whoop_eda

July 6, 2024

1 Setup

[92]:
```
import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
import os
```

[93]:
```
WHOOP_DIR = 'data/whoop/raw'
WHOOP_EXPORT_DIR = 'data/whoop/eda'
```

[94]:
```
def display_correlation_matrix(df: pd.DataFrame):
    plt.figure(figsize=(16, 12))
    correlation_matrix = df.corr(numeric_only=True)
    mask = np.triu(np.ones_like(correlation_matrix, dtype=bool))
    sns.heatmap(correlation_matrix, mask=mask, annot=True, cmap='Blues',␣
 ↪vmin=-1, vmax=1)
    plt.title('Correlation Matrix')
    plt.show()
```

[95]:
```
def get_correlated_features(df, threshold=0.5):
    correlation_matrix = df.corr(numeric_only=True)
    correlation_pairs = correlation_matrix.unstack()
    correlation_pairs = pd.DataFrame(correlation_pairs).reset_index()
    correlation_pairs.columns = ['Feature_1', 'Feature_2', 'Correlation']
    correlation_pairs = correlation_pairs[correlation_pairs['Feature_1'] !=␣
 ↪correlation_pairs['Feature_2']]
    correlated_features =␣
 ↪correlation_pairs[abs(correlation_pairs['Correlation']) > threshold]
    correlated_features = correlated_features.sort_values(by='Correlation',␣
 ↪ascending=False)

    return correlated_features

def display_pairplot(df, threshold=0.5):
    corr_features = get_correlated_features(df, threshold)['Feature_1'].unique()
    sns.pairplot(
        df[corr_features],
        kind='reg',
        diag_kind='kde',
        corner=True,
        plot_kws={
            'line_kws':{'color':'red', 'linestyle':'--'},
            'scatter_kws': {'alpha': 0.5}
        }
    )
```

2 Journal Entries

2.0.1 Options

- Include as factors to predict Recovery from strain & sleep metrics
  - Use only those with highest correlation & significance to Recovery

2.1 Import

[96]:

```
JOURNAL_PATH = os.path.join(WHOOP_DIR, 'journal_entries.csv')
```

[97]:

```
df_journal = pd.read_csv(JOURNAL_PATH)
df_journal
```

[97]:

```
          Cycle start time       Cycle end time Cycle timezone  \
0      2024-06-17 00:35:19                  NaN      UTC-05:00
1      2024-06-17 00:35:19                  NaN      UTC-05:00
2      2024-06-17 00:35:19                  NaN      UTC-05:00
3      2024-06-17 00:35:19                  NaN      UTC-05:00
4      2024-06-17 00:35:19                  NaN      UTC-05:00
…                    …                  …            …
10399  2022-09-14 00:00:00  2022-09-14 21:52:24      UTC-05:00
10400  2022-09-14 00:00:00  2022-09-14 21:52:24      UTC-05:00
10401  2022-09-14 00:00:00  2022-09-14 21:52:24      UTC-05:00
10402  2022-09-14 00:00:00  2022-09-14 21:52:24      UTC-05:00
10403  2022-09-14 00:00:00  2022-09-14 21:52:24      UTC-05:00

                             Question text  Answered yes
0                Have any alcoholic drinks?         False
1                       Have any caffeine?           True
2                         Consumed calcium          True
3                    Tracked your calories          True
4                   Consumed carbohydrates          True
…                                    …           …
10399              Use CBD oil in any form?         False
10400                      Felt recovered?          True
```

```
10401       Eat any food close to bedtime?          True
10402  See direct sunlight upon waking up?          True
10403  Take prescription sleep medication?          True

[10404 rows x 5 columns]
```

2.2 Drop columns

[98]:

```
df_journal.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10404 entries, 0 to 10403
Data columns (total 5 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Cycle start time  10404 non-null  object
 1   Cycle end time    10377 non-null  object
 2   Cycle timezone    10404 non-null  object
 3   Question text     10404 non-null  object
 4   Answered yes      10404 non-null  bool
dtypes: bool(1), object(4)
memory usage: 335.4+ KB
```

[99]:

```
JOURNAL_DROP_COLS = [
    'Cycle end time',
    'Cycle timezone'
]
```

[100]:

```
df_journal_sel = df_journal.copy().drop(columns=JOURNAL_DROP_COLS).
 ↪dropna(how='any')
df_journal_sel
```

[100]:

```
          Cycle start time                        Question text  Answered yes
0      2024-06-17 00:35:19          Have any alcoholic drinks?         False
1      2024-06-17 00:35:19                 Have any caffeine?          True
2      2024-06-17 00:35:19                    Consumed calcium          True
3      2024-06-17 00:35:19               Tracked your calories          True
4      2024-06-17 00:35:19             Consumed carbohydrates          True
…                    …                                  …           …
10399  2022-09-14 00:00:00           Use CBD oil in any form?         False
10400  2022-09-14 00:00:00                    Felt recovered?          True
10401  2022-09-14 00:00:00      Eat any food close to bedtime?          True
10402  2022-09-14 00:00:00  See direct sunlight upon waking up?          True
10403  2022-09-14 00:00:00  Take prescription sleep medication?          True

[10404 rows x 3 columns]
```

2.3 Dtypes

```
[101]: df_journal_dt = df_journal_sel.copy()
       df_journal_dt['Cycle start time'] = pd.to_datetime(df_journal_dt['Cycle start␣
        ↪time'])
```

```
[102]: df_journal_dt['Question text'].unique()
```

```
[102]: array(['Have any alcoholic drinks?', 'Have any caffeine? ',
              'Consumed calcium', 'Tracked your calories',
              'Consumed carbohydrates', 'Use CBD oil in any form?',
              'Avoid consuming processed foods?', 'Took a cold shower',
              'Did you utilize a CPAP Machine?', 'Consumed fats',
              'Consumed fiber', 'Take an ice bath?', 'Have an injury or wound',
              'Eat any food close to bedtime?', 'Consumed magnesium',
              'Used marijuana', 'Spend time outdoors?', 'Consumed protein?',
              'Felt recovered?', 'Use a sauna?', 'Engage in sexual activity?',
              'Slept in the same bed as usual', 'Wear a sleep mask?',
              'Consumed sodium?', 'Listen to noise while asleep?',
              'Spend time stretching?', 'Use tobacco in any form?',
              'Take prescription sleep medication?', 'Masturbate?',
              'Experiencing seasonal allergies',
              'Have a hot flash while sleeping?',
              'See direct sunlight upon waking up?'], dtype=object)
```

```
[103]: df_journal_dt
```

```
[103]:           Cycle start time                        Question text  Answered yes
       0     2024-06-17 00:35:19           Have any alcoholic drinks?         False
       1     2024-06-17 00:35:19                  Have any caffeine?           True
       2     2024-06-17 00:35:19                     Consumed calcium          True
       3     2024-06-17 00:35:19                Tracked your calories          True
       4     2024-06-17 00:35:19               Consumed carbohydrates          True
       …                       …                                    …           …
       10399 2022-09-14 00:00:00             Use CBD oil in any form?         False
       10400 2022-09-14 00:00:00                      Felt recovered?          True
       10401 2022-09-14 00:00:00       Eat any food close to bedtime?          True
       10402 2022-09-14 00:00:00  See direct sunlight upon waking up?          True
       10403 2022-09-14 00:00:00  Take prescription sleep medication?          True

       [10404 rows x 3 columns]
```

```
[104]: df_journal_dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10404 entries, 0 to 10403
Data columns (total 3 columns):
 #   Column            Non-Null Count  Dtype
```

```
---  ------             --------------  -----
 0   Cycle start time  10404 non-null  datetime64[ns]
 1   Question text     10404 non-null  object
 2   Answered yes      10404 non-null  bool
dtypes: bool(1), datetime64[ns](1), object(1)
memory usage: 172.8+ KB
```

2.4 Add cols

[105]:

```
df_journal_add = df_journal_dt.copy()
df_journal_add['cycle_date'] = df_journal_add['Cycle start time'].dt.date # do␣
 ↪not add 1 day here -- journal answers should reflect activities on previous␣
 ↪day
# df_journal_add.drop(columns='Cycle start time', inplace=True)
```

[106]:

```
journal_raw_question_counts = df_journal_dt.groupby('Question text')['Answered␣
 ↪yes'].nunique()
journal_valid_questions =␣
 ↪journal_raw_question_counts[journal_raw_question_counts > 1].index
df_journal_valid = df_journal_add[df_journal_add['Question text'].
 ↪isin(journal_valid_questions)]
df_journal_valid
```

[106]:

```
         Cycle start time                    Question text  Answered yes  \
0     2024-06-17 00:35:19          Have any alcoholic drinks?         False
1     2024-06-17 00:35:19               Have any caffeine?          True
3     2024-06-17 00:35:19             Tracked your calories          True
4     2024-06-17 00:35:19            Consumed carbohydrates          True
5     2024-06-17 00:35:19           Use CBD oil in any form?          True
…                     …                               …           …
10398 2022-09-14 00:00:00            Consumed carbohydrates         False
10399 2022-09-14 00:00:00           Use CBD oil in any form?         False
10401 2022-09-14 00:00:00       Eat any food close to bedtime?          True
10402 2022-09-14 00:00:00  See direct sunlight upon waking up?          True
10403 2022-09-14 00:00:00  Take prescription sleep medication?          True

       cycle_date
0      2024-06-17
1      2024-06-17
3      2024-06-17
4      2024-06-17
5      2024-06-17
…           …
10398  2022-09-14
10399  2022-09-14
10401  2022-09-14
10402  2022-09-14
```

```
10403  2022-09-14

[8476 rows x 4 columns]
```

```
[107]: df_journal_valid['Question text'].unique()
```

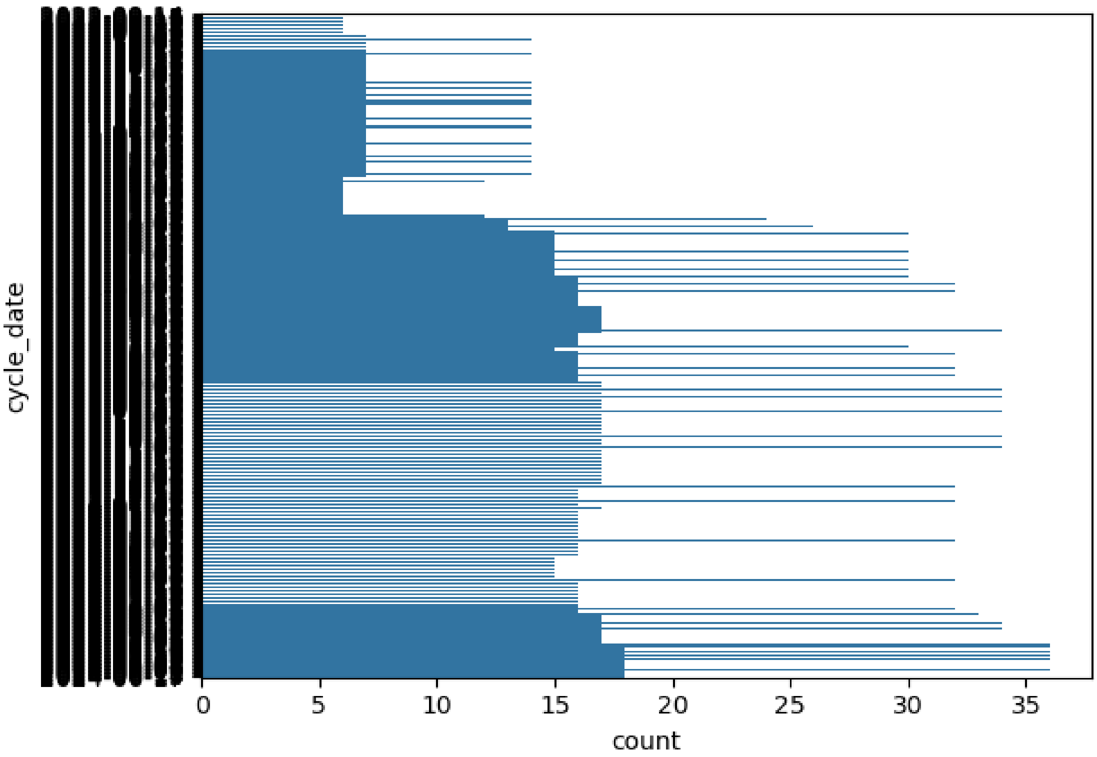
```
[107]: array(['Have any alcoholic drinks?', 'Have any caffeine? ',
              'Tracked your calories', 'Consumed carbohydrates',
              'Use CBD oil in any form?', 'Avoid consuming processed foods?',
              'Took a cold shower', 'Did you utilize a CPAP Machine?',
              'Take an ice bath?', 'Have an injury or wound',
              'Eat any food close to bedtime?', 'Used marijuana',
              'Spend time outdoors?', 'Use a sauna?',
              'Slept in the same bed as usual', 'Wear a sleep mask?',
              'Spend time stretching?', 'Use tobacco in any form?',
              'Take prescription sleep medication?', 'Masturbate?',
              'Experiencing seasonal allergies',
              'Have a hot flash while sleeping?',
              'See direct sunlight upon waking up?'], dtype=object)
```

2.5 EDA

2.5.1 Date Counts

```
[108]: journal_date_counts = df_journal_valid['cycle_date'].value_counts().
         ↪sort_index().reset_index()
       sns.barplot(data=journal_date_counts, x='count', y='cycle_date')
```

```
[108]: <Axes: xlabel='count', ylabel='cycle_date'>
```



2.5.2 Question Counts

```
[109]: journal_valid_question_counts = df_journal_valid.groupby('Question␣
        ↪text')['Answered yes'].value_counts().unstack(fill_value=0)
       journal_valid_question_counts.columns = ['false', 'true']
       journal_valid_question_counts
```

```
[109]:                                      false  true
       Question text
       Avoid consuming processed foods?       153   135
       Consumed carbohydrates                   3    65
       Did you utilize a CPAP Machine?          3   629
       Eat any food close to bedtime?         169   464
       Experiencing seasonal allergies        233    25
       Have a hot flash while sleeping?       111     9
       Have an injury or wound                395    48
       Have any alcoholic drinks?             526    89
       Have any caffeine?                       8   625
       Masturbate?                            202    75
       See direct sunlight upon waking up?    147    43
       Slept in the same bed as usual          31   412
       Spend time outdoors?                   375    68
       Spend time stretching?                  367   76
       Take an ice bath?                       315   42
       Take prescription sleep medication?     215  283
       Took a cold shower                      408   19
       Tracked your calories                     1  162
       Use CBD oil in any form?                160   32
       Use a sauna?                            346   81
       Use tobacco in any form?                225  158
       Used marijuana                           71   29
       Wear a sleep mask?                       11  432
```

[110]:

```
journal_valid_question_counts['total'] = journal_valid_question_counts['false']␣
  ↪+ journal_valid_question_counts['true']
journal_valid_question_counts['true_rate'] =␣
  ↪journal_valid_question_counts['true'] /␣
  ↪journal_valid_question_counts['total']
journal_valid_question_counts.sort_values(by='true_rate', inplace=True)
journal_valid_question_counts
```

[110]:

```
                                     false  true  total  true_rate
Question text
Took a cold shower                     408    19    427   0.044496
Have a hot flash while sleeping?       111     9    120   0.075000
Experiencing seasonal allergies        233    25    258   0.096899
Have an injury or wound                395    48    443   0.108352
Take an ice bath?                      315    42    357   0.117647
Have any alcoholic drinks?             526    89    615   0.144715
Spend time outdoors?                   375    68    443   0.153499
Use CBD oil in any form?               160    32    192   0.166667
Spend time stretching?                 367    76    443   0.171558
Use a sauna?                           346    81    427   0.189696
See direct sunlight upon waking up?    147    43    190   0.226316
Masturbate?                            202    75    277   0.270758
Used marijuana                          71    29    100   0.290000
Use tobacco in any form?               225   158    383   0.412533
Avoid consuming processed foods?       153   135    288   0.468750
Take prescription sleep medication?    215   283    498   0.568273
Eat any food close to bedtime?         169   464    633   0.733017
Slept in the same bed as usual          31   412    443   0.930023
Consumed carbohydrates                   3    65     68   0.955882
Wear a sleep mask?                      11   432    443   0.975169
Have any caffeine?                       8   625    633   0.987362
Tracked your calories                    1   162    163   0.993865
Did you utilize a CPAP Machine?          3   629    632   0.995253
```

[111]:

```
journal_valid_question_counts =␣
 ↪journal_valid_question_counts[(journal_valid_question_counts['true_rate'] >␣
 ↪0.1) & (journal_valid_question_counts['true_rate'] < 0.9)]
journal_valid_question_counts
```

[111]:

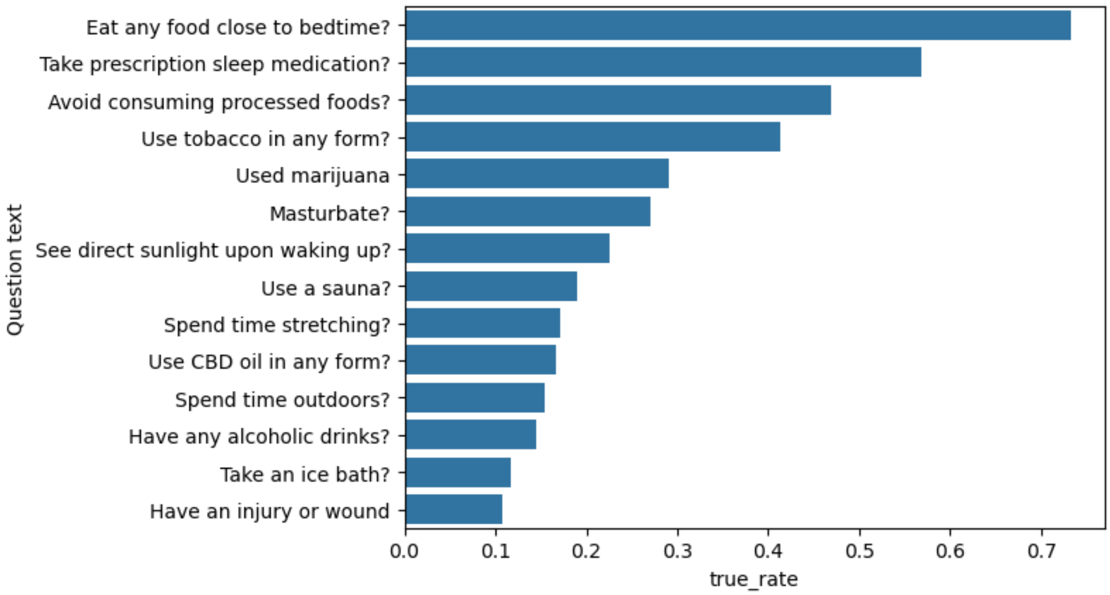
```
                                     false  true  total  true_rate
Question text
Have an injury or wound                395    48    443   0.108352
Take an ice bath?                      315    42    357   0.117647
Have any alcoholic drinks?             526    89    615   0.144715
Spend time outdoors?                   375    68    443   0.153499
Use CBD oil in any form?               160    32    192   0.166667
Spend time stretching?                 367    76    443   0.171558
Use a sauna?                           346    81    427   0.189696
See direct sunlight upon waking up?    147    43    190   0.226316
Masturbate?                            202    75    277   0.270758
Used marijuana                          71    29    100   0.290000
Use tobacco in any form?               225   158    383   0.412533
Avoid consuming processed foods?       153   135    288   0.468750
Take prescription sleep medication?    215   283    498   0.568273
Eat any food close to bedtime?         169   464    633   0.733017
```

[112]:

```
sns.barplot(data=journal_valid_question_counts.sort_values(by='true_rate',␣
 ↪ascending=False), x='true_rate', y='Question text')
```

[112]:

```
<Axes: xlabel='true_rate', ylabel='Question text'>
```



[113]:
```
df_journal_valid = df_journal_valid[df_journal_valid['Question text'].
 ↪isin(journal_valid_question_counts.index)]
df_journal_valid
```

[113]:
```
          Cycle start time                         Question text  Answered yes  \
0      2024-06-17 00:35:19           Have any alcoholic drinks?         False
5      2024-06-17 00:35:19              Use CBD oil in any form?          True
6      2024-06-17 00:35:19     Avoid consuming processed foods?         False
11     2024-06-17 00:35:19                     Take an ice bath?         False
12     2024-06-17 00:35:19                Have an injury or wound         False
…                      …                                     …            …
10396  2022-09-14 21:52:24  Take prescription sleep medication?          True
10399  2022-09-14 00:00:00              Use CBD oil in any form?         False
10401  2022-09-14 00:00:00         Eat any food close to bedtime?          True
10402  2022-09-14 00:00:00  See direct sunlight upon waking up?          True
10403  2022-09-14 00:00:00  Take prescription sleep medication?          True

       cycle_date
0      2024-06-17
5      2024-06-17
6      2024-06-17
11     2024-06-17
12     2024-06-17
…             …
10396  2022-09-14
10399  2022-09-14
10401  2022-09-14
10402  2022-09-14
10403  2022-09-14

[5289 rows x 4 columns]
```

2.6 Export

[114]:
```
df_journal_valid.to_csv(os.path.join(WHOOP_EXPORT_DIR, 'journal_valid.csv'),␣
 ↪index=False)
```

3 Physiological Cycles

3.0.1 Options

- Predict Recovery from strain & sleep metrics

3.1 Import

[115]:
```
PHYS_PATH = os.path.join(WHOOP_DIR, 'physiological_cycles.csv')
```

[116]:
```
df_phys = pd.read_csv(PHYS_PATH)
df_phys
```

[116]:
```
        Cycle start time       Cycle end time Cycle timezone  \
0    2024-06-17 00:35:19                  NaN      UTC-05:00
1    2024-06-15 23:12:59  2024-06-17 00:35:19      UTC-05:00
2    2024-06-14 23:39:45  2024-06-15 23:12:59      UTC-05:00
3    2024-06-14 00:07:47  2024-06-14 23:39:45      UTC-05:00
4    2024-06-13 00:12:09  2024-06-14 00:07:47      UTC-05:00
..                   …                    …            …
638  2022-09-17 22:36:20  2022-09-18 22:20:02      UTC-05:00
639  2022-09-16 21:53:01  2022-09-17 22:36:20      UTC-05:00
640  2022-09-15 22:58:46  2022-09-16 21:53:01      UTC-05:00
641  2022-09-14 21:52:24  2022-09-15 22:58:46      UTC-05:00
642  2022-09-14 00:00:00  2022-09-14 21:52:24      UTC-05:00

     Recovery score %  Resting heart rate (bpm)  Heart rate variability (ms)  \
0                95.0                      50.0                         99.0
1                84.0                      52.0                         92.0
2                94.0                      48.0                         95.0
3                83.0                      50.0                         87.0
4                73.0                      49.0                         80.0
..                …                       …                          …
638              65.0                      52.0                         85.0
639              78.0                      60.0                        102.0
640              97.0                      52.0                        132.0
641              68.0                      62.0                         75.0
642               NaN                       NaN                          NaN

     Skin temp (celsius)  Blood oxygen %  Day Strain  Energy burned (cal)  \
0                  33.80           98.88         NaN                  NaN
1                  33.30           96.14        12.6               2520.0
2                  34.30           98.89        17.3               3689.0
3                  34.00           97.82        14.8               2700.0
4                  33.20           99.00        12.8               2505.0
..                   …             …         …                  …
638                32.69           96.91        10.3               2104.0
639                32.35           96.40        15.5               2711.0
640                33.18           95.87        16.5               2584.0
641                32.73           97.36         5.6               1938.0
642                  NaN             NaN        14.6               1412.0

     …  Asleep duration (min)  In bed duration (min)  \
0    …                  507.0                  533.0
1    …                  533.0                  570.0
2    …                  433.0                  484.0
3    …                  535.0                  571.0
```

```
4    …                  488.0                  562.0
..   …                    …                    …
638  …                  448.0                  520.0
639  …                  495.0                  520.0
640  …                  411.0                  497.0
641  …                  529.0                  640.0
642  …                    NaN                    NaN

    Light sleep duration (min) Deep (SWS) duration (min)  REM duration (min)  \
0                        247.0                      94.0               166.0
1                        231.0                     137.0               165.0
2                        298.0                      95.0                40.0
3                        273.0                     126.0               136.0
4                        221.0                     123.0               144.0
..                         …                       …                 …
638                      267.0                     102.0                79.0
639                      357.0                      31.0               107.0
640                      244.0                      73.0                94.0
641                      362.0                      72.0                95.0
642                        NaN                       NaN                 NaN

     Awake duration (min)  Sleep need (min)  Sleep debt (min)  \
0                    26.0             489.0               7.0
1                    37.0             544.0              31.0
2                    51.0             495.0               0.0
3                    36.0             483.0               0.0
4                    74.0             472.0               0.0
..                    …               …               …
638                  72.0             526.0              23.0
639                  25.0             537.0              27.0
640                  86.0             468.0               2.0
641                 111.0             533.0               0.0
642                   NaN               NaN               NaN

     Sleep efficiency %  Sleep consistency %
0                  95.0                 78.0
1                  93.0                 77.0
2                  89.0                 90.0
3                  94.0                 87.0
4                  89.0                 81.0
..                  …                  …
638                86.0                 84.0
639                95.0                  NaN
640                82.0                  NaN
641                84.0                  NaN
642                 NaN                  NaN
```

```
[643 rows x 26 columns]
```

3.2 Drop columns

[117]:

```
df_phys.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 643 entries, 0 to 642
Data columns (total 26 columns):
 #   Column                       Non-Null Count  Dtype  
---  ------                       --------------  -----  
 0   Cycle start time             643 non-null    object 
 1   Cycle end time               642 non-null    object 
 2   Cycle timezone               643 non-null    object 
 3   Recovery score %             639 non-null    float64
 4   Resting heart rate (bpm)     639 non-null    float64
 5   Heart rate variability (ms)  639 non-null    float64
 6   Skin temp (celsius)          639 non-null    float64
 7   Blood oxygen %               638 non-null    float64
 8   Day Strain                   642 non-null    float64
 9   Energy burned (cal)          642 non-null    float64
 10  Max HR (bpm)                 642 non-null    float64
 11  Average HR (bpm)             642 non-null    float64
 12  Sleep onset                  639 non-null    object 
 13  Wake onset                   639 non-null    object 
 14  Sleep performance %          639 non-null    float64
 15  Respiratory rate (rpm)       637 non-null    float64
 16  Asleep duration (min)        639 non-null    float64
 17  In bed duration (min)        639 non-null    float64
 18  Light sleep duration (min)   639 non-null    float64
 19  Deep (SWS) duration (min)    639 non-null    float64
 20  REM duration (min)           639 non-null    float64
 21  Awake duration (min)         639 non-null    float64
 22  Sleep need (min)             639 non-null    float64
 23  Sleep debt (min)             639 non-null    float64
 24  Sleep efficiency %           639 non-null    float64
 25  Sleep consistency %          636 non-null    float64
dtypes: float64(21), object(5)
memory usage: 130.7+ KB
```

[118]:

```
PHYS_DROP_COLS =[
    'Cycle end time',
    'Cycle timezone',
    'Wake onset'
]
```

[119]:
```
df_phys_sel = df_phys.copy().drop(columns=PHYS_DROP_COLS).dropna(how='any')
df_phys_sel
```

[119]:
```
        Cycle start time  Recovery score %  Resting heart rate (bpm)  \
1    2024-06-15 23:12:59              84.0                      52.0
2    2024-06-14 23:39:45              94.0                      48.0
3    2024-06-14 00:07:47              83.0                      50.0
4    2024-06-13 00:12:09              73.0                      49.0
5    2024-06-11 23:11:54              74.0                      50.0
..                   …                 …                         …
634  2022-09-21 22:01:31              49.0                      46.0
635  2022-09-20 22:20:13              48.0                      47.0
636  2022-09-19 22:41:43              45.0                      49.0
637  2022-09-18 22:20:02              55.0                      49.0
638  2022-09-17 22:36:20              65.0                      52.0

     Heart rate variability (ms)  Skin temp (celsius)  Blood oxygen %  \
1                           92.0                33.30           96.14
2                           95.0                34.30           98.89
3                           87.0                34.00           97.82
4                           80.0                33.20           99.00
5                           78.0                33.90           98.40
..                           …                    …               …
634                         68.0                31.83           96.61
635                         65.0                32.80           96.69
636                         67.0                32.44           95.84
637                         68.0                33.30           97.33
638                         85.0                32.69           96.91

     Day Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  …  \
1          12.6               2520.0         153.0              73.0  …
2          17.3               3689.0         159.0              82.0  …
3          14.8               2700.0         184.0              73.0  …
4          12.8               2505.0         161.0              72.0  …
5          10.1               2202.0         145.0              72.0  …
..          …                  …             …                  …  …
634        15.5               2753.0         179.0              64.0  …
635         4.8               1794.0         119.0              57.0  …
636         6.1               1871.0         153.0              58.0  …
637        16.7               2863.0         175.0              69.0  …
638        10.3               2104.0         154.0              66.0  …

    Asleep duration (min)  In bed duration (min)  Light sleep duration (min)  \
1                   533.0                  570.0                       231.0
2                   433.0                  484.0                       298.0
3                   535.0                  571.0                       273.0
4                   488.0                  562.0                       221.0
```

```
5                   568.0                  623.0                       355.0
..                    …                    …                         …
634                 490.0                  557.0                       254.0
635                 481.0                  564.0                       258.0
636                 410.0                  487.0                       230.0
637                 482.0                  549.0                       182.0
638                 448.0                  520.0                       267.0

     Deep (SWS) duration (min)  REM duration (min)  Awake duration (min)  \
1                        137.0               165.0                  37.0
2                         95.0                40.0                  51.0
3                        126.0               136.0                  36.0
4                        123.0               144.0                  74.0
5                        131.0                82.0                  55.0
..                         …                 …                   …
634                      114.0               122.0                  67.0
635                      132.0                91.0                  83.0
636                       98.0                82.0                  77.0
637                      139.0               161.0                  67.0
638                      102.0                79.0                  72.0

     Sleep need (min)  Sleep debt (min)  Sleep efficiency %  \
1               544.0              31.0                93.0
2               495.0               0.0                89.0
3               483.0               0.0                94.0
4               472.0               0.0                89.0
5               501.0               0.0                92.0
..                …               …                 …
634             486.0              20.0                88.0
635             517.0              50.0                85.0
636             525.0              14.0                86.0
637             510.0              34.0                88.0
638             526.0              23.0                86.0

     Sleep consistency %
1                   77.0
2                   90.0
3                   87.0
4                   81.0
5                   81.0
..                   …
634                 87.0
635                 89.0
636                 89.0
637                 87.0
638                 84.0
```

```
[632 rows x 23 columns]
```

3.3 Dtypes

[120]:

```
df_phys_dt = df_phys_sel.copy()
df_phys_dt['Cycle start time'] = pd.to_datetime(df_phys_dt['Cycle start time'])

df_phys_dt['Recovery score %'] = df_phys_dt['Recovery score %'] / 100.
df_phys_dt['Blood oxygen %'] = df_phys_dt['Blood oxygen %'] / 100.
df_phys_dt['Sleep performance %'] = df_phys_dt['Sleep performance %'] / 100.
df_phys_dt['Sleep efficiency %'] = df_phys_dt['Sleep efficiency %'] / 100.
df_phys_dt['Sleep consistency %'] = df_phys_dt['Sleep consistency %'] / 100.
```

[121]:

```
df_phys_dt
```

[121]:

```
       Cycle start time  Recovery score %  Resting heart rate (bpm)  \
1   2024-06-15 23:12:59              0.84                      52.0
2   2024-06-14 23:39:45              0.94                      48.0
3   2024-06-14 00:07:47              0.83                      50.0
4   2024-06-13 00:12:09              0.73                      49.0
5   2024-06-11 23:11:54              0.74                      50.0
..                  …                 …                        …
634 2022-09-21 22:01:31              0.49                      46.0
635 2022-09-20 22:20:13              0.48                      47.0
636 2022-09-19 22:41:43              0.45                      49.0
637 2022-09-18 22:20:02              0.55                      49.0
638 2022-09-17 22:36:20              0.65                      52.0

     Heart rate variability (ms)  Skin temp (celsius)  Blood oxygen %  \
1                           92.0                33.30          0.9614
2                           95.0                34.30          0.9889
3                           87.0                34.00          0.9782
4                           80.0                33.20          0.9900
5                           78.0                33.90          0.9840
..                           …                  …               …
634                         68.0                31.83          0.9661
635                         65.0                32.80          0.9669
636                         67.0                32.44          0.9584
637                         68.0                33.30          0.9733
638                         85.0                32.69          0.9691

     Day Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  …  \
1          12.6               2520.0         153.0              73.0  …
2          17.3               3689.0         159.0              82.0  …
3          14.8               2700.0         184.0              73.0  …
4          12.8               2505.0         161.0              72.0  …
5          10.1               2202.0         145.0              72.0  …
```

```
..          …                  …           …               …  …
634        15.5               2753.0         179.0              64.0  …
635         4.8               1794.0         119.0              57.0  …
636         6.1               1871.0         153.0              58.0  …
637        16.7               2863.0         175.0              69.0  …
638        10.3               2104.0         154.0              66.0  …

    Asleep duration (min)  In bed duration (min)  Light sleep duration (min)  \
1                   533.0                  570.0                       231.0
2                   433.0                  484.0                       298.0
3                   535.0                  571.0                       273.0
4                   488.0                  562.0                       221.0
5                   568.0                  623.0                       355.0
..                    …                    …                         …
634                 490.0                  557.0                       254.0
635                 481.0                  564.0                       258.0
636                 410.0                  487.0                       230.0
637                 482.0                  549.0                       182.0
638                 448.0                  520.0                       267.0

     Deep (SWS) duration (min)  REM duration (min)  Awake duration (min)  \
1                        137.0               165.0                  37.0
2                         95.0                40.0                  51.0
3                        126.0               136.0                  36.0
4                        123.0               144.0                  74.0
5                        131.0                82.0                  55.0
..                         …                  …                   …
634                      114.0               122.0                  67.0
635                      132.0                91.0                  83.0
636                       98.0                82.0                  77.0
637                      139.0               161.0                  67.0
638                      102.0                79.0                  72.0

     Sleep need (min)  Sleep debt (min)  Sleep efficiency %  \
1               544.0              31.0                0.93
2               495.0               0.0                0.89
3               483.0               0.0                0.94
4               472.0               0.0                0.89
5               501.0               0.0                0.92
..                …               …                  …
634             486.0              20.0                0.88
635             517.0              50.0                0.85
636             525.0              14.0                0.86
637             510.0              34.0                0.88
638             526.0              23.0                0.86

     Sleep consistency %
```

```
1                    0.77
2                    0.90
3                    0.87
4                    0.81
5                    0.81
..                    …
634                  0.87
635                  0.89
636                  0.89
637                  0.87
638                  0.84

[632 rows x 23 columns]
```

[122]:

```
df_phys_dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 632 entries, 1 to 638
Data columns (total 23 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   Cycle start time               632 non-null    datetime64[ns]
 1   Recovery score %               632 non-null    float64
 2   Resting heart rate (bpm)       632 non-null    float64
 3   Heart rate variability (ms)    632 non-null    float64
 4   Skin temp (celsius)            632 non-null    float64
 5   Blood oxygen %                 632 non-null    float64
 6   Day Strain                     632 non-null    float64
 7   Energy burned (cal)            632 non-null    float64
 8   Max HR (bpm)                   632 non-null    float64
 9   Average HR (bpm)               632 non-null    float64
 10  Sleep onset                    632 non-null    object
 11  Sleep performance %            632 non-null    float64
 12  Respiratory rate (rpm)         632 non-null    float64
 13  Asleep duration (min)          632 non-null    float64
 14  In bed duration (min)          632 non-null    float64
 15  Light sleep duration (min)     632 non-null    float64
 16  Deep (SWS) duration (min)      632 non-null    float64
 17  REM duration (min)             632 non-null    float64
 18  Awake duration (min)           632 non-null    float64
 19  Sleep need (min)               632 non-null    float64
 20  Sleep debt (min)               632 non-null    float64
 21  Sleep efficiency %             632 non-null    float64
 22  Sleep consistency %            632 non-null    float64
dtypes: datetime64[ns](1), float64(21), object(1)
memory usage: 118.5+ KB
```

3.4 Add cols

[123]:

```
df_phys_dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 632 entries, 1 to 638
Data columns (total 23 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Cycle start time              632 non-null    datetime64[ns]
 1   Recovery score %              632 non-null    float64
 2   Resting heart rate (bpm)      632 non-null    float64
 3   Heart rate variability (ms)   632 non-null    float64
 4   Skin temp (celsius)           632 non-null    float64
 5   Blood oxygen %                632 non-null    float64
 6   Day Strain                    632 non-null    float64
 7   Energy burned (cal)           632 non-null    float64
 8   Max HR (bpm)                  632 non-null    float64
 9   Average HR (bpm)              632 non-null    float64
 10  Sleep onset                   632 non-null    object
 11  Sleep performance %           632 non-null    float64
 12  Respiratory rate (rpm)        632 non-null    float64
 13  Asleep duration (min)         632 non-null    float64
 14  In bed duration (min)         632 non-null    float64
 15  Light sleep duration (min)    632 non-null    float64
 16  Deep (SWS) duration (min)     632 non-null    float64
 17  REM duration (min)            632 non-null    float64
 18  Awake duration (min)          632 non-null    float64
 19  Sleep need (min)              632 non-null    float64
 20  Sleep debt (min)              632 non-null    float64
 21  Sleep efficiency %            632 non-null    float64
 22  Sleep consistency %           632 non-null    float64
dtypes: datetime64[ns](1), float64(21), object(1)
memory usage: 118.5+ KB
```

[124]:

```
df_phys_add = df_phys_dt.copy()
df_phys_add['Awake %'] = df_phys_add['Awake duration (min)'] / df_phys_add['In␣
  ↪bed duration (min)']
df_phys_add['Light sleep %'] = df_phys_add['Light sleep duration (min)'] /␣
  ↪df_phys_add['In bed duration (min)']
df_phys_add['Deep (SWS) %'] = df_phys_add['Deep (SWS) duration (min)'] /␣
  ↪df_phys_add['In bed duration (min)']
df_phys_add['REM %'] = df_phys_add['REM duration (min)'] / df_phys_add['In bed␣
  ↪duration (min)']
df_phys_add['Restorative sleep duration (min)'] = df_phys_add['Deep (SWS)␣
  ↪duration (min)'] + df_phys_add['REM duration (min)']
df_phys_add['Restorative sleep %'] = df_phys_add['Restorative sleep duration␣
  ↪(min)'] / df_phys_add['In bed duration (min)']
```

```
df_phys_add['cycle_start_time_adj'] = df_phys_add['Cycle start time'] - pd.
  ↪Timedelta(hours=9)
df_phys_add['cycle_date'] = df_phys_add['cycle_start_time_adj'].dt.date + pd.
  ↪Timedelta(days=1) # cycle starts at the beginning of sleep the previous␣
  ↪night -- adjustment matches metrics in app
```

```
[125]: df_phys_add
```

```
[125]:        Cycle start time  Recovery score %  Resting heart rate (bpm)  \
       1   2024-06-15 23:12:59              0.84                      52.0
       2   2024-06-14 23:39:45              0.94                      48.0
       3   2024-06-14 00:07:47              0.83                      50.0
       4   2024-06-13 00:12:09              0.73                      49.0
       5   2024-06-11 23:11:54              0.74                      50.0
       ..                  …                 …                         …
       634 2022-09-21 22:01:31              0.49                      46.0
       635 2022-09-20 22:20:13              0.48                      47.0
       636 2022-09-19 22:41:43              0.45                      49.0
       637 2022-09-18 22:20:02              0.55                      49.0
       638 2022-09-17 22:36:20              0.65                      52.0

            Heart rate variability (ms)  Skin temp (celsius)  Blood oxygen %  \
       1                           92.0                33.30          0.9614
       2                           95.0                34.30          0.9889
       3                           87.0                34.00          0.9782
       4                           80.0                33.20          0.9900
       5                           78.0                33.90          0.9840
       ..                           …                   …               …
       634                         68.0                31.83          0.9661
       635                         65.0                32.80          0.9669
       636                         67.0                32.44          0.9584
       637                         68.0                33.30          0.9733
       638                         85.0                32.69          0.9691

            Day Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  …  \
       1          12.6               2520.0         153.0              73.0  …
       2          17.3               3689.0         159.0              82.0  …
       3          14.8               2700.0         184.0              73.0  …
       4          12.8               2505.0         161.0              72.0  …
       5          10.1               2202.0         145.0              72.0  …
       ..          …                  …            …                …  …
       634        15.5               2753.0         179.0              64.0  …
       635         4.8               1794.0         119.0              57.0  …
       636         6.1               1871.0         153.0              58.0  …
       637        16.7               2863.0         175.0              69.0  …
       638        10.3               2104.0         154.0              66.0  …
```

```
    Sleep efficiency %  Sleep consistency %   Awake %  Light sleep %  \
1                 0.93                 0.77  0.064912       0.405263
2                 0.89                 0.90  0.105372       0.615702
3                 0.94                 0.87  0.063047       0.478109
4                 0.89                 0.81  0.131673       0.393238
5                 0.92                 0.81  0.088283       0.569823
..                 …                  …       …            …
634               0.88                 0.87  0.120287       0.456014
635               0.85                 0.89  0.147163       0.457447
636               0.86                 0.89  0.158111       0.472279
637               0.88                 0.87  0.122040       0.331512
638               0.86                 0.84  0.138462       0.513462

     Deep (SWS) %     REM %  Restorative sleep duration (min)  \
1        0.240351  0.289474                             302.0
2        0.196281  0.082645                             135.0
3        0.220665  0.238179                             262.0
4        0.218861  0.256228                             267.0
5        0.210273  0.131621                             213.0
..            …       …                               …
634      0.204668  0.219031                             236.0
635      0.234043  0.161348                             223.0
636      0.201232  0.168378                             180.0
637      0.253188  0.293260                             300.0
638      0.196154  0.151923                             181.0

     Restorative sleep %  cycle_start_time_adj  cycle_date
1               0.529825   2024-06-15 14:12:59  2024-06-16
2               0.278926   2024-06-14 14:39:45  2024-06-15
3               0.458844   2024-06-13 15:07:47  2024-06-14
4               0.475089   2024-06-12 15:12:09  2024-06-13
5               0.341894   2024-06-11 14:11:54  2024-06-12
..                  …                    …         …
634             0.423698   2022-09-21 13:01:31  2022-09-22
635             0.395390   2022-09-20 13:20:13  2022-09-21
636             0.369610   2022-09-19 13:41:43  2022-09-20
637             0.546448   2022-09-18 13:20:02  2022-09-19
638             0.348077   2022-09-17 13:36:20  2022-09-18

[632 rows x 31 columns]
```

3.5 Recovery vs Cycle Metrics

The Physiological Cycles data includes two categories of metrics.

First, it includes Recovery metrics that are calculated during the night of sleep at the beginning of the cycle. These metrics will be influenced by activites and strain metrics measured on the previous day.

Second, it includes activity and strain metrics that are measured on the day of the cycle. These metrics will be combined with journal entries and logged activities to calculate the recovery for the following cycle.

These two categories of metrics should be considered separately, and each be given appropriate date labels.

Finally, we re-merge the recovery and activity metrics such that the recovery metrics are paired with their respective causal activity metrics.

[126]:

```
df_phys_add.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 632 entries, 1 to 638
Data columns (total 31 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   Cycle start time                  632 non-null    datetime64[ns]
 1   Recovery score %                  632 non-null    float64
 2   Resting heart rate (bpm)          632 non-null    float64
 3   Heart rate variability (ms)       632 non-null    float64
 4   Skin temp (celsius)               632 non-null    float64
 5   Blood oxygen %                    632 non-null    float64
 6   Day Strain                        632 non-null    float64
 7   Energy burned (cal)               632 non-null    float64
 8   Max HR (bpm)                      632 non-null    float64
 9   Average HR (bpm)                  632 non-null    float64
 10  Sleep onset                       632 non-null    object
 11  Sleep performance %               632 non-null    float64
 12  Respiratory rate (rpm)            632 non-null    float64
 13  Asleep duration (min)             632 non-null    float64
 14  In bed duration (min)             632 non-null    float64
 15  Light sleep duration (min)        632 non-null    float64
 16  Deep (SWS) duration (min)         632 non-null    float64
 17  REM duration (min)                632 non-null    float64
 18  Awake duration (min)              632 non-null    float64
 19  Sleep need (min)                  632 non-null    float64
 20  Sleep debt (min)                  632 non-null    float64
 21  Sleep efficiency %                632 non-null    float64
 22  Sleep consistency %               632 non-null    float64
 23  Awake %                           632 non-null    float64
 24  Light sleep %                     632 non-null    float64
 25  Deep (SWS) %                      632 non-null    float64
 26  REM %                             632 non-null    float64
 27  Restorative sleep duration (min)  632 non-null    float64
 28  Restorative sleep %               632 non-null    float64
 29  cycle_start_time_adj              632 non-null    datetime64[ns]
 30  cycle_date                        632 non-null    object
```

```
dtypes: datetime64[ns](2), float64(27), object(2)
memory usage: 158.0+ KB
```

[129]:

```
PHYS_REC_METRICS =[
    'Cycle start time',
    'Recovery score %',
    'Resting heart rate (bpm)',
    'Heart rate variability (ms)',
    'Skin temp (celsius)',
    'Blood oxygen %',
    'Sleep onset',
    'Sleep performance %',
    'Respiratory rate (rpm)',
    'Asleep duration (min)',
    'In bed duration (min)',
    'Light sleep duration (min)',
    'Light sleep %',
    'Deep (SWS) duration (min)',
    'Deep (SWS) %',
    'REM duration (min)',
    'REM %',
    'Awake duration (min)',
    'Awake %',
    'Restorative sleep duration (min)',
    'Restorative sleep %',
    'Sleep need (min)',
    'Sleep debt (min)',
    'Sleep efficiency %',
    'Sleep consistency %',
    'cycle_start_time_adj',
    'cycle_date'
]

PHYS_CYCLE_METRICS =[
    'Cycle start time',
    'Day Strain',
    'Energy burned (cal)',
    'Max HR (bpm)',
    'Average HR (bpm)',
    'cycle_start_time_adj',
    'cycle_date'
]
```

[130]:

```
df_phys_rec = df_phys_add[PHYS_REC_METRICS].copy()
df_phys_rec['recovery_date'] = df_phys_rec['cycle_date'] - pd.Timedelta(days=1)␣
 ↪# set recovery cycle_date to the date where relevant activities and strain␣
 ↪metrics are recorded
```

```
df_phys_cycle = df_phys_add[PHYS_CYCLE_METRICS].copy()
```

[131]:
```
df_phys_rec
```

[131]:
```
        Cycle start time  Recovery score %  Resting heart rate (bpm)  \
1    2024-06-15 23:12:59              0.84                      52.0
2    2024-06-14 23:39:45              0.94                      48.0
3    2024-06-14 00:07:47              0.83                      50.0
4    2024-06-13 00:12:09              0.73                      49.0
5    2024-06-11 23:11:54              0.74                      50.0
..                   …                 …                        …
634  2022-09-21 22:01:31              0.49                      46.0
635  2022-09-20 22:20:13              0.48                      47.0
636  2022-09-19 22:41:43              0.45                      49.0
637  2022-09-18 22:20:02              0.55                      49.0
638  2022-09-17 22:36:20              0.65                      52.0

     Heart rate variability (ms)  Skin temp (celsius)  Blood oxygen %  \
1                           92.0                33.30          0.9614
2                           95.0                34.30          0.9889
3                           87.0                34.00          0.9782
4                           80.0                33.20          0.9900
5                           78.0                33.90          0.9840
..                           …                   …              …
634                         68.0                31.83          0.9661
635                         65.0                32.80          0.9669
636                         67.0                32.44          0.9584
637                         68.0                33.30          0.9733
638                         85.0                32.69          0.9691

             Sleep onset  Sleep performance %  Respiratory rate (rpm)  \
1    2024-06-15 23:12:59                 0.98                    14.8
2    2024-06-14 23:39:45                 0.87                    14.0
3    2024-06-14 00:07:47                 1.00                    14.4
4    2024-06-13 00:12:09                 1.00                    14.8
5    2024-06-11 23:11:54                 1.00                    14.3
..                   …                    …                      …
634  2022-09-21 22:01:31                 1.00                    14.2
635  2022-09-20 22:20:13                 0.93                    13.7
636  2022-09-19 22:41:43                 0.78                    13.7
637  2022-09-18 22:20:02                 0.95                    14.0
638  2022-09-17 22:36:20                 0.85                    13.8

     Asleep duration (min)  …   Awake %  Restorative sleep duration (min)  \
1                    533.0  …  0.064912                             302.0
2                    433.0  …  0.105372                             135.0
3                    535.0  …  0.063047                             262.0
```

```
4                    488.0  …  0.131673                             267.0
5                    568.0  …  0.088283                             213.0
..                     …  …       …                               …
634                  490.0  …  0.120287                             236.0
635                  481.0  …  0.147163                             223.0
636                  410.0  …  0.158111                             180.0
637                  482.0  …  0.122040                             300.0
638                  448.0  …  0.138462                             181.0

     Restorative sleep %  Sleep need (min)  Sleep debt (min)  \
1               0.529825             544.0              31.0
2               0.278926             495.0               0.0
3               0.458844             483.0               0.0
4               0.475089             472.0               0.0
5               0.341894             501.0               0.0
..                  …                …               …
634             0.423698             486.0              20.0
635             0.395390             517.0              50.0
636             0.369610             525.0              14.0
637             0.546448             510.0              34.0
638             0.348077             526.0              23.0

     Sleep efficiency %  Sleep consistency %  cycle_start_time_adj  \
1                  0.93                 0.77   2024-06-15 14:12:59
2                  0.89                 0.90   2024-06-14 14:39:45
3                  0.94                 0.87   2024-06-13 15:07:47
4                  0.89                 0.81   2024-06-12 15:12:09
5                  0.92                 0.81   2024-06-11 14:11:54
..                  …                  …                   …
634                0.88                 0.87   2022-09-21 13:01:31
635                0.85                 0.89   2022-09-20 13:20:13
636                0.86                 0.89   2022-09-19 13:41:43
637                0.88                 0.87   2022-09-18 13:20:02
638                0.86                 0.84   2022-09-17 13:36:20

     cycle_date  recovery_date
1    2024-06-16     2024-06-15
2    2024-06-15     2024-06-14
3    2024-06-14     2024-06-13
4    2024-06-13     2024-06-12
5    2024-06-12     2024-06-11
..          …            …
634  2022-09-22     2022-09-21
635  2022-09-21     2022-09-20
636  2022-09-20     2022-09-19
637  2022-09-19     2022-09-18
638  2022-09-18     2022-09-17
```

```
[632 rows x 28 columns]
```

[132]:

```
df_phys_cycle
```

[132]:

```
       Cycle start time  Day Strain  Energy burned (cal)  Max HR (bpm)  \
1   2024-06-15 23:12:59        12.6               2520.0         153.0
2   2024-06-14 23:39:45        17.3               3689.0         159.0
3   2024-06-14 00:07:47        14.8               2700.0         184.0
4   2024-06-13 00:12:09        12.8               2505.0         161.0
5   2024-06-11 23:11:54        10.1               2202.0         145.0
..                  …         …                  …           …
634 2022-09-21 22:01:31        15.5               2753.0         179.0
635 2022-09-20 22:20:13         4.8               1794.0         119.0
636 2022-09-19 22:41:43         6.1               1871.0         153.0
637 2022-09-18 22:20:02        16.7               2863.0         175.0
638 2022-09-17 22:36:20        10.3               2104.0         154.0

     Average HR (bpm) cycle_start_time_adj  cycle_date
1                 73.0  2024-06-15 14:12:59  2024-06-16
2                 82.0  2024-06-14 14:39:45  2024-06-15
3                 73.0  2024-06-13 15:07:47  2024-06-14
4                 72.0  2024-06-12 15:12:09  2024-06-13
5                 72.0  2024-06-11 14:11:54  2024-06-12
..                …                  …         …
634               64.0  2022-09-21 13:01:31  2022-09-22
635               57.0  2022-09-20 13:20:13  2022-09-21
636               58.0  2022-09-19 13:41:43  2022-09-20
637               69.0  2022-09-18 13:20:02  2022-09-19
638               66.0  2022-09-17 13:36:20  2022-09-18

[632 rows x 7 columns]
```

[133]:

```
df_phys_comb = pd.merge(
    df_phys_rec,
    df_phys_cycle,
    left_on='recovery_date',
    right_on='cycle_date',
    suffixes=('_rec', '_cycle')
)
df_phys_comb.drop(
    columns=[col for col in df_phys_comb.columns if col.endswith('_rec')],
    inplace=True
)
df_phys_comb.columns = [col.replace('_cycle', '') for col in df_phys_comb.
 ↪columns]
```

```
df_phys_comb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 625 entries, 0 to 624
Data columns (total 32 columns):
 #   Column                            Non-Null Count  Dtype
---  ------                            --------------  -----
 0   Recovery score %                  625 non-null    float64
 1   Resting heart rate (bpm)          625 non-null    float64
 2   Heart rate variability (ms)       625 non-null    float64
 3   Skin temp (celsius)               625 non-null    float64
 4   Blood oxygen %                    625 non-null    float64
 5   Sleep onset                       625 non-null    object
 6   Sleep performance %               625 non-null    float64
 7   Respiratory rate (rpm)            625 non-null    float64
 8   Asleep duration (min)             625 non-null    float64
 9   In bed duration (min)             625 non-null    float64
 10  Light sleep duration (min)        625 non-null    float64
 11  Light sleep %                     625 non-null    float64
 12  Deep (SWS) duration (min)         625 non-null    float64
 13  Deep (SWS) %                      625 non-null    float64
 14  REM duration (min)                625 non-null    float64
 15  REM %                             625 non-null    float64
 16  Awake duration (min)              625 non-null    float64
 17  Awake %                           625 non-null    float64
 18  Restorative sleep duration (min)  625 non-null    float64
 19  Restorative sleep %               625 non-null    float64
 20  Sleep need (min)                  625 non-null    float64
 21  Sleep debt (min)                  625 non-null    float64
 22  Sleep efficiency %                625 non-null    float64
 23  Sleep consistency %               625 non-null    float64
 24  recovery_date                     625 non-null    object
 25  Cycle start time                  625 non-null    datetime64[ns]
 26  Day Strain                        625 non-null    float64
 27  Energy burned (cal)               625 non-null    float64
 28  Max HR (bpm)                      625 non-null    float64
 29  Average HR (bpm)                  625 non-null    float64
 30  cycle_start_time_adj              625 non-null    datetime64[ns]
 31  cycle_date                        625 non-null    object
dtypes: datetime64[ns](2), float64(27), object(3)
memory usage: 156.4+ KB
```

```
[134]: df_phys_comb
```

```
[134]:     Recovery score %  Resting heart rate (bpm)  Heart rate variability (ms)  \
       0                0.84                      52.0                         92.0
       1                0.94                      48.0                         95.0
       2                0.83                      50.0                         87.0
```

```
3                0.73                      49.0                         80.0
4                0.74                      50.0                         78.0
..                …                       …                          …
620              0.57                      48.0                         74.0
621              0.49                      46.0                         68.0
622              0.48                      47.0                         65.0
623              0.45                      49.0                         67.0
624              0.55                      49.0                         68.0

     Skin temp (celsius)  Blood oxygen %          Sleep onset  \
0                  33.30          0.9614  2024-06-15 23:12:59
1                  34.30          0.9889  2024-06-14 23:39:45
2                  34.00          0.9782  2024-06-14 00:07:47
3                  33.20          0.9900  2024-06-13 00:12:09
4                  33.90          0.9840  2024-06-11 23:11:54
..                   …             …                  …
620                33.13          0.9581  2022-09-22 23:23:45
621                31.83          0.9661  2022-09-21 22:01:31
622                32.80          0.9669  2022-09-20 22:20:13
623                32.44          0.9584  2022-09-19 22:41:43
624                33.30          0.9733  2022-09-18 22:20:02

     Sleep performance %  Respiratory rate (rpm)  Asleep duration (min)  \
0                   0.98                    14.8                  533.0
1                   0.87                    14.0                  433.0
2                   1.00                    14.4                  535.0
3                   1.00                    14.8                  488.0
4                   1.00                    14.3                  568.0
..                   …                     …                    …
620                 0.93                    13.7                  453.0
621                 1.00                    14.2                  490.0
622                 0.93                    13.7                  481.0
623                 0.78                    13.7                  410.0
624                 0.95                    14.0                  482.0

     In bed duration (min)  …  Sleep efficiency %  Sleep consistency %  \
0                    570.0  …                0.93                 0.77
1                    484.0  …                0.89                 0.90
2                    571.0  …                0.94                 0.87
3                    562.0  …                0.89                 0.81
4                    623.0  …                0.92                 0.81
..                     …  …                 …                  …
620                  494.0  …                0.92                 0.79
621                  557.0  …                0.88                 0.87
622                  564.0  …                0.85                 0.89
623                  487.0  …                0.86                 0.89
624                  549.0  …                0.88                 0.87
```

```
     recovery_date    Cycle start time  Day Strain  Energy burned (cal)  \
0       2024-06-15 2024-06-14 23:39:45        17.3               3689.0
1       2024-06-14 2024-06-14 00:07:47        14.8               2700.0
2       2024-06-13 2024-06-13 00:12:09        12.8               2505.0
3       2024-06-12 2024-06-11 23:11:54        10.1               2202.0
4       2024-06-11 2024-06-10 23:15:19        15.8               3029.0
..             …                 …         …                  …
620     2022-09-22 2022-09-21 22:01:31        15.5               2753.0
621     2022-09-21 2022-09-20 22:20:13         4.8               1794.0
622     2022-09-20 2022-09-19 22:41:43         6.1               1871.0
623     2022-09-19 2022-09-18 22:20:02        16.7               2863.0
624     2022-09-18 2022-09-17 22:36:20        10.3               2104.0

     Max HR (bpm)  Average HR (bpm)  cycle_start_time_adj  cycle_date
0           159.0              82.0   2024-06-14 14:39:45  2024-06-15
1           184.0              73.0   2024-06-13 15:07:47  2024-06-14
2           161.0              72.0   2024-06-12 15:12:09  2024-06-13
3           145.0              72.0   2024-06-11 14:11:54  2024-06-12
4           177.0              73.0   2024-06-10 14:15:19  2024-06-11
..            …               …                   …         …
620         179.0              64.0   2022-09-21 13:01:31  2022-09-22
621         119.0              57.0   2022-09-20 13:20:13  2022-09-21
622         153.0              58.0   2022-09-19 13:41:43  2022-09-20
623         175.0              69.0   2022-09-18 13:20:02  2022-09-19
624         154.0              66.0   2022-09-17 13:36:20  2022-09-18

[625 rows x 32 columns]
```

3.6 EDA

```
[135]: df_phys_comb.describe()
```

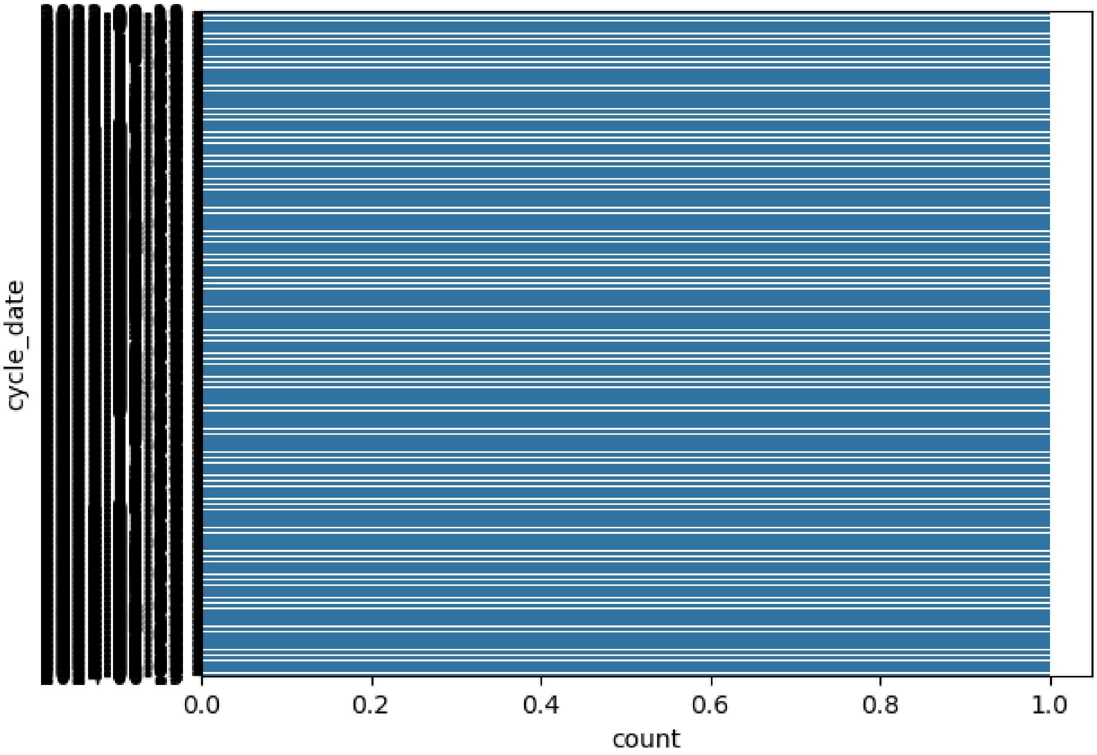
```
[135]:
       Recovery score %  Resting heart rate (bpm)  \
count        625.000000                625.000000
mean           0.651328                 52.003200
min            0.010000                 42.000000
25%            0.520000                 49.000000
50%            0.650000                 52.000000
75%            0.800000                 55.000000
max            0.990000                 75.000000
std            0.184094                  4.838745

       Heart rate variability (ms)  Skin temp (celsius)  Blood oxygen %  \
count                   625.000000           625.000000      625.000000
mean                     74.299200            33.907952        0.976309
min                      21.000000            31.300000        0.939000
```

```
25%                       66.000000             33.700000        0.970400
50%                       74.000000             34.000000        0.977000
75%                       82.000000             34.300000        0.983100
max                      133.000000             35.200000        0.998600
std                       13.472375              0.574598        0.009777

       Sleep performance %  Respiratory rate (rpm)  Asleep duration (min)  \
count           625.000000              625.000000             625.000000
mean              0.931024               14.503840             482.846400
min               0.540000               13.200000             268.000000
25%               0.890000               14.100000             453.000000
50%               0.960000               14.400000             486.000000
75%               1.000000               14.800000             515.000000
max               1.000000               17.300000             620.000000
std               0.082648                0.532698              49.748329

       In bed duration (min)  Light sleep duration (min)  …  \
count             625.000000                  625.000000  …
mean              539.080000                  229.862400  …
min               279.000000                  115.000000  …
25%               505.000000                  194.000000  …
50%               541.000000                  223.000000  …
75%               576.000000                  257.000000  …
max               680.000000                  546.000000  …
std                55.420508                   53.433522  …

       Sleep need (min)  Sleep debt (min)  Sleep efficiency %  \
count        625.000000        625.000000          625.000000
mean         505.200000         16.894400            0.900192
min          418.000000          0.000000            0.730000
25%          486.000000          0.000000            0.880000
50%          503.000000         11.000000            0.900000
75%          520.000000         27.000000            0.920000
max          630.000000        114.000000            0.970000
std           26.252656         19.976149            0.034175

       Sleep consistency %            Cycle start time  Day Strain  \
count           625.000000                         625  625.000000
mean              0.801552  2023-07-29 19:39:01.688000   12.869600
min               0.280000         2022-09-17 22:36:20    4.800000
25%               0.760000         2023-02-20 23:46:48   10.300000
50%               0.830000         2023-07-26 22:12:23   13.300000
75%               0.870000         2024-01-04 22:56:44   15.300000
max               0.950000         2024-06-14 23:39:45   20.400000
std               0.105093                         NaN    3.365517

       Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  \
count          625.000000   625.000000         625.000000
mean          2562.497600   158.748800          70.960000
min           1539.000000   119.000000          57.000000
25%           2147.000000   148.000000          68.000000
50%           2498.000000   158.000000          71.000000
75%           2871.000000   169.000000          74.000000
max           5251.000000   196.000000          91.000000
std            529.695695    15.483426           4.793826

             cycle_start_time_adj
count                         625
mean   2023-07-29 10:39:01.688000
min           2022-09-17 13:36:20
25%           2023-02-20 14:46:48
50%           2023-07-26 13:12:23
75%           2024-01-04 13:56:44
max           2024-06-14 14:39:45
std                           NaN

[8 rows x 29 columns]
```

3.6.1 Date Counts

```
[136]: phys_date_counts = df_phys_comb['cycle_date'].value_counts().sort_index().
        ↪reset_index()
       sns.barplot(data=phys_date_counts, x='count', y='cycle_date')
```

```
[136]: <Axes: xlabel='count', ylabel='cycle_date'>
```

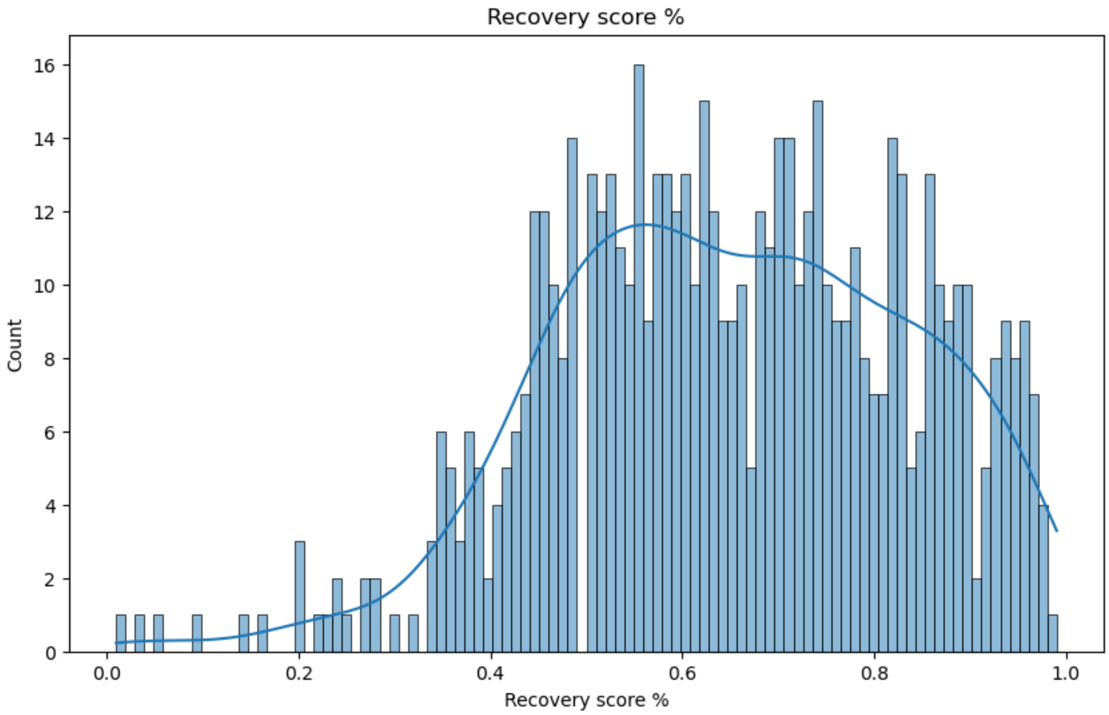


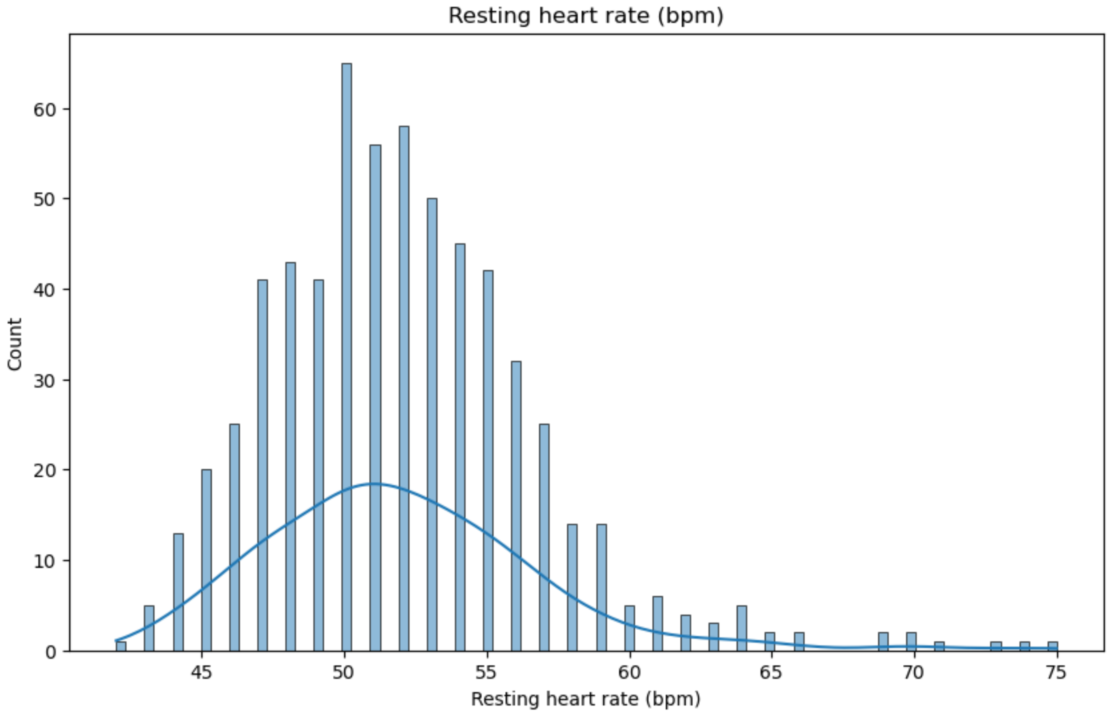
[137]:
```
phys_date_counts[phys_date_counts['count'] != 1]
```

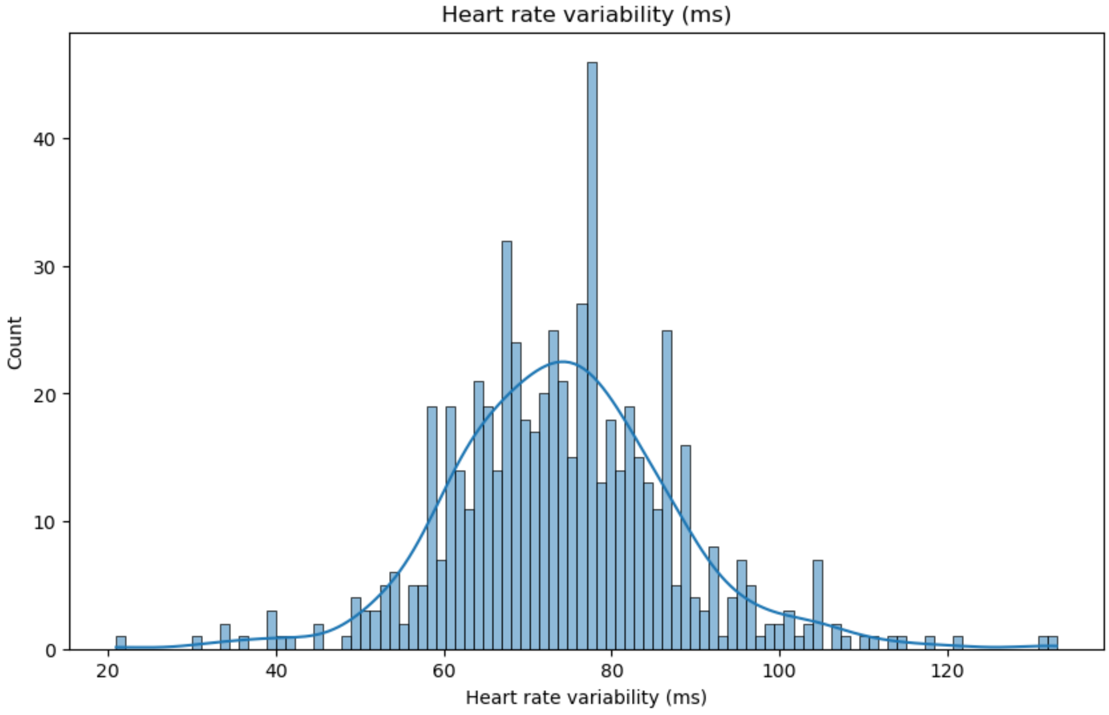
[137]:
```
Empty DataFrame
Columns: [cycle_date, count]
Index: []
```

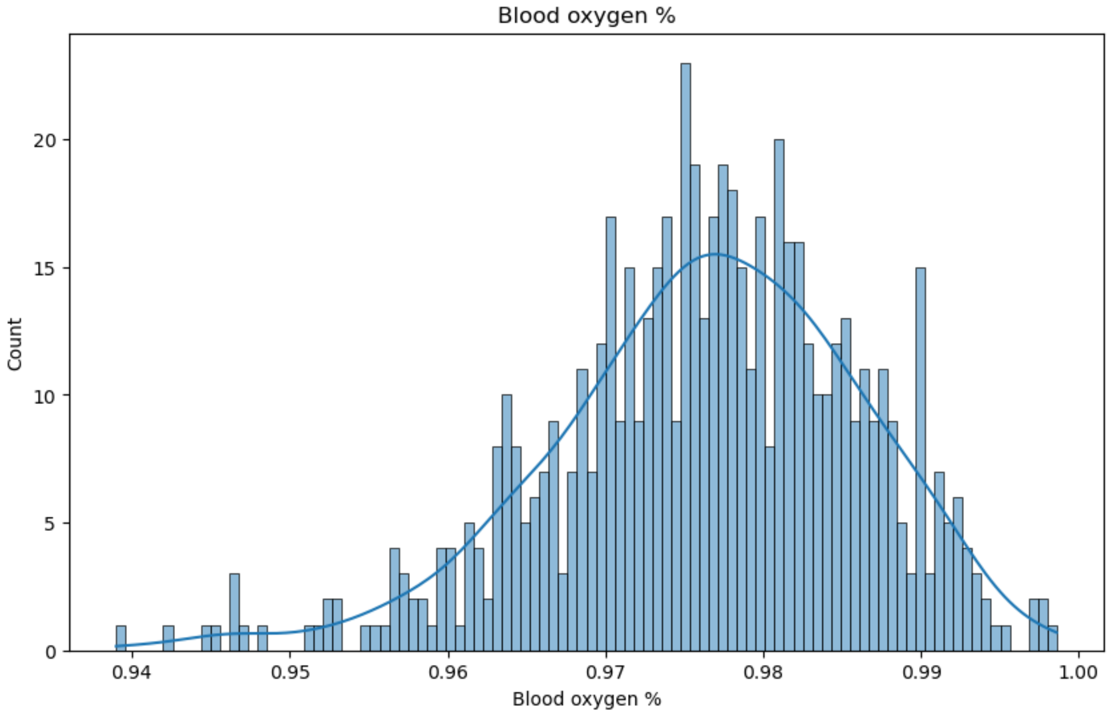
3.6.2 Distributions

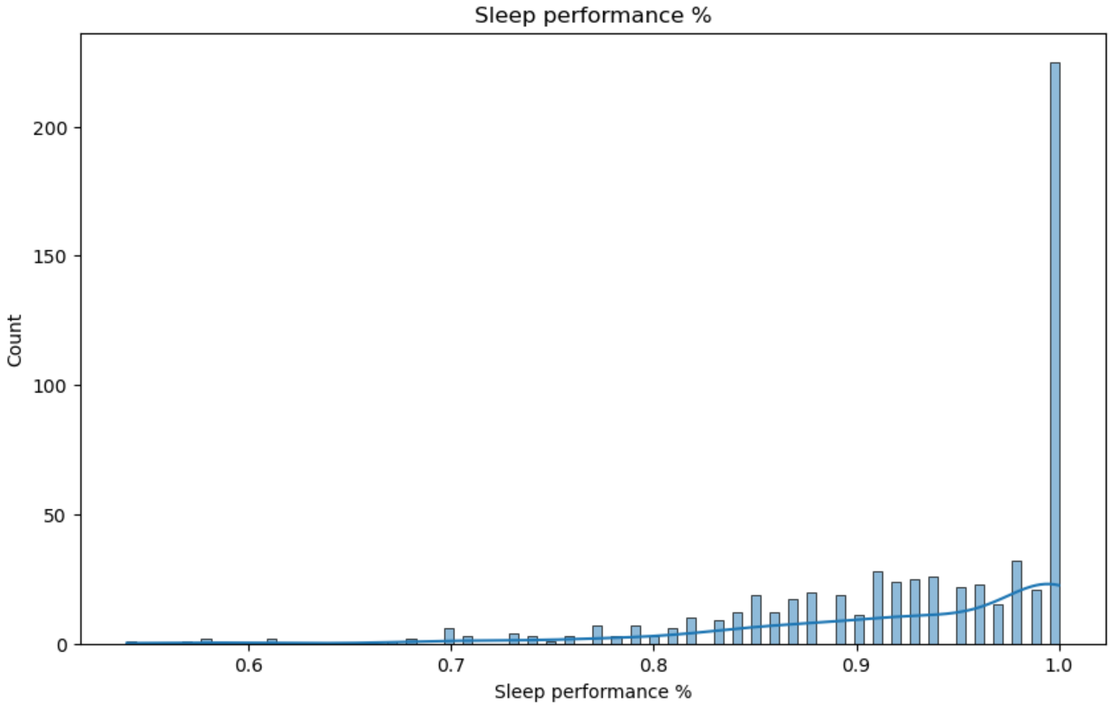
[138]:
```
PHYS_KDE_COLS = [
    # Recovery metrics
    'Recovery score %', # label
    'Resting heart rate (bpm)',
    'Heart rate variability (ms)',
    'Blood oxygen %',
    'Sleep performance %',
    'Asleep duration (min)', # corr w/ performance %
    'Sleep efficiency %',
    'Sleep consistency %',
    'Awake %',
    'Light sleep %',
    'Restorative sleep %',
    # Strain metrics
    'Day Strain', # label
    'Energy burned (cal)',
    'Max HR (bpm)',
    'Average HR (bpm)'
]

for column in PHYS_KDE_COLS:
    plt.figure(figsize=(10, 6))
    sns.histplot(df_phys_comb[column], bins=100, kde=True)
    plt.title(column)
    plt.show()
```

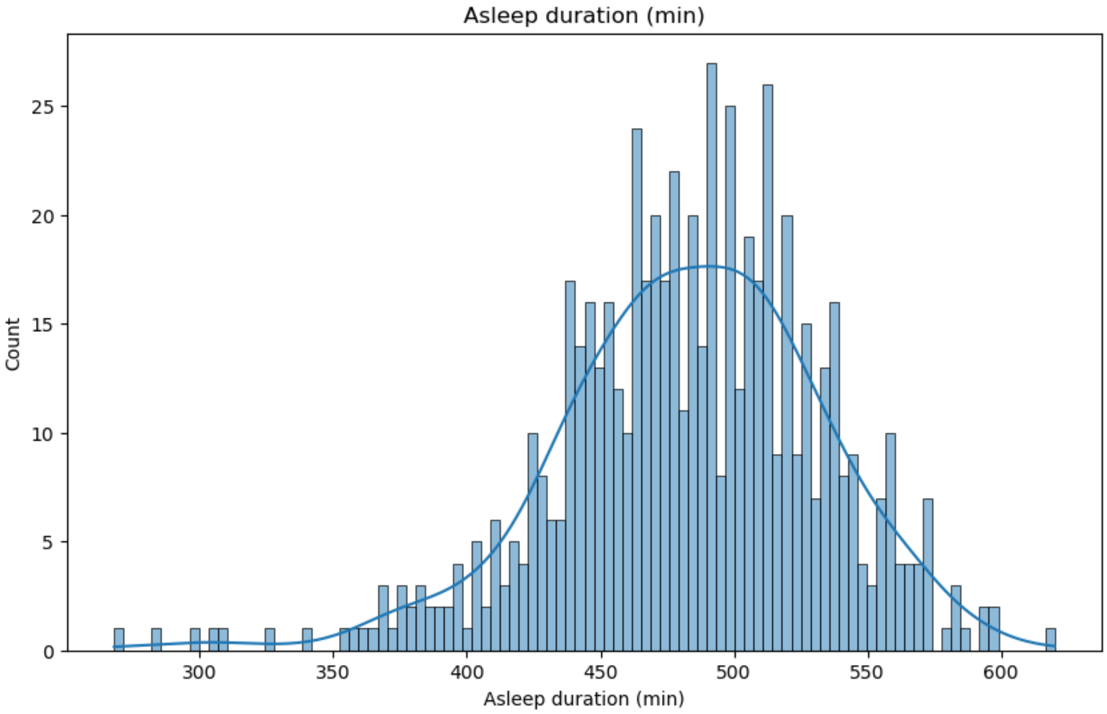


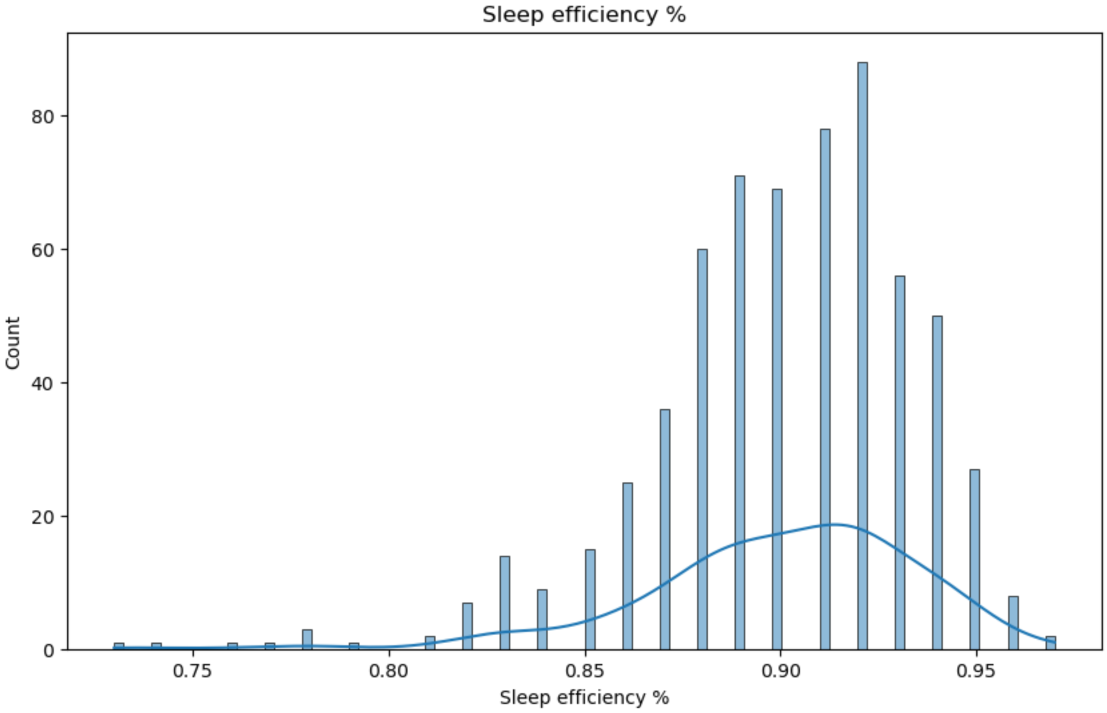

Resting heart rate (bpm)
Count
0
10
20
30
40
50
60
45
50
55
60
65
70
75
Resting heart rate (bpm)

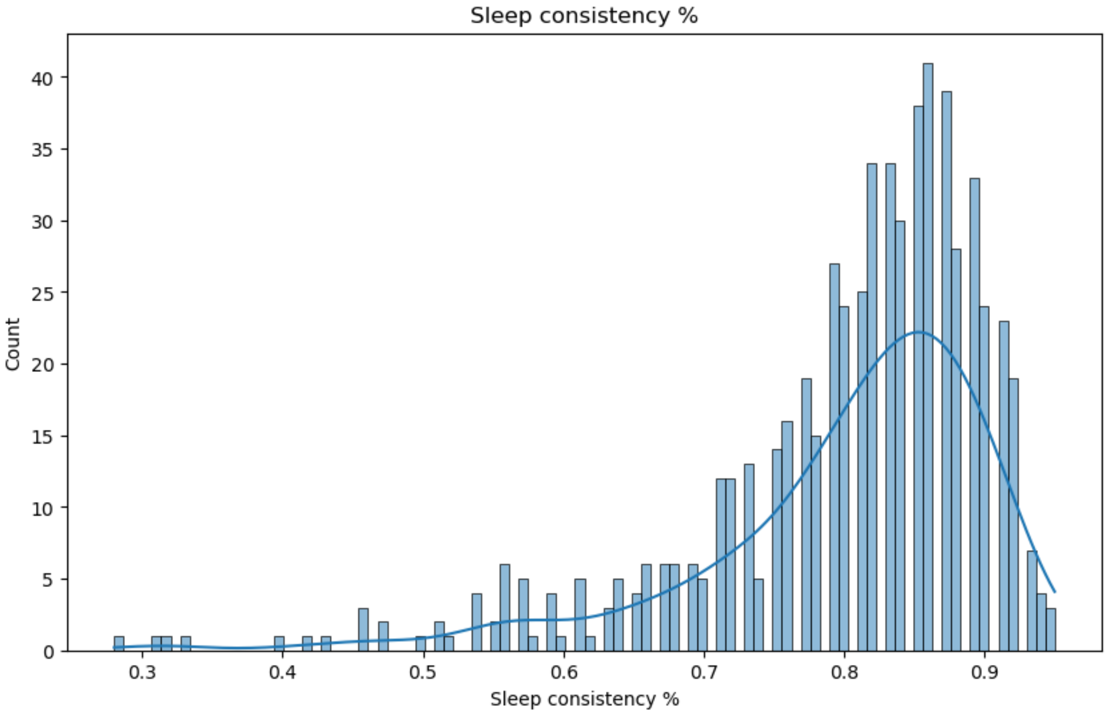

Heart rate variability (ms)
Count
0
10
20
30
40
20
40
60
80
100
120
Heart rate variability (ms)

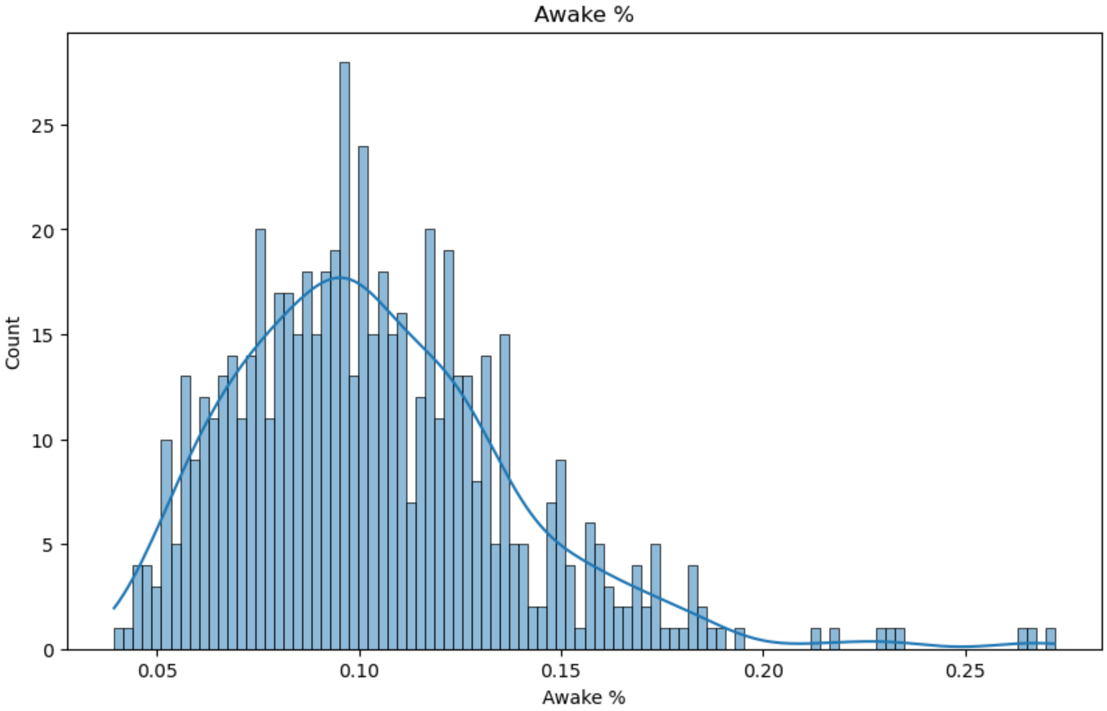

Blood oxygen %
Count
0
5
10
15
20
0.94
0.95
0.96
0.97
0.98
0.99
1.00
Blood oxygen %

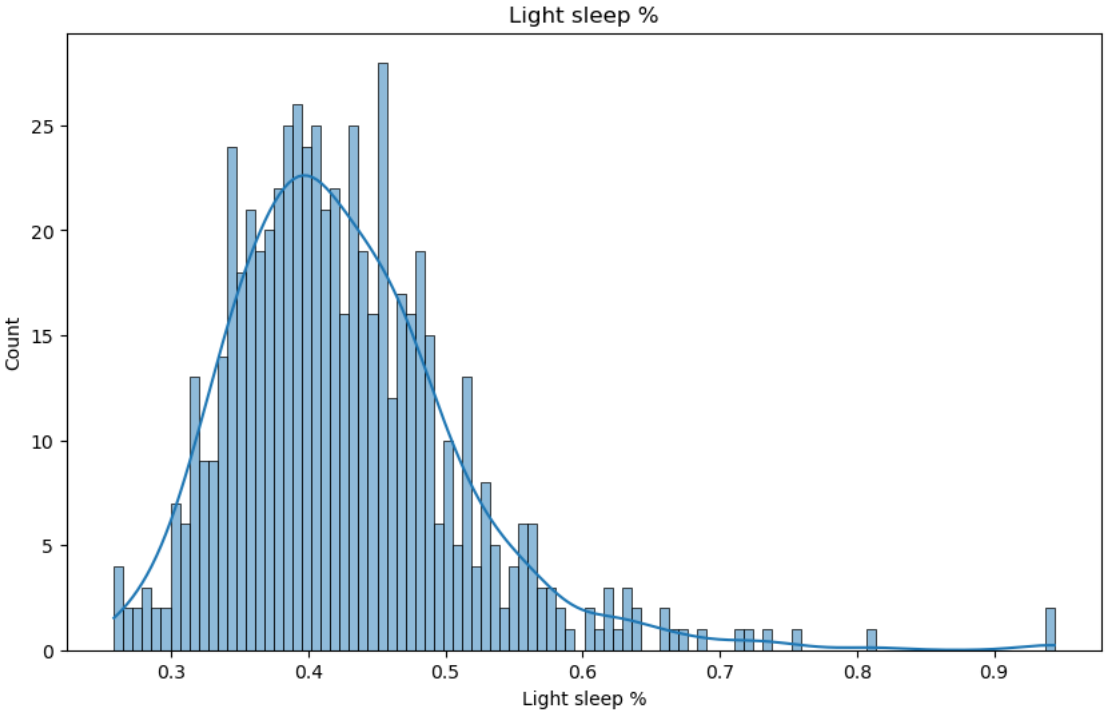

Sleep performance %
Count
0
50
100
150
200
0.6
0.7
0.8
0.9
1.0
Sleep performance %

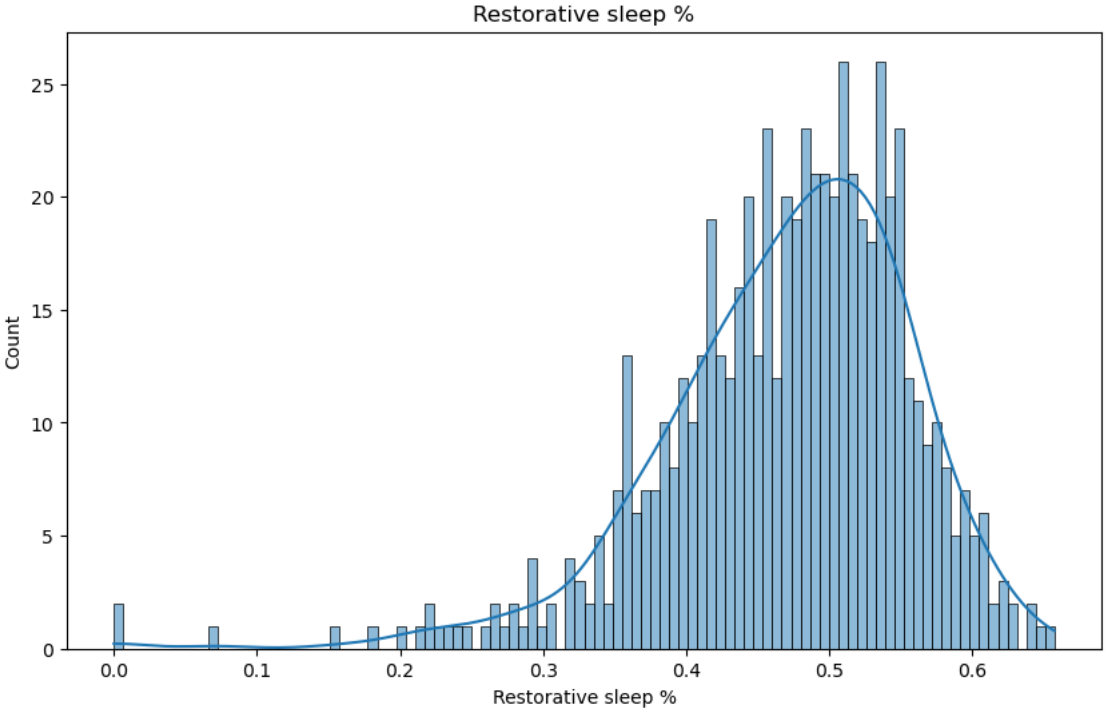

Asleep duration (min)
Count
0
5
10
15
20
25
300
350
400
450
500
550
600
Asleep duration (min)

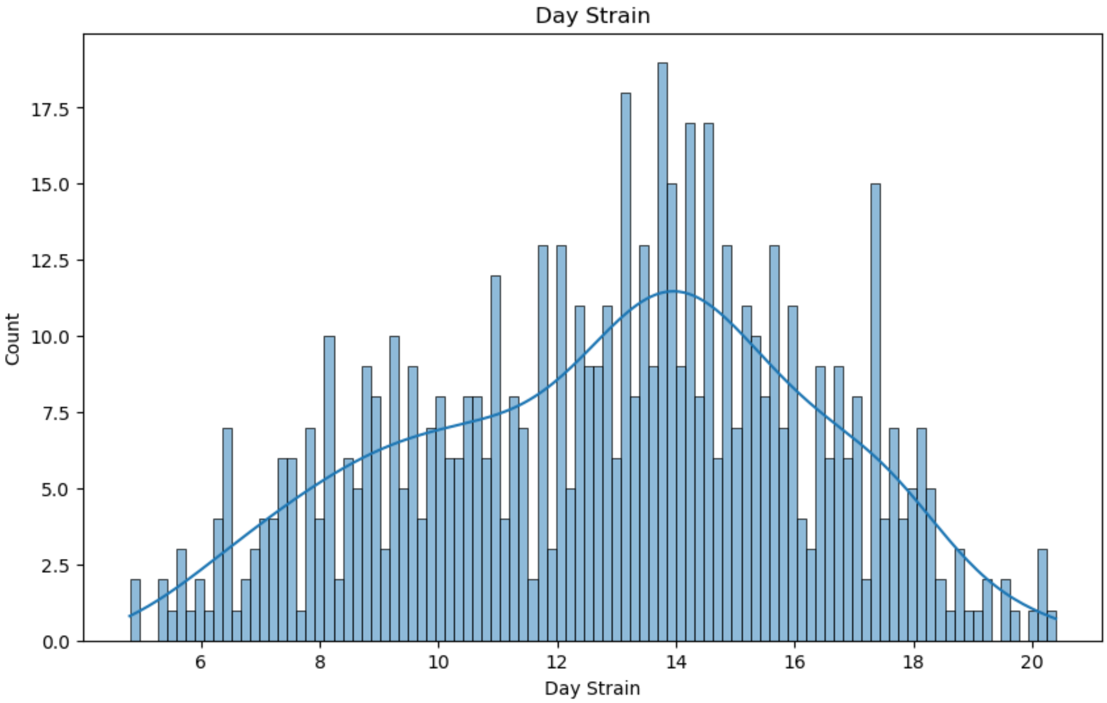

Sleep efficiency %
Count
0
20
40
60
80
0.75
0.80
0.85
0.90
0.95
Sleep efficiency %

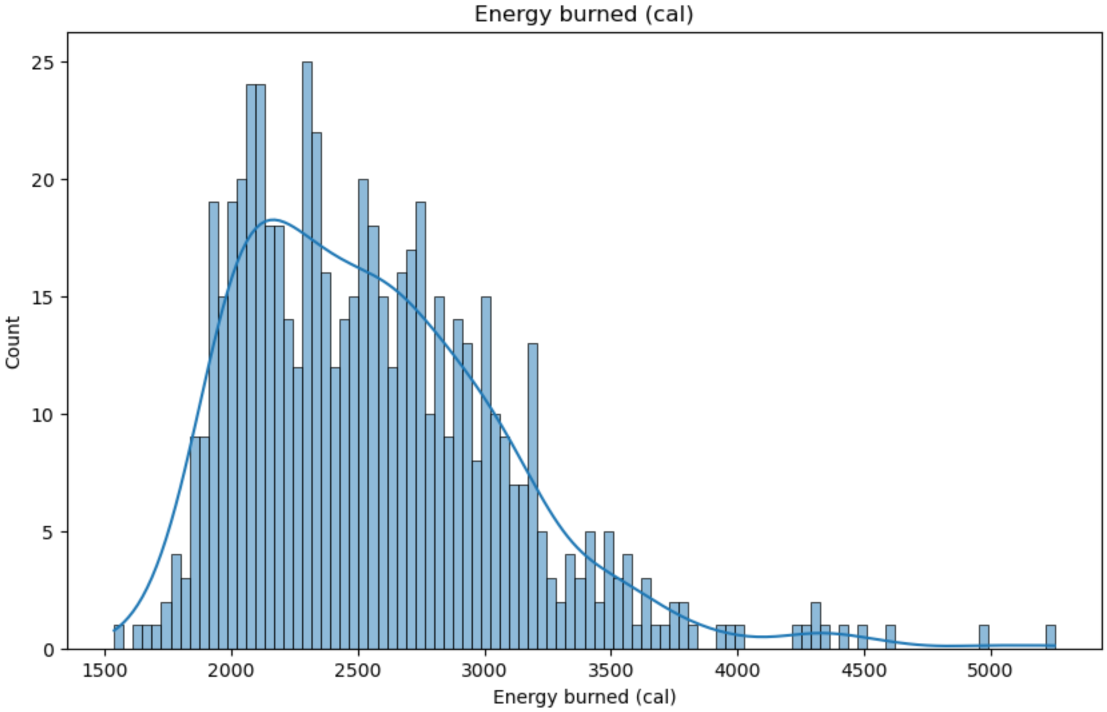

Sleep consistency %
Count
0
5
10
15
20
25
30
35
40
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Sleep consistency %

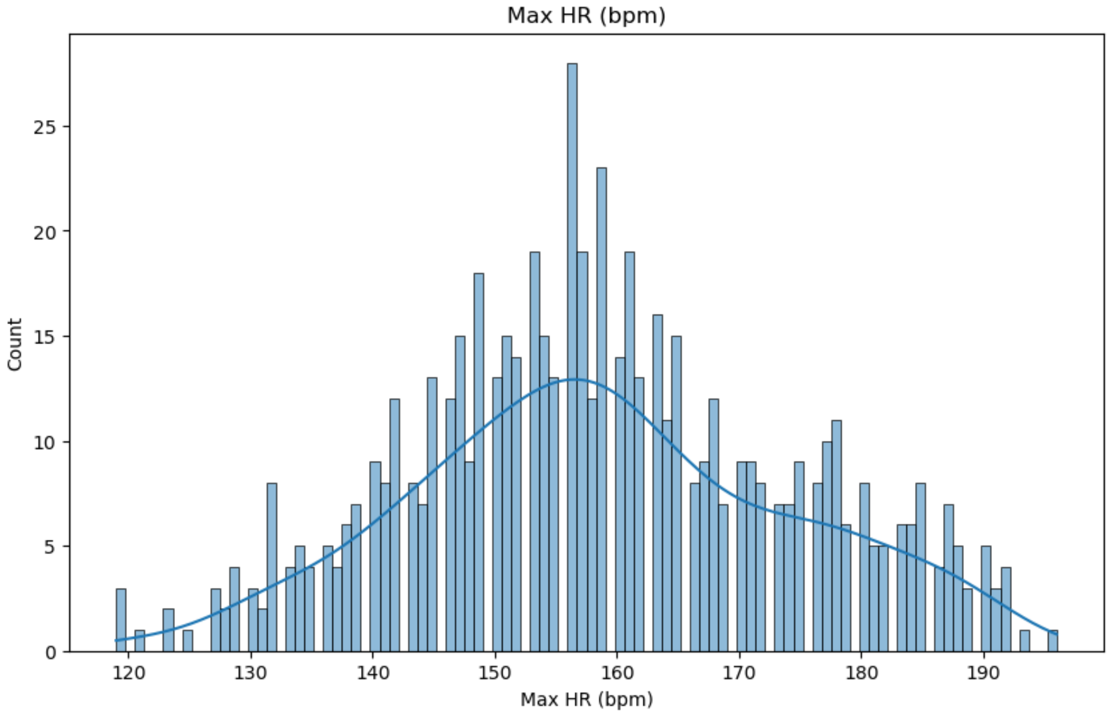

Awake %
Count
0
5
10
15
20
25
0.05
0.10
0.15
0.20
0.25
Awake %

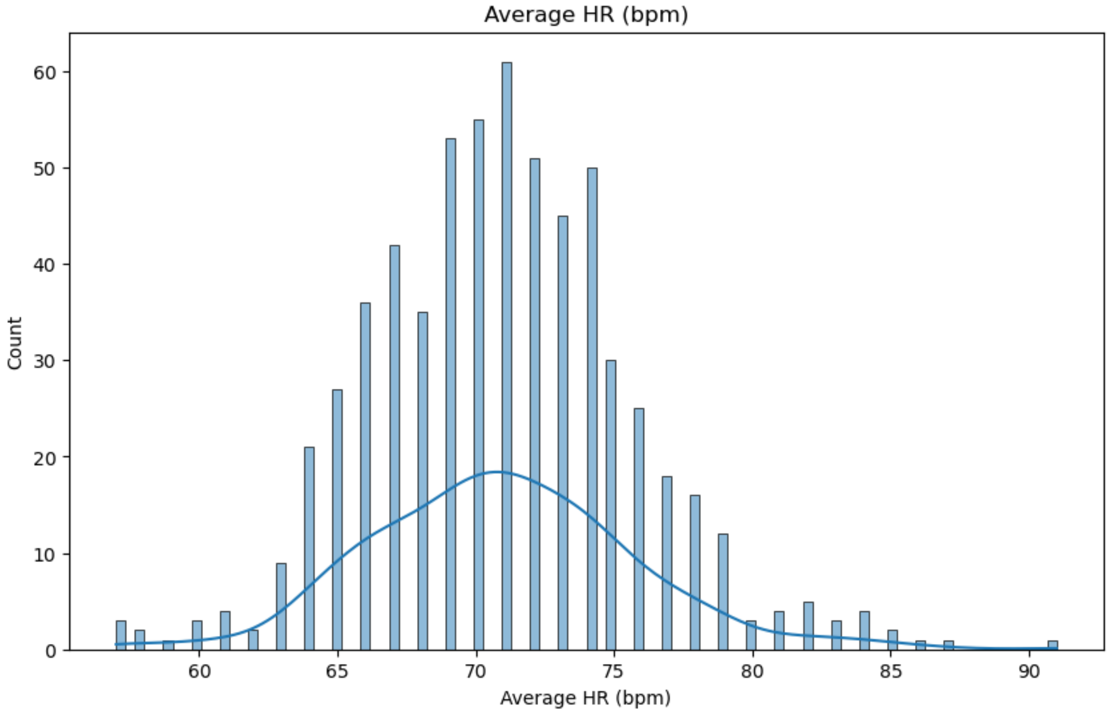

Light sleep %
Count
0
5
10
15
20
25
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Light sleep %


Restorative sleep %
Count
0
5
10
15
20
25
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Restorative sleep %


Day Strain
Count
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
6
8
10
12
14
16
18
20
Day Strain

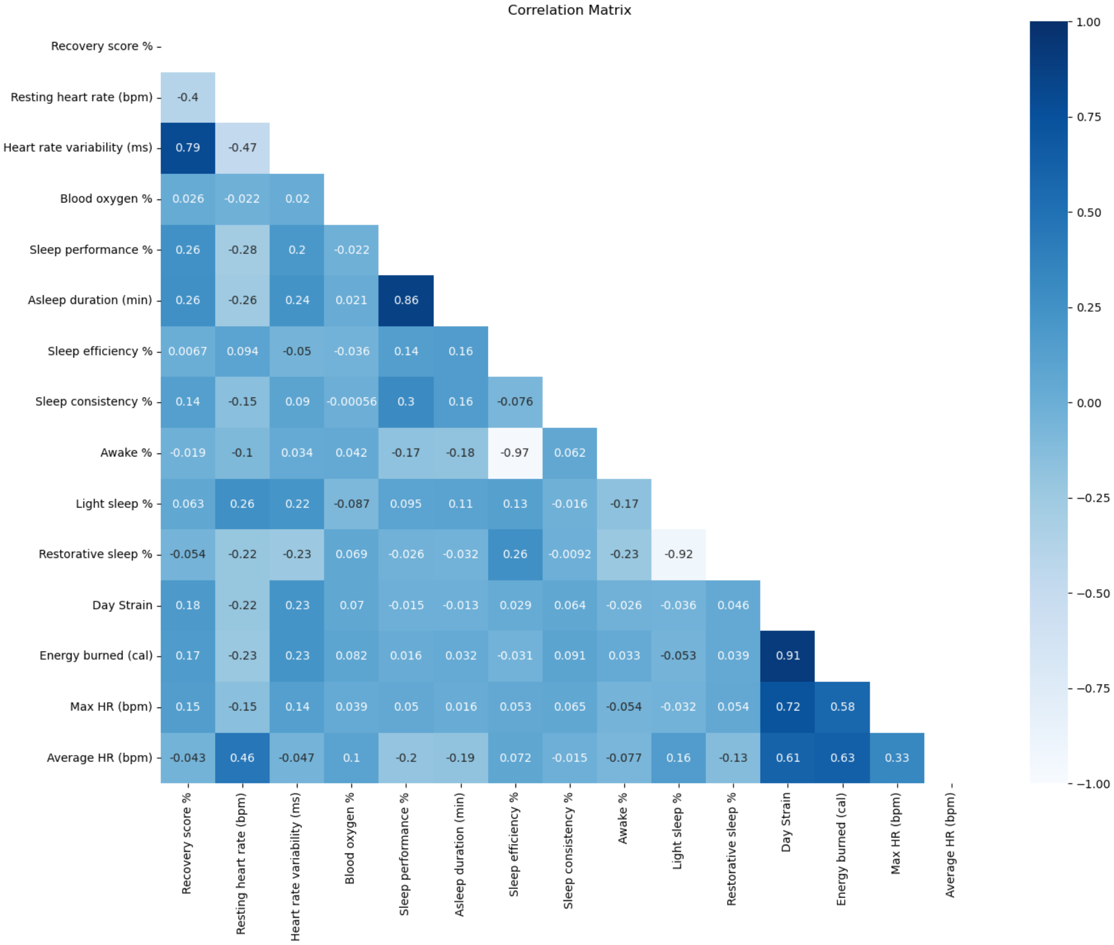

Energy burned (cal)
Count
0
5
10
15
20
25
1500
2000
2500
3000
3500
4000
4500
5000
Energy burned (cal)


Max HR (bpm)
Count
0
5
10
15
20
25
120
130
140
150
160
170
180
190
Max HR (bpm)

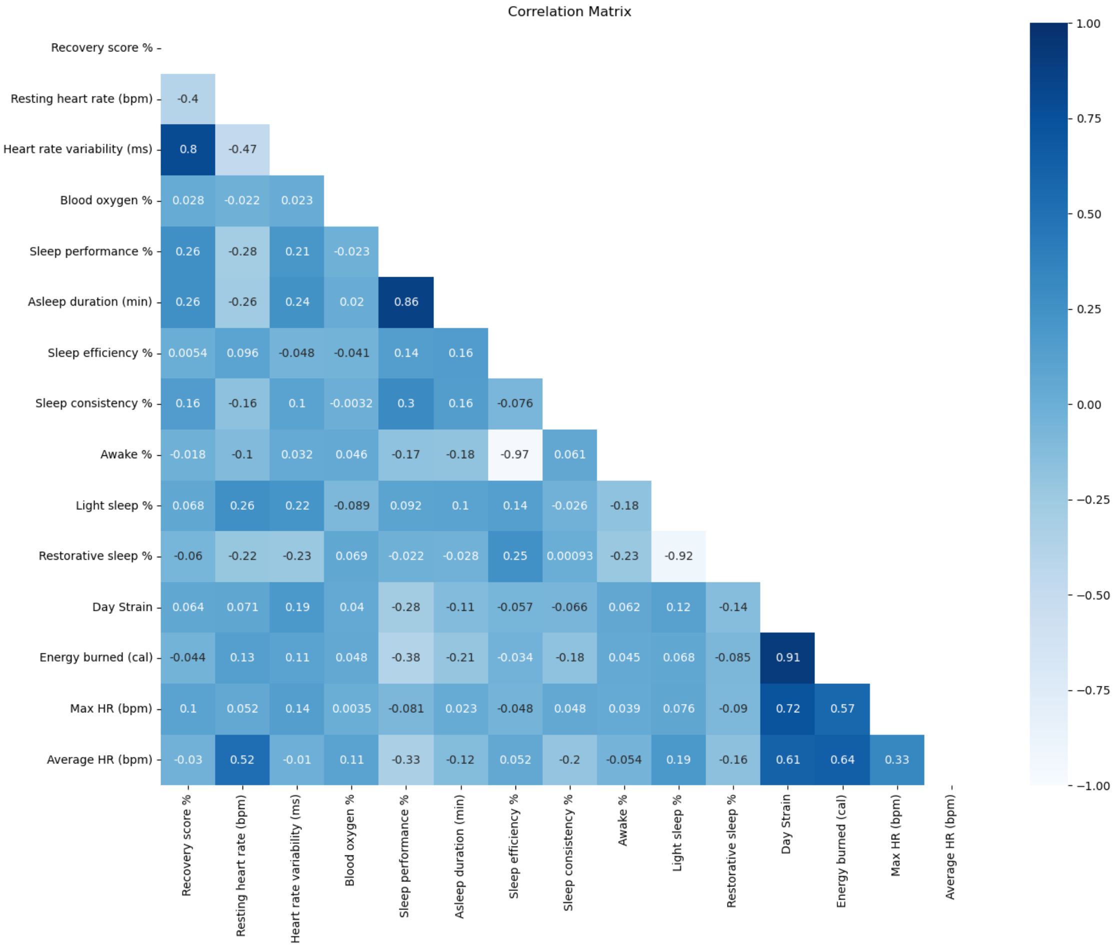

Average HR (bpm)
Count
0
10
20
30
40
50
60
60
65
70
75
80
85
90
Average HR (bpm)

3.6.3 Correlations

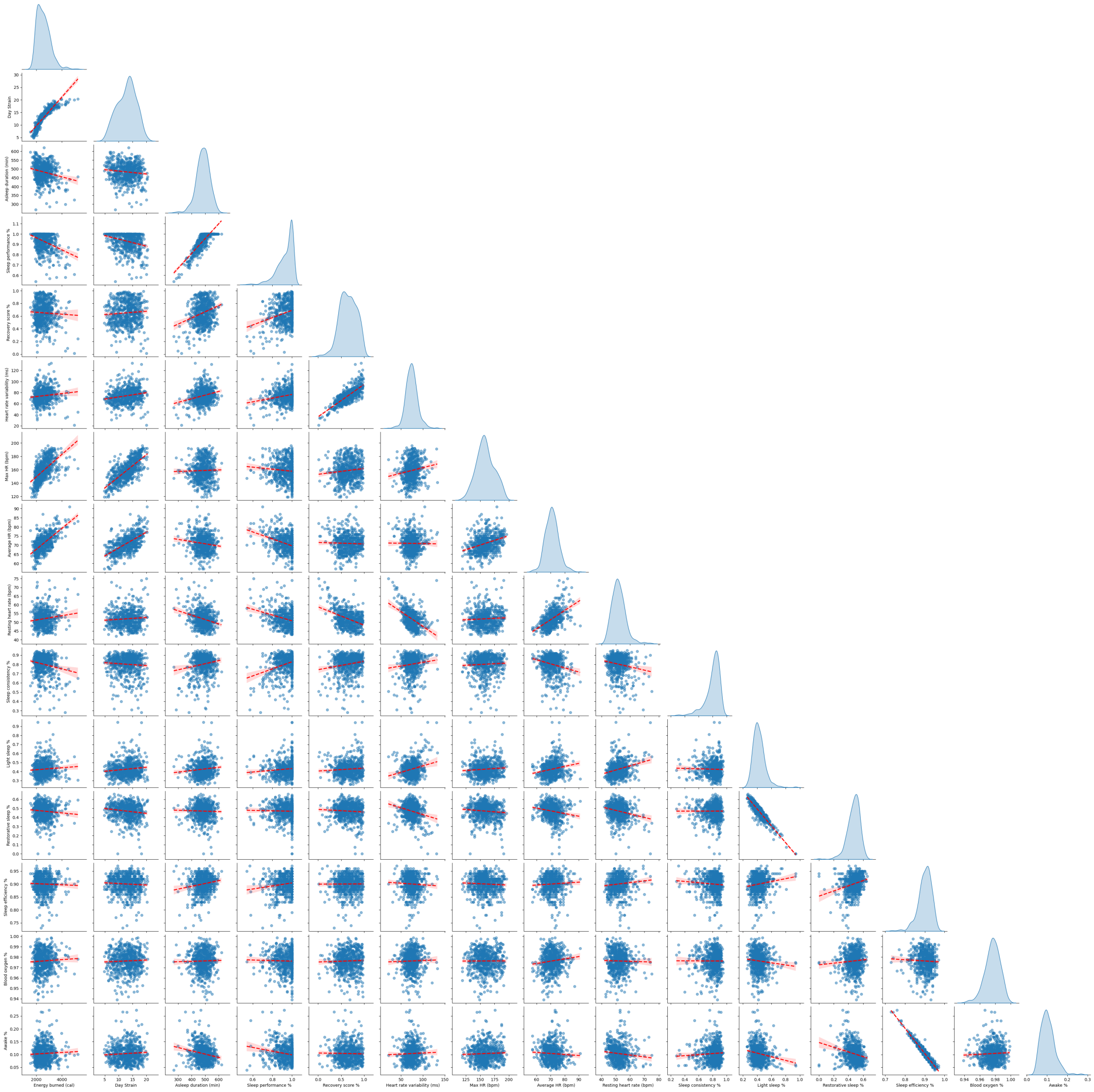
```
[139]: display_correlation_matrix(df_phys_add[PHYS_KDE_COLS]) # showing correlations between recovery metrics and *subsequent* strain metrics
```



```
[140]: display_correlation_matrix(df_phys_comb[PHYS_KDE_COLS]) # showing correlations between recovery metrics and *previous* strain metrics
```



```
[141]: display_pairplot(df_phys_comb[PHYS_KDE_COLS], threshold=0) # showing␣
        ↪relationships between recovery metrics and *previous* strain metrics
```



3.7 Export

[142]:
```
df_phys_add.to_csv(os.path.join(WHOOP_EXPORT_DIR, 'phys_add.csv'), index=False)
```

[143]:
```
df_phys_rec.to_csv(os.path.join(WHOOP_EXPORT_DIR, 'phys_rec.csv'), index=False)
```

[144]:
```
df_phys_cycle.to_csv(os.path.join(WHOOP_EXPORT_DIR, 'phys_cycle.csv'),␣
 ↪index=False)
```

[145]:
```
df_phys_comb.to_csv(os.path.join(WHOOP_EXPORT_DIR, 'phys_comb.csv'),␣
 ↪index=False)
```

4 Sleeps

No additional data not already in PHYS

4.1 Import

```
[ ]: SLEEP_PATH = os.path.join(WHOOP_DIR, 'sleeps.csv')
```

```
[ ]: df_sleep = pd.read_csv(SLEEP_PATH)
     df_sleep
```

```
[ ]:        Cycle start time       Cycle end time Cycle timezone  \
     0    2024-06-17 00:35:19                  NaN      UTC-05:00
     1    2024-06-15 23:12:59  2024-06-17 00:35:19      UTC-05:00
     2    2024-06-14 23:39:45  2024-06-15 23:12:59      UTC-05:00
     3    2024-06-14 00:07:47  2024-06-14 23:39:45      UTC-05:00
     4    2024-06-13 00:12:09  2024-06-14 00:07:47      UTC-05:00
     ..                   …                    …            …
     647  2022-09-18 22:20:02  2022-09-19 22:41:43      UTC-05:00
     648  2022-09-17 22:36:20  2022-09-18 22:20:02      UTC-05:00
     649  2022-09-16 21:53:01  2022-09-17 22:36:20      UTC-05:00
     650  2022-09-15 22:58:46  2022-09-16 21:53:01      UTC-05:00
     651  2022-09-14 21:52:24  2022-09-15 22:58:46      UTC-05:00

                 Sleep onset           Wake onset  Sleep performance %  \
     0    2024-06-17 00:35:19  2024-06-17 09:31:15                100.0
     1    2024-06-15 23:12:59  2024-06-16 08:49:16                 98.0
     2    2024-06-14 23:39:45  2024-06-15 07:54:03                 87.0
     3    2024-06-14 00:07:47  2024-06-14 09:55:04                100.0
     4    2024-06-13 00:12:09  2024-06-13 09:34:31                100.0
     ..                   …                    …                  …
     647  2022-09-18 22:20:02  2022-09-19 07:31:20                 95.0
     648  2022-09-17 22:36:20  2022-09-18 07:19:46                 85.0
     649  2022-09-16 21:53:01  2022-09-17 06:33:34                 92.0
     650  2022-09-15 22:58:46  2022-09-16 07:16:44                 88.0
     651  2022-09-14 21:52:24  2022-09-15 08:32:29                 99.0

          Respiratory rate (rpm)  Asleep duration (min)  In bed duration (min)  \
     0                      14.5                  507.0                  533.0
     1                      14.8                  533.0                  570.0
     2                      14.0                  433.0                  484.0
     3                      14.4                  535.0                  571.0
     4                      14.8                  488.0                  562.0
     ..                      …                      …                      …
     647                    14.0                  482.0                  549.0
     648                    13.8                  448.0                  520.0
     649                    14.0                  495.0                  520.0
     650                    13.6                  411.0                  497.0
```

```
651                    13.7                  529.0                  640.0

     Light sleep duration (min)  Deep (SWS) duration (min)  \
0                         247.0                       94.0   
1                         231.0                      137.0   
2                         298.0                       95.0   
3                         273.0                      126.0   
4                         221.0                      123.0   
..                          …                        …   
647                       182.0                      139.0   
648                       267.0                      102.0   
649                       357.0                       31.0   
650                       244.0                       73.0   
651                       362.0                       72.0   

     REM duration (min)  Awake duration (min)  Sleep need (min)  \
0                 166.0                  26.0             489.0   
1                 165.0                  37.0             544.0   
2                  40.0                  51.0             495.0   
3                 136.0                  36.0             483.0   
4                 144.0                  74.0             472.0   
..                  …                   …               …   
647               161.0                  67.0             510.0   
648                79.0                  72.0             526.0   
649               107.0                  25.0             537.0   
650                94.0                  86.0             468.0   
651                95.0                 111.0             533.0   

     Sleep debt (min)  Sleep efficiency %  Sleep consistency %    Nap
0                 7.0                95.0                 78.0  False
1                31.0                93.0                 77.0  False
2                 0.0                89.0                 90.0  False
3                 0.0                94.0                 87.0  False
4                 0.0                89.0                 81.0  False
..                  …                 …                  …    …
647              34.0                88.0                 87.0  False
648              23.0                86.0                 84.0  False
649              27.0                95.0                  NaN  False
650               2.0                82.0                  NaN  False
651               0.0                84.0                  NaN  False

[652 rows x 18 columns]
```

5 Workouts

5.0.1 Options

- Predict Strain from workout metrics
- Predict calories from workout metrics
- Predict Activity from workout metrics

