0.486120 NaN
-0.312714       50.000000 NaN  50.000000  50.000000            0.457262 NaN
-0.231091       57.776421 NaN  57.776421  57.776421            0.602178 NaN
-0.144216       56.311069 NaN  56.311069  56.311069            0.486120 NaN

                              PreInt_EduHx-computerinternet_hoursday      \
                min       max                                 mean std
sii
-0.760696  0.602178  0.602178                             3.000000 NaN
-0.314713  0.486120  0.486120                             3.000000 NaN
-0.312714  0.457262  0.457262                             2.814750 NaN
-0.231091  0.602178  0.602178                             1.475126 NaN
-0.144216  0.486120  0.486120                             1.318443 NaN

                                    sii                        Age_Group      \
                min       max      mean std       min       max      mean std
sii
-0.760696  3.000000  3.000000 -0.760696 NaN -0.760696 -0.760696  0.973770 NaN
-0.314713  3.000000  3.000000 -0.314713 NaN -0.314713 -0.314713  0.499136 NaN
-0.312714  2.814750  2.814750 -0.312714 NaN -0.312714 -0.312714  0.973770 NaN
-0.231091  1.475126  1.475126 -0.231091 NaN -0.231091 -0.231091  0.855589 NaN
-0.144216  1.318443  1.318443 -0.144216 NaN -0.144216 -0.144216  0.855589 NaN

                              PCIAT_Cluster                \
                min       max          mean std  min  max
sii
-0.760696  0.973770  0.973770           0.0 NaN  0.0  0.0
-0.314713  0.499136  0.499136           0.0 NaN  0.0  0.0
-0.312714  0.973770  0.973770           0.0 NaN  0.0  0.0
-0.231091  0.855589  0.855589           0.0 NaN  0.0  0.0
-0.144216  0.855589  0.855589           0.0 NaN  0.0  0.0

          CGAS-CGAS_Score_x_SDS-SDS_Total_Raw                            \
                                         mean std          min          max
sii
```

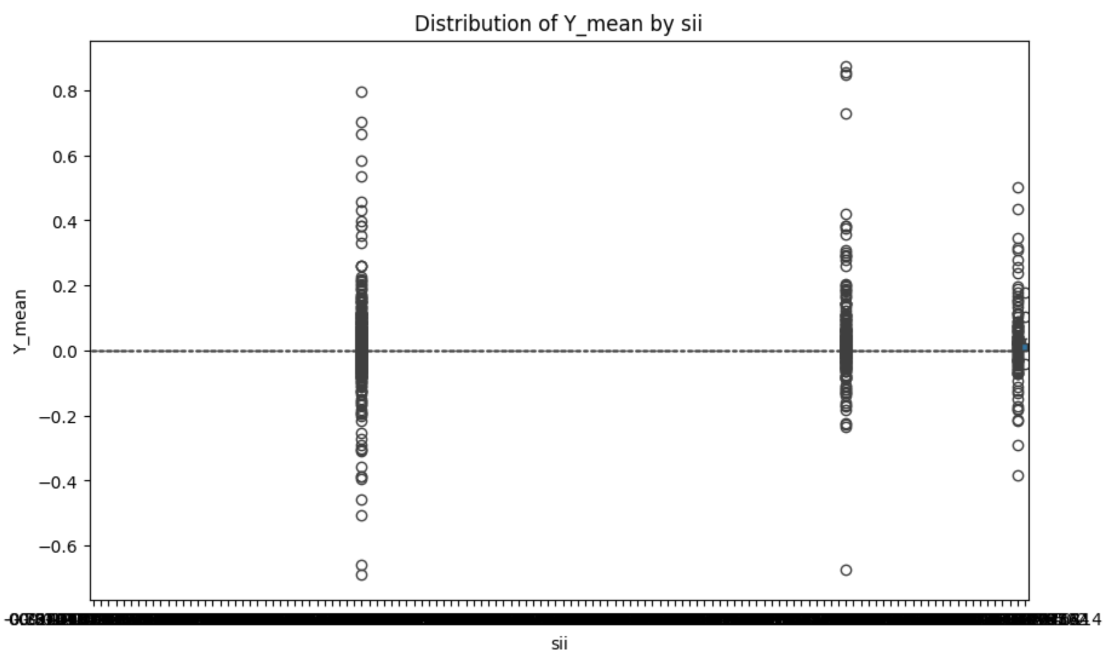
```
-0.760696                       2363.884081 NaN  2363.884081  2363.884081
-0.314713                       3018.565478 NaN  3018.565478  3018.565478
-0.312714                       2660.000000 NaN  2660.000000  2660.000000
-0.231091                       2447.538500 NaN  2447.538500  2447.538500
-0.144216                       2864.127094 NaN  2864.127094  2864.127094

          CGAS-CGAS_Score_x_Physical-BMI                                X_mean  \
                                    mean std          min          max   mean
sii
-0.760696                    1382.729398 NaN  1382.729398  1382.729398    0.0
-0.314713                    1954.256805 NaN  1954.256805  1954.256805    0.0
-0.312714                    2263.981151 NaN  2263.981151  2263.981151    0.0
-0.231091                    1011.853132 NaN  1011.853132  1011.853132    0.0
-0.144216                    1256.767284 NaN  1256.767284  1256.767284    0.0

                        X_std               X_min               X_max      \
          std  min  max  mean std  min  max  mean std  min  max  mean std
sii
-0.760696 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN
-0.314713 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN
-0.312714 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN
-0.231091 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN
-0.144216 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN

                    Y_mean               Y_std               Y_min           \
           min  max   mean std  min  max  mean std  min  max  mean std  min
sii
-0.760696  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0
-0.314713  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0
-0.312714  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0
-0.231091  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0
-0.144216  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0   0.0 NaN  0.0

               Y_max               Z_mean               Z_std                \
           max  mean std  min  max   mean std  min  max  mean std  min  max
sii
-0.760696  0.0   0.0 NaN  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0
-0.314713  0.0   0.0 NaN  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0
-0.312714  0.0   0.0 NaN  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0
-0.231091  0.0   0.0 NaN  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0
-0.144216  0.0   0.0 NaN  0.0  0.0    0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0

          Z_min               Z_max               enmo_mean                \
           mean std  min  max  mean std  min  max      mean std  min  max
sii
-0.760696   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0       0.0 NaN  0.0  0.0
-0.314713   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0       0.0 NaN  0.0  0.0
-0.312714   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0       0.0 NaN  0.0  0.0
```

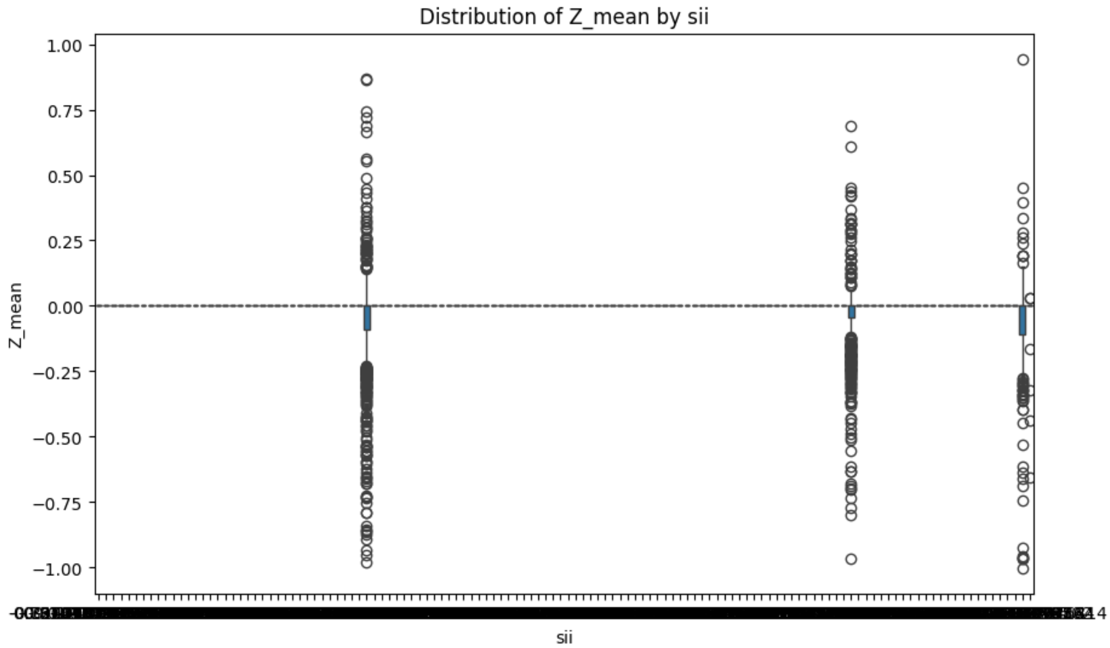
```
-0.231091   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0       0.0 NaN  0.0  0.0
-0.144216   0.0 NaN  0.0  0.0   0.0 NaN  0.0  0.0       0.0 NaN  0.0  0.0

          enmo_std               anglez_mean               anglez_std      \
              mean std  min  max        mean std  min  max       mean std
sii
-0.760696      0.0 NaN  0.0  0.0         0.0 NaN  0.0  0.0        0.0 NaN
-0.314713      0.0 NaN  0.0  0.0         0.0 NaN  0.0  0.0        0.0 NaN
-0.312714      0.0 NaN  0.0  0.0         0.0 NaN  0.0  0.0        0.0 NaN
-0.231091      0.0 NaN  0.0  0.0         0.0 NaN  0.0  0.0        0.0 NaN
-0.144216      0.0 NaN  0.0  0.0         0.0 NaN  0.0  0.0        0.0 NaN

                    non-wear_flag_mean               light_mean                \
           min  max               mean std  min  max       mean std  min  max
sii
-0.760696  0.0  0.0                0.0 NaN  0.0  0.0        0.0 NaN  0.0  0.0
-0.314713  0.0  0.0                0.0 NaN  0.0  0.0        0.0 NaN  0.0  0.0
-0.312714  0.0  0.0                0.0 NaN  0.0  0.0        0.0 NaN  0.0  0.0
-0.231091  0.0  0.0                0.0 NaN  0.0  0.0        0.0 NaN  0.0  0.0
-0.144216  0.0  0.0                0.0 NaN  0.0  0.0        0.0 NaN  0.0  0.0

          light_std               battery_voltage_mean                \
               mean std  min  max                 mean std  min  max
sii
-0.760696       0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.314713       0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.312714       0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.231091       0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.144216       0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0

          battery_voltage_std               time_of_day_mean                \
                         mean std  min  max             mean std  min  max
sii
-0.760696                 0.0 NaN  0.0  0.0              0.0 NaN  0.0  0.0
-0.314713                 0.0 NaN  0.0  0.0              0.0 NaN  0.0  0.0
-0.312714                 0.0 NaN  0.0  0.0              0.0 NaN  0.0  0.0
-0.231091                 0.0 NaN  0.0  0.0              0.0 NaN  0.0  0.0
-0.144216                 0.0 NaN  0.0  0.0              0.0 NaN  0.0  0.0

          weekday_mean               quarter_mean                \
                  mean std  min  max         mean std  min  max
sii
-0.760696          0.0 NaN  0.0  0.0          0.0 NaN  0.0  0.0
-0.314713          0.0 NaN  0.0  0.0          0.0 NaN  0.0  0.0
-0.312714          0.0 NaN  0.0  0.0          0.0 NaN  0.0  0.0
-0.231091          0.0 NaN  0.0  0.0          0.0 NaN  0.0  0.0
-0.144216          0.0 NaN  0.0  0.0          0.0 NaN  0.0  0.0
```

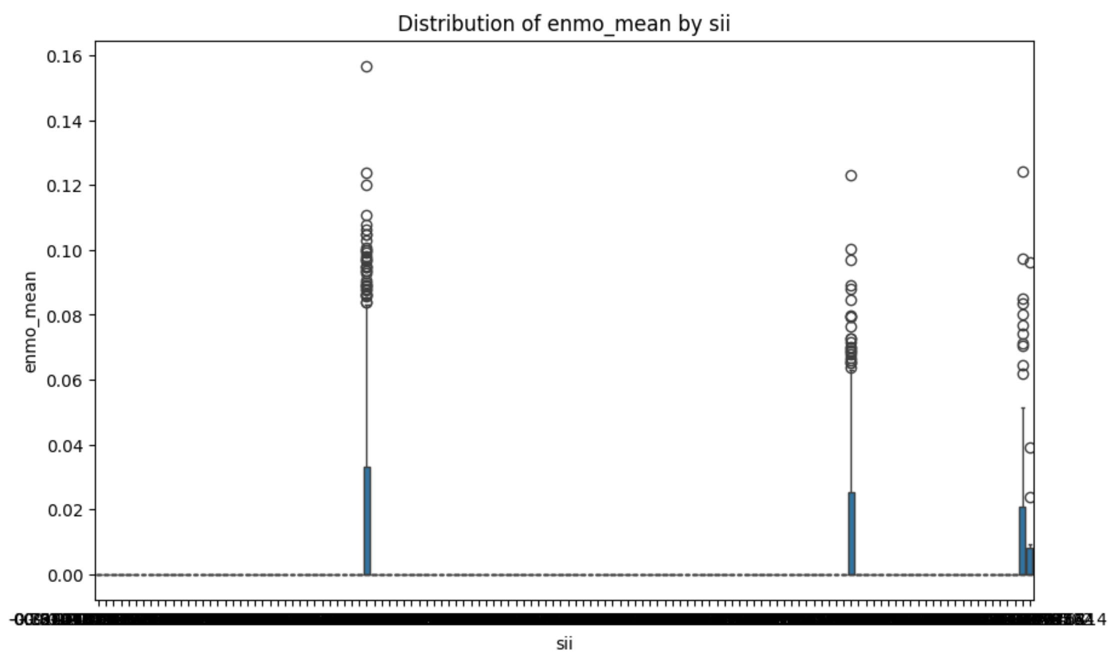
```
         relative_date_PCIAT_mean                fourier_X_real                \
                             mean std  min  max            mean std  min  max
sii
-0.760696                     0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.314713                     0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.312714                     0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.231091                     0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.144216                     0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0

         fourier_X_imag                fourier_Y_real                \
                   mean std  min  max            mean std  min  max
sii
-0.760696           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.314713           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.312714           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.231091           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.144216           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0

         fourier_Y_imag                fourier_Z_real                \
                   mean std  min  max            mean std  min  max
sii
-0.760696           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.314713           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.312714           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.231091           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0
-0.144216           0.0 NaN  0.0  0.0             0.0 NaN  0.0  0.0

         fourier_Z_imag                activity_during_day                \
                   mean std  min  max                 mean std  min  max
sii
-0.760696           0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.314713           0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.312714           0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.231091           0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0
-0.144216           0.0 NaN  0.0  0.0                  0.0 NaN  0.0  0.0

         activity_during_night                activity_ratio_day_night      \
                          mean std  min  max                      mean std
sii
-0.760696                  0.0 NaN  0.0  0.0                       0.0 NaN
-0.314713                  0.0 NaN  0.0  0.0                       0.0 NaN
-0.312714                  0.0 NaN  0.0  0.0                       0.0 NaN
-0.231091                  0.0 NaN  0.0  0.0                       0.0 NaN
-0.144216                  0.0 NaN  0.0  0.0                       0.0 NaN

                   non_wear_proportion                sedentary_proportion  \
          min  max                mean std  min  max                  mean
sii
```

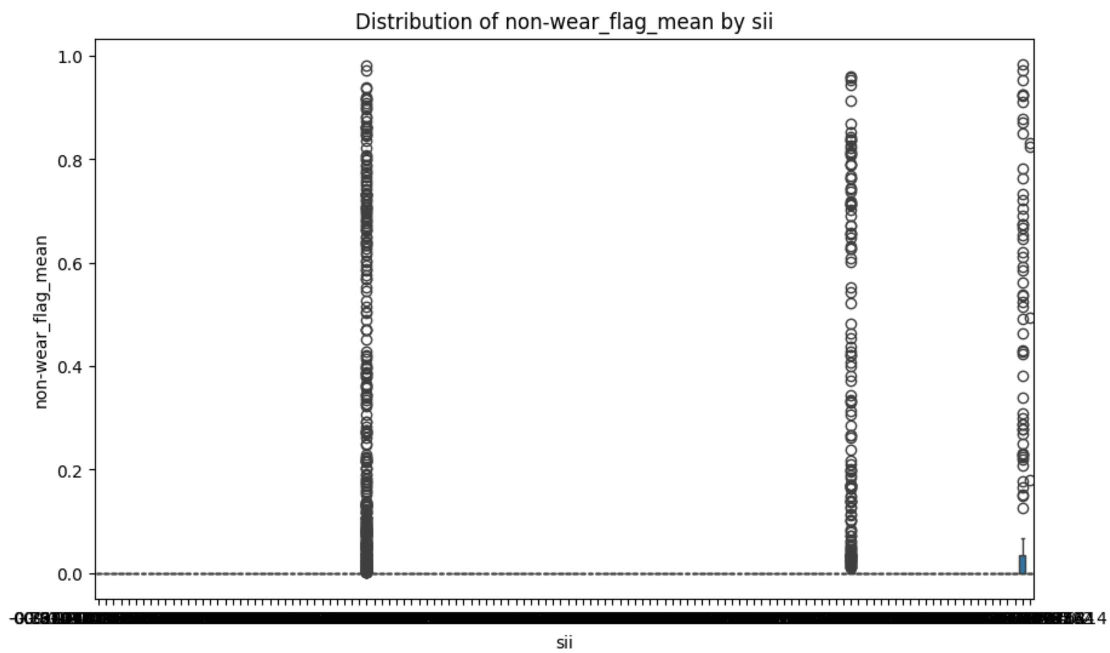
```
-0.760696  0.0  0.0                 0.0 NaN  0.0  0.0                  0.0
-0.314713  0.0  0.0                 0.0 NaN  0.0  0.0                  0.0
-0.312714  0.0  0.0                 0.0 NaN  0.0  0.0                  0.0
-0.231091  0.0  0.0                 0.0 NaN  0.0  0.0                  0.0
-0.144216  0.0  0.0                 0.0 NaN  0.0  0.0                  0.0

                        light_activity_proportion                \
          std  min  max                      mean std  min  max
sii
-0.760696 NaN  0.0  0.0                       0.0 NaN  0.0  0.0
-0.314713 NaN  0.0  0.0                       0.0 NaN  0.0  0.0
-0.312714 NaN  0.0  0.0                       0.0 NaN  0.0  0.0
-0.231091 NaN  0.0  0.0                       0.0 NaN  0.0  0.0
-0.144216 NaN  0.0  0.0                       0.0 NaN  0.0  0.0

          moderate_activity_proportion                \
                                  mean std  min  max
sii
-0.760696                          0.0 NaN  0.0  0.0
-0.314713                          0.0 NaN  0.0  0.0
-0.312714                          0.0 NaN  0.0  0.0
-0.231091                          0.0 NaN  0.0  0.0
-0.144216                          0.0 NaN  0.0  0.0

          vigorous_activity_proportion               actigraphy_present      \
                                  mean std  min  max               mean std
sii
-0.760696                          0.0 NaN  0.0  0.0                0.0 NaN
-0.314713                          0.0 NaN  0.0  0.0                0.0 NaN
-0.312714                          0.0 NaN  0.0  0.0                0.0 NaN
-0.231091                          0.0 NaN  0.0  0.0                0.0 NaN
-0.144216                          0.0 NaN  0.0  0.0                0.0 NaN


          min max
sii
-0.760696   0   0
-0.314713   0   0
-0.312714   0   0
-0.231091   0   0
-0.144216   0   0
```

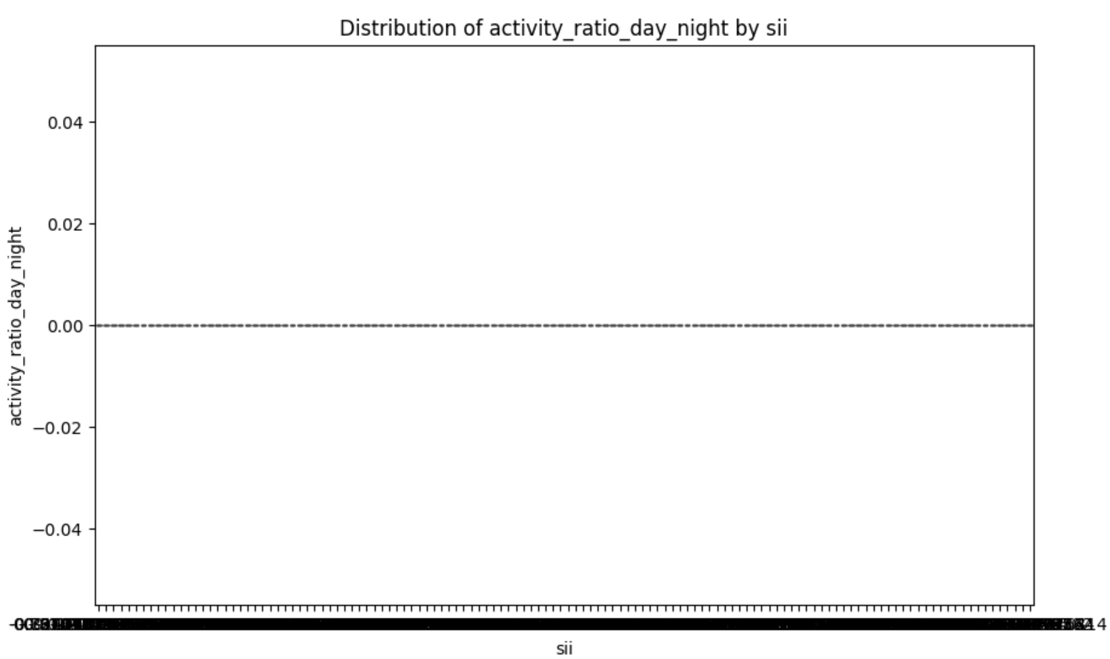

Distribution of X_mean by sii
X_mean
0.4
0.2
0.0
−0.2
−0.4
−0.6
sii

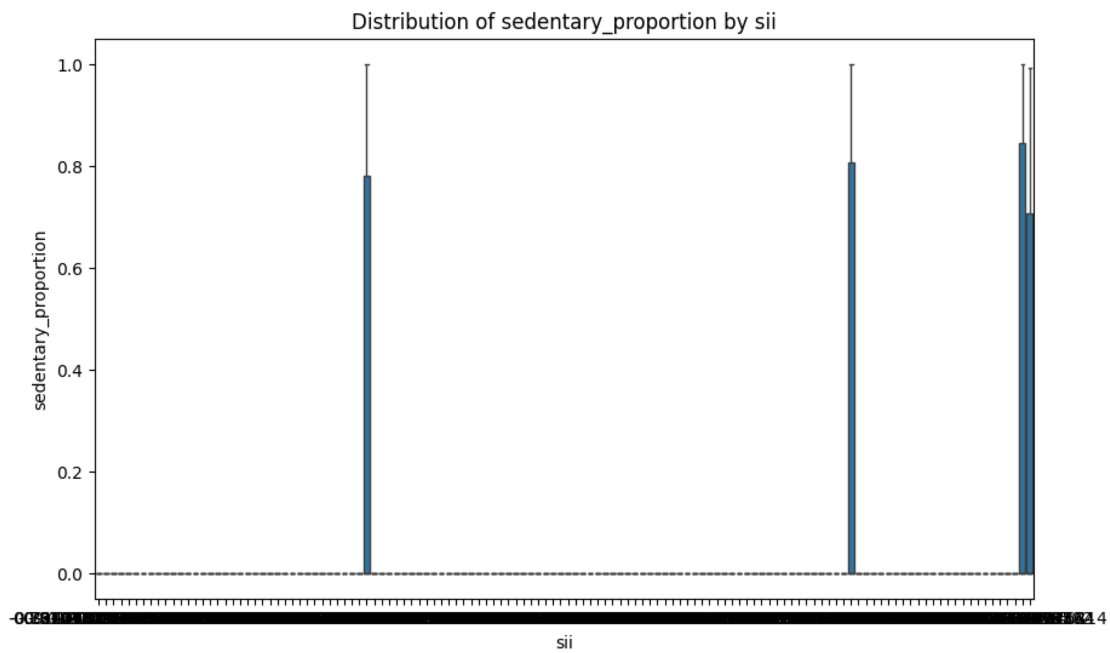

Distribution of Y_mean by sii
Y_mean
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
sii

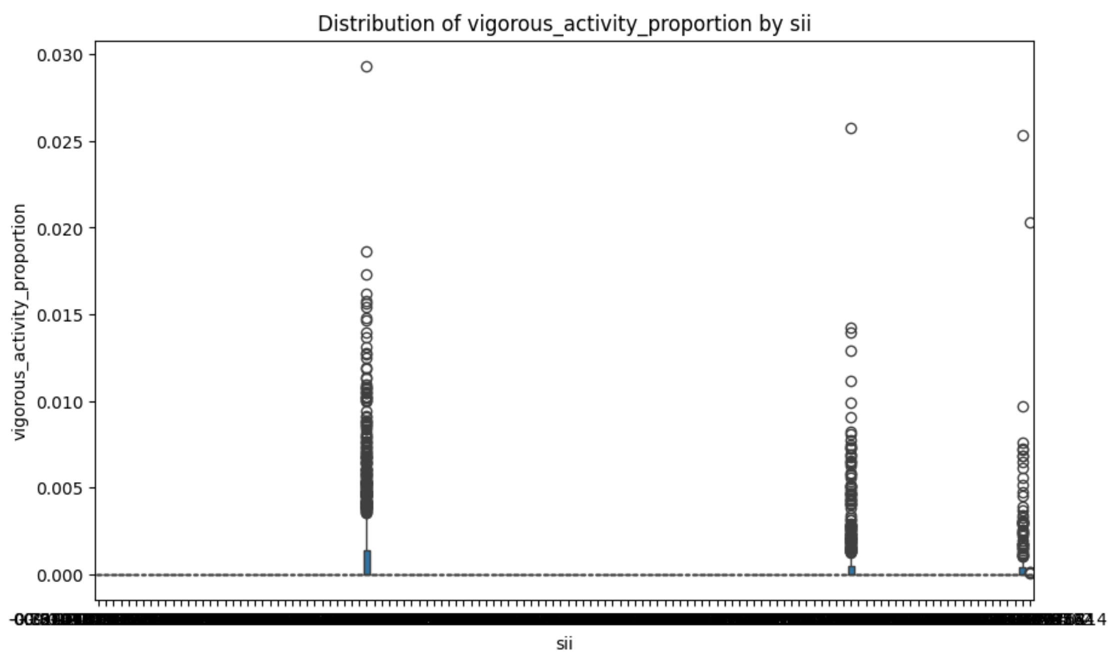

Distribution of Z_mean by sii
Z_mean
sii
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Distribution of enmo_mean by sii
enmo_mean
sii
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00


Distribution of non-wear_flag_mean by sii
non-wear_flag_mean
1.0
0.8
0.6
0.4
0.2
0.0
sii


Distribution of activity_ratio_day_night by sii
activity_ratio_day_night
0.04
0.02
0.00
−0.02
−0.04
sii





[46]:
```
merged_train_df.shape
```

[46]: (3045, 105)

[47]:
```
from imblearn.over_sampling import SMOTE
# Convert the 'sii' column to integer type to ensure it is treated as␣
↪categorical
merged_train_df['sii'] = merged_train_df['sii'].astype(int)

merged_train_df['id_numeric'] = pd.factorize(merged_train_df['id'])[0]
id_mapping = dict(zip(merged_train_df['id_numeric'], merged_train_df['id']))

# Extract the features and target variable 'sii' from the merged dataframe
features = merged_train_df.drop(columns=['sii', 'id'])  # Exclude 'sii' and␣
↪'id' columns
target = merged_train_df['sii']

# Apply SMOTE to balance the 'sii' classes
smote = SMOTE(random_state=42)
features_resampled, target_resampled = smote.fit_resample(features, target)

# Create a new DataFrame for the oversampled dataset
oversampled_df = pd.DataFrame(features_resampled, columns=features.columns)
oversampled_df['sii'] = target_resampled



# Display the class distribution after SMOTE
print("Class distribution after SMOTE oversampling:")
print(oversampled_df['sii'].value_counts())
```

```
Class distribution after SMOTE oversampling:
sii
2    2053
0    2053
1    2053
3    2053
Name: count, dtype: int64
```

```
/var/tmp/ipykernel_7796/2866646705.py:18: PerformanceWarning: DataFrame is
highly fragmented.  This is usually the result of calling `frame.insert` many
times, which has poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe =
frame.copy()`
  oversampled_df['sii'] = target_resampled
```

[48]:
```
oversampled_df['id'] = oversampled_df['id_numeric'].map(id_mapping)
oversampled_df = oversampled_df.drop(columns = ['id_numeric'])
```

```
/var/tmp/ipykernel_7796/3770237215.py:1: PerformanceWarning: DataFrame is highly
fragmented.  This is usually the result of calling `frame.insert` many times,
which has poor performance.  Consider joining all columns at once using
pd.concat(axis=1) instead. To get a de-fragmented frame, use `newframe =
```

```
frame.copy()`
  oversampled_df['id'] = oversampled_df['id_numeric'].map(id_mapping)
```

[49]:
```
oversampled_df.head()
```

[49]:
```
   Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  CGAS-Season  \
0                   0.559547              5.0              0.0     0.527092
1                   0.543858             10.0              1.0     0.538294
2                   0.598983              9.0              0.0     0.538294
3                   0.475815             13.0              1.0     0.527092
4                   0.559547             11.0              0.0     0.515994

   CGAS-CGAS_Score  Physical-Season  Physical-BMI  Physical-Height  \
0             51.0         0.526350     16.877316        46.000000
1             71.0         0.526350     16.648696        56.500000
2             71.0         0.516702     18.292347        56.000000
3             50.0         0.516702     22.279952        59.500000
4             66.0         0.471466     18.269958        55.568097

   Physical-Weight  Physical-Waist_Circumference  Physical-Diastolic_BP  \
0        50.800000                     23.154724              69.273255
1        75.600000                     25.755813              65.000000
2        81.600000                     30.237799              60.000000
3       112.200000                     33.812417              60.000000
4        83.565122                     26.013867              69.317251

   Physical-HeartRate  Physical-Systolic_BP  Fitness_Endurance-Season  \
0           86.942701            111.584992                  0.562024
1           94.000000            117.000000                  0.416561
2           97.000000            117.000000                  0.419733
3           73.000000            102.000000                  0.562024
4           84.110761            116.626216                  0.562024

   Fitness_Endurance-Max_Stage  Fitness_Endurance-Time_Mins  \
0                     4.661823                     5.902407
1                     5.000000                     7.000000
2                     6.000000                     9.000000
3                     3.781508                     6.496836
4                     4.861046                     8.019364

   Fitness_Endurance-Time_Sec  FGC-Season  FGC-FGC_CU  FGC-FGC_CU_Zone  \
0                   30.190591    0.539792    0.000000         0.000000
1                   33.000000    0.539792   20.000000         1.000000
2                   37.000000    0.531007   18.000000         1.000000
3                   24.040340    0.531007   12.000000         0.000000
4                   28.006224    0.552717   10.701686         0.455832
```

```
   FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  FGC-FGC_GSD_Zone  FGC-FGC_PU  \
0     12.457159           1.861092     7.105427          2.104217    0.000000
1     10.200000           1.000000    14.700000          2.000000    7.000000
2     15.489000           1.797822    17.595300          2.176744    5.000000
3     16.500000           2.000000    17.900000          2.000000    6.000000
4     19.270922           1.683190    19.325191          1.630797    5.096538

   FGC-FGC_PU_Zone  FGC-FGC_SRL  FGC-FGC_SRL_Zone  FGC-FGC_SRR  \
0         0.000000     7.000000           0.00000      6.00000
1         1.000000    10.000000           1.00000     10.00000
2         0.000000     7.000000           0.00000      7.00000
3         0.000000    10.000000           1.00000     11.00000
4         0.302039     8.751021           0.68659      8.92129

   FGC-FGC_SRR_Zone  FGC-FGC_TL  FGC-FGC_TL_Zone  BIA-Season  \
0          0.000000    6.000000         1.000000    0.534475
1          1.000000    5.000000         0.000000    0.490251
2          0.000000    7.000000         1.000000    0.566098
3          1.000000    8.000000         0.000000    0.566098
4          0.690292    9.021283         0.723361    0.490251

   BIA-BIA_Activity_Level_num  BIA-BIA_BMC  BIA-BIA_BMI  BIA-BIA_BMR  \
0                    2.000000     2.668550    16.879200   932.498000
1                    2.658187     7.660622    18.540298  1229.028902
2                    3.000000     3.841910    18.294300  1131.430000
3                    2.000000     4.330360    30.186500  1330.970000
4                    2.517398     7.660620    18.540307  1229.029934

   BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  BIA-BIA_FFMI  BIA-BIA_FMI  \
0  1492.000000     8.255980    41.586200     13.817700     3.061430
1  2004.524112    20.415826    73.170824     14.810861     3.729442
2  1923.440000    15.592500    62.775700     14.074000     4.220330
3  1996.450000    30.212400    84.028500     16.687700    13.498800
4  2020.016337    20.415797    73.170847     14.810861     3.729442

   BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  BIA-BIA_LDM  BIA-BIA_LST  \
0     9.213770           1.000000    24.434900     8.895360    38.917700
1     5.294715           1.531391    32.829003    19.926022    65.510201
2    18.824300           2.000000    30.404100    16.779000    58.933800
3    67.971500           2.000000    32.914100    20.902000    79.698200
4    10.512782           1.644971    32.829002    19.926031    65.510206

   BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  PAQ_A-PAQ_A_Total  PAQ_C-Season  \
0    19.541300    32.690900      0.480263           1.773888      0.460443
1    42.760486    53.244807      0.480263           1.991617      0.544231
2    26.479800    45.996600      0.480263           1.978289      0.716160
3    35.380400    63.126500      0.480263           1.968692      0.579612
```

```
4   29.831805   53.244809     0.480263           2.110197      0.716160

   PAQ_C-PAQ_C_Total  PCIAT-PCIAT_Total  SDS-Season  SDS-SDS_Total_Raw  \
0            2.68491               55.0    0.467407          42.186222
1            2.17000               28.0    0.515460          38.000000
2            2.45100               44.0    0.512539          31.000000
3            4.11000               34.0    0.512539          40.000000
4            1.10000               10.0    0.611293          42.000000

   SDS-SDS_Total_T  PreInt_EduHx-Season  \
0        59.348885             0.564223
1        54.000000             0.534033
2        45.000000             0.602178
3        56.000000             0.486120
4        59.000000             0.564223

   PreInt_EduHx-computerinternet_hoursday  Age_Group  PCIAT_Cluster  \
0                                     3.0   0.443345            5.0
1                                     2.0   0.362030            7.0
2                                     0.0   0.362030            4.0
3                                     0.0   0.855589            3.0
4                                     0.0   0.855589            3.0

   CGAS-CGAS_Score_x_SDS-SDS_Total_Raw  CGAS-CGAS_Score_x_Physical-BMI  \
0                          2156.462944                      860.743100
1                          2698.000000                     1182.057420
2                          2201.000000                     1298.756633
3                          2000.000000                     1113.997599
4                          2772.000000                     1206.032260

     X_mean     X_std     X_min     X_max    Y_mean     Y_std     Y_min  \
0  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
2 -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702 -2.905339
3 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812 -1.522690
4  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000

      Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  enmo_std  \
0  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
1  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
2  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388  0.106351
3  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926  0.024331
4  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000

   anglez_mean  anglez_std  non-wear_flag_mean  light_mean   light_std  \
0     0.000000    0.000000            0.000000     0.00000    0.000000
1     0.000000    0.000000            0.000000     0.00000    0.000000
```

```
2  -10.580416  42.947170            0.000000   42.29631  208.168976
3  -55.630768  50.303635            0.655708   16.77198   95.327438
4    0.000000   0.000000            0.000000    0.00000    0.000000

   battery_voltage_mean  battery_voltage_std  time_of_day_mean  weekday_mean  \
0              0.000000             0.000000      0.000000e+00      0.000000
1              0.000000             0.000000      0.000000e+00      0.000000
2           4053.578857           112.404037      5.046215e+13      4.470182
3           3838.189453           155.573868      4.321212e+13      3.909848
4              0.000000             0.000000      0.000000e+00      0.000000

   quarter_mean  relative_date_PCIAT_mean  fourier_X_real  fourier_X_imag  \
0           0.0                  0.000000        0.000000    0.000000e+00
1           0.0                  0.000000        0.000000    0.000000e+00
2           3.0                 53.201683    -2653.531675   -1.083578e-14
3           3.0                 79.435593     4103.865122   -1.680035e+03
4           0.0                  0.000000        0.000000    0.000000e+00

   fourier_Y_real  fourier_Y_imag  fourier_Z_real  fourier_Z_imag  \
0        0.000000        0.000000        0.000000    0.000000e+00
1        0.000000        0.000000        0.000000    0.000000e+00
2      705.803060     -207.446683     1995.401835   -5.684342e-15
3    14385.071425     3120.930726   -26765.674436   -8.731149e-12
4        0.000000        0.000000        0.000000    0.000000e+00

   activity_during_day  activity_during_night  activity_ratio_day_night  \
0                  0.0               0.000000                       0.0
1                  0.0               0.000000                       0.0
2                  0.0            2053.305176                       0.0
3                  0.0            4727.518555                       0.0
4                  0.0               0.000000                       0.0

   non_wear_proportion  sedentary_proportion  light_activity_proportion  \
0             0.000000              0.000000                   0.000000
1             0.000000              0.000000                   0.000000
2             0.000000              0.792453                   0.198131
3             0.655708              0.978501                   0.021171
4             0.000000              0.000000                   0.000000

   moderate_activity_proportion  vigorous_activity_proportion  \
0                      0.000000                      0.000000
1                      0.000000                      0.000000
2                      0.007870                      0.001546
3                      0.000288                      0.000040
4                      0.000000                      0.000000

   actigraphy_present  sii        id
```

```
0                    0    2  00008ff9
1                    0    0  00105258
2                    1    1  00115b9f
3                    1    1  001f3379
4                    0    0  00abe655
```

[50]:
```python
list(oversampled_df.columns)
```

[50]:
```
['Basic_Demos-Enroll_Season',
 'Basic_Demos-Age',
 'Basic_Demos-Sex',
 'CGAS-Season',
 'CGAS-CGAS_Score',
 'Physical-Season',
 'Physical-BMI',
 'Physical-Height',
 'Physical-Weight',
 'Physical-Waist_Circumference',
 'Physical-Diastolic_BP',
 'Physical-HeartRate',
 'Physical-Systolic_BP',
 'Fitness_Endurance-Season',
 'Fitness_Endurance-Max_Stage',
 'Fitness_Endurance-Time_Mins',
 'Fitness_Endurance-Time_Sec',
 'FGC-Season',
 'FGC-FGC_CU',
 'FGC-FGC_CU_Zone',
 'FGC-FGC_GSND',
 'FGC-FGC_GSND_Zone',
 'FGC-FGC_GSD',
 'FGC-FGC_GSD_Zone',
 'FGC-FGC_PU',
 'FGC-FGC_PU_Zone',
 'FGC-FGC_SRL',
 'FGC-FGC_SRL_Zone',
 'FGC-FGC_SRR',
 'FGC-FGC_SRR_Zone',
 'FGC-FGC_TL',
 'FGC-FGC_TL_Zone',
 'BIA-Season',
 'BIA-BIA_Activity_Level_num',
 'BIA-BIA_BMC',
 'BIA-BIA_BMI',
 'BIA-BIA_BMR',
 'BIA-BIA_DEE',
 'BIA-BIA_ECW',
```

```
'BIA-BIA_FFM',
'BIA-BIA_FFMI',
'BIA-BIA_FMI',
'BIA-BIA_Fat',
'BIA-BIA_Frame_num',
'BIA-BIA_ICW',
'BIA-BIA_LDM',
'BIA-BIA_LST',
'BIA-BIA_SMM',
'BIA-BIA_TBW',
'PAQ_A-Season',
'PAQ_A-PAQ_A_Total',
'PAQ_C-Season',
'PAQ_C-PAQ_C_Total',
'PCIAT-PCIAT_Total',
'SDS-Season',
'SDS-SDS_Total_Raw',
'SDS-SDS_Total_T',
'PreInt_EduHx-Season',
'PreInt_EduHx-computerinternet_hoursday',
'Age_Group',
'PCIAT_Cluster',
'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw',
'CGAS-CGAS_Score_x_Physical-BMI',
'X_mean',
'X_std',
'X_min',
'X_max',
'Y_mean',
'Y_std',
'Y_min',
'Y_max',
'Z_mean',
'Z_std',
'Z_min',
'Z_max',
'enmo_mean',
'enmo_std',
'anglez_mean',
'anglez_std',
'non-wear_flag_mean',
'light_mean',
'light_std',
'battery_voltage_mean',
'battery_voltage_std',
'time_of_day_mean',
'weekday_mean',
```

```
 'quarter_mean',
 'relative_date_PCIAT_mean',
 'fourier_X_real',
 'fourier_X_imag',
 'fourier_Y_real',
 'fourier_Y_imag',
 'fourier_Z_real',
 'fourier_Z_imag',
 'activity_during_day',
 'activity_during_night',
 'activity_ratio_day_night',
 'non_wear_proportion',
 'sedentary_proportion',
 'light_activity_proportion',
 'moderate_activity_proportion',
 'vigorous_activity_proportion',
 'actigraphy_present',
 'sii',
 'id']
```

[51]:
```
oversampled_df.shape
```

[51]:
```
(8212, 105)
```

[52]:
```
!pip install tensorflow
```

```
Requirement already satisfied: tensorflow in /opt/conda/lib/python3.10/site-
packages (2.17.0)
Requirement already satisfied: absl-py>=1.0.0 in /opt/conda/lib/python3.10/site-
packages (from tensorflow) (2.1.0)
Requirement already satisfied: astunparse>=1.6.0 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (1.6.3)
Requirement already satisfied: flatbuffers>=24.3.25 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (24.3.25)
Requirement already satisfied: gast!=0.5.0,!=0.5.1,!=0.5.2,>=0.2.1 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (0.6.0)
Requirement already satisfied: google-pasta>=0.1.1 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (0.2.0)
Requirement already satisfied: h5py>=3.10.0 in /opt/conda/lib/python3.10/site-
packages (from tensorflow) (3.12.1)
Requirement already satisfied: libclang>=13.0.0 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (18.1.1)
Requirement already satisfied: ml-dtypes<0.5.0,>=0.3.1 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (0.4.1)
Requirement already satisfied: opt-einsum>=2.3.2 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (3.4.0)
Requirement already satisfied: packaging in /opt/conda/lib/python3.10/site-
packages (from tensorflow) (24.1)
```

```
Requirement already satisfied:
protobuf!=4.21.0,!=4.21.1,!=4.21.2,!=4.21.3,!=4.21.4,!=4.21.5,<5.0.0dev,>=3.20.3
in /opt/conda/lib/python3.10/site-packages (from tensorflow) (3.20.3)
Requirement already satisfied: requests<3,>=2.21.0 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (2.32.3)
Requirement already satisfied: setuptools in /opt/conda/lib/python3.10/site-
packages (from tensorflow) (73.0.1)
Requirement already satisfied: six>=1.12.0 in /opt/conda/lib/python3.10/site-
packages (from tensorflow) (1.16.0)
Requirement already satisfied: termcolor>=1.1.0 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (2.4.0)
Requirement already satisfied: typing-extensions>=3.6.6 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (4.12.2)
Requirement already satisfied: wrapt>=1.11.0 in /opt/conda/lib/python3.10/site-
packages (from tensorflow) (1.16.0)
Requirement already satisfied: grpcio<2.0,>=1.24.3 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (1.66.1)
Requirement already satisfied: tensorboard<2.18,>=2.17 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (2.17.1)
Requirement already satisfied: keras>=3.2.0 in /opt/conda/lib/python3.10/site-
packages (from tensorflow) (3.6.0)
Requirement already satisfied: tensorflow-io-gcs-filesystem>=0.23.1 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (0.37.1)
Requirement already satisfied: numpy<2.0.0,>=1.23.5 in
/opt/conda/lib/python3.10/site-packages (from tensorflow) (1.26.4)
Requirement already satisfied: wheel<1.0,>=0.23.0 in
/opt/conda/lib/python3.10/site-packages (from astunparse>=1.6.0->tensorflow)
(0.44.0)
Requirement already satisfied: rich in /opt/conda/lib/python3.10/site-packages
(from keras>=3.2.0->tensorflow) (13.8.1)
Requirement already satisfied: namex in /opt/conda/lib/python3.10/site-packages
(from keras>=3.2.0->tensorflow) (0.0.8)
Requirement already satisfied: optree in /opt/conda/lib/python3.10/site-packages
(from keras>=3.2.0->tensorflow) (0.13.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/conda/lib/python3.10/site-packages (from requests<3,>=2.21.0->tensorflow)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /opt/conda/lib/python3.10/site-
packages (from requests<3,>=2.21.0->tensorflow) (3.8)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/opt/conda/lib/python3.10/site-packages (from requests<3,>=2.21.0->tensorflow)
(1.26.20)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/conda/lib/python3.10/site-packages (from requests<3,>=2.21.0->tensorflow)
(2024.8.30)
Requirement already satisfied: markdown>=2.6.8 in
/opt/conda/lib/python3.10/site-packages (from
tensorboard<2.18,>=2.17->tensorflow) (3.7)
```

```
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in
/opt/conda/lib/python3.10/site-packages (from
tensorboard<2.18,>=2.17->tensorflow) (0.7.2)
Requirement already satisfied: werkzeug>=1.0.1 in
/opt/conda/lib/python3.10/site-packages (from
tensorboard<2.18,>=2.17->tensorflow) (3.0.4)
Requirement already satisfied: MarkupSafe>=2.1.1 in
/opt/conda/lib/python3.10/site-packages (from
werkzeug>=1.0.1->tensorboard<2.18,>=2.17->tensorflow) (2.1.5)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/opt/conda/lib/python3.10/site-packages (from rich->keras>=3.2.0->tensorflow)
(3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/opt/conda/lib/python3.10/site-packages (from rich->keras>=3.2.0->tensorflow)
(2.18.0)
Requirement already satisfied: mdurl~=0.1 in /opt/conda/lib/python3.10/site-
packages (from markdown-it-py>=2.2.0->rich->keras>=3.2.0->tensorflow) (0.1.2)
```

[53]:
```
import tensorflow as tf

def focal_loss(gamma=2.0, alpha=0.25):
    """
    Focal loss implementation as described in the paper: https://arxiv.org/abs/
 ↪1708.02002

    Args:
        gamma: Focusing parameter. Default is 2.0.
        alpha: Balance parameter for class imbalance. Default is 0.25.

