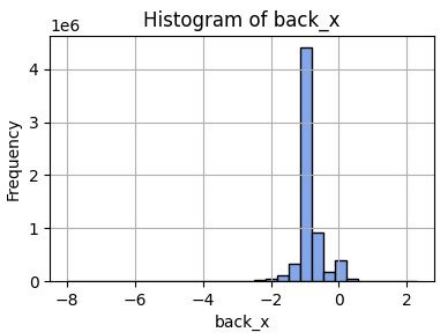
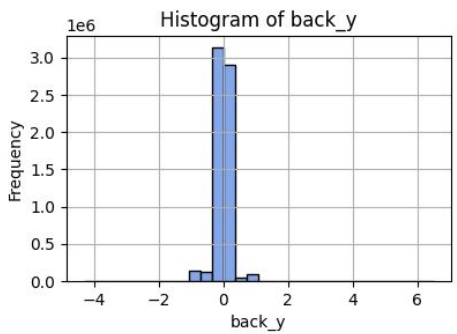
# HARTH dataset Import

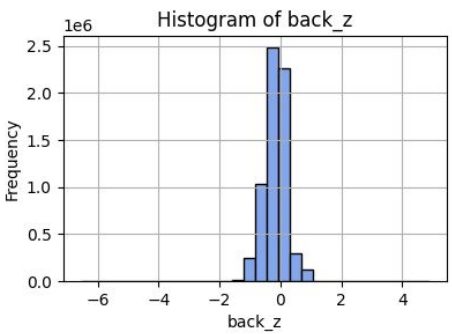
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn as sk
import os

harth_folder = "C://Users//cream//Downloads//CS_156//Demo2//harth"
data_frame = pd.DataFrame()

expected_columns = ['timestamp', 'back_x', 'back_y', 'back_z', 'thigh_x', 'thigh_y', 'thigh_z', 'label']

for file in os.listdir(harth_folder):
    if file.endswith(".csv"):
        file_path = os.path.join(harth_folder, file)
        temp = pd.read_csv(file_path)
        temp = temp[[col for col in temp.columns if col in expected_columns]]
        data_frame = pd.concat([data_frame, temp], ignore_index=True)

print(data_frame.head())
```

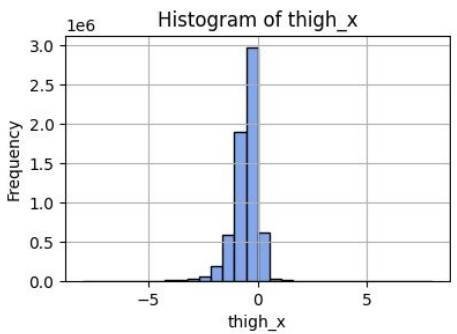
```
                timestamp    back_x    back_y  ...    thigh_y   thigh_z  label
0  2019-01-12 00:00:00.000 -0.760242  0.299570  ... -0.298644  0.709439      6
1  2019-01-12 00:00:00.010 -0.530138  0.281880  ...  0.286944  0.340309      6
2  2019-01-12 00:00:00.020 -1.170922  0.186353  ... -0.078423 -0.515212      6
3  2019-01-12 00:00:00.030 -0.648772  0.016579  ... -0.950978 -0.221140      6
4  2019-01-12 00:00:00.040 -0.355071 -0.051831  ...  0.140903 -0.653782      6
```

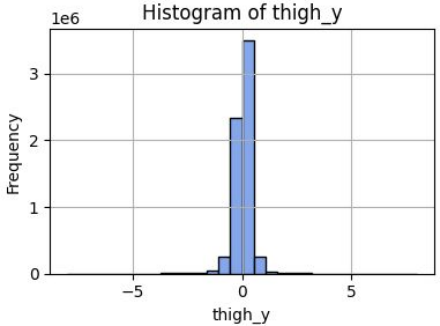
