In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df=pd.read_csv("nba.csv")
```

In [3]:
```
df
```

Out[3]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99.0	SF	25.0	6-6	235.0	Marquette	6796117.0
2	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28.0	SG	22.0	6-5	185.0	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8.0	PG	26.0	6-3	203.0	Butler	2433333.0
454	Raul Neto	Utah Jazz	25.0	PG	24.0	6-1	179.0	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21.0	C	26.0	7-3	256.0	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24.0	C	26.0	7-0	231.0	Kansas	947276.0
457	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

458 rows × 9 columns

In [4]:
```
df.head()
```

Out[4]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99.0	SF	25.0	6-6	235.0	Marquette	6796117.0
2	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28.0	SG	22.0	6-5	185.0	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0

In [5]:
```
df.tail()
```

Out[5]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
453	Shelvin Mack	Utah Jazz	8.0	PG	26.0	6-3	203.0	Butler	2433333.0
454	Raul Neto	Utah Jazz	25.0	PG	24.0	6-1	179.0	NaN	900000.0
455	Tibor Pleiss	Utah Jazz	21.0	C	26.0	7-3	256.0	NaN	2900000.0
456	Jeff Withey	Utah Jazz	24.0	C	26.0	7-0	231.0	Kansas	947276.0
457	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

In [6]:
```
df.describe()
```

Out[6]:

	Number	Age	Weight	Salary
count	457.000000	457.000000	457.000000	4.460000e+02
mean	17.678337	26.938731	221.522976	4.842684e+06
std	15.966090	4.404016	26.368343	5.229238e+06
min	0.000000	19.000000	161.000000	3.088800e+04
25%	5.000000	24.000000	200.000000	1.044792e+06
50%	13.000000	26.000000	220.000000	2.839073e+06
75%	25.000000	30.000000	240.000000	6.500000e+06
max	99.000000	40.000000	307.000000	2.500000e+07

In [7]:
```
df.dtypes
```

Out[7]:
```
Name          object
Team          object
Number       float64
Position      object
Age          float64
Height        object
Weight       float64
College       object
Salary       float64
dtype: object
```

In [8]:
```
df.size
```

Out[8]: 4122

In [9]:
```
df.shape
```

Out[9]: (458, 9)

In [17]:
```
#dividing into groups
#groupby() is used
groupheight=df.groupby(df['Height'])
```

In [19]:
```
groupheight.get_group('5-11')
```

Out[19]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
22	Shane Larkin	Brooklyn Nets	0.0	PG	23.0	5-11	175.0	Miami (FL)	1500000.0
130	Phil Pressey	Phoenix Suns	25.0	PG	25.0	5-11	175.0	Missouri	55722.0
203	Ty Lawson	Indiana Pacers	10.0	PG	28.0	5-11	195.0	North Carolina	211744.0

In [23]:
```
df.Age.value_counts()
```

Out[23]:
```
24.0    47
25.0    45
27.0    41
23.0    41
26.0    36
28.0    31
30.0    31
29.0    28
22.0    26
31.0    22
20.0    19
21.0    19
33.0    14
32.0    13
34.0    10
36.0    10
35.0     9
37.0     4
38.0     4
40.0     3
39.0     2
19.0     2
Name: Age, dtype: int64
```

In [24]:
```
#creating bins randomly select some sample points
bins=[19,25,31,36,40]
labels=['19-24','25-30','31-35','37-40'] #label fot new group
df['AgeGroup']=pd.cut(df['Age'],bins=bins,labels=labels,right=False) #c
```

In [25]:

Out[25]:

	Name	Team	Number	Position	Age	Height	Weight	College	Salary	AgeGroup
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0	25-30
1	Jae Crowder	Boston Celtics	99.0	SF	25.0	6-6	235.0	Marquette	6796117.0	25-30
2	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN	25-30
3	R.J. Hunter	Boston Celtics	28.0	SG	22.0	6-5	185.0	Georgia State	1148640.0	19-24
4	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0	25-30
...	...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8.0	PG	26.0	6-3	203.0	Butler	2433333.0	25-30
454	Raul Neto	Utah Jazz	25.0	PG	24.0	6-1	179.0	NaN	900000.0	19-24
455	Tibor Pleiss	Utah Jazz	21.0	C	26.0	7-3	256.0	NaN	2900000.0	25-30
456	Jeff Withey	Utah Jazz	24.0	C	26.0	7-0	231.0	Kansas	947276.0	25-30
457	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

458 rows × 10 columns

In [26]:
```
df.groupby('AgeGroup')['Salary'].mean()  #mean salary by age group
```

Out[26]:
```
AgeGroup
19-24    2.706212e+06
25-30    5.895838e+06
31-35    6.662026e+06
37-40    3.847918e+06
Name: Salary, dtype: float64
```

In [27]:
```
df.groupby('AgeGroup')['Salary'].describe()
```

Out[27]:

	count	mean	std	min	25%	50%	75%	
AgeGroup								
19-24	150.0	2.706212e+06	3.188468e+06	30888.0	960457.0	1654380.0	3095650.0	1640
25-30	207.0	5.895838e+06	5.535586e+06	55722.0	1100602.0	4000000.0	9106741.5	2235
31-35	67.0	6.662026e+06	6.281446e+06	200600.0	2092835.5	4500000.0	9697402.5	2297
37-40	19.0	3.847918e+06	5.516191e+06	222888.0	947276.0	2814000.0	3815259.5	2500

In [28]:
```
salarylist=list(df.groupby('AgeGroup')['Salary'])
```

In [29]:
```
salarylist
```

Out[29]:
```
[('19-24',
  3       1148640.0
  6       1170960.0
  8       1824360.0
  9       3431040.0
  10      2569260.0
            ...
  446    12000000.0
  447     1175880.0
  449     1348440.0
  452     2239800.0
  454      900000.0
  Name: Salary, Length: 154, dtype: float64),
 ('25-30',
  0       7730337.0
  1       6796117.0
  2             NaN
  4       5000000.0
  5      12000000.0
            ...
  450     2050000.0
  451      981348.0
  453     2433333.0
  455     2900000.0
  456      947276.0
  Name: Salary, Length: 212, dtype: float64),
 ('31-35',
  19      6300000.0
  31      1635476.0
  33     22875000.0
  34      7402812.0
  43       947276.0
            ...
  375    13000000.0
  394     4345000.0
  413     3750000.0
  415     3135000.0
  434     5016000.0
  Name: Salary, Length: 68, dtype: float64),
 ('37-40',
  46            NaN
  72      2900000.0
  93      5675000.0
  101     3376000.0
  102      947726.0
  109    25000000.0
  119      947276.0
  139     1449187.0
  183     2170465.0
  236     8333334.0
  256      947276.0
  259     5000000.0
  260     3542500.0
  261     4088019.0
  296      947276.0
  299     2814000.0
  343     2854940.0
```

```
392      947276.0
406      947276.0
420      222888.0
Name: Salary, dtype: float64)]
```

In [ ]: