```
Requirement already satisfied: pandas-profiling in d:\annoconda\lib\site-packages (2.9.0)
Requirement already \ satisfied: \ phik>=0.9.10 \ in \ d:\ lib\ site-packages \ (from \ pandas-profiling) \ (0.12.2)
Requirement already satisfied: numpy>=1.16.0 in d:\annoconda\lib\site-packages (from pandas-profiling) (1.20.3)
Requirement already satisfied: attrs>=19.3.0 in d:\annoconda\lib\site-packages (from pandas-profiling) (21.2.0)
Requirement already satisfied: tqdm>=4.43.0 in d:\annoconda\lib\site-packages (from pandas-profiling) (4.62.3)
Requirement already satisfied: jinja2>=2.11.1 in d:\annoconda\lib\site-packages (from pandas-profiling) (2.11.3)
Requirement already satisfied: matplotlib>=3.2.0 in d:\annoconda\lib\site-packages (from pandas-profiling) (3.4.3
Requirement already satisfied: pandas! = 1.0.0, != 1.0.1, != 1.0.2, != 1.1.0, >= 0.25.3 in d: \\ annoconda \\ lib \\ site-packages (lib \\ site-packages (l
from pandas-profiling) (1.3.4)
Requirement already satisfied: joblib in d:\annoconda\lib\site-packages (from pandas-profiling) (1.1.0)
Requirement already satisfied: visions[type image path] == 0.5.0 in d:\annoconda\lib\site-packages (from pandas-pro
filing) (0.5.0)
Requirement already satisfied: htmlmin>=0.1.12 in d:\annoconda\lib\site-packages (from pandas-profiling) (0.1.12)
Requirement already satisfied: tangled-up-in-unicode>=0.0.6 in d:\annoconda\lib\site-packages (from pandas-profil
ina) (0.2.0)
Requirement already satisfied: ipywidgets>=7.5.1 in d:\annoconda\lib\site-packages (from pandas-profiling) (7.6.5
Requirement already satisfied: confuse>=1.0.0 in d:\annoconda\lib\site-packages (from pandas-profiling) (1.7.0)
Requirement already satisfied: missingno>=0.4.2 in d:\annoconda\lib\site-packages (from pandas-profiling) (0.5.1)
Requirement already satisfied: scipy >= 1.4.1 in d:\annoconda\lib\site-packages (from pandas-profiling) (1.7.1)
Requirement already satisfied: seaborn>=0.10.1 in d:\annoconda\lib\site-packages (from pandas-profiling) (0.11.2)
Requirement already satisfied: requests >= 2.23.0 in c: \users \hp\appdata \roaming\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\python\
m pandas-profiling) (2.25.1)
Requirement already satisfied: networkx>=2.4 in d:\annoconda\lib\site-packages (from visions[type image path]==0.
5.0->pandas-profiling) (2.6.3)
Requirement already satisfied: Pillow in d:\annoconda\lib\site-packages (from visions[type_image_path]==0.5.0->pa
ndas-profiling) (8.4.0)
Requirement already satisfied: image hash in d:\annoconda\lib\site-packages (from visions[type\_image\_path] == 0.5.0-1.00 and the conditions of the condition 
>pandas-profiling) (4.2.1)
Requirement already satisfied: pyyaml in c:\users\hp\appdata\roaming\python\python39\site-packages (from confuse>
=1.0.0->pandas-profiling) (5.4.1)
Requirement already satisfied: widgetsnbextension~=3.5.0 in d:\annoconda\lib\site-packages (from ipywidgets>=7.5.
1->pandas-profiling) (3.5.1)
Requirement already satisfied: traitlets>=4.3.1 in d:\annoconda\lib\site-packages (from ipywidgets>=7.5.1->pandas
-profiling) (5.1.0)
Requirement already satisfied: ipykernel >= 4.5.1 in d: \\ annoconda \\ lib \\ site-packages (from ipywidgets >= 7.5.1- \\ >pandas \\ lib \\ > 1.5.1 \\ > 2.5.1 \\ > 2.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\ > 3.5.1 \\
 -profiling) (6.4.1)
Requirement already satisfied: ipython-genutils~=0.2.0 in d:\annoconda\lib\site-packages (from ipywidgets>=7.5.1-
