importing library pandas

```
import pandas as pd
```

importing into pandas dataframe

```
dataset = pd.read_csv("dirtydata.csv")

print(dataset)
```

prints the shape of dataset

```
dataset.shape

(32, 5)
```

shows datatypes dataset elements

```
dataset.dtypes

Duration      int64
Date         object
Pulse         int64
Maxpulse      int64
Calories    float64
dtype: object

dataset.columns

Index(['Duration', 'Date', 'Pulse', 'Maxpulse', 'Calories'],
dtype='object')
```

shows information of dataset

```
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Duration  32 non-null     int64
 1   Date      31 non-null     object
 2   Pulse     32 non-null     int64
 3   Maxpulse  32 non-null     int64
 4   Calories  30 non-null     float64
dtypes: float64(1), int64(3), object(1)
memory usage: 1.4+ KB
```

```
dataset.describe()
```

	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
50%	60.000000	102.500000	127.500000	282.200000
75%	60.000000	106.500000	132.250000	343.975000
max	450.000000	130.000000	175.000000	479.000000

```
dataset.isnull()
```

	Duration	Date	Pulse	Maxpulse	Calories
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
5	False	False	False	False	False
6	False	False	False	False	False
7	False	False	False	False	False
8	False	False	False	False	False
9	False	False	False	False	False
10	False	False	False	False	False
11	False	False	False	False	False
12	False	False	False	False	False
13	False	False	False	False	False
14	False	False	False	False	False
15	False	False	False	False	False
16	False	False	False	False	False
17	False	False	False	False	False
18	False	False	False	False	True
19	False	False	False	False	False
20	False	False	False	False	False
21	False	False	False	False	False
22	False	True	False	False	False
23	False	False	False	False	False
24	False	False	False	False	False
25	False	False	False	False	False
26	False	False	False	False	False
27	False	False	False	False	False
28	False	False	False	False	True
29	False	False	False	False	False
30	False	False	False	False	False
31	False	False	False	False	False

select down the null values of the dataset and sum

```
dataset.isnull().sum()
```

```
Duration    0
Date        1
Pulse       0
Maxpulse    0
Calories    2
dtype: int64
```

iloc shows the index location

```
dataset.iloc[:,1:3]
```

	Date	Pulse
0	'2020/12/01'	110
1	'2020/12/02'	117
2	'2020/12/03'	103
3	'2020/12/04'	109
4	'2020/12/05'	117
5	'2020/12/06'	102
6	'2020/12/07'	110
7	'2020/12/08'	104
8	'2020/12/09'	109
9	'2020/12/10'	98
10	'2020/12/11'	103
11	'2020/12/12'	100
12	'2020/12/12'	100
13	'2020/12/13'	106
14	'2020/12/14'	104
15	'2020/12/15'	98
16	'2020/12/16'	98
17	'2020/12/17'	100
18	'2020/12/18'	90
19	'2020/12/19'	103
20	'2020/12/20'	97
21	'2020/12/21'	108
22	NaN	100
23	'2020/12/23'	130
24	'2020/12/24'	105
25	'2020/12/25'	102
26	20201226	100
27	'2020/12/27'	92
28	'2020/12/28'	103
29	'2020/12/29'	100
30	'2020/12/30'	102
31	'2020/12/31'	92

```
dataset['Duration']
```

```
0      60
1      60
2      60
3      45
4      45
5      60
6      60
7     450
8      30
9      60
10     60
11     60
12     60
13     60
14     60
15     60
16     60
17     60
18     45
19     60
20     45
21     60
22     45
23     60
24     45
25     60
26     60
27     60
28     60
29     60
30     60
31     60
Name: Duration, dtype: int64
```

```
print(dataset.Duration,dataset.Date)
```

```
0      60
1      60
2      60
3      45
4      45
5      60
6      60
7     450
8      30
9      60
10     60
11     60
12     60
13     60
```

```
14     60
15     60
16     60
17     60
18     45
19     60
20     45
21     60
22     45
23     60
24     45
25     60
26     60
27     60
28     60
29     60
30     60
31     60
Name: Duration, dtype: int64 0     '2020/12/01'
1     '2020/12/02'
2     '2020/12/03'
3     '2020/12/04'
4     '2020/12/05'
5     '2020/12/06'
6     '2020/12/07'
7     '2020/12/08'
8     '2020/12/09'
9     '2020/12/10'
10    '2020/12/11'
11    '2020/12/12'
12    '2020/12/12'
13    '2020/12/13'
14    '2020/12/14'
15    '2020/12/15'
16    '2020/12/16'
17    '2020/12/17'
18    '2020/12/18'
19    '2020/12/19'
20    '2020/12/20'
21    '2020/12/21'
22             NaN
23    '2020/12/23'
24    '2020/12/24'
25    '2020/12/25'
26        20201226
27    '2020/12/27'
28    '2020/12/28'
29    '2020/12/29'
30    '2020/12/30'
```

```
31    '2020/12/31'
Name: Date, dtype: object
```

mean() function provides average

```
X = dataset['Calories'].mean()
```

fillna() is function fill the values inside the table . inplace is used to provide specific location

```
dataset['Calories'].fillna(X,inplace = True)

print(dataset)
```

```
    Duration          Date  Pulse  Maxpulse    Calories
0         60  '2020/12/01'    110       130  409.100000
1         60  '2020/12/02'    117       145  479.000000
2         60  '2020/12/03'    103       135  340.000000
3         45  '2020/12/04'    109       175  282.400000
4         45  '2020/12/05'    117       148  406.000000
5         60  '2020/12/06'    102       127 -300.000000
6         60  '2020/12/07'    110       136  374.000000
7        450  '2020/12/08'    104       134  253.300000
8         30  '2020/12/09'    109       133  195.100000
9         60  '2020/12/10'     98       124  269.000000
10        60  '2020/12/11'    103       147  329.300000
11        60  '2020/12/12'    100       120  250.700000
12        60  '2020/12/12'    100       120  250.700000
13        60  '2020/12/13'    106       128  345.300000
14        60  '2020/12/14'    104       132  379.300000
15        60  '2020/12/15'     98       123  275.000000
16        60  '2020/12/16'     98       120  215.200000
17        60  '2020/12/17'    100       120  300.000000
18        45  '2020/12/18'     90       112  266.013333
19        60  '2020/12/19'    103       123  323.000000
20        45  '2020/12/20'     97       125  243.000000
21        60  '2020/12/21'    108       131  364.200000
22        45           NaN    100       119  282.000000
23        60  '2020/12/23'    130       101  300.000000
24        45  '2020/12/24'    105       132  246.000000
25        60  '2020/12/25'    102       126  334.500000
26        60      20201226    100       120  250.000000
27        60  '2020/12/27'     92       118  241.000000
28        60  '2020/12/28'    103       132  266.013333
29        60  '2020/12/29'    100       132 -280.000000
30        60  '2020/12/30'    102       129  380.300000
31        60  '2020/12/31'     92       115  243.000000
```

abs() function convert negative values into positive values

```
dataset['Calories']=dataset['Calories'].abs()

dataset
```

	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.100000
1	60	'2020/12/02'	117	145	479.000000
2	60	'2020/12/03'	103	135	340.000000
3	45	'2020/12/04'	109	175	282.400000
4	45	'2020/12/05'	117	148	406.000000
5	60	'2020/12/06'	102	127	300.000000
6	60	'2020/12/07'	110	136	374.000000
7	450	'2020/12/08'	104	134	253.300000
8	30	'2020/12/09'	109	133	195.100000
9	60	'2020/12/10'	98	124	269.000000
10	60	'2020/12/11'	103	147	329.300000
11	60	'2020/12/12'	100	120	250.700000
12	60	'2020/12/12'	100	120	250.700000
13	60	'2020/12/13'	106	128	345.300000
14	60	'2020/12/14'	104	132	379.300000
15	60	'2020/12/15'	98	123	275.000000
16	60	'2020/12/16'	98	120	215.200000
17	60	'2020/12/17'	100	120	300.000000
18	45	'2020/12/18'	90	112	266.013333
19	60	'2020/12/19'	103	123	323.000000
20	45	'2020/12/20'	97	125	243.000000
21	60	'2020/12/21'	108	131	364.200000
22	45	NaN	100	119	282.000000
23	60	'2020/12/23'	130	101	300.000000
24	45	'2020/12/24'	105	132	246.000000
25	60	'2020/12/25'	102	126	334.500000
26	60	20201226	100	120	250.000000
27	60	'2020/12/27'	92	118	241.000000
28	60	'2020/12/28'	103	132	266.013333
29	60	'2020/12/29'	100	132	280.000000
30	60	'2020/12/30'	102	129	380.300000
31	60	'2020/12/31'	92	115	243.000000

dropna() function to drop operation

```
dataset.dropna(subset=['Date'],inplace = True)

dataset
```

	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.100000
1	60	'2020/12/02'	117	145	479.000000
2	60	'2020/12/03'	103	135	340.000000
3	45	'2020/12/04'	109	175	282.400000
4	45	'2020/12/05'	117	148	406.000000

```
5          60  '2020/12/06'    102        127  300.000000
6          60  '2020/12/07'    110        136  374.000000
7         450  '2020/12/08'    104        134  253.300000
8          30  '2020/12/09'    109        133  195.100000
9          60  '2020/12/10'     98        124  269.000000
10         60  '2020/12/11'    103        147  329.300000
11         60  '2020/12/12'    100        120  250.700000
12         60  '2020/12/12'    100        120  250.700000
13         60  '2020/12/13'    106        128  345.300000
14         60  '2020/12/14'    104        132  379.300000
15         60  '2020/12/15'     98        123  275.000000
16         60  '2020/12/16'     98        120  215.200000
17         60  '2020/12/17'    100        120  300.000000
18         45  '2020/12/18'     90        112  266.013333
19         60  '2020/12/19'    103        123  323.000000
20         45  '2020/12/20'     97        125  243.000000
21         60  '2020/12/21'    108        131  364.200000
23         60  '2020/12/23'    130        101  300.000000
24         45  '2020/12/24'    105        132  246.000000
25         60  '2020/12/25'    102        126  334.500000
26         60      20201226    100        120  250.000000
27         60  '2020/12/27'     92        118  241.000000
28         60  '2020/12/28'    103        132  266.013333
29         60  '2020/12/29'    100        132  280.000000
30         60  '2020/12/30'    102        129  380.300000
31         60  '2020/12/31'     92        115  243.000000
```

```
dataset.dtypes
```

```
Duration      int64
Date         object
Pulse         int64
Maxpulse      int64
Calories    float64
dtype: object
```

to update date in specified format use to date

```
dataset['Date'] = pd.to_datetime(dataset['Date'],format = 'mixed')
```

```
dataset
```

```
    Duration       Date  Pulse  Maxpulse    Calories
0         60 2020-12-01    110       130  409.100000
1         60 2020-12-02    117       145  479.000000
2         60 2020-12-03    103       135  340.000000
3         45 2020-12-04    109       175  282.400000
4         45 2020-12-05    117       148  406.000000
5         60 2020-12-06    102       127  300.000000
6         60 2020-12-07    110       136  374.000000
```

```
7         450 2020-12-08    104        134  253.300000
8          30 2020-12-09    109        133  195.100000
9          60 2020-12-10     98        124  269.000000
10         60 2020-12-11    103        147  329.300000
11         60 2020-12-12    100        120  250.700000
12         60 2020-12-12    100        120  250.700000
13         60 2020-12-13    106        128  345.300000
14         60 2020-12-14    104        132  379.300000
15         60 2020-12-15     98        123  275.000000
16         60 2020-12-16     98        120  215.200000
17         60 2020-12-17    100        120  300.000000
18         45 2020-12-18     90        112  266.013333
19         60 2020-12-19    103        123  323.000000
20         45 2020-12-20     97        125  243.000000
21         60 2020-12-21    108        131  364.200000
23         60 2020-12-23    130        101  300.000000
24         45 2020-12-24    105        132  246.000000
25         60 2020-12-25    102        126  334.500000
26         60 2020-12-26    100        120  250.000000
27         60 2020-12-27     92        118  241.000000
28         60 2020-12-28    103        132  266.013333
29         60 2020-12-29    100        132  280.000000
30         60 2020-12-30    102        129  380.300000
31         60 2020-12-31     92        115  243.000000
```

astype(int) convert data entries into integer

```
dataset['Calories']= dataset['Calories'].astype(int)

dataset
```

```
    Duration       Date  Pulse  Maxpulse  Calories
0         60 2020-12-01    110       130       409
1         60 2020-12-02    117       145       479
2         60 2020-12-03    103       135       340
3         45 2020-12-04    109       175       282
4         45 2020-12-05    117       148       406
5         60 2020-12-06    102       127       300
6         60 2020-12-07    110       136       374
7        450 2020-12-08    104       134       253
8         30 2020-12-09    109       133       195
9         60 2020-12-10     98       124       269
10        60 2020-12-11    103       147       329
11        60 2020-12-12    100       120       250
12        60 2020-12-12    100       120       250
13        60 2020-12-13    106       128       345
14        60 2020-12-14    104       132       379
15        60 2020-12-15     98       123       275
16        60 2020-12-16     98       120       215
```

```
17         60 2020-12-17    100        120        300
18         45 2020-12-18     90        112        266
19         60 2020-12-19    103        123        323
20         45 2020-12-20     97        125        243
21         60 2020-12-21    108        131        364
23         60 2020-12-23    130        101        300
24         45 2020-12-24    105        132        246
25         60 2020-12-25    102        126        334
26         60 2020-12-26    100        120        250
27         60 2020-12-27     92        118        241
28         60 2020-12-28    103        132        266
29         60 2020-12-29    100        132        280
30         60 2020-12-30    102        129        380
31         60 2020-12-31     92        115        243
```

updating the value of specific location

```
dataset.loc[7,'Duration']=45

dataset
```

```
    Duration       Date  Pulse  Maxpulse  Calories
0         60 2020-12-01    110        130        409
1         60 2020-12-02    117        145        479
2         60 2020-12-03    103        135        340
3         45 2020-12-04    109        175        282
4         45 2020-12-05    117        148        406
5         60 2020-12-06    102        127        300
6         60 2020-12-07    110        136        374
7         45 2020-12-08    104        134        253
8         30 2020-12-09    109        133        195
9         60 2020-12-10     98        124        269
10        60 2020-12-11    103        147        329
11        60 2020-12-12    100        120        250
12        60 2020-12-12    100        120        250
13        60 2020-12-13    106        128        345
14        60 2020-12-14    104        132        379
15        60 2020-12-15     98        123        275
16        60 2020-12-16     98        120        215
17        60 2020-12-17    100        120        300
18        45 2020-12-18     90        112        266
19        60 2020-12-19    103        123        323
20        45 2020-12-20     97        125        243
21        60 2020-12-21    108        131        364
23        60 2020-12-23    130        101        300
24        45 2020-12-24    105        132        246
25        60 2020-12-25    102        126        334
26        60 2020-12-26    100        120        250
27        60 2020-12-27     92        118        241
```

```
28         60 2020-12-28    103        132        266
29         60 2020-12-29    100        132        280
30         60 2020-12-30    102        129        380
31         60 2020-12-31     92        115        243
```

```
dataset.duplicated()
```

```
0     False
1     False
2     False
3     False
4     False
5     False
6     False
7     False
8     False
9     False
10    False
11    False
12     True
13    False
14    False
15    False
16    False
17    False
18    False
19    False
20    False
21    False
23    False
24    False
25    False
26    False
27    False
28    False
29    False
30    False
31    False
dtype: bool
```

```
dataset.duplicated().sum()
```

```
1
```

```
dataset.drop_duplicates(inplace = True)
```

```
dataset
```

```
    Duration       Date  Pulse  Maxpulse  Calories
0         60 2020-12-01    110       130       409
1         60 2020-12-02    117       145       479
```

```
2          60 2020-12-03    103        135        340
3          45 2020-12-04    109        175        282
4          45 2020-12-05    117        148        406
5          60 2020-12-06    102        127        300
6          60 2020-12-07    110        136        374
7          45 2020-12-08    104        134        253
8          30 2020-12-09    109        133        195
9          60 2020-12-10     98        124        269
10         60 2020-12-11    103        147        329
11         60 2020-12-12    100        120        250
13         60 2020-12-13    106        128        345
14         60 2020-12-14    104        132        379
15         60 2020-12-15     98        123        275
16         60 2020-12-16     98        120        215
17         60 2020-12-17    100        120        300
18         45 2020-12-18     90        112        266
19         60 2020-12-19    103        123        323
20         45 2020-12-20     97        125        243
21         60 2020-12-21    108        131        364
23         60 2020-12-23    130        101        300
24         45 2020-12-24    105        132        246
25         60 2020-12-25    102        126        334
26         60 2020-12-26    100        120        250
27         60 2020-12-27     92        118        241
28         60 2020-12-28    103        132        266
29         60 2020-12-29    100        132        280
30         60 2020-12-30    102        129        380
31         60 2020-12-31     92        115        243
```

```
dataset.duplicated().sum()
```

```
0
```

performing the or operation

```
dataset[(dataset['Duration']==45)|(dataset['Maxpulse']==120)]
```

```
    Duration       Date  Pulse  Maxpulse  Calories
3          45 2020-12-04    109        175        282
4          45 2020-12-05    117        148        406
7          45 2020-12-08    104        134        253
11         60 2020-12-12    100        120        250
16         60 2020-12-16     98        120        215
17         60 2020-12-17    100        120        300
18         45 2020-12-18     90        112        266
20         45 2020-12-20     97        125        243
24         45 2020-12-24    105        132        246
26         60 2020-12-26    100        120        250
```

```
df1 = dataset.loc[:,['Duration','Calories']]
```

```
df1
```

	Duration	Calories
0	60	409
1	60	479
2	60	340
3	45	282
4	45	406
5	60	300
6	60	374
7	45	253
8	30	195
9	60	269
10	60	329
11	60	250
13	60	345
14	60	379
15	60	275
16	60	215
17	60	300
18	45	266
19	60	323
20	45	243
21	60	364
23	60	300
24	45	246
25	60	334
26	60	250
27	60	241
28	60	266
29	60	280
30	60	380
31	60	243

selecting all colums except maxpulse

```
df2 = dataset.loc[:,dataset.columns!= 'Maxpulse']

df2
```

	Duration	Date	Pulse	Calories
0	60	2020-12-01	110	409
1	60	2020-12-02	117	479
2	60	2020-12-03	103	340
3	45	2020-12-04	109	282
4	45	2020-12-05	117	406
5	60	2020-12-06	102	300
6	60	2020-12-07	110	374
7	45	2020-12-08	104	253
8	30	2020-12-09	109	195

```
9          60 2020-12-10     98        269
10         60 2020-12-11    103        329
11         60 2020-12-12    100        250
13         60 2020-12-13    106        345
14         60 2020-12-14    104        379
15         60 2020-12-15     98        275
16         60 2020-12-16     98        215
17         60 2020-12-17    100        300
18         45 2020-12-18     90        266
19         60 2020-12-19    103        323
20         45 2020-12-20     97        243
21         60 2020-12-21    108        364
23         60 2020-12-23    130        300
24         45 2020-12-24    105        246
25         60 2020-12-25    102        334
26         60 2020-12-26    100        250
27         60 2020-12-27     92        241
28         60 2020-12-28    103        266
29         60 2020-12-29    100        280
30         60 2020-12-30    102        380
31         60 2020-12-31     92        243
```

updating the name of column maxpulse to new

```
df2['new']=dataset['Maxpulse']

df2
```

```
    Duration       Date  Pulse  Calories  new
0          60 2020-12-01    110        409  130
1          60 2020-12-02    117        479  145
2          60 2020-12-03    103        340  135
3          45 2020-12-04    109        282  175
4          45 2020-12-05    117        406  148
5          60 2020-12-06    102        300  127
6          60 2020-12-07    110        374  136
7          45 2020-12-08    104        253  134
8          30 2020-12-09    109        195  133
9          60 2020-12-10     98        269  124
10         60 2020-12-11    103        329  147
11         60 2020-12-12    100        250  120
13         60 2020-12-13    106        345  128
14         60 2020-12-14    104        379  132
15         60 2020-12-15     98        275  123
16         60 2020-12-16     98        215  120
17         60 2020-12-17    100        300  120
18         45 2020-12-18     90        266  112
19         60 2020-12-19    103        323  123
20         45 2020-12-20     97        243  125
```

```
21         60 2020-12-21    108        364  131
23         60 2020-12-23    130        300  101
24         45 2020-12-24    105        246  132
25         60 2020-12-25    102        334  126
26         60 2020-12-26    100        250  120
27         60 2020-12-27     92        241  118
28         60 2020-12-28    103        266  132
29         60 2020-12-29    100        280  132
30         60 2020-12-30    102        380  129
31         60 2020-12-31     92        243  115
```

```
dataset
```

```
    Duration       Date  Pulse  Maxpulse  Calories
0         60 2020-12-01    110       130       409
1         60 2020-12-02    117       145       479
2         60 2020-12-03    103       135       340
3         45 2020-12-04    109       175       282
4         45 2020-12-05    117       148       406
5         60 2020-12-06    102       127       300
6         60 2020-12-07    110       136       374
7         45 2020-12-08    104       134       253
8         30 2020-12-09    109       133       195
9         60 2020-12-10     98       124       269
10        60 2020-12-11    103       147       329
11        60 2020-12-12    100       120       250
13        60 2020-12-13    106       128       345
14        60 2020-12-14    104       132       379
15        60 2020-12-15     98       123       275
16        60 2020-12-16     98       120       215
17        60 2020-12-17    100       120       300
18        45 2020-12-18     90       112       266
19        60 2020-12-19    103       123       323
20        45 2020-12-20     97       125       243
21        60 2020-12-21    108       131       364
23        60 2020-12-23    130       101       300
24        45 2020-12-24    105       132       246
25        60 2020-12-25    102       126       334
26        60 2020-12-26    100       120       250
27        60 2020-12-27     92       118       241
28        60 2020-12-28    103       132       266
29        60 2020-12-29    100       132       280
30        60 2020-12-30    102       129       380
31        60 2020-12-31     92       115       243
```

various operations on nba dataset

```
import pandas as pd

df = pd.read_csv("nba.csv")
```

```
df
```

```
                Name             Team  Number Position   Age Height
Weight  \
0    Avery Bradley  Boston Celtics     0.0       PG  25.0    6-2
180.0
1      Jae Crowder  Boston Celtics    99.0       SF  25.0    6-6
235.0
2     John Holland  Boston Celtics    30.0       SG  27.0    6-5
205.0
3      R.J. Hunter  Boston Celtics    28.0       SG  22.0    6-5
185.0
4    Jonas Jerebko  Boston Celtics     8.0       PF  29.0   6-10
231.0
..               ...              ...     ...      ...   ...    ...      .
..
453   Shelvin Mack       Utah Jazz     8.0       PG  26.0    6-3
203.0
454      Raul Neto       Utah Jazz    25.0       PG  24.0    6-1
179.0
455   Tibor Pleiss       Utah Jazz    21.0        C  26.0    7-3
256.0
456    Jeff Withey       Utah Jazz    24.0        C  26.0    7-0
231.0
457            NaN              NaN     NaN      NaN   NaN    NaN
NaN

                College     Salary
0                 Texas  7730337.0
1             Marquette  6796117.0
2    Boston University        NaN
3        Georgia State  1148640.0
4                   NaN  5000000.0
..                  ...        ...
453             Butler  2433333.0
454                NaN   900000.0
455                NaN  2900000.0
456             Kansas   947276.0
457                NaN        NaN

[458 rows x 9 columns]
```

```
df.dtypes
```

```
Name         object
Team         object
Number        int64
Position     object
Age           int64
Height       object
```

```
Weight        int64
College      object
Salary      float64
dtype: object
```

```
df['Position'].unique()
```

```
array(['PG', 'SF', 'SG', 'PF', 'C'], dtype=object)
```

```
dfn = df
dfn['Position'].replace(['PG','SF','SG','PF','C'],[0,1,2,3,4],inplace
= True)
```

```
/tmp/ipykernel_5112/2764792939.py:2: FutureWarning: A value is trying
to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.


  dfn['Position'].replace(['PG','SF','SG','PF','C'],
[0,1,2,3,4],inplace = True)
/tmp/ipykernel_5112/2764792939.py:2: FutureWarning: Downcasting
behavior in `replace` is deprecated and will be removed in a future
version. To retain the old behavior, explicitly call
`result.infer_objects(copy=False)`. To opt-in to the future behavior,
set `pd.set_option('future.no_silent_downcasting', True)`
  dfn['Position'].replace(['PG','SF','SG','PF','C'],
[0,1,2,3,4],inplace = True)
```

```
dfn
```

```
              Name            Team  Number  Position  Age  Height
Weight  \
0    Avery Bradley  Boston Celtics       0         0   25   2-Jun
180
1      Jae Crowder  Boston Celtics      99         1   25   6-Jun
235
2     John Holland  Boston Celtics      30         2   27   5-Jun
205
3      R.J. Hunter  Boston Celtics      28         2   22   5-Jun
185
4    Jonas Jerebko  Boston Celtics       8         3   29  10-Jun
231
..             ...             ...     ...       ...  ...     ...
```

```
...
452     Trey Lyles       Utah Jazz      41          3   20  10-Jun
234
453   Shelvin Mack       Utah Jazz       8          0   26   3-Jun
203
454      Raul Neto       Utah Jazz      25          0   24   1-Jun
179
455   Tibor Pleiss       Utah Jazz      21          4   26   3-Jul
256
456    Jeff Withey       Utah Jazz      24          4   26  Jul-00
231

                College     Salary
0                 Texas  7730337.0
1             Marquette  6796117.0
2    Boston University        NaN
3        Georgia State  1148640.0
4                   NaN  5000000.0
..                  ...        ...
452           Kentucky  2239800.0
453             Butler  2433333.0
454                NaN   900000.0
455                NaN  2900000.0
456             Kansas   947276.0

[457 rows x 9 columns]
```

```
from sklearn import preprocessing

lbt = preprocessing.LabelEncoder()

df['Position']= lbt.fit_transform(df['Position'])

df
```

```
                Name            Team  Number  Position  Age  Height
Weight  \
0    Avery Bradley  Boston Celtics       0         0   25   2-Jun
180
1      Jae Crowder  Boston Celtics      99         1   25   6-Jun
235
2     John Holland  Boston Celtics      30         2   27   5-Jun
205
3      R.J. Hunter  Boston Celtics      28         2   22   5-Jun
185
4    Jonas Jerebko  Boston Celtics       8         3   29  10-Jun
231
..             ...             ...     ...       ...  ...     ...
...
452     Trey Lyles       Utah Jazz      41         3   20  10-Jun
234
```

```
453   Shelvin Mack       Utah Jazz       8          0   26   3-Jun
203
454      Raul Neto       Utah Jazz      25          0   24   1-Jun
179
455   Tibor Pleiss       Utah Jazz      21          4   26   3-Jul
256
456    Jeff Withey       Utah Jazz      24          4   26  Jul-00
231

                College     Salary
0                 Texas  7730337.0
1             Marquette  6796117.0
2    Boston University        NaN
3        Georgia State  1148640.0
4                   NaN  5000000.0
..                  ...        ...
452           Kentucky  2239800.0
453             Butler  2433333.0
454                NaN   900000.0
455                NaN  2900000.0
456             Kansas   947276.0

[457 rows x 9 columns]
```

```
df['Position'].unique()
```

```
array([0, 1, 2, 3, 4])
```

```
df['Position'].value_counts()
```

```
Position
2    102
3    100
0     92
1     85
4     78
Name: count, dtype: int64
```

```
df['Age'].dropna(inplace = True)
```

```
df
```

```
               Name            Team  Number  Position  Age  Height
Weight  \
0    Avery Bradley  Boston Celtics       0         0   25   2-Jun
180
1      Jae Crowder  Boston Celtics      99         1   25   6-Jun
235
2     John Holland  Boston Celtics      30         2   27   5-Jun
205
3      R.J. Hunter  Boston Celtics      28         2   22   5-Jun
```

```
185
4    Jonas Jerebko  Boston Celtics       8         3   29  10-Jun
231
..             ...             ...     ...       ...  ...     ...
...
452     Trey Lyles       Utah Jazz      41         3   20  10-Jun
234
453   Shelvin Mack       Utah Jazz       8         0   26   3-Jun
203
454      Raul Neto       Utah Jazz      25         0   24   1-Jun
179
455   Tibor Pleiss       Utah Jazz      21         4   26   3-Jul
256
456    Jeff Withey       Utah Jazz      24         4   26  Jul-00
231

                College     Salary
0                 Texas  7730337.0
1             Marquette  6796117.0
2    Boston University        NaN
3        Georgia State  1148640.0
4                   NaN  5000000.0
..                  ...        ...
452           Kentucky  2239800.0
453             Butler  2433333.0
454                NaN   900000.0
455                NaN  2900000.0
456             Kansas   947276.0

[457 rows x 9 columns]
```

```
df['Age'].isnull().sum()
```

```
0
```

```
category =
pd.cut(df.Age,bins=[19,25,30,35,45],labels=['A','B','C','D'])

df.insert(5,'Age1',category)

df
```

```
               Name            Team  Number  Position  Age Age1  Height
\
0    Avery Bradley  Boston Celtics       0         0   25    A   2-Jun

1      Jae Crowder  Boston Celtics      99         1   25    A   6-Jun

2     John Holland  Boston Celtics      30         2   27    B   5-Jun

3      R.J. Hunter  Boston Celtics      28         2   22    A   5-Jun
```

```
4    Jonas Jerebko  Boston Celtics       8         3   29    B  10-Jun

..             ...             ...     ...       ...  ...  ...     ...

452     Trey Lyles       Utah Jazz      41         3   20    A  10-Jun

453   Shelvin Mack       Utah Jazz       8         0   26    B   3-Jun

454      Raul Neto       Utah Jazz      25         0   24    A   1-Jun

455   Tibor Pleiss       Utah Jazz      21         4   26    B   3-Jul

456    Jeff Withey       Utah Jazz      24         4   26    B  Jul-00


     Weight            College     Salary
0       180              Texas  7730337.0
1       235          Marquette  6796117.0
2       205  Boston University        NaN
3       185      Georgia State  1148640.0
4       231                NaN  5000000.0
..      ...                ...        ...
452     234           Kentucky  2239800.0
453     203             Butler  2433333.0
454     179                NaN   900000.0
455     256                NaN  2900000.0
456     231             Kansas   947276.0

[457 rows x 10 columns]

odf = df.drop('College' , axis =1)

odf

               Name            Team  Number        Position   Age
Height  \
0    Avery Bradley  Boston Celtics     0.0              PG  25.0    6-
2
1      Jae Crowder  Boston Celtics    99.0              SF  25.0    6-
6
2     John Holland  Boston Celtics    30.0              SG  27.0    6-
5
3      R.J. Hunter  Boston Celtics    28.0              SG  22.0    6-
5
4    Jonas Jerebko  Boston Celtics     8.0              PF  29.0   6-
10
..             ...             ...     ...             ...   ...    ..
.
453   Shelvin Mack       Utah Jazz     8.0              PG  26.0    6-
3
```

```
454      Raul Neto       Utah Jazz    25.0              PG  24.0    6-
1
455   Tibor Pleiss       Utah Jazz    21.0               C  26.0    7-
3
456    Jeff Withey       Utah Jazz    24.0               C  26.0    7-
0
457             NaN              NaN     NaN  5077829.215909   NaN
NaN

     Weight        Salary
0     180.0  7.730337e+06
1     235.0  6.796117e+06
2     205.0  5.077829e+06
3     185.0  1.148640e+06
4     231.0  5.000000e+06
..      ...           ...
453   203.0  2.433333e+06
454   179.0  9.000000e+05
455   256.0  2.900000e+06
456   231.0  9.472760e+05
457     NaN  5.077829e+06

[458 rows x 8 columns]
```

```
PG = df[(df['Position']=='PG')]

PG
```

```
               Name                     Team  Number Position   Age
Height  \
0    Avery Bradley          Boston Celtics     0.0       PG  25.0
6-2
8     Terry Rozier          Boston Celtics    12.0       PG  22.0
6-2
9     Marcus Smart          Boston Celtics    36.0       PG  22.0
6-4
11   Isaiah Thomas          Boston Celtics     4.0       PG  27.0
5-9
19    Jarrett Jack           Brooklyn Nets     2.0       PG  32.0
6-3
..             ...                     ...     ...      ...   ...    .
..
440  Brian Roberts  Portland Trail Blazers     2.0       PG  30.0
6-1
443     Trey Burke               Utah Jazz     3.0       PG  23.0
6-1
445     Dante Exum               Utah Jazz    11.0       PG  20.0
6-6
453   Shelvin Mack               Utah Jazz     8.0       PG  26.0
6-3
```

```
454      Raul Neto                Utah Jazz    25.0       PG  24.0
6-1
```

```
     Weight          College     Salary
0     180.0            Texas  7730337.0
8     190.0       Louisville  1824360.0
9     220.0  Oklahoma State  3431040.0
11    185.0       Washington  6912869.0
19    200.0    Georgia Tech  6300000.0
..      ...              ...        ...
440   173.0           Dayton  2854940.0
443   191.0         Michigan  2658240.0
445   190.0              NaN  3777720.0
453   203.0           Butler  2433333.0
454   179.0              NaN   900000.0

[92 rows x 9 columns]
```

```
X = PG['Salary'].mean()
```

```
X
```

```
5077829.215909091
```

```
PG['Salary'].fillna(X,inplace = True)
```

```
/tmp/ipykernel_6266/2443542409.py:1: FutureWarning: A value is trying
to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.


  PG['Salary'].fillna(X,inplace = True)
/tmp/ipykernel_6266/2443542409.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  PG['Salary'].fillna(X,inplace = True)
```

```
PG
```

```
               Name                     Team  Number Position   Age
Height  \
0     Avery Bradley          Boston Celtics     0.0       PG  25.0
6-2
8      Terry Rozier          Boston Celtics    12.0       PG  22.0
6-2
9      Marcus Smart          Boston Celtics    36.0       PG  22.0
6-4
11    Isaiah Thomas          Boston Celtics     4.0       PG  27.0
5-9
19     Jarrett Jack           Brooklyn Nets     2.0       PG  32.0
6-3
..              ...                     ...     ...      ...   ...    .
..
440   Brian Roberts  Portland Trail Blazers     2.0       PG  30.0
6-1
443      Trey Burke               Utah Jazz     3.0       PG  23.0
6-1
445      Dante Exum               Utah Jazz    11.0       PG  20.0
6-6
453    Shelvin Mack               Utah Jazz     8.0       PG  26.0
6-3
454       Raul Neto               Utah Jazz    25.0       PG  24.0
6-1

     Weight         College     Salary
0     180.0           Texas  7730337.0
8     190.0      Louisville  1824360.0
9     220.0  Oklahoma State  3431040.0
11    185.0      Washington  6912869.0
19    200.0    Georgia Tech  6300000.0
..      ...             ...        ...
440   173.0          Dayton  2854940.0
443   191.0        Michigan  2658240.0
445   190.0             NaN  3777720.0
453   203.0          Butler  2433333.0
454   179.0             NaN   900000.0

[92 rows x 9 columns]
```

```
C = df[(df['Position']=='C')]
SF = df[(df['Position']=='SF')]
SG = df[(df['Position']=='SG')]
y= C['Salary'].mean()
z = SF['Salary'].mean()
w=SG['Salary'].mean()

SG['Salary'].fillna(w,inplace = True)
SF['Salary'].fillna(z,inplace = True)
C['Salary'].fillna(y,inplace = True)
```

```
/tmp/ipykernel_6266/3803027798.py:1: FutureWarning: A value is trying
to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.


  SG['Salary'].fillna(w,inplace = True)
/tmp/ipykernel_6266/3803027798.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  SG['Salary'].fillna(w,inplace = True)
/tmp/ipykernel_6266/3803027798.py:2: FutureWarning: A value is trying
to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
df[col].method(value) instead, to perform the operation inplace on the
original object.


  SF['Salary'].fillna(z,inplace = True)
/tmp/ipykernel_6266/3803027798.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  SF['Salary'].fillna(z,inplace = True)
/tmp/ipykernel_6266/3803027798.py:3: FutureWarning: A value is trying
to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never
work because the intermediate object on which we are setting values
always behaves as a copy.

For example, when doing 'df[col].method(value, inplace=True)', try
using 'df.method({col: value}, inplace=True)' or df[col] =
```

```
df[col].method(value) instead, to perform the operation inplace on the
original object.


  C['Salary'].fillna(y,inplace = True)
/tmp/ipykernel_6266/3803027798.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  C['Salary'].fillna(y,inplace = True)

SF
SG
C

                 Name                      Team  Number Position   Age
Height  \
7       Kelly Olynyk          Boston Celtics    41.0        C  25.0
7-0
10   Jared Sullinger          Boston Celtics     7.0        C  24.0
6-9
14      Tyler Zeller          Boston Celtics    44.0        C  26.0
7-0
23       Brook Lopez           Brooklyn Nets    11.0        C  28.0
7-0
27        Henry Sims           Brooklyn Nets    14.0        C  26.0
6-10
..               ...                     ...     ...      ...   ...
...
434      Chris Kaman  Portland Trail Blazers    35.0        C  34.0
7-0
439    Mason Plumlee  Portland Trail Blazers    24.0        C  26.0
6-11
447      Rudy Gobert              Utah Jazz    27.0        C  23.0
7-1
455     Tibor Pleiss              Utah Jazz    21.0        C  26.0
7-3
456      Jeff Withey              Utah Jazz    24.0        C  26.0
7-0

     Weight           College      Salary
7     238.0           Gonzaga   2165160.0
10    260.0        Ohio State   2569260.0
14    253.0    North Carolina   2616975.0
23    275.0          Stanford  19689000.0
27    248.0        Georgetown    947276.0
..      ...               ...         ...
434   265.0  Central Michigan   5016000.0
```

```
439   235.0              Duke   1415520.0
447   245.0               NaN   1175880.0
455   256.0               NaN   2900000.0
456   231.0            Kansas    947276.0

[78 rows x 9 columns]
```

SG

```
                  Name                     Team  Number Position   Age
Height  \
2         John Holland          Boston Celtics    30.0       SG  27.0
6-5
3          R.J. Hunter          Boston Celtics    28.0       SG  22.0
6-5
12         Evan Turner          Boston Celtics    11.0       SG  27.0
6-7
13         James Young          Boston Celtics    13.0       SG  20.0
6-6
15   Bojan Bogdanovic           Brooklyn Nets    44.0       SG  27.0
6-8
..                ...                     ...     ...      ...   ...
...
433  Gerald Henderson  Portland Trail Blazers     9.0       SG  28.0
6-5
437     C.J. McCollum  Portland Trail Blazers     3.0       SG  24.0
6-4
438      Luis Montero  Portland Trail Blazers    44.0       SG  23.0
6-7
444        Alec Burks              Utah Jazz    10.0       SG  24.0
6-6
449       Rodney Hood              Utah Jazz     5.0       SG  23.0
6-8

     Weight            College        Salary
2     205.0  Boston University  5.077829e+06
3     185.0      Georgia State  1.148640e+06
12    220.0         Ohio State  3.425510e+06
13    215.0           Kentucky  1.749840e+06
15    216.0                NaN  3.425510e+06
..      ...                ...           ...
433   215.0               Duke  6.000000e+06
437   200.0             Lehigh  2.525160e+06
438   185.0     Westchester CC  5.250930e+05
444   214.0           Colorado  9.463484e+06
449   206.0               Duke  1.348440e+06

[102 rows x 9 columns]
```