    Returns:
        loss function to be used in model.compile()
    """
    def focal_loss_fixed(y_true, y_pred):
        # Clip the predictions to avoid log(0) errors
        y_pred = tf.clip_by_value(y_pred, 1e-7, 1.0 - 1e-7)

        # Calculate cross entropy
        cross_entropy = -y_true * tf.math.log(y_pred)

        # Calculate focal loss
        loss = alpha * tf.pow(1 - y_pred, gamma) * cross_entropy
        return tf.reduce_mean(tf.reduce_sum(loss, axis=-1))

    return focal_loss_fixed
```

```
2024-10-10 07:35:38.725310: I tensorflow/core/util/port.cc:153] oneDNN custom
operations are on. You may see slightly different numerical results due to
floating-point round-off errors from different computation orders. To turn them
```

```
off, set the environment variable `TF_ENABLE_ONEDNN_OPTS=0`.
2024-10-10 07:35:38.767850: I external/local_xla/xla/tsl/cuda/cudart_stub.cc:32]
Could not find cuda drivers on your machine, GPU will not be used.
2024-10-10 07:35:40.214014: I external/local_xla/xla/tsl/cuda/cudart_stub.cc:32]
Could not find cuda drivers on your machine, GPU will not be used.
2024-10-10 07:35:40.520301: E
external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:485] Unable to register
cuFFT factory: Attempting to register factory for plugin cuFFT when one has
already been registered
2024-10-10 07:35:40.899942: E
external/local_xla/xla/stream_executor/cuda/cuda_dnn.cc:8454] Unable to register
cuDNN factory: Attempting to register factory for plugin cuDNN when one has
already been registered
2024-10-10 07:35:40.973865: E
external/local_xla/xla/stream_executor/cuda/cuda_blas.cc:1452] Unable to
register cuBLAS factory: Attempting to register factory for plugin cuBLAS when
one has already been registered
2024-10-10 07:35:41.925048: I tensorflow/core/platform/cpu_feature_guard.cc:210]
This TensorFlow binary is optimized to use available CPU instructions in
performance-critical operations.
To enable the following instructions: AVX2 AVX512F AVX512_VNNI FMA, in other
operations, rebuild TensorFlow with the appropriate compiler flags.
```

[54]:

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.models import Model
from tensorflow.keras.layers import Dense, Dropout, Input, BatchNormalization,␣
↪LeakyReLU, LSTM, Conv1D, GlobalAveragePooling1D, MultiHeadAttention, Reshape
from tensorflow.keras.optimizers import Adam
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score, f1_score
from tensorflow.keras.regularizers import l2, l1_l2
from tensorflow.keras.losses import CategoricalCrossentropy
#from tensorflow_addons.losses import SigmoidFocalCrossEntropy

# Separate features and target from the merged dataset
teacher_features = oversampled_df.drop(columns=['sii', 'id',␣
↪'actigraphy_present'])  # All columns except 'sii' and 'id'
teacher_target = oversampled_df['sii']  # Target variable

# Split the oversampled dataset into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(
    oversampled_df.drop(columns=['sii', 'id']), oversampled_df['sii'],␣
↪test_size=0.2, random_state=42
)

def create_teacher_model(architecture='A', input_shape=None):
```

```
model = Sequential()

# Dynamically set input shape if not provided
if input_shape is None:
    raise ValueError("Input shape must be specified for model creation.")

# Add the input layer
model.add(Input(shape=(input_shape,)))

# Choose architecture A or B
if architecture == 'A':
    # OLD Model A - Previous architectures and classification reports
    # TEST #1 ~ 62% accuracy
    # Teacher Model Performance on Validation Set: (model A)
    # Accuracy: 0.6207
    # F1-Score: 0.6437
    # Classification Report:
    #               precision    recall  f1-score   support
    #            0       0.86      0.70      0.77       401
    #            1       0.36      0.51      0.42       154
    #            2       0.31      0.40      0.35        50
    #            3       0.00      0.00      0.00         4
    #
    #     accuracy                           0.62       609
    #    macro avg       0.38      0.40      0.39       609
    # weighted avg       0.68      0.62      0.64       609

    # TEST #2 ~ 25% accuracy
    # Accuracy: 0.2447
    # F1-Score: 0.0962
    # Classification Report:
    #               precision    recall  f1-score   support
    #            0       0.00      0.00      0.00       418
    #            1       0.00      0.00      0.00       407
    #            2       0.00      0.00      0.00       416
    #            3       0.24      1.00      0.39       402
    #
    #     accuracy                           0.24      1643
    #    macro avg       0.06      0.25      0.10      1643
    # weighted avg       0.06      0.24      0.10      1643

    # TEST #3 ~ 30% accuracy
    # Accuracy: 0.3001
    # F1-Score: 0.2369
    # Classification Report:
    #               precision    recall  f1-score   support
    #            0       0.27      0.73      0.40       418
```

```
#           1       0.00      0.00      0.00       407
#           2       0.37      0.30      0.33       416
#           3       0.34      0.16      0.22       402
#
#    accuracy                           0.30      1643
#    macro avg       0.25      0.30      0.24      1643
# weighted avg       0.25      0.30      0.24      1643

# OLD Model A:
 # - Accuracy: 0.3774
 # - F1-Score: 0.3484
 # Classification Report:
 #              precision    recall  f1-score   support
 #           0       0.40      0.69      0.51       418
 #           1       0.21      0.30      0.25       407
 #           2       0.27      0.04      0.07       416
 #           3       0.71      0.48      0.57       402
 #    accuracy                           0.38      1643
 #    macro avg       0.40      0.38      0.35      1643
 # weighted avg       0.40      0.38      0.35      1643

 # OLD Model A:
 # - Accuracy: 0.2447
 # - F1-Score: 0.0962
 # Classification Report:
 #              precision    recall  f1-score   support
 #           0       0.00      0.00      0.00       418
 #           1       0.00      0.00      0.00       407
 #           2       0.00      0.00      0.00       416
 #           3       0.24      1.00      0.39       402
 #    accuracy                           0.24      1643
 #    macro avg       0.06      0.25      0.10      1643
 # weighted avg       0.06      0.24      0.10      1643

 # Model A: Sequential Model with a Single Hidden Layer
 # Model A: Sequential Model with Multiple Hidden Layers for Stability
 # Previous Performance (Baseline Model A):
 # - Accuracy: 0.5362
 # - F1-Score: 0.5498
 # Classification Report:
 #              precision    recall  f1-score   support
 #           0       0.50      0.72      0.59       418
 #           1       0.26      0.35      0.30       407
 #           2       0.96      0.38      0.54       416
 #           3       0.87      0.70      0.77       402
 #    accuracy                           0.54      1643
 #    macro avg       0.65      0.54      0.55      1643
```

```
        # weighted avg       0.65      0.54      0.55      1643
        # Enhanced Model A: Improved Regularization and Learning Rate␣
↪Adjustments

        # New Model A: Enhanced baseline with additional hidden layers
        inputs = Input(shape=(input_shape,))
        # Architecture A: 3 hidden layers with 128, 64, and 32 neurons
        x = Dense(128, activation='relu', kernel_regularizer=l2(0.001))(inputs)
        x = Dense(64, activation='relu', kernel_regularizer=l2(0.001))(x)
        #x = LSTM(16, return_sequences=True)(x)
        #x = Dense(32, activation='relu', kernel_regularizer=l2(0.001))(x)
        x = Dense(32, activation='relu', kernel_regularizer=l2(0.001))(x)

        outputs = Dense(4, activation='softmax')(x)  # Output layer for 4␣
↪classes

        # Create the model
        model = Model(inputs=inputs, outputs=outputs)

        # Compile the model with a standard loss function
        optimizer = Adam(learning_rate=0.001)
        model.compile(optimizer=optimizer, loss=CategoricalCrossentropy(),␣
↪metrics=['accuracy'])

    elif architecture == 'B':
        # OLD Model B - Previous architectures and classification reports
        # TEST #1 ~ 68% accuracy
        # Accuracy: 0.6552
        # F1-Score: 0.5218
        # Classification Report:
        #               precision    recall  f1-score   support
        #           0       0.66      1.00      0.79       401
        #           1       0.00      0.00      0.00       154
        #           2       0.00      0.00      0.00        50
        #           3       0.00      0.00      0.00         4
        #
        #    accuracy                           0.66       609
        #    macro avg       0.16      0.25      0.20       609
        # weighted avg       0.43      0.66      0.52       609

        # TEST #2 ~ 25% accuracy
        # Accuracy: 0.2532
        # F1-Score: 0.1023
        # Classification Report:
        #               precision    recall  f1-score   support
        #           0       0.00      0.00      0.00       418
        #           1       0.00      0.00      0.00       407
```

```
#           2       0.25      1.00      0.40       416
#           3       0.00      0.00      0.00       402
#
#    accuracy                           0.25      1643
#    macro avg       0.06      0.25      0.10      1643
# weighted avg       0.06      0.25      0.10      1643

# TEST #3 ~ 29% accuracy
# Accuracy: 0.2879
# F1-Score: 0.1912
# Classification Report:
#              precision    recall  f1-score   support
#           0       0.27      0.73      0.40       418
#           1       0.00      0.00      0.00       407
#           2       0.32      0.40      0.36       416
#           3       0.00      0.00      0.00       402
#
#    accuracy                           0.29      1643
#    macro avg       0.15      0.28      0.19      1643
# weighted avg       0.15      0.29      0.19      1643

# OLD Model B:
# - Accuracy: 0.2672
# - F1-Score: 0.1807
# Classification Report:
#              precision    recall  f1-score   support
#           0       0.30      0.72      0.43       418
#           1       0.00      0.00      0.00       407
#           2       0.03      0.01      0.01       416
#           3       0.25      0.33      0.28       402
#    accuracy                           0.27      1643
#    macro avg       0.14      0.27      0.18      1643
# weighted avg       0.14      0.27      0.18      1643

# OLD Model B:
# - Accuracy: 0.2465
# - F1-Score: 0.1005
# Classification Report:
#              precision    recall  f1-score   support
#           0       0.00      0.00      0.00       418
#           1       0.00      0.00      0.00       407
#           2       0.17      0.01      0.01       416
#           3       0.25      1.00      0.40       402
#    accuracy                           0.25      1643
#    macro avg       0.10      0.25      0.10      1643
# weighted avg       0.10      0.25      0.10      1643
```

```
    # Model B: Functional Model with CNN-LSTM and Multi-Head Attention
# Previous Performance (Overfitting Model B):
# - Accuracy: 0.9884
# - F1-Score: 0.9884
# Classification Report:
#               precision    recall  f1-score   support
#            0       0.99      0.99      0.99       418
#            1       0.97      0.98      0.98       407
#            2       0.99      0.98      0.99       416
#            3       1.00      1.00      1.00       402
#     accuracy                           0.99      1643
#    macro avg       0.99      0.99      0.99      1643
# weighted avg       0.99      0.99      0.99      1643

# New Model B: Reduced complexity with fewer attention heads and dense␣
↪layers
inputs = Input(shape=(input_shape,))
reshaped = Reshape((input_shape, 1))(inputs)  # Reshape for 1D Conv
x = Dense(128, activation='relu', kernel_regularizer=l2(0.
↪001))(reshaped)
x = Conv1D(8, kernel_size=3, activation='relu', padding='same')(x)
x = LSTM(16, return_sequences=True)(x)  # LSTM to capture sequence info

# Define query and value tensors for Multi-Head Attention
query = Dense(32)(x)  # Define the query tensor
value = Dense(32)(x)  # Define the value tensor

# Use Multi-Head Attention with query and value tensors
attention_output = MultiHeadAttention(num_heads=2,␣
↪key_dim=32)(query=query, value=value)

# Pool the attention output and add the final dense layer
pooled_output = GlobalAveragePooling1D()(attention_output)
x = Dropout(0.3)(pooled_output)
outputs = Dense(4, activation='softmax')(x)  # Output layer for 4␣
↪classes

# Create the model
model = Model(inputs=inputs, outputs=outputs)

# Compile the model with a standard loss function
optimizer = Adam(learning_rate=0.0004)  # Reduced learning rate to␣
↪control overfitting
model.compile(optimizer=optimizer, loss=CategoricalCrossentropy(),␣
↪metrics=['accuracy'])
```

```
    else:
        raise ValueError("Unsupported architecture specified. Choose 'A' or 'B'.
↪")

    return model


# Create models for each architecture type
teacher_model_A = create_teacher_model(architecture='A',␣
↪input_shape=teacher_features.shape[1])
teacher_model_B = create_teacher_model(architecture='B',␣
↪input_shape=teacher_features.shape[1])

# Print the summaries of both models to verify architecture
print("Teacher Model A Summary:")
teacher_model_A.summary()

print("\nTeacher Model B Summary:")
teacher_model_B.summary()
```

```
Teacher Model A Summary:
```

Model: "functional"

Layer (type)	Output Shape	Param #
input_layer_1 (InputLayer)	(None, 102)	0
dense (Dense)	(None, 128)	13,184
dense_1 (Dense)	(None, 64)	8,256
dense_2 (Dense)	(None, 32)	2,080
dense_3 (Dense)	(None, 4)	132

Total params: 23,652 (92.39 KB)

Trainable params: 23,652 (92.39 KB)

Non-trainable params: 0 (0.00 B)

Teacher Model B Summary:

Model: "functional_1"

Layer (type)	Output Shape	Param #	Connected to
input_layer_3 (InputLayer)	(None, 102)	0	-
reshape (Reshape)	(None, 102, 1)	0	input_layer_3[0]…
dense_4 (Dense)	(None, 102, 128)	256	reshape[0][0]
conv1d (Conv1D)	(None, 102, 8)	3,080	dense_4[0][0]
lstm (LSTM)	(None, 102, 16)	1,600	conv1d[0][0]
dense_5 (Dense)	(None, 102, 32)	544	lstm[0][0]
dense_6 (Dense)	(None, 102, 32)	544	lstm[0][0]
multi_head_attenti… (MultiHeadAttentio…	(None, 102, 32)	8,416	dense_5[0][0], dense_6[0][0]
global_average_poo… (GlobalAveragePool…	(None, 32)	0	multi_head_atten…
dropout_1 (Dropout)	(None, 32)	0	global_average_p…
dense_7 (Dense)	(None, 4)	132	dropout_1[0][0]

Total params: 14,572 (56.92 KB)

Trainable params: 14,572 (56.92 KB)

Non-trainable params: 0 (0.00 B)

[80]:
```
# Import necessary libraries
from tensorflow.keras.utils import to_categorical
from tensorflow.keras.callbacks import EarlyStopping
import matplotlib.pyplot as plt
from sklearn.utils.class_weight import compute_class_weight
```

```
# Split the data into training and validation sets (already balanced using␣
↪SMOTE earlier)
X_train, X_val, y_train, y_val = train_test_split(teacher_features,␣
↪teacher_target, test_size=0.2, random_state=42)

print(X_train)
# Configure training hyperparameters
batch_size = 64
epochs = 50
learning_rate = 0.001

# Create the teacher model (choose architecture A or B)
teacher_model = teacher_model_A

# Ensure that y_train and y_val are converted to NumPy arrays before flattening␣
↪or processing
y_train_np = y_train.values  # Convert to NumPy array
y_val_np = y_val.values      # Convert to NumPy array

# Print shapes of the features and target to ensure they are correct
print(f"X_train shape: {X_train.shape}")
print(f"X_val shape: {X_val.shape}")
print(f"y_train shape: {y_train_np.shape}")
print(f"y_val shape: {y_val_np.shape}")

# Check the data types and unique values in y_train and y_val to ensure they␣
↪are categorical
print(f"y_train data type: {y_train_np.dtype}, unique values: {set(y_train_np.
↪flatten())}")
print(f"y_val data type: {y_val_np.dtype}, unique values: {set(y_val_np.
↪flatten())}")

# One-hot encode the labels using to_categorical if not already one-hot encoded
y_train_categorical = to_categorical(y_train_np, num_classes=4)
y_val_categorical = to_categorical(y_val_np, num_classes=4)

# Print the shapes after encoding to ensure everything is correct
print(f"y_train_categorical shape: {y_train_categorical.shape}")
print(f"y_val_categorical shape: {y_val_categorical.shape}")

# Ensure that the model is compiled with the appropriate loss and metrics
teacher_model = teacher_model_A  # Choose the appropriate model

teacher_model.compile(
    optimizer='adam',  # or any other optimizer of your choice
```

```
    loss='categorical_crossentropy',  # Use categorical_crossentropy since␣
↪labels are one-hot encoded
    metrics=['accuracy']  # Monitor accuracy during training
)

# Compute class weights
class_weights = compute_class_weight(
    class_weight='balanced',
    classes=np.unique(y_train_np),
    y=y_train_np
)

# Convert to dictionary format
class_weight_dict = dict(enumerate(class_weights))

# Remove or update the early stopping callback to monitor a valid metric
early_stopping = EarlyStopping(monitor='val_loss', patience=8,␣
↪restore_best_weights=True)  # Updated patience to 4 as requested

# Train the teacher model
history = teacher_model.fit(
    X_train, y_train_categorical,  # Use the one-hot encoded labels for training
    validation_data=(X_val, y_val_categorical),
    epochs=epochs,
    batch_size=batch_size,
    callbacks=[early_stopping],
    class_weight=class_weight_dict,  # Apply class weights here
    verbose=1
)

# Plot training and validation loss
plt.figure(figsize=(10, 5))
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title('Teacher Model Training & Validation Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()

# Evaluate the teacher model on the validation set
teacher_val_predictions = teacher_model.predict(X_val)
teacher_val_pred_labels = teacher_val_predictions.argmax(axis=1)  # Convert␣
↪probabilities to class labels

# Print evaluation metrics
print("Teacher Model Performance on Validation Set:")
```

```
print(f"Accuracy: {accuracy_score(y_val_np, teacher_val_pred_labels):.4f}")
print(f"F1-Score: {f1_score(y_val_np, teacher_val_pred_labels,␣
↪average='weighted'):.4f}")
print(f"Classification Report:\n{classification_report(y_val_np,␣
↪teacher_val_pred_labels)}")

teacher_val_preds = teacher_model_A.predict(X_val)
```

```
      Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
4700                   0.502580         7.565392         0.000000
1181                   0.598983        14.000000         1.000000
1375                   0.475815        15.000000         0.000000
5787                   0.503763         9.907644         0.226911
2115                   0.460940         9.604295         0.367842
…                           …              …              …
5734                   0.536519        13.275019         0.000000
5191                   0.477339        13.981799         0.000000
5390                   0.541954        12.000000         0.027975
860                    0.559547         8.000000         0.000000
7270                   0.500848        15.103731         0.367910

      CGAS-Season  CGAS-CGAS_Score  Physical-Season  Physical-BMI  \
4700     0.515994        61.738432         0.518235     16.694711
1181     0.581124        60.000000         0.523758     20.732314
1375     0.515994        65.000000         0.516702     22.967517
5787     0.515994        66.546178         0.525762     19.552118
2115     0.460940        65.102546         0.471466     18.495364
…             …              …              …            …
5734     0.569345        64.375094         0.580097     19.951395
5191     0.539074        67.908993         0.525172     17.364689
5390     0.528603        62.776202         0.526350     20.241677
860      0.527092        80.000000         0.526350     17.758884
7270     0.534173        68.735820         0.491659     19.287629

      Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
4700        51.630784        63.409656                     20.390689
1181        62.500000       115.200000                     28.473320
1375        75.250000       185.000000                     35.152447
5787        56.382565        91.823578                     26.835510
2115        54.579849        82.240167                     26.032637
…                …              …                              …
5734        69.431245       136.849510                     31.374684
5191        69.254550       118.473234                     28.084724
5390        57.139874        94.019621                     40.339213
860         51.500000        67.000000                     27.150807
7270        60.229286       101.171800                     27.613246
```

	Physical-Diastolic_BP	Physical-HeartRate	Physical-Systolic_BP	\
4700	63.086520	82.565392	110.217304	
1181	56.000000	66.000000	95.000000	
1375	64.000000	81.000000	136.000000	
5787	60.092356	86.000000	111.680733	
2115	68.984227	82.140101	116.171643	
…	…	…	…	
5734	73.599699	84.424755	124.499623	
5191	62.291221	75.236617	110.364026	
5390	59.671394	75.664303	110.727343	
860	74.000000	89.000000	132.000000	
7270	72.992900	86.798660	120.941474	

	Fitness_Endurance-Season	Fitness_Endurance-Max_Stage	\
4700	0.450653	4.246104	
1181	0.562024	5.045249	
1375	0.562024	2.585017	
5787	0.449568	5.040249	
2115	0.562024	5.022398	
…	…	…	
5734	0.562024	4.524442	
5191	0.562024	4.237524	
5390	0.420630	5.000368	
860	0.416561	6.000000	
7270	0.562024	4.817933	

	Fitness_Endurance-Time_Mins	Fitness_Endurance-Time_Sec	FGC-Season	\
4700	6.249104	18.435388	0.532606	
1181	7.281378	27.478715	0.552717	
1375	12.795107	36.652067	0.531007	
5787	7.492937	25.325949	0.538366	
2115	7.641242	26.788222	0.472214	
…	…	…	…	
5734	8.162711	20.767827	0.548770	
5191	10.583830	21.916051	0.538627	
5390	6.975520	41.548811	0.539792	
860	9.000000	35.000000	0.539792	
7270	7.042091	23.761783	0.538891	

	FGC-FGC_CU	FGC-FGC_CU_Zone	FGC-FGC_GSND	FGC-FGC_GSND_Zone	\
4700	7.826960	0.782696	18.176607	2.211874	
1181	16.763021	0.351448	22.147415	1.827399	
1375	18.000000	0.000000	22.200000	1.000000	
5787	14.213739	0.753985	20.318358	2.133178	
2115	11.067916	0.525539	18.121700	1.823866	
…	…	…	…	…	
5734	24.676073	0.275019	28.360045	2.000000	
5191	64.962526	0.981799	30.056317	2.000000	

5390	3.195823	0.000000	16.510481	1.972025
860	10.000000	1.000000	14.991645	1.789701
7270	25.190724	0.698734	22.647991	1.654444

	FGC-FGC_GSD	FGC-FGC_GSD_Zone	FGC-FGC_PU	FGC-FGC_PU_Zone	FGC-FGC_SRL
4700	13.077850	1.426754	5.565392	1.000000	11.217304
1181	22.270642	2.093320	9.518281	0.475742	9.062625
1375	49.600000	3.000000	9.000000	0.000000	7.783728
5787	18.001850	2.129147	8.115063	0.616528	10.238470
2115	17.314493	2.016925	5.842512	0.360343	8.892884
…	…	…	…	…	…
5734	28.287622	2.000000	6.375094	0.000000	3.624906
5191	32.881692	2.000000	9.908993	0.000000	0.091007
5390	15.344051	1.972025	3.888101	0.000000	7.055949
860	14.107620	2.297390	0.000000	0.000000	8.500000
7270	20.886962	1.550619	8.923425	0.464942	8.657967

	FGC-FGC_SRL_Zone	FGC-FGC_SRR	FGC-FGC_SRR_Zone	FGC-FGC_TL
4700	1.000000	12.000000	1.000000	9.782696
1181	0.578585	9.036522	0.570378	10.257168
1375	0.608529	8.536847	0.702746	8.000000
5787	0.718168	10.341462	0.779268	8.868513
2115	0.678605	9.004501	0.680599	9.016459
…	…	…	…	…
5734	0.000000	4.349887	0.000000	8.724981
5191	0.000000	0.109208	0.000000	8.018201
5390	0.000000	6.569937	0.000000	8.055949
860	1.000000	7.500000	1.000000	8.000000
7270	0.336244	8.679490	0.419282	9.572363

	FGC-FGC_TL_Zone	BIA-Season	BIA-BIA_Activity_Level_num	BIA-BIA_BMC
4700	1.000000	0.566098	3.000000	3.317999
1181	0.726999	0.490251	3.000127	7.660636
1375	0.000000	0.566098	3.000000	8.322940
5787	0.631474	0.563331	2.773089	4.139876
2115	0.772281	0.490251	2.687674	7.660624
…	…	…	…	…
5734	0.724981	0.542911	3.000000	6.642455
5191	0.018201	0.565516	3.000000	6.410864
5390	0.027975	0.534475	1.972025	9.755144
860	1.000000	0.534475	4.000000	7.134250
7270	0.777310	0.506521	2.728329	5.298298

	BIA-BIA_BMI	BIA-BIA_BMR	BIA-BIA_DEE	BIA-BIA_ECW	BIA-BIA_FFM
4700	16.696570	1035.152246	1759.756600	11.941833	52.520217
1181	18.540285	1229.030094	2084.783491	20.415851	73.170816
1375	22.970000	1863.730000	3168.330000	44.472800	140.774000
5787	19.554283	1192.572076	1960.892484	20.535850	69.287832

```
2115    18.540302  1229.029946  2055.943863    20.415803    73.170848
…           …           …           …           …           …
5734    19.953576  1560.797970  2653.357374    33.209356   108.508055
5191    17.366620  1475.213996  2507.866739    29.309628    99.392644
5390    20.243845  1160.136643  1845.637392    16.047988    65.832307
860     17.760800  1029.600000  2162.150000    11.490600    51.928600
7270    19.763566  1276.004739  2080.516891    23.732340    78.174404

      BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
4700     13.829303     2.867249    10.889446           1.000000    27.850559
1181     14.810861     3.729442    38.333226           1.823437    32.828975
1375     17.478900     5.491160    44.225700           3.000000    58.332500
5787     15.097468     4.456760    22.535746           2.000000    32.374804
2115     14.810861     3.729442     9.164517           1.663348    32.829002
…             …            …            …                …            …
5734     15.822819     4.130711    28.341166           1.724981    46.512153
5191     14.569769     2.796792    19.080583           1.018201    43.256442
5390     14.174343     6.069581    28.187314           1.972025    30.204337
860      13.765600     3.995230    15.071400           2.000000    26.351600
7270     14.928128     4.835409    23.609686           1.678369    33.655141

      BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  \
4700    12.727817    49.202221    24.608922    39.792399      0.480263
1181    19.926072    65.510248    42.408203    53.244832      0.480263
1375    37.969000   132.451000    65.539300   102.805000      0.867459
5787    16.377178    65.147995    30.339432    52.910654      0.480263
2115    19.926029    65.510205    33.972700    53.244809      0.480263
…           …           …           …           …           …
5734    28.786835   101.866013    47.542115    79.721509      0.603595
5191    26.826581    92.981790    44.731676    72.566070      0.920551
5390    19.579983    56.077124    25.595848    46.252325      0.480263
860     14.086400    44.794300    20.822800    37.842200      0.480263
7270    20.786931    72.876101    34.943804    57.387476      0.649018

      PAQ_A-PAQ_A_Total  PAQ_C-Season  PAQ_C-PAQ_C_Total  PCIAT-PCIAT_Total  \
4700           1.920678      0.526024           2.475277          51.000000
1181           2.185024      0.579612           2.640000          27.000000
1375           4.309000      0.460443         -37.127996          13.000000
5787           2.019716      0.610596           3.053284          54.773089
2115           2.101212      0.460443           2.540356           8.770492
…                   …             …                  …                  …
5734           2.429442      0.588854          -2.360806          58.550038
5191           3.044169      0.463667         -10.972151          59.963597
5390           1.859503      0.549041           1.637975          57.832152
860            2.203806      0.460443           2.703067          20.000000
7270           1.735669      0.513405           5.289773          81.000000

      SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  PreInt_EduHx-Season  \
```

4700	0.518336	67.434608	91.651912	0.511340
1181	0.611293	46.000000	64.000000	0.602178
1375	0.512539	31.000000	45.000000	0.486120
5787	0.539217	31.361466	44.815288	0.512454
2115	0.467407	40.896130	57.473724	0.457262
…	…	…	…	…
5734	0.584134	31.300226	45.125282	0.542743
5191	0.514336	39.781584	55.726980	0.487541
5390	0.515378	46.055949	64.083924	0.532693
860	0.515460	45.000000	63.000000	0.564223
7270	0.513613	51.471641	71.207461	0.503747

	PreInt_EduHx-computerinternet_hoursday	Age_Group	PCIAT_Cluster	\
4700	0.217304	0.362030	4.869216	
1181	3.000000	0.855589	0.000000	
1375	0.000000	0.855589	0.000000	
5787	1.226911	0.474024	4.000000	
2115	0.695464	0.443345	0.000000	
…	…	…	…	
5734	1.449962	0.855589	8.275019	
5191	0.036403	0.855589	8.981799	
5390	2.000000	0.855589	4.111899	
860	0.000000	0.362030	0.000000	
7270	2.000000	0.899069	8.000000	

	CGAS-CGAS_Score_x_SDS-SDS_Total_Raw	CGAS-CGAS_Score_x_Physical-BMI	\
4700	4166.028261	1030.154161	
1181	2760.000000	1243.938816	
1375	2015.000000	1492.888599	
5787	2084.880616	1297.875366	
2115	2645.540688	1204.515066	
…	…	…	
5734	2026.918002	1280.724352	
5191	2702.599551	1178.891546	
5390	2890.782516	1270.780770	
860	3600.000000	1420.710718	
7270	3539.805865	1329.224327	

	X_mean	X_std	X_min	X_max	Y_mean	Y_std	Y_min	\
4700	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1181	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
1375	0.059223	0.119636	-0.949533	1.042885	0.160075	0.413583	-0.997729	
5787	-0.264147	0.586831	-2.053381	1.867559	-0.001036	0.448261	-2.929571	
2115	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
…	…	…	…	…	…	…	…	
5734	-0.016970	0.305961	-1.571900	1.315234	0.028214	0.297028	-2.764179	
5191	0.027425	0.095131	-0.970502	1.002803	0.042165	0.162715	-1.039631	
5390	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	

```
860   0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
7270  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000

         Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  enmo_std  \
4700  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
1181  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
1375  1.035902 -0.621916  0.653727 -1.023129  1.013462   0.015610  0.009221
5787  2.699965 -0.067758  0.559380 -1.011406  1.623062   0.036143  0.126210
2115  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
…          …         …         …         …         …          …         …
5734  2.026787 -0.518336  0.595110 -1.020474  1.045226   0.018623  0.064642
5191  1.023706 -0.950779  0.188632 -1.014648  0.980170   0.001053  0.006825
5390  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
860   0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000
7270  0.000000  0.000000  0.000000  0.000000  0.000000   0.000000  0.000000

      anglez_mean  anglez_std  non-wear_flag_mean  light_mean   light_std  \
4700     0.000000    0.000000            0.000000    0.000000    0.000000
1181     0.000000    0.000000            0.000000    0.000000    0.000000
1375   -53.168716   51.672413            0.908472   11.131063   17.960384
5787    -4.987710   40.240898            0.036958   28.955147  143.626431
2115     0.000000    0.000000            0.000000    0.000000    0.000000
…             …           …                   …           …           …
5734   -44.395025   50.065777            0.586121   10.875281   41.560736
5191   -83.972223   17.170857            0.903722    6.160095   59.511204
5390     0.000000    0.000000            0.000000    0.000000    0.000000
860      0.000000    0.000000            0.000000    0.000000    0.000000
7270     0.000000    0.000000            0.000000    0.000000    0.000000

      battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
4700              0.000000             0.000000      0.000000e+00
1181              0.000000             0.000000      0.000000e+00
1375           3844.954346           163.499146      4.315404e+13
5787           3845.312609           159.261820      4.328247e+13
2115              0.000000             0.000000      0.000000e+00
…                      …                    …                 …
5734           3854.832900           164.385136      4.323829e+13
5191           3856.278209           162.648632      4.323598e+13
5390              0.000000             0.000000      0.000000e+00
860               0.000000             0.000000      0.000000e+00
7270              0.000000             0.000000      0.000000e+00

      weekday_mean  quarter_mean  relative_date_PCIAT_mean  fourier_X_real  \
4700      0.000000      0.000000                  0.000000        0.000000
1181      0.000000      0.000000                  0.000000        0.000000
1375      4.186077      2.905743                 18.562656     1548.397978
5787      4.007968      3.546178                149.066992   -42939.590174
2115      0.000000      0.000000                  0.000000        0.000000
```

…	…	…	…	…
5734	4.187859	1.716959	54.741550	10162.399097
5191	4.204346	3.559497	111.615809	-308.758777
5390	0.000000	0.000000	0.000000	0.000000
860	0.000000	0.000000	0.000000	0.000000
7270	0.000000	0.000000	0.000000	0.000000

	fourier_X_imag	fourier_Y_real	fourier_Y_imag	fourier_Z_real	\
4700	0.000000e+00	0.000000	0.000000	0.000000	
1181	0.000000e+00	0.000000	0.000000	0.000000	
1375	3.637979e-13	17745.827957	0.000000	-16342.842688	
5787	-6.858480e-13	801.332181	184.197653	-8737.058914	
2115	0.000000e+00	0.000000	0.000000	0.000000	
…	…	…	…	…	
5734	-2.917840e+03	3456.750701	291.801908	-59359.227833	
5191	-7.325500e+01	8149.463251	7.325950	-70943.255998	
5390	0.000000e+00	0.000000	0.000000	0.000000	
860	0.000000e+00	0.000000	0.000000	0.000000	
7270	0.000000e+00	0.000000	0.000000	0.000000	

	fourier_Z_imag	activity_during_day	activity_during_night	\
4700	0.000000e+00	0.0	0.000000	
1181	0.000000e+00	0.0	0.000000	
1375	5.820766e-12	0.0	6689.417480	
5787	-2.446071e+03	0.0	14316.060811	
2115	0.000000e+00	0.0	0.000000	
…	…	…	…	
5734	3.488591e-13	0.0	8078.098284	
5191	3.634228e-13	0.0	446.840576	
5390	0.000000e+00	0.0	0.000000	
860	0.000000e+00	0.0	0.000000	
7270	0.000000e+00	0.0	0.000000	

	activity_ratio_day_night	non_wear_proportion	sedentary_proportion	\
4700	0.0	0.000000	0.000000	
1181	0.0	0.000000	0.000000	
1375	0.0	0.908472	0.995669	
5787	0.0	0.036958	0.859015	
2115	0.0	0.000000	0.000000	
…	…	…	…	
5734	0.0	0.586121	0.919501	
5191	0.0	0.903722	0.996632	
5390	0.0	0.000000	0.000000	
860	0.0	0.000000	0.000000	
7270	0.0	0.000000	0.000000	

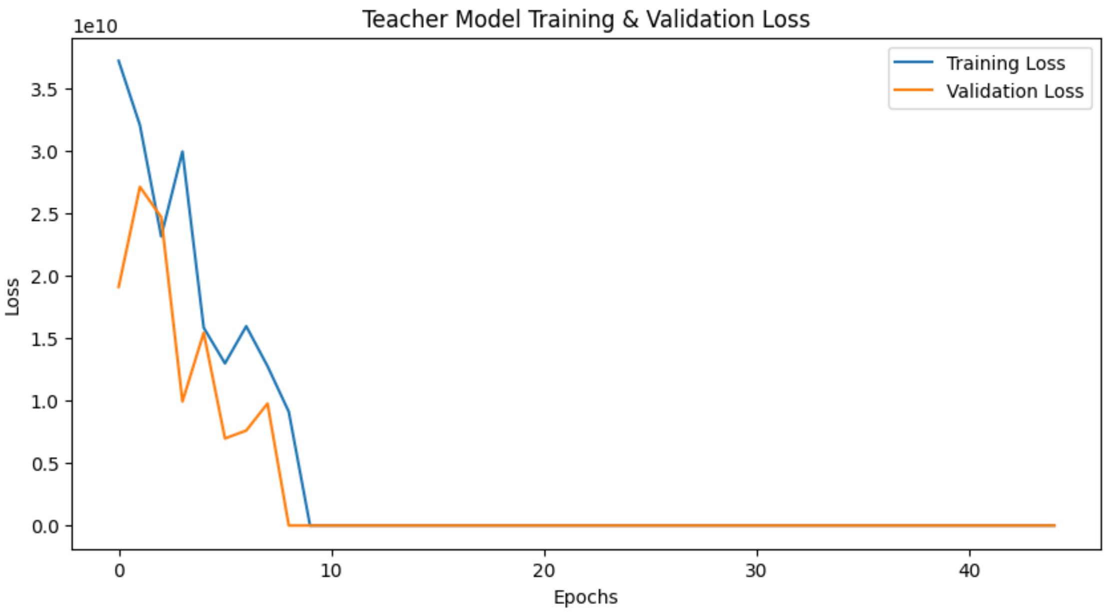
	light_activity_proportion	moderate_activity_proportion	\
4700	0.000000	0.000000	

```
1181                  0.000000                      0.000000
1375                  0.004329                      0.000002
5787                  0.129437                      0.007394
2115                  0.000000                      0.000000
…                        …                            …
5734                  0.077019                      0.002208
5191                  0.003271                      0.000065
5390                  0.000000                      0.000000
860                   0.000000                      0.000000
7270                  0.000000                      0.000000

      vigorous_activity_proportion
4700                     0.000000
1181                     0.000000
1375                     0.000000
5787                     0.004154
2115                     0.000000
…                           …
5734                     0.001271
5191                     0.000032
5390                     0.000000
860                      0.000000
7270                     0.000000

[6569 rows x 102 columns]
X_train shape: (6569, 102)
X_val shape: (1643, 102)
y_train shape: (6569,)
y_val shape: (1643,)
y_train data type: int64, unique values: {0, 1, 2, 3}
y_val data type: int64, unique values: {0, 1, 2, 3}
y_train_categorical shape: (6569, 4)
y_val_categorical shape: (1643, 4)
Epoch 1/50
103/103             2s 4ms/step -
accuracy: 0.4369 - loss: 50564001792.0000 - val_accuracy: 0.4419 - val_loss:
19116826624.0000
Epoch 2/50
103/103             0s 2ms/step -
accuracy: 0.4420 - loss: 34283069440.0000 - val_accuracy: 0.4437 - val_loss:
27138234368.0000
Epoch 3/50
103/103             0s 2ms/step -
accuracy: 0.5087 - loss: 21594140672.0000 - val_accuracy: 0.4784 - val_loss:
24713201664.0000
Epoch 4/50
103/103             0s 2ms/step -
accuracy: 0.5392 - loss: 31784050688.0000 - val_accuracy: 0.4912 - val_loss:
```

```
9934446592.0000
Epoch 5/50
103/103            0s 2ms/step -
accuracy: 0.5239 - loss: 20208961536.0000 - val_accuracy: 0.6336 - val_loss:
15447905280.0000
Epoch 6/50
103/103            0s 2ms/step -
accuracy: 0.5981 - loss: 13002493952.0000 - val_accuracy: 0.5247 - val_loss:
6976550912.0000
Epoch 7/50
103/103            0s 2ms/step -
accuracy: 0.5450 - loss: 15250896896.0000 - val_accuracy: 0.6409 - val_loss:
7611130880.0000
Epoch 8/50
103/103            0s 2ms/step -
accuracy: 0.5765 - loss: 12817714176.0000 - val_accuracy: 0.6494 - val_loss:
9775671296.0000
Epoch 9/50
103/103            0s 2ms/step -
accuracy: 0.6289 - loss: 13025772544.0000 - val_accuracy: 0.6738 - val_loss:
0.9720
Epoch 10/50
103/103            0s 2ms/step -
accuracy: 0.6812 - loss: 0.9475 - val_accuracy: 0.7127 - val_loss: 0.8914
Epoch 11/50
103/103            0s 2ms/step -
accuracy: 0.7152 - loss: 0.8489 - val_accuracy: 0.7365 - val_loss: 0.8106
Epoch 12/50
103/103            0s 2ms/step -
accuracy: 0.7550 - loss: 0.7769 - val_accuracy: 0.7419 - val_loss: 0.7919
Epoch 13/50
103/103            0s 2ms/step -
accuracy: 0.7519 - loss: 0.7780 - val_accuracy: 0.7529 - val_loss: 0.7728
Epoch 14/50
103/103            0s 2ms/step -
accuracy: 0.7530 - loss: 0.7633 - val_accuracy: 0.7389 - val_loss: 0.7977
Epoch 15/50
103/103            0s 2ms/step -
accuracy: 0.7429 - loss: 0.7767 - val_accuracy: 0.7492 - val_loss: 0.7557
Epoch 16/50
103/103            0s 2ms/step -
accuracy: 0.7373 - loss: 0.7771 - val_accuracy: 0.7407 - val_loss: 0.7703
Epoch 17/50
103/103            0s 2ms/step -
accuracy: 0.7528 - loss: 0.7400 - val_accuracy: 0.6391 - val_loss: 0.9627
Epoch 18/50
103/103            0s 2ms/step -
accuracy: 0.7267 - loss: 0.7951 - val_accuracy: 0.7553 - val_loss: 0.7473
```

```
Epoch 19/50
103/103             0s 2ms/step -
accuracy: 0.7625 - loss: 0.7289 - val_accuracy: 0.7578 - val_loss: 0.7279
Epoch 20/50
103/103             0s 2ms/step -
accuracy: 0.7549 - loss: 0.7363 - val_accuracy: 0.7596 - val_loss: 0.7279
Epoch 21/50
103/103             0s 2ms/step -
accuracy: 0.7636 - loss: 0.7157 - val_accuracy: 0.7614 - val_loss: 0.7247
Epoch 22/50
103/103             0s 2ms/step -
accuracy: 0.7654 - loss: 0.7074 - val_accuracy: 0.7389 - val_loss: 0.7648
Epoch 23/50
103/103             0s 2ms/step -
accuracy: 0.7522 - loss: 0.7291 - val_accuracy: 0.7395 - val_loss: 0.7697
Epoch 24/50
103/103             0s 2ms/step -
accuracy: 0.7385 - loss: 0.7564 - val_accuracy: 0.7529 - val_loss: 0.7278
Epoch 25/50
103/103             0s 2ms/step -
accuracy: 0.7536 - loss: 0.7208 - val_accuracy: 0.7675 - val_loss: 0.6954
Epoch 26/50
103/103             0s 2ms/step -
accuracy: 0.7570 - loss: 0.7128 - val_accuracy: 0.7602 - val_loss: 0.7077
Epoch 27/50
103/103             0s 2ms/step -
accuracy: 0.7625 - loss: 0.7021 - val_accuracy: 0.7529 - val_loss: 0.7084
Epoch 28/50
103/103             0s 2ms/step -
accuracy: 0.7573 - loss: 0.7023 - val_accuracy: 0.7681 - val_loss: 0.6909
Epoch 29/50
103/103             0s 2ms/step -
accuracy: 0.7512 - loss: 0.7244 - val_accuracy: 0.7590 - val_loss: 0.7083
Epoch 30/50
103/103             0s 2ms/step -
accuracy: 0.7394 - loss: 0.7494 - val_accuracy: 0.7304 - val_loss: 0.7666
Epoch 31/50
103/103             0s 2ms/step -
accuracy: 0.7663 - loss: 0.7054 - val_accuracy: 0.7626 - val_loss: 0.6941
Epoch 32/50
103/103             0s 2ms/step -
accuracy: 0.7606 - loss: 0.6994 - val_accuracy: 0.6963 - val_loss: 0.9199
Epoch 33/50
103/103             0s 2ms/step -
accuracy: 0.7139 - loss: 0.8451 - val_accuracy: 0.7645 - val_loss: 0.6888
Epoch 34/50
103/103             0s 2ms/step -
accuracy: 0.7648 - loss: 0.7026 - val_accuracy: 0.7371 - val_loss: 0.7481
```

```
Epoch 35/50
103/103            0s 2ms/step -
accuracy: 0.7585 - loss: 0.7193 - val_accuracy: 0.6275 - val_loss: 1.0134
Epoch 36/50
103/103            0s 2ms/step -
accuracy: 0.7183 - loss: 0.8031 - val_accuracy: 0.7687 - val_loss: 0.6738
Epoch 37/50
103/103            0s 2ms/step -
accuracy: 0.7751 - loss: 0.6666 - val_accuracy: 0.7718 - val_loss: 0.6642
Epoch 38/50
103/103            0s 2ms/step -
accuracy: 0.7667 - loss: 0.6746 - val_accuracy: 0.7614 - val_loss: 0.6877
Epoch 39/50
103/103            0s 2ms/step -
accuracy: 0.7677 - loss: 0.6771 - val_accuracy: 0.7273 - val_loss: 0.7783
Epoch 40/50
103/103            0s 2ms/step -
accuracy: 0.7620 - loss: 0.7038 - val_accuracy: 0.7541 - val_loss: 0.7097
Epoch 41/50
103/103            0s 2ms/step -
accuracy: 0.7567 - loss: 0.6888 - val_accuracy: 0.7438 - val_loss: 0.7402
Epoch 42/50
103/103            0s 2ms/step -
accuracy: 0.7520 - loss: 0.6994 - val_accuracy: 0.7663 - val_loss: 0.6777
Epoch 43/50
103/103            0s 2ms/step -
accuracy: 0.7211 - loss: 1.1159 - val_accuracy: 0.7365 - val_loss: 1.3660
Epoch 44/50
103/103            0s 2ms/step -
accuracy: 0.7394 - loss: 0.8639 - val_accuracy: 0.7048 - val_loss: 0.8435
Epoch 45/50
103/103            0s 2ms/step -
accuracy: 0.7344 - loss: 0.7639 - val_accuracy: 0.6939 - val_loss: 0.8696
```




```
52/52             0s 2ms/step
Teacher Model Performance on Validation Set:
Accuracy: 0.7718
F1-Score: 0.7853
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.72      0.83       418
           1       0.95      0.69      0.80       407
           2       0.53      0.98      0.69       416
           3       1.00      0.70      0.82       402

    accuracy                           0.77      1643
   macro avg       0.87      0.77      0.79      1643
weighted avg       0.87      0.77      0.79      1643

52/52             0s 1ms/step
```

[58]:
```
# training data preprocessing
# Import necessary libraries
import pandas as pd
import numpy as np
from scipy.stats import skew, kurtosis

# Step 2: Filter only the numeric columns from the test data
numeric_cols = test_df.select_dtypes(include=[np.number]).columns
```

```python
# Step 3: Create a summary statistics DataFrame
summary_stats = test_df[numeric_cols].describe().T

# Step 4: Add skewness and kurtosis to the summary statistics DataFrame
summary_stats['Skewness'] = test_df[numeric_cols].apply(skew)
summary_stats['Kurtosis'] = test_df[numeric_cols].apply(kurtosis)

# Step 5: Identify potential outliers using the IQR method
Q1 = test_df[numeric_cols].quantile(0.25)
Q3 = test_df[numeric_cols].quantile(0.75)
IQR = Q3 - Q1
summary_stats['Lower_Bound'] = Q1 - 1.5 * IQR
summary_stats['Upper_Bound'] = Q3 + 1.5 * IQR

# Calculate the number of outliers for each column
summary_stats['Outliers'] = test_df[numeric_cols].apply(
    lambda x: ((x < summary_stats.loc[x.name, 'Lower_Bound']) |
               (x > summary_stats.loc[x.name, 'Upper_Bound'])).sum()
)

# Step 6: Display the summary statistics DataFrame
print("\nNumeric Feature Summary Analysis for Test Data:\n")
print(summary_stats)

# Step 7: Print skewness and kurtosis separately to understand distributions
print("\nSkewness of Numeric Features in Test Data:\n")
print(test_df[numeric_cols].apply(skew))

print("\nKurtosis of Numeric Features in Test Data:\n")
print(test_df[numeric_cols].apply(kurtosis))
```

```
Numeric Feature Summary Analysis for Test Data:

                                     count        mean        std  \
Basic_Demos-Age                       20.0   10.750000   3.725799
Basic_Demos-Sex                       20.0    0.400000   0.502625
CGAS-CGAS_Score                        8.0   62.500000  11.275764
Physical-BMI                          13.0   19.835939   4.927625
Physical-Height                       13.0   52.961538   6.942357
Physical-Weight                       13.0   79.200000  23.632181
Physical-Waist_Circumference           5.0   25.400000   3.130495
Physical-Diastolic_BP                 11.0   70.545455  18.806189
Physical-HeartRate                    12.0   81.666667   9.316001
Physical-Systolic_BP                  11.0  117.545455  21.262002
Fitness_Endurance-Max_Stage            3.0    5.000000   1.000000
Fitness_Endurance-Time_Mins            3.0    7.000000   2.000000
Fitness_Endurance-Time_Sec             3.0   34.000000   2.645751
```

FGC-FGC_CU	13.0	8.692308	7.899205
FGC-FGC_CU_Zone	13.0	0.461538	0.518875
FGC-FGC_GSND	5.0	16.160000	4.879857
FGC-FGC_GSND_Zone	5.0	1.600000	0.547723
FGC-FGC_GSD	5.0	16.740000	3.990363
FGC-FGC_GSD_Zone	5.0	1.600000	0.547723
FGC-FGC_PU	13.0	4.000000	5.627314
FGC-FGC_PU_Zone	13.0	0.153846	0.375534
FGC-FGC_SRL	13.0	7.500000	4.000000
FGC-FGC_SRL_Zone	13.0	0.538462	0.518875
FGC-FGC_SRR	13.0	7.961538	4.436879
FGC-FGC_SRR_Zone	13.0	0.615385	0.506370
FGC-FGC_TL	13.0	7.961538	3.152126
FGC-FGC_TL_Zone	13.0	0.692308	0.480384
BIA-BIA_Activity_Level_num	8.0	2.625000	1.060660
BIA-BIA_BMC	8.0	3.636360	0.898087
BIA-BIA_BMI	8.0	19.284788	4.876077
BIA-BIA_BMR	8.0	1111.248000	143.724879
BIA-BIA_DEE	8.0	1886.912500	486.140935
BIA-BIA_ECW	8.0	16.681051	7.651128
BIA-BIA_FFM	8.0	60.625612	15.308597
BIA-BIA_FFMI	8.0	14.432937	1.227543
BIA-BIA_FMI	8.0	4.851857	3.728203
BIA-BIA_Fat	8.0	21.799390	19.920902
BIA-BIA_Frame_num	8.0	1.625000	0.517549
BIA-BIA_ICW	8.0	28.486750	5.099449
BIA-BIA_LDM	8.0	15.457795	4.021153
BIA-BIA_LST	8.0	56.989275	14.490362
BIA-BIA_SMM	8.0	25.985962	7.479799
BIA-BIA_TBW	8.0	45.167825	11.940000
PAQ_A-PAQ_A_Total	1.0	1.040000	NaN
PAQ_C-PAQ_C_Total	9.0	2.372333	1.080099
SDS-SDS_Total_Raw	10.0	36.800000	5.533735
SDS-SDS_Total_T	10.0	52.300000	7.024560
PreInt_EduHx-computerinternet_hoursday	16.0	1.437500	1.152895

	min	25%	50%	\
Basic_Demos-Age	5.00000	9.000000	10.000000	
Basic_Demos-Sex	0.00000	0.000000	0.000000	
CGAS-CGAS_Score	50.00000	51.000000	63.000000	
Physical-BMI	14.03559	16.861286	18.292347	
Physical-Height	37.50000	48.000000	55.000000	
Physical-Weight	46.00000	60.200000	81.600000	
Physical-Waist_Circumference	22.00000	24.000000	24.000000	
Physical-Diastolic_BP	57.00000	60.500000	63.000000	
Physical-HeartRate	70.00000	74.500000	80.000000	
Physical-Systolic_BP	95.00000	102.500000	116.000000	
Fitness_Endurance-Max_Stage	4.00000	4.500000	5.000000	

Fitness_Endurance-Time_Mins	5.00000	6.000000	7.000000
Fitness_Endurance-Time_Sec	32.00000	32.500000	33.000000
FGC-FGC_CU	0.00000	3.000000	6.000000
FGC-FGC_CU_Zone	0.00000	0.000000	0.000000
FGC-FGC_GSND	10.20000	12.600000	16.500000
FGC-FGC_GSND_Zone	1.00000	1.000000	2.000000
FGC-FGC_GSD	11.10000	14.700000	17.900000
FGC-FGC_GSD_Zone	1.00000	1.000000	2.000000
FGC-FGC_PU	0.00000	0.000000	2.000000
FGC-FGC_PU_Zone	0.00000	0.000000	0.000000
FGC-FGC_SRL	0.00000	7.000000	8.000000
FGC-FGC_SRL_Zone	0.00000	0.000000	1.000000
FGC-FGC_SRR	0.00000	6.000000	9.500000
FGC-FGC_SRR_Zone	0.00000	0.000000	1.000000
FGC-FGC_TL	3.00000	6.000000	7.000000
FGC-FGC_TL_Zone	0.00000	0.000000	1.000000
BIA-BIA_Activity_Level_num	2.00000	2.000000	2.000000
BIA-BIA_BMC	2.57949	2.729900	3.812310
BIA-BIA_BMI	14.03710	16.875175	17.784050
BIA-BIA_BMR	932.49800	986.466500	1133.645000
BIA-BIA_DEE	1492.00000	1503.120000	1852.720000
BIA-BIA_ECW	6.01993	13.423195	15.960000
BIA-BIA_FFM	41.58620	47.334825	63.011350
BIA-BIA_FFMI	12.82540	13.765575	14.081900
BIA-BIA_FMI	1.21172	3.153410	3.737140
BIA-BIA_Fat	3.97085	10.625893	17.535850
BIA-BIA_Frame_num	1.00000	1.000000	2.000000
BIA-BIA_ICW	21.03520	24.230725	29.470400
BIA-BIA_LDM	8.89536	13.815400	16.402450
BIA-BIA_LST	38.91770	44.627250	59.199050
BIA-BIA_SMM	15.41070	19.801775	26.337750
BIA-BIA_TBW	27.05520	37.245575	46.608850
PAQ_A-PAQ_A_Total	1.04000	1.040000	1.040000
PAQ_C-PAQ_C_Total	1.10000	1.270000	2.340000
SDS-SDS_Total_Raw	27.00000	33.500000	37.500000
SDS-SDS_Total_T	40.00000	47.750000	53.500000
PreInt_EduHx-computerinternet_hoursday	0.00000	0.000000	2.000000

	75%	max	Skewness \
Basic_Demos-Age	12.250000	19.000000	0.514853
Basic_Demos-Sex	1.000000	1.000000	0.408248
CGAS-CGAS_Score	71.000000	80.000000	NaN
Physical-BMI	21.079065	30.094649	NaN
Physical-Height	57.750000	60.000000	NaN
Physical-Weight	85.600000	121.600000	NaN
Physical-Waist_Circumference	27.000000	30.000000	NaN
Physical-Diastolic_BP	73.000000	123.000000	NaN
Physical-HeartRate	90.250000	97.000000	NaN

Physical-Systolic_BP	119.500000	163.000000	NaN
Fitness_Endurance-Max_Stage	5.500000	6.000000	NaN
Fitness_Endurance-Time_Mins	8.000000	9.000000	NaN
Fitness_Endurance-Time_Sec	35.000000	37.000000	NaN
FGC-FGC_CU	12.000000	24.000000	NaN
FGC-FGC_CU_Zone	1.000000	1.000000	NaN
FGC-FGC_GSND	19.200000	22.300000	NaN
FGC-FGC_GSND_Zone	2.000000	2.000000	NaN
FGC-FGC_GSD	18.400000	21.600000	NaN
FGC-FGC_GSD_Zone	2.000000	2.000000	NaN
FGC-FGC_PU	6.000000	20.000000	NaN
FGC-FGC_PU_Zone	0.000000	1.000000	NaN
FGC-FGC_SRL	10.500000	12.000000	NaN
FGC-FGC_SRL_Zone	1.000000	1.000000	NaN
FGC-FGC_SRR	11.000000	15.000000	NaN
FGC-FGC_SRR_Zone	1.000000	1.000000	NaN
FGC-FGC_TL	11.000000	12.500000	NaN
FGC-FGC_TL_Zone	1.000000	1.000000	NaN
BIA-BIA_Activity_Level_num	3.000000	5.000000	NaN
BIA-BIA_BMC	4.125535	5.080250	NaN
BIA-BIA_BMI	20.017525	30.186500	NaN
BIA-BIA_BMR	1194.895000	1330.970000	NaN
BIA-BIA_DEE	1941.692500	2974.710000	NaN
BIA-BIA_ECW	20.450875	30.212400	NaN
BIA-BIA_FFM	69.535100	84.028500	NaN
BIA-BIA_FFMI	14.939925	16.687700	NaN
BIA-BIA_FMI	5.077595	13.498800	NaN
BIA-BIA_Fat	22.444175	67.971500	NaN
BIA-BIA_Frame_num	2.000000	2.000000	NaN
BIA-BIA_ICW	31.398725	36.057200	NaN
BIA-BIA_LDM	17.674625	20.902000	NaN
BIA-BIA_LST	65.222050	79.698200	NaN
BIA-BIA_SMM	30.421100	36.223200	NaN
BIA-BIA_TBW	51.860475	63.126500	NaN
PAQ_A-PAQ_A_Total	1.040000	1.040000	NaN
PAQ_C-PAQ_C_Total	3.020000	4.110000	NaN
SDS-SDS_Total_Raw	39.750000	46.000000	NaN
SDS-SDS_Total_T	55.750000	64.000000	NaN
PreInt_EduHx-computerinternet_hoursday	2.000000	3.000000	NaN

	Kurtosis	Lower_Bound	Upper_Bound	\
Basic_Demos-Age	0.073754	4.125000	17.125000	
Basic_Demos-Sex	-1.833333	-1.500000	2.500000	
CGAS-CGAS_Score	NaN	21.000000	101.000000	
Physical-BMI	NaN	10.534618	27.405733	
Physical-Height	NaN	33.375000	72.375000	
Physical-Weight	NaN	22.100000	123.700000	
Physical-Waist_Circumference	NaN	19.500000	31.500000	

Physical-Diastolic_BP	NaN	41.750000	91.750000
Physical-HeartRate	NaN	50.875000	113.875000
Physical-Systolic_BP	NaN	77.000000	145.000000
Fitness_Endurance-Max_Stage	NaN	3.000000	7.000000
Fitness_Endurance-Time_Mins	NaN	3.000000	11.000000
Fitness_Endurance-Time_Sec	NaN	28.750000	38.750000
FGC-FGC_CU	NaN	-10.500000	25.500000
FGC-FGC_CU_Zone	NaN	-1.500000	2.500000
FGC-FGC_GSND	NaN	2.700000	29.100000
FGC-FGC_GSND_Zone	NaN	-0.500000	3.500000
FGC-FGC_GSD	NaN	9.150000	23.950000
FGC-FGC_GSD_Zone	NaN	-0.500000	3.500000
FGC-FGC_PU	NaN	-9.000000	15.000000
FGC-FGC_PU_Zone	NaN	0.000000	0.000000
FGC-FGC_SRL	NaN	1.750000	15.750000
FGC-FGC_SRL_Zone	NaN	-1.500000	2.500000
FGC-FGC_SRR	NaN	-1.500000	18.500000
FGC-FGC_SRR_Zone	NaN	-1.500000	2.500000
FGC-FGC_TL	NaN	-1.500000	18.500000
FGC-FGC_TL_Zone	NaN	-1.500000	2.500000
BIA-BIA_Activity_Level_num	NaN	0.500000	4.500000
BIA-BIA_BMC	NaN	0.636448	6.218987
BIA-BIA_BMI	NaN	12.161650	24.731050
BIA-BIA_BMR	NaN	673.823750	1507.537750
BIA-BIA_DEE	NaN	845.261250	2599.551250
BIA-BIA_ECW	NaN	2.881675	30.992395
BIA-BIA_FFM	NaN	14.034413	102.835512
BIA-BIA_FFMI	NaN	12.004050	16.701450
BIA-BIA_FMI	NaN	0.267133	7.963872
BIA-BIA_Fat	NaN	-7.101531	40.171599
BIA-BIA_Frame_num	NaN	-0.500000	3.500000
BIA-BIA_ICW	NaN	13.478725	42.150725
BIA-BIA_LDM	NaN	8.026563	23.463462
BIA-BIA_LST	NaN	13.735050	96.114250
BIA-BIA_SMM	NaN	3.872787	46.350088
BIA-BIA_TBW	NaN	15.323225	73.782825
PAQ_A-PAQ_A_Total	NaN	1.040000	1.040000
PAQ_C-PAQ_C_Total	NaN	-1.355000	5.645000
SDS-SDS_Total_Raw	NaN	24.125000	49.125000
SDS-SDS_Total_T	NaN	35.750000	67.750000
PreInt_EduHx-computerinternet_hoursday	NaN	-3.000000	5.000000

	Outliers
Basic_Demos-Age	2
Basic_Demos-Sex	0
CGAS-CGAS_Score	0
Physical-BMI	2
Physical-Height	0

```
Physical-Weight                              0
Physical-Waist_Circumference                 0
Physical-Diastolic_BP                        1
Physical-HeartRate                           0
Physical-Systolic_BP                         2
Fitness_Endurance-Max_Stage                  0
Fitness_Endurance-Time_Mins                  0
Fitness_Endurance-Time_Sec                   0
FGC-FGC_CU                                   0
FGC-FGC_CU_Zone                              0
FGC-FGC_GSND                                 0
FGC-FGC_GSND_Zone                            0
FGC-FGC_GSD                                  0
FGC-FGC_GSD_Zone                             0
FGC-FGC_PU                                   1
FGC-FGC_PU_Zone                              2
FGC-FGC_SRL                                  2
FGC-FGC_SRL_Zone                             0
FGC-FGC_SRR                                  0
FGC-FGC_SRR_Zone                             0
FGC-FGC_TL                                   0
FGC-FGC_TL_Zone                              0
BIA-BIA_Activity_Level_num                   1
BIA-BIA_BMC                                  0
BIA-BIA_BMI                                  1
BIA-BIA_BMR                                  0
BIA-BIA_DEE                                  1
BIA-BIA_ECW                                  0
BIA-BIA_FFM                                  0
BIA-BIA_FFMI                                 0
BIA-BIA_FMI                                  1
BIA-BIA_Fat                                  1
BIA-BIA_Frame_num                            0
BIA-BIA_ICW                                  0
BIA-BIA_LDM                                  0
BIA-BIA_LST                                  0
BIA-BIA_SMM                                  0
BIA-BIA_TBW                                  0
PAQ_A-PAQ_A_Total                            0
PAQ_C-PAQ_C_Total                            0
SDS-SDS_Total_Raw                            0
SDS-SDS_Total_T                              0
PreInt_EduHx-computerinternet_hoursday       0

Skewness of Numeric Features in Test Data:

Basic_Demos-Age                         0.514853
Basic_Demos-Sex                         0.408248
```

```
CGAS-CGAS_Score                                 NaN
Physical-BMI                                    NaN
Physical-Height                                 NaN
Physical-Weight                                 NaN
Physical-Waist_Circumference                    NaN
Physical-Diastolic_BP                           NaN
Physical-HeartRate                              NaN
Physical-Systolic_BP                            NaN
Fitness_Endurance-Max_Stage                     NaN
Fitness_Endurance-Time_Mins                     NaN
Fitness_Endurance-Time_Sec                      NaN
FGC-FGC_CU                                      NaN
FGC-FGC_CU_Zone                                 NaN
FGC-FGC_GSND                                    NaN
FGC-FGC_GSND_Zone                               NaN
FGC-FGC_GSD                                     NaN
FGC-FGC_GSD_Zone                                NaN
FGC-FGC_PU                                      NaN
FGC-FGC_PU_Zone                                 NaN
FGC-FGC_SRL                                     NaN
FGC-FGC_SRL_Zone                                NaN
FGC-FGC_SRR                                     NaN
FGC-FGC_SRR_Zone                                NaN
FGC-FGC_TL                                      NaN
FGC-FGC_TL_Zone                                 NaN
BIA-BIA_Activity_Level_num                      NaN
BIA-BIA_BMC                                     NaN
BIA-BIA_BMI                                     NaN
BIA-BIA_BMR                                     NaN
BIA-BIA_DEE                                     NaN
BIA-BIA_ECW                                     NaN
BIA-BIA_FFM                                     NaN
BIA-BIA_FFMI                                    NaN
BIA-BIA_FMI                                     NaN
BIA-BIA_Fat                                     NaN
BIA-BIA_Frame_num                               NaN
BIA-BIA_ICW                                     NaN
BIA-BIA_LDM                                     NaN
BIA-BIA_LST                                     NaN
BIA-BIA_SMM                                     NaN
BIA-BIA_TBW                                     NaN
PAQ_A-PAQ_A_Total                               NaN
PAQ_C-PAQ_C_Total                               NaN
SDS-SDS_Total_Raw                               NaN
SDS-SDS_Total_T                                 NaN
PreInt_EduHx-computerinternet_hoursday          NaN
dtype: float64
```

```
Kurtosis of Numeric Features in Test Data:

Basic_Demos-Age                              0.073754
Basic_Demos-Sex                             -1.833333
CGAS-CGAS_Score                                   NaN
Physical-BMI                                      NaN
Physical-Height                                   NaN
Physical-Weight                                   NaN
Physical-Waist_Circumference                      NaN
Physical-Diastolic_BP                             NaN
Physical-HeartRate                                NaN
Physical-Systolic_BP                              NaN
Fitness_Endurance-Max_Stage                       NaN
Fitness_Endurance-Time_Mins                       NaN
Fitness_Endurance-Time_Sec                        NaN
FGC-FGC_CU                                        NaN
FGC-FGC_CU_Zone                                   NaN
FGC-FGC_GSND                                      NaN
FGC-FGC_GSND_Zone                                 NaN
FGC-FGC_GSD                                       NaN
FGC-FGC_GSD_Zone                                  NaN
FGC-FGC_PU                                        NaN
FGC-FGC_PU_Zone                                   NaN
FGC-FGC_SRL                                       NaN
FGC-FGC_SRL_Zone                                  NaN
FGC-FGC_SRR                                       NaN
FGC-FGC_SRR_Zone                                  NaN
FGC-FGC_TL                                        NaN
FGC-FGC_TL_Zone                                   NaN
BIA-BIA_Activity_Level_num                        NaN
BIA-BIA_BMC                                       NaN
BIA-BIA_BMI                                       NaN
BIA-BIA_BMR                                       NaN
BIA-BIA_DEE                                       NaN
BIA-BIA_ECW                                       NaN
BIA-BIA_FFM                                       NaN
BIA-BIA_FFMI                                      NaN
BIA-BIA_FMI                                       NaN
BIA-BIA_Fat                                       NaN
BIA-BIA_Frame_num                                 NaN
BIA-BIA_ICW                                       NaN
BIA-BIA_LDM                                       NaN
BIA-BIA_LST                                       NaN
BIA-BIA_SMM                                       NaN
BIA-BIA_TBW                                       NaN
PAQ_A-PAQ_A_Total                                 NaN
PAQ_C-PAQ_C_Total                                 NaN
SDS-SDS_Total_Raw                                 NaN
```

```
SDS-SDS_Total_T                               NaN
PreInt_EduHx-computerinternet_hoursday        NaN
dtype: float64
```

[59]:

```
ID_arr = list(test_df['id'])
test_df.head()
```

[59]:

```
         id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0  00008ff9                      Fall                5                0
1  000fd460                    Summer                9                0
2  00105258                    Summer               10                1
3  00115b9f                    Winter                9                0
4  0016bb22                    Spring               18                1

  CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  Physical-Height  \
0      Winter             51.0            Fall     16.877316             46.0
1         NaN              NaN            Fall     14.035590             48.0
2        Fall             71.0            Fall     16.648696             56.5
3        Fall             71.0          Summer     18.292347             56.0
4      Summer              NaN             NaN           NaN              NaN

   Physical-Weight  Physical-Waist_Circumference  Physical-Diastolic_BP  \
0             50.8                           NaN                    NaN
1             46.0                          22.0                   75.0
2             75.6                           NaN                   65.0
3             81.6                           NaN                   60.0
4              NaN                           NaN                    NaN

   Physical-HeartRate  Physical-Systolic_BP Fitness_Endurance-Season  \
0                 NaN                   NaN                      NaN
1                70.0                 122.0                      NaN
2                94.0                 117.0                     Fall
3                97.0                 117.0                   Summer
4                 NaN                   NaN                      NaN

   Fitness_Endurance-Max_Stage  Fitness_Endurance-Time_Mins  \
0                          NaN                          NaN
1                          NaN                          NaN
2                          5.0                          7.0
3                          6.0                          9.0
4                          NaN                          NaN

   Fitness_Endurance-Time_Sec FGC-Season  FGC-FGC_CU  FGC-FGC_CU_Zone  \
0                         NaN       Fall         0.0              0.0
1                         NaN       Fall         3.0              0.0
2                        33.0       Fall        20.0              1.0
3                        37.0     Summer        18.0              1.0
```

```
4                        NaN         NaN          NaN               NaN

   FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  FGC-FGC_GSD_Zone  FGC-FGC_PU  \
0           NaN                NaN          NaN               NaN         0.0
1           NaN                NaN          NaN               NaN         5.0
2          10.2                1.0         14.7               2.0         7.0
3           NaN                NaN          NaN               NaN         5.0
4           NaN                NaN          NaN               NaN         NaN

   FGC-FGC_PU_Zone  FGC-FGC_SRL  FGC-FGC_SRL_Zone  FGC-FGC_SRR  \
0              0.0          7.0               0.0          6.0
1              0.0         11.0               1.0         11.0
2              1.0         10.0               1.0         10.0
3              0.0          7.0               0.0          7.0
4              NaN          NaN               NaN          NaN

   FGC-FGC_SRR_Zone  FGC-FGC_TL  FGC-FGC_TL_Zone BIA-Season  \
0               0.0         6.0              1.0       Fall
1               1.0         3.0              0.0     Winter
2               1.0         5.0              0.0        NaN
3               0.0         7.0              1.0     Summer
4               NaN         NaN              NaN        NaN

   BIA-BIA_Activity_Level_num  BIA-BIA_BMC  BIA-BIA_BMI  BIA-BIA_BMR  \
0                         2.0      2.66855      16.8792      932.498
1                         2.0      2.57949      14.0371      936.656
2                         NaN          NaN          NaN          NaN
3                         3.0      3.84191      18.2943     1131.430
4                         NaN          NaN          NaN          NaN

   BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  BIA-BIA_FFMI  BIA-BIA_FMI  \
0      1492.00      8.25598      41.5862       13.8177      3.06143
1      1498.65      6.01993      42.0291       12.8254      1.21172
2          NaN          NaN          NaN           NaN          NaN
3      1923.44     15.59250      62.7757       14.0740      4.22033
4          NaN          NaN          NaN           NaN          NaN

   BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  BIA-BIA_LDM  BIA-BIA_LST  \
0      9.21377                1.0      24.4349      8.89536      38.9177
1      3.97085                1.0      21.0352     14.97400      39.4497
2          NaN                NaN          NaN          NaN          NaN
3     18.82430                2.0      30.4041     16.77900      58.9338
4          NaN                NaN          NaN          NaN          NaN

   BIA-BIA_SMM  BIA-BIA_TBW PAQ_A-Season  PAQ_A-PAQ_A_Total PAQ_C-Season  \
0      19.5413      32.6909          NaN                NaN          NaN
1      15.4107      27.0552          NaN                NaN         Fall
```

```
2         NaN         NaN         NaN                   NaN       Summer
3     26.4798     45.9966         NaN                   NaN       Winter
4         NaN         NaN      Summer                  1.04          NaN

   PAQ_C-PAQ_C_Total SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  \
0                NaN        NaN                NaN              NaN
1              2.340       Fall               46.0             64.0
2              2.170       Fall               38.0             54.0
3              2.451     Summer               31.0             45.0
4                NaN        NaN                NaN              NaN

  PreInt_EduHx-Season  PreInt_EduHx-computerinternet_hoursday
0                Fall                                     3.0
1              Summer                                     0.0
2              Summer                                     2.0
3              Winter                                     0.0
4                 NaN                                     NaN
```

[60]:

```python
# Step 1: Handle Extreme Outliers in CGAS-CGAS_Score for test_df

# Calculate the IQR for CGAS-CGAS_Score
Q1_cgas_test = test_df['CGAS-CGAS_Score'].quantile(0.25)
Q3_cgas_test = test_df['CGAS-CGAS_Score'].quantile(0.75)
IQR_cgas_test = Q3_cgas_test - Q1_cgas_test

# Define bounds for outlier removal
lower_bound_cgas_test = Q1_cgas_test - 1.5 * IQR_cgas_test
upper_bound_cgas_test = Q3_cgas_test + 1.5 * IQR_cgas_test

# Remove extreme outliers in CGAS-CGAS_Score
test_df = test_df[(test_df['CGAS-CGAS_Score'] >= lower_bound_cgas_test) &
                  (test_df['CGAS-CGAS_Score'] <= upper_bound_cgas_test)]

# Display the updated CGAS-CGAS_Score statistics
print("\nUpdated CGAS-CGAS_Score Summary After Outlier Removal:")
print(test_df['CGAS-CGAS_Score'].describe())
```

```
Updated CGAS-CGAS_Score Summary After Outlier Removal:
count     8.000000
mean     62.500000
std      11.275764
min      50.000000
25%      51.000000
50%      63.000000
75%      71.000000
max      80.000000
Name: CGAS-CGAS_Score, dtype: float64
```

[61]:

```python
# Step 2: Handle Outliers in Physical-BMI Based on Age Group Thresholds for␣
↪test_df

# Create age group segments for test_df
age_bins_test = [5, 10, 15, 20, 25]
age_labels_test = ['5-10', '11-15', '16-20', '21-25']
test_df['Age_Group'] = pd.cut(test_df['Basic_Demos-Age'], bins=age_bins_test,␣
↪labels=age_labels_test)

# Define the scientifically accepted BMI thresholds for each age group (same as␣
↪train_df)
bmi_thresholds_test = {
    '5-10': (14, 24),  # Min, Max BMI for age group 5-10
    '11-15': (16, 28),
    '16-20': (18, 30),
    '21-25': (19, 32)
}

# Remove BMI outliers based on the thresholds defined above
for age_group_test, (bmi_min_test, bmi_max_test) in bmi_thresholds_test.items():
    # Define the condition for outlier removal based on BMI thresholds
    condition_test = (
        (test_df['Age_Group'] == age_group_test) &
        ((test_df['Physical-BMI'] < bmi_min_test) | (test_df['Physical-BMI'] >␣
↪bmi_max_test))
    )
    # Remove outliers for the given age group
    test_df = test_df[~condition_test]

# Display the updated Physical-BMI statistics by age group
print("\nUpdated Physical-BMI Summary After Outlier Removal (Test Data):")
print(test_df.groupby('Age_Group')['Physical-BMI'].describe())
```

```
Updated Physical-BMI Summary After Outlier Removal (Test Data):
           count       mean       std        min        25%        50%  \
Age_Group
5-10         2.0  17.470522  1.162237  16.648696  17.059609  17.470522
11-15        1.0  22.279952       NaN  22.279952  22.279952  22.279952
16-20        0.0        NaN       NaN        NaN        NaN        NaN
21-25        0.0        NaN       NaN        NaN        NaN        NaN

                 75%        max
Age_Group
5-10       17.881434  18.292347
11-15      22.279952  22.279952
16-20            NaN        NaN
```

```
21-25             NaN        NaN
```

```
/var/tmp/ipykernel_7796/3117403607.py:28: FutureWarning: The default of
observed=False is deprecated and will be changed to True in a future version of
pandas. Pass observed=False to retain current behavior or observed=True to adopt
the future default and silence this warning.
  print(test_df.groupby('Age_Group')['Physical-BMI'].describe())
```

[62]:
```
# Step 3: Age Group-Based Analysis for test_df

# Update numeric_cols to only include columns that are still in test_df
numeric_cols_test = [col for col in numeric_cols if col in test_df.columns]

# Create a summary DataFrame to store age group-based analysis results for␣
 ↪test_df
age_group_summary_test = test_df.groupby('Age_Group')[numeric_cols_test].
 ↪agg(['mean', 'median', 'std'])

# Display the summary statistics for numeric features based on age group for␣
 ↪test_df
print("\nAge Group-Based Summary Statistics for Numeric Features (Test Data):")
print(age_group_summary_test)
```

```
Age Group-Based Summary Statistics for Numeric Features (Test Data):
          Basic_Demos-Age                    Basic_Demos-Sex                   \
                     mean median       std              mean median       std
Age_Group
5-10                  9.5    9.5  0.707107               0.5    0.5  0.707107
11-15                12.0   12.0  1.414214               0.5    0.5  0.707107
16-20                 NaN    NaN       NaN               NaN    NaN       NaN
21-25                 NaN    NaN       NaN               NaN    NaN       NaN

          CGAS-CGAS_Score                     Physical-BMI                      \
                     mean median        std           mean     median       std
Age_Group
5-10                 71.0   71.0   0.000000      17.470522  17.470522  1.162237
11-15                58.0   58.0  11.313708      22.279952  22.279952       NaN
16-20                 NaN    NaN        NaN            NaN        NaN       NaN
21-25                 NaN    NaN        NaN            NaN        NaN       NaN

          Physical-Height                    Physical-Weight                   \
                     mean median       std              mean median       std
Age_Group
5-10                56.25  56.25  0.353553              78.6   78.6  4.242641
11-15               59.50  59.50       NaN             112.2  112.2       NaN
16-20                 NaN    NaN       NaN               NaN    NaN       NaN
21-25                 NaN    NaN       NaN               NaN    NaN       NaN
```

```
          Physical-Waist_Circumference             Physical-Diastolic_BP  \
                                  mean median std                  mean
Age_Group
5-10                               NaN    NaN NaN                  62.5
11-15                              NaN    NaN NaN                  60.0
16-20                              NaN    NaN NaN                   NaN
21-25                              NaN    NaN NaN                   NaN

                           Physical-HeartRate                  \
          median       std               mean median      std
Age_Group
5-10        62.5  3.535534               95.5   95.5  2.12132
11-15       60.0       NaN               73.0   73.0      NaN
16-20        NaN       NaN                NaN    NaN      NaN
21-25        NaN       NaN                NaN    NaN      NaN

          Physical-Systolic_BP             Fitness_Endurance-Max_Stage         \
                          mean median  std                        mean median
Age_Group
5-10                     117.0  117.0  0.0                         5.5    5.5
11-15                    102.0  102.0  NaN                         NaN    NaN
16-20                      NaN    NaN  NaN                         NaN    NaN
21-25                      NaN    NaN  NaN                         NaN    NaN

                    Fitness_Endurance-Time_Mins                  \
                std                        mean median       std
Age_Group
5-10       0.707107                         8.0    8.0  1.414214
11-15           NaN                         NaN    NaN       NaN
16-20           NaN                         NaN    NaN       NaN
21-25           NaN                         NaN    NaN       NaN

          Fitness_Endurance-Time_Sec                    FGC-FGC_CU         \
                                mean median       std       mean median
Age_Group
5-10                            35.0   35.0  2.828427       19.0   19.0
11-15                            NaN    NaN       NaN       12.0   12.0
16-20                            NaN    NaN       NaN        NaN    NaN
21-25                            NaN    NaN       NaN        NaN    NaN

                    FGC-FGC_CU_Zone             FGC-FGC_GSND               \
                std           mean median  std          mean median std
Age_Group
5-10       1.414214            1.0    1.0  0.0          10.2   10.2 NaN
11-15           NaN            0.0    0.0  NaN          16.5   16.5 NaN
16-20           NaN            NaN    NaN  NaN           NaN    NaN NaN
21-25           NaN            NaN    NaN  NaN           NaN    NaN NaN
```

	FGC-FGC_GSND_Zone			FGC-FGC_GSD			\
	mean	median	std	mean	median	std	
Age_Group							
5-10	1.0	1.0	NaN	14.7	14.7	NaN	
11-15	2.0	2.0	NaN	17.9	17.9	NaN	
16-20	NaN	NaN	NaN	NaN	NaN	NaN	
21-25	NaN	NaN	NaN	NaN	NaN	NaN	

	FGC-FGC_GSD_Zone			FGC-FGC_PU			\
	mean	median	std	mean	median	std	
Age_Group							
5-10	2.0	2.0	NaN	6.0	6.0	1.414214	
11-15	2.0	2.0	NaN	6.0	6.0	NaN	
16-20	NaN	NaN	NaN	NaN	NaN	NaN	
21-25	NaN	NaN	NaN	NaN	NaN	NaN	

	FGC-FGC_PU_Zone			FGC-FGC_SRL			\
	mean	median	std	mean	median	std	
Age_Group							
5-10	0.5	0.5	0.707107	8.5	8.5	2.12132	
11-15	0.0	0.0	NaN	10.0	10.0	NaN	
16-20	NaN	NaN	NaN	NaN	NaN	NaN	
21-25	NaN	NaN	NaN	NaN	NaN	NaN	

	FGC-FGC_SRL_Zone			FGC-FGC_SRR			\
	mean	median	std	mean	median	std	
Age_Group							
5-10	0.5	0.5	0.707107	8.5	8.5	2.12132	
11-15	1.0	1.0	NaN	11.0	11.0	NaN	
16-20	NaN	NaN	NaN	NaN	NaN	NaN	
21-25	NaN	NaN	NaN	NaN	NaN	NaN	

	FGC-FGC_SRR_Zone			FGC-FGC_TL			\
	mean	median	std	mean	median	std	
Age_Group							
5-10	0.5	0.5	0.707107	6.0	6.0	1.414214	
11-15	1.0	1.0	NaN	8.0	8.0	NaN	
16-20	NaN	NaN	NaN	NaN	NaN	NaN	
21-25	NaN	NaN	NaN	NaN	NaN	NaN	

	FGC-FGC_TL_Zone			BIA-BIA_Activity_Level_num		\
	mean	median	std	mean	median	
Age_Group						
5-10	0.5	0.5	0.707107	3.0	3.0	
11-15	0.0	0.0	NaN	2.0	2.0	
16-20	NaN	NaN	NaN	NaN	NaN	
21-25	NaN	NaN	NaN	NaN	NaN	

```
          BIA-BIA_BMC             BIA-BIA_BMI             BIA-BIA_BMR  \
      std        mean   median std        mean   median std        mean
Age_Group
5-10      NaN     3.84191  3.84191 NaN     18.2943  18.2943 NaN     1131.43
11-15     NaN     4.33036  4.33036 NaN     30.1865  30.1865 NaN     1330.97
16-20     NaN         NaN      NaN NaN         NaN      NaN NaN         NaN
21-25     NaN         NaN      NaN NaN         NaN      NaN NaN         NaN

                   BIA-BIA_DEE             BIA-BIA_ECW             \
            median std        mean   median std        mean   median std
Age_Group
5-10        1131.43 NaN     1923.44  1923.44 NaN     15.5925  15.5925 NaN
11-15       1330.97 NaN     1996.45  1996.45 NaN     30.2124  30.2124 NaN
16-20           NaN NaN         NaN      NaN NaN         NaN      NaN NaN
21-25           NaN NaN         NaN      NaN NaN         NaN      NaN NaN

          BIA-BIA_FFM             BIA-BIA_FFMI             BIA-BIA_FMI  \
                 mean   median std        mean   median std        mean
Age_Group
5-10          62.7757  62.7757 NaN     14.0740  14.0740 NaN     4.22033
11-15         84.0285  84.0285 NaN     16.6877  16.6877 NaN    13.49880
16-20             NaN      NaN NaN         NaN      NaN NaN         NaN
21-25             NaN      NaN NaN         NaN      NaN NaN         NaN

                    BIA-BIA_Fat             BIA-BIA_Frame_num             \
             median std        mean   median std              mean median std
Age_Group
5-10         4.22033 NaN     18.8243  18.8243 NaN               2.0    2.0 NaN
11-15       13.49880 NaN     67.9715  67.9715 NaN               2.0    2.0 NaN
16-20            NaN NaN         NaN      NaN NaN               NaN    NaN NaN
21-25            NaN NaN         NaN      NaN NaN               NaN    NaN NaN

          BIA-BIA_ICW             BIA-BIA_LDM            BIA-BIA_LST  \
                 mean   median std        mean  median std        mean
Age_Group
5-10          30.4041  30.4041 NaN      16.779  16.779 NaN     58.9338
11-15         32.9141  32.9141 NaN      20.902  20.902 NaN     79.6982
16-20             NaN      NaN NaN         NaN     NaN NaN         NaN
21-25             NaN      NaN NaN         NaN     NaN NaN         NaN

                   BIA-BIA_SMM             BIA-BIA_TBW             \
            median std        mean   median std        mean   median std
Age_Group
5-10        58.9338 NaN     26.4798  26.4798 NaN     45.9966  45.9966 NaN
11-15       79.6982 NaN     35.3804  35.3804 NaN     63.1265  63.1265 NaN
16-20           NaN NaN         NaN      NaN NaN         NaN      NaN NaN
21-25           NaN NaN         NaN      NaN NaN         NaN      NaN NaN
```

```
         PAQ_A-PAQ_A_Total            PAQ_C-PAQ_C_Total                    \
                      mean median std              mean  median       std
Age_Group
5-10                   NaN    NaN NaN            2.3105  2.3105  0.198697
11-15                  NaN    NaN NaN            2.6050  2.6050  2.128391
16-20                  NaN    NaN NaN               NaN     NaN       NaN
21-25                  NaN    NaN NaN               NaN     NaN       NaN

         SDS-SDS_Total_Raw                 SDS-SDS_Total_T                   \
                      mean median       std            mean median       std
Age_Group
5-10                  34.5   34.5  4.949747            49.5   49.5  6.363961
11-15                 41.0   41.0  1.414214            57.5   57.5  2.121320
16-20                  NaN    NaN       NaN             NaN    NaN       NaN
21-25                  NaN    NaN       NaN             NaN    NaN       NaN

         PreInt_EduHx-computerinternet_hoursday
                                           mean median       std
Age_Group
5-10                                        1.0    1.0  1.414214
11-15                                       0.0    0.0  0.000000
16-20                                       NaN    NaN       NaN
21-25                                       NaN    NaN       NaN
```

```
/var/tmp/ipykernel_7796/1008554566.py:7: FutureWarning: The default of
observed=False is deprecated and will be changed to True in a future version of
pandas. Pass observed=False to retain current behavior or observed=True to adopt
the future default and silence this warning.
  age_group_summary_test =
test_df.groupby('Age_Group')[numeric_cols_test].agg(['mean', 'median', 'std'])
```

[63]:

```python
# Step 1: Get categorical columns in test_df
categorical_cols_test = test_df.select_dtypes(include=['object']).columns

# Step 2: Create a summary dictionary to store analysis results for each␣
 ↪categorical feature in test_df
categorical_summary_test = {}

for col in categorical_cols_test:
    # Get the value counts and missing percentage for test_df
    value_counts_test = test_df[col].value_counts(dropna=False)
    missing_percentage_test = test_df[col].isnull().mean() * 100

    # Store the summary statistics in the dictionary for test_df
    categorical_summary_test[col] = {
        'Value Counts': value_counts_test,
        'Missing Percentage': missing_percentage_test
```

```
    }

# Step 3: Create a DataFrame to display categorical feature summaries for␣
 ↪test_df
categorical_summary_test_df = pd.DataFrame({
    'Feature': [col for col in categorical_summary_test.keys()],
    'Value Counts': [str(categorical_summary_test[col]['Value Counts']) for col␣
 ↪in categorical_summary_test.keys()],
    'Missing Percentage (%)': [categorical_summary_test[col]['Missing␣
 ↪Percentage'] for col in categorical_summary_test.keys()]
})

# Display the summary DataFrame for test_df
print("Categorical Feature Analysis for Test Data:")
print(categorical_summary_test_df)

# Step 4: Check for low-frequency categories (less than 5% frequency) to␣
 ↪identify potential consolidation opportunities in test_df
low_freq_categories_test = {}
for col in categorical_cols_test:
    # Identify categories that make up less than 5% of the total count
    low_freq_test = categorical_summary_test[col]['Value␣
 ↪Counts'][categorical_summary_test[col]['Value Counts'] < 0.05 *␣
 ↪len(test_df)].index.tolist()
    if low_freq_test:
        low_freq_categories_test[col] = low_freq_test

# Display low-frequency categories if found for test_df
if low_freq_categories_test:
    print("\nLow-Frequency Categories Identified for Potential Consolidation in␣
 ↪Test Data:")
    for col, low_freq_test in low_freq_categories_test.items():
        print(f"Feature: {col}")
        print(f"Low-Frequency Categories: {low_freq_test}")
else:
    print("\nNo Low-Frequency Categories Identified for Consolidation in Test␣
 ↪Data.")
```

```
Categorical Feature Analysis for Test Data:
                      Feature  \
0                          id
1    Basic_Demos-Enroll_Season
2                  CGAS-Season
3              Physical-Season
4     Fitness_Endurance-Season
5                   FGC-Season
6                   BIA-Season
```

```
7                PAQ_A-Season
8                PAQ_C-Season
9                  SDS-Season
10        PreInt_EduHx-Season

                                                    Value Counts  Missing Percentage (%)
0   id\n00008ff9    1\n00105258    1\n00115b9f    …                 0.000000
1   Basic_Demos-Enroll_Season\nFall      2\nSpring…                 0.000000
2   CGAS-Season\nWinter    2\nFall      2\nSummer …                 0.000000
3   Physical-Season\nFall      2\nSummer    2\nNaN…                16.666667
4   Fitness_Endurance-Season\nNaN       3\nFall    …               50.000000
5   FGC-Season\nFall      2\nSummer    2\nWinter  …                 0.000000
6   BIA-Season\nNaN       3\nSummer    2\nFall    …                50.000000
7   PAQ_A-Season\nNaN    6\nName: count, dtype: int64               100.000000
8   PAQ_C-Season\nNaN       2\nWinter    2\nSummer…                33.333333
9   SDS-Season\nSummer    2\nNaN       1\nFall    …                16.666667
10  PreInt_EduHx-Season\nFall      2\nSpring    2\…                 0.000000

No Low-Frequency Categories Identified for Consolidation in Test Data.
```

[64]:

```python
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Extract numeric columns from test_df
numeric_features_test = test_df.select_dtypes(include=[np.number]).columns

# Step 2: Compute the correlation matrix for the numeric features in test_df
correlation_matrix_test = test_df[numeric_features_test].corr()

# Step 3: Display the correlation matrix to identify relationships between␣
↪features in test_df
print("Correlation Matrix for Numeric Features (Test Data):")
print(correlation_matrix_test)

# Step 4: Visualize the correlation matrix using a heatmap to identify␣
↪multicollinearity or strong relationships in test_df
plt.figure(figsize=(18, 12))
sns.heatmap(correlation_matrix_test, annot=True, cmap='coolwarm', linewidths=0.
↪5, fmt='.2f')
plt.title('Correlation Matrix Heatmap for Numeric Features (Test Data)')
plt.show()

# Step 5: Identify highly correlated pairs in test_df (correlation > 0.8 or <␣
↪-0.8)
highly_correlated_pairs_test = []
threshold_test = 0.8
```

```python
for i in range(len(correlation_matrix_test.columns)):
    for j in range(i + 1, len(correlation_matrix_test.columns)):
        if abs(correlation_matrix_test.iloc[i, j]) > threshold_test:
            highly_correlated_pairs_test.append(
                (correlation_matrix_test.columns[i], correlation_matrix_test.
 ↪columns[j], correlation_matrix_test.iloc[i, j])
            )

# Step 6: Display highly correlated pairs if found for test_df
if highly_correlated_pairs_test:
    print("\nHighly Correlated Pairs in Test Data (|correlation| > 0.8):")
    for pair in highly_correlated_pairs_test:
        print(f"Feature 1: {pair[0]}, Feature 2: {pair[1]}, Correlation:␣
 ↪{pair[2]:.2f}")
else:
    print("\nNo highly correlated pairs found with |correlation| > 0.8 in Test␣
 ↪Data.")
```

```
Correlation Matrix for Numeric Features (Test Data):
                                       Basic_Demos-Age  Basic_Demos-Sex  \
Basic_Demos-Age                           1.000000e+00         0.168497
Basic_Demos-Sex                           1.684968e-01         1.000000
CGAS-CGAS_Score                          -2.882110e-01         0.197949
Physical-BMI                              7.497721e-01         0.271616
Physical-Height                           9.640227e-01         0.184610
Physical-Weight                           9.749089e-01         0.256066
Physical-Waist_Circumference                       NaN              NaN
Physical-Diastolic_BP                    -2.119044e-02         0.420084
Physical-HeartRate                       -2.471208e-02        -0.653197
Physical-Systolic_BP                     -3.105422e-01        -0.530314
Fitness_Endurance-Max_Stage              -1.000000e+00        -1.000000
Fitness_Endurance-Time_Mins              -1.000000e+00        -1.000000
Fitness_Endurance-Time_Sec               -1.000000e+00        -1.000000
FGC-FGC_CU                                7.435859e-01         0.095173
FGC-FGC_CU_Zone                           2.923217e-01        -0.166667
FGC-FGC_GSND                              1.000000e+00              NaN
FGC-FGC_GSND_Zone                         1.000000e+00              NaN
FGC-FGC_GSD                               1.000000e+00              NaN
FGC-FGC_GSD_Zone                                   NaN              NaN
FGC-FGC_PU                                9.049739e-01         0.298719
FGC-FGC_PU_Zone                           2.603778e-01         0.408248
FGC-FGC_SRL                               2.792054e-01         0.993146
FGC-FGC_SRL_Zone                          3.720458e-01         1.000000
FGC-FGC_SRR                               5.505464e-01         0.968475
FGC-FGC_SRR_Zone                          3.720458e-01         1.000000
FGC-FGC_TL                                3.063858e-01         0.080064
FGC-FGC_TL_Zone                          -8.238157e-01        -0.666667
```

BIA-BIA_Activity_Level_num	-1.922963e-16	-0.500000
BIA-BIA_BMC	9.728373e-01	0.726757
BIA-BIA_BMI	9.103617e-01	0.995303
BIA-BIA_BMR	9.999996e-01	0.866466
BIA-BIA_DEE	9.251778e-01	0.611460
BIA-BIA_ECW	9.821499e-01	0.944617
BIA-BIA_FFM	9.999996e-01	0.866456
BIA-BIA_FFMI	9.035463e-01	0.996739
BIA-BIA_FMI	9.122150e-01	0.994857
BIA-BIA_Fat	9.321317e-01	0.988309
BIA-BIA_Frame_num	8.660254e-01	0.500000
BIA-BIA_ICW	9.733643e-01	0.728326
BIA-BIA_LDM	9.840399e-01	0.763230
BIA-BIA_LST	9.999439e-01	0.871274
BIA-BIA_SMM	9.974522e-01	0.899488
BIA-BIA_TBW	9.973791e-01	0.899932
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	5.649018e-01	0.632346
SDS-SDS_Total_Raw	4.295478e-01	0.241423
SDS-SDS_Total_T	3.838790e-01	0.244600
PreInt_EduHx-computerinternet_hoursday	-4.243216e-01	-0.137361

	CGAS-CGAS_Score	Physical-BMI \
Basic_Demos-Age	-0.288211	0.749772
Basic_Demos-Sex	0.197949	0.271616
CGAS-CGAS_Score	1.000000	-0.522120
Physical-BMI	-0.522120	1.000000
Physical-Height	-0.260793	0.607281
Physical-Weight	-0.419738	0.856963
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	0.297044	-0.622059
Physical-HeartRate	0.365655	-0.576780
Physical-Systolic_BP	0.630416	-0.803882
Fitness_Endurance-Max_Stage	NaN	1.000000
Fitness_Endurance-Time_Mins	NaN	1.000000
Fitness_Endurance-Time_Sec	NaN	1.000000
FGC-FGC_CU	0.132362	0.176773
FGC-FGC_CU_Zone	0.436314	-0.316501
FGC-FGC_GSND	-1.000000	1.000000
FGC-FGC_GSND_Zone	-1.000000	1.000000
FGC-FGC_GSD	-1.000000	1.000000
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	-0.059984	0.406402
FGC-FGC_PU_Zone	0.267187	-0.391428
FGC-FGC_SRL	0.313233	0.225014
FGC-FGC_SRL_Zone	0.245427	0.271616
FGC-FGC_SRR	0.177406	0.473849
FGC-FGC_SRR_Zone	0.245427	0.271616

```
FGC-FGC_TL                                 -0.225973    0.816606
FGC-FGC_TL_Zone                             0.279514   -0.466387
BIA-BIA_Activity_Level_num                  0.999109   -0.265082
BIA-BIA_BMC                                 0.190222    0.876671
BIA-BIA_BMI                                -0.451869    0.987489
BIA-BIA_BMR                                -0.043088    0.964459
BIA-BIA_DEE                                 0.340146    0.791473
BIA-BIA_ECW                                -0.229386    0.996876
BIA-BIA_FFM                                -0.043068    0.964454
BIA-BIA_FFMI                               -0.466245    0.984808
BIA-BIA_FMI                                -0.447849    0.988189
BIA-BIA_Fat                                -0.401140    0.994777
BIA-BIA_Frame_num                           0.463002    0.702503
BIA-BIA_ICW                                 0.187976    0.877769
BIA-BIA_LDM                                 0.136255    0.901666
BIA-BIA_LST                                -0.052789    0.966980
BIA-BIA_SMM                                -0.113375    0.980680
BIA-BIA_TBW                                -0.114386    0.980878
PAQ_A-PAQ_A_Total                                NaN         NaN
PAQ_C-PAQ_C_Total                          -0.750481    0.988267
SDS-SDS_Total_Raw                          -0.422847    0.374065
SDS-SDS_Total_T                            -0.372952    0.318156
PreInt_EduHx-computerinternet_hoursday     -0.366025   -0.566751

                                        Physical-Height  Physical-Weight  \
Basic_Demos-Age                                0.964023         0.974909
Basic_Demos-Sex                                0.184610         0.256066
CGAS-CGAS_Score                               -0.260793        -0.419738
Physical-BMI                                   0.607281         0.856963
Physical-Height                                1.000000         0.929547
Physical-Weight                                0.929547         1.000000
Physical-Waist_Circumference                        NaN              NaN
Physical-Diastolic_BP                          0.051154        -0.263705
Physical-HeartRate                             0.280378        -0.118328
Physical-Systolic_BP                          -0.016846        -0.419514
Fitness_Endurance-Max_Stage                   -1.000000         1.000000
Fitness_Endurance-Time_Mins                   -1.000000         1.000000
Fitness_Endurance-Time_Sec                    -1.000000         1.000000
FGC-FGC_CU                                     0.873480         0.644895
FGC-FGC_CU_Zone                                0.507679         0.175349
FGC-FGC_GSND                                   1.000000         1.000000
FGC-FGC_GSND_Zone                              1.000000         1.000000
FGC-FGC_GSD                                    1.000000         1.000000
FGC-FGC_GSD_Zone                                    NaN              NaN
FGC-FGC_PU                                     0.959337         0.811553
FGC-FGC_PU_Zone                                0.331076         0.043463
FGC-FGC_SRL                                    0.086848         0.170110
FGC-FGC_SRL_Zone                               0.184610         0.256066
```

FGC-FGC_SRR	0.364799	0.463933
FGC-FGC_SRR_Zone	0.184610	0.256066
FGC-FGC_TL	0.167866	0.479004
FGC-FGC_TL_Zone	-0.738442	-0.707660
BIA-BIA_Activity_Level_num	0.267828	0.001881
BIA-BIA_BMC	0.999296	0.973271
BIA-BIA_BMI	0.766273	0.909582
BIA-BIA_BMR	0.963230	0.999996
BIA-BIA_DEE	0.993028	0.925890
BIA-BIA_ECW	0.895890	0.981794
BIA-BIA_FFM	0.963236	0.999996
BIA-BIA_FFMI	0.755775	0.902739
BIA-BIA_FMI	0.769157	0.911443
BIA-BIA_Fat	0.801092	0.931449
BIA-BIA_Frame_num	0.968301	0.866964
BIA-BIA_ICW	0.999207	0.973794
BIA-BIA_LDM	0.995749	0.984373
BIA-BIA_LST	0.960575	0.999922
BIA-BIA_SMM	0.941906	0.997316
BIA-BIA_TBW	0.941563	0.997241
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	0.964600	0.999819
SDS-SDS_Total_Raw	0.088035	0.263425
SDS-SDS_Total_T	0.030251	0.198970
PreInt_EduHx-computerinternet_hoursday	-0.280890	-0.432538

	Physical-Waist_Circumference \
Basic_Demos-Age	NaN
Basic_Demos-Sex	NaN
CGAS-CGAS_Score	NaN
Physical-BMI	NaN
Physical-Height	NaN
Physical-Weight	NaN
Physical-Waist_Circumference	NaN
Physical-Diastolic_BP	NaN
Physical-HeartRate	NaN
Physical-Systolic_BP	NaN
Fitness_Endurance-Max_Stage	NaN
Fitness_Endurance-Time_Mins	NaN
Fitness_Endurance-Time_Sec	NaN
FGC-FGC_CU	NaN
FGC-FGC_CU_Zone	NaN
FGC-FGC_GSND	NaN
FGC-FGC_GSND_Zone	NaN
FGC-FGC_GSD	NaN
FGC-FGC_GSD_Zone	NaN
FGC-FGC_PU	NaN
FGC-FGC_PU_Zone	NaN

```
FGC-FGC_SRL                                             NaN
FGC-FGC_SRL_Zone                                        NaN
FGC-FGC_SRR                                             NaN
FGC-FGC_SRR_Zone                                        NaN
FGC-FGC_TL                                              NaN
FGC-FGC_TL_Zone                                         NaN
BIA-BIA_Activity_Level_num                              NaN
BIA-BIA_BMC                                             NaN
BIA-BIA_BMI                                             NaN
BIA-BIA_BMR                                             NaN
BIA-BIA_DEE                                             NaN
BIA-BIA_ECW                                             NaN
BIA-BIA_FFM                                             NaN
BIA-BIA_FFMI                                            NaN
BIA-BIA_FMI                                             NaN
BIA-BIA_Fat                                             NaN
BIA-BIA_Frame_num                                       NaN
BIA-BIA_ICW                                             NaN
BIA-BIA_LDM                                             NaN
BIA-BIA_LST                                             NaN
BIA-BIA_SMM                                             NaN
BIA-BIA_TBW                                             NaN
PAQ_A-PAQ_A_Total                                       NaN
PAQ_C-PAQ_C_Total                                       NaN
SDS-SDS_Total_Raw                                       NaN
SDS-SDS_Total_T                                         NaN
PreInt_EduHx-computerinternet_hoursday                  NaN

                                        Physical-Diastolic_BP  \
Basic_Demos-Age                                     -0.021190
Basic_Demos-Sex                                      0.420084
CGAS-CGAS_Score                                      0.297044
Physical-BMI                                        -0.622059
Physical-Height                                      0.051154
Physical-Weight                                     -0.263705
Physical-Waist_Circumference                              NaN
Physical-Diastolic_BP                                1.000000
Physical-HeartRate                                   0.411597
Physical-Systolic_BP                                 0.513355
Fitness_Endurance-Max_Stage                         -1.000000
Fitness_Endurance-Time_Mins                         -1.000000
Fitness_Endurance-Time_Sec                          -1.000000
FGC-FGC_CU                                           0.388967
FGC-FGC_CU_Zone                                      0.485071
FGC-FGC_GSND                                        -1.000000
FGC-FGC_GSND_Zone                                   -1.000000
FGC-FGC_GSD                                         -1.000000
FGC-FGC_GSD_Zone                                          NaN
```

FGC-FGC_PU	0.360302
FGC-FGC_PU_Zone	0.980196
FGC-FGC_SRL	0.371768
FGC-FGC_SRL_Zone	0.420084
FGC-FGC_SRR	0.242536
FGC-FGC_SRR_Zone	0.420084
FGC-FGC_TL	-0.945905
FGC-FGC_TL_Zone	-0.485071
BIA-BIA_Activity_Level_num	NaN
BIA-BIA_BMC	NaN
BIA-BIA_BMI	NaN
BIA-BIA_BMR	NaN
BIA-BIA_DEE	NaN
BIA-BIA_ECW	NaN
BIA-BIA_FFM	NaN
BIA-BIA_FFMI	NaN
BIA-BIA_FMI	NaN
BIA-BIA_Fat	NaN
BIA-BIA_Frame_num	NaN
BIA-BIA_ICW	NaN
BIA-BIA_LDM	NaN
BIA-BIA_LST	NaN
BIA-BIA_SMM	NaN
BIA-BIA_TBW	NaN
PAQ_A-PAQ_A_Total	NaN
PAQ_C-PAQ_C_Total	-0.611549
SDS-SDS_Total_Raw	0.289241
SDS-SDS_Total_T	0.318874
PreInt_EduHx-computerinternet_hoursday	0.980196

	Physical-HeartRate \
Basic_Demos-Age	-0.024712
Basic_Demos-Sex	-0.653197
CGAS-CGAS_Score	0.365655
Physical-BMI	-0.576780
Physical-Height	0.280378
Physical-Weight	-0.118328
Physical-Waist_Circumference	NaN
Physical-Diastolic_BP	0.411597
Physical-HeartRate	1.000000
Physical-Systolic_BP	0.948835
Fitness_Endurance-Max_Stage	1.000000
Fitness_Endurance-Time_Mins	1.000000
Fitness_Endurance-Time_Sec	1.000000
FGC-FGC_CU	0.762063
FGC-FGC_CU_Zone	0.989949
FGC-FGC_GSND	-1.000000
FGC-FGC_GSND_Zone	-1.000000

```
FGC-FGC_GSD                                  -1.000000
FGC-FGC_GSD_Zone                                   NaN
FGC-FGC_PU                                    0.446442
FGC-FGC_PU_Zone                               0.489898
FGC-FGC_SRL                                  -0.688583
FGC-FGC_SRL_Zone                             -0.653197
FGC-FGC_SRR                                  -0.754247
FGC-FGC_SRR_Zone                             -0.653197
FGC-FGC_TL                                   -0.648886
FGC-FGC_TL_Zone                               0.141421
BIA-BIA_Activity_Level_num                    1.000000
BIA-BIA_BMC                                  -1.000000
BIA-BIA_BMI                                  -1.000000
BIA-BIA_BMR                                  -1.000000
BIA-BIA_DEE                                  -1.000000
BIA-BIA_ECW                                  -1.000000
BIA-BIA_FFM                                  -1.000000
BIA-BIA_FFMI                                 -1.000000
BIA-BIA_FMI                                  -1.000000
BIA-BIA_Fat                                  -1.000000
BIA-BIA_Frame_num                                  NaN
BIA-BIA_ICW                                  -1.000000
BIA-BIA_LDM                                  -1.000000
BIA-BIA_LST                                  -1.000000
BIA-BIA_SMM                                  -1.000000
BIA-BIA_TBW                                  -1.000000
PAQ_A-PAQ_A_Total                                  NaN
PAQ_C-PAQ_C_Total                            -0.969066
SDS-SDS_Total_Raw                            -0.695701
SDS-SDS_Total_T                              -0.693026
PreInt_EduHx-computerinternet_hoursday        0.489898

                                        Physical-Systolic_BP  \
Basic_Demos-Age                                    -0.310542
Basic_Demos-Sex                                    -0.530314
CGAS-CGAS_Score                                     0.630416
Physical-BMI                                       -0.803882
Physical-Height                                    -0.016846
Physical-Weight                                    -0.419514
Physical-Waist_Circumference                             NaN
Physical-Diastolic_BP                               0.513355
Physical-HeartRate                                  0.948835
Physical-Systolic_BP                                1.000000
Fitness_Endurance-Max_Stage                              NaN
Fitness_Endurance-Time_Mins                              NaN
Fitness_Endurance-Time_Sec                               NaN
FGC-FGC_CU                                          0.550810
FGC-FGC_CU_Zone                                     0.918532
```

FGC-FGC_GSND	-1.000000
FGC-FGC_GSND_Zone	-1.000000
FGC-FGC_GSD	-1.000000
FGC-FGC_GSD_Zone	NaN
FGC-FGC_PU	0.200231
FGC-FGC_PU_Zone	0.530314
FGC-FGC_SRL	-0.529335
FGC-FGC_SRL_Zone	-0.530314
FGC-FGC_SRR	-0.705539
FGC-FGC_SRR_Zone	-0.530314
FGC-FGC_TL	-0.760444
FGC-FGC_TL_Zone	0.279553
BIA-BIA_Activity_Level_num	1.000000
BIA-BIA_BMC	-1.000000
BIA-BIA_BMI	-1.000000
BIA-BIA_BMR	-1.000000
BIA-BIA_DEE	-1.000000
BIA-BIA_ECW	-1.000000
BIA-BIA_FFM	-1.000000
BIA-BIA_FFMI	-1.000000
BIA-BIA_FMI	-1.000000
BIA-BIA_Fat	-1.000000
BIA-BIA_Frame_num	NaN
BIA-BIA_ICW	-1.000000
BIA-BIA_LDM	-1.000000
BIA-BIA_LST	-1.000000
BIA-BIA_SMM	-1.000000
BIA-BIA_TBW	-1.000000
PAQ_A-PAQ_A_Total	NaN
PAQ_C-PAQ_C_Total	-0.990979
SDS-SDS_Total_Raw	-0.672726
SDS-SDS_Total_T	-0.649167
PreInt_EduHx-computerinternet_hoursday	0.530314

	Fitness_Endurance-Max_Stage \
Basic_Demos-Age	-1.0
Basic_Demos-Sex	-1.0
CGAS-CGAS_Score	NaN
Physical-BMI	1.0
Physical-Height	-1.0
Physical-Weight	1.0
Physical-Waist_Circumference	NaN
Physical-Diastolic_BP	-1.0
Physical-HeartRate	1.0
Physical-Systolic_BP	NaN
Fitness_Endurance-Max_Stage	1.0
Fitness_Endurance-Time_Mins	1.0
Fitness_Endurance-Time_Sec	1.0

```
FGC-FGC_CU                                                     -1.0
FGC-FGC_CU_Zone                                                 NaN
FGC-FGC_GSND                                                    NaN
FGC-FGC_GSND_Zone                                               NaN
FGC-FGC_GSD                                                     NaN
FGC-FGC_GSD_Zone                                                NaN
FGC-FGC_PU                                                     -1.0
FGC-FGC_PU_Zone                                                -1.0
FGC-FGC_SRL                                                    -1.0
FGC-FGC_SRL_Zone                                               -1.0
FGC-FGC_SRR                                                    -1.0
FGC-FGC_SRR_Zone                                               -1.0
FGC-FGC_TL                                                      1.0
FGC-FGC_TL_Zone                                                 1.0
BIA-BIA_Activity_Level_num                                      NaN
BIA-BIA_BMC                                                     NaN
BIA-BIA_BMI                                                     NaN
BIA-BIA_BMR                                                     NaN
BIA-BIA_DEE                                                     NaN
BIA-BIA_ECW                                                     NaN
BIA-BIA_FFM                                                     NaN
BIA-BIA_FFMI                                                    NaN
BIA-BIA_FMI                                                     NaN
BIA-BIA_Fat                                                     NaN
BIA-BIA_Frame_num                                               NaN
BIA-BIA_ICW                                                     NaN
BIA-BIA_LDM                                                     NaN
BIA-BIA_LST                                                     NaN
BIA-BIA_SMM                                                     NaN
BIA-BIA_TBW                                                     NaN
PAQ_A-PAQ_A_Total                                               NaN
PAQ_C-PAQ_C_Total                                               1.0
SDS-SDS_Total_Raw                                              -1.0
SDS-SDS_Total_T                                                -1.0
PreInt_EduHx-computerinternet_hoursday                         -1.0

                                        Fitness_Endurance-Time_Mins  \
Basic_Demos-Age                                                -1.0
Basic_Demos-Sex                                                -1.0
CGAS-CGAS_Score                                                 NaN
Physical-BMI                                                    1.0
Physical-Height                                                -1.0
Physical-Weight                                                 1.0
Physical-Waist_Circumference                                    NaN
Physical-Diastolic_BP                                          -1.0
Physical-HeartRate                                              1.0
Physical-Systolic_BP                                            NaN
Fitness_Endurance-Max_Stage                                     1.0
```

```
Fitness_Endurance-Time_Mins                          1.0
Fitness_Endurance-Time_Sec                           1.0
FGC-FGC_CU                                          -1.0
FGC-FGC_CU_Zone                                      NaN
FGC-FGC_GSND                                         NaN
FGC-FGC_GSND_Zone                                    NaN
FGC-FGC_GSD                                          NaN
FGC-FGC_GSD_Zone                                     NaN
FGC-FGC_PU                                          -1.0
FGC-FGC_PU_Zone                                     -1.0
FGC-FGC_SRL                                         -1.0
FGC-FGC_SRL_Zone                                    -1.0
FGC-FGC_SRR                                         -1.0
FGC-FGC_SRR_Zone                                    -1.0
FGC-FGC_TL                                           1.0
FGC-FGC_TL_Zone                                      1.0
BIA-BIA_Activity_Level_num                           NaN
BIA-BIA_BMC                                          NaN
BIA-BIA_BMI                                          NaN
BIA-BIA_BMR                                          NaN
BIA-BIA_DEE                                          NaN
BIA-BIA_ECW                                          NaN
BIA-BIA_FFM                                          NaN
BIA-BIA_FFMI                                         NaN
BIA-BIA_FMI                                          NaN
BIA-BIA_Fat                                          NaN
BIA-BIA_Frame_num                                    NaN
BIA-BIA_ICW                                          NaN
BIA-BIA_LDM                                          NaN
BIA-BIA_LST                                          NaN
BIA-BIA_SMM                                          NaN
BIA-BIA_TBW                                          NaN
PAQ_A-PAQ_A_Total                                    NaN
PAQ_C-PAQ_C_Total                                    1.0
SDS-SDS_Total_Raw                                   -1.0
SDS-SDS_Total_T                                     -1.0
PreInt_EduHx-computerinternet_hoursday              -1.0

                                        Fitness_Endurance-Time_Sec  \
Basic_Demos-Age                                              -1.0
Basic_Demos-Sex                                              -1.0
CGAS-CGAS_Score                                               NaN
Physical-BMI                                                  1.0
Physical-Height                                              -1.0
Physical-Weight                                               1.0
Physical-Waist_Circumference                                  NaN
Physical-Diastolic_BP                                        -1.0
Physical-HeartRate                                            1.0
```

```
Physical-Systolic_BP                                              NaN
Fitness_Endurance-Max_Stage                                       1.0
Fitness_Endurance-Time_Mins                                       1.0
Fitness_Endurance-Time_Sec                                        1.0
FGC-FGC_CU                                                       -1.0
FGC-FGC_CU_Zone                                                   NaN
FGC-FGC_GSND                                                      NaN
FGC-FGC_GSND_Zone                                                 NaN
FGC-FGC_GSD                                                       NaN
FGC-FGC_GSD_Zone                                                  NaN
FGC-FGC_PU                                                       -1.0
FGC-FGC_PU_Zone                                                  -1.0
FGC-FGC_SRL                                                      -1.0
FGC-FGC_SRL_Zone                                                 -1.0
FGC-FGC_SRR                                                      -1.0
FGC-FGC_SRR_Zone                                                 -1.0
FGC-FGC_TL                                                        1.0
FGC-FGC_TL_Zone                                                   1.0
BIA-BIA_Activity_Level_num                                        NaN
BIA-BIA_BMC                                                       NaN
BIA-BIA_BMI                                                       NaN
BIA-BIA_BMR                                                       NaN
BIA-BIA_DEE                                                       NaN
BIA-BIA_ECW                                                       NaN
BIA-BIA_FFM                                                       NaN
BIA-BIA_FFMI                                                      NaN
BIA-BIA_FMI                                                       NaN
BIA-BIA_Fat                                                       NaN
BIA-BIA_Frame_num                                                 NaN
BIA-BIA_ICW                                                       NaN
BIA-BIA_LDM                                                       NaN
BIA-BIA_LST                                                       NaN
BIA-BIA_SMM                                                       NaN
BIA-BIA_TBW                                                       NaN
PAQ_A-PAQ_A_Total                                                 NaN
PAQ_C-PAQ_C_Total                                                 1.0
SDS-SDS_Total_Raw                                                -1.0
SDS-SDS_Total_T                                                  -1.0
PreInt_EduHx-computerinternet_hoursday                           -1.0

                                        FGC-FGC_CU  FGC-FGC_CU_Zone  \
Basic_Demos-Age                           0.743586     2.923217e-01
Basic_Demos-Sex                           0.095173    -1.666667e-01
CGAS-CGAS_Score                           0.132362     4.363141e-01
Physical-BMI                              0.176773    -3.165009e-01
Physical-Height                           0.873480     5.076788e-01
Physical-Weight                           0.644895     1.753494e-01
Physical-Waist_Circumference                   NaN              NaN
```

Physical-Diastolic_BP	0.388967	4.850713e-01
Physical-HeartRate	0.762063	9.899495e-01
Physical-Systolic_BP	0.550810	9.185315e-01
Fitness_Endurance-Max_Stage	-1.000000	NaN
Fitness_Endurance-Time_Mins	-1.000000	NaN
Fitness_Endurance-Time_Sec	-1.000000	NaN
FGC-FGC_CU	1.000000	8.565604e-01
FGC-FGC_CU_Zone	0.856560	1.000000e+00
FGC-FGC_GSND	-1.000000	-1.000000e+00
FGC-FGC_GSND_Zone	-1.000000	-1.000000e+00
FGC-FGC_GSD	-1.000000	-1.000000e+00
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	0.945947	6.517505e-01
FGC-FGC_PU_Zone	0.582816	6.123724e-01
FGC-FGC_SRL	0.014924	-2.090833e-01
FGC-FGC_SRL_Zone	0.095173	-1.666667e-01
FGC-FGC_SRR	0.216406	-1.263228e-01
FGC-FGC_SRR_Zone	0.095173	-1.666667e-01
FGC-FGC_TL	-0.182879	-4.803845e-01
FGC-FGC_TL_Zone	-0.571040	-1.666667e-01
BIA-BIA_Activity_Level_num	0.755929	1.000000e+00
BIA-BIA_BMC	0.811861	2.314897e-01
BIA-BIA_BMI	0.283158	-4.138134e-01
BIA-BIA_BMR	0.653987	-8.809372e-04
BIA-BIA_DEE	0.892572	3.795341e-01
BIA-BIA_ECW	0.500778	-1.880999e-01
BIA-BIA_FFM	0.654003	-8.610810e-04
BIA-BIA_FFMI	0.267602	-4.284905e-01
BIA-BIA_FMI	0.287472	-4.097119e-01
BIA-BIA_Fat	0.336487	-3.621194e-01
BIA-BIA_Frame_num	0.944911	5.000000e-01
BIA-BIA_ICW	0.810524	2.292639e-01
BIA-BIA_LDM	0.778721	1.779477e-01
BIA-BIA_LST	0.646609	-1.059347e-02
BIA-BIA_SMM	0.599059	-7.133819e-02
BIA-BIA_TBW	0.598244	-7.235330e-02
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	-0.994158	-9.909795e-01
SDS-SDS_Total_Raw	-0.258199	-5.962848e-01
SDS-SDS_Total_T	-0.291361	-5.976143e-01
PreInt_EduHx-computerinternet_hoursday	-0.184302	-7.166459e-17

	FGC-FGC_GSND	FGC-FGC_GSND_Zone	\
Basic_Demos-Age	1.0	1.0	
Basic_Demos-Sex	NaN	NaN	
CGAS-CGAS_Score	-1.0	-1.0	
Physical-BMI	1.0	1.0	
Physical-Height	1.0	1.0	

Physical-Weight	1.0	1.0
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	-1.0	-1.0
Physical-HeartRate	-1.0	-1.0
Physical-Systolic_BP	-1.0	-1.0
Fitness_Endurance-Max_Stage	NaN	NaN
Fitness_Endurance-Time_Mins	NaN	NaN
Fitness_Endurance-Time_Sec	NaN	NaN
FGC-FGC_CU	-1.0	-1.0
FGC-FGC_CU_Zone	-1.0	-1.0
FGC-FGC_GSND	1.0	1.0
FGC-FGC_GSND_Zone	1.0	1.0
FGC-FGC_GSD	1.0	1.0
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	-1.0	-1.0
FGC-FGC_PU_Zone	-1.0	-1.0
FGC-FGC_SRL	NaN	NaN
FGC-FGC_SRL_Zone	NaN	NaN
FGC-FGC_SRR	1.0	1.0
FGC-FGC_SRR_Zone	NaN	NaN
FGC-FGC_TL	1.0	1.0
FGC-FGC_TL_Zone	NaN	NaN
BIA-BIA_Activity_Level_num	NaN	NaN
BIA-BIA_BMC	NaN	NaN
BIA-BIA_BMI	NaN	NaN
BIA-BIA_BMR	NaN	NaN
BIA-BIA_DEE	NaN	NaN
BIA-BIA_ECW	NaN	NaN
BIA-BIA_FFM	NaN	NaN
BIA-BIA_FFMI	NaN	NaN
BIA-BIA_FMI	NaN	NaN
BIA-BIA_Fat	NaN	NaN
BIA-BIA_Frame_num	NaN	NaN
BIA-BIA_ICW	NaN	NaN
BIA-BIA_LDM	NaN	NaN
BIA-BIA_LST	NaN	NaN
BIA-BIA_SMM	NaN	NaN
BIA-BIA_TBW	NaN	NaN
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	1.0	1.0
SDS-SDS_Total_Raw	1.0	1.0
SDS-SDS_Total_T	1.0	1.0
PreInt_EduHx-computerinternet_hoursday	-1.0	-1.0