>pandas-profiling) (0.2.0)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in d:\annoconda\lib\site-packages (from ipywidgets>=7.5.
1->pandas-profiling) (1.0.0)
Requirement already satisfied: ipython>=4.0.0 in d:\annoconda\lib\site-packages (from ipywidgets>=7.5.1->pandas-p
rofiling) (7.29.0)
Requirement already satisfied: nbformat>=4.2.0 in d:\annoconda\lib\site-packages (from ipywidgets>=7.5.1->pandas-
profiling) (5.1.3)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in d:\annoconda\lib\site-packages (from ipykernel>=4.5.1->ipyw
idgets>=7.5.1->pandas-profiling) (1.4.1)
Requirement already satisfied: tornado<7.0,>=4.2 in d:\annoconda\lib\site-packages (from ipykernel>=4.5.1->ipywid
gets>=7.5.1->pandas-profiling) (6.1)
Requirement already satisfied: matplotlib-inline<0.2.0,>=0.1.0 in d:\annoconda\lib\site-packages (from ipykernel>
=4.5.1->ipywidgets>=7.5.1->pandas-profiling) (0.1.2)
Requirement already satisfied: jupyter-client<8.0 in d:\annoconda\lib\site-packages (from ipykernel>=4.5.1->ipywi
dgets>=7.5.1->pandas-profiling) (6.1.12)
Requirement already satisfied: setuptools>=18.5 in d:\annoconda\lib\site-packages (from ipython>=4.0.0->ipywidget
s>=7.5.1->pandas-profiling) (58.0.4)
Requirement already satisfied: pickleshare in d:\annoconda\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.
5.1->pandas-profiling) (0.7.5)
Requirement already satisfied: backcall in d:\annoconda\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.1
 ->pandas-profiling) (0.2.0)
Requirement already satisfied: colorama in d:\annoconda\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.1
->pandas-profiling) (0.4.4)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in d:\annoconda\lib\site-packages (fr
om ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling) (3.0.20)
Requirement already satisfied: jedi>=0.16 in d:\annoconda\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5
 .1->pandas-profiling) (0.18.0)
Requirement already satisfied: decorator in d:\annoconda\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.
1->pandas-profiling) (5.1.0)
Requirement already satisfied: pygments in d:\annoconda\lib\site-packages (from ipython>=4.0.0->ipywidgets>=7.5.1
 ->pandas-profiling) (2.10.0)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in d:\annoconda\lib\site-packages (from jedi>=0.16->ipvthon>=4
 .0.0->ipywidgets>=7.5.1->pandas-profiling) (0.8.2)
Requirement already satisfied: MarkupSafe>=0.23 in d:\annoconda\lib\site-packages (from jinja2>=2.11.1->pandas-pr
ofiling) (2.0.1)
Requirement already satisfied: python-dateutil>=2.1 in d:\annoconda\lib\site-packages (from jupyter-client<8.0->i
pykernel>=4.5.1->ipywidgets>=7.5.1->pandas-profiling) (2.8.2)
Requirement already satisfied: pyzmq>=13 in d:\annoconda\lib\site-packages (from jupyter-client<8.0->ipykernel>=4
 .5.1->ipywidgets>=7.5.1->pandas-profiling) (22.2.1)
Requirement already \ satisfied: jupyter-core>=4.6.0 \ in \ d:\ annoconda\\ lib\ site-packages \ (from \ jupyter-client<8.0->ip \ del{lib}
ykernel>=4.5.1->ipywidgets>=7.5.1->pandas-profiling) (4.8.1)
Requirement already satisfied: pywin32>=1.0 in d:\annoconda\lib\site-packages (from jupyter-core>=4.6.0->jupyter-
client<8.0->ipykernel>=4.5.1->ipywidgets>=7.5.1->pandas-profiling) (228)
Requirement already satisfied: cycler>=0.10 in d:\annoconda\lib\site-packages (from matplotlib>=3.2.0->pandas-pro
filing) (0.10.0)
```

Requirement already satisfied: pyparsing>=2.2.1 in d:\annoconda\lib\site-packages (from matplotlib>=3.2.0->pandas

```
-profiling) (3.0.4)
           Requirement already satisfied: kiwisolver>=1.0.1 in d:\annoconda\lib\site-packages (from matplotlib>=3.2.0->panda
           s-profiling) (1.3.1)
           Requirement already satisfied: six in d:\annoconda\lib\site-packages (from cycler>=0.10->matplotlib>=3.2.0->panda
           s-profiling) (1.16.0)
           Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in d:\annoconda\lib\site-packages (from nbformat>=4.2.0->i
           pywidgets>=7.5.1->pandas-profiling) (3.2.0)
           Requirement already satisfied: pyrsistent>=0.14.0 in d:\annoconda\lib\site-packages (from jsonschema!=2.5.0,>=2.4
           ->nbformat>=4.2.0->ipywidgets>=7.5.1->pandas-profiling) (0.18.0)
           Requirement already satisfied: pytz>=2017.3 in d:\annoconda\lib\site-packages (from pandas!=1.0.0,!=1.0.1,!=1.0.2
           ,!=1.1.0,>=0.25.3->pandas-profiling) (2021.3)
           Requirement already satisfied: wcwidth in d:\annoconda\lib\site-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1
           .0,>=2.0.0->ipython>=4.0.0->ipywidgets>=7.5.1->pandas-profiling) \enskip (0.2.5)
           Requirement already satisfied: idna<3,>=2.5 in c:\users\hp\appdata\roaming\python\python39\site-packages (from re
           quests>=2.23.0->pandas-profiling) (2.10)
           Requirement already satisfied: certifi>=2017.4.17 in d:\annoconda\lib\site-packages (from requests>=2.23.0->panda
           s-profiling) (2021.10.8)
           Requirement already satisfied: chardet<5,>=3.0.2 in d:\annoconda\lib\site-packages (from requests>=2.23.0->pandas
           -profiling) (4.0.0)
           Requirement already satisfied: urllib3<1.27,>=1.21.1 in d:\annoconda\lib\site-packages (from requests>=2.23.0->pa
           ndas-profiling) (1.26.7)
           Requirement already satisfied: notebook>=4.4.1 in d:\annoconda\lib\site-packages (from widgetsnbextension~=3.5.0-
           >ipywidgets>=7.5.1->pandas-profiling) (6.4.5)
           Requirement already satisfied: nbconvert in d:\annoconda\lib\site-packages (from notebook>=4.4.1->widgetsnbextens
           ion~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (6.1.0)
           Requirement already satisfied: argon2-cffi in d:\annoconda\lib\site-packages (from notebook>=4.4.1->widgetsnbexte
           nsion~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (20.1.0)
           Requirement already satisfied: Send2Trash>=1.5.0 in d:\annoconda\lib\site-packages (from notebook>=4.4.1->widgets
           nbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (1.8.0)
           Requirement already satisfied: terminado>=0.8.3 in d:\annoconda\lib\site-packages (from notebook>=4.4.1->widgetsn
           bextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (0.9.4)
           Requirement already satisfied: prometheus-client in d:\annoconda\lib\site-packages (from notebook>=4.4.1->widgets
           nbextension \sim = 3.5.0 - > ipywidgets > = 7.5.1 - > pandas - profiling) \ (0.11.0)
           Requirement already satisfied: pywinpty>=0.5 in d:\annoconda\lib\site-packages (from terminado>=0.8.3->notebook>=
           4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (0.5.7)
           Requirement already satisfied: cffi>=1.0.0 in d:\annoconda\lib\site-packages (from argon2-cffi->notebook>=4.4.1->
           widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (1.14.6)
           Requirement already satisfied: pycparser in d:\annoconda\lib\site-packages (from cffi>=1.0.0->argon2-cffi->notebo
           ok>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (2.20)
           Requirement already satisfied: PyWavelets in d:\annoconda\lib\site-packages (from imagehash->visions[type_image_p
           ath]==0.5.0->pandas-profiling) (1.1.1)
           Requirement already satisfied: mistune<2,>=0.8.1 in d:\annoconda\lib\site-packages (from nbconvert->notebook>=4.4
           .1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (0.8.4)