SF

```
                        Name                     Team  Number Position
Age  \
1               Jae Crowder          Boston Celtics    99.0       SF
25.0
32   Thanasis Antetokounmpo         New York Knicks    43.0       SF
23.0
33          Carmelo Anthony         New York Knicks     7.0       SF
32.0
35         Cleanthony Early         New York Knicks    11.0       SF
25.0
42             Lance Thomas         New York Knicks    42.0       SF
28.0
..                      ...                     ...     ...      ...
...
428         Al-Farouq Aminu  Portland Trail Blazers     8.0       SF
25.0
432        Maurice Harkless  Portland Trail Blazers     4.0       SF
23.0
448          Gordon Hayward               Utah Jazz    20.0       SF
26.0
450              Joe Ingles               Utah Jazz     2.0       SF
28.0
451           Chris Johnson               Utah Jazz    23.0       SF
26.0

    Height  Weight        College      Salary
1      6-6   235.0      Marquette   6796117.0
32     6-7   205.0            NaN     30888.0
33     6-8   240.0       Syracuse  22875000.0
35     6-8   210.0  Wichita State    845059.0
42     6-8   235.0           Duke   1636842.0
..     ...     ...            ...         ...
428    6-9   215.0    Wake Forest   8042895.0
432    6-9   215.0     St. John's   2894059.0
448    6-8   226.0         Butler  15409570.0
450    6-8   226.0            NaN   2050000.0
451    6-6   206.0         Dayton    981348.0

[85 rows x 9 columns]
```

```
SF.isnull().sum()
```

```
Name          0
Team          0
Number        0
Position      0
Age           0
Height        0
Weight        0
College      14
```

```
Salary       0
dtype: int64
```

```
comb_df = pd.concat([PG, SF,C,SG])
```

```
comb_df
```

```
                  Name                     Team  Number Position   Age
Height  \
0       Avery Bradley          Boston Celtics     0.0        PG  25.0
6-2
8        Terry Rozier          Boston Celtics    12.0        PG  22.0
6-2
9        Marcus Smart          Boston Celtics    36.0        PG  22.0
6-4
11      Isaiah Thomas          Boston Celtics     4.0        PG  27.0
5-9
19       Jarrett Jack           Brooklyn Nets     2.0        PG  32.0
6-3
..                ...                     ...     ...       ...   ...
...
433  Gerald Henderson  Portland Trail Blazers     9.0        SG  28.0
6-5
437     C.J. McCollum  Portland Trail Blazers     3.0        SG  24.0
6-4
438      Luis Montero  Portland Trail Blazers    44.0        SG  23.0
6-7
444        Alec Burks               Utah Jazz    10.0        SG  24.0
6-6
449       Rodney Hood               Utah Jazz     5.0        SG  23.0
6-8

     Weight         College     Salary
0     180.0           Texas  7730337.0
8     190.0      Louisville  1824360.0
9     220.0  Oklahoma State  3431040.0
11    185.0      Washington  6912869.0
19    200.0    Georgia Tech  6300000.0
..      ...             ...        ...
433   215.0            Duke  6000000.0
437   200.0          Lehigh  2525160.0
438   185.0  Westchester CC   525093.0
444   214.0        Colorado  9463484.0
449   206.0            Duke  1348440.0

[357 rows x 9 columns]
```

```
comb_df.sort_index(axis = 0)
```

```
              Name            Team  Number Position   Age Height
Weight  \
```

```
0    Avery Bradley  Boston Celtics     0.0       PG  25.0    6-2
180.0
1      Jae Crowder  Boston Celtics    99.0       SF  25.0    6-6
235.0
2     John Holland  Boston Celtics    30.0       SG  27.0    6-5
205.0
3      R.J. Hunter  Boston Celtics    28.0       SG  22.0    6-5
185.0
7     Kelly Olynyk  Boston Celtics    41.0        C  25.0    7-0
238.0
..             ...             ...     ...      ...   ...    ...     .
..
451  Chris Johnson       Utah Jazz    23.0       SF  26.0    6-6
206.0
453   Shelvin Mack       Utah Jazz     8.0       PG  26.0    6-3
203.0
454      Raul Neto       Utah Jazz    25.0       PG  24.0    6-1
179.0
455   Tibor Pleiss       Utah Jazz    21.0        C  26.0    7-3
256.0
456    Jeff Withey       Utah Jazz    24.0        C  26.0    7-0
231.0

                College        Salary
0                 Texas  7.730337e+06
1             Marquette  6.796117e+06
2    Boston University  5.077829e+06
3        Georgia State  1.148640e+06
7              Gonzaga  2.165160e+06
..                 ...           ...
451             Dayton  9.813480e+05
453             Butler  2.433333e+06
454                NaN  9.000000e+05
455                NaN  2.900000e+06
456             Kansas  9.472760e+05

[357 rows x 9 columns]
```