

In [1]:

```
import pandas as pd
```

Find mean, median, mode and describe

In [5]:

```
df1=pd.read_csv("E:/Datasets/2_2015.csv")
df2=pd.read_csv("E:/Datasets/3_Fitness-1.csv")
df3=pd.read_csv("E:/Datasets/4_drug200.csv")
df4=pd.read_csv("E:/Datasets/8_BreastCancerPrediction.csv")
df5=pd.read_csv("E:/Datasets/6_Salesworkload1.csv")
```

In [7]:

```
print(df1.describe())
print(df2.describe())
print(df3.describe())
print(df4.describe())
print(df5.describe())
```

	Happiness Rank	Happiness Score	Standard Error	\
count	158.000000	158.000000	158.000000	
mean	79.493671	5.375734	0.047885	
std	45.754363	1.145010	0.017146	
min	1.000000	2.839000	0.018480	
25%	40.250000	4.526000	0.037268	
50%	79.500000	5.232500	0.043940	
75%	118.750000	6.243750	0.052300	
max	158.000000	7.587000	0.136930	

	Economy (GDP per Capita)	Family	Health (Life Expectancy)
\			
count	158.000000	158.000000	158.000000
mean	0.846137	0.991046	0.630259
std	0.403121	0.272369	0.247078
min	0.000000	0.000000	0.000000
25%	0.545808	0.856823	0.439185
50%	0.910245	1.029510	0.696705
75%	1.158448	1.214405	0.811013
max	1.600000	1.400000	1.000000

Mean

In [8]:

```
print(df1["Happiness Rank"].mean())  
print(df2["Sum of Total Sales"].mean())  
print(df3["Na_to_K"].mean())  
print(df4["perimeter_mean"].mean())  
print(df5["Sales units"].mean())
```

```
79.49367088607595  
255.55555555555554  
16.084485  
91.96903339191564  
1076470.714379085
```

Median

In [10]:

```
print(df1["Happiness Rank"].median())  
print(df2["Sum of Total Sales"].median())  
print(df3["Na_to_K"].median())  
print(df4["perimeter_mean"].median())  
print(df5["Sales units"].median())
```

```
79.5  
167.0  
13.9365  
86.24  
293230.0
```

Mode

In [11]:

```
print(df1["Happiness Rank"].mode())
print(df2["Sum of Total Sales"].mode())
print(df3["Na_to_K"].mode())
print(df4["perimeter_mean"].mode())
print(df5["Sales units"].mode())
```

```
0    82
Name: Happiness Rank, dtype: int64
0     75
1    101
2    127
3    160
4    167
5    170
6    171
7    179
8   1150
Name: Sum of Total Sales, dtype: int64
0    12.006
1    18.295
Name: Na_to_K, dtype: float64
0     82.61
1     87.76
2    134.70
Name: perimeter_mean, dtype: float64
0     0.0
Name: Sales units, dtype: float64
```

Find sum(), cumsum(), count, min and max values

sum()

In [12]:

```
print(df1["Happiness Rank"].sum())
print(df2["Sum of Total Sales"].sum())
print(df3["Na_to_K"].sum())
print(df4["perimeter_mean"].sum())
print(df5["Sales units"].sum())
```

```
12560
2300
3216.897
52330.380000000005
8235000965.0
```

cumsum()

In [13]:

```
print(df1["Happiness Rank"].cumsum())  
print(df2["Sum of Total Sales"].cumsum())  
print(df3["Na_to_K"].cumsum())  
print(df4["perimeter_mean"].cumsum())  
print(df5["Sales units"].cumsum())
```

0	1
1	3
2	6
3	10
4	15

...

153	11934
154	12089
155	12245
156	12402
157	12560

Name: Happiness Rank, Length: 158, dtype: int64

0	75
1	235
2	336
3	463
4	642
5	809
6	980
7	1150
8	2300

Name: Sum of Total Sales, dtype: int64

0	25.355
1	38.448
2	48.562
3	56.360
4	74.403

...