           Requirement already satisfied: testpath in d:\annoconda\lib\site-packages (from nbconvert->notebook>=4.4.1->widge
           tsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (0.5.0)
           Requirement already satisfied: entrypoints>=0.2.2 in d:\annoconda\lib\site-packages (from nbconvert->notebook>=4.
           4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (0.3)
           Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in d:\annoconda\lib\site-packages (from nbconvert->notebook
           >=4.4.1- yidgetsnbextension=3.5.0- ipywidgets=7.5.1- pandas-profiling) (0.5.3)
           Requirement already satisfied: defusedxml in d:\annoconda\lib\site-packages (from nbconvert->notebook>=4.4.1->wid
           getsnbextension \sim = 3.5.0 - ipywidgets > = 7.5.1 - pandas - profiling) \quad (0.7.1)
           Requirement already satisfied: jupyterlab-pygments in d:\annoconda\lib\site-packages (from nbconvert->notebook>=4
           .4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (0.1.2)
           Requirement already satisfied: pandocfilters>=1.4.1 in d:\annoconda\lib\site-packages (from nbconvert->notebook>=
           4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (1.4.3)
           Requirement already satisfied: bleach in d:\annoconda\lib\site-packages (from nbconvert->notebook>=4.4.1->widgets
           nbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (4.0.0)
           Requirement already satisfied: nest-asyncio in d:\annoconda\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconda (from nbclient<0.6.0,>=0.5.0
           vert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (1.5.1)
           Requirement already satisfied: async-generator in d:\annoconda\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nb
           convert->notebook>=4.4.1->widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (1.10)
           Requirement already satisfied: webencodings in d:\annoconda\lib\site-packages (from bleach->nbconvert->notebook>=
           4.4.1- widgetsnbextension~=3.5.0->ipywidgets>=7.5.1->pandas-profiling) (0.5.1)
           Requirement already satisfied: packaging in d:\annoconda\lib\site-packages (from bleach->nbconvert->notebook>=4.4
           .1-\verb|widgetsnbextension|=3.5.0-\verb|sipywidgets|=7.5.1-\verb|spandas-profiling|| (21.0)
In [2]:
            import numpy as np
            import pandas as pd
            import seaborn as sns
In [3]:
            df=pd.read_csv("E:\FitBit data.csv")
In [4]:
            df.head()
                        Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
Out[4]:
           0 1503960366
                               3/25/2016
                                               11004
                                                                7.11
                                                                                   7.11
                                                                                                               0.0
                                                                                                                                    2.57
                                                                                                                                                                0.46
              1503960366
                               3/26/2016
                                               17609
                                                               11.55
                                                                                 11.55
                                                                                                               0.0
                                                                                                                                    6.92
                                                                                                                                                                0.73
```

