fitbit-analysis-part1-1

February 2, 2024

1 FITBIT DATA ANALYSIS PROJECT (DIPEN PRADHAN & ROSHANI SINGH)

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import altair as alt
```

[2]:
```
! pip install altair
```

```
Requirement already satisfied: altair in c:\users\avili\anaconda3\lib\site-
packages (5.2.0)
Requirement already satisfied: jinja2 in c:\users\avili\anaconda3\lib\site-
packages (from altair) (3.1.2)
Requirement already satisfied: jsonschema>=3.0 in
c:\users\avili\anaconda3\lib\site-packages (from altair) (4.17.3)
Requirement already satisfied: numpy in c:\users\avili\anaconda3\lib\site-
packages (from altair) (1.24.3)
Requirement already satisfied: packaging in c:\users\avili\anaconda3\lib\site-
packages (from altair) (23.1)
Requirement already satisfied: pandas>=0.25 in
c:\users\avili\anaconda3\lib\site-packages (from altair) (2.0.3)
Requirement already satisfied: toolz in c:\users\avili\anaconda3\lib\site-
packages (from altair) (0.12.0)
Requirement already satisfied: attrs>=17.4.0 in
c:\users\avili\anaconda3\lib\site-packages (from jsonschema>=3.0->altair)
(22.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
c:\users\avili\anaconda3\lib\site-packages (from jsonschema>=3.0->altair)
(0.18.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\avili\anaconda3\lib\site-packages (from pandas>=0.25->altair) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\avili\anaconda3\lib\site-packages (from pandas>=0.25->altair)
(2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in
```

```
c:\users\avili\anaconda3\lib\site-packages (from pandas>=0.25->altair) (2023.3)
Requirement already satisfied: MarkupSafe>=2.0 in
c:\users\avili\anaconda3\lib\site-packages (from jinja2->altair) (2.1.1)
Requirement already satisfied: six>=1.5 in c:\users\avili\anaconda3\lib\site-
packages (from python-dateutil>=2.8.2->pandas>=0.25->altair) (1.16.0)
```

[3]:
```
dailyActivity = pd.read_csv("C:\\Users\\avili\\Downloads\\dailyActivity_merged.
 ↪csv")
dailycalories = pd.read_csv("C:\\Users\\avili\\Downloads\\dailyCalories_merged.
 ↪csv")
dailyIntensities  = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\dailyIntensities_merged.csv")
dailySteps = pd.read_csv("C:\\Users\\avili\\Downloads\\dailySteps_merged.csv")
heartrate_seconds = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\heartrate_seconds_merged (1).csv")
hourlyCalories = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\hourlyCalories_merged.csv")
hourlyIntensities = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\hourlyIntensities_merged.csv")
hourlySteps = pd.read_csv("C:\\Users\\avili\\Downloads\\hourlySteps_merged.csv")
minuteCaloriesNarrow = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\minuteCaloriesNarrow_merged (2).csv")
minuteCaloriesWide = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\minuteCaloriesWide_merged.csv")
minuteIntensitiesNarrow = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\minuteIntensitiesNarrow_merged.csv")
minuteIntensitiesWide = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\minuteIntensitiesWide_merged.csv")
minuteMETsNarrow = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\minuteMETsNarrow_merged.csv")
minuteSleep = pd.read_csv("C:\\Users\\avili\\Downloads\\minuteSleep_merged.csv")
minuteStepsNarrow = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\minuteStepsNarrow_merged.csv")
minuteStepsWide = pd.read_csv("C:
 ↪\\Users\\avili\\Downloads\\minuteStepsWide_merged.csv")
sleepDay = pd.read_csv("C:\\Users\\avili\\Downloads\\sleepDay_merged (1).csv")
weightLogInfo = pd.read_csv("C:\\Users\\avili\\Downloads\\weightLogInfo_merged␣
 ↪(1).csv")
```

[4]:
```
sleepDay.head()
```

[4]:

	Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	\
0	1503960366	4/12/2016 12:00:00 AM	1	327	
1	1503960366	4/13/2016 12:00:00 AM	2	384	
2	1503960366	4/15/2016 12:00:00 AM	1	412	
3	1503960366	4/16/2016 12:00:00 AM	2	340	

```
4  1503960366  4/17/2016 12:00:00 AM                1                  700

   TotalTimeInBed
0             346
1             407
2             442
3             367
4             712
```

[5]:
```
sleepDay['Id'].unique()
```

[5]:
```
array([1503960366, 1644430081, 1844505072, 1927972279, 2026352035,
       2320127002, 2347167796, 3977333714, 4020332650, 4319703577,
       4388161847, 4445114986, 4558609924, 4702921684, 5553957443,
       5577150313, 6117666160, 6775888955, 6962181067, 7007744171,
       7086361926, 8053475328, 8378563200, 8792009665], dtype=int64)
```

[6]:
```
weightLogInfo.head(30)
```

[6]:

	Id	Date	WeightKg	WeightPounds	Fat	\
0	1503960366	5/2/2016 11:59:59 PM	52.599998	115.963147	22.0	
1	1503960366	5/3/2016 11:59:59 PM	52.599998	115.963147	NaN	
2	1927972279	4/13/2016 1:08:52 AM	133.500000	294.317120	NaN	
3	2873212765	4/21/2016 11:59:59 PM	56.700001	125.002104	NaN	
4	2873212765	5/12/2016 11:59:59 PM	57.299999	126.324875	NaN	
5	4319703577	4/17/2016 11:59:59 PM	72.400002	159.614681	25.0	
6	4319703577	5/4/2016 11:59:59 PM	72.300003	159.394222	NaN	
7	4558609924	4/18/2016 11:59:59 PM	69.699997	153.662190	NaN	
8	4558609924	4/25/2016 11:59:59 PM	70.300003	154.984977	NaN	
9	4558609924	5/1/2016 11:59:59 PM	69.900002	154.103125	NaN	
10	4558609924	5/2/2016 11:59:59 PM	69.199997	152.559879	NaN	
11	4558609924	5/9/2016 11:59:59 PM	69.099998	152.339420	NaN	
12	5577150313	4/17/2016 9:17:55 AM	90.699997	199.959265	NaN	
13	6962181067	4/12/2016 11:59:59 PM	62.500000	137.788914	NaN	
14	6962181067	4/13/2016 11:59:59 PM	62.099998	136.907061	NaN	
15	6962181067	4/14/2016 11:59:59 PM	61.700001	136.025217	NaN	
16	6962181067	4/15/2016 11:59:59 PM	61.500000	135.584291	NaN	
17	6962181067	4/16/2016 11:59:59 PM	62.000000	136.686603	NaN	
18	6962181067	4/17/2016 11:59:59 PM	61.400002	135.363832	NaN	
19	6962181067	4/18/2016 11:59:59 PM	61.200001	134.922906	NaN	
20	6962181067	4/19/2016 11:59:59 PM	61.400002	135.363832	NaN	
21	6962181067	4/20/2016 11:59:59 PM	61.700001	136.025217	NaN	
22	6962181067	4/21/2016 11:59:59 PM	61.400002	135.363832	NaN	
23	6962181067	4/22/2016 11:59:59 PM	61.400002	135.363832	NaN	
24	6962181067	4/23/2016 11:59:59 PM	61.500000	135.584291	NaN	
25	6962181067	4/24/2016 11:59:59 PM	61.500000	135.584291	NaN	
26	6962181067	4/25/2016 11:59:59 PM	61.700001	136.025217	NaN	

```
27  6962181067  4/27/2016 11:59:59 PM   61.200001    134.922906   NaN
28  6962181067  4/28/2016 11:59:59 PM   61.200001    134.922906   NaN
29  6962181067  4/29/2016 11:59:59 PM   61.400002    135.363832   NaN
```

	BMI	IsManualReport	LogId
0	22.650000	True	1462233599000
1	22.650000	True	1462319999000
2	47.540001	False	1460509732000
3	21.450001	True	1461283199000
4	21.690001	True	1463097599000
5	27.450001	True	1460937599000
6	27.379999	True	1462406399000
7	27.250000	True	1461023999000
8	27.459999	True	1461628799000
9	27.320000	True	1462147199000
10	27.040001	True	1462233599000
11	27.000000	True	1462838399000
12	28.000000	False	1460884675000
13	24.389999	True	1460505599000
14	24.240000	True	1460591999000
15	24.100000	True	1460678399000
16	24.000000	True	1460764799000
17	24.209999	True	1460851199000
18	23.959999	True	1460937599000
19	23.889999	True	1461023999000
20	23.959999	True	1461110399000
21	24.100000	True	1461196799000
22	23.959999	True	1461283199000
23	23.959999	True	1461369599000
24	24.000000	True	1461455999000
25	24.000000	True	1461542399000
26	24.100000	True	1461628799000
27	23.889999	True	1461801599000
28	23.889999	True	1461887999000
29	23.959999	True	1461974399000

[7]:
```
weightLogInfo.shape
```

[7]:
```
(67, 8)
```

[8]:
```
weightLogInfo['Id'].value_counts()
```

[8]:
```
Id
6962181067    30
8877689391    24
4558609924     5
1503960366     2
```

```
2873212765     2
4319703577     2
1927972279     1
5577150313     1
Name: count, dtype: int64
```

[9]:
```
# to find out how many records are manual reporting
weightLogInfo['IsManualReport'].value_counts()
```

[9]:
```
IsManualReport
True     41
False    26
Name: count, dtype: int64
```

[10]:
```
#Out of 67 records, 41 records were mannually reported. It means the fitness␣
 ↪tracker is not automatically tracking the weight for all the users.
#It depends on what kind of Fitness tracker is being used. Weight is a␣
 ↪essential metrics for a healthly body,
#and we can track the weight of the participants using a smart device connected␣
 ↪to the user's account.
```

[11]:
```
weightLogInfo['Id'].dtype
```

[11]:
```
dtype('int64')
```

[12]:
```
weightLogInfo['Id'].astype('object')
```

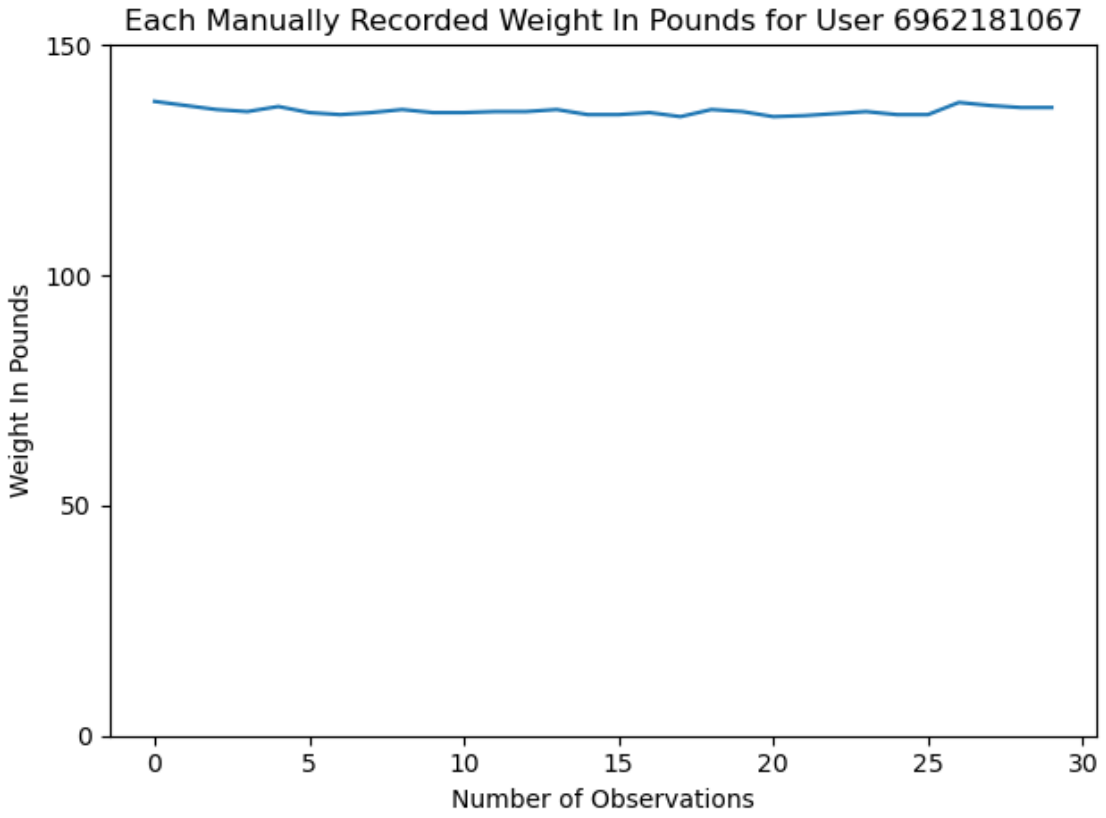
[12]:
```
0     1503960366
1     1503960366
2     1927972279
3     2873212765
4     2873212765
         …
62    8877689391
63    8877689391
64    8877689391
65    8877689391
66    8877689391
Name: Id, Length: 67, dtype: object
```

[13]:
```
weightLogInfo['Fat'].isnull().sum()
```

[13]:
```
65
```

[14]:
```
# let's check the weight records for one particular id who has 30 records to␣
 ↪see how the data is.
weightLogInfo1 = weightLogInfo[weightLogInfo['Id']==6962181067].reset_index()
```

```
weightLogInfo1['WeightPounds'].plot(kind='line')
plt.yticks(ticks=range(0,160,50),labels=range(0,160,50))
plt.title('Each Manually Recorded Weight In Pounds for User 6962181067')
plt.xlabel('Number of Observations')
plt.ylabel('Weight In Pounds')
plt.tight_layout()
```



[15]:
```
# Finding the average weight of each user from the dataset in ascending order␣
 ↪by WeightPounds with their respective average BMI
df1 = weightLogInfo.groupby('Id')[['WeightPounds','BMI']].mean().
 ↪sort_values(by='WeightPounds').reset_index()
df1
```

[15]:

	Id	WeightPounds	BMI
0	1503960366	115.963147	22.650000
1	2873212765	125.663489	21.570001
2	6962181067	135.701872	24.028000
3	4558609924	153.529918	27.214000
4	4319703577	159.504452	27.415000
5	8877689391	187.714432	25.487083

```
6  5577150313    199.959265  28.000000
7  1927972279    294.317120  47.540001
```

[16]:
```
# Calculate the BMI range
#BMI Range to measure Overall Health
# Underweight (Below 18.5)
#Healthy Weight (18.5 - 24.9)#
#Overweight (25.0 - 29.9)
#Obesity (30.0 and Above)
for i in df1.index:
    if (df1.loc[i,'BMI']>24.9):
        df1.loc[i,'Overweight'] = 'Yes'
    else:
        df1.loc[i,'Overweight'] = 'No'
# to check the new column and the data
df1
```

[16]:
```
           Id  WeightPounds        BMI Overweight
0  1503960366    115.963147  22.650000         No
1  2873212765    125.663489  21.570001         No
2  6962181067    135.701872  24.028000         No
3  4558609924    153.529918  27.214000        Yes
4  4319703577    159.504452  27.415000        Yes
5  8877689391    187.714432  25.487083        Yes
6  5577150313    199.959265  28.000000        Yes
7  1927972279    294.317120  47.540001        Yes
```

[17]:
```
# the Id(1927972279) hs the highest weight pound which is overweight it should␣
 ↪be Analysis
```

[18]:
```
# Checking the average sleeptime and average total time in bed of the id with␣
 ↪heaviest weight and highest BMI
new_df = sleepDay[sleepDay['Id']==1927972279]
new_df.groupby('Id')[['TotalMinutesAsleep','TotalTimeInBed']].mean()
```

[18]:
```
            TotalMinutesAsleep  TotalTimeInBed
Id
1927972279               417.0           437.8
```

[19]:
```
dailyActivity.head()
```

[19]:
```
           Id ActivityDate  TotalSteps  TotalDistance  TrackerDistance  \
0  1503960366    4/12/2016       13162           8.50             8.50
1  1503960366    4/13/2016       10735           6.97             6.97
2  1503960366    4/14/2016       10460           6.74             6.74
3  1503960366    4/15/2016        9762           6.28             6.28
4  1503960366    4/16/2016       12669           8.16             8.16
```

```
   LoggedActivitiesDistance  VeryActiveDistance  ModeratelyActiveDistance  \
0                       0.0                1.88                      0.55
1                       0.0                1.57                      0.69
2                       0.0                2.44                      0.40
3                       0.0                2.14                      1.26
4                       0.0                2.71                      0.41

   LightActiveDistance  SedentaryActiveDistance  VeryActiveMinutes  \
0                 6.06                      0.0                 25
1                 4.71                      0.0                 21
2                 3.91                      0.0                 30
3                 2.83                      0.0                 29
4                 5.04                      0.0                 36

   FairlyActiveMinutes  LightlyActiveMinutes  SedentaryMinutes  Calories
0                   13                   328               728      1985
1                   19                   217               776      1797
2                   11                   181              1218      1776
3                   34                   209               726      1745
4                   10                   221               773      1863
```

[20]:
```
dailycalories.head()
```

[20]:
```
           Id ActivityDay  Calories
0  1503960366   4/12/2016      1985
1  1503960366   4/13/2016      1797
2  1503960366   4/14/2016      1776
3  1503960366   4/15/2016      1745
4  1503960366   4/16/2016      1863
```

[21]:
```
dailyActivity['Id'].value_counts()
```

[21]:
```
Id
1503960366    31
4319703577    31
8583815059    31
8378563200    31
8053475328    31
7086361926    31
6962181067    31
5553957443    31
4702921684    31
4558609924    31
1624580081    31
4388161847    31
4445114986    31
```

```
8877689391    31
1927972279    31
2873212765    31
2320127002    31
4020332650    31
2026352035    31
1844505072    31
2022484408    31
3977333714    30
1644430081    30
5577150313    30
8792009665    29
6290855005    29
6117666160    28
6775888955    26
7007744171    26
3372868164    20
8253242879    19
2347167796    18
4057192912     4
Name: count, dtype: int64
```

[22]:
```
dailycalories['Id'].value_counts()
```

[22]:
```
Id
1503960366    31
4319703577    31
8583815059    31
8378563200    31
8053475328    31
7086361926    31
6962181067    31
5553957443    31
4702921684    31
4558609924    31
1624580081    31
4388161847    31
4445114986    31
8877689391    31
1927972279    31
2873212765    31
2320127002    31
4020332650    31
2026352035    31
1844505072    31
2022484408    31
3977333714    30
```

```
1644430081    30
5577150313    30
8792009665    29
6290855005    29
6117666160    28
6775888955    26
7007744171    26
3372868164    20
8253242879    19
2347167796    18
4057192912     4
Name: count, dtype: int64
```

[23]:
```
activitycalorymerged = dailyActivity.
↪groupby('Id')[['TotalSteps','TotalDistance','TrackerDistance','LoggedActivitiesDistance','V
↪'ModeratelyActiveDistance','LightActiveDistance','SedentaryActiveDistance','VeryActiveMinut
↪'FairlyActiveMinutes','LightlyActiveMinutes','SedentaryMinutes','Calories']].
↪mean().reset_index()
```

[24]:
```
activitycalorymerged
```

[24]:

	Id	TotalSteps	TotalDistance	TrackerDistance	\
0	1503960366	12116.741935	7.809677	7.809677	
1	1624580081	5743.903226	3.914839	3.914839	
2	1644430081	7282.966667	5.295333	5.295333	
3	1844505072	2580.064516	1.706129	1.706129	
4	1927972279	916.129032	0.634516	0.634516	
5	2022484408	11370.645161	8.084193	8.084193	
6	2026352035	5566.870968	3.454839	3.454839	
7	2320127002	4716.870968	3.187742	3.187742	
8	2347167796	9519.666667	6.355556	6.355556	
9	2873212765	7555.774194	5.101613	5.101613	
10	3372868164	6861.650000	4.707000	4.707000	
11	3977333714	10984.566667	7.517000	7.517000	
12	4020332650	2267.225806	1.626129	1.626129	
13	4057192912	3838.000000	2.862500	2.862500	
14	4319703577	7268.838710	4.892258	4.892258	
15	4388161847	10813.935484	8.393226	8.393226	
16	4445114986	4796.548387	3.245806	3.245806	
17	4558609924	7685.129032	5.080645	5.080645	
18	4702921684	8572.064516	6.955161	6.955161	
19	5553957443	8612.580645	5.639677	5.639677	
20	5577150313	8304.433333	6.213333	6.213333	
21	6117666160	7046.714286	5.342143	5.342143	
22	6290855005	5649.551724	4.272414	4.272414	

```
23  6775888955   2519.692308       1.813462         1.813462
24  6962181067   9794.806452       6.585806         6.519355
25  7007744171  11323.423077       8.015385         7.575769
26  7086361926   9371.774194       6.388065         6.388065
27  8053475328  14763.290323      11.475161        11.475161
28  8253242879   6482.157895       4.667368         4.667368
29  8378563200   8717.709677       6.913548         6.913548
30  8583815059   7198.516129       5.615484         5.615484
31  8792009665   1853.724138       1.186552         1.186552
32  8877689391  16040.032258      13.212903        13.212903

    LoggedActivitiesDistance  VeryActiveDistance  ModeratelyActiveDistance  \
0                   0.000000            2.858387                  0.794194
1                   0.000000            0.939355                  0.360645
2                   0.000000            0.730000                  0.951000
3                   0.000000            0.008387                  0.049032
4                   0.000000            0.095806                  0.031290
5                   0.000000            2.421613                  0.720000
6                   0.000000            0.006129                  0.011290
7                   0.000000            0.106774                  0.097742
8                   0.000000            1.059444                  1.075000
9                   0.000000            0.676129                  0.276129
10                  0.000000            0.629500                  0.153000
11                  0.000000            1.615000                  2.751000
12                  0.000000            0.142258                  0.129677
13                  0.000000            0.052500                  0.065000
14                  0.000000            0.278065                  0.502258
15                  0.000000            1.719355                  0.901935
16                  0.000000            0.523226                  0.075484
17                  0.000000            0.549355                  0.682258
18                  0.000000            0.417419                  1.304839
19                  0.000000            1.464194                  0.669032
20                  0.000000            3.113667                  0.658000
21                  0.000000            0.128214                  0.083929
22                  0.000000            0.085517                  0.128276
23                  0.075369            0.709231                  0.384231
24                  0.323700            1.616452                  0.960000
25                  2.118673            2.415000                  0.738462
26                  0.000000            2.781290                  0.773226
27                  0.000000            8.514839                  0.423871
28                  0.000000            2.214210                  0.695789
29                  1.116158            2.503548                  0.519032
30                  0.000000            0.798065                  1.020645
31                  0.000000            0.024828                  0.058276
32                  0.000000            6.637419                  0.337742

    LightActiveDistance  SedentaryActiveDistance  VeryActiveMinutes  \
```

0	4.152903	0.000000	38.709677
1	2.606774	0.006129	8.677419
2	3.609000	0.004000	9.566667
3	1.647419	0.000000	0.129032
4	0.507097	0.000000	1.322581
5	4.942581	0.000000	36.290323
6	3.436129	0.000000	0.096774
7	2.980323	0.000000	1.354839
8	4.221667	0.000000	13.500000
9	4.143548	0.005161	14.096774
10	3.910000	0.011000	9.150000
11	3.134333	0.000000	18.900000
12	1.308387	0.005484	5.193548
13	2.687500	0.000000	0.750000
14	3.768710	0.000000	3.580645
15	5.396129	0.000000	23.161290
16	2.644839	0.000000	6.612903
17	3.847742	0.000000	10.387097
18	5.225484	0.000000	5.129032
19	3.504516	0.000000	23.419355
20	2.428000	0.000000	87.333333
21	4.843214	0.000000	1.571429
22	4.048621	0.008621	2.758621
23	0.711538	0.000000	11.000000
24	4.001613	0.006774	22.806452
25	4.861538	0.000769	31.038462
26	2.818710	0.000000	42.580645
27	2.533871	0.000000	85.161290
28	1.754737	0.000526	20.526316
29	3.889355	0.000000	58.677419
30	2.617419	0.000000	9.677419
31	1.103448	0.000000	0.965517
32	6.188710	0.005161	66.064516

	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes	Calories
0	19.161290	219.935484	848.161290	1816.419355
1	5.806452	153.483871	1257.741935	1483.354839
2	21.366667	178.466667	1161.866667	2811.300000
3	1.290323	115.451613	1206.612903	1573.483871
4	0.774194	38.580645	1317.419355	2172.806452
5	19.354839	257.451613	1112.580645	2509.967742
6	0.258065	256.645161	689.419355	1540.645161
7	2.580645	198.193548	1220.096774	1724.161290
8	20.555556	252.500000	687.166667	2043.444444
9	6.129032	308.000000	1097.193548	1916.967742
10	4.100000	327.900000	1077.550000	1933.100000
11	61.266667	174.766667	707.533333	1513.666667

```
12              5.354839              76.935484       1237.258065  2385.806452
13              1.500000             103.000000       1217.250000  1973.750000
14             12.322581             228.774194        735.806452  2037.677419
15             20.354839             229.354839        836.677419  3093.870968
16              1.741935             209.096774        829.903226  2186.193548
17             13.709677             284.967742       1093.612903  2033.258065
18             26.032258             237.483871        766.419355  2965.548387
19             13.000000             206.193548        668.354839  1875.677419
20             29.833333             147.933333        754.433333  3359.633333
21              2.035714             288.357143        796.285714  2261.142857
22              3.793103             227.448276       1193.034483  2599.620690
23             14.807692              40.153846       1299.423077  2131.769231
24             18.516129             245.806452        662.322581  1982.032258
25             16.269231             280.730769       1055.346154  2544.000000
26             25.354839             143.838710        850.451613  2566.354839
27              9.580645             150.967742       1148.000000  2945.806452
28             14.315789             116.894737       1287.368421  1788.000000
29             10.258065             156.096774        716.129032  3436.580645
30             22.193548             138.290323       1267.225806  2732.032258
31              4.034483              91.793103       1060.482759  1962.310345
32              9.935484             234.709677       1112.870968  3420.258065
```

[25]:
```
dailyIntensities.head()
```

[25]:
```
           Id ActivityDay  SedentaryMinutes  LightlyActiveMinutes  \
0  1503960366   4/12/2016               728                   328
1  1503960366   4/13/2016               776                   217
2  1503960366   4/14/2016              1218                   181
3  1503960366   4/15/2016               726                   209
4  1503960366   4/16/2016               773                   221

   FairlyActiveMinutes  VeryActiveMinutes  SedentaryActiveDistance  \
0                   13                 25                      0.0
1                   19                 21                      0.0
2                   11                 30                      0.0
3                   34                 29                      0.0
4                   10                 36                      0.0

   LightActiveDistance  ModeratelyActiveDistance  VeryActiveDistance
0                 6.06                      0.55                1.88
1                 4.71                      0.69                1.57
2                 3.91                      0.40                2.44
3                 2.83                      1.26                2.14
4                 5.04                      0.41                2.71
```

[26]:
```
activitycalorymerged.describe()
```

[26]:

	Id	TotalSteps	TotalDistance	TrackerDistance
count	3.300000e+01	33.000000	33.000000	33.000000
mean	4.857201e+09	7519.272678	5.398953	5.383618
std	2.433765e+09	3576.340125	2.772293	2.759479
min	1.503960e+09	916.129032	0.634516	0.634516
25%	2.347168e+09	5566.870968	3.454839	3.454839
50%	4.445115e+09	7282.966667	5.295333	5.295333
75%	6.962181e+09	9519.666667	6.913548	6.913548
max	8.877689e+09	16040.032258	13.212903	13.212903

	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActiveDistance
count	33.000000	33.000000	33.000000
mean	0.110118	1.449551	0.557039
std	0.412496	1.866091	0.537638
min	0.000000	0.006129	0.011290
25%	0.000000	0.142258	0.128276
50%	0.000000	0.730000	0.502258
75%	0.000000	2.214210	0.773226
max	2.118673	8.514839	2.751000

	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes
count	33.000000	33.000000	33.000000
mean	3.317450	0.001625	20.308769
std	1.375683	0.003020	23.803214
min	0.507097	0.000000	0.096774
25%	2.606774	0.000000	3.580645
50%	3.504516	0.000000	10.387097
75%	4.143548	0.000769	23.419355
max	6.188710	0.011000	87.333333

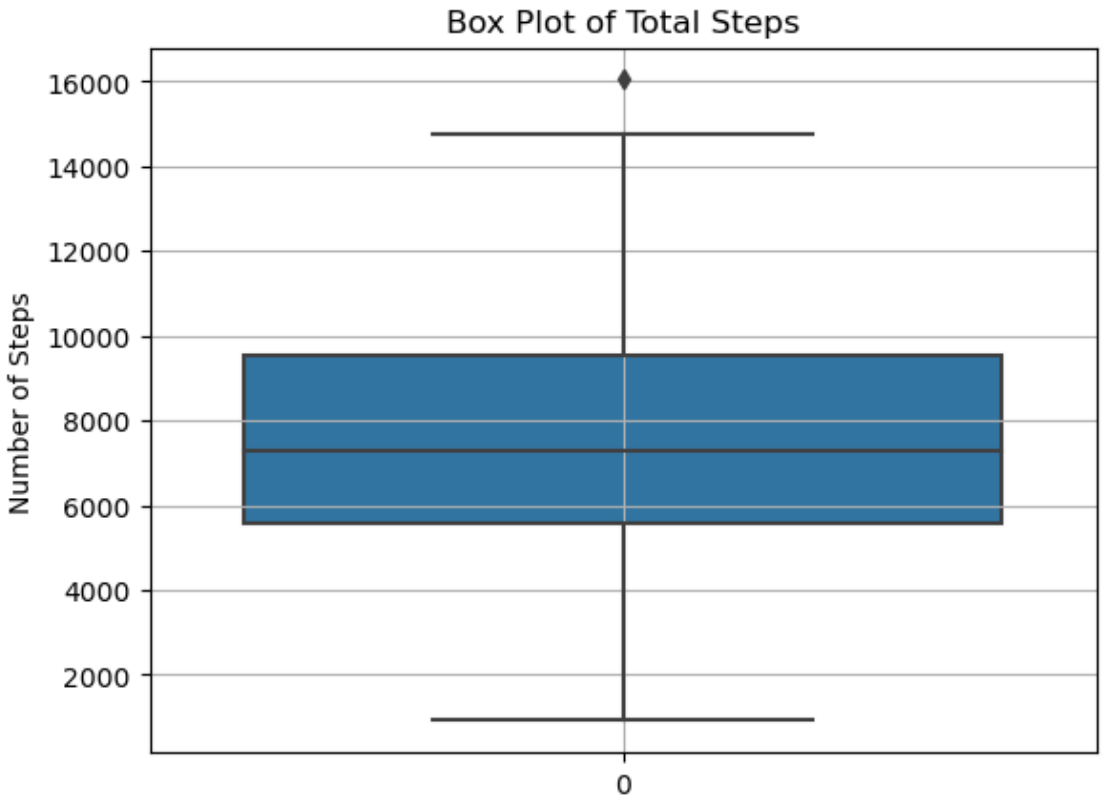
	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes
count	33.000000	33.000000	33.000000
mean	13.260240	191.521291	999.151475
std	12.108217	75.689747	227.678526
min	0.258065	38.580645	662.322581
25%	4.034483	143.838710	766.419355
50%	12.322581	206.193548	1077.550000
75%	19.354839	245.806452	1206.612903
max	61.266667	327.900000	1317.419355

	Calories
count	33.000000
mean	2282.443660
std	562.761632
min	1483.354839
25%	1916.967742
50%	2131.769231

```
75%     2599.620690
max     3436.580645
```

[27]:
```
#The average number of daily total steps of the participants are 7,519, which␣
 ↪is less
# the CDC recommended 8,000 steps. The maximum number of Total Steps is␣
 ↪16040,is very unusual and may be outlier data.
#We can remove or replace outlier data to clean the data for better analysis
```

[28]:
```
sns.boxplot(data=activitycalorymerged['TotalSteps'])
plt.grid(True)
plt.ylabel('Number of Steps')
plt.title('Box Plot of Total Steps')
plt.show()
```



[29]:
```
q3 = activitycalorymerged['TotalSteps'].quantile(0.75)
q1 = activitycalorymerged['TotalSteps'].quantile(0.25)
# now, i will calculate the inter quartile range
iqr = q3-q1
# now i will calculate the max value of the data, anything away from the max is␣
 ↪considered outliers value
```

```
outliers = q3 + (1.5* iqr)
```

[30]:
```
len(activitycalorymerged[activitycalorymerged['TotalSteps']>outliers])
```

[30]: 1

[31]:
```
len(activitycalorymerged['TotalSteps'])
```

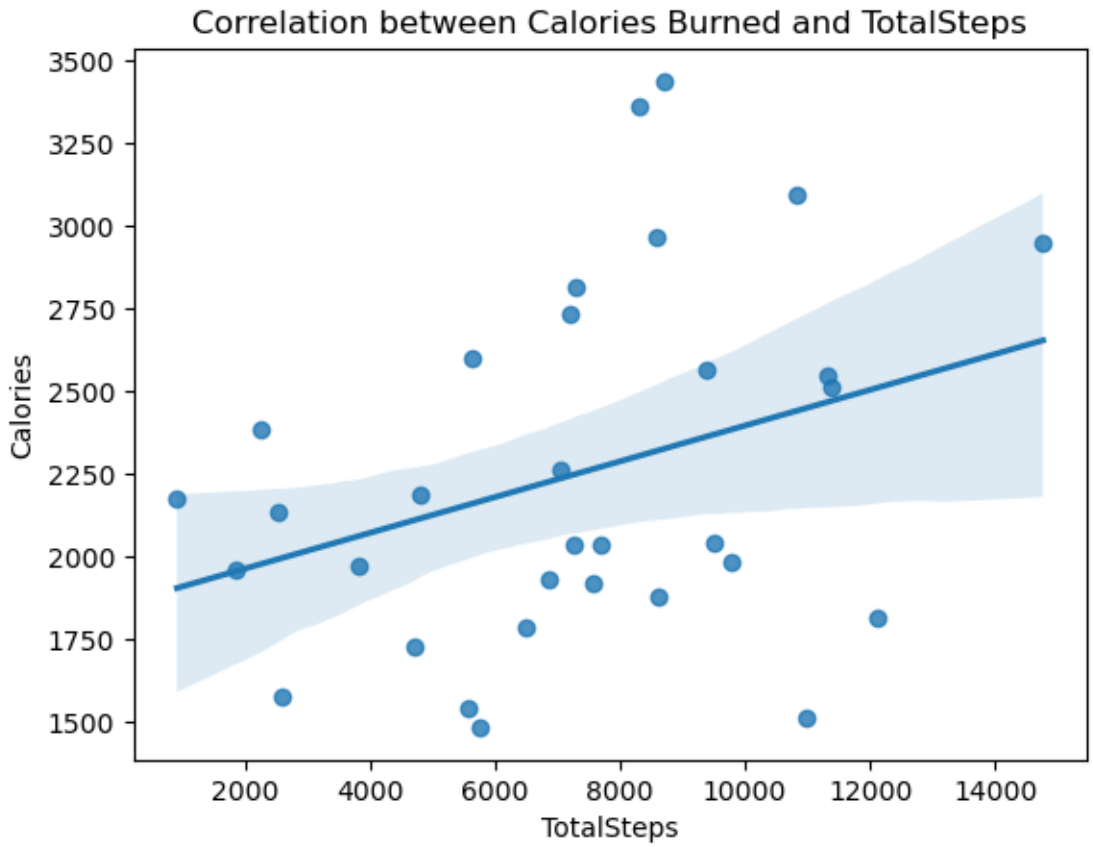
[31]: 33

[32]:
```
activitycalorymerged.
  ↪drop(activitycalorymerged[activitycalorymerged['TotalSteps']>outliers].
  ↪index,inplace=True)
```

[33]:
```
len(activitycalorymerged['TotalSteps'])
```

[33]: 32

[34]:
```
sns.regplot(data=activitycalorymerged,y='Calories',x='TotalSteps')
plt.title('Correlation between Calories Burned and TotalSteps')
plt.show()
```



[35]:
```
#The scatterplot shows us there is a direct proportional relationship between␣
 ↪Calories burned and Total daily steps.
#The more number of daily steps led to more calories burned.
```

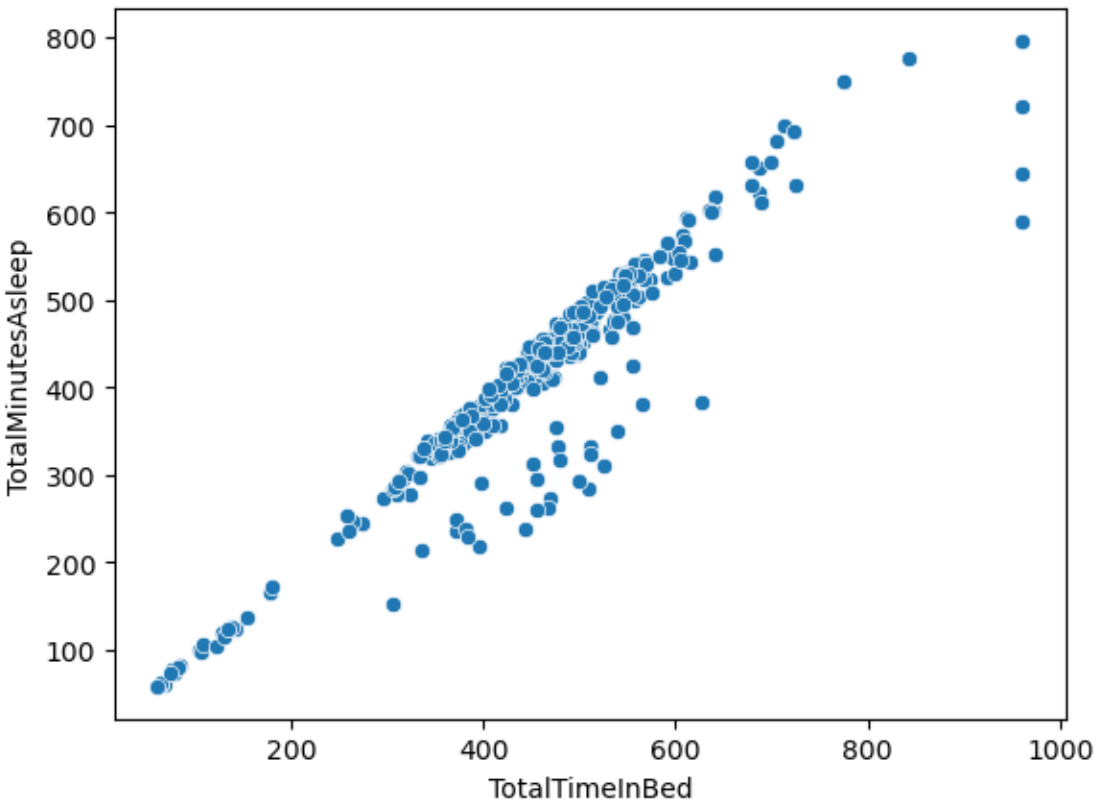
[36]:
```
#Checking statistical Analysis for sleepday dataset
sleepDay.describe()
```

[36]:

	Id	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed
count	4.130000e+02	413.000000	413.000000	413.000000
mean	5.000979e+09	1.118644	419.467312	458.639225
std	2.060360e+09	0.345521	118.344679	127.101607
min	1.503960e+09	1.000000	58.000000	61.000000
25%	3.977334e+09	1.000000	361.000000	403.000000
50%	4.702922e+09	1.000000	433.000000	463.000000
75%	6.962181e+09	1.000000	490.000000	526.000000
max	8.792010e+09	3.000000	796.000000	961.000000

[37]:
```
sns.scatterplot(data=sleepDay, x="TotalTimeInBed", y="TotalMinutesAsleep")
plt
```

[37]:
```
<module 'matplotlib.pyplot' from 'C:\\Users\\avili\\anaconda3\\Lib\\site-
packages\\matplotlib\\pyplot.py'>
```



[38]:
```
df_sleeptime = sleepDay.groupby('Id')[['TotalMinutesAsleep','TotalTimeInBed']].
↪mean().sort_values(by='TotalMinutesAsleep').reset_index()
df_sleeptime.head()
```

[38]:

	Id	TotalMinutesAsleep	TotalTimeInBed
0	2320127002	61.000000	69.000000
1	7007744171	68.500000	71.500000
2	4558609924	127.600000	140.000000
3	3977333714	293.642857	461.142857
4	1644430081	294.000000	346.000000

[39]:
```
# Create a new column Overweight to find if a user is Overweight or not.
for i in df_sleeptime .index:
    if (df_sleeptime .loc[i,'TotalMinutesAsleep']<480):
        df_sleeptime .loc[i,'lack_of_sleep'] = 'Yes'
    else:
        df_sleeptime .loc[i,'lack_of_sleep'] = 'No'
df_sleeptime .head()
```

[39]:

	Id	TotalMinutesAsleep	TotalTimeInBed	lack_of_sleep
0	2320127002	61.000000	69.000000	Yes

```
1  7007744171          68.500000       71.500000          Yes
2  4558609924         127.600000      140.000000          Yes
3  3977333714         293.642857      461.142857          Yes
4  1644430081         294.000000      346.000000          Yes
```

[40]:
```
df_sleeptime['lack_of_sleep'].value_counts()
```

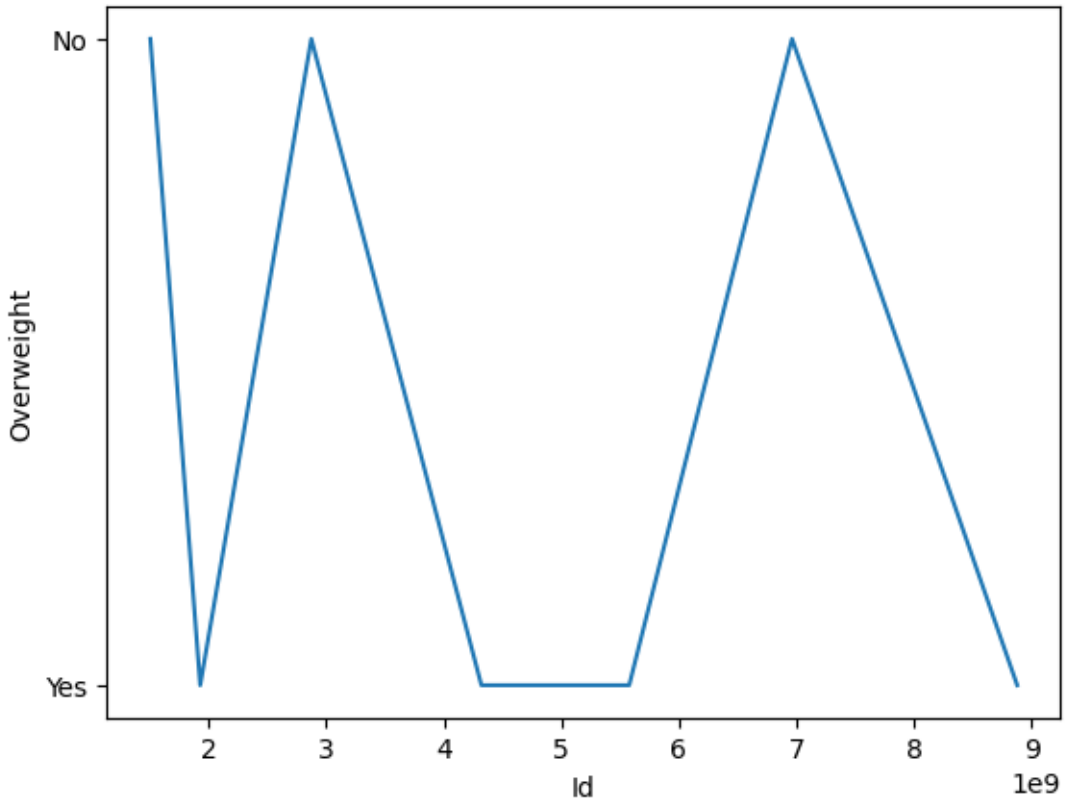
[40]:
```
lack_of_sleep
Yes    22
No      2
Name: count, dtype: int64
```

[41]:
```
df_sleep_weightmerge = df1.merge(df_sleeptime,on='Id',how="left")
df_sleep_weightmerge
```

[41]:
```
           Id  WeightPounds        BMI Overweight  TotalMinutesAsleep  \
0  1503960366    115.963147  22.650000         No          360.280000
1  2873212765    125.663489  21.570001         No                 NaN
2  6962181067    135.701872  24.028000         No          448.000000
3  4558609924    153.529918  27.214000        Yes          127.600000
4  4319703577    159.504452  27.415000        Yes          476.653846
5  8877689391    187.714432  25.487083        Yes                 NaN
6  5577150313    199.959265  28.000000        Yes          432.000000
7  1927972279    294.317120  47.540001        Yes          417.000000

   TotalTimeInBed lack_of_sleep
0      383.200000           Yes
1             NaN           NaN
2      466.129032           Yes
3      140.000000           Yes
4      501.961538           Yes
5             NaN           NaN
6      460.615385           Yes
7      437.800000           Yes
```

[42]:
```
sns.lineplot(data=df_sleep_weightmerge, x="Id", y="Overweight")
```

[42]:
```
<Axes: xlabel='Id', ylabel='Overweight'>
```



[43]:
```
# Join the grouped id_sleep_weightmerge dataframe and activitycalorymerged␣
 ↪dataframe
df_sleepweight_calories = df_sleep_weightmerge.
 ↪merge(activitycalorymerged,on='Id',how="left")
df_sleepweight_calories
```

[43]:

	Id	WeightPounds	BMI	Overweight	TotalMinutesAsleep	\
0	1503960366	115.963147	22.650000	No	360.280000	
1	2873212765	125.663489	21.570001	No	NaN	
2	6962181067	135.701872	24.028000	No	448.000000	
3	4558609924	153.529918	27.214000	Yes	127.600000	
4	4319703577	159.504452	27.415000	Yes	476.653846	
5	8877689391	187.714432	25.487083	Yes	NaN	
6	5577150313	199.959265	28.000000	Yes	432.000000	
7	1927972279	294.317120	47.540001	Yes	417.000000	

	TotalTimeInBed	lack_of_sleep	TotalSteps	TotalDistance	TrackerDistance	\
0	383.200000	Yes	12116.741935	7.809677	7.809677	
1	NaN	NaN	7555.774194	5.101613	5.101613	
2	466.129032	Yes	9794.806452	6.585806	6.519355	
3	140.000000	Yes	7685.129032	5.080645	5.080645	

```
4      501.961538           Yes   7268.838710       4.892258         4.892258
5             NaN           NaN           NaN            NaN              NaN
6      460.615385           Yes   8304.433333       6.213333         6.213333
7      437.800000           Yes    916.129032       0.634516         0.634516

   LoggedActivitiesDistance  VeryActiveDistance  ModeratelyActiveDistance  \
0                    0.0000            2.858387                  0.794194
1                    0.0000            0.676129                  0.276129
2                    0.3237            1.616452                  0.960000
3                    0.0000            0.549355                  0.682258
4                    0.0000            0.278065                  0.502258
5                       NaN                 NaN                       NaN
6                    0.0000            3.113667                  0.658000
7                    0.0000            0.095806                  0.031290

   LightActiveDistance  SedentaryActiveDistance  VeryActiveMinutes  \
0             4.152903                 0.000000          38.709677
1             4.143548                 0.005161          14.096774
2             4.001613                 0.006774          22.806452
3             3.847742                 0.000000          10.387097
4             3.768710                 0.000000           3.580645
5                  NaN                      NaN                NaN
6             2.428000                 0.000000          87.333333
7             0.507097                 0.000000           1.322581