	FGC-FGC_GSD	FGC-FGC_GSD_Zone \
Basic_Demos-Age	1.0	NaN
Basic_Demos-Sex	NaN	NaN
CGAS-CGAS_Score	-1.0	NaN

Physical-BMI	1.0	NaN
Physical-Height	1.0	NaN
Physical-Weight	1.0	NaN
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	-1.0	NaN
Physical-HeartRate	-1.0	NaN
Physical-Systolic_BP	-1.0	NaN
Fitness_Endurance-Max_Stage	NaN	NaN
Fitness_Endurance-Time_Mins	NaN	NaN
Fitness_Endurance-Time_Sec	NaN	NaN
FGC-FGC_CU	-1.0	NaN
FGC-FGC_CU_Zone	-1.0	NaN
FGC-FGC_GSND	1.0	NaN
FGC-FGC_GSND_Zone	1.0	NaN
FGC-FGC_GSD	1.0	NaN
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	-1.0	NaN
FGC-FGC_PU_Zone	-1.0	NaN
FGC-FGC_SRL	NaN	NaN
FGC-FGC_SRL_Zone	NaN	NaN
FGC-FGC_SRR	1.0	NaN
FGC-FGC_SRR_Zone	NaN	NaN
FGC-FGC_TL	1.0	NaN
FGC-FGC_TL_Zone	NaN	NaN
BIA-BIA_Activity_Level_num	NaN	NaN
BIA-BIA_BMC	NaN	NaN
BIA-BIA_BMI	NaN	NaN
BIA-BIA_BMR	NaN	NaN
BIA-BIA_DEE	NaN	NaN
BIA-BIA_ECW	NaN	NaN
BIA-BIA_FFM	NaN	NaN
BIA-BIA_FFMI	NaN	NaN
BIA-BIA_FMI	NaN	NaN
BIA-BIA_Fat	NaN	NaN
BIA-BIA_Frame_num	NaN	NaN
BIA-BIA_ICW	NaN	NaN
BIA-BIA_LDM	NaN	NaN
BIA-BIA_LST	NaN	NaN
BIA-BIA_SMM	NaN	NaN
BIA-BIA_TBW	NaN	NaN
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	1.0	NaN
SDS-SDS_Total_Raw	1.0	NaN
SDS-SDS_Total_T	1.0	NaN
PreInt_EduHx-computerinternet_hoursday	-1.0	NaN

	FGC-FGC_PU	FGC-FGC_PU_Zone \
Basic_Demos-Age	0.904974	0.260378

Basic_Demos-Sex	0.298719	0.408248
CGAS-CGAS_Score	-0.059984	0.267187
Physical-BMI	0.406402	-0.391428
Physical-Height	0.959337	0.331076
Physical-Weight	0.811553	0.043463
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	0.360302	0.980196
Physical-HeartRate	0.446442	0.489898
Physical-Systolic_BP	0.200231	0.530314
Fitness_Endurance-Max_Stage	-1.000000	-1.000000
Fitness_Endurance-Time_Mins	-1.000000	-1.000000
Fitness_Endurance-Time_Sec	-1.000000	-1.000000
FGC-FGC_CU	0.945947	0.582816
FGC-FGC_CU_Zone	0.651751	0.612372
FGC-FGC_GSND	-1.000000	-1.000000
FGC-FGC_GSND_Zone	-1.000000	-1.000000
FGC-FGC_GSD	-1.000000	-1.000000
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	1.000000	0.565412
FGC-FGC_PU_Zone	0.565412	1.000000
FGC-FGC_SRL	0.204405	0.352101
FGC-FGC_SRL_Zone	0.298719	0.408248
FGC-FGC_SRR	0.432237	0.309426
FGC-FGC_SRR_Zone	0.298719	0.408248
FGC-FGC_TL	-0.052182	-0.784465
FGC-FGC_TL_Zone	-0.787532	-0.612372
BIA-BIA_Activity_Level_num	0.359211	NaN
BIA-BIA_BMC	0.991060	NaN
BIA-BIA_BMI	0.700955	NaN
BIA-BIA_BMR	0.932940	NaN
BIA-BIA_DEE	0.999761	NaN
BIA-BIA_ECW	0.849030	NaN
BIA-BIA_FFM	0.932947	NaN
BIA-BIA_FFMI	0.689322	NaN
BIA-BIA_FMI	0.704158	NaN
BIA-BIA_Fat	0.739841	NaN
BIA-BIA_Frame_num	0.987829	NaN
BIA-BIA_ICW	0.990753	NaN
BIA-BIA_LDM	0.982282	NaN
BIA-BIA_LST	0.929399	NaN
BIA-BIA_SMM	0.905253	NaN
BIA-BIA_TBW	0.904820	NaN
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	-0.134014	-0.611549
SDS-SDS_Total_Raw	0.110727	0.258199
SDS-SDS_Total_T	0.066585	0.276026
PreInt_EduHx-computerinternet_hoursday	-0.210352	0.395285

	FGC-FGC_SRL	FGC-FGC_SRL_Zone \
Basic_Demos-Age	0.279205	0.372046
Basic_Demos-Sex	0.993146	1.000000
CGAS-CGAS_Score	0.313233	0.245427
Physical-BMI	0.225014	0.271616
Physical-Height	0.086848	0.184610
Physical-Weight	0.170110	0.256066
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	0.371768	0.420084
Physical-HeartRate	-0.688583	-0.653197
Physical-Systolic_BP	-0.529335	-0.530314
Fitness_Endurance-Max_Stage	-1.000000	-1.000000
Fitness_Endurance-Time_Mins	-1.000000	-1.000000
Fitness_Endurance-Time_Sec	-1.000000	-1.000000
FGC-FGC_CU	0.014924	0.095173
FGC-FGC_CU_Zone	-0.209083	-0.166667
FGC-FGC_GSND	NaN	NaN
FGC-FGC_GSND_Zone	NaN	NaN
FGC-FGC_GSD	NaN	NaN
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	0.204405	0.298719
FGC-FGC_PU_Zone	0.352101	0.408248
FGC-FGC_SRL	1.000000	0.993146
FGC-FGC_SRL_Zone	0.993146	1.000000
FGC-FGC_SRR	0.950832	0.968475
FGC-FGC_SRR_Zone	0.993146	1.000000
FGC-FGC_TL	0.100440	0.080064
FGC-FGC_TL_Zone	-0.574979	-0.666667
BIA-BIA_Activity_Level_num	-0.500000	-0.500000
BIA-BIA_BMC	0.726757	0.726757
BIA-BIA_BMI	0.995303	0.995303
BIA-BIA_BMR	0.866466	0.866466
BIA-BIA_DEE	0.611460	0.611460
BIA-BIA_ECW	0.944617	0.944617
BIA-BIA_FFM	0.866456	0.866456
BIA-BIA_FFMI	0.996739	0.996739
BIA-BIA_FMI	0.994857	0.994857
BIA-BIA_Fat	0.988309	0.988309
BIA-BIA_Frame_num	0.500000	0.500000
BIA-BIA_ICW	0.728326	0.728326
BIA-BIA_LDM	0.763230	0.763230
BIA-BIA_LST	0.871274	0.871274
BIA-BIA_SMM	0.899488	0.899488
BIA-BIA_TBW	0.899932	0.899932
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	0.379430	0.379430
SDS-SDS_Total_Raw	0.900567	0.946729
SDS-SDS_Total_T	0.926823	0.966092

PreInt_EduHx-computerinternet_hoursday	-0.354277	-0.322749

	FGC-FGC_SRR	FGC-FGC_SRR_Zone \
Basic_Demos-Age	0.550546	0.372046
Basic_Demos-Sex	0.968475	1.000000
CGAS-CGAS_Score	0.177406	0.245427
Physical-BMI	0.473849	0.271616
Physical-Height	0.364799	0.184610
Physical-Weight	0.463933	0.256066
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	0.242536	0.420084
Physical-HeartRate	-0.754247	-0.653197
Physical-Systolic_BP	-0.705539	-0.530314
Fitness_Endurance-Max_Stage	-1.000000	-1.000000
Fitness_Endurance-Time_Mins	-1.000000	-1.000000
Fitness_Endurance-Time_Sec	-1.000000	-1.000000
FGC-FGC_CU	0.216406	0.095173
FGC-FGC_CU_Zone	-0.126323	-0.166667
FGC-FGC_GSND	1.000000	NaN
FGC-FGC_GSND_Zone	1.000000	NaN
FGC-FGC_GSD	1.000000	NaN
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	0.432237	0.298719
FGC-FGC_PU_Zone	0.309426	0.408248
FGC-FGC_SRL	0.950832	0.993146
FGC-FGC_SRL_Zone	0.968475	1.000000
FGC-FGC_SRR	1.000000	0.968475
FGC-FGC_SRR_Zone	0.968475	1.000000
FGC-FGC_TL	0.262962	0.080064
FGC-FGC_TL_Zone	-0.715829	-0.666667
BIA-BIA_Activity_Level_num	-0.327327	-0.500000
BIA-BIA_BMC	0.843472	0.726757
BIA-BIA_BMI	0.995663	0.995303
BIA-BIA_BMR	0.945199	0.866466
BIA-BIA_DEE	0.749979	0.611460
BIA-BIA_ECW	0.989615	0.944617
BIA-BIA_FFM	0.945193	0.866456
BIA-BIA_FFMI	0.994027	0.996739
BIA-BIA_FMI	0.996072	0.994857
BIA-BIA_Fat	0.999313	0.988309
BIA-BIA_Frame_num	0.654654	0.500000
BIA-BIA_ICW	0.844699	0.728326
BIA-BIA_LDM	0.871583	0.763230
BIA-BIA_LST	0.948326	0.871274
BIA-BIA_SMM	0.965855	0.899488
BIA-BIA_TBW	0.966118	0.899932
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	0.590553	0.379430

SDS-SDS_Total_Raw	0.993808	0.946729
SDS-SDS_Total_T	0.996024	0.966092
PreInt_EduHx-computerinternet_hoursday	-0.489246	-0.322749

	FGC-FGC_TL	FGC-FGC_TL_Zone	\
Basic_Demos-Age	3.063858e-01	-8.238157e-01	
Basic_Demos-Sex	8.006408e-02	-6.666667e-01	
CGAS-CGAS_Score	-2.259734e-01	2.795137e-01	
Physical-BMI	8.166062e-01	-4.663872e-01	
Physical-Height	1.678662e-01	-7.384419e-01	
Physical-Weight	4.790038e-01	-7.076602e-01	
Physical-Waist_Circumference	NaN	NaN	
Physical-Diastolic_BP	-9.459053e-01	-4.850713e-01	
Physical-HeartRate	-6.488857e-01	1.414214e-01	
Physical-Systolic_BP	-7.604445e-01	2.795531e-01	
Fitness_Endurance-Max_Stage	1.000000e+00	1.000000e+00	
Fitness_Endurance-Time_Mins	1.000000e+00	1.000000e+00	
Fitness_Endurance-Time_Sec	1.000000e+00	1.000000e+00	
FGC-FGC_CU	-1.828792e-01	-5.710402e-01	
FGC-FGC_CU_Zone	-4.803845e-01	-1.666667e-01	
FGC-FGC_GSND	1.000000e+00	NaN	
FGC-FGC_GSND_Zone	1.000000e+00	NaN	
FGC-FGC_GSD	1.000000e+00	NaN	
FGC-FGC_GSD_Zone	NaN	NaN	
FGC-FGC_PU	-5.218181e-02	-7.875319e-01	
FGC-FGC_PU_Zone	-7.844645e-01	-6.123724e-01	
FGC-FGC_SRL	1.004404e-01	-5.749792e-01	
FGC-FGC_SRL_Zone	8.006408e-02	-6.666667e-01	
FGC-FGC_SRR	2.629619e-01	-7.158291e-01	
FGC-FGC_SRR_Zone	8.006408e-02	-6.666667e-01	
FGC-FGC_TL	1.000000e+00	8.006408e-02	
FGC-FGC_TL_Zone	8.006408e-02	1.000000e+00	
BIA-BIA_Activity_Level_num	-1.922963e-16	5.000000e-01	
BIA-BIA_BMC	9.728373e-01	-7.267570e-01	
BIA-BIA_BMI	9.103617e-01	-9.953031e-01	
BIA-BIA_BMR	9.999996e-01	-8.664655e-01	
BIA-BIA_DEE	9.251778e-01	-6.114604e-01	
BIA-BIA_ECW	9.821499e-01	-9.446167e-01	
BIA-BIA_FFM	9.999996e-01	-8.664556e-01	
BIA-BIA_FFMI	9.035463e-01	-9.967393e-01	
BIA-BIA_FMI	9.122150e-01	-9.948573e-01	
BIA-BIA_Fat	9.321317e-01	-9.883094e-01	
BIA-BIA_Frame_num	8.660254e-01	-5.000000e-01	
BIA-BIA_ICW	9.733643e-01	-7.283263e-01	
BIA-BIA_LDM	9.840399e-01	-7.632297e-01	
BIA-BIA_LST	9.999439e-01	-8.712735e-01	
BIA-BIA_SMM	9.974522e-01	-8.994880e-01	
BIA-BIA_TBW	9.973791e-01	-8.999323e-01	

```
PAQ_A-PAQ_A_Total                                    NaN           NaN
PAQ_C-PAQ_C_Total                           8.368427e-01 -3.794303e-01
SDS-SDS_Total_Raw                           3.419928e-02 -7.453560e-01
SDS-SDS_Total_T                             2.435423e-16 -7.171372e-01
PreInt_EduHx-computerinternet_hoursday -7.752171e-01 -3.583229e-17

                                        BIA-BIA_Activity_Level_num  \
Basic_Demos-Age                                      -1.922963e-16
Basic_Demos-Sex                                      -5.000000e-01
CGAS-CGAS_Score                                       9.991089e-01
Physical-BMI                                         -2.650820e-01
Physical-Height                                       2.678278e-01
Physical-Weight                                       1.880616e-03
Physical-Waist_Circumference                                   NaN
Physical-Diastolic_BP                                          NaN
Physical-HeartRate                                    1.000000e+00
Physical-Systolic_BP                                  1.000000e+00
Fitness_Endurance-Max_Stage                                    NaN
Fitness_Endurance-Time_Mins                                    NaN
Fitness_Endurance-Time_Sec                                     NaN
FGC-FGC_CU                                            7.559289e-01
FGC-FGC_CU_Zone                                       1.000000e+00
FGC-FGC_GSND                                                   NaN
FGC-FGC_GSND_Zone                                              NaN
FGC-FGC_GSD                                                    NaN
FGC-FGC_GSD_Zone                                               NaN
FGC-FGC_PU                                            3.592106e-01
FGC-FGC_PU_Zone                                                NaN
FGC-FGC_SRL                                          -5.000000e-01
FGC-FGC_SRL_Zone                                     -5.000000e-01
FGC-FGC_SRR                                          -3.273268e-01
FGC-FGC_SRR_Zone                                     -5.000000e-01
FGC-FGC_TL                                           -1.922963e-16
FGC-FGC_TL_Zone                                       5.000000e-01
BIA-BIA_Activity_Level_num                            1.000000e+00
BIA-BIA_BMC                                           2.314897e-01
BIA-BIA_BMI                                          -4.138134e-01
BIA-BIA_BMR                                          -8.809372e-04
BIA-BIA_DEE                                           3.795341e-01
BIA-BIA_ECW                                          -1.880999e-01
BIA-BIA_FFM                                          -8.610810e-04
BIA-BIA_FFMI                                         -4.284905e-01
BIA-BIA_FMI                                          -4.097119e-01
BIA-BIA_Fat                                          -3.621194e-01
BIA-BIA_Frame_num                                     5.000000e-01
BIA-BIA_ICW                                           2.292639e-01
BIA-BIA_LDM                                           1.779477e-01
BIA-BIA_LST                                          -1.059347e-02
```

BIA-BIA_SMM	-7.133819e-02
BIA-BIA_TBW	-7.235330e-02
PAQ_A-PAQ_A_Total	NaN
PAQ_C-PAQ_C_Total	-1.000000e+00
SDS-SDS_Total_Raw	-1.000000e+00
SDS-SDS_Total_T	-1.000000e+00
PreInt_EduHx-computerinternet_hoursday	-5.000000e-01

	BIA-BIA_BMC	BIA-BIA_BMI	BIA-BIA_BMR	\
Basic_Demos-Age	0.972837	0.910362	1.000000	
Basic_Demos-Sex	0.726757	0.995303	0.866466	
CGAS-CGAS_Score	0.190222	-0.451869	-0.043088	
Physical-BMI	0.876671	0.987489	0.964459	
Physical-Height	0.999296	0.766273	0.963230	
Physical-Weight	0.973271	0.909582	0.999996	
Physical-Waist_Circumference	NaN	NaN	NaN	
Physical-Diastolic_BP	NaN	NaN	NaN	
Physical-HeartRate	-1.000000	-1.000000	-1.000000	
Physical-Systolic_BP	-1.000000	-1.000000	-1.000000	
Fitness_Endurance-Max_Stage	NaN	NaN	NaN	
Fitness_Endurance-Time_Mins	NaN	NaN	NaN	
Fitness_Endurance-Time_Sec	NaN	NaN	NaN	
FGC-FGC_CU	0.811861	0.283158	0.653987	
FGC-FGC_CU_Zone	0.231490	-0.413813	-0.000881	
FGC-FGC_GSND	NaN	NaN	NaN	
FGC-FGC_GSND_Zone	NaN	NaN	NaN	
FGC-FGC_GSD	NaN	NaN	NaN	
FGC-FGC_GSD_Zone	NaN	NaN	NaN	
FGC-FGC_PU	0.991060	0.700955	0.932940	
FGC-FGC_PU_Zone	NaN	NaN	NaN	
FGC-FGC_SRL	0.726757	0.995303	0.866466	
FGC-FGC_SRL_Zone	0.726757	0.995303	0.866466	
FGC-FGC_SRR	0.843472	0.995663	0.945199	
FGC-FGC_SRR_Zone	0.726757	0.995303	0.866466	
FGC-FGC_TL	0.972837	0.910362	1.000000	
FGC-FGC_TL_Zone	-0.726757	-0.995303	-0.866466	
BIA-BIA_Activity_Level_num	0.231490	-0.413813	-0.000881	
BIA-BIA_BMC	1.000000	0.789840	0.972633	
BIA-BIA_BMI	0.789840	1.000000	0.910726	
BIA-BIA_BMR	0.972633	0.910726	1.000000	
BIA-BIA_DEE	0.987906	0.685190	0.924843	
BIA-BIA_ECW	0.911929	0.971950	0.982315	
BIA-BIA_FFM	0.972638	0.910718	1.000000	
BIA-BIA_FFMI	0.779812	0.999869	0.903923	
BIA-BIA_FMI	0.792593	0.999990	0.912576	
BIA-BIA_Fat	0.822986	0.998427	0.932450	
BIA-BIA_Frame_num	0.958247	0.581490	0.865585	
BIA-BIA_ICW	0.999997	0.791241	0.973162	

```
BIA-BIA_LDM                                 0.998504    0.822195    0.983883
BIA-BIA_LST                                 0.970330    0.914694    0.999953
BIA-BIA_SMM                                 0.953845    0.937563    0.997515
BIA-BIA_TBW                                 0.953539    0.937917    0.997442
PAQ_A-PAQ_A_Total                                NaN         NaN         NaN
PAQ_C-PAQ_C_Total                           1.000000    1.000000    1.000000
SDS-SDS_Total_Raw                           1.000000    1.000000    1.000000
SDS-SDS_Total_T                             1.000000    1.000000    1.000000
PreInt_EduHx-computerinternet_hoursday     -0.958247   -0.581490   -0.865585

                                         BIA-BIA_DEE BIA-BIA_ECW BIA-BIA_FFM  \
Basic_Demos-Age                             0.925178    0.982150    1.000000
Basic_Demos-Sex                             0.611460    0.944617    0.866456
CGAS-CGAS_Score                             0.340146   -0.229386   -0.043068
Physical-BMI                                0.791473    0.996876    0.964454
Physical-Height                             0.993028    0.895890    0.963236
Physical-Weight                             0.925890    0.981794    0.999996
Physical-Waist_Circumference                     NaN         NaN         NaN
Physical-Diastolic_BP                            NaN         NaN         NaN
Physical-HeartRate                         -1.000000   -1.000000   -1.000000
Physical-Systolic_BP                       -1.000000   -1.000000   -1.000000
Fitness_Endurance-Max_Stage                      NaN         NaN         NaN
Fitness_Endurance-Time_Mins                      NaN         NaN         NaN
Fitness_Endurance-Time_Sec                       NaN         NaN         NaN
FGC-FGC_CU                                  0.892572    0.500778    0.654003
FGC-FGC_CU_Zone                             0.379534   -0.188100   -0.000861
FGC-FGC_GSND                                     NaN         NaN         NaN
FGC-FGC_GSND_Zone                                NaN         NaN         NaN
FGC-FGC_GSD                                      NaN         NaN         NaN
FGC-FGC_GSD_Zone                                 NaN         NaN         NaN
FGC-FGC_PU                                  0.999761    0.849030    0.932947
FGC-FGC_PU_Zone                                  NaN         NaN         NaN
FGC-FGC_SRL                                 0.611460    0.944617    0.866456
FGC-FGC_SRL_Zone                            0.611460    0.944617    0.866456
FGC-FGC_SRR                                 0.749979    0.989615    0.945193
FGC-FGC_SRR_Zone                            0.611460    0.944617    0.866456
FGC-FGC_TL                                  0.925178    0.982150    1.000000
FGC-FGC_TL_Zone                            -0.611460   -0.944617   -0.866456
BIA-BIA_Activity_Level_num                  0.379534   -0.188100   -0.000861
BIA-BIA_BMC                                 0.987906    0.911929    0.972638
BIA-BIA_BMI                                 0.685190    0.971950    0.910718
BIA-BIA_BMR                                 0.924843    0.982315    1.000000
BIA-BIA_DEE                                 1.000000    0.837273    0.924851
BIA-BIA_ECW                                 0.837273    1.000000    0.982312
BIA-BIA_FFM                                 0.924851    0.982312    1.000000
BIA-BIA_FFMI                                0.673314    0.968017    0.903915
BIA-BIA_FMI                                 0.688461    0.972999    0.912567
BIA-BIA_Fat                                 0.724951    0.983608    0.932443
```

BIA-BIA_Frame_num	0.990994	0.756517	0.865595
BIA-BIA_ICW	0.987548	0.912865	0.973167
BIA-BIA_LDM	0.977949	0.933003	0.983886
BIA-BIA_LST	0.921105	0.984087	0.999953
BIA-BIA_SMM	0.895745	0.993066	0.997513
BIA-BIA_TBW	0.895292	0.993185	0.997441
PAQ_A-PAQ_A_Total	NaN	NaN	NaN
PAQ_C-PAQ_C_Total	1.000000	1.000000	1.000000
SDS-SDS_Total_Raw	1.000000	1.000000	1.000000
SDS-SDS_Total_T	1.000000	1.000000	1.000000
PreInt_EduHx-computerinternet_hoursday	-0.990994	-0.756517	-0.865595

	BIA-BIA_FFMI	BIA-BIA_FMI	\
Basic_Demos-Age	0.903546	0.912215	
Basic_Demos-Sex	0.996739	0.994857	
CGAS-CGAS_Score	-0.466245	-0.447849	
Physical-BMI	0.984808	0.988189	
Physical-Height	0.755775	0.769157	
Physical-Weight	0.902739	0.911443	
Physical-Waist_Circumference	NaN	NaN	
Physical-Diastolic_BP	NaN	NaN	
Physical-HeartRate	-1.000000	-1.000000	
Physical-Systolic_BP	-1.000000	-1.000000	
Fitness_Endurance-Max_Stage	NaN	NaN	
Fitness_Endurance-Time_Mins	NaN	NaN	
Fitness_Endurance-Time_Sec	NaN	NaN	
FGC-FGC_CU	0.267602	0.287472	
FGC-FGC_CU_Zone	-0.428490	-0.409712	
FGC-FGC_GSND	NaN	NaN	
FGC-FGC_GSND_Zone	NaN	NaN	
FGC-FGC_GSD	NaN	NaN	
FGC-FGC_GSD_Zone	NaN	NaN	
FGC-FGC_PU	0.689322	0.704158	
FGC-FGC_PU_Zone	NaN	NaN	
FGC-FGC_SRL	0.996739	0.994857	
FGC-FGC_SRL_Zone	0.996739	0.994857	
FGC-FGC_SRR	0.994027	0.996072	
FGC-FGC_SRR_Zone	0.996739	0.994857	
FGC-FGC_TL	0.903546	0.912215	
FGC-FGC_TL_Zone	-0.996739	-0.994857	
BIA-BIA_Activity_Level_num	-0.428490	-0.409712	
BIA-BIA_BMC	0.779812	0.792593	
BIA-BIA_BMI	0.999869	0.999990	
BIA-BIA_BMR	0.903923	0.912576	
BIA-BIA_DEE	0.673314	0.688461	
BIA-BIA_ECW	0.968017	0.972999	
BIA-BIA_FFM	0.903915	0.912567	
BIA-BIA_FFMI	1.000000	0.999786	

BIA-BIA_FMI	0.999786	1.000000
BIA-BIA_Fat	0.997389	0.998669
BIA-BIA_Frame_num	0.568249	0.585145
BIA-BIA_ICW	0.781242	0.793985
BIA-BIA_LDM	0.812877	0.824749
BIA-BIA_LST	0.908035	0.916504
BIA-BIA_SMM	0.931812	0.939119
BIA-BIA_TBW	0.932181	0.939468
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	1.000000	1.000000
SDS-SDS_Total_Raw	1.000000	1.000000
SDS-SDS_Total_T	1.000000	1.000000
PreInt_EduHx-computerinternet_hoursday	-0.568249	-0.585145

	BIA-BIA_Fat	BIA-BIA_Frame_num \
Basic_Demos-Age	0.932132	0.866025
Basic_Demos-Sex	0.988309	0.500000
CGAS-CGAS_Score	-0.401140	0.463002
Physical-BMI	0.994777	0.702503
Physical-Height	0.801092	0.968301
Physical-Weight	0.931449	0.866964
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	NaN	NaN
Physical-HeartRate	-1.000000	NaN
Physical-Systolic_BP	-1.000000	NaN
Fitness_Endurance-Max_Stage	NaN	NaN
Fitness_Endurance-Time_Mins	NaN	NaN
Fitness_Endurance-Time_Sec	NaN	NaN
FGC-FGC_CU	0.336487	0.944911
FGC-FGC_CU_Zone	-0.362119	0.500000
FGC-FGC_GSND	NaN	NaN
FGC-FGC_GSND_Zone	NaN	NaN
FGC-FGC_GSD	NaN	NaN
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	0.739841	0.987829
FGC-FGC_PU_Zone	NaN	NaN
FGC-FGC_SRL	0.988309	0.500000
FGC-FGC_SRL_Zone	0.988309	0.500000
FGC-FGC_SRR	0.999313	0.654654
FGC-FGC_SRR_Zone	0.988309	0.500000
FGC-FGC_TL	0.932132	0.866025
FGC-FGC_TL_Zone	-0.988309	-0.500000
BIA-BIA_Activity_Level_num	-0.362119	0.500000
BIA-BIA_BMC	0.822986	0.958247
BIA-BIA_BMI	0.998427	0.581490
BIA-BIA_BMR	0.932450	0.865585
BIA-BIA_DEE	0.724951	0.990994
BIA-BIA_ECW	0.983608	0.756517

BIA-BIA_FFM	0.932443	0.865595
BIA-BIA_FFMI	0.997389	0.568249
BIA-BIA_FMI	0.998669	0.585145
BIA-BIA_Fat	1.000000	0.626190
BIA-BIA_Frame_num	0.626190	1.000000
BIA-BIA_ICW	0.824283	0.957590
BIA-BIA_LDM	0.852817	0.941177
BIA-BIA_LST	0.935916	0.860680
BIA-BIA_SMM	0.955590	0.828150
BIA-BIA_TBW	0.955889	0.827579
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	1.000000	NaN
SDS-SDS_Total_Raw	1.000000	NaN
SDS-SDS_Total_T	1.000000	NaN
PreInt_EduHx-computerinternet_hoursday	-0.626190	-1.000000

	BIA-BIA_ICW	BIA-BIA_LDM	BIA-BIA_LST \
Basic_Demos-Age	0.973364	0.984040	0.999944
Basic_Demos-Sex	0.728326	0.763230	0.871274
CGAS-CGAS_Score	0.187976	0.136255	-0.052789
Physical-BMI	0.877769	0.901666	0.966980
Physical-Height	0.999207	0.995749	0.960575
Physical-Weight	0.973794	0.984373	0.999922
Physical-Waist_Circumference	NaN	NaN	NaN
Physical-Diastolic_BP	NaN	NaN	NaN
Physical-HeartRate	-1.000000	-1.000000	-1.000000
Physical-Systolic_BP	-1.000000	-1.000000	-1.000000
Fitness_Endurance-Max_Stage	NaN	NaN	NaN
Fitness_Endurance-Time_Mins	NaN	NaN	NaN
Fitness_Endurance-Time_Sec	NaN	NaN	NaN
FGC-FGC_CU	0.810524	0.778721	0.646609
FGC-FGC_CU_Zone	0.229264	0.177948	-0.010593
FGC-FGC_GSND	NaN	NaN	NaN
FGC-FGC_GSND_Zone	NaN	NaN	NaN
FGC-FGC_GSD	NaN	NaN	NaN
FGC-FGC_GSD_Zone	NaN	NaN	NaN
FGC-FGC_PU	0.990753	0.982282	0.929399
FGC-FGC_PU_Zone	NaN	NaN	NaN
FGC-FGC_SRL	0.728326	0.763230	0.871274
FGC-FGC_SRL_Zone	0.728326	0.763230	0.871274
FGC-FGC_SRR	0.844699	0.871583	0.948326
FGC-FGC_SRR_Zone	0.728326	0.763230	0.871274
FGC-FGC_TL	0.973364	0.984040	0.999944
FGC-FGC_TL_Zone	-0.728326	-0.763230	-0.871274
BIA-BIA_Activity_Level_num	0.229264	0.177948	-0.010593
BIA-BIA_BMC	0.999997	0.998504	0.970330
BIA-BIA_BMI	0.791241	0.822195	0.914694
BIA-BIA_BMR	0.973162	0.983883	0.999953

BIA-BIA_DEE	0.987548	0.977949	0.921105
BIA-BIA_ECW	0.912865	0.933003	0.984087
BIA-BIA_FFM	0.973167	0.983886	0.999953
BIA-BIA_FFMI	0.781242	0.812877	0.908035
BIA-BIA_FMI	0.793985	0.824749	0.916504
BIA-BIA_Fat	0.824283	0.852817	0.935916
BIA-BIA_Frame_num	0.957590	0.941177	0.860680
BIA-BIA_ICW	1.000000	0.998626	0.970881
BIA-BIA_LDM	0.998626	1.000000	0.982100
BIA-BIA_LST	0.970881	0.982100	1.000000
BIA-BIA_SMM	0.954529	0.968838	0.998152
BIA-BIA_TBW	0.954225	0.968586	0.998090
PAQ_A-PAQ_A_Total	NaN	NaN	NaN
PAQ_C-PAQ_C_Total	1.000000	1.000000	1.000000
SDS-SDS_Total_Raw	1.000000	1.000000	1.000000
SDS-SDS_Total_T	1.000000	1.000000	1.000000
PreInt_EduHx-computerinternet_hoursday	-0.957590	-0.941177	-0.860680

	BIA-BIA_SMM	BIA-BIA_TBW	\
Basic_Demos-Age	0.997452	0.997379	
Basic_Demos-Sex	0.899488	0.899932	
CGAS-CGAS_Score	-0.113375	-0.114386	
Physical-BMI	0.980680	0.980878	
Physical-Height	0.941906	0.941563	
Physical-Weight	0.997316	0.997241	
Physical-Waist_Circumference	NaN	NaN	
Physical-Diastolic_BP	NaN	NaN	
Physical-HeartRate	-1.000000	-1.000000	
Physical-Systolic_BP	-1.000000	-1.000000	
Fitness_Endurance-Max_Stage	NaN	NaN	
Fitness_Endurance-Time_Mins	NaN	NaN	
Fitness_Endurance-Time_Sec	NaN	NaN	
FGC-FGC_CU	0.599059	0.598244	
FGC-FGC_CU_Zone	-0.071338	-0.072353	
FGC-FGC_GSND	NaN	NaN	
FGC-FGC_GSND_Zone	NaN	NaN	
FGC-FGC_GSD	NaN	NaN	
FGC-FGC_GSD_Zone	NaN	NaN	
FGC-FGC_PU	0.905253	0.904820	
FGC-FGC_PU_Zone	NaN	NaN	
FGC-FGC_SRL	0.899488	0.899932	
FGC-FGC_SRL_Zone	0.899488	0.899932	
FGC-FGC_SRR	0.965855	0.966118	
FGC-FGC_SRR_Zone	0.899488	0.899932	
FGC-FGC_TL	0.997452	0.997379	
FGC-FGC_TL_Zone	-0.899488	-0.899932	
BIA-BIA_Activity_Level_num	-0.071338	-0.072353	
BIA-BIA_BMC	0.953845	0.953539	

BIA-BIA_BMI	0.937563	0.937917
BIA-BIA_BMR	0.997515	0.997442
BIA-BIA_DEE	0.895745	0.895292
BIA-BIA_ECW	0.993066	0.993185
BIA-BIA_FFM	0.997513	0.997441
BIA-BIA_FFMI	0.931812	0.932181
BIA-BIA_FMI	0.939119	0.939468
BIA-BIA_Fat	0.955590	0.955889
BIA-BIA_Frame_num	0.828150	0.827579
BIA-BIA_ICW	0.954529	0.954225
BIA-BIA_LDM	0.968838	0.968586
BIA-BIA_LST	0.998152	0.998090
BIA-BIA_SMM	1.000000	0.999999
BIA-BIA_TBW	0.999999	1.000000
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	1.000000	1.000000
SDS-SDS_Total_Raw	1.000000	1.000000
SDS-SDS_Total_T	1.000000	1.000000
PreInt_EduHx-computerinternet_hoursday	-0.828150	-0.827579

	PAQ_A-PAQ_A_Total	PAQ_C-PAQ_C_Total	\
Basic_Demos-Age	NaN	0.564902	
Basic_Demos-Sex	NaN	0.632346	
CGAS-CGAS_Score	NaN	-0.750481	
Physical-BMI	NaN	0.988267	
Physical-Height	NaN	0.964600	
Physical-Weight	NaN	0.999819	
Physical-Waist_Circumference	NaN	NaN	
Physical-Diastolic_BP	NaN	-0.611549	
Physical-HeartRate	NaN	-0.969066	
Physical-Systolic_BP	NaN	-0.990979	
Fitness_Endurance-Max_Stage	NaN	1.000000	
Fitness_Endurance-Time_Mins	NaN	1.000000	
Fitness_Endurance-Time_Sec	NaN	1.000000	
FGC-FGC_CU	NaN	-0.994158	
FGC-FGC_CU_Zone	NaN	-0.990979	
FGC-FGC_GSND	NaN	1.000000	
FGC-FGC_GSND_Zone	NaN	1.000000	
FGC-FGC_GSD	NaN	1.000000	
FGC-FGC_GSD_Zone	NaN	NaN	
FGC-FGC_PU	NaN	-0.134014	
FGC-FGC_PU_Zone	NaN	-0.611549	
FGC-FGC_SRL	NaN	0.379430	
FGC-FGC_SRL_Zone	NaN	0.379430	
FGC-FGC_SRR	NaN	0.590553	
FGC-FGC_SRR_Zone	NaN	0.379430	
FGC-FGC_TL	NaN	0.836843	
FGC-FGC_TL_Zone	NaN	-0.379430	

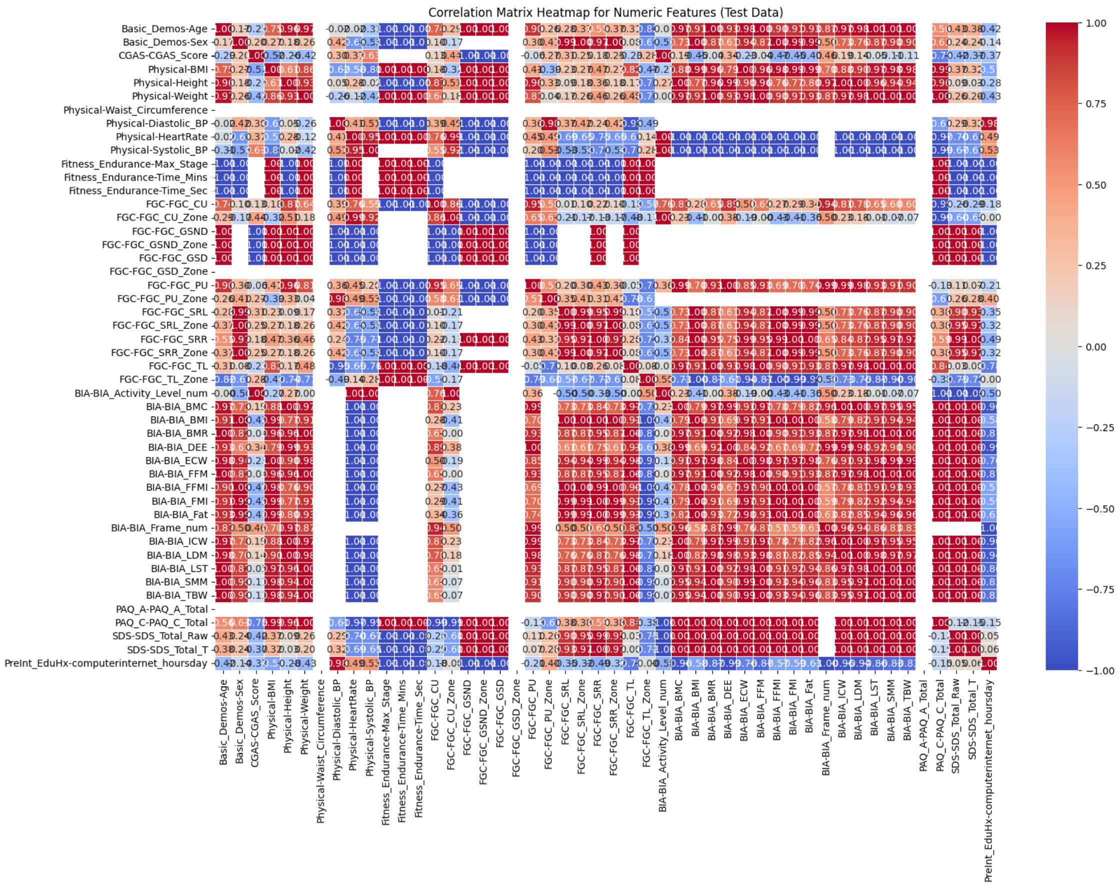
BIA-BIA_Activity_Level_num	NaN	-1.000000
BIA-BIA_BMC	NaN	1.000000
BIA-BIA_BMI	NaN	1.000000
BIA-BIA_BMR	NaN	1.000000
BIA-BIA_DEE	NaN	1.000000
BIA-BIA_ECW	NaN	1.000000
BIA-BIA_FFM	NaN	1.000000
BIA-BIA_FFMI	NaN	1.000000
BIA-BIA_FMI	NaN	1.000000
BIA-BIA_Fat	NaN	1.000000
BIA-BIA_Frame_num	NaN	NaN
BIA-BIA_ICW	NaN	1.000000
BIA-BIA_LDM	NaN	1.000000
BIA-BIA_LST	NaN	1.000000
BIA-BIA_SMM	NaN	1.000000
BIA-BIA_TBW	NaN	1.000000
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	NaN	1.000000
SDS-SDS_Total_Raw	NaN	-0.116212
SDS-SDS_Total_T	NaN	-0.151963
PreInt_EduHx-computerinternet_hoursday	NaN	-0.153981

	SDS-SDS_Total_Raw	SDS-SDS_Total_T \
Basic_Demos-Age	0.429548	3.838790e-01
Basic_Demos-Sex	0.241423	2.445998e-01
CGAS-CGAS_Score	-0.422847	-3.729518e-01
Physical-BMI	0.374065	3.181555e-01
Physical-Height	0.088035	3.025061e-02
Physical-Weight	0.263425	1.989699e-01
Physical-Waist_Circumference	NaN	NaN
Physical-Diastolic_BP	0.289241	3.188741e-01
Physical-HeartRate	-0.695701	-6.930265e-01
Physical-Systolic_BP	-0.672726	-6.491665e-01
Fitness_Endurance-Max_Stage	-1.000000	-1.000000e+00
Fitness_Endurance-Time_Mins	-1.000000	-1.000000e+00
Fitness_Endurance-Time_Sec	-1.000000	-1.000000e+00
FGC-FGC_CU	-0.258199	-2.913610e-01
FGC-FGC_CU_Zone	-0.596285	-5.976143e-01
FGC-FGC_GSND	1.000000	1.000000e+00
FGC-FGC_GSND_Zone	1.000000	1.000000e+00
FGC-FGC_GSD	1.000000	1.000000e+00
FGC-FGC_GSD_Zone	NaN	NaN
FGC-FGC_PU	0.110727	6.658451e-02
FGC-FGC_PU_Zone	0.258199	2.760262e-01
FGC-FGC_SRL	0.900567	9.268232e-01
FGC-FGC_SRL_Zone	0.946729	9.660918e-01
FGC-FGC_SRR	0.993808	9.960238e-01
FGC-FGC_SRR_Zone	0.946729	9.660918e-01

FGC-FGC_TL	0.034199	2.435423e-16
FGC-FGC_TL_Zone	-0.745356	-7.171372e-01
BIA-BIA_Activity_Level_num	-1.000000	-1.000000e+00
BIA-BIA_BMC	1.000000	1.000000e+00
BIA-BIA_BMI	1.000000	1.000000e+00
BIA-BIA_BMR	1.000000	1.000000e+00
BIA-BIA_DEE	1.000000	1.000000e+00
BIA-BIA_ECW	1.000000	1.000000e+00
BIA-BIA_FFM	1.000000	1.000000e+00
BIA-BIA_FFMI	1.000000	1.000000e+00
BIA-BIA_FMI	1.000000	1.000000e+00
BIA-BIA_Fat	1.000000	1.000000e+00
BIA-BIA_Frame_num	NaN	NaN
BIA-BIA_ICW	1.000000	1.000000e+00
BIA-BIA_LDM	1.000000	1.000000e+00
BIA-BIA_LST	1.000000	1.000000e+00
BIA-BIA_SMM	1.000000	1.000000e+00
BIA-BIA_TBW	1.000000	1.000000e+00
PAQ_A-PAQ_A_Total	NaN	NaN
PAQ_C-PAQ_C_Total	-0.116212	-1.519632e-01
SDS-SDS_Total_Raw	1.000000	9.985168e-01
SDS-SDS_Total_T	0.998517	1.000000e+00
PreInt_EduHx-computerinternet_hoursday	0.053760	6.419407e-02

	PreInt_EduHx-computerinternet_hoursday
Basic_Demos-Age	-4.243216e-01
Basic_Demos-Sex	-1.373606e-01
CGAS-CGAS_Score	-3.660247e-01
Physical-BMI	-5.667507e-01
Physical-Height	-2.808902e-01
Physical-Weight	-4.325382e-01
Physical-Waist_Circumference	NaN
Physical-Diastolic_BP	9.801961e-01
Physical-HeartRate	4.898979e-01
Physical-Systolic_BP	5.303144e-01
Fitness_Endurance-Max_Stage	-1.000000e+00
Fitness_Endurance-Time_Mins	-1.000000e+00
Fitness_Endurance-Time_Sec	-1.000000e+00
FGC-FGC_CU	-1.843024e-01
FGC-FGC_CU_Zone	-7.166459e-17
FGC-FGC_GSND	-1.000000e+00
FGC-FGC_GSND_Zone	-1.000000e+00
FGC-FGC_GSD	-1.000000e+00
FGC-FGC_GSD_Zone	NaN
FGC-FGC_PU	-2.103516e-01
FGC-FGC_PU_Zone	3.952847e-01
FGC-FGC_SRL	-3.542771e-01
FGC-FGC_SRL_Zone	-3.227486e-01

FGC-FGC_SRR	-4.892461e-01
FGC-FGC_SRR_Zone	-3.227486e-01
FGC-FGC_TL	-7.752171e-01
FGC-FGC_TL_Zone	-3.583229e-17
BIA-BIA_Activity_Level_num	-5.000000e-01
BIA-BIA_BMC	-9.582467e-01
BIA-BIA_BMI	-5.814897e-01
BIA-BIA_BMR	-8.655846e-01
BIA-BIA_DEE	-9.909945e-01
BIA-BIA_ECW	-7.565168e-01
BIA-BIA_FFM	-8.655945e-01
BIA-BIA_FFMI	-5.682488e-01
BIA-BIA_FMI	-5.851454e-01
BIA-BIA_Fat	-6.261901e-01
BIA-BIA_Frame_num	-1.000000e+00
BIA-BIA_ICW	-9.575901e-01
BIA-BIA_LDM	-9.411774e-01
BIA-BIA_LST	-8.606801e-01
BIA-BIA_SMM	-8.281498e-01
BIA-BIA_TBW	-8.275790e-01
PAQ_A-PAQ_A_Total	NaN
PAQ_C-PAQ_C_Total	-1.539806e-01
SDS-SDS_Total_Raw	5.376033e-02
SDS-SDS_Total_T	6.419407e-02
PreInt_EduHx-computerinternet_hoursday	1.000000e+00




```
Highly Correlated Pairs in Test Data (|correlation| > 0.8):
Feature 1: Basic_Demos-Age, Feature 2: Physical-Height, Correlation: 0.96
Feature 1: Basic_Demos-Age, Feature 2: Physical-Weight, Correlation: 0.97
Feature 1: Basic_Demos-Age, Feature 2: Fitness_Endurance-Max_Stage, Correlation:
-1.00
Feature 1: Basic_Demos-Age, Feature 2: Fitness_Endurance-Time_Mins, Correlation:
-1.00
Feature 1: Basic_Demos-Age, Feature 2: Fitness_Endurance-Time_Sec, Correlation:
-1.00
Feature 1: Basic_Demos-Age, Feature 2: FGC-FGC_GSND, Correlation: 1.00
Feature 1: Basic_Demos-Age, Feature 2: FGC-FGC_GSND_Zone, Correlation: 1.00
Feature 1: Basic_Demos-Age, Feature 2: FGC-FGC_GSD, Correlation: 1.00
Feature 1: Basic_Demos-Age, Feature 2: FGC-FGC_PU, Correlation: 0.90
Feature 1: Basic_Demos-Age, Feature 2: FGC-FGC_TL_Zone, Correlation: -0.82
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_BMC, Correlation: 0.97
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_BMI, Correlation: 0.91
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_BMR, Correlation: 1.00
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_DEE, Correlation: 0.93
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_ECW, Correlation: 0.98
```

```
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_FFM, Correlation: 1.00
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_FFMI, Correlation: 0.90
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_FMI, Correlation: 0.91
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_Fat, Correlation: 0.93
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_Frame_num, Correlation: 0.87
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_ICW, Correlation: 0.97
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_LDM, Correlation: 0.98
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_LST, Correlation: 1.00
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_SMM, Correlation: 1.00
Feature 1: Basic_Demos-Age, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: Basic_Demos-Sex, Feature 2: Fitness_Endurance-Max_Stage, Correlation:
-1.00
Feature 1: Basic_Demos-Sex, Feature 2: Fitness_Endurance-Time_Mins, Correlation:
-1.00
Feature 1: Basic_Demos-Sex, Feature 2: Fitness_Endurance-Time_Sec, Correlation:
-1.00
Feature 1: Basic_Demos-Sex, Feature 2: FGC-FGC_SRL, Correlation: 0.99
Feature 1: Basic_Demos-Sex, Feature 2: FGC-FGC_SRL_Zone, Correlation: 1.00
Feature 1: Basic_Demos-Sex, Feature 2: FGC-FGC_SRR, Correlation: 0.97
Feature 1: Basic_Demos-Sex, Feature 2: FGC-FGC_SRR_Zone, Correlation: 1.00
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_BMI, Correlation: 1.00
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_BMR, Correlation: 0.87
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_ECW, Correlation: 0.94
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_FFM, Correlation: 0.87
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_FFMI, Correlation: 1.00
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_FMI, Correlation: 0.99
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_Fat, Correlation: 0.99
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_LST, Correlation: 0.87
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_SMM, Correlation: 0.90
Feature 1: Basic_Demos-Sex, Feature 2: BIA-BIA_TBW, Correlation: 0.90
Feature 1: CGAS-CGAS_Score, Feature 2: FGC-FGC_GSND, Correlation: -1.00
Feature 1: CGAS-CGAS_Score, Feature 2: FGC-FGC_GSND_Zone, Correlation: -1.00
Feature 1: CGAS-CGAS_Score, Feature 2: FGC-FGC_GSD, Correlation: -1.00
Feature 1: CGAS-CGAS_Score, Feature 2: BIA-BIA_Activity_Level_num, Correlation:
1.00
Feature 1: Physical-BMI, Feature 2: Physical-Weight, Correlation: 0.86
Feature 1: Physical-BMI, Feature 2: Physical-Systolic_BP, Correlation: -0.80
Feature 1: Physical-BMI, Feature 2: Fitness_Endurance-Max_Stage, Correlation:
1.00
Feature 1: Physical-BMI, Feature 2: Fitness_Endurance-Time_Mins, Correlation:
1.00
Feature 1: Physical-BMI, Feature 2: Fitness_Endurance-Time_Sec, Correlation:
1.00
Feature 1: Physical-BMI, Feature 2: FGC-FGC_GSND, Correlation: 1.00
Feature 1: Physical-BMI, Feature 2: FGC-FGC_GSND_Zone, Correlation: 1.00
Feature 1: Physical-BMI, Feature 2: FGC-FGC_GSD, Correlation: 1.00
Feature 1: Physical-BMI, Feature 2: FGC-FGC_TL, Correlation: 0.82
Feature 1: Physical-BMI, Feature 2: BIA-BIA_BMC, Correlation: 0.88
```

Feature 1: Physical-BMI, Feature 2: BIA-BIA_BMI, Correlation: 0.99
Feature 1: Physical-BMI, Feature 2: BIA-BIA_BMR, Correlation: 0.96
Feature 1: Physical-BMI, Feature 2: BIA-BIA_ECW, Correlation: 1.00
Feature 1: Physical-BMI, Feature 2: BIA-BIA_FFM, Correlation: 0.96
Feature 1: Physical-BMI, Feature 2: BIA-BIA_FFMI, Correlation: 0.98
Feature 1: Physical-BMI, Feature 2: BIA-BIA_FMI, Correlation: 0.99
Feature 1: Physical-BMI, Feature 2: BIA-BIA_Fat, Correlation: 0.99
Feature 1: Physical-BMI, Feature 2: BIA-BIA_ICW, Correlation: 0.88
Feature 1: Physical-BMI, Feature 2: BIA-BIA_LDM, Correlation: 0.90
Feature 1: Physical-BMI, Feature 2: BIA-BIA_LST, Correlation: 0.97
Feature 1: Physical-BMI, Feature 2: BIA-BIA_SMM, Correlation: 0.98
Feature 1: Physical-BMI, Feature 2: BIA-BIA_TBW, Correlation: 0.98
Feature 1: Physical-BMI, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 0.99
Feature 1: Physical-Height, Feature 2: Physical-Weight, Correlation: 0.93
Feature 1: Physical-Height, Feature 2: Fitness_Endurance-Max_Stage, Correlation: -1.00
Feature 1: Physical-Height, Feature 2: Fitness_Endurance-Time_Mins, Correlation: -1.00
Feature 1: Physical-Height, Feature 2: Fitness_Endurance-Time_Sec, Correlation: -1.00
Feature 1: Physical-Height, Feature 2: FGC-FGC_CU, Correlation: 0.87
Feature 1: Physical-Height, Feature 2: FGC-FGC_GSND, Correlation: 1.00
Feature 1: Physical-Height, Feature 2: FGC-FGC_GSND_Zone, Correlation: 1.00
Feature 1: Physical-Height, Feature 2: FGC-FGC_GSD, Correlation: 1.00
Feature 1: Physical-Height, Feature 2: FGC-FGC_PU, Correlation: 0.96
Feature 1: Physical-Height, Feature 2: BIA-BIA_BMC, Correlation: 1.00
Feature 1: Physical-Height, Feature 2: BIA-BIA_BMR, Correlation: 0.96
Feature 1: Physical-Height, Feature 2: BIA-BIA_DEE, Correlation: 0.99
Feature 1: Physical-Height, Feature 2: BIA-BIA_ECW, Correlation: 0.90
Feature 1: Physical-Height, Feature 2: BIA-BIA_FFM, Correlation: 0.96
Feature 1: Physical-Height, Feature 2: BIA-BIA_Fat, Correlation: 0.80
Feature 1: Physical-Height, Feature 2: BIA-BIA_Frame_num, Correlation: 0.97
Feature 1: Physical-Height, Feature 2: BIA-BIA_ICW, Correlation: 1.00
Feature 1: Physical-Height, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: Physical-Height, Feature 2: BIA-BIA_LST, Correlation: 0.96
Feature 1: Physical-Height, Feature 2: BIA-BIA_SMM, Correlation: 0.94
Feature 1: Physical-Height, Feature 2: BIA-BIA_TBW, Correlation: 0.94
Feature 1: Physical-Height, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 0.96
Feature 1: Physical-Weight, Feature 2: Fitness_Endurance-Max_Stage, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: Fitness_Endurance-Time_Mins, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: Fitness_Endurance-Time_Sec, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: FGC-FGC_GSND, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: FGC-FGC_GSND_Zone, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: FGC-FGC_GSD, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: FGC-FGC_PU, Correlation: 0.81

Feature 1: Physical-Weight, Feature 2: BIA-BIA_BMC, Correlation: 0.97
Feature 1: Physical-Weight, Feature 2: BIA-BIA_BMI, Correlation: 0.91
Feature 1: Physical-Weight, Feature 2: BIA-BIA_BMR, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: BIA-BIA_DEE, Correlation: 0.93
Feature 1: Physical-Weight, Feature 2: BIA-BIA_ECW, Correlation: 0.98
Feature 1: Physical-Weight, Feature 2: BIA-BIA_FFM, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: BIA-BIA_FFMI, Correlation: 0.90
Feature 1: Physical-Weight, Feature 2: BIA-BIA_FMI, Correlation: 0.91
Feature 1: Physical-Weight, Feature 2: BIA-BIA_Fat, Correlation: 0.93
Feature 1: Physical-Weight, Feature 2: BIA-BIA_Frame_num, Correlation: 0.87
Feature 1: Physical-Weight, Feature 2: BIA-BIA_ICW, Correlation: 0.97
Feature 1: Physical-Weight, Feature 2: BIA-BIA_LDM, Correlation: 0.98
Feature 1: Physical-Weight, Feature 2: BIA-BIA_LST, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: BIA-BIA_SMM, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: Physical-Weight, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: Physical-Diastolic_BP, Feature 2: Fitness_Endurance-Max_Stage, Correlation: -1.00
Feature 1: Physical-Diastolic_BP, Feature 2: Fitness_Endurance-Time_Mins, Correlation: -1.00
Feature 1: Physical-Diastolic_BP, Feature 2: Fitness_Endurance-Time_Sec, Correlation: -1.00
Feature 1: Physical-Diastolic_BP, Feature 2: FGC-FGC_GSND, Correlation: -1.00
Feature 1: Physical-Diastolic_BP, Feature 2: FGC-FGC_GSND_Zone, Correlation: -1.00
Feature 1: Physical-Diastolic_BP, Feature 2: FGC-FGC_GSD, Correlation: -1.00
Feature 1: Physical-Diastolic_BP, Feature 2: FGC-FGC_PU_Zone, Correlation: 0.98
Feature 1: Physical-Diastolic_BP, Feature 2: FGC-FGC_TL, Correlation: -0.95
Feature 1: Physical-Diastolic_BP, Feature 2: PreInt_EduHx-computerinternet_hoursday, Correlation: 0.98
Feature 1: Physical-HeartRate, Feature 2: Physical-Systolic_BP, Correlation: 0.95
Feature 1: Physical-HeartRate, Feature 2: Fitness_Endurance-Max_Stage, Correlation: 1.00
Feature 1: Physical-HeartRate, Feature 2: Fitness_Endurance-Time_Mins, Correlation: 1.00
Feature 1: Physical-HeartRate, Feature 2: Fitness_Endurance-Time_Sec, Correlation: 1.00
Feature 1: Physical-HeartRate, Feature 2: FGC-FGC_CU_Zone, Correlation: 0.99
Feature 1: Physical-HeartRate, Feature 2: FGC-FGC_GSND, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: FGC-FGC_GSND_Zone, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: FGC-FGC_GSD, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_Activity_Level_num, Correlation: 1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_BMC, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_BMI, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_BMR, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_DEE, Correlation: -1.00

```
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_ECW, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_FFM, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_FFMI, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_FMI, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_Fat, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_ICW, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_LDM, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_LST, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_SMM, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: BIA-BIA_TBW, Correlation: -1.00
Feature 1: Physical-HeartRate, Feature 2: PAQ_C-PAQ_C_Total, Correlation: -0.97
Feature 1: Physical-Systolic_BP, Feature 2: FGC-FGC_CU_Zone, Correlation: 0.92
Feature 1: Physical-Systolic_BP, Feature 2: FGC-FGC_GSND, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: FGC-FGC_GSND_Zone, Correlation:
-1.00
Feature 1: Physical-Systolic_BP, Feature 2: FGC-FGC_GSD, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_Activity_Level_num,
Correlation: 1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_BMC, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_BMI, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_BMR, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_DEE, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_ECW, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_FFM, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_FFMI, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_FMI, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_Fat, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_ICW, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_LDM, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_LST, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_SMM, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: BIA-BIA_TBW, Correlation: -1.00
Feature 1: Physical-Systolic_BP, Feature 2: PAQ_C-PAQ_C_Total, Correlation:
-0.99
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: Fitness_Endurance-Time_Mins,
Correlation: 1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: Fitness_Endurance-Time_Sec,
Correlation: 1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_CU, Correlation:
-1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_PU, Correlation:
-1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_PU_Zone, Correlation:
-1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_SRL, Correlation:
-1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_SRL_Zone,
Correlation: -1.00
```

Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_SRR, Correlation: -1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_SRR_Zone, Correlation: -1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_TL, Correlation: 1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: FGC-FGC_TL_Zone, Correlation: 1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: SDS-SDS_Total_Raw, Correlation: -1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: SDS-SDS_Total_T, Correlation: -1.00
Feature 1: Fitness_Endurance-Max_Stage, Feature 2: PreInt_EduHx-computerinternet_hoursday, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: Fitness_Endurance-Time_Sec, Correlation: 1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_CU, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_PU, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_PU_Zone, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_SRL, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_SRL_Zone, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_SRR, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_SRR_Zone, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_TL, Correlation: 1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: FGC-FGC_TL_Zone, Correlation: 1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: SDS-SDS_Total_Raw, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: SDS-SDS_Total_T, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Mins, Feature 2: PreInt_EduHx-computerinternet_hoursday, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_CU, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_PU, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_PU_Zone, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_SRL, Correlation: -1.00

Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_SRL_Zone, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_SRR, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_SRR_Zone, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_TL, Correlation: 1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: FGC-FGC_TL_Zone, Correlation: 1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: SDS-SDS_Total_Raw, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: SDS-SDS_Total_T, Correlation: -1.00
Feature 1: Fitness_Endurance-Time_Sec, Feature 2: PreInt_EduHx-computerinternet_hoursday, Correlation: -1.00
Feature 1: FGC-FGC_CU, Feature 2: FGC-FGC_CU_Zone, Correlation: 0.86
Feature 1: FGC-FGC_CU, Feature 2: FGC-FGC_GSND, Correlation: -1.00
Feature 1: FGC-FGC_CU, Feature 2: FGC-FGC_GSND_Zone, Correlation: -1.00
Feature 1: FGC-FGC_CU, Feature 2: FGC-FGC_GSD, Correlation: -1.00
Feature 1: FGC-FGC_CU, Feature 2: FGC-FGC_PU, Correlation: 0.95
Feature 1: FGC-FGC_CU, Feature 2: BIA-BIA_BMC, Correlation: 0.81
Feature 1: FGC-FGC_CU, Feature 2: BIA-BIA_DEE, Correlation: 0.89
Feature 1: FGC-FGC_CU, Feature 2: BIA-BIA_Frame_num, Correlation: 0.94
Feature 1: FGC-FGC_CU, Feature 2: BIA-BIA_ICW, Correlation: 0.81
Feature 1: FGC-FGC_CU, Feature 2: PAQ_C-PAQ_C_Total, Correlation: -0.99
Feature 1: FGC-FGC_CU_Zone, Feature 2: FGC-FGC_GSND, Correlation: -1.00
Feature 1: FGC-FGC_CU_Zone, Feature 2: FGC-FGC_GSND_Zone, Correlation: -1.00
Feature 1: FGC-FGC_CU_Zone, Feature 2: FGC-FGC_GSD, Correlation: -1.00
Feature 1: FGC-FGC_CU_Zone, Feature 2: BIA-BIA_Activity_Level_num, Correlation: 1.00
Feature 1: FGC-FGC_CU_Zone, Feature 2: PAQ_C-PAQ_C_Total, Correlation: -0.99
Feature 1: FGC-FGC_GSND, Feature 2: FGC-FGC_GSND_Zone, Correlation: 1.00
Feature 1: FGC-FGC_GSND, Feature 2: FGC-FGC_GSD, Correlation: 1.00
Feature 1: FGC-FGC_GSND, Feature 2: FGC-FGC_PU, Correlation: -1.00
Feature 1: FGC-FGC_GSND, Feature 2: FGC-FGC_PU_Zone, Correlation: -1.00
Feature 1: FGC-FGC_GSND, Feature 2: FGC-FGC_SRR, Correlation: 1.00
Feature 1: FGC-FGC_GSND, Feature 2: FGC-FGC_TL, Correlation: 1.00
Feature 1: FGC-FGC_GSND, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: FGC-FGC_GSND, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: FGC-FGC_GSND, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: FGC-FGC_GSND, Feature 2: PreInt_EduHx-computerinternet_hoursday, Correlation: -1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: FGC-FGC_GSD, Correlation: 1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: FGC-FGC_PU, Correlation: -1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: FGC-FGC_PU_Zone, Correlation: -1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: FGC-FGC_SRR, Correlation: 1.00

```
Feature 1: FGC-FGC_GSND_Zone, Feature 2: FGC-FGC_TL, Correlation: 1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: FGC-FGC_GSND_Zone, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -1.00
Feature 1: FGC-FGC_GSD, Feature 2: FGC-FGC_PU, Correlation: -1.00
Feature 1: FGC-FGC_GSD, Feature 2: FGC-FGC_PU_Zone, Correlation: -1.00
Feature 1: FGC-FGC_GSD, Feature 2: FGC-FGC_SRR, Correlation: 1.00
Feature 1: FGC-FGC_GSD, Feature 2: FGC-FGC_TL, Correlation: 1.00
Feature 1: FGC-FGC_GSD, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: FGC-FGC_GSD, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: FGC-FGC_GSD, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: FGC-FGC_GSD, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -1.00
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_BMC, Correlation: 0.99
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_BMR, Correlation: 0.93
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_DEE, Correlation: 1.00
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_ECW, Correlation: 0.85
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_FFM, Correlation: 0.93
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_Frame_num, Correlation: 0.99
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_ICW, Correlation: 0.99
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_LDM, Correlation: 0.98
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_LST, Correlation: 0.93
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_SMM, Correlation: 0.91
Feature 1: FGC-FGC_PU, Feature 2: BIA-BIA_TBW, Correlation: 0.90
Feature 1: FGC-FGC_SRL, Feature 2: FGC-FGC_SRL_Zone, Correlation: 0.99
Feature 1: FGC-FGC_SRL, Feature 2: FGC-FGC_SRR, Correlation: 0.95
Feature 1: FGC-FGC_SRL, Feature 2: FGC-FGC_SRR_Zone, Correlation: 0.99
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_BMI, Correlation: 1.00
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_BMR, Correlation: 0.87
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_ECW, Correlation: 0.94
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_FFM, Correlation: 0.87
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_FFMI, Correlation: 1.00
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_FMI, Correlation: 0.99
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_Fat, Correlation: 0.99
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_LST, Correlation: 0.87
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_SMM, Correlation: 0.90
Feature 1: FGC-FGC_SRL, Feature 2: BIA-BIA_TBW, Correlation: 0.90
Feature 1: FGC-FGC_SRL, Feature 2: SDS-SDS_Total_Raw, Correlation: 0.90
Feature 1: FGC-FGC_SRL, Feature 2: SDS-SDS_Total_T, Correlation: 0.93
Feature 1: FGC-FGC_SRL_Zone, Feature 2: FGC-FGC_SRR, Correlation: 0.97
Feature 1: FGC-FGC_SRL_Zone, Feature 2: FGC-FGC_SRR_Zone, Correlation: 1.00
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_BMI, Correlation: 1.00
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_BMR, Correlation: 0.87
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_ECW, Correlation: 0.94
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_FFM, Correlation: 0.87
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_FFMI, Correlation: 1.00
```

```
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_FMI, Correlation: 0.99
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_Fat, Correlation: 0.99
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_LST, Correlation: 0.87
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_SMM, Correlation: 0.90
Feature 1: FGC-FGC_SRL_Zone, Feature 2: BIA-BIA_TBW, Correlation: 0.90
Feature 1: FGC-FGC_SRL_Zone, Feature 2: SDS-SDS_Total_Raw, Correlation: 0.95
Feature 1: FGC-FGC_SRL_Zone, Feature 2: SDS-SDS_Total_T, Correlation: 0.97
Feature 1: FGC-FGC_SRR, Feature 2: FGC-FGC_SRR_Zone, Correlation: 0.97
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_BMC, Correlation: 0.84
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_BMI, Correlation: 1.00
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_BMR, Correlation: 0.95
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_ECW, Correlation: 0.99
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_FFM, Correlation: 0.95
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_FFMI, Correlation: 0.99
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_FMI, Correlation: 1.00
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_Fat, Correlation: 1.00
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_ICW, Correlation: 0.84
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_LDM, Correlation: 0.87
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_LST, Correlation: 0.95
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_SMM, Correlation: 0.97
Feature 1: FGC-FGC_SRR, Feature 2: BIA-BIA_TBW, Correlation: 0.97
Feature 1: FGC-FGC_SRR, Feature 2: SDS-SDS_Total_Raw, Correlation: 0.99
Feature 1: FGC-FGC_SRR, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_BMI, Correlation: 1.00
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_BMR, Correlation: 0.87
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_ECW, Correlation: 0.94
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_FFM, Correlation: 0.87
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_FFMI, Correlation: 1.00
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_FMI, Correlation: 0.99
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_Fat, Correlation: 0.99
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_LST, Correlation: 0.87
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_SMM, Correlation: 0.90
Feature 1: FGC-FGC_SRR_Zone, Feature 2: BIA-BIA_TBW, Correlation: 0.90
Feature 1: FGC-FGC_SRR_Zone, Feature 2: SDS-SDS_Total_Raw, Correlation: 0.95
Feature 1: FGC-FGC_SRR_Zone, Feature 2: SDS-SDS_Total_T, Correlation: 0.97
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_BMC, Correlation: 0.97
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_BMI, Correlation: 0.91
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_BMR, Correlation: 1.00
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_DEE, Correlation: 0.93
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_ECW, Correlation: 0.98
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_FFM, Correlation: 1.00
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_FFMI, Correlation: 0.90
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_FMI, Correlation: 0.91
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_Fat, Correlation: 0.93
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_Frame_num, Correlation: 0.87
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_ICW, Correlation: 0.97
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_LDM, Correlation: 0.98
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_LST, Correlation: 1.00
```

```
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_SMM, Correlation: 1.00
Feature 1: FGC-FGC_TL, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: FGC-FGC_TL, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 0.84
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_BMI, Correlation: -1.00
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_BMR, Correlation: -0.87
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_ECW, Correlation: -0.94
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_FFM, Correlation: -0.87
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_FFMI, Correlation: -1.00
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_FMI, Correlation: -0.99
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_Fat, Correlation: -0.99
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_LST, Correlation: -0.87
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_SMM, Correlation: -0.90
Feature 1: FGC-FGC_TL_Zone, Feature 2: BIA-BIA_TBW, Correlation: -0.90
Feature 1: BIA-BIA_Activity_Level_num, Feature 2: PAQ_C-PAQ_C_Total,
Correlation: -1.00
Feature 1: BIA-BIA_Activity_Level_num, Feature 2: SDS-SDS_Total_Raw,
Correlation: -1.00
Feature 1: BIA-BIA_Activity_Level_num, Feature 2: SDS-SDS_Total_T, Correlation:
-1.00
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_BMR, Correlation: 0.97
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_DEE, Correlation: 0.99
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_ECW, Correlation: 0.91
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_FFM, Correlation: 0.97
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_Fat, Correlation: 0.82
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_Frame_num, Correlation: 0.96
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_ICW, Correlation: 1.00
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_LST, Correlation: 0.97
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_SMM, Correlation: 0.95
Feature 1: BIA-BIA_BMC, Feature 2: BIA-BIA_TBW, Correlation: 0.95
Feature 1: BIA-BIA_BMC, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_BMC, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_BMC, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_BMC, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.96
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_BMR, Correlation: 0.91
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_ECW, Correlation: 0.97
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_FFM, Correlation: 0.91
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_FFMI, Correlation: 1.00
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_FMI, Correlation: 1.00
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_Fat, Correlation: 1.00
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_LDM, Correlation: 0.82
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_LST, Correlation: 0.91
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_SMM, Correlation: 0.94
Feature 1: BIA-BIA_BMI, Feature 2: BIA-BIA_TBW, Correlation: 0.94
Feature 1: BIA-BIA_BMI, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_BMI, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_BMI, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
```

```
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_DEE, Correlation: 0.92
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_ECW, Correlation: 0.98
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_FFM, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_FFMI, Correlation: 0.90
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_FMI, Correlation: 0.91
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_Fat, Correlation: 0.93
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_Frame_num, Correlation: 0.87
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_ICW, Correlation: 0.97
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_LDM, Correlation: 0.98
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_LST, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_SMM, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_BMR, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.87
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_ECW, Correlation: 0.84
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_FFM, Correlation: 0.92
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_Frame_num, Correlation: 0.99
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_ICW, Correlation: 0.99
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_LDM, Correlation: 0.98
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_LST, Correlation: 0.92
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_SMM, Correlation: 0.90
Feature 1: BIA-BIA_DEE, Feature 2: BIA-BIA_TBW, Correlation: 0.90
Feature 1: BIA-BIA_DEE, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_DEE, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_DEE, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_DEE, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.99
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_FFM, Correlation: 0.98
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_FFMI, Correlation: 0.97
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_FMI, Correlation: 0.97
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_Fat, Correlation: 0.98
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_ICW, Correlation: 0.91
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_LDM, Correlation: 0.93
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_LST, Correlation: 0.98
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_SMM, Correlation: 0.99
Feature 1: BIA-BIA_ECW, Feature 2: BIA-BIA_TBW, Correlation: 0.99
Feature 1: BIA-BIA_ECW, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_ECW, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_ECW, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_FFMI, Correlation: 0.90
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_FMI, Correlation: 0.91
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_Fat, Correlation: 0.93
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_Frame_num, Correlation: 0.87
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_ICW, Correlation: 0.97
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_LDM, Correlation: 0.98
```

```
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_LST, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_SMM, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_FFM, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.87
Feature 1: BIA-BIA_FFMI, Feature 2: BIA-BIA_FMI, Correlation: 1.00
Feature 1: BIA-BIA_FFMI, Feature 2: BIA-BIA_Fat, Correlation: 1.00
Feature 1: BIA-BIA_FFMI, Feature 2: BIA-BIA_LDM, Correlation: 0.81
Feature 1: BIA-BIA_FFMI, Feature 2: BIA-BIA_LST, Correlation: 0.91
Feature 1: BIA-BIA_FFMI, Feature 2: BIA-BIA_SMM, Correlation: 0.93
Feature 1: BIA-BIA_FFMI, Feature 2: BIA-BIA_TBW, Correlation: 0.93
Feature 1: BIA-BIA_FFMI, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_FFMI, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_FFMI, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_FMI, Feature 2: BIA-BIA_Fat, Correlation: 1.00
Feature 1: BIA-BIA_FMI, Feature 2: BIA-BIA_LDM, Correlation: 0.82
Feature 1: BIA-BIA_FMI, Feature 2: BIA-BIA_LST, Correlation: 0.92
Feature 1: BIA-BIA_FMI, Feature 2: BIA-BIA_SMM, Correlation: 0.94
Feature 1: BIA-BIA_FMI, Feature 2: BIA-BIA_TBW, Correlation: 0.94
Feature 1: BIA-BIA_FMI, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_FMI, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_FMI, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_ICW, Correlation: 0.82
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_LDM, Correlation: 0.85
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_LST, Correlation: 0.94
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_SMM, Correlation: 0.96
Feature 1: BIA-BIA_Fat, Feature 2: BIA-BIA_TBW, Correlation: 0.96
Feature 1: BIA-BIA_Fat, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_Fat, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_Fat, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_Frame_num, Feature 2: BIA-BIA_ICW, Correlation: 0.96
Feature 1: BIA-BIA_Frame_num, Feature 2: BIA-BIA_LDM, Correlation: 0.94
Feature 1: BIA-BIA_Frame_num, Feature 2: BIA-BIA_LST, Correlation: 0.86
Feature 1: BIA-BIA_Frame_num, Feature 2: BIA-BIA_SMM, Correlation: 0.83
Feature 1: BIA-BIA_Frame_num, Feature 2: BIA-BIA_TBW, Correlation: 0.83
Feature 1: BIA-BIA_Frame_num, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -1.00
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_LDM, Correlation: 1.00
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_LST, Correlation: 0.97
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_SMM, Correlation: 0.95
Feature 1: BIA-BIA_ICW, Feature 2: BIA-BIA_TBW, Correlation: 0.95
Feature 1: BIA-BIA_ICW, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_ICW, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_ICW, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_ICW, Feature 2: PreInt_EduHx-computerinternet_hoursday,
```

```
Correlation: -0.96
Feature 1: BIA-BIA_LDM, Feature 2: BIA-BIA_LST, Correlation: 0.98
Feature 1: BIA-BIA_LDM, Feature 2: BIA-BIA_SMM, Correlation: 0.97
Feature 1: BIA-BIA_LDM, Feature 2: BIA-BIA_TBW, Correlation: 0.97
Feature 1: BIA-BIA_LDM, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_LDM, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_LDM, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_LDM, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.94
Feature 1: BIA-BIA_LST, Feature 2: BIA-BIA_SMM, Correlation: 1.00
Feature 1: BIA-BIA_LST, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_LST, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_LST, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_LST, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_LST, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.86
Feature 1: BIA-BIA_SMM, Feature 2: BIA-BIA_TBW, Correlation: 1.00
Feature 1: BIA-BIA_SMM, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_SMM, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_SMM, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_SMM, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.83
Feature 1: BIA-BIA_TBW, Feature 2: PAQ_C-PAQ_C_Total, Correlation: 1.00
Feature 1: BIA-BIA_TBW, Feature 2: SDS-SDS_Total_Raw, Correlation: 1.00
Feature 1: BIA-BIA_TBW, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
Feature 1: BIA-BIA_TBW, Feature 2: PreInt_EduHx-computerinternet_hoursday,
Correlation: -0.83
Feature 1: SDS-SDS_Total_Raw, Feature 2: SDS-SDS_Total_T, Correlation: 1.00
```

[65]:
```python
test_df.head()
```

[65]:
```
          id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0   00008ff9                      Fall                5                0
2   00105258                    Summer               10                1
3   00115b9f                    Winter                9                0
5   001f3379                    Spring               13                1
11  00abe655                      Fall               11                0

   CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  \
0       Winter             51.0            Fall     16.877316
2         Fall             71.0            Fall     16.648696
3         Fall             71.0          Summer     18.292347
5       Winter             50.0          Summer     22.279952
11      Summer             66.0             NaN           NaN

    Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0              46.0             50.8                           NaN
```

```
2             56.5             75.6                              NaN
3             56.0             81.6                              NaN
5             59.5            112.2                              NaN
11             NaN              NaN                              NaN

    Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
0                     NaN                 NaN                   NaN
2                    65.0                94.0                 117.0
3                    60.0                97.0                 117.0
5                    60.0                73.0                 102.0
11                    NaN                 NaN                   NaN

   Fitness_Endurance-Season  Fitness_Endurance-Max_Stage  \
0                       NaN                          NaN
2                      Fall                          5.0
3                    Summer                          6.0
5                       NaN                          NaN
11                      NaN                          NaN

    Fitness_Endurance-Time_Mins  Fitness_Endurance-Time_Sec FGC-Season  \
0                           NaN                         NaN       Fall
2                           7.0                        33.0       Fall
3                           9.0                        37.0     Summer
5                           NaN                         NaN     Summer
11                          NaN                         NaN     Winter

    FGC-FGC_CU  FGC-FGC_CU_Zone  FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  \
0          0.0              0.0           NaN                NaN          NaN
2         20.0              1.0          10.2                1.0         14.7
3         18.0              1.0           NaN                NaN          NaN
5         12.0              0.0          16.5                2.0         17.9
11         NaN              NaN           NaN                NaN          NaN

    FGC-FGC_GSD_Zone  FGC-FGC_PU  FGC-FGC_PU_Zone  FGC-FGC_SRL  \
0                NaN         0.0              0.0          7.0
2                2.0         7.0              1.0         10.0
3                NaN         5.0              0.0          7.0
5                2.0         6.0              0.0         10.0
11               NaN         NaN              NaN          NaN

    FGC-FGC_SRL_Zone  FGC-FGC_SRR  FGC-FGC_SRR_Zone  FGC-FGC_TL  \
0                0.0          6.0               0.0         6.0
2                1.0         10.0               1.0         5.0
3                0.0          7.0               0.0         7.0
5                1.0         11.0               1.0         8.0
11               NaN          NaN               NaN         NaN
```

```
   FGC-FGC_TL_Zone BIA-Season  BIA-BIA_Activity_Level_num  BIA-BIA_BMC  \
0              1.0       Fall                         2.0      2.66855
2              0.0        NaN                         NaN          NaN
3              1.0     Summer                         3.0      3.84191
5              0.0     Summer                         2.0      4.33036
11             NaN        NaN                         NaN          NaN

   BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0      16.8792      932.498      1492.00      8.25598      41.5862
2          NaN          NaN          NaN          NaN          NaN
3      18.2943     1131.430      1923.44     15.59250      62.7757
5      30.1865     1330.970      1996.45     30.21240      84.0285
11         NaN          NaN          NaN          NaN          NaN

   BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0       13.8177      3.06143      9.21377                1.0      24.4349
2           NaN          NaN          NaN                NaN          NaN
3       14.0740      4.22033     18.82430                2.0      30.4041
5       16.6877     13.49880     67.97150                2.0      32.9141
11          NaN          NaN          NaN                NaN          NaN

   BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW PAQ_A-Season  \
0      8.89536      38.9177      19.5413      32.6909          NaN
2          NaN          NaN          NaN          NaN          NaN
3     16.77900      58.9338      26.4798      45.9966          NaN
5     20.90200      79.6982      35.3804      63.1265          NaN
11         NaN          NaN          NaN          NaN          NaN

   PAQ_A-PAQ_A_Total PAQ_C-Season  PAQ_C-PAQ_C_Total SDS-Season  \
0                NaN          NaN                NaN        NaN
2                NaN       Summer              2.170       Fall
3                NaN       Winter              2.451     Summer
5                NaN       Spring              4.110     Summer
11               NaN       Winter              1.100     Winter

   SDS-SDS_Total_Raw  SDS-SDS_Total_T PreInt_EduHx-Season  \
0                NaN              NaN                Fall
2               38.0             54.0              Summer
3               31.0             45.0              Winter
5               40.0             56.0              Spring
11              42.0             59.0                Fall

   PreInt_EduHx-computerinternet_hoursday Age_Group
0                                     3.0       NaN
2                                     2.0      5-10
3                                     0.0      5-10
5                                     0.0     11-15
```

```
11                                       0.0      11-15
```

[66]:
```
test_df.columns
```

[66]:
```
Index(['id', 'Basic_Demos-Enroll_Season', 'Basic_Demos-Age', 'Basic_Demos-Sex',
       'CGAS-Season', 'CGAS-CGAS_Score', 'Physical-Season', 'Physical-BMI',
       'Physical-Height', 'Physical-Weight', 'Physical-Waist_Circumference',
       'Physical-Diastolic_BP', 'Physical-HeartRate', 'Physical-Systolic_BP',
       'Fitness_Endurance-Season', 'Fitness_Endurance-Max_Stage',
       'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec',
       'FGC-Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND',
       'FGC-FGC_GSND_Zone', 'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU',
       'FGC-FGC_PU_Zone', 'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR',
       'FGC-FGC_SRR_Zone', 'FGC-FGC_TL', 'FGC-FGC_TL_Zone', 'BIA-Season',
       'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC', 'BIA-BIA_BMI',
       'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM',
       'BIA-BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num',
       'BIA-BIA_ICW', 'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM',
       'BIA-BIA_TBW', 'PAQ_A-Season', 'PAQ_A-PAQ_A_Total', 'PAQ_C-Season',
       'PAQ_C-PAQ_C_Total', 'SDS-Season', 'SDS-SDS_Total_Raw',
       'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
       'PreInt_EduHx-computerinternet_hoursday', 'Age_Group'],
      dtype='object')
```