5.1 Import

```
[ ]: WO_PATH = os.path.join(WHOOP_DIR, 'workouts.csv')
```

```
[ ]: df_wo = pd.read_csv(WO_PATH)
     df_wo
```

```
[ ]:        Cycle start time       Cycle end time Cycle timezone  \
     0    2024-06-14 23:39:45  2024-06-15 23:12:59      UTC-05:00
     1    2024-06-14 00:07:47  2024-06-14 23:39:45      UTC-05:00
     2    2024-06-13 00:12:09  2024-06-14 00:07:47      UTC-05:00
     3    2024-06-10 23:15:19  2024-06-11 23:11:54      UTC-05:00
     4    2024-06-10 23:15:19  2024-06-11 23:11:54      UTC-05:00
     ..                   …                    …            …
     445  2022-09-18 22:20:02  2022-09-19 22:41:43      UTC-05:00
     446  2022-09-16 21:53:01  2022-09-17 22:36:20      UTC-05:00
     447  2022-09-16 21:53:01  2022-09-17 22:36:20      UTC-05:00
     448  2022-09-15 22:58:46  2022-09-16 21:53:01      UTC-05:00
     449  2022-09-14 00:00:00  2022-09-14 21:52:24      UTC-05:00

          Workout start time     Workout end time  Duration (min)  Activity name  \
     0    2024-06-15 12:52:30  2024-06-15 14:22:29              89       Activity
     1    2024-06-14 17:30:00  2024-06-14 18:57:00              87      Jiu Jitsu
     2    2024-06-13 18:08:57  2024-06-13 19:25:28              76  Weightlifting
     3    2024-06-11 17:30:00  2024-06-11 18:50:00              80      Jiu Jitsu
     4    2024-06-11 16:30:00  2024-06-11 17:29:00              59   Martial Arts
     ..                   …                    …              …              …
     445  2022-09-19 17:25:36  2022-09-19 18:43:21              77    Box Fitness
     446  2022-09-17 10:40:25  2022-09-17 11:17:11              36   Powerlifting
     447  2022-09-17 09:27:55  2022-09-17 10:29:45              61    Box Fitness
     448  2022-09-16 16:33:25  2022-09-16 17:40:16              66    Box Fitness
     449  2022-09-14 17:22:28  2022-09-14 18:48:51              86    Box Fitness

          Activity Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  \
     0               11.7                705.0           159               127
     1               12.8                638.0           184               125
     2                9.6                467.0           161               119
     3               10.1                511.0           167               121
     4               11.8                534.0           177               134
     ..               …                    …            …                 …
```

```
445             15.7                884.0           175                144
446              7.4                247.0           153                122
447             13.5                556.0           179                133
448             15.9                779.0           180                147
449             13.5                723.0           174                130

     HR Zone 1 %  HR Zone 2 %  HR Zone 3 %  HR Zone 4 %  HR Zone 5 %  \
0              4           78           18            0            0
1             28           47           11            9            4
2             36           45           19            0            0
3             30           48           17            2            0
4              4           51           29           15            1
..          …          …          …          …          …
445            0           15           33           49            3
446            8           70           20            2            0
447            8           44           19           19           10
448            2           14           24           35           25
449            8           41           30           19            2

     GPS enabled  Distance (meters)  Altitude gain (meters)  \
0          False                NaN                     NaN
1          False                NaN                     NaN
2          False                NaN                     NaN
3          False                NaN                     NaN
4          False                NaN                     NaN
..          …                …                     …
445        False                NaN                     NaN
446        False                NaN                     NaN
447        False                NaN                     NaN
448        False                NaN                     NaN
449        False                NaN                     NaN

     Altitude change (meters)
0                         NaN
1                         NaN
2                         NaN
3                         NaN
4                         NaN
..                        …
445                       NaN
446                       NaN
447                       NaN
448                       NaN
449                       NaN

[450 rows x 20 columns]
```

5.2 Drop columns

[ ]:

```
df_wo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 450 entries, 0 to 449
Data columns (total 20 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Cycle start time         450 non-null    object
 1   Cycle end time           450 non-null    object
 2   Cycle timezone           450 non-null    object
 3   Workout start time       450 non-null    object
 4   Workout end time         450 non-null    object
 5   Duration (min)           450 non-null    int64
 6   Activity name            450 non-null    object
 7   Activity Strain          450 non-null    float64
 8   Energy burned (cal)      450 non-null    float64
 9   Max HR (bpm)             450 non-null    int64
 10  Average HR (bpm)         450 non-null    int64
 11  HR Zone 1 %              450 non-null    int64
 12  HR Zone 2 %              450 non-null    int64
 13  HR Zone 3 %              450 non-null    int64
 14  HR Zone 4 %              450 non-null    int64
 15  HR Zone 5 %              450 non-null    int64
 16  GPS enabled              450 non-null    bool
 17  Distance (meters)        0 non-null      float64
 18  Altitude gain (meters)   0 non-null      float64
 19  Altitude change (meters) 0 non-null      float64
dtypes: bool(1), float64(5), int64(8), object(6)
memory usage: 67.4+ KB
```

[ ]:

```
WO_DROP_COLS = [
    'Cycle end time',
    'Cycle timezone',
    'Workout end time',
    'GPS enabled',
    'Distance (meters)',
    'Altitude gain (meters)',
    'Altitude change (meters)'
]
```

[ ]:

```
df_wo_sel = df_wo.drop(columns=WO_DROP_COLS)
df_wo_sel
```

[ ]:

```
         Cycle start time   Workout start time  Duration (min)  Activity name  \
0     2024-06-14 23:39:45  2024-06-15 12:52:30              89       Activity
1     2024-06-14 00:07:47  2024-06-14 17:30:00              87     Jiu Jitsu
```

```
2    2024-06-13 00:12:09  2024-06-13 18:08:57              76  Weightlifting
3    2024-06-10 23:15:19  2024-06-11 17:30:00              80      Jiu Jitsu
4    2024-06-10 23:15:19  2024-06-11 16:30:00              59   Martial Arts
..                    …                    …             …             …
445  2022-09-18 22:20:02  2022-09-19 17:25:36              77    Box Fitness
446  2022-09-16 21:53:01  2022-09-17 10:40:25              36   Powerlifting
447  2022-09-16 21:53:01  2022-09-17 09:27:55              61    Box Fitness
448  2022-09-15 22:58:46  2022-09-16 16:33:25              66    Box Fitness
449  2022-09-14 00:00:00  2022-09-14 17:22:28              86    Box Fitness

     Activity Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  \
0               11.7                705.0           159               127
1               12.8                638.0           184               125
2                9.6                467.0           161               119
3               10.1                511.0           167               121
4               11.8                534.0           177               134
..                 …                    …           …                 …
445             15.7                884.0           175               144
446              7.4                247.0           153               122
447             13.5                556.0           179               133
448             15.9                779.0           180               147
449             13.5                723.0           174               130

     HR Zone 1 %  HR Zone 2 %  HR Zone 3 %  HR Zone 4 %  HR Zone 5 %
0              4           78           18            0            0
1             28           47           11            9            4
2             36           45           19            0            0
3             30           48           17            2            0
4              4           51           29           15            1
..          …          …          …          …          …
445            0           15           33           49            3
446            8           70           20            2            0
447            8           44           19           19           10
448            2           14           24           35           25
449            8           41           30           19            2

[450 rows x 13 columns]
```

5.3 Dtypes

```
[ ]: df_wo_sel.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 450 entries, 0 to 449
Data columns (total 13 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Cycle start time     450 non-null    object
 1   Workout start time   450 non-null    object
 2   Duration (min)       450 non-null    int64
 3   Activity name        450 non-null    object
 4   Activity Strain      450 non-null    float64
 5   Energy burned (cal)  450 non-null    float64
 6   Max HR (bpm)         450 non-null    int64
 7   Average HR (bpm)     450 non-null    int64
 8   HR Zone 1 %          450 non-null    int64
 9   HR Zone 2 %          450 non-null    int64
 10  HR Zone 3 %          450 non-null    int64
 11  HR Zone 4 %          450 non-null    int64
 12  HR Zone 5 %          450 non-null    int64
dtypes: float64(2), int64(8), object(3)
memory usage: 45.8+ KB
```

[ ]:

```python
df_wo_dt = df_wo_sel.copy()
df_wo_dt['Cycle start time'] = pd.to_datetime(df_wo_dt['Cycle start time'])
df_wo_dt['Workout start time'] = pd.to_datetime(df_wo_dt['Workout start time'])
```

[ ]:

```python
for i in range(5):
    df_wo_dt[f'HR Zone {i+1} %'] = df_wo_dt[f'HR Zone {i+1} %'] / 100.
```

[ ]:

```python
df_wo_dt
```

[ ]:

```
       Cycle start time  Workout start time  Duration (min)  Activity name  \
0   2024-06-14 23:39:45 2024-06-15 12:52:30              89       Activity
1   2024-06-14 00:07:47 2024-06-14 17:30:00              87      Jiu Jitsu
2   2024-06-13 00:12:09 2024-06-13 18:08:57              76  Weightlifting
3   2024-06-10 23:15:19 2024-06-11 17:30:00              80      Jiu Jitsu
4   2024-06-10 23:15:19 2024-06-11 16:30:00              59   Martial Arts
..                  …                 …             …            …
445 2022-09-18 22:20:02 2022-09-19 17:25:36              77    Box Fitness
446 2022-09-16 21:53:01 2022-09-17 10:40:25              36   Powerlifting
447 2022-09-16 21:53:01 2022-09-17 09:27:55              61    Box Fitness
448 2022-09-15 22:58:46 2022-09-16 16:33:25              66    Box Fitness
449 2022-09-14 00:00:00 2022-09-14 17:22:28              86    Box Fitness

     Activity Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  \
0               11.7                705.0           159               127
1               12.8                638.0           184               125
2                9.6                467.0           161               119
3               10.1                511.0           167               121
4               11.8                534.0           177               134
..                 …                  …           …                …
445             15.7                884.0           175               144
446              7.4                247.0           153               122
447             13.5                556.0           179               133
448             15.9                779.0           180               147
```

```
449             13.5                723.0           174                 130

     HR Zone 1 %  HR Zone 2 %  HR Zone 3 %  HR Zone 4 %  HR Zone 5 %
0           0.04         0.78         0.18         0.00         0.00
1           0.28         0.47         0.11         0.09         0.04
2           0.36         0.45         0.19         0.00         0.00
3           0.30         0.48         0.17         0.02         0.00
4           0.04         0.51         0.29         0.15         0.01
..           …          …          …          …          …
445         0.00         0.15         0.33         0.49         0.03
446         0.08         0.70         0.20         0.02         0.00
447         0.08         0.44         0.19         0.19         0.10
448         0.02         0.14         0.24         0.35         0.25
449         0.08         0.41         0.30         0.19         0.02

[450 rows x 13 columns]
```

```
[ ]: df_wo_dt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 450 entries, 0 to 449
Data columns (total 13 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Cycle start time     450 non-null    datetime64[ns]
 1   Workout start time   450 non-null    datetime64[ns]
 2   Duration (min)       450 non-null    int64
 3   Activity name        450 non-null    object
 4   Activity Strain      450 non-null    float64
 5   Energy burned (cal)  450 non-null    float64
 6   Max HR (bpm)         450 non-null    int64
 7   Average HR (bpm)     450 non-null    int64
 8   HR Zone 1 %          450 non-null    float64
 9   HR Zone 2 %          450 non-null    float64
 10  HR Zone 3 %          450 non-null    float64
 11  HR Zone 4 %          450 non-null    float64
 12  HR Zone 5 %          450 non-null    float64
dtypes: datetime64[ns](2), float64(7), int64(3), object(1)
memory usage: 45.8+ KB
```

5.4 Add cols

```
[ ]: df_wo_add = df_wo_dt.copy()
     df_wo_add['cycle_start_time_adj'] = df_wo_add['Cycle start time'] - pd.
      ↪Timedelta(hours=9)
     df_wo_add['cycle_date'] = df_wo_add['cycle_start_time_adj'].dt.date + pd.
      ↪Timedelta(days=1) # cycle starts at the beginning of sleep the previous night
     df_wo_add['wo_date'] = df_wo_add['Workout start time'].dt.date
```

```
[ ]: df_wo_add[df_wo_add['cycle_date'] != df_wo_add['wo_date']] # rare event -- ACL␣
      ↪festival -- can drop
```

```
[ ]:        Cycle start time  Workout start time  Duration (min) Activity name  \
     222 2023-08-09 22:37:56 2023-08-11 01:31:30              34      Activity

          Activity Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  \
     222              7.1                234.0           159               118

          HR Zone 1 %  HR Zone 2 %  HR Zone 3 %  HR Zone 4 %  HR Zone 5 %  \
     222         0.16         0.75         0.06         0.03          0.0

         cycle_start_time_adj  cycle_date     wo_date
     222  2023-08-09 13:37:56  2023-08-10  2023-08-11
```

```
[ ]: df_wo_add = df_wo_add.drop(index=222).reset_index(drop=True)
```

```
[ ]: df_wo_add
```

```
[ ]:        Cycle start time  Workout start time  Duration (min)  Activity name  \
     0   2024-06-14 23:39:45 2024-06-15 12:52:30              89       Activity
     1   2024-06-14 00:07:47 2024-06-14 17:30:00              87      Jiu Jitsu
     2   2024-06-13 00:12:09 2024-06-13 18:08:57              76  Weightlifting
     3   2024-06-10 23:15:19 2024-06-11 17:30:00              80      Jiu Jitsu
     4   2024-06-10 23:15:19 2024-06-11 16:30:00              59   Martial Arts
     ..                  …                   …              …              …
     444 2022-09-18 22:20:02 2022-09-19 17:25:36              77    Box Fitness
     445 2022-09-16 21:53:01 2022-09-17 10:40:25              36   Powerlifting
     446 2022-09-16 21:53:01 2022-09-17 09:27:55              61    Box Fitness
     447 2022-09-15 22:58:46 2022-09-16 16:33:25              66    Box Fitness
     448 2022-09-14 00:00:00 2022-09-14 17:22:28              86    Box Fitness

          Activity Strain  Energy burned (cal)  Max HR (bpm)  Average HR (bpm)  \
     0               11.7                705.0           159               127
     1               12.8                638.0           184               125
     2                9.6                467.0           161               119
     3               10.1                511.0           167               121
     4               11.8                534.0           177               134
     ..               …                  …            …                …
     444             15.7                884.0           175               144
```

```
445              7.4                247.0           153               122
446             13.5                556.0           179               133
447             15.9                779.0           180               147
448             13.5                723.0           174               130

     HR Zone 1 %  HR Zone 2 %  HR Zone 3 %  HR Zone 4 %  HR Zone 5 %  \
0           0.04         0.78         0.18         0.00         0.00
1           0.28         0.47         0.11         0.09         0.04
2           0.36         0.45         0.19         0.00         0.00
3           0.30         0.48         0.17         0.02         0.00
4           0.04         0.51         0.29         0.15         0.01
..           …           …          …          …          …
444         0.00         0.15         0.33         0.49         0.03
445         0.08         0.70         0.20         0.02         0.00
446         0.08         0.44         0.19         0.19         0.10
447         0.02         0.14         0.24         0.35         0.25
448         0.08         0.41         0.30         0.19         0.02

    cycle_start_time_adj  cycle_date     wo_date
0    2024-06-14 14:39:45  2024-06-15  2024-06-15
1    2024-06-13 15:07:47  2024-06-14  2024-06-14
2    2024-06-12 15:12:09  2024-06-13  2024-06-13
3    2024-06-10 14:15:19  2024-06-11  2024-06-11
4    2024-06-10 14:15:19  2024-06-11  2024-06-11
..                   …         …         …
444  2022-09-18 13:20:02  2022-09-19  2022-09-19
445  2022-09-16 12:53:01  2022-09-17  2022-09-17
446  2022-09-16 12:53:01  2022-09-17  2022-09-17
447  2022-09-15 13:58:46  2022-09-16  2022-09-16
448  2022-09-13 15:00:00  2022-09-14  2022-09-14

[449 rows x 16 columns]
```

5.5 EDA

5.5.1 Class counts

```
[ ]: df_wo_class_counts = df_wo_add['Activity name'].value_counts(sort=True).
      ↪reset_index()
     df_wo_class_counts
```

```
[ ]:             Activity name  count
     0           Weightlifting    161
     1               Jiu Jitsu     96
     2            Powerlifting     35
     3                Activity     33
     4             Box Fitness     33
     5               Wrestling     20
```

```
6                     Yoga     17
7               Kickboxing     12
8             Martial Arts     12
9                     Spin      7
10            Assault Bike      7
11            Manual Labor      4
12                  Boxing      4
13  Operations - Tactical       3
14               Yard Work      2
15                   Dance      1
16               Paintball      1
17                Lacrosse      1
```

```
[ ]: df_wo_keep_classes = df_wo_class_counts['Activity␣
      ↪name'][df_wo_class_counts['count'] >= 5].to_list()
     df_wo_keep_classes.remove('Activity') # 'Activity' represents various unlabeled␣
      ↪activites, and should therefore be excluded
     df_wo_keep_classes
```

```
[ ]: ['Weightlifting',
      'Jiu Jitsu',
      'Powerlifting',
      'Box Fitness',
      'Wrestling',
      'Yoga',
      'Kickboxing',
      'Martial Arts',
      'Spin',
      'Assault Bike']
```

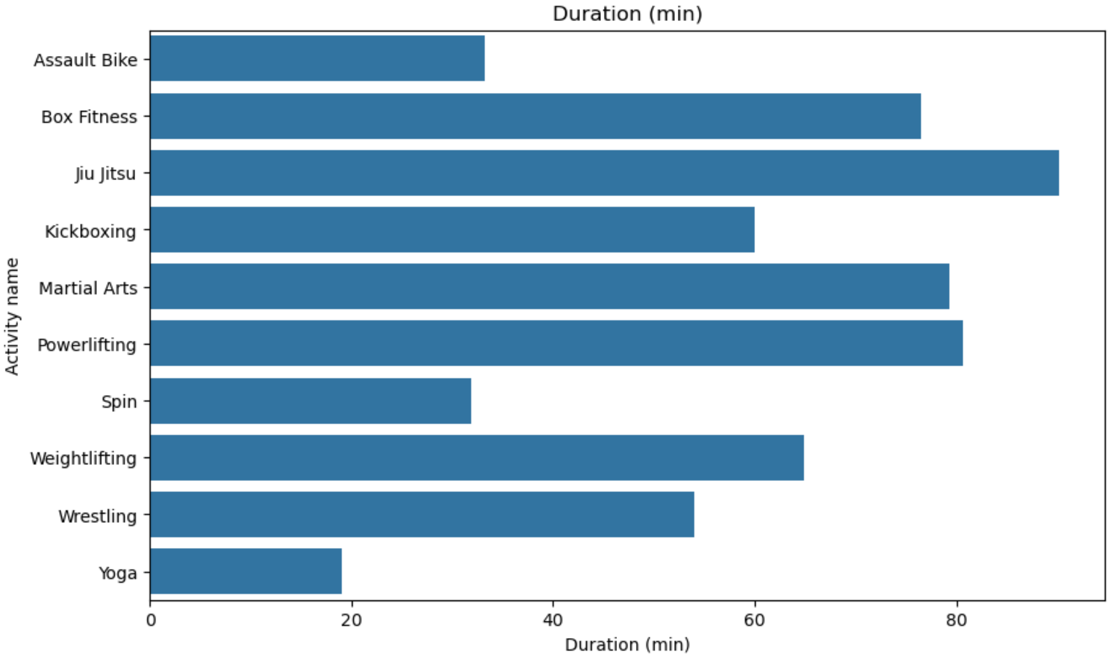
5.5.2 Class Averages

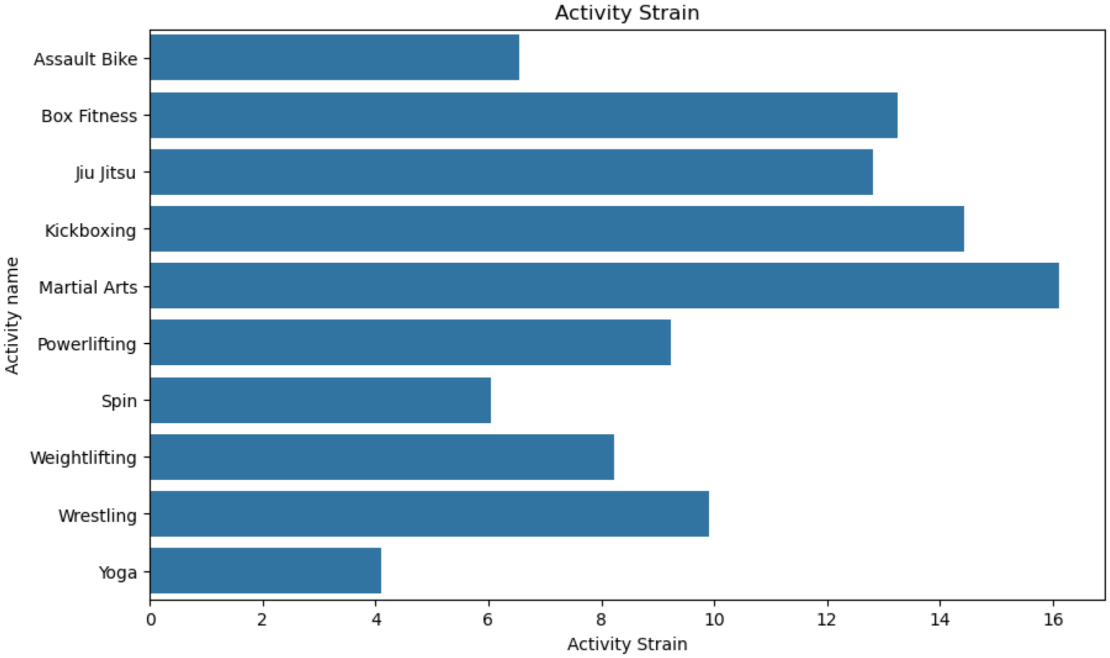
```
[ ]: WO_AVG_COLS = [
         'Duration (min)',
         'Activity Strain',
         'Energy burned (cal)',
         'Max HR (bpm)',
         'Average HR (bpm)',
         'HR Zone 1 %',
         'HR Zone 2 %',
         'HR Zone 3 %',
         'HR Zone 4 %',
         'HR Zone 5 %'
     ]

     df_wo_avg = df_wo_add[df_wo_add['Activity name'].isin(df_wo_keep_classes)].
      ↪groupby(by='Activity name')[WO_AVG_COLS].mean().reset_index()
```

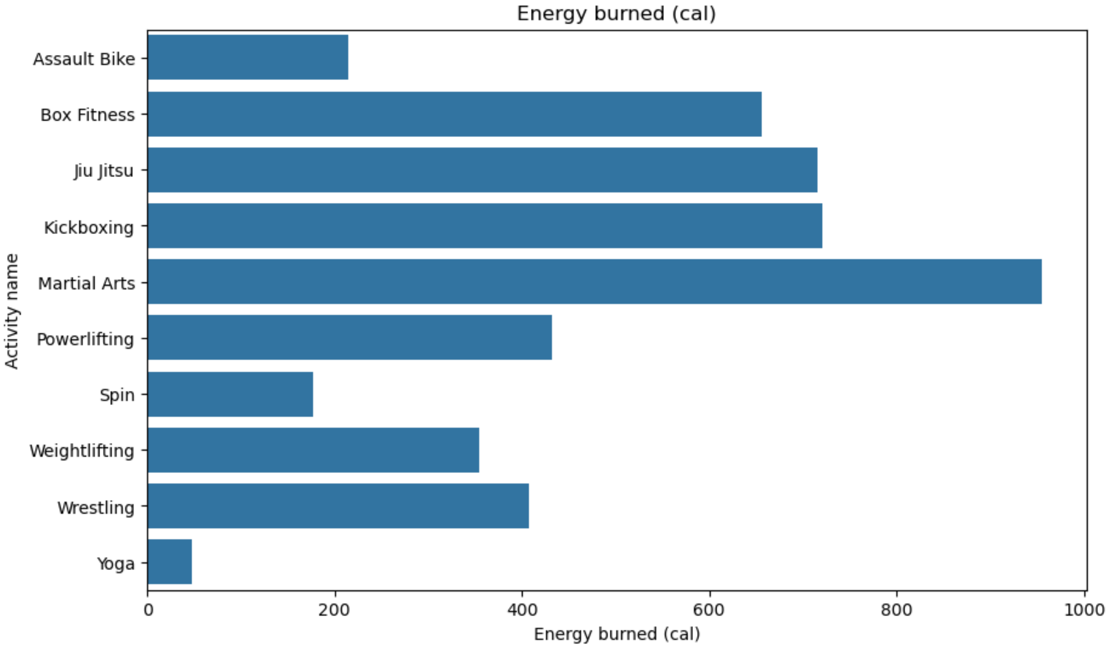
```
df_wo_avg
```

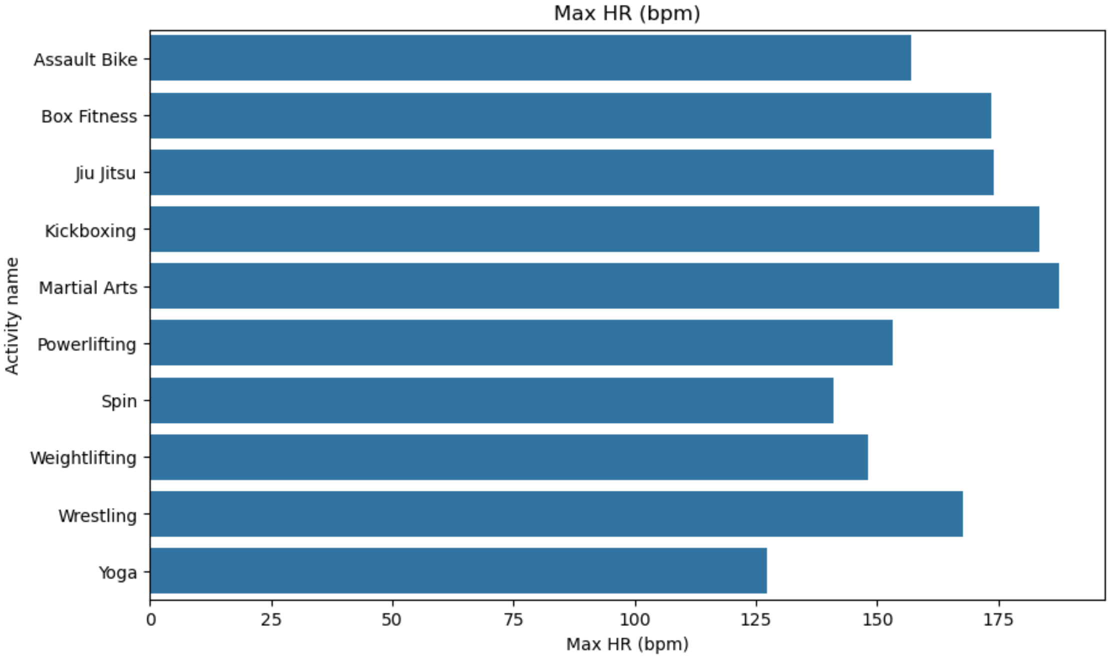
```
[ ]:    Activity name  Duration (min)  Activity Strain  Energy burned (cal)  \
     0    Assault Bike       33.285714         6.542857           214.285714
     1     Box Fitness       76.484848        13.248485           656.424242
     2       Jiu Jitsu       90.218750        12.814583           716.031250
     3      Kickboxing       60.000000        14.425000           720.583333
     4    Martial Arts       79.333333        16.116667           955.416667
     5    Powerlifting       80.628571         9.228571           431.742857
     6            Spin       31.857143         6.042857           177.285714
     7   Weightlifting       64.931677         8.222360           354.031056
     8       Wrestling       54.000000         9.905000           407.150000
     9            Yoga       19.058824         4.100000            47.764706

        Max HR (bpm)  Average HR (bpm)  HR Zone 1 %  HR Zone 2 %  HR Zone 3 %  \
     0    157.000000        120.000000     0.175714     0.708571     0.104286
     1    173.666667        127.939394     0.173636     0.297576     0.270909
     2    174.062500        126.020833     0.216042     0.358125     0.241667
     3    183.416667        146.750000     0.025000     0.189167     0.346667
     4    187.583333        147.250000     0.055833     0.203333     0.306667
     5    153.114286        112.885714     0.322857     0.432000     0.182286
     6    141.000000        114.000000     0.402857     0.530000     0.044286
     7    148.093168        113.031056     0.352236     0.460559     0.121988
     8    167.800000        125.150000     0.215500     0.373000     0.258500
     9    127.176471         87.529412     0.256471     0.076471     0.050588

        HR Zone 4 %  HR Zone 5 %
     0     0.002857     0.000000
     1     0.198182     0.037273
     2     0.110625     0.038646
     3     0.303333     0.128333
     4     0.255000     0.178333
     5     0.013429     0.000000
     6     0.000000     0.000000
     7     0.006770     0.000000
     8     0.093000     0.017500
     9     0.008235     0.000000
```

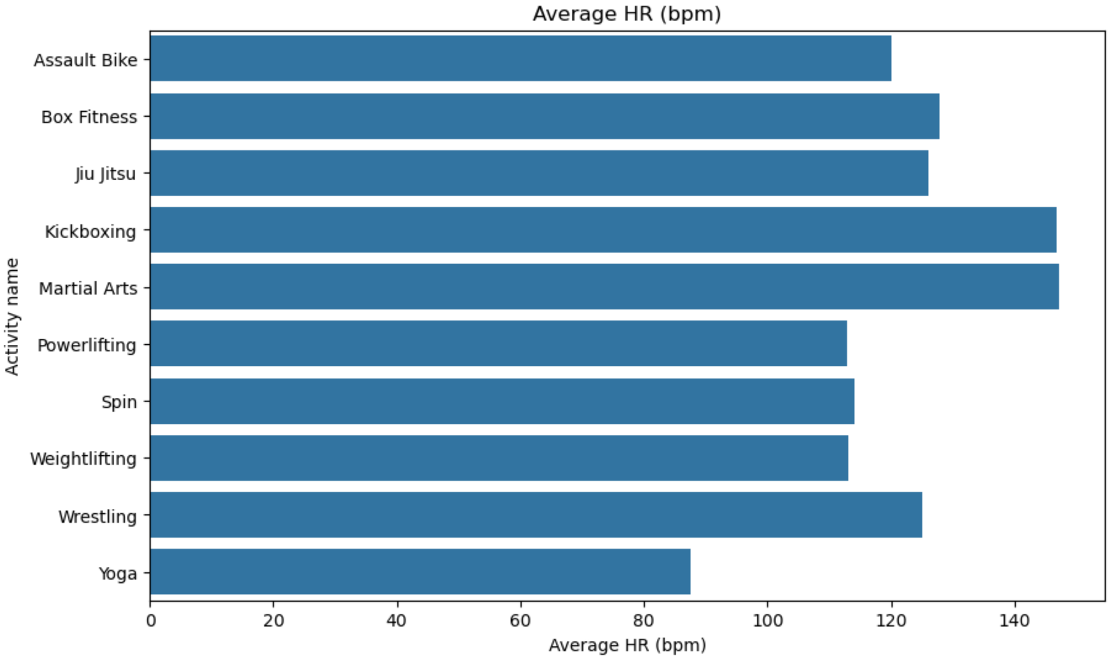
```
[ ]: for column in WO_AVG_COLS:
         plt.figure(figsize=(10, 6))
         sns.barplot(data=df_wo_avg, x=column, y='Activity name')
         plt.title(column)
         plt.show()
```

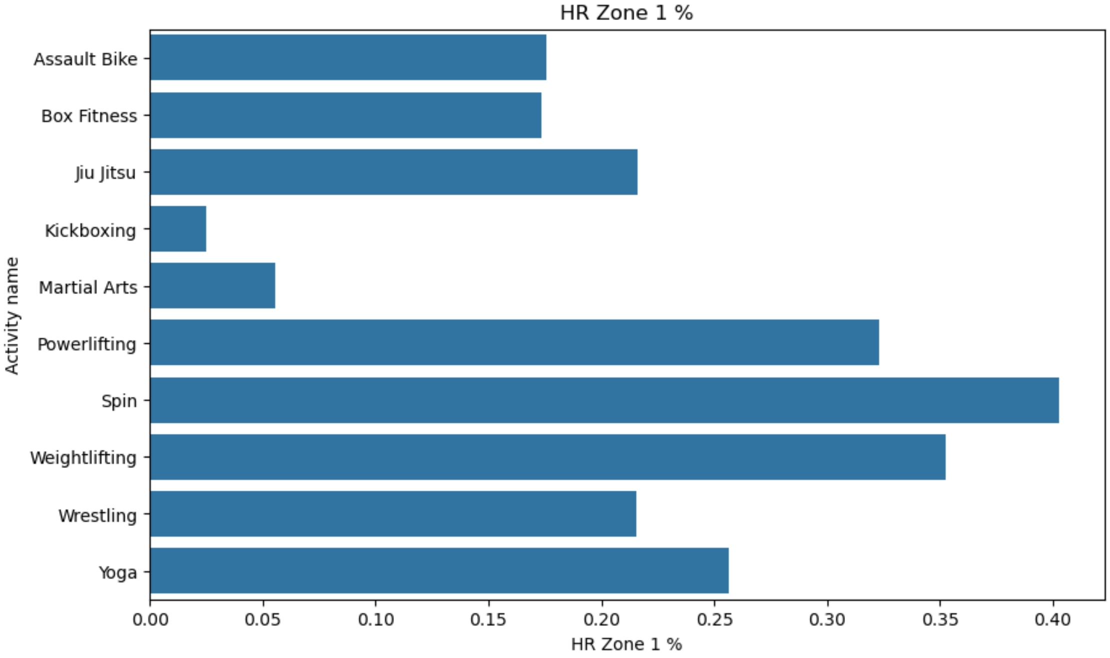

Duration (min)
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0
20
40
60
80
Duration (min)

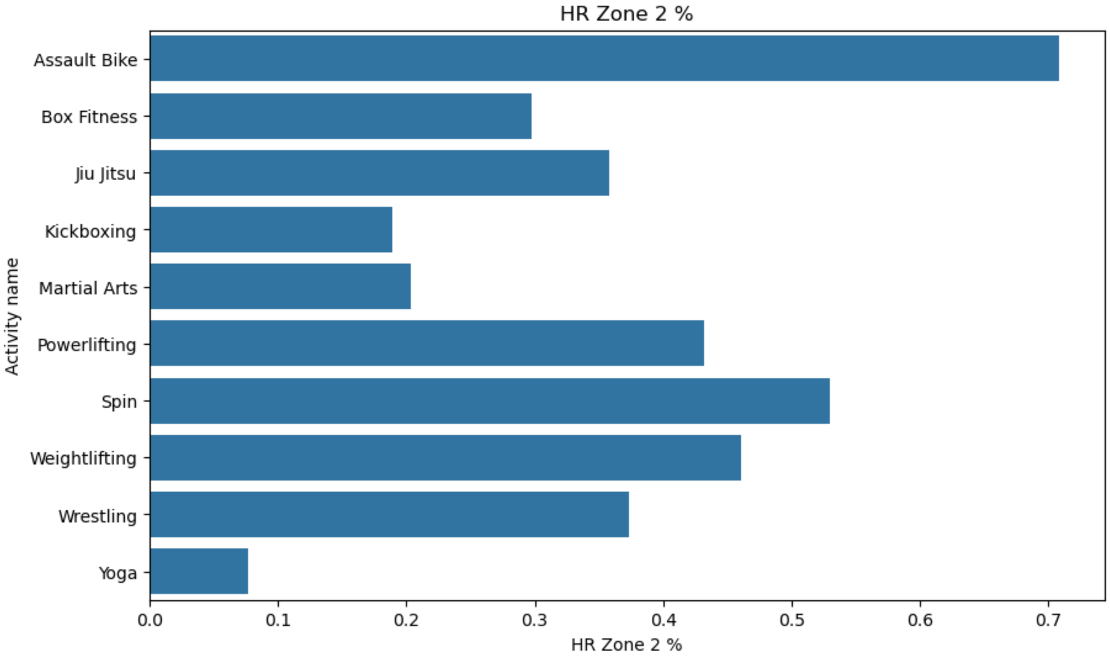

Activity Strain
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0
2
4
6
8
10
12
14
16
Activity Strain

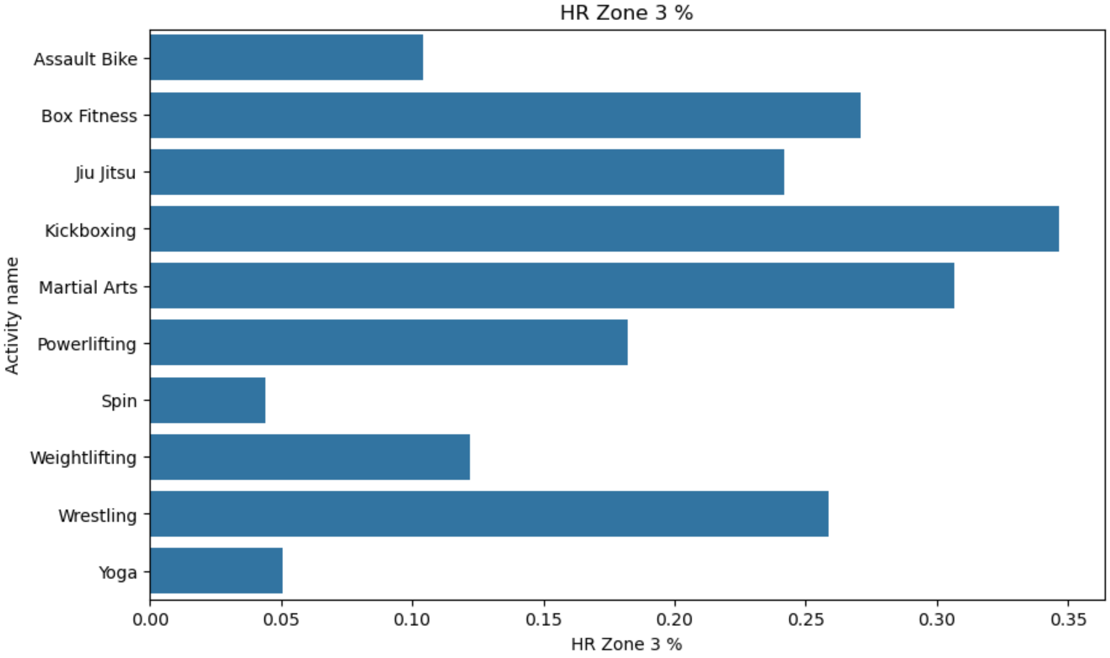

Energy burned (cal)
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0
200
400
600
800
1000
Energy burned (cal)

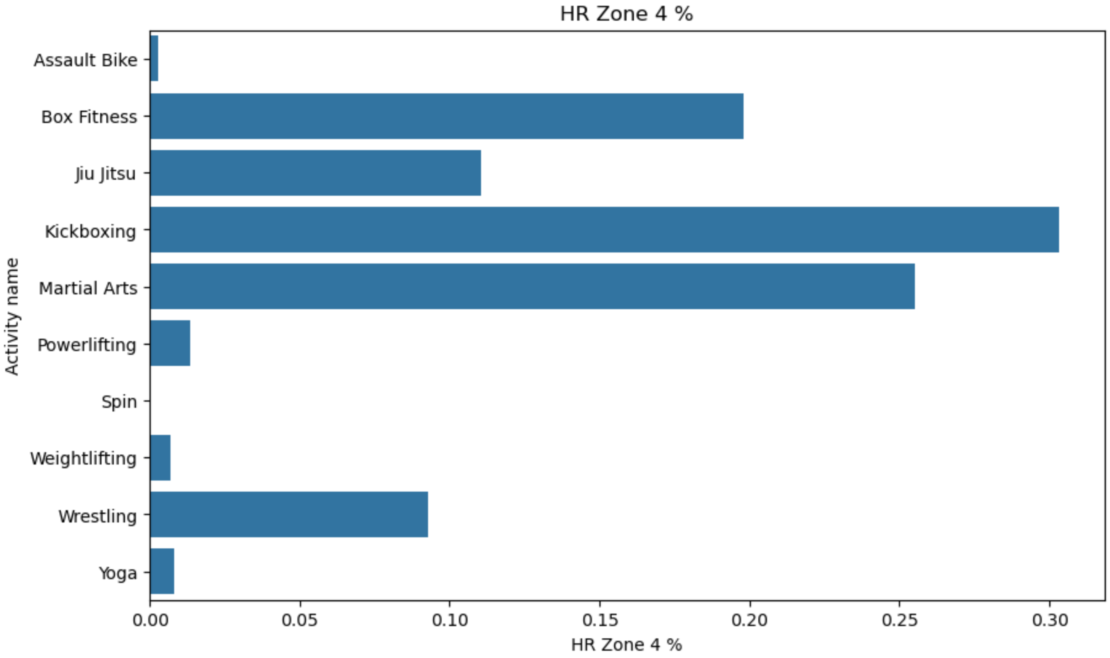

Max HR (bpm)
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0
25
50
75
100
125
150
175
Max HR (bpm)

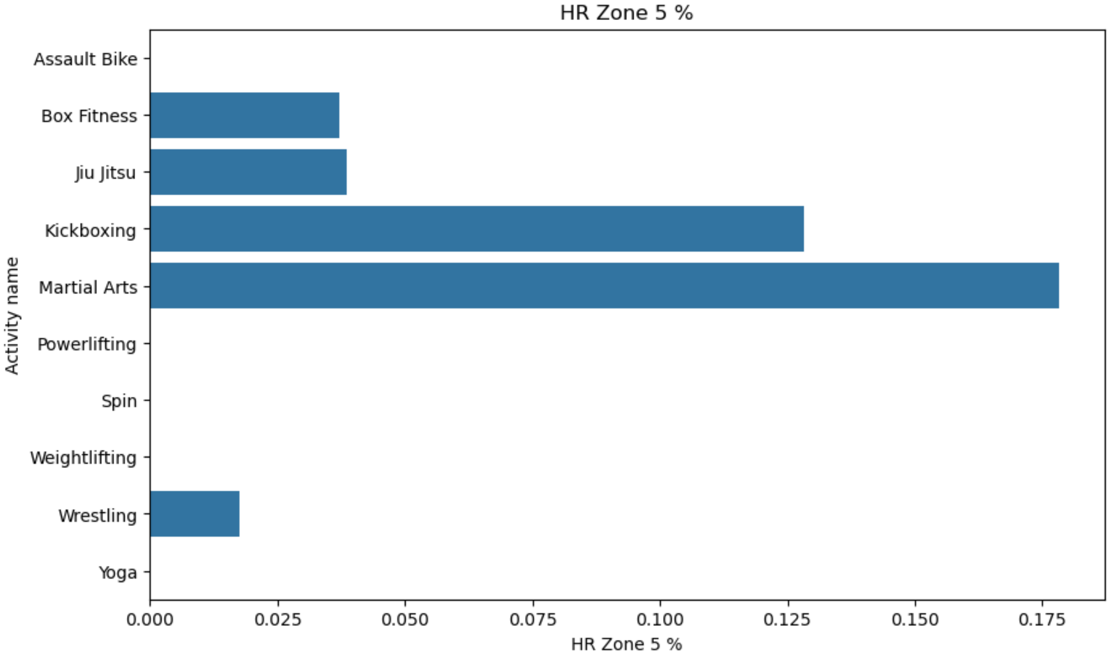

Average HR (bpm)
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0
20
40
60
80
100
120
140
Average HR (bpm)


HR Zone 1 %
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
HR Zone 1 %


HR Zone 2 %
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
HR Zone 2 %

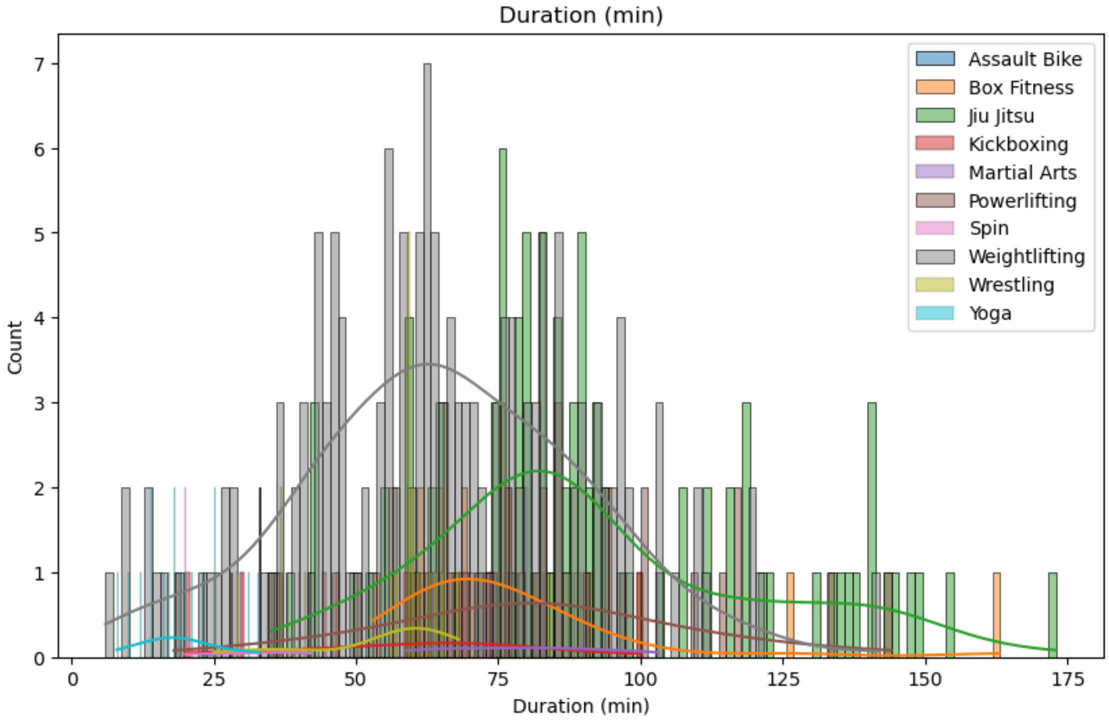

HR Zone 3 %
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
HR Zone 3 %

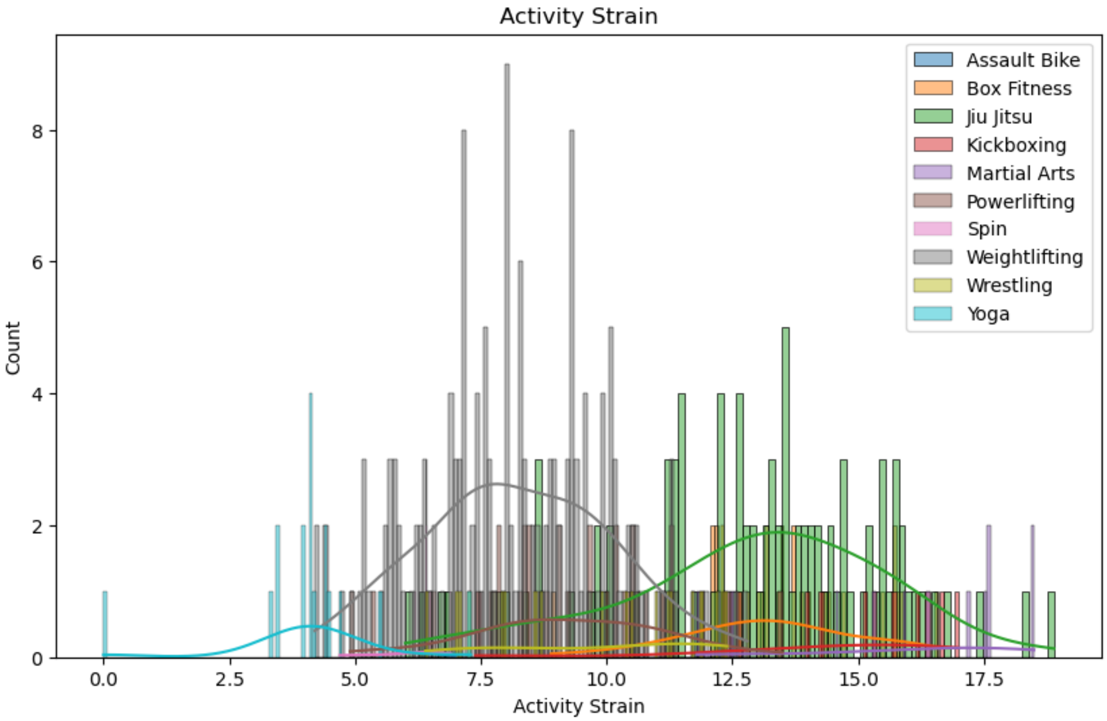

HR Zone 4 %
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0.00
0.05
0.10
0.15
0.20
0.25
0.30
HR Zone 4 %


HR Zone 5 %
Activity name
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
HR Zone 5 %