2 1503960366

**3** 1503960366

3/27/2016

3/28/2016

12736

13231

8 53

8.93

8 53

8.93

0.0

0.0

4 66

3.19

0.16

0.79

# **EDA** by Pandas Profiling

In []:
In [7]:
from pandas\_profiling import ProfileReport
prof=ProfileReport(df)
prof.to\_file(output\_file='output.html')

In [8]: from pandas\_profiling import profile\_report
 df.profile\_report()

## Overview

Dataset statistics	
Number of variables	15
Number of observations	457
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	53.7 KiB

Average record size in memory

### Variable types

NUM	14
CAT	1

#### Warnings

TotalDistance is highly correlated with TotalSteps and 1 other fields (TotalSteps,

TrackerDistance)

TotalSteps is highly correlated with TotalDistance and 1 other fields (TotalDistance,

TrackerDistance)

TrackerDistance is highly correlated with TotalSteps and 1 other fields (TotalSteps,

TotalDistance)

TotalSteps has 61 (13.3%) zeros

TotalDistance has 63 (13.8%) zeros

120.3 B

Out[8]:

4

### **EDA STEP BY STEP**

```
Id
                                                                                                   0
  Out[9]:
                         ActivityDate
                                                                                                   0
                         TotalSteps
                                                                                                   0
                                                                                                   0
                         TotalDistance
                         TrackerDistance
                                                                                                   0
                        LoggedActivitiesDistance
                                                                                                   0
                         VeryActiveDistance
                                                                                                   0
                         ModeratelyActiveDistance
                                                                                                   0
                         LightActiveDistance
                                                                                                   0
                         SedentaryActiveDistance
                                                                                                   0
                         VeryActiveMinutes
                                                                                                   0
                                                                                                   0
                         FairlyActiveMinutes
                        LightlyActiveMinutes
                                                                                                   0
                         SedentaryMinutes
                         Calories
                                                                                                   0
                         dtype: int64
In [11]:
                           df.ActivityDate.unique()
                       array(['3/25/2016', '3/26/2016', '3/27/2016', '3/28/2016', '3/29/2016', '3/30/2016', '3/31/2016', '4/1/2016', '4/2/2016', '4/3/2016', '4/4/2016', '4/5/2016', '4/6/2016', '4/7/2016', '4/8/2016', '4/9/2016', '4/10/2016', '4/11/2016', '4/12/2016', '3/12/2016', '3/13/2016', '3/14/2016', '3/15/2016', '3/16/2016', '3/17/2016', '3/18/2016', '3/19/2016', '3/20/2016', '3/21/2016', '3/22/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016', '3/2016'
                                            '3/23/2016', '3/24/2016'], dtype=object)
In [15]:
                           df['Date']=pd.DatetimeIndex(df['ActivityDate'])
In [17]:
                           df.drop('date',axis=1,inplace=True)
In [18]:
                           df.head(2)
Out[18]:
                                                  Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
                         0 1503960366
                                                                3/25/2016
                                                                                                11004
                                                                                                                                                                     7.11
                                                                                                                                                                                                                             0.0
                                                                                                                                                                                                                                                                      2.57
                                                                                                                                                                                                                                                                                                                            0.46
                                                                                                                                 7.11
                               1503960366
                                                                3/26/2016
                                                                                                17609
                                                                                                                                11.55
                                                                                                                                                                   11.55
                                                                                                                                                                                                                              0.0
                                                                                                                                                                                                                                                                      6.92
                                                                                                                                                                                                                                                                                                                             0.73
In [19]:
                           df['Year']=pd.DatetimeIndex(df['ActivityDate']).year
                           df['Month']=pd.DatetimeIndex(df['ActivityDate']).month
                           df['Day']=pd.DatetimeIndex(df['ActivityDate']).day
In [20]:
                           df.head(2)
                                                  Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
                         0 1503960366
                                                                3/25/2016
                                                                                                11004
                                                                                                                                  7 11
                                                                                                                                                                     7 11
                                                                                                                                                                                                                             0.0
                                                                                                                                                                                                                                                                      2 57
                                                                                                                                                                                                                                                                                                                            0.46
                                                                                                                                                                                                                              0.0
                              1503960366
                                                                3/26/2016
                                                                                                17609
                                                                                                                                11.55
                                                                                                                                                                   11.55
                                                                                                                                                                                                                                                                      6.92
                                                                                                                                                                                                                                                                                                                             0.73
In [23]:
                           df[['ActivityDate','Day','Month','Year']].head()
Out[23]:
                               ActivityDate Day Month Year
                                     3/25/2016
                                                                                 3 2016
                         0
                                                              25
                                     3/26/2016
                                                              26
                                                                                 3 2016
                                     3/27/2016
                                                              27
                                                                                     2016
                                     3/28/2016
                                                              28
                                                                                 3 2016
                                     3/29/2016
                                                              29
                                                                                 3 2016
```

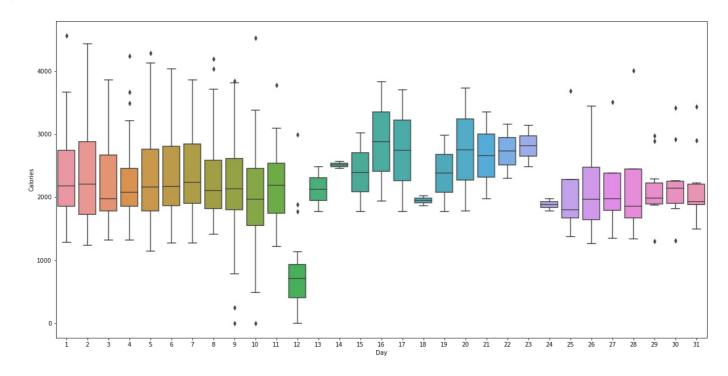
In [25]: ...

dt.drop(['ActivityDate'],axis=1).head() Id TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance LightActiveD 0 1503960366 11004 7 11 0.0 2 57 0.46 7 11 1503960366 17609 11.55 11.55 0.0 6.92 0.73 1503960366 12736 8.53 8.53 0.0 4.66 0.16 1503960366 13231 8.93 8.93 0.0 3.19 0.79 1503960366 12041 7.85 7.85 0.0 2.16 1.09

import matplotlib.pyplot as plt

In [29]: plt.figure(figsize=(20,10))
 sns.boxplot(x='Day',y='Calories',data=df)

Out[29]: <AxesSubplot:xlabel='Day', ylabel='Calories'>



In [30]: df['Week']=pd.DatetimeIndex(df['ActivityDate']).week

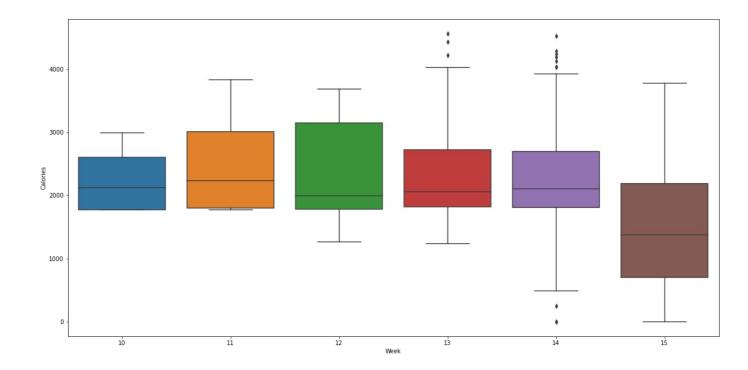
C:\Users\hp\AppData\Local\Temp/ipykernel\_11492/1452975531.py:1: FutureWarning: weekofyear and week have been depr ecated, please use DatetimeIndex.isocalendar().week instead, which returns a Series. To exactly reproduce the be havior of week and weekofyear and return an Index, you may call pd.Int64Index(idx.isocalendar().week) df['Week']=pd.DatetimeIndex(df['ActivityDate']).week

In [31]: df.head(2)

Out[31]: Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance 1503960366 3/25/2016 11004 7.11 7.11 0.0 2.57 0.46 1503960366 3/26/2016 17609 11.55 11.55 0.0 6.92 0.73

In [32]:
 plt.figure(figsize=(20,10))
 sns.boxplot(x='Week',y='Calories',data=df)

Out[32]: <AxesSubplot:xlabel='Week', ylabel='Calories'>

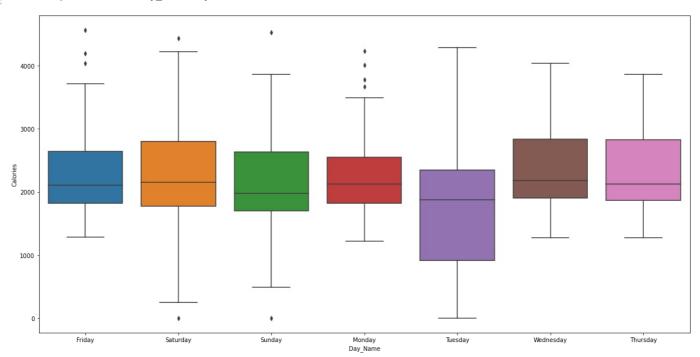


In [33]: df['Day Name']=pd.DatetimeIndex(df['ActivityDate']).day\_name() In [34]: df.head(2) Id ActivityDate TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance Out[34]: **0** 1503960366 3/25/2016 11004 7.11 7.11 0.0 2.57 0.46 **1** 1503960366 3/26/2016 17609 0.0 6.92 0.73 11.55 11.55

2 rows × 21 columns

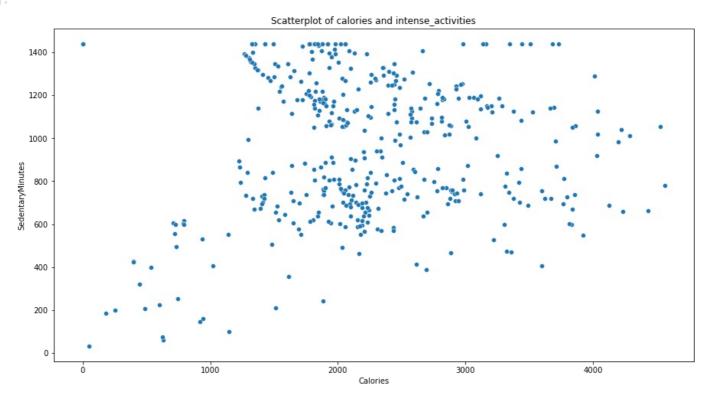
In [35]:
 plt.figure(figsize=(20,10))
 sns.boxplot(x='Day\_Name',y='Calories',data=df)