[67]:
```
# Step 1: Import necessary libraries
from sklearn.experimental import enable_iterative_imputer
from sklearn.impute import IterativeImputer
import category_encoders as ce  # for target encoding if needed
import numpy as np
import pandas as pd

# Step 2: Handle Missing Values in Numeric Columns
# Select numeric columns in test_df for imputation
numeric_features_test = test_df.select_dtypes(include=[np.number]).columns

# Remove columns with only missing values
non_missing_numeric_features_test = [col for col in numeric_features_test if␣
 ↪test_df[col].notnull().any()]

# Use IterativeImputer to perform MICE imputation on numeric columns in test_df
mice_imputer_test = IterativeImputer(max_iter=10, random_state=42)
imputed_values_test = mice_imputer_test.
 ↪fit_transform(test_df[non_missing_numeric_features_test])

# Replace original DataFrame's numeric columns (that have non-missing values)␣
 ↪with MICE-imputed values
test_df[non_missing_numeric_features_test] = imputed_values_test
```

```
# Step 3: Handle Missing Values in Categorical Features
# Select categorical features in test_df
categorical_features_test = test_df.select_dtypes(include=['category',␣
↪'object']).columns

# Check if categorical features exist in test_df
if len(categorical_features_test) > 0:
    print(f"Categorical features identified in test_df:␣
↪{categorical_features_test}")

    # Apply simple mode imputation for categorical columns (as we don't have a␣
↪target variable for encoding)
    test_df[categorical_features_test] = test_df[categorical_features_test].
↪fillna(test_df[categorical_features_test].mode().iloc[0])
else:
    print("No categorical features found for encoding in test_df.")

# Step 4: Create Missingness Indicator Columns
# Define threshold for missingness indicator creation (10% missingness)
missing_threshold_test = 0.1

# Create missingness indicator columns
for col in test_df.columns:
    missing_ratio_test = test_df[col].isnull().mean()
    if missing_ratio_test >= missing_threshold_test:
        # Create missingness indicator
        missing_indicator_col_test = f"{col}_missing"
        test_df[missing_indicator_col_test] = test_df[col].isnull().astype(int)
        print(f"Created missingness indicator for {col} in test_df (Missing␣
↪Ratio: {missing_ratio_test:.2f})")

# Final output: Print remaining columns and missing values status
print(f"Remaining columns after missing value handling in test_df: {test_df.
↪columns}")
print("\nRemaining missing values in test_df after imputation:")
print(test_df.isnull().sum())
```

```
Categorical features identified in test_df: Index(['id', 'Basic_Demos-
Enroll_Season', 'CGAS-Season', 'Physical-Season',
       'Fitness_Endurance-Season', 'FGC-Season', 'BIA-Season', 'PAQ_A-Season',
       'PAQ_C-Season', 'SDS-Season', 'PreInt_EduHx-Season', 'Age_Group'],
      dtype='object')
Created missingness indicator for Physical-Waist_Circumference in test_df
(Missing Ratio: 1.00)
Created missingness indicator for PAQ_A-Season in test_df (Missing Ratio: 1.00)
Created missingness indicator for PAQ_A-PAQ_A_Total in test_df (Missing Ratio:
```

```
1.00)
Remaining columns after missing value handling in test_df: Index(['id',
'Basic_Demos-Enroll_Season', 'Basic_Demos-Age', 'Basic_Demos-Sex',
       'CGAS-Season', 'CGAS-CGAS_Score', 'Physical-Season', 'Physical-BMI',
       'Physical-Height', 'Physical-Weight', 'Physical-Waist_Circumference',
       'Physical-Diastolic_BP', 'Physical-HeartRate', 'Physical-Systolic_BP',
       'Fitness_Endurance-Season', 'Fitness_Endurance-Max_Stage',
       'Fitness_Endurance-Time_Mins', 'Fitness_Endurance-Time_Sec',
       'FGC-Season', 'FGC-FGC_CU', 'FGC-FGC_CU_Zone', 'FGC-FGC_GSND',
       'FGC-FGC_GSND_Zone', 'FGC-FGC_GSD', 'FGC-FGC_GSD_Zone', 'FGC-FGC_PU',
       'FGC-FGC_PU_Zone', 'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'FGC-FGC_SRR',
       'FGC-FGC_SRR_Zone', 'FGC-FGC_TL', 'FGC-FGC_TL_Zone', 'BIA-Season',
       'BIA-BIA_Activity_Level_num', 'BIA-BIA_BMC', 'BIA-BIA_BMI',
       'BIA-BIA_BMR', 'BIA-BIA_DEE', 'BIA-BIA_ECW', 'BIA-BIA_FFM',
       'BIA-BIA_FFMI', 'BIA-BIA_FMI', 'BIA-BIA_Fat', 'BIA-BIA_Frame_num',
       'BIA-BIA_ICW', 'BIA-BIA_LDM', 'BIA-BIA_LST', 'BIA-BIA_SMM',
       'BIA-BIA_TBW', 'PAQ_A-Season', 'PAQ_A-PAQ_A_Total', 'PAQ_C-Season',
       'PAQ_C-PAQ_C_Total', 'SDS-Season', 'SDS-SDS_Total_Raw',
       'SDS-SDS_Total_T', 'PreInt_EduHx-Season',
       'PreInt_EduHx-computerinternet_hoursday', 'Age_Group',
       'Physical-Waist_Circumference_missing', 'PAQ_A-Season_missing',
       'PAQ_A-PAQ_A_Total_missing'],
      dtype='object')

Remaining missing values in test_df after imputation:
id                                        0
Basic_Demos-Enroll_Season                 0
Basic_Demos-Age                           0
Basic_Demos-Sex                           0
CGAS-Season                               0
                                         ..
PreInt_EduHx-computerinternet_hoursday    0
Age_Group                                 0
Physical-Waist_Circumference_missing      0
PAQ_A-Season_missing                      0
PAQ_A-PAQ_A_Total_missing                 0
Length: 63, dtype: int64

/opt/conda/lib/python3.10/site-packages/sklearn/impute/_iterative.py:825:
ConvergenceWarning: [IterativeImputer] Early stopping criterion not reached.
  warnings.warn(
/var/tmp/ipykernel_7796/2051824373.py:31: FutureWarning: Downcasting object
dtype arrays on .fillna, .ffill, .bfill is deprecated and will change in a
future version. Call result.infer_objects(copy=False) instead. To opt-in to the
future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
  test_df[categorical_features_test] = test_df[categorical_features_test].fillna
(test_df[categorical_features_test].mode().iloc[0])
```

[68]:
```
test_df.head()
```

[68]:
```
          id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0   00008ff9                      Fall              5.0              0.0
2   00105258                    Summer             10.0              1.0
3   00115b9f                    Winter              9.0              0.0
5   001f3379                    Spring             13.0              1.0
11  00abe655                      Fall             11.0              0.0

   CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  \
0       Winter             51.0            Fall     16.877316
2         Fall             71.0            Fall     16.648696
3         Fall             71.0          Summer     18.292347
5       Winter             50.0          Summer     22.279952
11      Summer             66.0            Fall     17.864055

    Physical-Height  Physical-Weight  Physical-Waist_Circumference  \
0         46.000000        50.800000                           NaN
2         56.500000        75.600000                           NaN
3         56.000000        81.600000                           NaN
5         59.500000       112.200000                           NaN
11        53.089389        73.360103                           NaN

    Physical-Diastolic_BP  Physical-HeartRate  Physical-Systolic_BP  \
0               76.659994           42.625084             95.537106
2               65.000000           94.000000            117.000000
3               60.000000           97.000000            117.000000
5               60.000000           73.000000            102.000000
11              64.906547           79.417997            109.671015

   Fitness_Endurance-Season  Fitness_Endurance-Max_Stage  \
0                      Fall                     0.108192
2                      Fall                     5.000000
3                    Summer                     6.000000
5                      Fall                     8.161510
11                     Fall                     4.434398

    Fitness_Endurance-Time_Mins  Fitness_Endurance-Time_Sec FGC-Season  \
0                     -2.783656                   13.432611       Fall
2                      7.000000                   33.000000       Fall
3                      9.000000                   37.000000     Summer
5                     13.323058                   45.646307     Summer
11                     5.868766                   30.737339     Winter

    FGC-FGC_CU  FGC-FGC_CU_Zone  FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  \
0     0.000000          0.00000      3.706899          -0.030646    11.401920
2    20.000000          1.00000     10.200000           1.000000    14.700000
```

```
3   18.000000         1.00000    11.694644          1.237246    15.459184
5   12.000000         0.00000    16.500000          2.000000    17.900000
11  11.771994         0.48317     9.702796          0.921081    14.447450

   FGC-FGC_GSD_Zone  FGC-FGC_PU  FGC-FGC_PU_Zone  FGC-FGC_SRL  \
0               2.0    0.000000         0.000000     7.000000
2               2.0    7.000000         1.000000    10.000000
3               2.0    5.000000         0.000000     7.000000
5               2.0    6.000000         0.000000    10.000000
11              2.0    4.413705         0.458711     9.593142

   FGC-FGC_SRL_Zone  FGC-FGC_SRR  FGC-FGC_SRR_Zone  FGC-FGC_TL  \
0          0.000000     6.000000          0.000000     6.00000
2          1.000000    10.000000          1.000000     5.00000
3          0.000000     7.000000          0.000000     7.00000
5          1.000000    11.000000          1.000000     8.00000
11         0.832517     9.539816          0.834634     6.11333

   FGC-FGC_TL_Zone BIA-Season  BIA-BIA_Activity_Level_num  BIA-BIA_BMC  \
0         1.000000       Fall                    2.000000     2.668550
2         0.000000     Summer                    2.940449     3.641048
3         1.000000     Summer                    3.000000     3.841910
5         0.000000     Summer                    2.000000     4.330360
11        0.357534     Summer                    2.634960     3.539088

   BIA-BIA_BMI  BIA-BIA_BMR  BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  \
0    16.879200   932.498000  1492.000000     8.255980    41.586200
2    17.160848  1087.854533  1858.417991    13.394408    58.139277
3    18.294300  1131.430000  1923.440000    15.592500    62.775700
5    30.186500  1330.970000  1996.450000    30.212400    84.028500
11   18.782182  1087.500178  1803.297738    14.468455    58.086989

   BIA-BIA_FFMI  BIA-BIA_FMI  BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  \
0     13.817700     3.061430     9.213770           1.000000    24.434900
2     13.825440     3.303771    13.482399           1.863178    29.381756
3     14.074000     4.220330    18.824300           2.000000    30.404100
5     16.687700    13.498800    67.971500           2.000000    32.914100
11    14.178944     4.520419    19.377546           1.710098    28.844178

   BIA-BIA_LDM  BIA-BIA_LST  BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  \
0     8.895360    38.917700    19.541300    32.690900           NaN
2    15.362565    54.493534    24.801015    42.771962           NaN
3    16.779000    58.933800    26.479800    45.996600           NaN
5    20.902000    79.698200    35.380400    63.126500           NaN
11   14.773667    54.540391    25.072255    43.300657           NaN

   PAQ_A-PAQ_A_Total PAQ_C-Season  PAQ_C-PAQ_C_Total SDS-Season  \
```

```
0                  NaN       Winter          -2.892886     Summer
2                  NaN       Summer           2.170000       Fall
3                  NaN       Winter           2.451000     Summer
5                  NaN       Spring           4.110000     Summer
11                 NaN       Winter           1.100000     Winter

    SDS-SDS_Total_Raw  SDS-SDS_Total_T PreInt_EduHx-Season  \
0           65.892149        89.323116                Fall
2           38.000000        54.000000              Summer
3           31.000000        45.000000              Winter
5           40.000000        56.000000              Spring
11          42.000000        59.000000                Fall

    PreInt_EduHx-computerinternet_hoursday Age_Group  \
0                                      3.0      5-10
2                                      2.0      5-10
3                                      0.0      5-10
5                                      0.0     11-15
11                                     0.0     11-15

    Physical-Waist_Circumference_missing  PAQ_A-Season_missing  \
0                                      1                     1
2                                      1                     1
3                                      1                     1
5                                      1                     1
11                                     1                     1

    PAQ_A-PAQ_A_Total_missing
0                           1
2                           1
3                           1
5                           1
11                          1
```

[69]:

```
# Aggregate actigraphy data by ID
actigraphy_test_summary_df = (
    actigraphy_test_df
    .groupby('ID')  # Group by ID
    .agg({
        'X': ['mean', 'std', 'min', 'max'],  # Aggregate X with mean, std, min,␣
 ↪max
        'Y': ['mean', 'std', 'min', 'max'],  # Aggregate Y with mean, std, min,␣
 ↪max
        'Z': ['mean', 'std', 'min', 'max'],  # Aggregate Z with mean, std, min,␣
 ↪max
        'enmo': ['mean', 'std'],  # Aggregate enmo with mean, std
        'anglez': ['mean', 'std'],  # Aggregate anglez with mean, std
```

```
        'non-wear_flag': 'mean',  # Proportion of non-wear time
        'light': ['mean', 'std'],  # Aggregate light with mean, std
        'battery_voltage': ['mean', 'std'],  # Aggregate battery voltage
        'time_of_day': 'mean',  # Mean time of day
        'weekday': 'mean',  # Average weekday for the activity readings
        'quarter': 'mean',  # Average quarter for the activity readings
        'relative_date_PCIAT': 'mean'  # Average relative date PCIAT
    })
)

# Flatten MultiIndex columns
actigraphy_test_summary_df.columns = ['_'.join(col) for col in␣
 ↪actigraphy_test_summary_df.columns]

# Reset index to make 'ID' a column again
actigraphy_test_summary_df.reset_index(inplace=True)

# Display first few rows of the aggregated summary
actigraphy_test_summary_df.head()
```

[69]:

```
         ID    X_mean     X_std     X_min     X_max    Y_mean     Y_std  \
0  00115b9f -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702
1  001f3379 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812

      Y_min     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  \
0 -2.905339  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388
1 -1.522690  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926

   enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  light_mean  \
0  0.106351   -10.580416   42.947170            0.000000    42.29631
1  0.024331   -55.630768   50.303635            0.655708    16.77198

    light_std  battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
0  208.168976           4053.578857           112.404037      5.046215e+13
1   95.327438           3838.189453           155.573868      4.321212e+13

   weekday_mean  quarter_mean  relative_date_PCIAT_mean
0      4.470182           3.0                 53.201683
1      3.909848           3.0                 79.435593
```

[70]:

```
import numpy as np

# Function to compute top n Fourier coefficients
def compute_top_n_fourier_coefficients(signal, n=5):
    # Perform the Fourier Transform on the signal
    fft_coeffs = np.fft.fft(signal)
    # Calculate magnitudes of the FFT coefficients
```

```
    magnitudes = np.abs(fft_coeffs)
    # Get top n coefficients based on magnitudes (excluding the DC component at␣
↪index 0)
    top_n_indices = np.argsort(magnitudes)[-n:]
    # Return the real and imaginary parts of the top n coefficients
    top_n_real_parts = fft_coeffs[top_n_indices].real
    top_n_imag_parts = fft_coeffs[top_n_indices].imag
    return top_n_real_parts, top_n_imag_parts

# List to store Fourier features for each participant
fourier_features_list = []

# Calculate Fourier Transforms for each ID in the actigraphy training data
for participant_id, participant_data in actigraphy_test_df.groupby('ID'):
    # Calculate Fourier coefficients for each axis
    fourier_X_real, fourier_X_imag =␣
↪compute_top_n_fourier_coefficients(participant_data['X'].values)
    fourier_Y_real, fourier_Y_imag =␣
↪compute_top_n_fourier_coefficients(participant_data['Y'].values)
    fourier_Z_real, fourier_Z_imag =␣
↪compute_top_n_fourier_coefficients(participant_data['Z'].values)

    # Create a dictionary to store Fourier features for this participant
    fourier_features = {
        'ID': participant_id,
        'fourier_X_real': fourier_X_real.mean(),  # Using mean of the top real␣
↪parts as a feature
        'fourier_X_imag': fourier_X_imag.mean(),  # Using mean of the top␣
↪imaginary parts as a feature
        'fourier_Y_real': fourier_Y_real.mean(),
        'fourier_Y_imag': fourier_Y_imag.mean(),
        'fourier_Z_real': fourier_Z_real.mean(),
        'fourier_Z_imag': fourier_Z_imag.mean(),
    }
    fourier_features_list.append(fourier_features)

# Convert the list of Fourier features into a DataFrame
fourier_features_df = pd.DataFrame(fourier_features_list)

# Merge the Fourier features into the existing actigraphy summary dataframe
actigraphy_test_summary_with_fourier_df = pd.merge(
    actigraphy_test_summary_df,
    fourier_features_df,
    on='ID',
    how='left'
)
```

```python
# Display first few rows of the combined summary
actigraphy_train_summary_with_fourier_df.head()
```

[70]:

```
         ID    X_mean     X_std     X_min     X_max    Y_mean     Y_std  \
0  00115b9f -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702
1  001f3379 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812
2  00f332d1  0.208036  0.486977 -1.952594  1.666465  0.057094  0.443755
3  01085eb3 -0.343396  0.516126 -2.284304  1.000692 -0.055826  0.424303
4  012cadd8  0.018670  0.595251 -2.143912  3.341210  0.071660  0.508311

      Y_min     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  \
0 -2.905339  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388
1 -1.522690  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926
2 -2.361866  1.016429  0.141550  0.683114 -1.016758  2.239939   0.030255
3 -2.276082  1.011419 -0.254433  0.564593 -1.022549  1.299293   0.032946
4 -3.373025  4.442658 -0.061682  0.578022 -1.003249  2.321265   0.058280

   enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  light_mean  \
0  0.106351   -10.580416   42.947170            0.000000   42.296310
1  0.024331   -55.630768   50.303635            0.655708   16.771980
2  0.104136     6.687339   52.754208            0.171246   66.563393
3  0.083798   -17.589037   39.895645            0.035210   17.800735
4  0.197285    -5.059758   39.994808            0.000000   54.893402

    light_std  battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
0  208.168976           4053.578857           112.404037      5.046215e+13
1   95.327438           3838.189453           155.573868      4.321212e+13
2  286.916595           3848.583252           166.968582      4.318680e+13
3   73.023468           3849.650146           171.100159      4.338433e+13
4  230.972397           3974.910889           119.525154      4.343573e+13

   weekday_mean  quarter_mean  relative_date_PCIAT_mean  fourier_X_real  \
0      4.470182           3.0                 53.201683    -2653.531675
1      3.909848           3.0                 79.435593     4103.865122
2      3.832677           2.0                 26.152903    14903.301506
3      3.963284           4.0                 49.910686   -25796.646158
4      4.168412           4.0                 -1.168288     6637.890207

   fourier_X_imag  fourier_Y_real  fourier_Y_imag  fourier_Z_real  \
0   -1.083578e-14      705.803060   -2.074467e+02     1995.401835
1   -1.680035e+03    14385.071425    3.120931e+03   -26765.674436
2   -3.637979e-13     -923.242685    7.275958e-13    -5261.764351
3   -1.093525e-12   -10776.376135   -1.989520e-13    -8614.531458
4    5.618541e+02     2279.335449   -3.637979e-13    -4597.967535

   fourier_Z_imag
```

```
0   -5.684342e-15
1   -8.731149e-12
2    7.275958e-13
3   -9.563905e-13
4    0.000000e+00
```

[71]:
```python
# Function to calculate day and night activity based on time_of_day
def calculate_day_night_activity(df, day_start=6, day_end=18):
    # Calculate day activity (time_of_day between day_start and day_end)
    day_activity = df[(df['time_of_day'] >= day_start) & (df['time_of_day'] <␣
↪day_end)]['enmo'].sum()
    # Calculate night activity (time_of_day outside the day range)
    night_activity = df[(df['time_of_day'] < day_start) | (df['time_of_day'] >=␣
↪day_end)]['enmo'].sum()
    return day_activity, night_activity

# List to store time-domain features for each participant
time_features_list = []

# Calculate time-domain features for each participant
for participant_id, participant_data in actigraphy_test_df.groupby('ID'):
    # Calculate day and night activity
    day_activity, night_activity =␣
↪calculate_day_night_activity(participant_data)
    activity_ratio_day_night = day_activity / (night_activity + 1e-6)  # Avoid␣
↪division by zero

    # Calculate proportion of non-wear time
    non_wear_proportion = participant_data['non-wear_flag'].mean()

    # Calculate proportion of time spent in different activity levels based on␣
↪'enmo'
    sedentary_time = participant_data[participant_data['enmo'] < 0.05].shape[0]
    light_activity_time = participant_data[(participant_data['enmo'] >= 0.05) &␣
↪(participant_data['enmo'] < 0.5)].shape[0]
    moderate_activity_time = participant_data[(participant_data['enmo'] >= 0.5)␣
↪& (participant_data['enmo'] < 1)].shape[0]
    vigorous_activity_time = participant_data[participant_data['enmo'] >= 1].
↪shape[0]
    total_time = participant_data.shape[0]

    # Calculate proportions
    sedentary_proportion = sedentary_time / total_time
    light_activity_proportion = light_activity_time / total_time
    moderate_activity_proportion = moderate_activity_time / total_time
    vigorous_activity_proportion = vigorous_activity_time / total_time
```

```
    # Create a dictionary of time domain features
    time_features = {
        'ID': participant_id,
        'activity_during_day': day_activity,
        'activity_during_night': night_activity,
        'activity_ratio_day_night': activity_ratio_day_night,
        'non_wear_proportion': non_wear_proportion,
        'sedentary_proportion': sedentary_proportion,
        'light_activity_proportion': light_activity_proportion,
        'moderate_activity_proportion': moderate_activity_proportion,
        'vigorous_activity_proportion': vigorous_activity_proportion
    }
    time_features_list.append(time_features)

# Convert the list of time domain features into a DataFrame
time_features_df = pd.DataFrame(time_features_list)

# Merge the time-domain features into the actigraphy summary with Fourier␣
 ↪features
actigraphy_combined_features_test_df = pd.merge(
    actigraphy_test_summary_with_fourier_df,  # The dataframe with Fourier␣
 ↪features
    time_features_df,  # The new time-domain features
    on='ID',
    how='left'
)

# Display the first few rows of the final combined feature set
actigraphy_combined_features_test_df.head(10)
```

[71]:

```
         ID    X_mean     X_std     X_min     X_max    Y_mean     Y_std  \
0  00115b9f -0.316384  0.453665 -1.746094  1.507865  0.016009  0.502702
1  001f3379 -0.004272  0.351545 -1.038711  1.034351  0.016859  0.303812

      Y_min     Y_max    Z_mean     Z_std     Z_min     Z_max  enmo_mean  \
0 -2.905339  1.666354 -0.167890  0.585710 -1.048372  1.546979   0.047388
1 -1.522690  1.946303 -0.631731  0.623476 -1.018787  1.146284   0.011926

   enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  light_mean  \
0  0.106351   -10.580416   42.947170            0.000000    42.29631
1  0.024331   -55.630768   50.303635            0.655708    16.77198

    light_std  battery_voltage_mean  battery_voltage_std  time_of_day_mean  \
0  208.168976           4053.578857           112.404037      5.046215e+13
1   95.327438           3838.189453           155.573868      4.321212e+13
```

```
   weekday_mean  quarter_mean  relative_date_PCIAT_mean  fourier_X_real  \
0      4.470182           3.0                 53.201683    -2653.531675
1      3.909848           3.0                 79.435593     4103.865122

   fourier_X_imag  fourier_Y_real  fourier_Y_imag  fourier_Z_real  \
0   -1.083578e-14      705.803060     -207.446683     1995.401835
1   -1.680035e+03    14385.071425     3120.930726   -26765.674436

   fourier_Z_imag  activity_during_day  activity_during_night  \
0   -5.684342e-15                  0.0            2053.305176
1   -8.731149e-12                  0.0            4727.518555

   activity_ratio_day_night  non_wear_proportion  sedentary_proportion  \
0                       0.0             0.000000              0.792453
1                       0.0             0.655708              0.978501

   light_activity_proportion  moderate_activity_proportion  \
0                   0.198131                      0.007870
1                   0.021171                      0.000288

   vigorous_activity_proportion
0                      0.001546
1                      0.000040
```

[72]:

```python
# Merging the main training dataset with actigraphy_combined_features_df based␣
↪on the ID column
merged_test_df = pd.merge(test_df, actigraphy_combined_features_test_df,␣
↪how='left', left_on='id', right_on='ID')

# Dropping the redundant ID column from actigraphy_combined_features_df if␣
↪needed
merged_test_df = merged_test_df.drop(columns=['ID'])

# Display the first few rows of the merged DataFrame to confirm the merge
#print(merged_train_df.head())
```

[73]:

```python
# Define the list of actigraphy feature columns
actigraphy_columns = [
    'X_mean', 'X_std', 'X_min', 'X_max', 'Y_mean', 'Y_std', 'Y_min', 'Y_max',
    'Z_mean', 'Z_std', 'Z_min', 'Z_max', 'enmo_mean', 'enmo_std', 'anglez_mean',
    'anglez_std', 'non-wear_flag_mean', 'light_mean', 'light_std',␣
↪'battery_voltage_mean',
    'battery_voltage_std', 'time_of_day_mean', 'weekday_mean', 'quarter_mean',
    'relative_date_PCIAT_mean', 'fourier_X_real', 'fourier_X_imag',␣
↪'fourier_Y_real',
    'fourier_Y_imag', 'fourier_Z_real', 'fourier_Z_imag', 'activity_during_day',
    'activity_during_night', 'activity_ratio_day_night', 'non_wear_proportion',
```

```
    'sedentary_proportion', 'light_activity_proportion',␣
 ↪'moderate_activity_proportion',
    'vigorous_activity_proportion'
]

# Fill NaN values in actigraphy columns with zeros if all actigraphy columns␣
 ↪are NaN in a row
merged_test_df[actigraphy_columns] = merged_test_df[actigraphy_columns].
 ↪apply(lambda row: row.fillna(0) if row.isnull().all() else row, axis=1)

# Create actigraphy_present column
# If all actigraphy columns have 0 values for a given row, set␣
 ↪actigraphy_present to 0, otherwise 1
merged_test_df['actigraphy_present'] = merged_test_df[actigraphy_columns].
 ↪apply(lambda row: 0 if (row == 0).all() else 1, axis=1)

# Display the first few rows to verify the results
merged_test_df.head()
```

[73]:

```
         id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0  00008ff9                      Fall              5.0              0.0
1  00105258                    Summer             10.0              1.0
2  00115b9f                    Winter              9.0              0.0
3  001f3379                    Spring             13.0              1.0
4  00abe655                      Fall             11.0              0.0

  CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  Physical-Height  \
0      Winter             51.0            Fall     16.877316        46.000000
1        Fall             71.0            Fall     16.648696        56.500000
2        Fall             71.0          Summer     18.292347        56.000000
3      Winter             50.0          Summer     22.279952        59.500000
4      Summer             66.0            Fall     17.864055        53.089389

   Physical-Weight  Physical-Waist_Circumference  Physical-Diastolic_BP  \
0        50.800000                           NaN              76.659994
1        75.600000                           NaN              65.000000
2        81.600000                           NaN              60.000000
3       112.200000                           NaN              60.000000
4        73.360103                           NaN              64.906547

   Physical-HeartRate  Physical-Systolic_BP Fitness_Endurance-Season  \
0           42.625084             95.537106                     Fall
1           94.000000            117.000000                     Fall
2           97.000000            117.000000                   Summer
3           73.000000            102.000000                     Fall
4           79.417997            109.671015                     Fall
```

```
   Fitness_Endurance-Max_Stage  Fitness_Endurance-Time_Mins  \
0                     0.108192                    -2.783656
1                     5.000000                     7.000000
2                     6.000000                     9.000000
3                     8.161510                    13.323058
4                     4.434398                     5.868766

   Fitness_Endurance-Time_Sec FGC-Season  FGC-FGC_CU  FGC-FGC_CU_Zone  \
0                   13.432611       Fall    0.000000          0.00000
1                   33.000000       Fall   20.000000          1.00000
2                   37.000000     Summer   18.000000          1.00000
3                   45.646307     Summer   12.000000          0.00000
4                   30.737339     Winter   11.771994          0.48317

   FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  FGC-FGC_GSD_Zone  FGC-FGC_PU  \
0      3.706899          -0.030646    11.401920               2.0    0.000000
1     10.200000           1.000000    14.700000               2.0    7.000000
2     11.694644           1.237246    15.459184               2.0    5.000000
3     16.500000           2.000000    17.900000               2.0    6.000000
4      9.702796           0.921081    14.447450               2.0    4.413705

   FGC-FGC_PU_Zone  FGC-FGC_SRL  FGC-FGC_SRL_Zone  FGC-FGC_SRR  \
0         0.000000     7.000000          0.000000     6.000000
1         1.000000    10.000000          1.000000    10.000000
2         0.000000     7.000000          0.000000     7.000000
3         0.000000    10.000000          1.000000    11.000000
4         0.458711     9.593142          0.832517     9.539816

   FGC-FGC_SRR_Zone  FGC-FGC_TL  FGC-FGC_TL_Zone BIA-Season  \
0          0.000000     6.00000         1.000000       Fall
1          1.000000     5.00000         0.000000     Summer
2          0.000000     7.00000         1.000000     Summer
3          1.000000     8.00000         0.000000     Summer
4          0.834634     6.11333         0.357534     Summer

   BIA-BIA_Activity_Level_num  BIA-BIA_BMC  BIA-BIA_BMI  BIA-BIA_BMR  \
0                    2.000000     2.668550    16.879200   932.498000
1                    2.940449     3.641048    17.160848  1087.854533
2                    3.000000     3.841910    18.294300  1131.430000
3                    2.000000     4.330360    30.186500  1330.970000
4                    2.634960     3.539088    18.782182  1087.500178

   BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  BIA-BIA_FFMI  BIA-BIA_FMI  \
0  1492.000000     8.255980    41.586200     13.817700     3.061430
1  1858.417991    13.394408    58.139277     13.825440     3.303771
2  1923.440000    15.592500    62.775700     14.074000     4.220330
3  1996.450000    30.212400    84.028500     16.687700    13.498800
```

```
4  1803.297738    14.468455    58.086989     14.178944     4.520419

   BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  BIA-BIA_LDM  BIA-BIA_LST  \
0     9.213770           1.000000    24.434900     8.895360    38.917700
1    13.482399           1.863178    29.381756    15.362565    54.493534
2    18.824300           2.000000    30.404100    16.779000    58.933800
3    67.971500           2.000000    32.914100    20.902000    79.698200
4    19.377546           1.710098    28.844178    14.773667    54.540391

   BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  PAQ_A-PAQ_A_Total PAQ_C-Season  \
0    19.541300    32.690900           NaN                NaN       Winter
1    24.801015    42.771962           NaN                NaN       Summer
2    26.479800    45.996600           NaN                NaN       Winter
3    35.380400    63.126500           NaN                NaN       Spring
4    25.072255    43.300657           NaN                NaN       Winter

   PAQ_C-PAQ_C_Total SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  \
0          -2.892886     Summer          65.892149        89.323116
1           2.170000       Fall          38.000000        54.000000
2           2.451000     Summer          31.000000        45.000000
3           4.110000     Summer          40.000000        56.000000
4           1.100000     Winter          42.000000        59.000000

  PreInt_EduHx-Season  PreInt_EduHx-computerinternet_hoursday Age_Group  \
0                Fall                                     3.0      5-10
1              Summer                                     2.0      5-10
2              Winter                                     0.0      5-10
3              Spring                                     0.0     11-15
4                Fall                                     0.0     11-15

   Physical-Waist_Circumference_missing  PAQ_A-Season_missing  \
0                                     1                     1
1                                     1                     1
2                                     1                     1
3                                     1                     1
4                                     1                     1

   PAQ_A-PAQ_A_Total_missing    X_mean     X_std     X_min     X_max  \
0                          1  0.000000  0.000000  0.000000  0.000000
1                          1  0.000000  0.000000  0.000000  0.000000
2                          1 -0.316384  0.453665 -1.746094  1.507865
3                          1 -0.004272  0.351545 -1.038711  1.034351
4                          1  0.000000  0.000000  0.000000  0.000000

     Y_mean     Y_std     Y_min     Y_max    Z_mean     Z_std     Z_min  \
0  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
```

```
2  0.016009  0.502702 -2.905339  1.666354 -0.167890  0.585710 -1.048372
3  0.016859  0.303812 -1.522690  1.946303 -0.631731  0.623476 -1.018787
4  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000

      Z_max  enmo_mean  enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  \
0  0.000000   0.000000  0.000000     0.000000    0.000000            0.000000
1  0.000000   0.000000  0.000000     0.000000    0.000000            0.000000
2  1.546979   0.047388  0.106351   -10.580416   42.947170            0.000000
3  1.146284   0.011926  0.024331   -55.630768   50.303635            0.655708
4  0.000000   0.000000  0.000000     0.000000    0.000000            0.000000

   light_mean   light_std  battery_voltage_mean  battery_voltage_std  \
0     0.00000    0.000000              0.000000             0.000000
1     0.00000    0.000000              0.000000             0.000000
2    42.29631  208.168976           4053.578857           112.404037
3    16.77198   95.327438           3838.189453           155.573868
4     0.00000    0.000000              0.000000             0.000000

   time_of_day_mean  weekday_mean  quarter_mean  relative_date_PCIAT_mean  \
0      0.000000e+00      0.000000           0.0                  0.000000
1      0.000000e+00      0.000000           0.0                  0.000000
2      5.046215e+13      4.470182           3.0                 53.201683
3      4.321212e+13      3.909848           3.0                 79.435593
4      0.000000e+00      0.000000           0.0                  0.000000

   fourier_X_real  fourier_X_imag  fourier_Y_real  fourier_Y_imag  \
0        0.000000    0.000000e+00        0.000000        0.000000
1        0.000000    0.000000e+00        0.000000        0.000000
2    -2653.531675   -1.083578e-14      705.803060     -207.446683
3     4103.865122   -1.680035e+03    14385.071425     3120.930726
4        0.000000    0.000000e+00        0.000000        0.000000

   fourier_Z_real  fourier_Z_imag  activity_during_day  activity_during_night  \
0        0.000000    0.000000e+00                  0.0               0.000000
1        0.000000    0.000000e+00                  0.0               0.000000
2     1995.401835   -5.684342e-15                  0.0            2053.305176
3   -26765.674436   -8.731149e-12                  0.0            4727.518555
4        0.000000    0.000000e+00                  0.0               0.000000

   activity_ratio_day_night  non_wear_proportion  sedentary_proportion  \
0                       0.0             0.000000              0.000000
1                       0.0             0.000000              0.000000
2                       0.0             0.000000              0.792453
3                       0.0             0.655708              0.978501
4                       0.0             0.000000              0.000000

   light_activity_proportion  moderate_activity_proportion  \
```

```
0                    0.000000                      0.000000
1                    0.000000                      0.000000
2                    0.198131                      0.007870
3                    0.021171                      0.000288
4                    0.000000                      0.000000

   vigorous_activity_proportion  actigraphy_present
0                      0.000000                   0
1                      0.000000                   0
2                      0.001546                   1
3                      0.000040                   1
4                      0.000000                   0
```

[74]:
```
list(oversampled_df.columns)
```

[74]:
```
['Basic_Demos-Enroll_Season',
 'Basic_Demos-Age',
 'Basic_Demos-Sex',
 'CGAS-Season',
 'CGAS-CGAS_Score',
 'Physical-Season',
 'Physical-BMI',
 'Physical-Height',
 'Physical-Weight',
 'Physical-Waist_Circumference',
 'Physical-Diastolic_BP',
 'Physical-HeartRate',
 'Physical-Systolic_BP',
 'Fitness_Endurance-Season',
 'Fitness_Endurance-Max_Stage',
 'Fitness_Endurance-Time_Mins',
 'Fitness_Endurance-Time_Sec',
 'FGC-Season',
 'FGC-FGC_CU',
 'FGC-FGC_CU_Zone',
 'FGC-FGC_GSND',
 'FGC-FGC_GSND_Zone',
 'FGC-FGC_GSD',
 'FGC-FGC_GSD_Zone',
 'FGC-FGC_PU',
 'FGC-FGC_PU_Zone',
 'FGC-FGC_SRL',
 'FGC-FGC_SRL_Zone',
 'FGC-FGC_SRR',
 'FGC-FGC_SRR_Zone',
 'FGC-FGC_TL',
 'FGC-FGC_TL_Zone',
```

```
'BIA-Season',
'BIA-BIA_Activity_Level_num',
'BIA-BIA_BMC',
'BIA-BIA_BMI',
'BIA-BIA_BMR',
'BIA-BIA_DEE',
'BIA-BIA_ECW',
'BIA-BIA_FFM',
'BIA-BIA_FFMI',
'BIA-BIA_FMI',
'BIA-BIA_Fat',
'BIA-BIA_Frame_num',
'BIA-BIA_ICW',
'BIA-BIA_LDM',
'BIA-BIA_LST',
'BIA-BIA_SMM',
'BIA-BIA_TBW',
'PAQ_A-Season',
'PAQ_A-PAQ_A_Total',
'PAQ_C-Season',
'PAQ_C-PAQ_C_Total',
'PCIAT-PCIAT_Total',
'SDS-Season',
'SDS-SDS_Total_Raw',
'SDS-SDS_Total_T',
'PreInt_EduHx-Season',
'PreInt_EduHx-computerinternet_hoursday',
'Age_Group',
'PCIAT_Cluster',
'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw',
'CGAS-CGAS_Score_x_Physical-BMI',
'X_mean',
'X_std',
'X_min',
'X_max',
'Y_mean',
'Y_std',
'Y_min',
'Y_max',
'Z_mean',
'Z_std',
'Z_min',
'Z_max',
'enmo_mean',
'enmo_std',
'anglez_mean',
'anglez_std',
```

```
 'non-wear_flag_mean',
 'light_mean',
 'light_std',
 'battery_voltage_mean',
 'battery_voltage_std',
 'time_of_day_mean',
 'weekday_mean',
 'quarter_mean',
 'relative_date_PCIAT_mean',
 'fourier_X_real',
 'fourier_X_imag',
 'fourier_Y_real',
 'fourier_Y_imag',
 'fourier_Z_real',
 'fourier_Z_imag',
 'activity_during_day',
 'activity_during_night',
 'activity_ratio_day_night',
 'non_wear_proportion',
 'sedentary_proportion',
 'light_activity_proportion',
 'moderate_activity_proportion',
 'vigorous_activity_proportion',
 'actigraphy_present',
 'sii',
 'id']
```

[75]:
```
# Step 1: Create sets of columns for easier comparison
oversampled_columns_set = set(oversampled_df.columns)
merged_test_columns_set = set(merged_test_df.columns)

# Step 2: Identify shared columns
shared_columns = list(oversampled_columns_set.
 ↪intersection(merged_test_columns_set))
print(f"Shared Columns between oversampled_df and merged_test_df:
 ↪\n{shared_columns}\n")

# Step 3: Identify columns present in oversampled_df but missing in␣
 ↪merged_test_df
missing_in_test_df = list(oversampled_columns_set - merged_test_columns_set)
print(f"Columns in oversampled_df but missing in merged_test_df:
 ↪\n{missing_in_test_df}\n")

# Step 4: Identify additional columns present in merged_test_df but not in␣
 ↪oversampled_df
additional_in_test_df = list(merged_test_columns_set - oversampled_columns_set)
```

```
print(f"Additional Columns in merged_test_df not in oversampled_df:
↪\n{additional_in_test_df}\n")

# Step 5: Filter out columns that cannot be used in training due to missing␣
↪target variables like 'sii'
# Remove columns such as 'sii' which are only relevant during training, not in␣
↪the test set
excluded_columns = ['sii', 'PCIAT-PCIAT_Total', 'PCIAT_Cluster']

# Determine final columns for training based on shared columns, excluding any␣
↪that are in the excluded list
final_student_training_columns = [col for col in shared_columns if col not in␣
↪excluded_columns]
print(f"Final Columns to be used for training the student model:
↪\n{final_student_training_columns}\n")

# Step 6: (Optional) Create placeholders for missing columns if necessary
# For columns that are required but missing, we can fill with zeros or mean␣
↪values (if numeric), or a special placeholder for categoricals
for col in missing_in_test_df:
    if col not in excluded_columns:  # Avoid creating placeholders for excluded␣
↪columns
        # Create a placeholder column (e.g., filling with zeros)
        if col in oversampled_df.select_dtypes(include=[np.number]).columns:
            merged_test_df[col] = 0  # Fill missing numeric columns with zero
        else:
            merged_test_df[col] = 'missing'  # Fill missing categorical columns␣
↪with a placeholder
        print(f"Created placeholder for missing column: {col}")

# Step 7: Select the final subset of columns from merged_test_df for training␣
↪the student model
final_student_model_df = merged_test_df[final_student_training_columns]
print(f"Shape of final dataframe for student model training:␣
↪{final_student_model_df.shape}")
```

```
Shared Columns between oversampled_df and merged_test_df:
['BIA-BIA_SMM', 'BIA-BIA_FMI', 'Fitness_Endurance-Season', 'Y_mean', 'Z_min',
'CGAS-Season', 'BIA-BIA_Fat', 'FGC-FGC_GSD', 'X_max', 'battery_voltage_mean',
'PreInt_EduHx-computerinternet_hoursday', 'quarter_mean', 'fourier_Z_real',
'Fitness_Endurance-Time_Mins', 'BIA-BIA_DEE', 'X_min', 'Y_min', 'Age_Group',
'BIA-BIA_ECW', 'fourier_Y_imag', 'FGC-FGC_GSND_Zone', 'anglez_std', 'enmo_std',
'battery_voltage_std', 'X_mean', 'BIA-BIA_Activity_Level_num', 'FGC-
FGC_SRR_Zone', 'CGAS-CGAS_Score', 'Y_max', 'SDS-SDS_Total_Raw', 'PAQ_C-Season',
'FGC-FGC_TL', 'BIA-BIA_BMC', 'fourier_Z_imag', 'PAQ_C-PAQ_C_Total', 'FGC-
FGC_CU_Zone', 'Basic_Demos-Age', 'FGC-FGC_GSD_Zone', 'FGC-FGC_TL_Zone', 'non-
wear_flag_mean', 'Physical-Waist_Circumference', 'fourier_X_imag', 'BIA-
```

```
BIA_BMI', 'actigraphy_present', 'FGC-FGC_CU', 'fourier_Y_real', 'BIA-BIA_TBW',
'activity_ratio_day_night', 'FGC-FGC_PU_Zone', 'Physical-Season', 'Physical-
Weight', 'Physical-HeartRate', 'activity_during_night', 'BIA-BIA_BMR',
'fourier_X_real', 'light_activity_proportion', 'BIA-BIA_FFMI',
'moderate_activity_proportion', 'Z_std', 'Z_mean', 'anglez_mean', 'Physical-
Diastolic_BP', 'BIA-BIA_Frame_num', 'Physical-BMI', 'sedentary_proportion',
'PAQ_A-Season', 'FGC-FGC_SRR', 'weekday_mean', 'Basic_Demos-Enroll_Season',
'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'BIA-BIA_ICW', 'Z_max',
'activity_during_day', 'SDS-SDS_Total_T', 'time_of_day_mean', 'PAQ_A-
PAQ_A_Total', 'FGC-FGC_GSND', 'Physical-Systolic_BP', 'BIA-BIA_LDM', 'FGC-
Season', 'id', 'relative_date_PCIAT_mean', 'BIA-Season', 'BIA-BIA_FFM',
'non_wear_proportion', 'Fitness_Endurance-Time_Sec', 'light_std', 'Physical-
Height', 'PreInt_EduHx-Season', 'enmo_mean', 'Fitness_Endurance-Max_Stage',
'light_mean', 'Basic_Demos-Sex', 'Y_std', 'X_std', 'BIA-BIA_LST', 'FGC-FGC_PU',
'vigorous_activity_proportion', 'SDS-Season']

Columns in oversampled_df but missing in merged_test_df:
['PCIAT_Cluster', 'CGAS-CGAS_Score_x_SDS-SDS_Total_Raw', 'CGAS-
CGAS_Score_x_Physical-BMI', 'sii', 'PCIAT-PCIAT_Total']

Additional Columns in merged_test_df not in oversampled_df:
['Physical-Waist_Circumference_missing', 'PAQ_A-PAQ_A_Total_missing', 'PAQ_A-
Season_missing']

Final Columns to be used for training the student model:
['BIA-BIA_SMM', 'BIA-BIA_FMI', 'Fitness_Endurance-Season', 'Y_mean', 'Z_min',
'CGAS-Season', 'BIA-BIA_Fat', 'FGC-FGC_GSD', 'X_max', 'battery_voltage_mean',
'PreInt_EduHx-computerinternet_hoursday', 'quarter_mean', 'fourier_Z_real',
'Fitness_Endurance-Time_Mins', 'BIA-BIA_DEE', 'X_min', 'Y_min', 'Age_Group',
'BIA-BIA_ECW', 'fourier_Y_imag', 'FGC-FGC_GSND_Zone', 'anglez_std', 'enmo_std',
'battery_voltage_std', 'X_mean', 'BIA-BIA_Activity_Level_num', 'FGC-
FGC_SRR_Zone', 'CGAS-CGAS_Score', 'Y_max', 'SDS-SDS_Total_Raw', 'PAQ_C-Season',
'FGC-FGC_TL', 'BIA-BIA_BMC', 'fourier_Z_imag', 'PAQ_C-PAQ_C_Total', 'FGC-
FGC_CU_Zone', 'Basic_Demos-Age', 'FGC-FGC_GSD_Zone', 'FGC-FGC_TL_Zone', 'non-
wear_flag_mean', 'Physical-Waist_Circumference', 'fourier_X_imag', 'BIA-
BIA_BMI', 'actigraphy_present', 'FGC-FGC_CU', 'fourier_Y_real', 'BIA-BIA_TBW',
'activity_ratio_day_night', 'FGC-FGC_PU_Zone', 'Physical-Season', 'Physical-
Weight', 'Physical-HeartRate', 'activity_during_night', 'BIA-BIA_BMR',
'fourier_X_real', 'light_activity_proportion', 'BIA-BIA_FFMI',
'moderate_activity_proportion', 'Z_std', 'Z_mean', 'anglez_mean', 'Physical-
Diastolic_BP', 'BIA-BIA_Frame_num', 'Physical-BMI', 'sedentary_proportion',
'PAQ_A-Season', 'FGC-FGC_SRR', 'weekday_mean', 'Basic_Demos-Enroll_Season',
'FGC-FGC_SRL', 'FGC-FGC_SRL_Zone', 'BIA-BIA_ICW', 'Z_max',
'activity_during_day', 'SDS-SDS_Total_T', 'time_of_day_mean', 'PAQ_A-
PAQ_A_Total', 'FGC-FGC_GSND', 'Physical-Systolic_BP', 'BIA-BIA_LDM', 'FGC-
Season', 'id', 'relative_date_PCIAT_mean', 'BIA-Season', 'BIA-BIA_FFM',
'non_wear_proportion', 'Fitness_Endurance-Time_Sec', 'light_std', 'Physical-
Height', 'PreInt_EduHx-Season', 'enmo_mean', 'Fitness_Endurance-Max_Stage',
```

```
'light_mean', 'Basic_Demos-Sex', 'Y_std', 'X_std', 'BIA-BIA_LST', 'FGC-FGC_PU',
'vigorous_activity_proportion', 'SDS-Season']

Created placeholder for missing column: CGAS-CGAS_Score_x_SDS-SDS_Total_Raw
Created placeholder for missing column: CGAS-CGAS_Score_x_Physical-BMI
Shape of final dataframe for student model training: (6, 100)
```

