Duration Pulse Maxpulse Calories	count 32.000000 32.000000 32.000000 30.000000	mean 68.437500 103.500000 128.500000 266.013333	std 70.039591 7.832933 12.998759 164.876415		min 30.000000 90.000000 101.000000 -300.000000	25% 60.000000 100.000000 120.000000 247.000000	E0% 60 000000 102 500000 127 500000 382 200000	000000:121 000000:201 000000:00	75% 60.000000 106.500000 132.250000 343.975000	max 450 000000 130 00000 175 000000 479 000000			df.isnull().sum()		ion		Pulse 0	Maxpulse 0	Calories 2	dtype: int64		df.dtypes		ion	O		Maxpurse int64	hier	منهالات معارده	<pre>df['Calories'] = df['Calories'].abs()</pre>	print(df)					
Out[55]:													In [56]:		Out[56]:							In [57]:		Out[57]:						In [70]:						
						Calories	409.1	479.0	340.0	282.4 406 0	9.00.6	374.0	253.3	195.1	269.0	329.3	250.7	250.7	345,3	379.3	275.0	215.2	300.0	NaN	323.0	243.6	364.2	282.8	246.0	334.5	250.0	241.0	NaN	-280.0	380.3	243.0
						Maxpulse	130	145	135	1/5	127	136	134	133	124	147	120	120	128	132	123	120	120	112	123	125	131	191	132	126	120	118	132	132	129	115
		ignore')	2 2 2 2	ta.csv')		Pulse	110	117	103	109	102	110	104	109	86	103	100	100	106	104	86	86	100	96	103	/6	108	130	105	102	100	95	103	100	102	92
	import pandas as pd import numpy as np	import warnings warnings_filterwarnings('ignore')		In [69]: df = pd.read csv('dirtydata.csv')	(df)	Duration Date	60 2020/12/01			45 2020/12/04:		60 2020/12/07									•						60 2020/12/21	אמאו ל- 45 אמראר אם א			60 20201226	60 2020/12/27	60 2020/12/28	60 2020/12/29		60 2020/12/31'
	In [52]: impor	impor	3	In [69]: df =	print(df)	Dur	0	1	5 2	w 4	. 7	9	7	∞	6	10	11	12	13	14	15	16	17	18	19	97	21	22 در	23	25	26	27	28	29	30	31

```
If['Calories'] = df['Calories'].abs()
inint(df)
                                                                                                     int64
object
int64
int64
float64
f.isnull().sum()
                                                                                                                                                      Itype: object
                           Date 1
Pulse 0
Maxpulse 0
Calories 2
dtype: int64
                 Juration
                                                                                  if.dtypes
                                                                                                     Duration
Date
Pulse
Maxpulse
Calories
```

In [55]: df.describe()

In [54]: df.shape Out[54]: (32, 5)

```
print(df)
                                                                                                                                                                                                                                                                                                                                             *
                           3/9/24, 7:46 AM
                                                                                                                                                                                                                                                                                                                                             L
                         409.1
479.0
2840.0
2840.0
3840.0
3960.0
374.0
374.0
374.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
3840.0
                                                                                                                                                                                                                                                                                                                 380.3
                                                                                                                                                                                                                                                                                                        280.0
                                                                                                                                                                                                                                                                                               NaN
                Calories
Assignment 1
                Maxpulse
                         df['Calories'].mean()
                                                                                                                                                                                               2020/12/18'
2020/12/19'
2020/12/20'
2020/12/21'
NaN
2020/12/23'
2020/12/24'
                                                                                                                                                                                                                                                                                             2020/12/28'
2020/12/29'
2020/12/30'
2020/12/31'
                       2020/12/01'
2020/12/02'
2020/12/03'
2020/12/04'
2020/12/05'
2020/12/05'
2020/12/06'
                                                                                         2020/12/08'
2020/12/09'
                                                                                                                               2020/12/12'
2020/12/12'
2020/12/13'
                                                                                                                                                                    2020/12/15'
2020/12/16'
2020/12/17'
                                                                                                                                                                                                                                                                            20201226
                                                                                                                                                                                                                                                                                     2020/12/27
                                                                                                              2020/12/10
                                                                                                                      2020/12/11
                                                                                                                                                           2020/12/14
             89
                                                                                                                                                                                                                                                                                                                                                                         304
                          Out[59]:
3/9/24, 7:46 AM
                                                                                                                                                                                                                                                                                                                                             In
```

df['Calories'].fillna(x, inplace=True) Filling null values print(df) # In

```
df['Date'] = pd.to_datetime(df['Date'], format='mixed')
Calories 409.10 479.00 340.00 282.44 406.00 374.00 253.30 195.00 329.30 259.00 329.30 275.00 245.00 384.50 251.00 241.00 384.68 280.00 243.00 243.00 241.00 243.00 243.00 243.00 243.00 243.00 244.68 280.00 243.00 243.00 244.68 280.00 243.00 243.00 243.00 384.50 241.00 384.50 243.00 244.68 280.00 243.00 244.68 280.00 244.00 384.50 243.00 244.68 280.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 243.00 
                    Convert the date into a "Date" format
 2020/12/18'
2020/12/19'
2020/12/20'
2020/12/21'
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               2020/12/28'
2020/12/29'
2020/12/30'
2020/12/31'
                2020/12/01'
2020/12/02'
2020/12/03'
2020/12/04'
2020/12/06'
2020/12/06'
                                                                                                                                                                                                                                                                                                           2020/12/15'
2020/12/16'
2020/12/17'
                                                                                                                                                                                                                                                                                                                                                                                                                                                                  2020/12/23'
2020/12/24'
                                                                                                                                                                                                                                                                                                                                                                                                                                                   NaN
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            20201226
                                                                                                                                                        2020/12/08
                                                                                                                                                                                              2020/12/10
                                                                                                                                                                                                                                                                       2020/12/13
                                                                                                                                                                                                                                                                                          2020/12/14
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         2020/12/25
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              2020/12/27
                                                                                                                                                                           2020/12/09
                                                                                                                                                                                                               2020/12/11
                                                                                                                                                                                                                                  2020/12/12
                                                                                                                                                                                                                                                    2020/12/12
```

Assignment 1

Maxpulse

Assignment 1	Calories	409.10	479.00	340.00	282.40	406.00	300.00	374.00	253.30	195.10	269.00	329.30	250.70	250.70	345.30	379.30	275.00	215.20	300.00	304.68	323.00	243.00	364.20	282.00	300.00	246.00	334.50	250.00	241.00	304.68	280.00	380.30	243.00
•	Maxpulse	130	145	135	175	148	127	136	134	133	124	147	120	120	128	132	123	120	120	112	123	125	131	119	101	132	126	120	118	132	132	129	115
	Pulse	110	117	103	109	117	102	110	104	109	86	103	100	100	106	104	86	86	100	96	103	46	108	100	130	105	102	100	95	103	100	102	92
	Duration Date	60 2020-12-01	60 2020-12-02	60 2020-12-03	45 2020-12-04	45 2020-12-05	60 2020-12-06	60 2020-12-07	450 2020-12-08	30 2020-12-09	60 2020-12-10	60 2020-12-11	60 2020-12-12	60 2020-12-12	60 2020-12-13	60 2020-12-14	60 2020-12-15	60 2020-12-16	60 2020-12-17	45 2020-12-18	60 2020-12-19	45 2020-12-20	60 2020-12-21	45 NaT	60 2020-12-23	45 2020-12-24	60 2020-12-25	60 2020-12-26	60 2020-12-27	60 2020-12-28	60 2020-12-29	60 2020-12-30	60 2020-12-31
		0	П	7	3	4	2	9	7	∞	6	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	56	27	28	29	30	31
3/9/24, 7:46 AM																																	

In [73]: # 7th Location's duration value is 450 which is an outlier
df.loc[7, 'Duration'] = 45
print(df)

 Duration
 Date of 2020-12-01
 Pulse of 2020-12-01
 Maxpulse of 2020-10-00

 1
 60 2020-12-02
 117
 145
 449-10

 2
 60 2020-12-03
 117
 148
 440-00

 3
 45 2020-12-04
 109
 175
 282-40

 4
 5 2020-12-05
 117
 148
 406.00

 5
 60 2020-12-06
 102
 117
 148
 406.00

 6
 2020-12-07
 110
 113
 340.00

 7
 45 2020-12-09
 110
 113
 374.00

 8
 30 2020-12-09
 110
 113
 359.30

 9
 60 2020-12-10
 98
 124
 269.00

 10
 60 2020-12-11
 100
 120
 250.70

 11
 60 2020-12-12
 98
 124
 359.30

 12
 60 2020-12-13
 100
 120
 250.70

 13
 60 2020-12-13
 100
 120
 250.00

 14
 60 2020-12-13
 100
 120
 <t

Assignment 1

3/9/24, 7:46 AM

[63]: # Finding duplicates
df.duplicated().sum()

L

Out[63]: 1

In [74]: df.drop_duplicates(inplace=True)
 print(df)

Assignment 1
3/9/24, 7:46 AM

Calories	409.10	479.00	340.00	282.40	406.00	300.00	374.00	253.30	195.10	269.00	329.30	250.70	345.30	379.30	275.00	215.20	300.00	304.68	323.00	243.00	364.20	282.00	300.00	246.00	334.50	250.00	241.00	304.68	280.00	380.30	243.00
Maxpulse	130	145	135	175	148	127	136	134	133	124	147	120	128	132	123	120	120	112	123	125	131	119	101	132	126	120	118	132	132		115
Pulse	110	117	103	109	117	102	110	104	109	86	103	100	106	104	86	86	100	96	103	97	108	100	130	105	102	100	95	103	100	102	92
Date	2020-12-01	2020-12-02	2020-12-03	2020-12-04	2020-12-05	2020-12-06	2020-12-07	2020-12-08	2020-12-09	2020-12-10	2020-12-11	2020-12-12	2020-12-13	2020-12-14	2020-12-15	2020-12-16	2020-12-17	2020-12-18	2020-12-19	2020-12-20	2020-12-21	NaT	2020-12-23	2020-12-24	2020-12-25	2020-12-26	2020-12-27	2020-12-28	2020-12-29	2020-12-30	2020-12-31
Duration	99	99	99	45	45	99	99	45	30	99	99	99	99	99	99	99	99	45	99	45	99	45	99	45	99	99	99	99	99	99	99
	0	1	2	m	4	2	9	7	∞	6	10	11	13	14	15	16	17	18	19	20	21	22	23	24	25	56	27	28	59	30	31