Out[35]. <AxesSubplot:xlabel='Day\_Name', ylabel='Calories'>



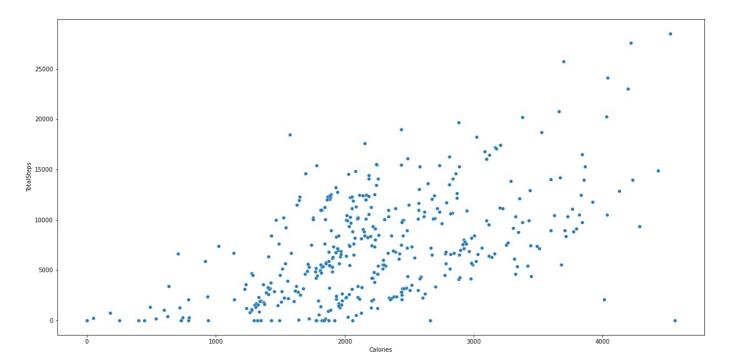
```
ax = sns.scatterplot(x='Calories', y='SedentaryMinutes', data=df)
ax.set_title('Scatterplot of calories and intense_activities')
```

Out[37]: Text(0.5, 1.0, 'Scatterplot of calories and intense\_activities')



```
In [38]:
    plt.figure(figsize=(20,10))
    sns.scatterplot(x='Calories',y='TotalSteps',data=df)
```

Out[38]: <AxesSubplot:xlabel='Calories', ylabel='TotalSteps'>



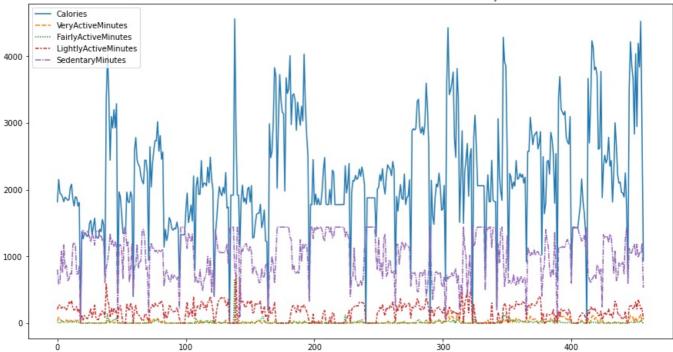
```
plt.figure(figsize=(20,20))
col_select = ['Calories','VeryActiveMinutes','FairlyActiveMinutes','LightlyActiveMinutes','SedentaryMinutes']
wide_df = df[col_select]

# figure size
plt.figure(figsize=(15,8))
# timeseries plot using lineplot
```

```
ax = sns.lineplot(data=wide_df)
ax.set_title('Un-normalized value of calories and different activities based on activity minutes')
```

Out[41]: Text(0.5, 1.0, 'Un-normalized value of calories and different activities based on activity minutes') <Figure size 1440x1440 with 0 Axes>





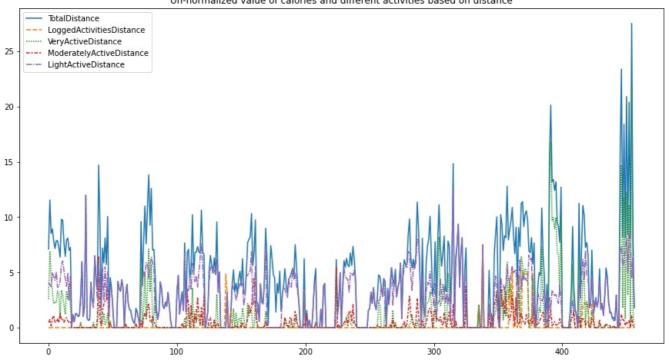
```
rol_select = ['TotalDistance', 'LoggedActivitiesDistance', 'VeryActiveDistance', 'ModeratelyActiveDistance', 'LightAwide_df1 = df[rol_select]

# figure size
plt.figure(figsize=(15,8))

# timeseries plot using lineplot
ax = sns.lineplot(data=wide_df1)
ax.set_title('Un-normalized value of calories and different activities based on distance')
```

Out[42]: Text(0.5, 1.0, 'Un-normalized value of calories and different activities based on distance')

#### Un-normalized value of calories and different activities based on distance



The EDA here gives us the insight about the relation between the active hours, the distance for which the user has moderate and intense activity and the calories burnt during that period.

In [ ]:

Loading [MathJax]/extensions/Safe.js