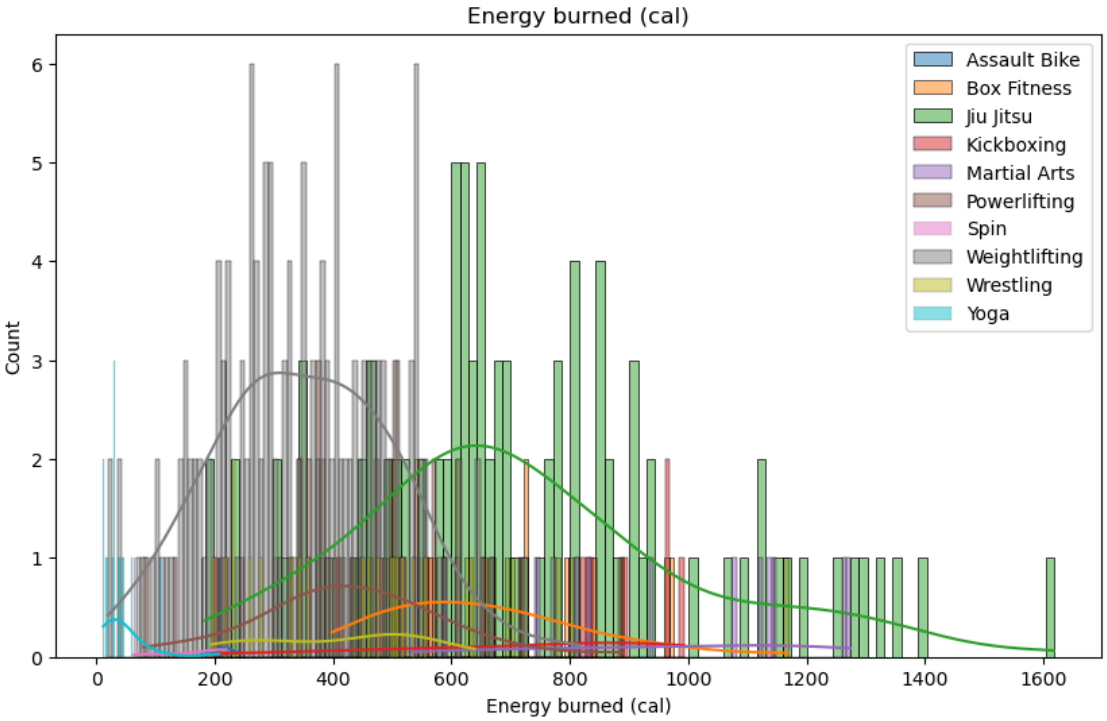
5.5.3 Class Distributions

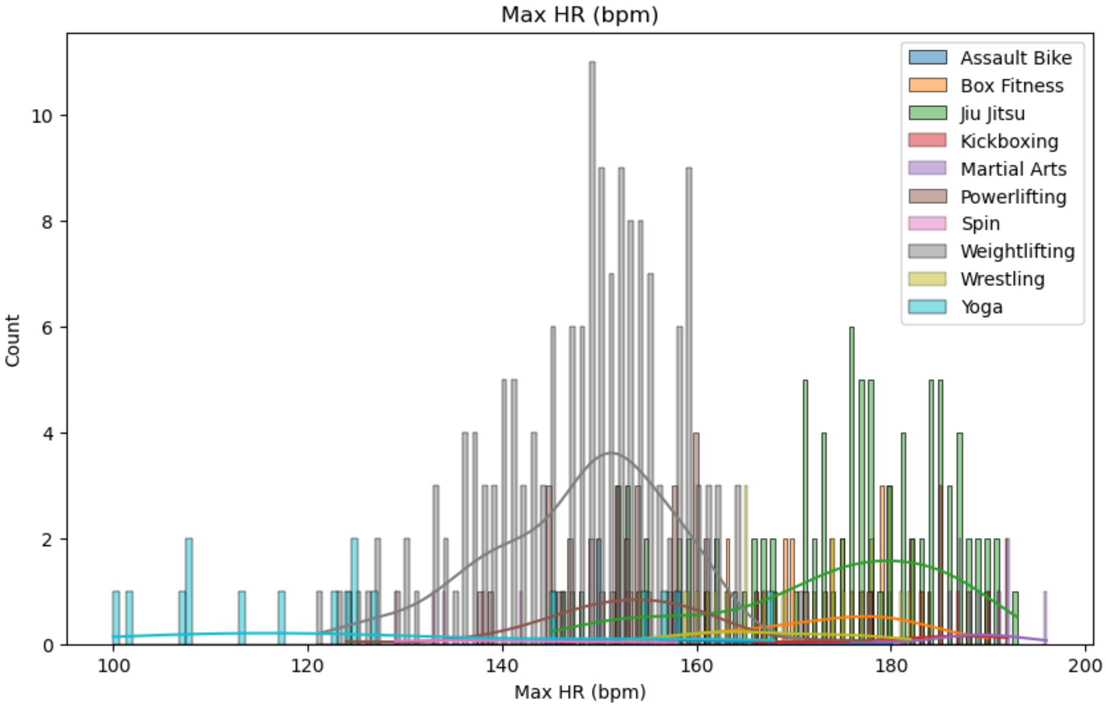
```
[ ]: for column in WO_AVG_COLS:
         plt.figure(figsize=(10, 6))

         for activity in df_wo_avg['Activity name'].unique():
             sns.histplot(
                 df_wo_add[column][df_wo_add['Activity name'] == activity],
                 bins=100,
                 kde=True,
                 label=activity
             )

         plt.legend()
         plt.title(f'{column}')
         plt.show()
```

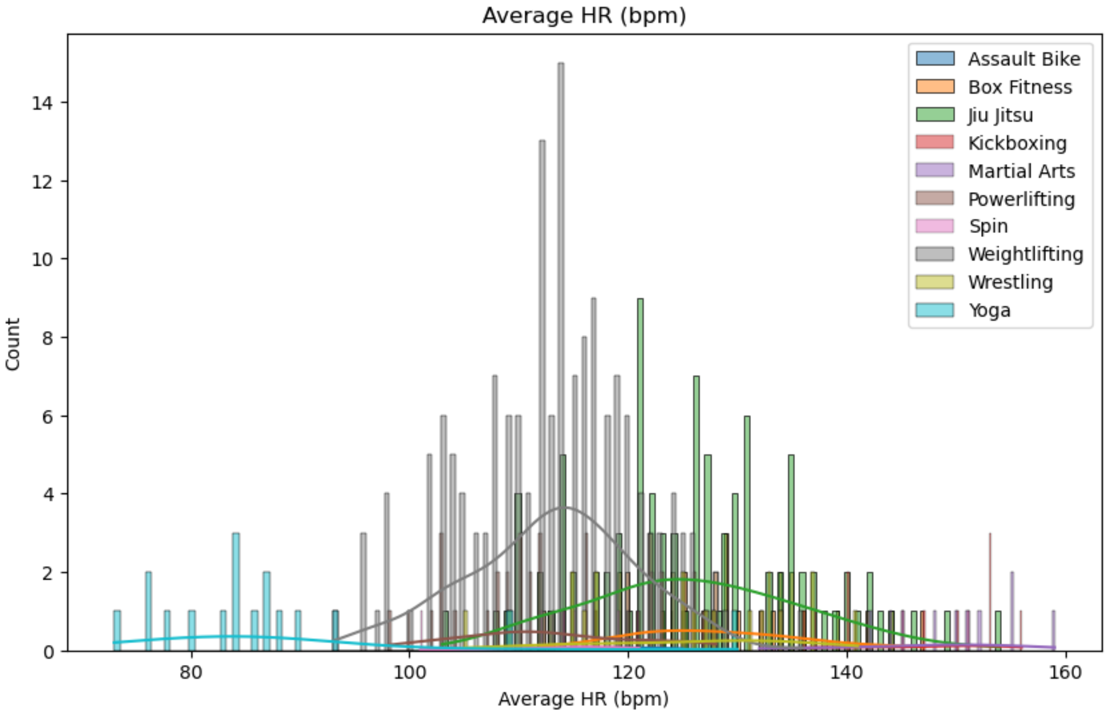


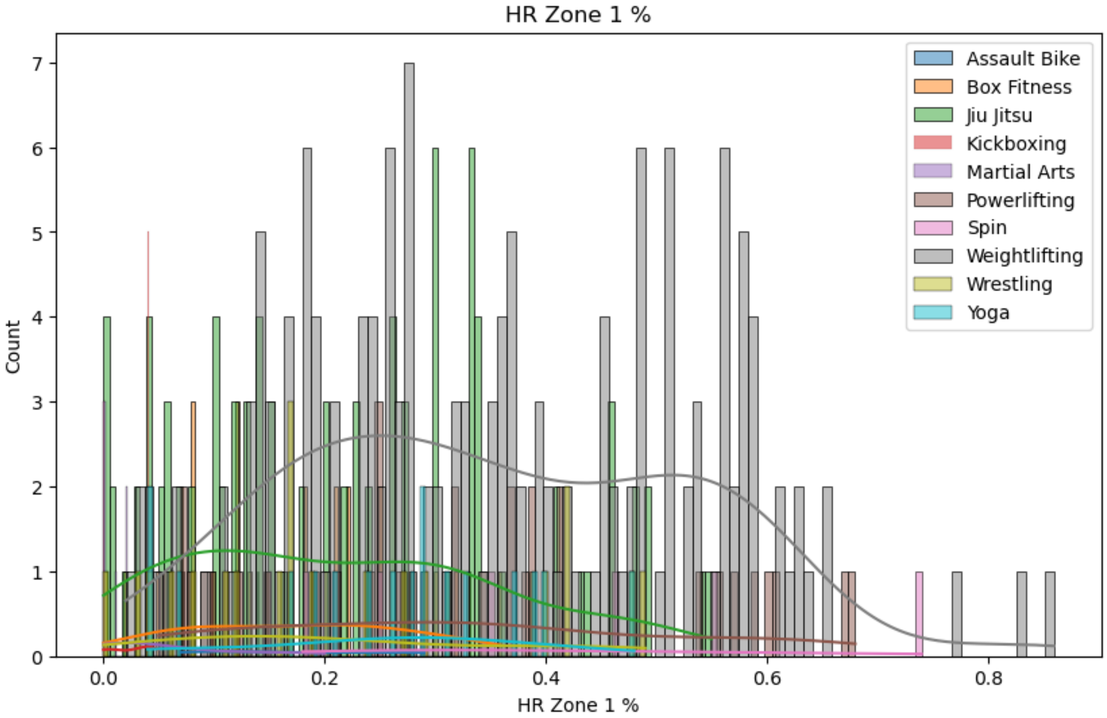


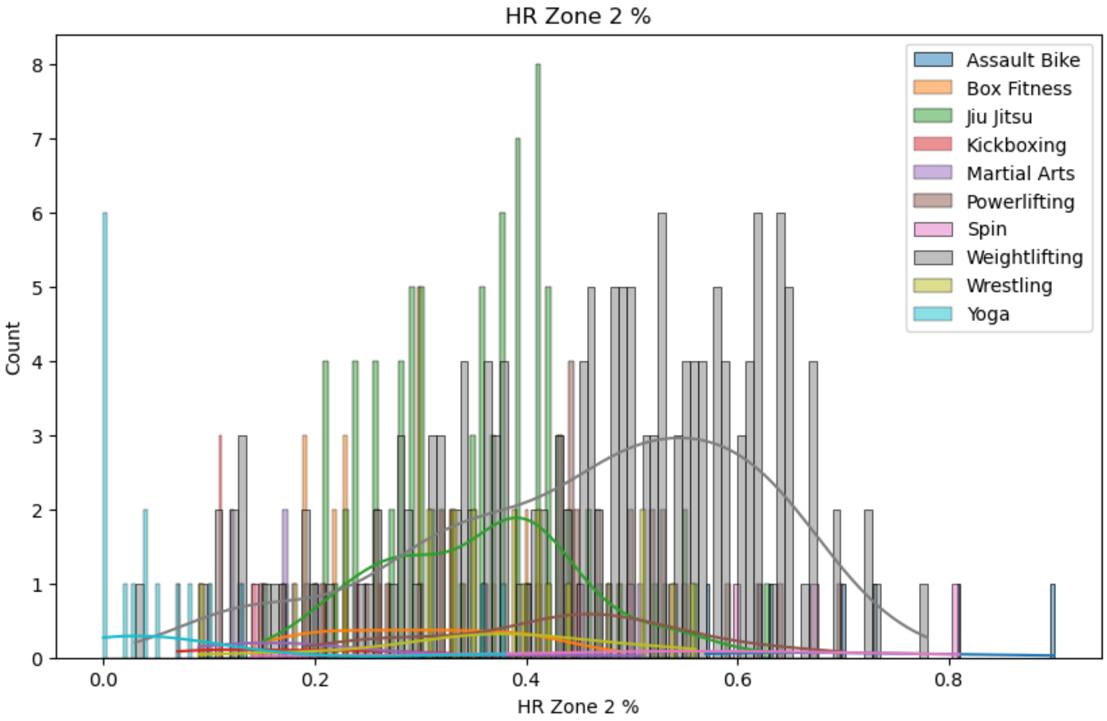
```
Exception ignored in: <function WeakKeyDictionary.__init__.<locals>.remove at
0x137ea3910>
Traceback (most recent call last):
  File "/Users/tyler/miniconda3/envs/ds310/lib/python3.10/weakref.py", line 370,
in remove
    def remove(k, selfref=ref(self)):
KeyboardInterrupt:
```

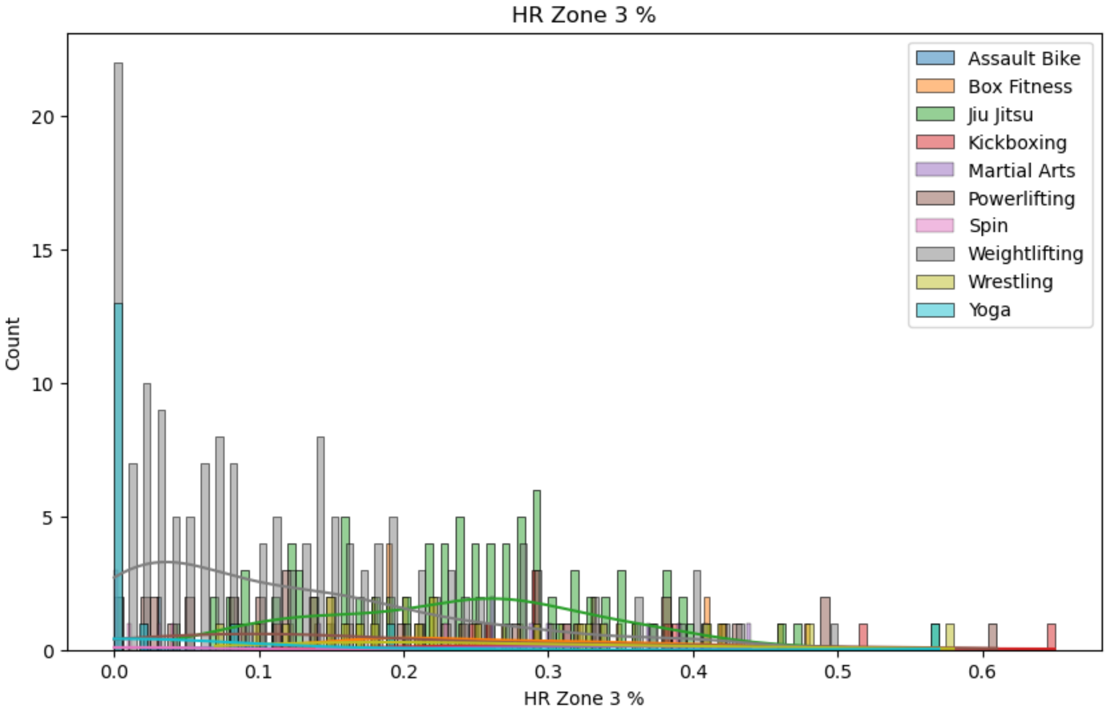

Energy burned (cal)
Count
0
1
2
3
4
5
6
0
200
400
600
800
1000
1200
1400
1600
Energy burned (cal)
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga

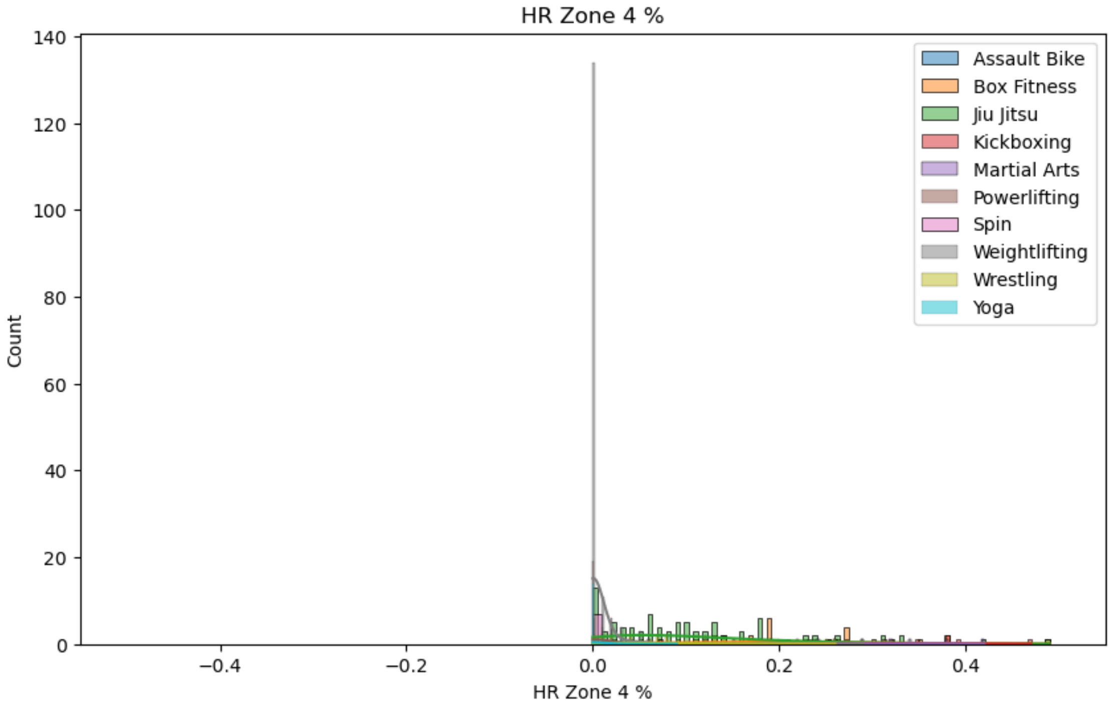

Max HR (bpm)
Count
0
2
4
6
8
10
100
120
140
160
180
200
Max HR (bpm)
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga

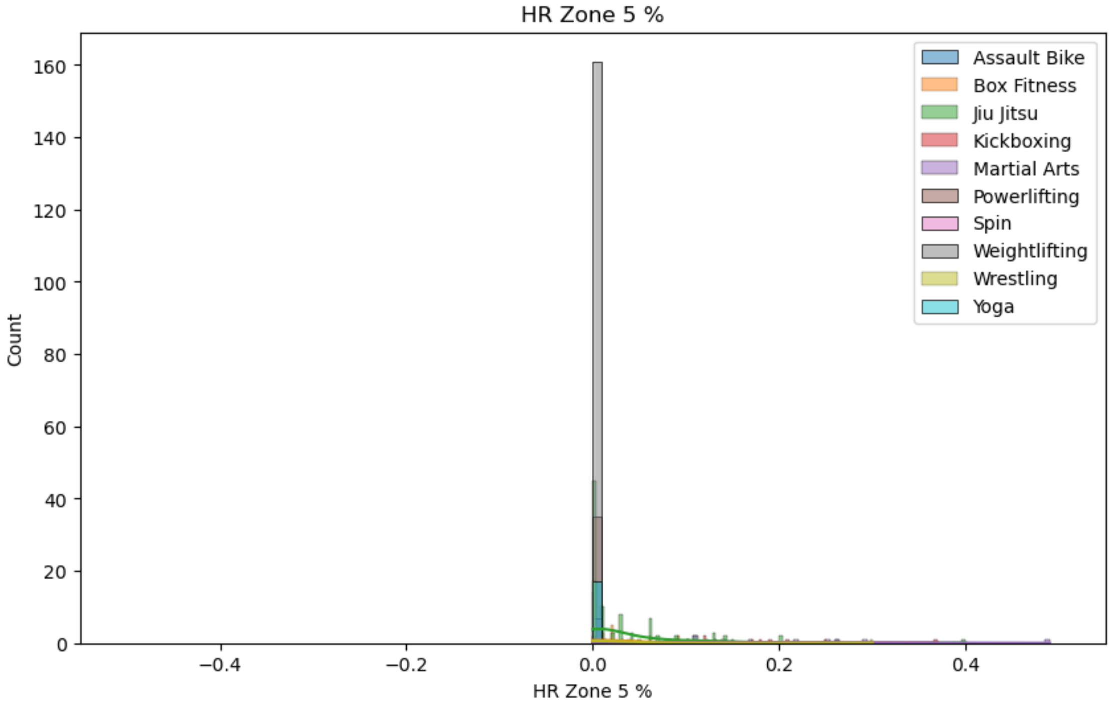

Average HR (bpm)
Count
Average HR (bpm)
0
2
4
6
8
10
12
14
80
100
120
140
160
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga


HR Zone 1 %
Count
HR Zone 1 %
0
1
2
3
4
5
6
7
0.0
0.2
0.4
0.6
0.8
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga


HR Zone 2 %
Count
0
1
2
3
4
5
6
7
8
0.0
0.2
0.4
0.6
0.8
HR Zone 2 %
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga

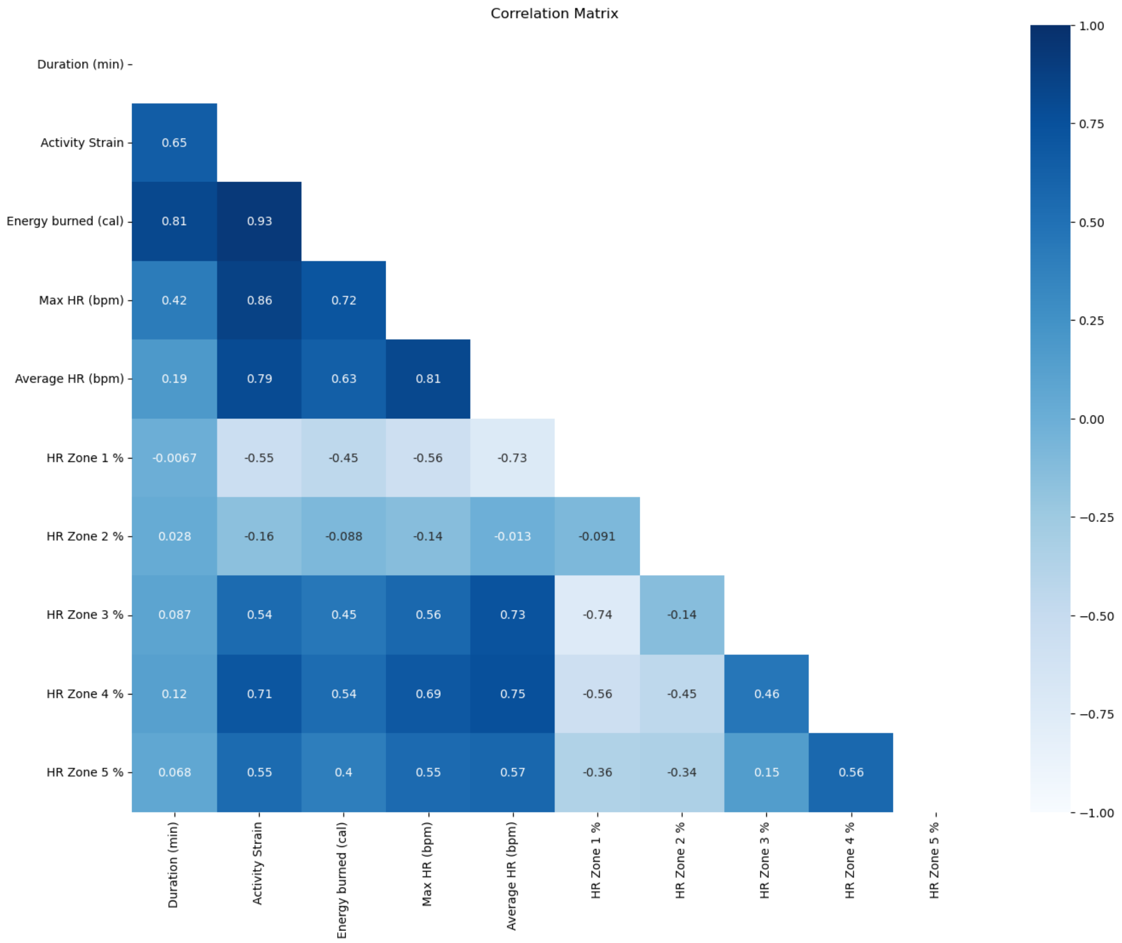

HR Zone 3 %
Count
0
5
10
15
20
0.0
0.1
0.2
0.3
0.4
0.5
0.6
HR Zone 3 %
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga


HR Zone 4 %
Count
140
120
100
80
60
40
20
0
−0.4
−0.2
0.0
0.2
0.4
HR Zone 4 %
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga


HR Zone 5 %
Count
160
140
120
100
80
60
40
20
0
−0.4
−0.2
0.0
0.2
0.4
HR Zone 5 %
Assault Bike
Box Fitness
Jiu Jitsu
Kickboxing
Martial Arts
Powerlifting
Spin
Weightlifting
Wrestling
Yoga

5.5.4 Correlations

```
[ ]: display_correlation_matrix(df_wo_add)
```



5.6 Export

```
[ ]: df_wo_add.to_csv(os.path.join(WHOOP_EXPORT_DIR, 'wo_add.csv'), index=False)
```