[76]:

```python
# Import necessary libraries for feature selection
from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import SelectKBest, f_classif
import pandas as pd

# Given: Excluded actigraphy-related columns
included_columns =['FGC-FGC_PU', 'Z_std', 'BIA-BIA_DEE', 'SDS-Season',␣
 ↪'light_mean',
                    'BIA-BIA_BMR', 'BIA-BIA_ICW', 'FGC-FGC_GSD_Zone',␣
 ↪'Physical-BMI',
                    'Y_min', 'sedentary_proportion', 'time_of_day_mean',␣
 ↪'BIA-BIA_FMI',
                    'PAQ_C-PAQ_C_Total', 'Physical-Weight',␣
 ↪'Fitness_Endurance-Max_Stage',
                    'FGC-FGC_TL', 'PAQ_A-PAQ_A_Total', 'X_mean',␣
 ↪'vigorous_activity_proportion',
                    'BIA-BIA_ECW', 'non-wear_flag_mean', 'actigraphy_present',␣
 ↪'FGC-FGC_SRR_Zone',
                    'fourier_X_imag', 'Y_std', 'SDS-SDS_Total_T',␣
 ↪'Basic_Demos-Enroll_Season',
                    'BIA-BIA_LDM', 'Fitness_Endurance-Time_Mins',␣
 ↪'Physical-Systolic_BP', 'Basic_Demos-Sex',
                    'FGC-FGC_SRL', 'Y_max', 'Z_max', 'fourier_Y_real',␣
 ↪'FGC-FGC_CU_Zone', 'PAQ_C-Season',
                    'non_wear_proportion', 'Age_Group', 'FGC-FGC_GSND_Zone',␣
 ↪'quarter_mean', 'light_std',
                    'X_max', 'CGAS-Season', 'FGC-FGC_SRL_Zone', 'anglez_mean',␣
 ↪'anglez_std', 'fourier_Z_real',
                    'enmo_std', 'FGC-FGC_GSD', 'X_min', 'BIA-Season',␣
 ↪'FGC-FGC_PU_Zone', 'X_std', 'BIA-BIA_TBW',
                    'Physical-HeartRate', 'FGC-Season',␣
 ↪'Fitness_Endurance-Time_Sec', 'activity_during_day',
                    'Y_mean', 'BIA-BIA_Activity_Level_num', 'FGC-FGC_SRR',␣
 ↪'Z_min', 'PreInt_EduHx-Season',
                    'fourier_X_real', 'BIA-BIA_FFMI', 'BIA-BIA_FFM',␣
 ↪'BIA-BIA_SMM', 'activity_during_night',
                    'PAQ_A-Season', 'fourier_Z_imag', 'Physical-Diastolic_BP',␣
 ↪'battery_voltage_mean',
```

```
                    'moderate_activity_proportion', 'Physical-Season', 'battery_voltage_std', 'BIA-BIA_Fat',
                    'BIA-BIA_LST', 'PreInt_EduHx-computerinternet_hoursday', 'enmo_mean', 'FGC-FGC_TL_Zone',
                    'BIA-BIA_Frame_num', 'weekday_mean', 'Fitness_Endurance-Season', 'Physical-Height',
                    'FGC-FGC_GSND', 'fourier_Y_imag', 'BIA-BIA_BMI', 'SDS-SDS_Total_Raw',
                    'Physical-Waist_Circumference', 'CGAS-CGAS_Score', 'relative_date_PCIAT_mean',
                    'FGC-FGC_CU', 'light_activity_proportion', 'activity_ratio_day_night', 'Z_mean',
                    'BIA-BIA_BMC', 'Basic_Demos-Age']
# Load the oversampled_df DataFrame
# Make sure to replace 'oversampled_df' with your actual DataFrame variable
# Note: Exclude 'sii' (target) and 'id' (identifier) from features
#features_to_exclude = ['sii', 'id', 'actigraphy_present'] + actigraphy_columns
features_for_selection = oversampled_df[included_columns]
print(list(features_for_selection.columns))
target = oversampled_df['sii']

# Perform Tree-based Feature Selection using RandomForestClassifier
forest_model = RandomForestClassifier(random_state=42)
forest_model.fit(features_for_selection, target)

# Extract feature importances and sort them
feature_importances = pd.Series(forest_model.feature_importances_, index=features_for_selection.columns)
sorted_importances = feature_importances.sort_values(ascending=False)

# Perform Univariate Feature Selection using SelectKBest with ANOVA F-test
k_best_selector = SelectKBest(score_func=f_classif, k=10)  # Select top 10 features
k_best_selector.fit(features_for_selection, target)

# Extract selected features
k_best_features = features_for_selection.columns[k_best_selector.get_support()]

# Store the results in a DataFrame for display
selected_features_df = pd.DataFrame({
    'Feature': features_for_selection.columns,
    'Importance_Score': feature_importances,
    'Selected_By_KBest': k_best_selector.get_support()
})

# Sort the DataFrame by importance score
```

```python
selected_features_df = selected_features_df.sort_values(by='Importance_Score',␣
    ↪ascending=False)

# Display the sorted feature importances and selected features using pandas
print("Feature Selection Results:")
print(selected_features_df)
```

```
['FGC-FGC_PU', 'Z_std', 'BIA-BIA_DEE', 'SDS-Season', 'light_mean', 'BIA-
BIA_BMR', 'BIA-BIA_ICW', 'FGC-FGC_GSD_Zone', 'Physical-BMI', 'Y_min',
'sedentary_proportion', 'time_of_day_mean', 'BIA-BIA_FMI', 'PAQ_C-PAQ_C_Total',
'Physical-Weight', 'Fitness_Endurance-Max_Stage', 'FGC-FGC_TL', 'PAQ_A-
PAQ_A_Total', 'X_mean', 'vigorous_activity_proportion', 'BIA-BIA_ECW', 'non-
wear_flag_mean', 'actigraphy_present', 'FGC-FGC_SRR_Zone', 'fourier_X_imag',
'Y_std', 'SDS-SDS_Total_T', 'Basic_Demos-Enroll_Season', 'BIA-BIA_LDM',
'Fitness_Endurance-Time_Mins', 'Physical-Systolic_BP', 'Basic_Demos-Sex', 'FGC-
FGC_SRL', 'Y_max', 'Z_max', 'fourier_Y_real', 'FGC-FGC_CU_Zone', 'PAQ_C-Season',
'non_wear_proportion', 'Age_Group', 'FGC-FGC_GSND_Zone', 'quarter_mean',
'light_std', 'X_max', 'CGAS-Season', 'FGC-FGC_SRL_Zone', 'anglez_mean',
'anglez_std', 'fourier_Z_real', 'enmo_std', 'FGC-FGC_GSD', 'X_min', 'BIA-
Season', 'FGC-FGC_PU_Zone', 'X_std', 'BIA-BIA_TBW', 'Physical-HeartRate', 'FGC-
Season', 'Fitness_Endurance-Time_Sec', 'activity_during_day', 'Y_mean', 'BIA-
BIA_Activity_Level_num', 'FGC-FGC_SRR', 'Z_min', 'PreInt_EduHx-Season',
'fourier_X_real', 'BIA-BIA_FFMI', 'BIA-BIA_FFM', 'BIA-BIA_SMM',
'activity_during_night', 'PAQ_A-Season', 'fourier_Z_imag', 'Physical-
Diastolic_BP', 'battery_voltage_mean', 'moderate_activity_proportion',
'Physical-Season', 'battery_voltage_std', 'BIA-BIA_Fat', 'BIA-BIA_LST',
'PreInt_EduHx-computerinternet_hoursday', 'enmo_mean', 'FGC-FGC_TL_Zone', 'BIA-
BIA_Frame_num', 'weekday_mean', 'Fitness_Endurance-Season', 'Physical-Height',
'FGC-FGC_GSND', 'fourier_Y_imag', 'BIA-BIA_BMI', 'SDS-SDS_Total_Raw', 'Physical-
Waist_Circumference', 'CGAS-CGAS_Score', 'relative_date_PCIAT_mean', 'FGC-
FGC_CU', 'light_activity_proportion', 'activity_ratio_day_night', 'Z_mean',
'BIA-BIA_BMC', 'Basic_Demos-Age']
Feature Selection Results:
                                                                       Feature
\
Basic_Demos-Age                                                Basic_Demos-Age
BIA-BIA_LDM                                                        BIA-BIA_LDM
Age_Group                                                            Age_Group
PreInt_EduHx-computerinternet_hoursday  PreInt_EduHx-computerinternet_hoursday
Physical-Weight                                                Physical-Weight
…                                                                            …
fourier_X_imag                                                  fourier_X_imag
non-wear_flag_mean                                          non-wear_flag_mean
actigraphy_present                                          actigraphy_present
activity_ratio_day_night                              activity_ratio_day_night
activity_during_day                                        activity_during_day

                                        Importance_Score  Selected_By_KBest
```

```
Basic_Demos-Age                                 0.045561             True
BIA-BIA_LDM                                     0.034236            False
Age_Group                                       0.032241             True
PreInt_EduHx-computerinternet_hoursday          0.025768             True
Physical-Weight                                 0.025649             True
…                                                  …              …
fourier_X_imag                                  0.002257            False
non-wear_flag_mean                              0.002081            False
actigraphy_present                              0.000332            False
activity_ratio_day_night                        0.000000            False
activity_during_day                             0.000000            False

[99 rows x 3 columns]
```

```
/opt/conda/lib/python3.10/site-
packages/sklearn/feature_selection/_univariate_selection.py:112: UserWarning:
Features [59 95] are constant.
  warnings.warn("Features %s are constant." % constant_features_idx,
UserWarning)
/opt/conda/lib/python3.10/site-
packages/sklearn/feature_selection/_univariate_selection.py:113: RuntimeWarning:
invalid value encountered in divide
  f = msb / msw
```

[77]:
```
selected_features_df.head()
```

[77]:
```
                                                                        Feature  \
Basic_Demos-Age                                                 Basic_Demos-Age
BIA-BIA_LDM                                                         BIA-BIA_LDM
Age_Group                                                             Age_Group
PreInt_EduHx-computerinternet_hoursday  PreInt_EduHx-computerinternet_hoursday
Physical-Weight                                                 Physical-Weight

                                        Importance_Score  Selected_By_KBest
Basic_Demos-Age                                 0.045561               True
BIA-BIA_LDM                                     0.034236              False
Age_Group                                       0.032241               True
PreInt_EduHx-computerinternet_hoursday          0.025768               True
Physical-Weight                                 0.025649               True
```

[82]:
```
selected_features = features_for_selection.loc[:, k_best_selector.get_support()]
selected_features
```

[82]:
```
      Physical-BMI  Physical-Weight  Age_Group  PAQ_A-Season  \
0        16.877316        50.800000   0.443345      0.480263
1        16.648696        75.600000   0.362030      0.480263
2        18.292347        81.600000   0.362030      0.480263
```

```
3        22.279952       112.200000    0.855589      0.480263
4        18.269958        83.565122    0.855589      0.480263
…              …               …             …             …
8207     21.591746       122.849976    0.965907      0.908433
8208     24.131142       159.969205    0.855589      0.893266
8209     23.879108       158.833890    0.865555      1.017185
8210     24.046040       133.743377    0.855589      0.480263
8211     19.869775       101.793592    0.855589      0.480263

      PreInt_EduHx-computerinternet_hoursday  Physical-Height  FGC-FGC_GSND  \
0                                   3.000000        46.000000     12.457159
1                                   2.000000        56.500000     10.200000
2                                   0.000000        56.000000     15.489000
3                                   0.000000        59.500000     16.500000
4                                   0.000000        55.568097     19.270922
…                                        …               …            …
8207                                2.000000        63.155757     24.573501
8208                                2.000000        68.269205     37.768842
8209                                2.084328        68.373508     39.950756
8210                                0.767371        61.822798     23.801864
8211                                2.753714        59.891713     16.584285

      BIA-BIA_BMI  FGC-FGC_CU  Basic_Demos-Age
0       16.879200    0.000000         5.000000
1       18.540298   20.000000        10.000000
2       18.294300   18.000000         9.000000
3       30.186500   12.000000        13.000000
4       18.540307   10.701686        11.000000
…            …           …              …
8207    21.644003   35.757071        16.800417
8208    24.184481   14.154304        14.538411
8209    23.888294   11.517900        15.168656
8210    24.336307   16.272720        13.383686
8211    19.771615   27.462860        12.507428

[8212 rows x 10 columns]
```

[81]:
```
from tensorflow.keras.layers import Dense, Dropout, BatchNormalization, Input,
  ↪LeakyReLU, ELU  # Include ELU here
from tensorflow.keras.models import Model
from tensorflow.keras.losses import CategoricalCrossentropy, KLDivergence
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score, f1_score
import numpy as np
```

```
# Extract the selected features (using SelectKBest results)
selected_features = features_for_selection.loc[:, k_best_selector.get_support()]

print(selected_features.columns)
# Define input features and target
X = oversampled_df[selected_features.columns]  # Use columns of selected␣
↪features (non-actigraphy features)
y = oversampled_df['sii']

# Split the data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)

# One-hot encode the labels for training
y_train_onehot = tf.keras.utils.to_categorical(y_train, num_classes=4)
y_val_onehot = tf.keras.utils.to_categorical(y_val, num_classes=4)

# Precompute teacher model predictions using the full feature set (102 features)
teacher_full_X = oversampled_df.drop(columns=['sii', 'id',␣
↪'actigraphy_present'])  # Use all features except target and ID
teacher_full_preds = teacher_model_A.predict(teacher_full_X)  # Get softmax␣
↪outputs

# Match the teacher predictions to the training and validation sets based on␣
↪indices
teacher_train_preds = teacher_full_preds[y_train.index]  # Teacher predictions␣
↪for training data
teacher_val_preds = teacher_full_preds[y_val.index]      # Teacher predictions␣
↪for validation data

# Create a TensorFlow variable to hold the teacher predictions for the current␣
↪batch
teacher_train_preds_tensor = tf.convert_to_tensor(teacher_train_preds, dtype=tf.
↪float32)

# Custom loss function combining categorical cross-entropy and KL Divergence
def custom_loss(teacher_preds_tensor):
    @tf.function
    def loss_fn(y_true, y_pred):
        # Get the current batch size
        batch_size = tf.shape(y_pred)[0]

        # Select the corresponding teacher predictions for the current batch
        teacher_preds_batch = teacher_preds_tensor[:batch_size]

        # Calculate categorical cross-entropy loss
```

```
        cce_loss = CategoricalCrossentropy()(y_true, y_pred)

        # Calculate KL Divergence loss between student predictions and teacher␣
↪outputs
        kl_loss = KLDivergence()(teacher_preds_batch, y_pred)

        # Combine the two losses with equal weight (0.5 each)
        return 0.5 * cce_loss + 0.5 * kl_loss

    return loss_fn

# Create the student model
def create_student_model(input_shape):
    inputs = Input(shape=(input_shape,))
    x = Dense(64, activation='linear', kernel_regularizer=tf.keras.regularizers.
↪l2(0.002))(inputs)
    #x = LeakyReLU(alpha=0.1)(x)
    x = BatchNormalization()(x)
    x = Dropout(0.4)(x)

    x = Dense(32, activation='linear', kernel_regularizer=tf.keras.regularizers.
↪l2(0.002))(x)
    x = LeakyReLU(alpha=0.1)(x)
    #x = BatchNormalization()(x)
    #x = Dropout(0.6)(x)

    # Additional layer for better capacity
    x = Dense(16, activation='linear', kernel_regularizer=tf.keras.regularizers.
↪l2(0.002))(x)
    x = ELU(alpha=1.0)(x)
    #x = BatchNormalization()(x)
    #x = Dropout(0.5)(x)

    x = Dense(8, activation='linear', kernel_regularizer=tf.keras.regularizers.
↪l2(0.002))(x)
    x = ELU(alpha=1.0)(x)

    outputs = Dense(4, activation='softmax')(x)

    model = Model(inputs=inputs, outputs=outputs)

    # Compile the model with the custom loss function and the specified␣
↪optimizer
    model.compile(optimizer=Adam(learning_rate=0.001),␣
↪loss=custom_loss(teacher_train_preds_tensor), metrics=['accuracy'])
```

```python
    return model

# Create the student model using the dynamically determined input shape
student_model = create_student_model(X_train.shape[1])

improved_class_weight_dict = {0: 1.0, 1: 1.0, 2: 1.2, 3: 1.0}

# Early stopping callback
early_stopping = EarlyStopping(monitor='val_loss', patience=6, restore_best_weights=True)

# Train the student model
history = student_model.fit(
    X_train, y_train_onehot,
    validation_data=(X_val, y_val_onehot),
    epochs=50,
    batch_size=64,
    callbacks=[early_stopping],
    class_weight=improved_class_weight_dict,
    verbose=1
)

# Evaluate the student model on the validation set
student_val_predictions = student_model.predict(X_val)
student_val_pred_labels = np.argmax(student_val_predictions, axis=1)

# Calculate evaluation metrics
student_accuracy = accuracy_score(y_val, student_val_pred_labels)
student_f1 = f1_score(y_val, student_val_pred_labels, average='weighted')

print(f"Student Model Accuracy: {student_accuracy:.4f}")
print(f"Student Model F1-Score: {student_f1:.4f}")
print(f"Classification Report:\n{classification_report(y_val, student_val_pred_labels)}")

# Implement Quadratic Weighted Kappa Evaluation
def quadratic_weighted_kappa(y_true, y_pred, N=4):
    # Create matrices O, W, and E
    O = np.zeros((N, N))
    for a, p in zip(y_true, y_pred):
        O[a][p] += 1

    W = np.zeros((N, N))
    for i in range(N):
        for j in range(N):
            W[i][j] = ((i - j) ** 2) / (N - 1) ** 2
```

```python
    actual_hist = np.histogram(y_true, bins=np.arange(N + 1))[0]
    pred_hist = np.histogram(y_pred, bins=np.arange(N + 1))[0]
    E = np.outer(actual_hist, pred_hist) / np.sum(actual_hist)

    # Calculate QWK
    kappa = 1 - (np.sum(W * O) / np.sum(W * E))
    return kappa

# Compute QWK for the student model
student_qwk = quadratic_weighted_kappa(y_val, student_val_pred_labels)
print(f"Student Model Quadratic Weighted Kappa (QWK): {student_qwk:.4f}")
```

```
Index(['Physical-BMI', 'Physical-Weight', 'Age_Group', 'PAQ_A-Season',
       'PreInt_EduHx-computerinternet_hoursday', 'Physical-Height',
       'FGC-FGC_GSND', 'BIA-BIA_BMI', 'FGC-FGC_CU', 'Basic_Demos-Age'],
      dtype='object')
257/257              0s 1ms/step
Epoch 1/50
```

```
/opt/conda/lib/python3.10/site-
packages/keras/src/layers/activations/leaky_relu.py:41: UserWarning: Argument
`alpha` is deprecated. Use `negative_slope` instead.
  warnings.warn(
```

```
103/103              2s 5ms/step -
accuracy: 0.3437 - loss: 1.4683 - val_accuracy: 0.3488 - val_loss: 1.3621
Epoch 2/50
103/103              0s 2ms/step -
accuracy: 0.4743 - loss: 1.2830 - val_accuracy: 0.4090 - val_loss: 1.2159
Epoch 3/50
103/103              0s 2ms/step -
accuracy: 0.4867 - loss: 1.2510 - val_accuracy: 0.4577 - val_loss: 1.1786
Epoch 4/50
103/103              0s 2ms/step -
accuracy: 0.5152 - loss: 1.2273 - val_accuracy: 0.5362 - val_loss: 1.1637
Epoch 5/50
103/103              0s 2ms/step -
accuracy: 0.5296 - loss: 1.2100 - val_accuracy: 0.4741 - val_loss: 1.1449
Epoch 6/50
103/103              0s 2ms/step -
accuracy: 0.5491 - loss: 1.1971 - val_accuracy: 0.4425 - val_loss: 1.1373
Epoch 7/50
103/103              0s 2ms/step -
accuracy: 0.5445 - loss: 1.1842 - val_accuracy: 0.5727 - val_loss: 1.1212
Epoch 8/50
103/103              0s 2ms/step -
accuracy: 0.5573 - loss: 1.1675 - val_accuracy: 0.5417 - val_loss: 1.1136
Epoch 9/50
103/103              0s 2ms/step -
```

```
accuracy: 0.5690 - loss: 1.1575 - val_accuracy: 0.5265 - val_loss: 1.1198
Epoch 10/50
103/103            0s 2ms/step -
accuracy: 0.5644 - loss: 1.1515 - val_accuracy: 0.4760 - val_loss: 1.1160
Epoch 11/50
103/103            0s 2ms/step -
accuracy: 0.5778 - loss: 1.1421 - val_accuracy: 0.5575 - val_loss: 1.0925
Epoch 12/50
103/103            0s 2ms/step -
accuracy: 0.5773 - loss: 1.1288 - val_accuracy: 0.5703 - val_loss: 1.0971
Epoch 13/50
103/103            0s 2ms/step -
accuracy: 0.5818 - loss: 1.1259 - val_accuracy: 0.4632 - val_loss: 1.0959
Epoch 14/50
103/103            0s 2ms/step -
accuracy: 0.5789 - loss: 1.1276 - val_accuracy: 0.4948 - val_loss: 1.0995
Epoch 15/50
103/103            0s 2ms/step -
accuracy: 0.5880 - loss: 1.1257 - val_accuracy: 0.5721 - val_loss: 1.0725
Epoch 16/50
103/103            0s 2ms/step -
accuracy: 0.5907 - loss: 1.1176 - val_accuracy: 0.3743 - val_loss: 1.1253
Epoch 17/50
103/103            0s 2ms/step -
accuracy: 0.5964 - loss: 1.1171 - val_accuracy: 0.5405 - val_loss: 1.0851
Epoch 18/50
103/103            0s 2ms/step -
accuracy: 0.5943 - loss: 1.1162 - val_accuracy: 0.4680 - val_loss: 1.1174
Epoch 19/50
103/103            0s 2ms/step -
accuracy: 0.5717 - loss: 1.1184 - val_accuracy: 0.5697 - val_loss: 1.0663
Epoch 20/50
103/103            0s 2ms/step -
accuracy: 0.5796 - loss: 1.1154 - val_accuracy: 0.5161 - val_loss: 1.1162
Epoch 21/50
103/103            0s 2ms/step -
accuracy: 0.5939 - loss: 1.1137 - val_accuracy: 0.5612 - val_loss: 1.0818
Epoch 22/50
103/103            0s 2ms/step -
accuracy: 0.5962 - loss: 1.1084 - val_accuracy: 0.6403 - val_loss: 1.0640
Epoch 23/50
103/103            0s 2ms/step -
accuracy: 0.5893 - loss: 1.1123 - val_accuracy: 0.5770 - val_loss: 1.0610
Epoch 24/50
103/103            0s 2ms/step -
accuracy: 0.5934 - loss: 1.1090 - val_accuracy: 0.5247 - val_loss: 1.0870
Epoch 25/50
103/103            0s 2ms/step -
```

```
accuracy: 0.5992 - loss: 1.1081 - val_accuracy: 0.5222 - val_loss: 1.0904
Epoch 26/50
103/103            0s 2ms/step -
accuracy: 0.5926 - loss: 1.1090 - val_accuracy: 0.4857 - val_loss: 1.0692
Epoch 27/50
103/103            0s 2ms/step -
accuracy: 0.5957 - loss: 1.1108 - val_accuracy: 0.5654 - val_loss: 1.0614
Epoch 28/50
103/103            0s 2ms/step -
accuracy: 0.5954 - loss: 1.1112 - val_accuracy: 0.4997 - val_loss: 1.0842
Epoch 29/50
103/103            0s 2ms/step -
accuracy: 0.5899 - loss: 1.1117 - val_accuracy: 0.4334 - val_loss: 1.1042
52/52            0s 2ms/step
Student Model Accuracy: 0.5770
Student Model F1-Score: 0.5355
Classification Report:
              precision    recall  f1-score   support

           0       0.52      0.82      0.64       418
           1       0.45      0.17      0.25       407
           2       0.62      0.40      0.49       416
           3       0.66      0.91      0.77       402

    accuracy                           0.58      1643
   macro avg       0.56      0.58      0.54      1643
weighted avg       0.56      0.58      0.54      1643

Student Model Quadratic Weighted Kappa (QWK): 0.7245
```

[79]:
```
oversampled_df['sii'].value_counts()
```

[79]:
```
sii
2    2053
0    2053
1    2053
3    2053
Name: count, dtype: int64
```

[94]:
```
merged_test_df.head(20)
```

[94]:
```
         id Basic_Demos-Enroll_Season  Basic_Demos-Age  Basic_Demos-Sex  \
0  00008ff9                      Fall              5.0              0.0
1  00105258                    Summer             10.0              1.0
2  00115b9f                    Winter              9.0              0.0
3  001f3379                    Spring             13.0              1.0
4  00abe655                      Fall             11.0              0.0
5  00d56d4b                    Spring              5.0              1.0
```

```
  CGAS-Season  CGAS-CGAS_Score Physical-Season  Physical-BMI  Physical-Height  \
0      Winter             51.0            Fall     16.877316        46.000000
1        Fall             71.0            Fall     16.648696        56.500000
2        Fall             71.0          Summer     18.292347        56.000000
3      Winter             50.0          Summer     22.279952        59.500000
4      Summer             66.0            Fall     17.864055        53.089389
5      Summer             80.0          Spring     17.284504        44.000000

   Physical-Weight  Physical-Waist_Circumference  Physical-Diastolic_BP  \
0        50.800000                           NaN              76.659994
1        75.600000                           NaN              65.000000
2        81.600000                           NaN              60.000000
3       112.200000                           NaN              60.000000
4        73.360103                           NaN              64.906547
5        47.600000                           NaN              61.000000

   Physical-HeartRate  Physical-Systolic_BP Fitness_Endurance-Season  \
0           42.625084             95.537106                     Fall
1           94.000000            117.000000                     Fall
2           97.000000            117.000000                   Summer
3           73.000000            102.000000                     Fall
4           79.417997            109.671015                     Fall
5           76.000000            109.000000                   Spring

   Fitness_Endurance-Max_Stage  Fitness_Endurance-Time_Mins  \
0                     0.108192                    -2.783656
1                     5.000000                     7.000000
2                     6.000000                     9.000000
3                     8.161510                    13.323058
4                     4.434398                     5.868766
5                     2.625106                     2.250025

   Fitness_Endurance-Time_Sec FGC-Season  FGC-FGC_CU  FGC-FGC_CU_Zone  \
0                   13.432611       Fall    0.000000          0.00000
1                   33.000000       Fall   20.000000          1.00000
2                   37.000000     Summer   18.000000          1.00000
3                   45.646307     Summer   12.000000          0.00000
4                   30.737339     Winter   11.771994          0.48317
5                   23.498683     Spring    0.000000          0.00000

   FGC-FGC_GSND  FGC-FGC_GSND_Zone  FGC-FGC_GSD  FGC-FGC_GSD_Zone  FGC-FGC_PU  \
0      3.706899          -0.030646    11.401920               2.0    0.000000
1     10.200000           1.000000    14.700000               2.0    7.000000
2     11.694644           1.237246    15.459184               2.0    5.000000
3     16.500000           2.000000    17.900000               2.0    6.000000
4      9.702796           0.921081    14.447450               2.0    4.413705
```

```
5     6.372775           0.392507    12.755998              2.0    0.000000

   FGC-FGC_PU_Zone  FGC-FGC_SRL  FGC-FGC_SRL_Zone  FGC-FGC_SRR  \
0         0.000000     7.000000          0.000000     6.000000
1         1.000000    10.000000          1.000000    10.000000
2         0.000000     7.000000          0.000000     7.000000
3         0.000000    10.000000          1.000000    11.000000
4         0.458711     9.593142          0.832517     9.539816
5         0.000000    10.500000          1.000000    10.000000

   FGC-FGC_SRR_Zone  FGC-FGC_TL  FGC-FGC_TL_Zone BIA-Season  \
0          0.000000     6.00000         1.000000       Fall
1          1.000000     5.00000         0.000000     Summer
2          0.000000     7.00000         1.000000     Summer
3          1.000000     8.00000         0.000000     Summer
4          0.834634     6.11333         0.357534     Summer
5          1.000000     7.00000         1.000000     Summer

   BIA-BIA_Activity_Level_num  BIA-BIA_BMC  BIA-BIA_BMI  BIA-BIA_BMR  \
0                    2.000000     2.668550    16.879200   932.498000
1                    2.940449     3.641048    17.160848  1087.854533
2                    3.000000     3.841910    18.294300  1131.430000
3                    2.000000     4.330360    30.186500  1330.970000
4                    2.634960     3.539088    18.782182  1087.500178
5                    2.932333     3.170967    13.459508   968.943255

   BIA-BIA_DEE  BIA-BIA_ECW  BIA-BIA_FFM  BIA-BIA_FFMI  BIA-BIA_FMI  \
0  1492.000000     8.255980    41.586200     13.817700     3.061430
1  1858.417991    13.394408    58.139277     13.825440     3.303771
2  1923.440000    15.592500    62.775700     14.074000     4.220330
3  1996.450000    30.212400    84.028500     16.687700    13.498800
4  1803.297738    14.468455    58.086989     14.178944     4.520419
5  1707.997278     6.780333    45.387093     12.955624     0.141155

   BIA-BIA_Fat  BIA-BIA_Frame_num  BIA-BIA_ICW  BIA-BIA_LDM  BIA-BIA_LST  \
0     9.213770           1.000000    24.434900     8.895360    38.917700
1    13.482399           1.863178    29.381756    15.362565    54.493534
2    18.824300           2.000000    30.404100    16.779000    58.933800
3    67.971500           2.000000    32.914100    20.902000    79.698200
4    19.377546           1.710098    28.844178    14.773667    54.540391
5    -4.415591           1.566602    26.842912    11.761994    42.191669

   BIA-BIA_SMM  BIA-BIA_TBW  PAQ_A-Season  PAQ_A-PAQ_A_Total PAQ_C-Season  \
0    19.541300    32.690900           NaN                NaN       Winter
1    24.801015    42.771962           NaN                NaN       Summer
2    26.479800    45.996600           NaN                NaN       Winter
3    35.380400    63.126500           NaN                NaN       Spring
```

```
4   25.072255   43.300657           NaN                NaN      Winter
5   20.013959   33.568832           NaN                NaN      Winter

   PAQ_C-PAQ_C_Total SDS-Season  SDS-SDS_Total_Raw  SDS-SDS_Total_T  \
0          -2.892886     Summer          65.892149        89.323116
1           2.170000       Fall          38.000000        54.000000
2           2.451000     Summer          31.000000        45.000000
3           4.110000     Summer          40.000000        56.000000
4           1.100000     Winter          42.000000        59.000000
5          -1.349256     Spring          37.000000        53.000000

  PreInt_EduHx-Season  PreInt_EduHx-computerinternet_hoursday Age_Group  \
0                Fall                                     3.0      5-10
1              Summer                                     2.0      5-10
2              Winter                                     0.0      5-10
3              Spring                                     0.0     11-15
4                Fall                                     0.0     11-15
5              Spring                                     0.0      5-10

   Physical-Waist_Circumference_missing  PAQ_A-Season_missing  \
0                                     1                     1
1                                     1                     1
2                                     1                     1
3                                     1                     1
4                                     1                     1
5                                     1                     1

   PAQ_A-PAQ_A_Total_missing    X_mean     X_std     X_min     X_max  \
0                          1  0.000000  0.000000  0.000000  0.000000
1                          1  0.000000  0.000000  0.000000  0.000000
2                          1 -0.316384  0.453665 -1.746094  1.507865
3                          1 -0.004272  0.351545 -1.038711  1.034351
4                          1  0.000000  0.000000  0.000000  0.000000
5                          1  0.000000  0.000000  0.000000  0.000000

     Y_mean     Y_std     Y_min     Y_max    Z_mean     Z_std     Z_min  \
0  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
1  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
2  0.016009  0.502702 -2.905339  1.666354 -0.167890  0.585710 -1.048372
3  0.016859  0.303812 -1.522690  1.946303 -0.631731  0.623476 -1.018787
4  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000
5  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000  0.000000

      Z_max  enmo_mean  enmo_std  anglez_mean  anglez_std  non-wear_flag_mean  \
0  0.000000   0.000000  0.000000     0.000000    0.000000            0.000000
1  0.000000   0.000000  0.000000     0.000000    0.000000            0.000000
2  1.546979   0.047388  0.106351   -10.580416   42.947170            0.000000
```

```
3  1.146284   0.011926  0.024331   -55.630768   50.303635            0.655708
4  0.000000   0.000000  0.000000     0.000000    0.000000            0.000000
5  0.000000   0.000000  0.000000     0.000000    0.000000            0.000000

   light_mean   light_std  battery_voltage_mean  battery_voltage_std  \
0     0.00000    0.000000              0.000000             0.000000
1     0.00000    0.000000              0.000000             0.000000
2    42.29631  208.168976           4053.578857           112.404037
3    16.77198   95.327438           3838.189453           155.573868
4     0.00000    0.000000              0.000000             0.000000
5     0.00000    0.000000              0.000000             0.000000

   time_of_day_mean  weekday_mean  quarter_mean  relative_date_PCIAT_mean  \
0      0.000000e+00      0.000000           0.0                  0.000000
1      0.000000e+00      0.000000           0.0                  0.000000
2      5.046215e+13      4.470182           3.0                 53.201683
3      4.321212e+13      3.909848           3.0                 79.435593
4      0.000000e+00      0.000000           0.0                  0.000000
5      0.000000e+00      0.000000           0.0                  0.000000

   fourier_X_real  fourier_X_imag  fourier_Y_real  fourier_Y_imag  \
0        0.000000    0.000000e+00        0.000000        0.000000
1        0.000000    0.000000e+00        0.000000        0.000000
2    -2653.531675   -1.083578e-14      705.803060     -207.446683
3     4103.865122   -1.680035e+03    14385.071425     3120.930726
4        0.000000    0.000000e+00        0.000000        0.000000
5        0.000000    0.000000e+00        0.000000        0.000000

   fourier_Z_real  fourier_Z_imag  activity_during_day  activity_during_night  \
0        0.000000    0.000000e+00                  0.0               0.000000
1        0.000000    0.000000e+00                  0.0               0.000000
2     1995.401835   -5.684342e-15                  0.0            2053.305176
3   -26765.674436   -8.731149e-12                  0.0            4727.518555
4        0.000000    0.000000e+00                  0.0               0.000000
5        0.000000    0.000000e+00                  0.0               0.000000

   activity_ratio_day_night  non_wear_proportion  sedentary_proportion  \
0                       0.0             0.000000              0.000000
1                       0.0             0.000000              0.000000
2                       0.0             0.000000              0.792453
3                       0.0             0.655708              0.978501
4                       0.0             0.000000              0.000000
5                       0.0             0.000000              0.000000

   light_activity_proportion  moderate_activity_proportion  \
0                   0.000000                      0.000000
1                   0.000000                      0.000000
```

```
2                    0.198131                    0.007870
3                    0.021171                    0.000288
4                    0.000000                    0.000000
5                    0.000000                    0.000000

   vigorous_activity_proportion  actigraphy_present  \
0                      0.000000                   0
1                      0.000000                   0
2                      0.001546                   1
3                      0.000040                   1
4                      0.000000                   0
5                      0.000000                   0

   CGAS-CGAS_Score_x_SDS-SDS_Total_Raw  CGAS-CGAS_Score_x_Physical-BMI
0                                    0                               0
1                                    0                               0
2                                    0                               0
3                                    0                               0
4                                    0                               0
5                                    0                               0
```

[97]:
```
# Ensure that the feature columns are of numeric type (float32) before making␣
↪predictions
X_test = merged_test_df[selected_features_list].copy()

# Convert categorical columns in X_test to numerical types if necessary
for col in X_test.select_dtypes(include=['category', 'object']).columns:
    # Convert categorical values to numeric using the category codes or an␣
↪appropriate mapping
    X_test[col] = X_test[col].astype('category').cat.codes

# Convert the entire DataFrame to float32 type to match the model's expected␣
↪input type
X_test = X_test.astype('float32')

# Step 2: Use the final student model to make predictions on merged_test_df
test_predictions = student_model.predict(X_test)

# Step 3: Convert softmax predictions to class labels (integers)
predicted_labels = np.argmax(test_predictions, axis=1)

# Step 4: Create a DataFrame with 'id' and 'sii' columns
submission_df = merged_test_df[['id']].copy()  # Copy the 'id' column from␣
↪merged_test_df
submission_df['sii'] = predicted_labels  # Add the predicted 'sii' values

# Step 5: Save the result as a CSV file with header
```

```
#submission_df.to_csv('imported/output/sample_submission.csv', index=False)


#print("sample_submission.csv created successfully.")
```

```
1/1            0s 21ms/step
```

[100]:

```
X_test.head(10)
```

[100]:

```
   Physical-BMI  Physical-Weight  Age_Group  PAQ_A-Season  \
0     16.877316        50.799999        0.0           NaN
1     16.648697        75.599998        0.0           NaN
2     18.292347        81.599998        0.0           NaN
3     22.279951       112.199997        1.0           NaN
4     17.864056        73.360100        1.0           NaN
5     17.284504        47.599998        0.0           NaN

   PreInt_EduHx-computerinternet_hoursday  Physical-Height  FGC-FGC_GSND  \
0                                     3.0         46.00000      3.706898
1                                     2.0         56.50000     10.200000
2                                     0.0         56.00000     11.694644
3                                     0.0         59.50000     16.500000
4                                     0.0         53.08939      9.702796
5                                     0.0         44.00000      6.372775

   BIA-BIA_BMI  FGC-FGC_CU  Basic_Demos-Age
0    16.879200    0.000000              5.0
1    17.160849   20.000000             10.0
2    18.294300   18.000000              9.0
3    30.186501   12.000000             13.0
4    18.782181   11.771994             11.0
5    13.459508    0.000000              5.0
```

[98]:

```
submission_df.head(12)
```

[98]:

```
         id  sii
0  00008ff9    0
1  00105258    0
2  00115b9f    0
3  001f3379    0
4  00abe655    0
5  00d56d4b    0
```