   FairlyActiveMinutes  LightlyActiveMinutes  SedentaryMinutes     Calories
0            19.161290            219.935484        848.161290  1816.419355
1             6.129032            308.000000       1097.193548  1916.967742
2            18.516129            245.806452        662.322581  1982.032258
3            13.709677            284.967742       1093.612903  2033.258065
4            12.322581            228.774194        735.806452  2037.677419
5                  NaN                   NaN               NaN          NaN
6            29.833333            147.933333        754.433333  3359.633333
7             0.774194             38.580645       1317.419355  2172.806452
```

[44]:
```
df_sleepweight_calories.describe()
```

[44]:
```
                 Id  WeightPounds        BMI  TotalMinutesAsleep  \
count  8.000000e+00      8.000000   8.000000            6.000000
mean   4.575060e+09    171.544212  27.988011          376.922308
std    2.524611e+09     57.433835   8.236066          128.122419
min    1.503960e+09    115.963147  21.570001          127.600000
25%    2.636903e+09    133.192276  23.683500          374.460000
50%    4.439157e+09    156.517185  26.350542          424.500000
75%    5.923408e+09    190.775641  27.561250          444.000000
max    8.877689e+09    294.317120  47.540001          476.653846
```

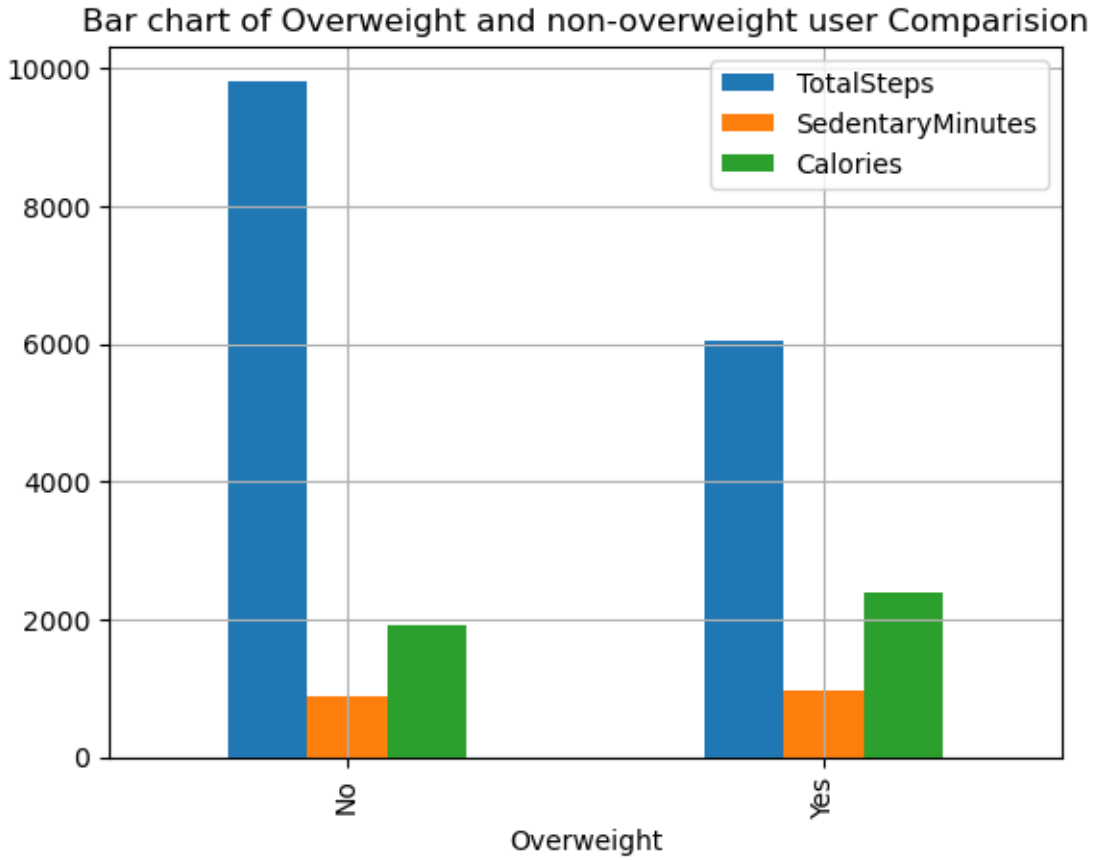
```
       TotalTimeInBed    TotalSteps  TotalDistance  TrackerDistance  \
count        6.000000      7.000000       7.000000         7.000000
mean       398.284326   7663.121813       5.188264         5.178771
std        132.466485   3427.328178       2.262706         2.255995
min        140.000000    916.129032       0.634516         0.634516
25%        396.850000   7412.306452       4.986452         4.986452
50%        449.207692   7685.129032       5.101613         5.101613
75%        464.750620   9049.619892       6.399570         6.366344
max        501.961538  12116.741935       7.809677         7.809677

       LoggedActivitiesDistance  VeryActiveDistance  ModeratelyActiveDistance  \
count                  7.000000            7.000000                  7.000000
mean                   0.046243            1.312551                  0.557733
std                    0.122347            1.242586                  0.316968
min                    0.000000            0.095806                  0.031290
25%                    0.000000            0.413710                  0.389194
50%                    0.000000            0.676129                  0.658000
75%                    0.000000            2.237419                  0.738226
max                    0.323700            3.113667                  0.960000

       LightActiveDistance  SedentaryActiveDistance  VeryActiveMinutes  \
count             7.000000                 7.000000           7.000000
mean              3.264230                 0.001705          25.462366
std               1.354174                 0.002949          30.079176
min               0.507097                 0.000000           1.322581
25%               3.098355                 0.000000           6.983871
50%               3.847742                 0.000000          14.096774
75%               4.072581                 0.002581          30.758065
max               4.152903                 0.006774          87.333333

       FairlyActiveMinutes  LightlyActiveMinutes  SedentaryMinutes  \
count             7.000000              7.000000          7.000000
mean             14.349462            210.571121        929.849923
std               9.457325             91.462225        242.108800
min               0.774194             38.580645        662.322581
25%               9.225806            183.934409        745.119892
50%              13.709677            228.774194        848.161290
75%              18.838710            265.387097       1095.403226
max              29.833333            308.000000       1317.419355

          Calories
count     7.000000
mean   2188.399232
std     528.117681
min    1816.419355
25%    1949.500000
50%    2033.258065
```

```
75%     2105.241935
max     3359.633333
```

[45]:
```
# let's group the joined dataframe data into by overweight yes or no to show␣
 ↪TotalSteps, SedentaryMinutes and Calories.
df_overweight_group = df_sleepweight_calories.
 ↪groupby('Overweight')[['TotalSteps','SedentaryMinutes','Calories']].mean()
```

[46]:
```
plt.figure(figsize=(10,10))
df_overweight_group.plot(kind='bar',grid=True)
plt.title('Bar chart of Overweight and non-overweight user Comparision')
plt.show()
```

```
<Figure size 1000x1000 with 0 Axes>
```



[47]:
```
#Bar chart shows us that the Overweight user's take less steps which means less␣
 ↪movement,
#more sedentary time, and more weight gain. Although the calories burned for␣
 ↪overweight is more than non-overweight users,
```

```
#a overweight person needs to burn more calories than other people depending on␣
 ↪their present body weight.
```