3/9/24, 7:46 AM

Assignment 1

Out[76]:		Name	Team	Number	Position	Age	Height	Weight	College	Salary
	0	Avery Bradley	Boston Celtics	0	PG	25	2-Jun	180	Texas	7730337.0
	_	Jae Crowder	Boston Celtics	66	SF	25	e-Jun	235	Marquette	6796117.0
	7	John Holland	Boston Celtics	30	SG	27	5-Jun	205	Boston University	NaN
	m	R.J. Hunter	Boston Celtics	28	SG	22	5-Jun	185	Georgia State	1148640.0
	4	Jonas Jerebko	Boston Celtics	80	PF	29	10-Jun	231	NaN	5000000.0
	:	÷	:	:	ŧ	:	:	:	:	:
	452	Trey Lyles	Utah Jazz	41	PF	20	10-Jun	234	Kentucky	2239800.0
	453	Shelvin Mack	Utah Jazz	80	PG	26	3-Jun	203	Butler	2433333.0
	454	Raul Neto	Utah Jazz	25	PG	24	1-Jun	179	NaN	9000000
	455	Tibor Pleiss	Utah Jazz	21	U	26	3-Jul	256	NaN	2900000.0
	456	Jeff Withey	Utah Jazz	24	U	26	Jul-00	231	Kansas	947276.0

457 rows \times 9 columns

```
In [77]: data['Position'].value_counts()

Out[77]: Position
SG 102
PF 100
PG 92
SF 85
C 78
Name: count, dtype: int64
```

In [78]: data['Position'].replace(['SG','PF','PG','SF','C'], [1,2,3,4,5], inplace=True)
data

7/10

Assignment 1	
3/9/24, 7:46 AM	

Out[78]:		Name	Team	Number	Position	Age	Height	Weight	College	Salary
	0	Avery Bradley	Boston Celtics	0	С	25	2-Jun	180	Texas	7730337.0
	-	Jae Crowder	Boston Celtics	66	4	25	e-Jun	235	Marquette	6796117.0
	7	John Holland	Boston Celtics	30	-	27	5-Jun	205	Boston University	NaN
	m	R.J. Hunter	Boston Celtics	28	-	22	5-Jun	185	Georgia State	1148640.0
	4	Jonas Jerebko	Boston Celtics	∞	2	29	10-Jun	231	NaN	5000000.0
	:	:	:	:	:	÷	:	:	:	:
	452	Trey Lyles	Utah Jazz	41	7	20	10-Jun	234	Kentucky	2239800.0
	453	Shelvin Mack	Utah Jazz	∞	33	26	3-Jun	203	Butler	2433333.0
	454	Raul Neto	Utah Jazz	25	9	24	1-Jun	179	NaN	0.000006
	455	Tibor Pleiss	Utah Jazz	21	5	26	3-Jul	256	NaN	2900000.0
	456	Jeff Withey	Utah Jazz	24	5	26	Jul-00	231	Kansas	947276.0

457 rows × 9 columns

In [79]: category = pd.cut(data.Age, bins=[19,25,30,35,45], labels=['A','B','C','D'])
data.insert(5, 'Age_group', category)
data

3/9/24, 7:46 AM Assignment 1

Out[79]:	Name		Team Number	Position	Age	Age_group Height Weight	Height	Weight	College	
0	Avery Bradley	Boston Celtics	0	m	25	A	2-Jun	180	Texas	77
-	Jae Crowder	Boston Celtics	66	4	25	A	e-Jun	235	Marquette	29
7	John Holland	Boston Celtics	30	~	27	В	5-Jun	205	Boston University	
m	R.J. Hunter	Boston Celtics	28	~	22	A	5-Jun	185	Georgia State	
4	Jonas Jerebko	Boston Celtics	ω	2	29	В	10-Jun	231	NaN	50
:	1	:	:	:	:	:	:	:	:	
452	Trey Lyles	Utah Jazz	41	7	20	A	10-Jun	234	Kentucky	22
453	Shelvin Mack	Utah Jazz	∞	m	26	В	3-Jun	203	Butler	24
454	Raul Neto	Utah Jazz	25	m	24	A	1-Jun	179	NaN	6
455	Tibor Pleiss	Utah Jazz	21	5	26	В	3-Jul	256	NaN	29
456	Jeff Withey	Utah Jazz	24	2	26	В	Jul-00	231	Kansas	6
457 r	457 rows × 10 columns	columns								
		ı	l	ı	ı		ı	ı	ı	4

file:///C:/Users/PV/Downloads/Assignment 1.html