195	3169.628
196	3181.634
197	3191.528
198	3205.548
199	3216.897

Name: Na_to_K, Length: 200, dtype: float64

0	122.80
1	255.70
2	385.70
3	463.28
4	598.38

...

564	51902.86
565	52034.06
566	52142.36
567	52282.46
568	52330.38

Name: perimeter_mean, Length: 569, dtype: float64

0	3.985600e+05
1	4.812850e+05
2	9.196850e+05
3	1.229110e+06
4	1.394625e+06

...

7653	8.231114e+09
7654	8.231114e+09
7655	8.231114e+09
7656	8.231114e+09
7657	8.235001e+09

Name: Sales units, Length: 7658, dtype: float64

Count()

In [14]:

```
print(df1["Happiness Rank"].count())
print(df2["Sum of Total Sales"].count())
print(df3["Na_to_K"].count())
print(df4["perimeter_mean"].count())
print(df5["Sales units"].count())
```

```
158
9
200
569
7650
```

min()

In [16]:

```
print(df1["Happiness Rank"].min())
print(df2["Sum of Total Sales"].min())
print(df3["Na_to_K"].min())
print(df4["perimeter_mean"].min())
print(df5["Sales units"].min())
```

```
1
75
6.269
43.79
0.0
```

max()

In [17]:

```
print(df1["Happiness Rank"].max())
print(df2["Sum of Total Sales"].max())
print(df3["Na_to_K"].max())
print(df4["perimeter_mean"].max())
print(df5["Sales units"].max())
```

```
158
1150
38.247
188.5
11242955.0
```

Find covariance and correlation (spearman and pearsons)

In [20]:

```
import scipy
import numpy
from scipy.stats import pearsonr
from scipy.stats import spearmanr
from numpy import cov
```

In [26]:

```
print(cov(df1["Happiness Rank"],df1["Happiness Score"]))
#print(cov(df2["Sum of Jan"],df2["Sum of Feb"]))
print(cov(df3["Age"],df3["Na_to_K"]))
print(cov(df4["perimeter_mean"],df4["area_mean"]))
print(cov(df5["Sales units"],df5["Turnover"]))
```

```
[[ 2.09346174e+03 -5.19756132e+01]
 [-5.19756132e+01  1.31104821e+00]]
[[273.71434673 -7.54375153]
 [-7.54375153  52.18553348]]
[[ 590.44047952  8435.77234508]
 [ 8435.77234508 123843.55431768]]
[[nan nan]
 [nan nan]]
```

In [31]:

```
print(pearsonr(df1["Happiness Rank"],df1["Happiness Score"]))
#print(pearsonr(df2["Sum of Jan"],df2["Sum of Feb"]))
print(pearsonr(df3["Age"],df3["Na_to_K"]))
print(pearsonr(df4["perimeter_mean"],df4["area_mean"]))
#print(pearsonr(df5["Sales units"],df5["Turnover"]))
```

```
PearsonRResult(statistic=-0.9921053148284925, pvalue=1.4013759581556859e-1
42)
PearsonRResult(statistic=-0.0631194972677259, pvalue=0.37457563990343007)
PearsonRResult(statistic=0.9865068039913898, pvalue=0.0)
```

In [29]:

```
print(spearmanr(df1["Happiness Rank"],df1["Happiness Score"]))
print(spearmanr(df2["Sum of Jan"],df2["Sum of Feb"]))
print(spearmanr(df3["Age"],df3["Na_to_K"]))
print(spearmanr(df4["perimeter_mean"],df4["area_mean"]))
print(spearmanr(df5["Sales units"],df5["Turnover"]))
```

```
SignificanceResult(statistic=-0.9999999999999999, pvalue=0.0)
SignificanceResult(statistic=-0.049999999999999996, pvalue=0.8983528043506
301)
SignificanceResult(statistic=-0.047273882688479915, pvalue=0.5062200581387
418)
SignificanceResult(statistic=0.9970682695182411, pvalue=0.0)
SignificanceResult(statistic=nan, pvalue=nan)
```

In []: