

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.pyplot as plt
from sklearn.metrics import train_test_split
from sklearn import XGBRegressor
import seaborn as sns

In [2]: calories_data=pd.read_csv("calories.csv")

In [3]: calories_data.head()

Out[3]:
   User_ID  Calories
0  14733363    231.0
1  14861698     66.0
2  11179863     26.0
3  16180408     73.0
4  1771927    35.0

In [4]: exercise_data=pd.read_csv("exercise.csv")

In [5]: exr_data.head()

Out[5]:
   User_ID Gender  Age  Height  Weight  Duration  Heart_Rate  Body_Temp
0  14733363    male   68   190.0    94.0     29.0     105.0     40.8
1  14861698  female   20   166.0    60.0     14.0     94.0     40.3
2  11179863    male   69   179.0    79.0      5.0     88.0     38.7
3  16180408  female   34   179.0    71.0     13.0     100.0     40.5
4  1771927    female   27   154.0    58.0     10.0     81.0     39.8

In [6]: exercise_data.shape

Out[6]:
(15808, 8)

In [7]: calories_data.shape

Out[7]:
(15808, 2)

In [8]: data = pd.concat([exercise_data,calories_data['Calories']],axis=1)

In [9]: data.head()

Out[9]:
   User_ID Gender  Age  Height  Weight  Duration  Heart_Rate  Body_Temp  Calories
0  14733363    male   68   190.0    94.0     29.0     105.0     40.8    231.0
1  14861698  female   20   166.0    60.0     14.0     94.0     40.3     66.0
2  11179863    male   69   179.0    79.0      5.0     88.0     38.7     26.0
3  16180408  female   34   179.0    71.0     13.0     100.0     40.5     71.0
4  1771927    female   27   154.0    58.0     10.0     81.0     39.8     35.0

In [10]: data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15808 entries, 0 to 14999
Data columns (total 9 columns):
#  Column                Non-Null Count  Dtype
--  --
0  User_ID                15808 non-null    int64
1  Gender                15808 non-null    object
2  Age                  15808 non-null    int64
3  Height               15808 non-null    float64
4  Weight               15808 non-null    float64
5  Duration             15808 non-null    float64
6  Heart_Rate           15808 non-null    float64
7  Body_Temp            15808 non-null    float64
8  Calories              15808 non-null    float64
dtypes: float64(6), int64(2), object(1)
memory usage: 1.8+ MB

In [11]: data.describe()

Out[11]:
   User_ID      Age      Height      Weight      Duration      Heart_Rate      Body_Temp      Calories
count  1500000e+04  15000.00000  15000.00000  15000.00000  15000.00000  15000.00000  15000.00000  15000.00000
mean    1.497728e+07    42.788004    174.462153    74.960867    85.538600    96.518633    40.025453    69.538633
std     2.022514e+05    16.880254    14.250114    15.030567    8.310030    9.503320    0.779200    62.452978
min     1.005116e+07    20.000000    123.000000    36.000000    1.000000    67.000000    37.100000    1.000000
25%    1.247419e+07    28.000000    164.000000    63.000000    8.000000    88.000000    39.600000    35.000000
50%    1.497728e+07    39.000000    175.000000    74.000000    16.000000    96.000000    40.200000    79.000000
75%    1.744925e+07    56.000000    185.000000    87.000000    23.000000    103.000000    40.600000    138.000000
max     1.999995e+07    79.000000    222.000000    132.000000    30.000000    128.000000    41.500000    314.000000

In [14]: plt.figure(figsize=(5,5))
sns.countplot(x=data['Gender'])
plt.show()

<Figure size 500x500 with 0 Axes>

<img alt="A bar chart showing the count of users for each gender. The x-axis is labeled 'Gender' with categories 'male' and 'female'. The y-axis is labeled 'count' and ranges from 0 to 7000. The 'male' bar is blue and has a count of approximately 7000. The 'female' bar is orange and has a count of approximately 7000." data-bbox="60 170 360 200"/>
```