# Dataset Visualization

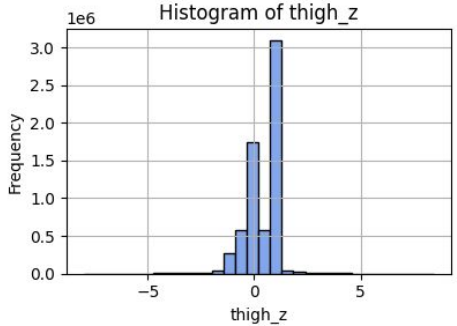


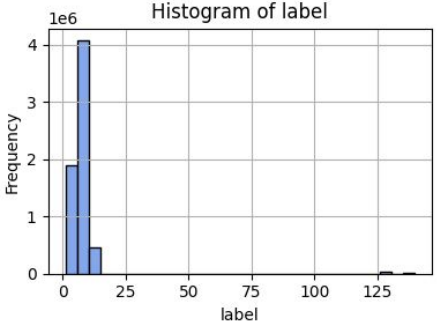



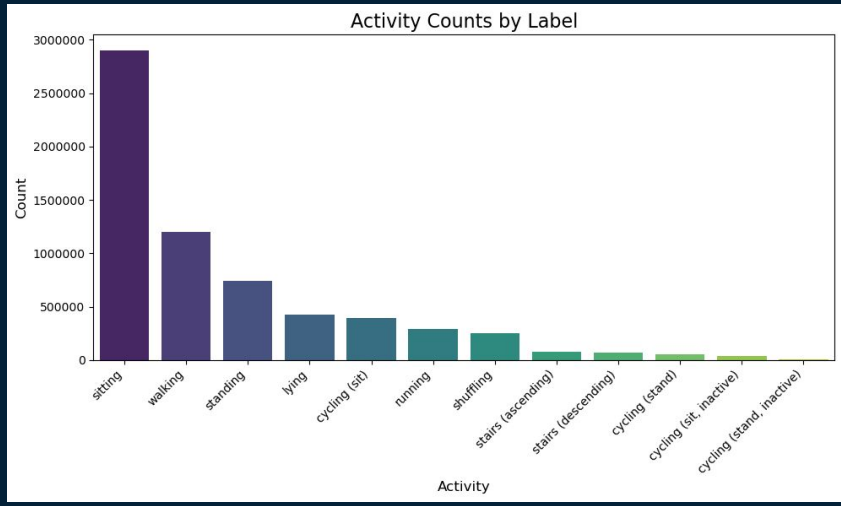



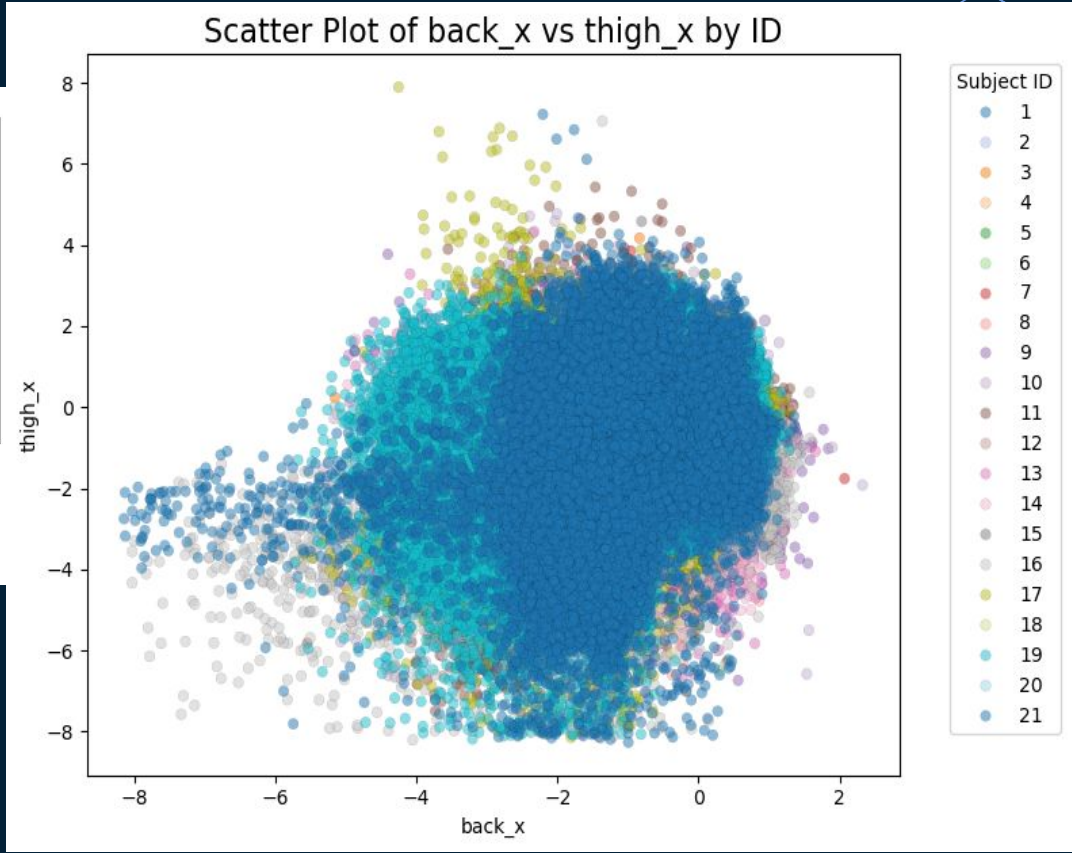



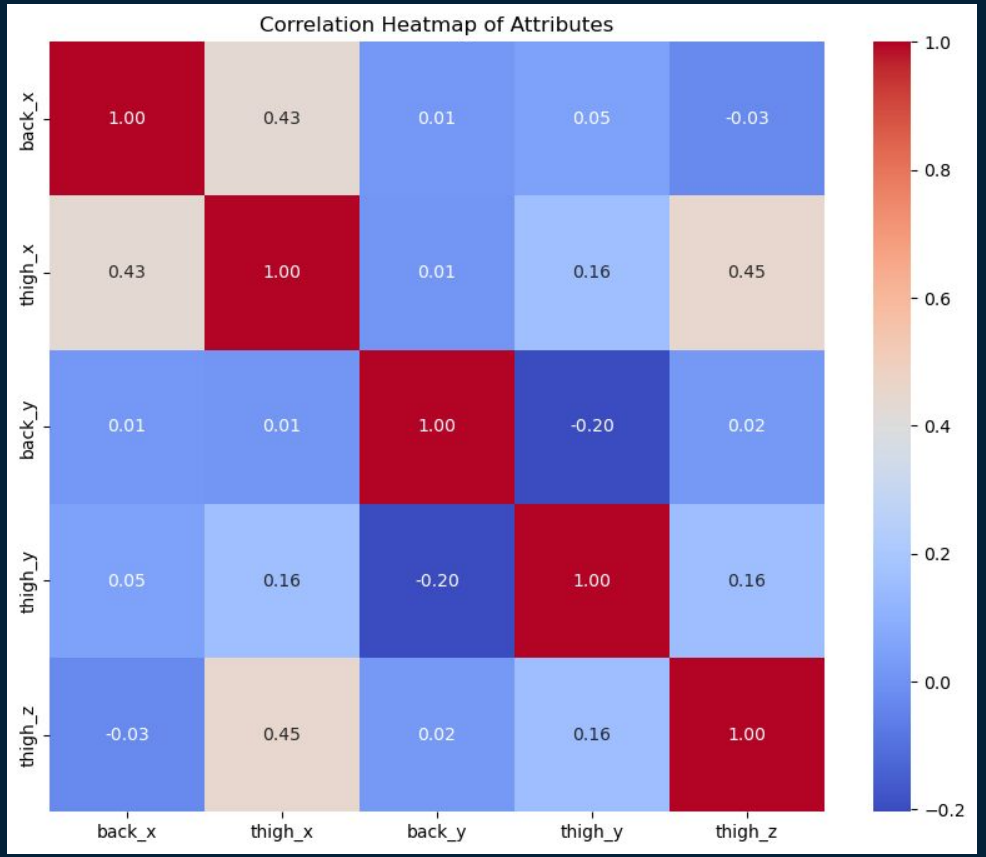



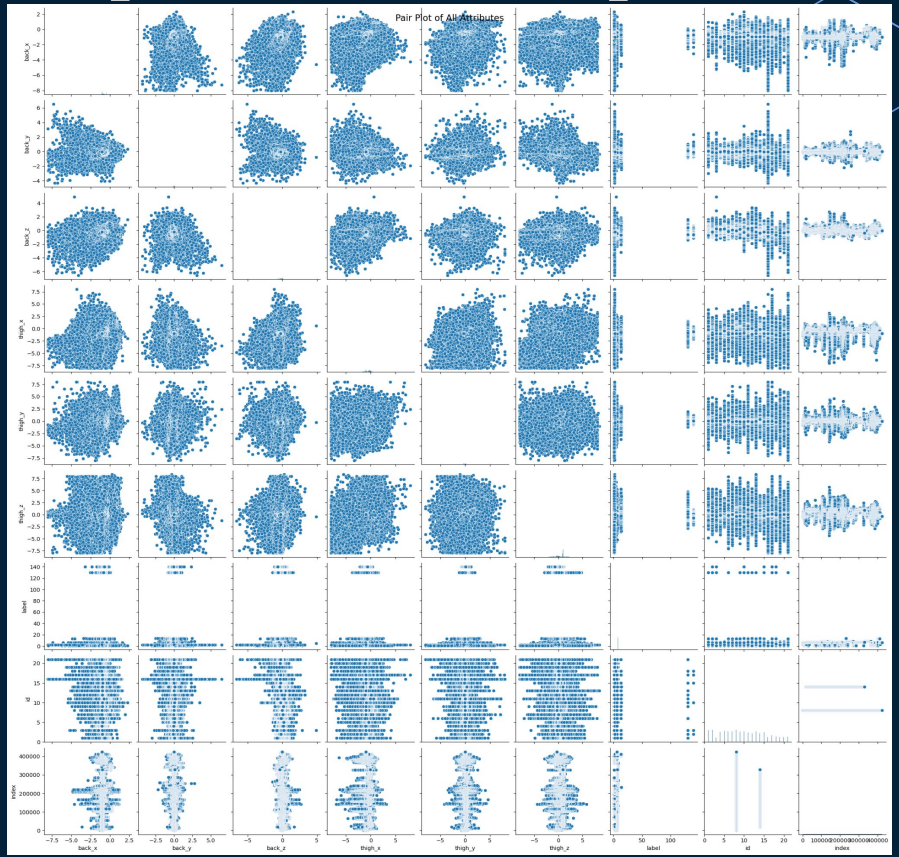
# Dataset Visualization 2



# Data Visualization 2 (Continued)



# Dataset Preprocessing

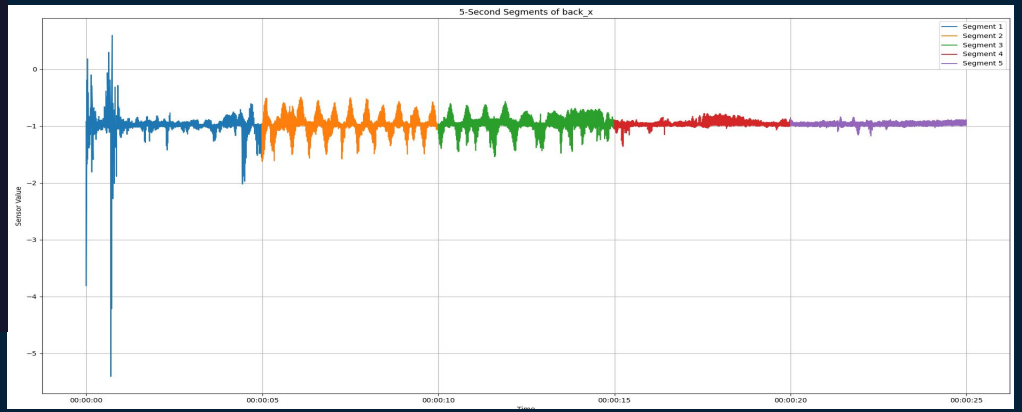
```
[5 rows x 8 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6461328 entries, 0 to 6461327
Data columns (total 8 columns):
 #   Column     Dtype
---  ------     -----
 0   timestamp  object
 1   back_x     float64
 2   back_y     float64
 3   back_z     float64
 4   thigh_x    float64
 5   thigh_y    float64
 6   thigh_z    float64
 7   label      int64
dtypes: float64(6), int64(1), object(1)
memory usage: 394.4+ MB
Columns and num of missing values :
timestamp    0
back_x       0
back_y       0
back_z       0
thigh_x      0
thigh_y      0
thigh_z      0
label        0
```

- Harth Dataset has no missing values.

**Alternative Cases where values are missing?**

- Replaces missing values with median value.
- Delete the rows that has too many missing values.
- Replaces missing values with custom values that are based on neighboring values.

**Data Segmentation**



# Dataset Preprocessing 2

## Noise Reduction and Normalization