

FITBIT DATA ANALYSIS

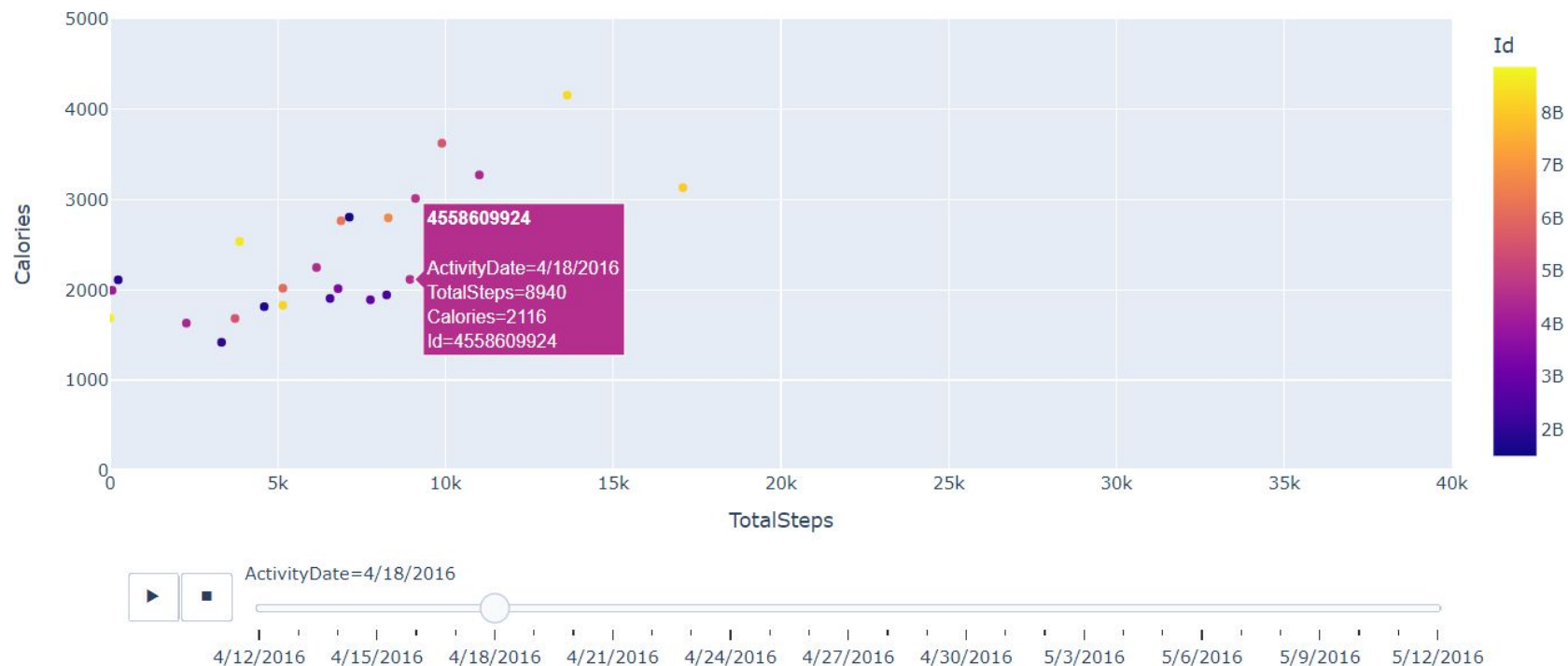
- By SRIJA BASAK

Part 1: Daily Activity Analysis

Here, the features contributing to burning calories is analyzed.

Plotting an animation plot to show the activity of all the 31 fitbit users on each of the given Activity Dates provided in the given dataset. Hovering over the data points in the plot will display ID of the fitbit user, Total Steps, Calories burnt and Activity Date

Animation Plot



Finding 5 most significant features contributing to burning Calories using the Chi-square test

```
✓ [16] #Selecting important activities contributing to burning calories - using Chi square test
0s from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import chi2
best_features=SelectKBest(chi2,k=5)
dfDaily_best_features=best_features.fit_transform(X,Y)

print(dfDaily_best_features)
```

```
[[1.3162e+04 2.5000e+01 1.3000e+01 3.2800e+02 7.2800e+02]
 [1.0735e+04 2.1000e+01 1.9000e+01 2.1700e+02 7.7600e+02]
 [1.0460e+04 3.0000e+01 1.1000e+01 1.8100e+02 1.2180e+03]
 ...
 [1.0733e+04 1.8000e+01 1.1000e+01 2.2400e+02 1.1870e+03]
 [2.1420e+04 8.8000e+01 1.2000e+01 2.1300e+02 1.1270e+03]
 [8.0640e+03 2.3000e+01 1.0000e+00 1.3700e+02 7.7000e+02]]
```

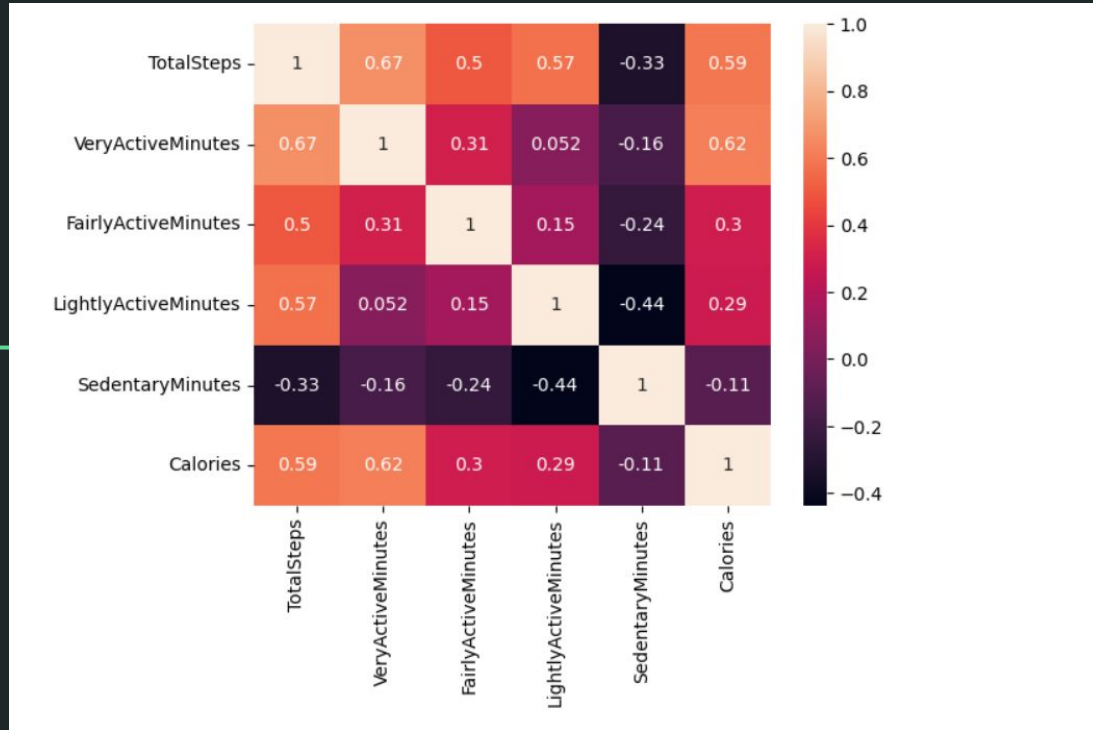
```
✓ [17] X.head()
```

ActiveDistance	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes
0.55	6.06	0.0	25	13	328	728
0.69	4.71	0.0	21	19	217	776
0.40	3.91	0.0	30	11	181	1218
1.26	2.83	0.0	29	34	209	726
0.41	5.04	0.0	36	10	221	773

Comparing the best features values given by chi-square test to the values in the feature set to know the selected features

From the previous 2 results it can be seen that TotalSteps, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes are the 5 best selected features for 'Calories' column according to the Chi-square test

Correlation heatmap of these 5 features along with 'Calories'



2 most significant features for 'Calories'

```
[35] #Selecting 2 most important features contributing to burning calories - using Chi square test
```

```
best_2_features=SelectKBest(chi2,k=2)  
dfDaily_best_2_features=best_2_features.fit_transform(X,Y)
```

```
print(dfDaily_best_2_features)
```

```
[[13162.  728.]  
 [10735.  776.]  
 [10460. 1218.]  
 ...  
 [10733. 1187.]  
 [21420. 1127.]  
 [ 8064.  770.]]
```

Best 2 features - TotalSteps, SedentaryMinutes

Correlation Heatmap for these 2 features along with 'Calories'

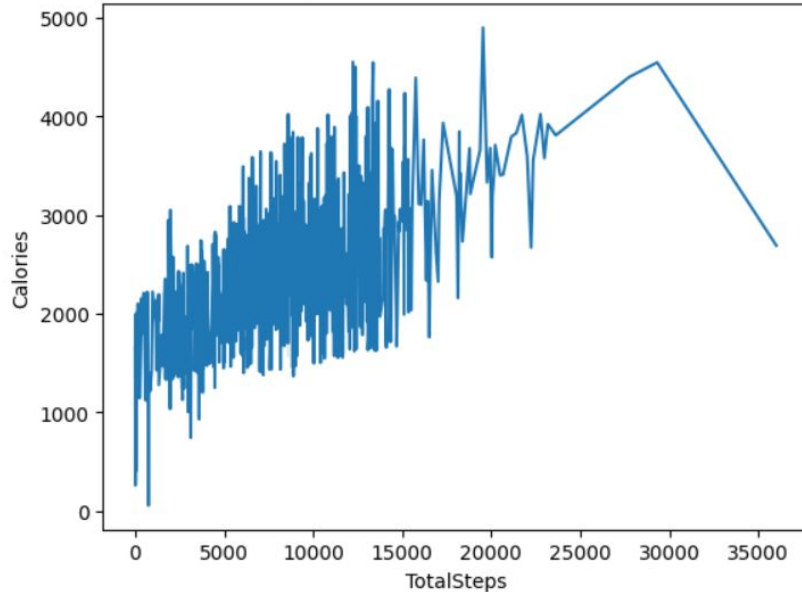


As per the above heatmap it can be concluded that 'SedentaryMinutes' has is negatively correlated to 'Calories' which means that is Sedentary Minutes increases, Calorie burning will decrease. Also it can be concluded that 'TotalSteps' is positively correlated to 'Calories' which means that as Total Steps increases, Calorie burning will also increase

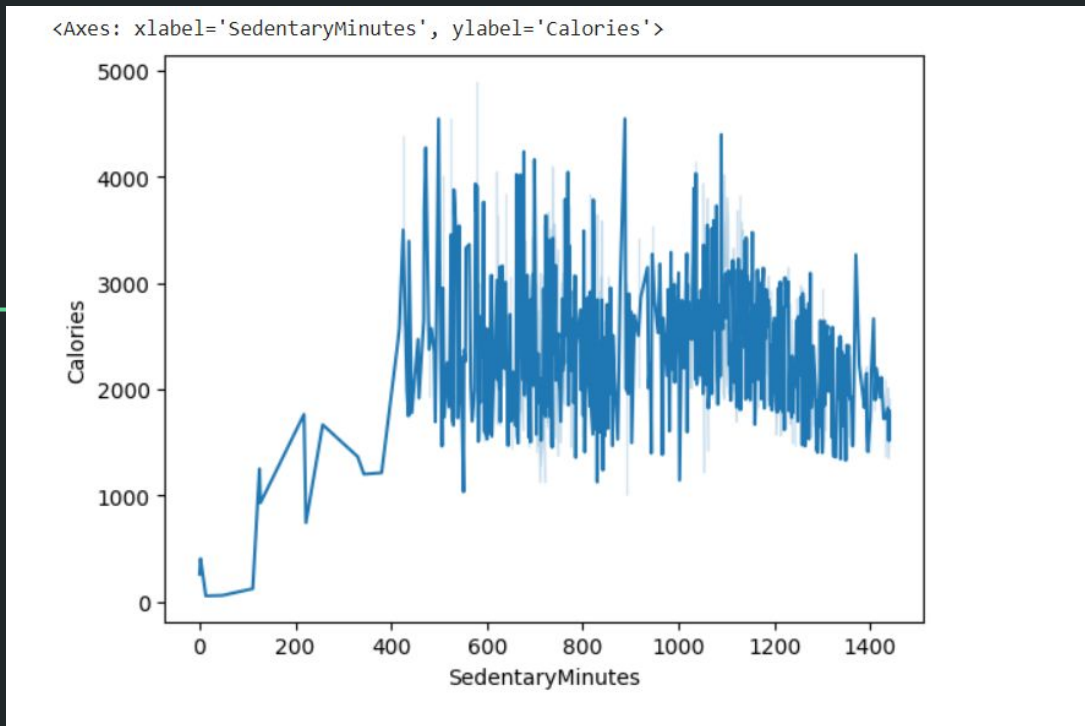
Total Steps vs Calories Lineplot

```
✓ [49] sns.lineplot(x=dfDaily['TotalSteps'],y=dfDaily['Calories'])  
2s
```

<Axes: xlabel='TotalSteps', ylabel='Calories'>



SedentaryMinutes vs Calories Lineplot



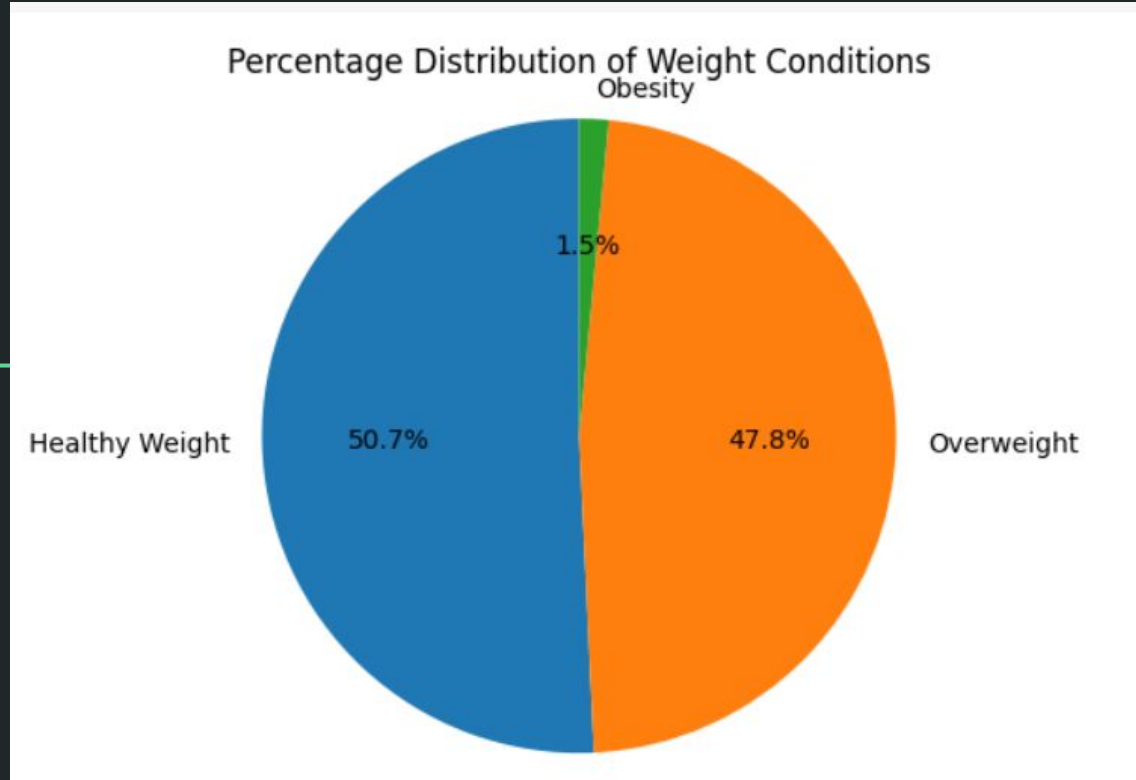
Part 2: Analysis of Sleep Patterns and Weight

ANALYSIS OF WEIGHT:

A new column 'Weight Condition' has been added to the existing data 'weightLogInfo_merged.csv'. In this column 'Weight Condition', categories include 'Underweight', 'Healthy Weight', 'Overweight', 'Obesity' based on BMI

```
if bmi < 18.5:  
    return 'Underweight'  
elif 18.5 <= bmi < 24.9:  
    return 'Healthy Weight'  
elif 25 <= bmi < 29.9:  
    return 'Overweight'  
else:  
    return 'Obesity'
```

Pie chart of 'Weight Condition' column



ANALYSIS OF DAY WISE SLEEP PATTERNS:

An animation plot has been plotted to display 'TotalTimeInBed' vs 'TotalMinutesAsleep'. Dataset used : sleepDay_merged.csv

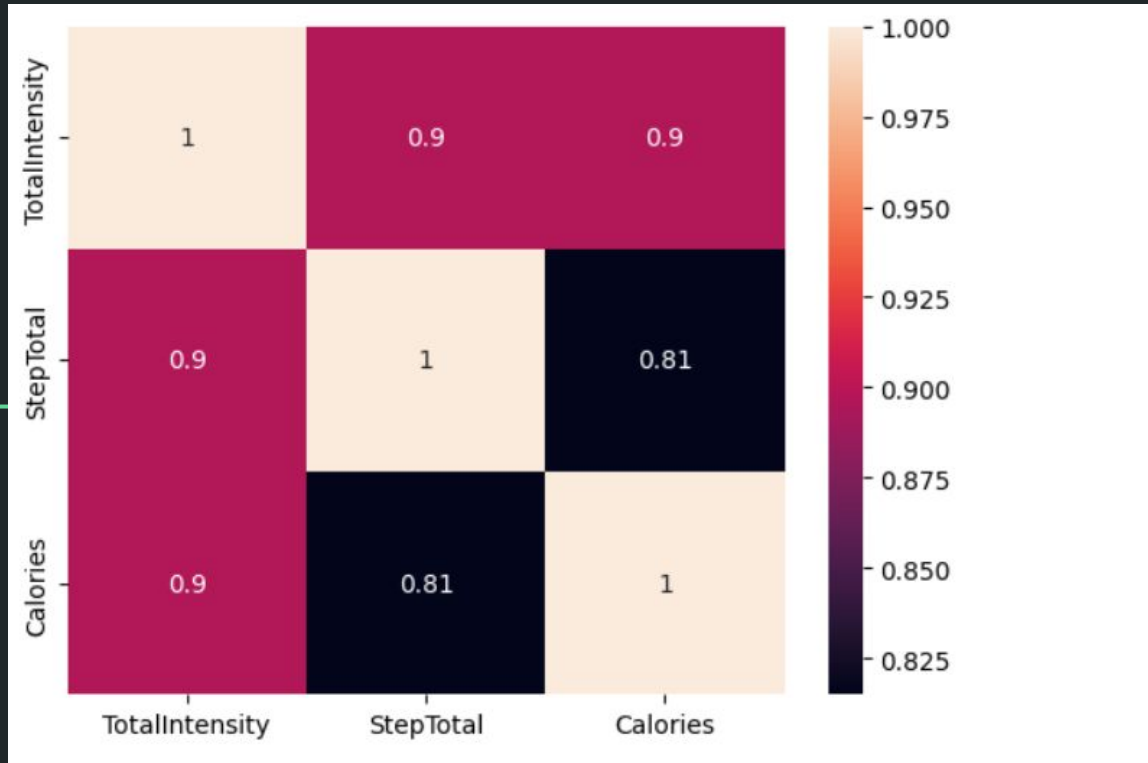


DATA ANALYSIS OF HOURLY ACTIVITY

The datasets – hourlyCalories_merged.csv, hourlyIntensities_merged.csv and hourlySteps_merged.csv have been merged on columns 'Id' and 'ActivityHour'

The 2 most significant features for calculating calorie burning include 'StepTotal' and 'TotalIntensity'. This has been found out using Chi-square test.

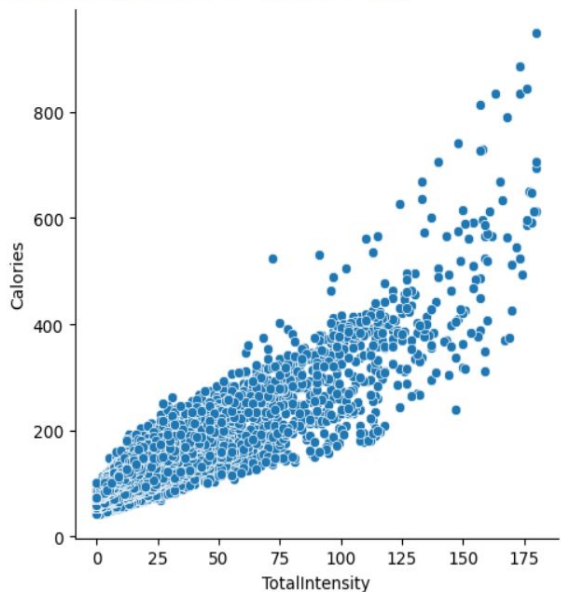
Correlation heatmap of Significant features



Plots of Total Intensity vs Calories without handling outliers

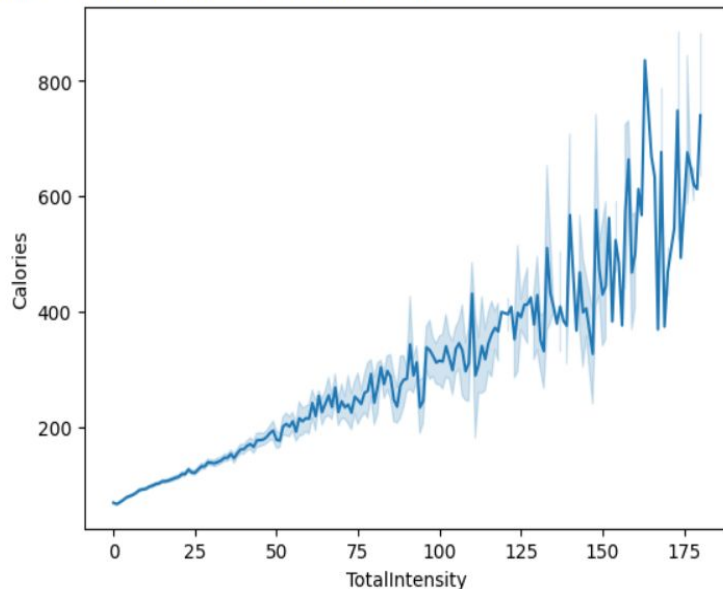
```
[118] sns.relplot(x="TotalIntensity",y="Calories",data=df)
```

<seaborn.axisgrid.FacetGrid at 0x79ab53e3a7a0>



```
] sns.lineplot(x="TotalIntensity",y="Calories",data=df)
```

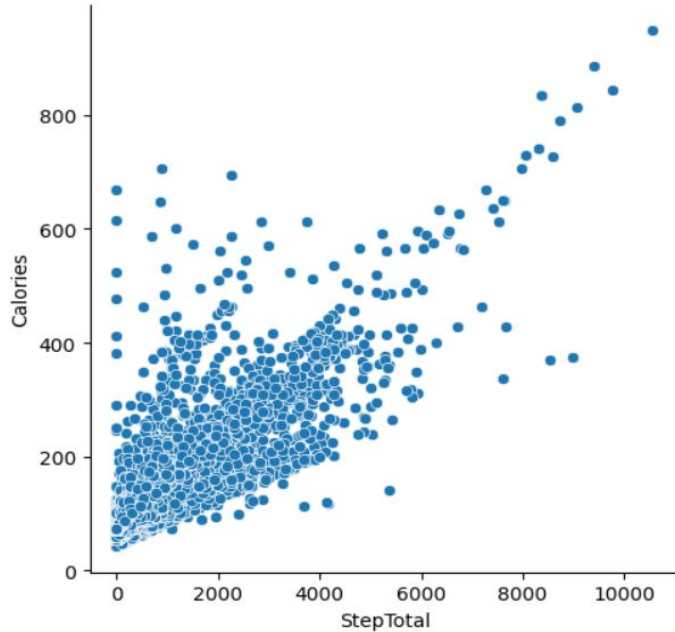
<Axes: xlabel='TotalIntensity', ylabel='Calories'>



Plots of Total Steps vs Calories

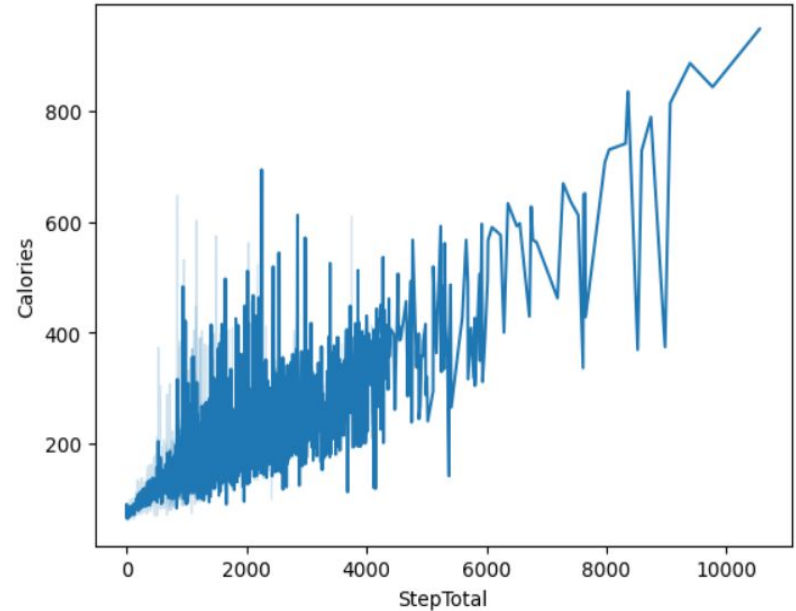
```
[120] sns.relplot(x="StepTotal",y="Calories",data=df)
```

<seaborn.axisgrid.FacetGrid at 0x79ab53cf2e60>



```
[121] sns.lineplot(x="StepTotal",y="Calories",data=df)
```

<Axes: xlabel='StepTotal', ylabel='Calories'>



From the plots in the previous slide, it can be concluded that as Total Steps increases, Calorie burning also increases. And also, Calorie burning also increases with Total Intensity.

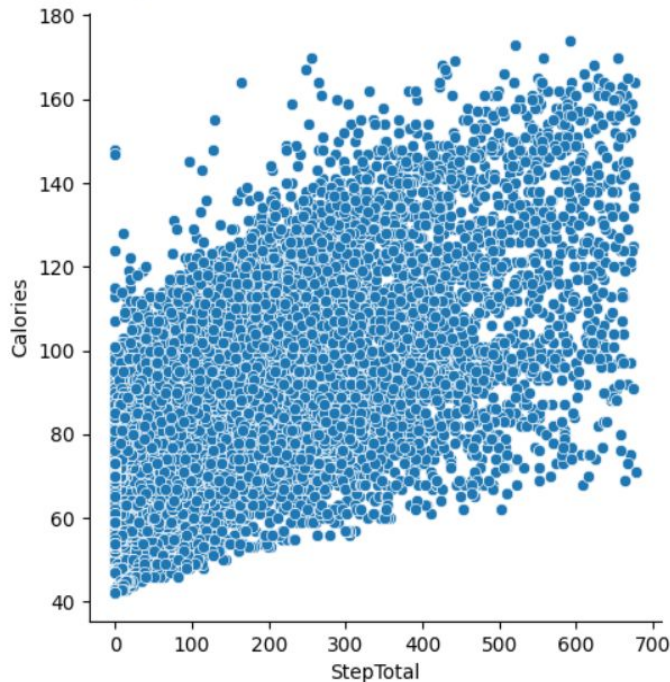
After this step, the outliers have been detected and handled using IQR(Interquartile range method). The outliers are handled so that while plotting the graphs to analyze trends in Calorie burning with respect to Total Steps and Total Intensity, there is less randomness in the data points.

The plots obtained after handling the outliers are shown on the next slide.

Plots of Total Steps vs Calories after handling outliers

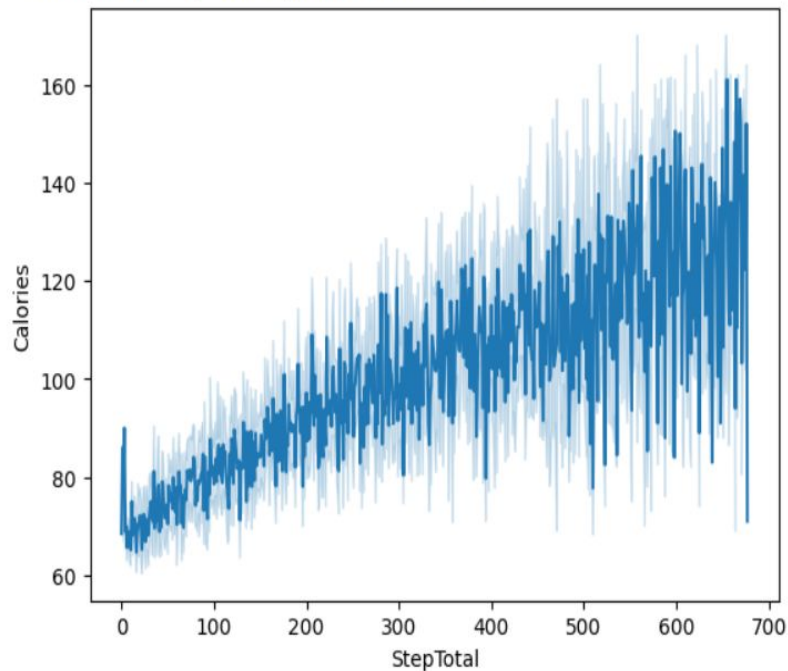
```
[144] sns.relplot(x="StepTotal",y="Calories",data=df)
```

<seaborn.axisgrid.FacetGrid at 0x79ab53b46dd0>



```
[145] sns.lineplot(x="StepTotal",y="Calories",data=df)
```

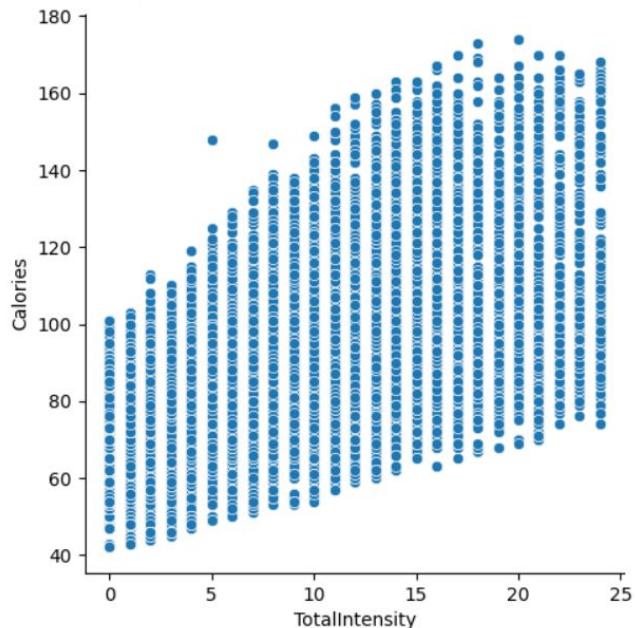
<Axes: xlabel='StepTotal', ylabel='Calories'>



Plots of Total Intensity vs Calories after handling outliers

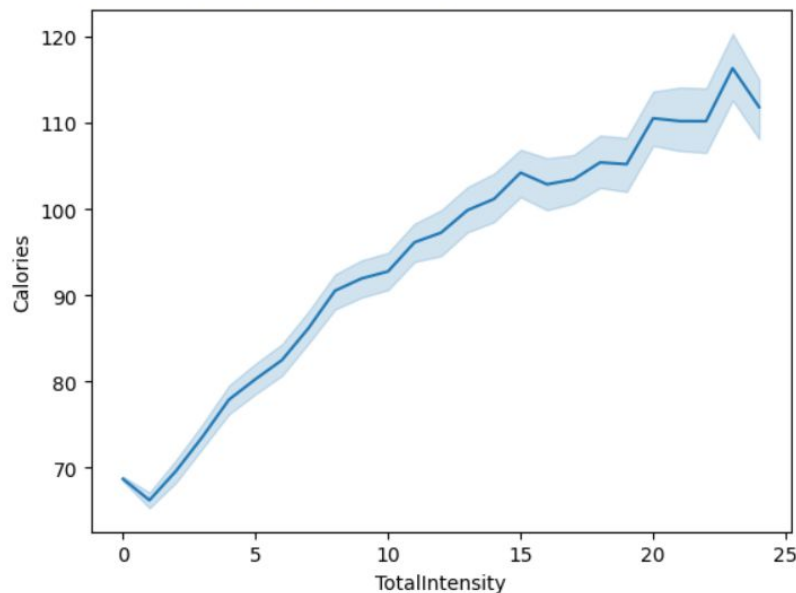
```
sns.relplot(x="TotalIntensity",y="Calories",data=df)
```

```
<seaborn.axisgrid.FacetGrid at 0x79ab539eb4c0>
```



```
[147] sns.lineplot(x="TotalIntensity",y="Calories",data=df)
```

```
<Axes: xlabel='TotalIntensity', ylabel='Calories'>
```



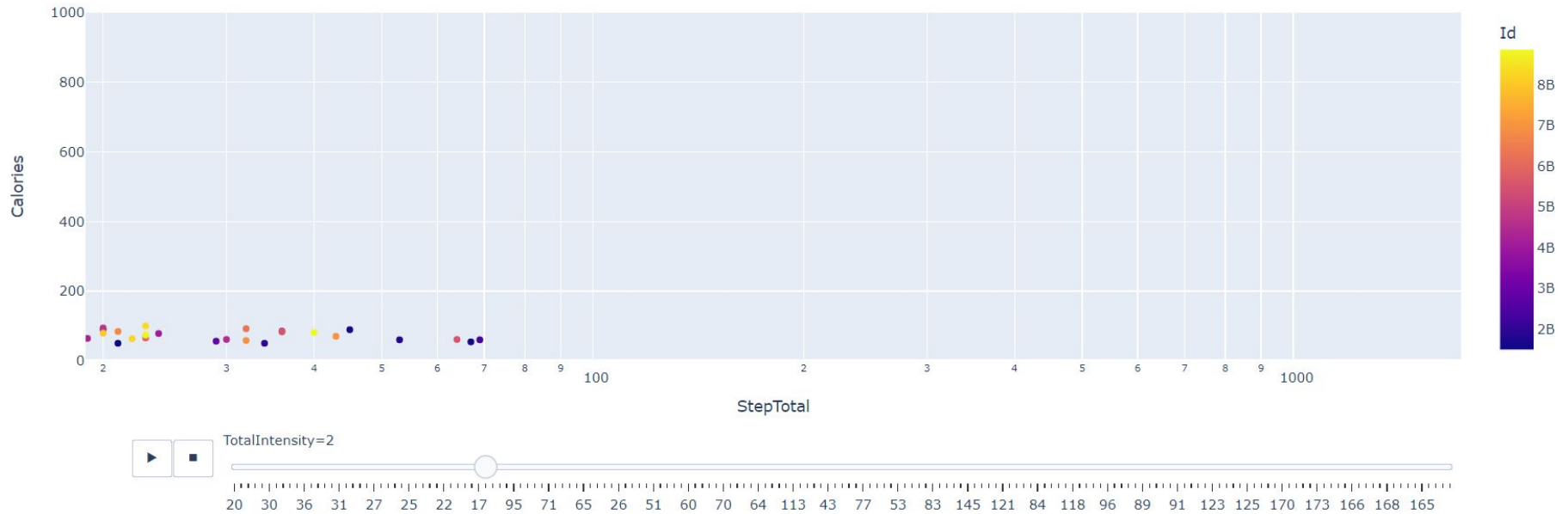
The above two graphs do not show a steep increase in Calorie burning wrt Total Steps and also wrt Total Intensity but overall there is an increase in calorie burning wrt Total Steps and also wrt Total Intensity

Outliers have been removed in order to reduce the randomness of data points which hence helps to observe, analyze and visualize the trends (For eg. 'StepTotal' vs Calorie burning, etc.) in a more generalized sense

As compared to the previous result where outliers were not handled in the given dataset, the graphs plotted after handling the outliers have shown less randomness in the data points

Hence, based on the two above plots, as Total Steps Taken increases, Calorie burning also increases

Animation Plot showing Total Steps vs Calories with Animation frame as Total Intensity



Similar Data analysis techniques have been followed for Minutes Activity as well.
