In [1]:
```
import csv
```

In [2]:
```
# test csv reader and peek the dataset
with open('db_lec4.txt', newline='') as csvfile:
    reader = csv.reader(csvfile, delimiter='\t')
    for row in reader:
        print(",".join(row))
```

```
Pt,BP,Age,Weight,BSA,Dur,Pulse,Stress
1,105,47,85.4,1.75,5.1,63,33
2,115,49,94.2,2.10,3.8,70,14
3,116,49,95.3,1.98,8.2,72,10
4,117,50,94.7,2.01,5.8,73,99
5,112,51,89.4,1.89,7.0,72,95
6,121,48,99.5,2.25,9.3,71,10
7,121,49,99.8,2.25,2.5,69,42
8,110,47,90.9,1.90,6.2,66,8
9,110,49,89.2,1.83,7.1,69,62
10,114,48,92.7,2.07,5.6,64,35
11,114,47,94.4,2.07,5.3,74,90
12,115,49,94.1,1.98,5.6,71,21
13,114,50,91.6,2.05,10.2,68,47
14,106,45,87.1,1.92,5.6,67,80
15,125,52,101.3,2.19,10.0,76,98
16,114,46,94.5,1.98,7.4,69,95
17,106,46,87.0,1.87,3.6,62,18
18,113,46,94.5,1.90,4.3,70,12
19,110,48,90.5,1.88,9.0,71,99
20,122,56,95.7,2.09,7.0,75,99
```

In [3]:
```
import pandas as pd
import numpy as np
```

In [4]:
```
df = pd.read_csv('db_lec4.txt', delimiter='\t')
df
```

Out[4]:

	Pt	BP	Age	Weight	BSA	Dur	Pulse	Stress
0	1	105	47	85.4	1.75	5.1	63	33
1	2	115	49	94.2	2.10	3.8	70	14
2	3	116	49	95.3	1.98	8.2	72	10
3	4	117	50	94.7	2.01	5.8	73	99
4	5	112	51	89.4	1.89	7.0	72	95
5	6	121	48	99.5	2.25	9.3	71	10
6	7	121	49	99.8	2.25	2.5	69	42
7	8	110	47	90.9	1.90	6.2	66	8
8	9	110	49	89.2	1.83	7.1	69	62
9	10	114	48	92.7	2.07	5.6	64	35
10	11	114	47	94.4	2.07	5.3	74	90
11	12	115	49	94.1	1.98	5.6	71	21
12	13	114	50	91.6	2.05	10.2	68	47
13	14	106	45	87.1	1.92	5.6	67	80
14	15	125	52	101.3	2.19	10.0	76	98
15	16	114	46	94.5	1.98	7.4	69	95
16	17	106	46	87.0	1.87	3.6	62	18
17	18	113	46	94.5	1.90	4.3	70	12
18	19	110	48	90.5	1.88	9.0	71	99
19	20	122	56	95.7	2.09	7.0	75	99

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

Out[5]:

	Pt	BP	Age	Weight	BSA	Dur	Pulse	Stress
Pt	1.000000	0.031135	0.042694	0.024857	-0.031288	0.176246	0.112285	0.343152
BP	0.031135	1.000000	0.659093	0.950068	0.865879	0.292834	0.721413	0.163901
Age	0.042694	0.659093	1.000000	0.407349	0.378455	0.343792	0.618764	0.368224
Weight	0.024857	0.950068	0.407349	1.000000	0.875305	0.200650	0.659340	0.034355
BSA	-0.031288	0.865879	0.378455	0.875305	1.000000	0.130540	0.464819	0.018446
Dur	0.176246	0.292834	0.343792	0.200650	0.130540	1.000000	0.401514	0.311640
Pulse	0.112285	0.721413	0.618764	0.659340	0.464819	0.401514	1.000000	0.506310
Stress	0.343152	0.163901	0.368224	0.034355	0.018446	0.311640	0.506310	1.000000

In [6]:
```
corr = df.corr()
corr.style.background_gradient(cmap='coolwarm')
```

Out[6]:

	Pt	BP	Age	Weight	BSA	Dur	Pulse	Stress
Pt	1.000000	0.031135	0.042694	0.024857	-0.031288	0.176246	0.112285	0.343152
BP	0.031135	1.000000	0.659093	0.950068	0.865879	0.292834	0.721413	0.163901
Age	0.042694	0.659093	1.000000	0.407349	0.378455	0.343792	0.618764	0.368224
Weight	0.024857	0.950068	0.407349	1.000000	0.875305	0.200650	0.659340	0.034355
BSA	-0.031288	0.865879	0.378455	0.875305	1.000000	0.130540	0.464819	0.018446
Dur	0.176246	0.292834	0.343792	0.200650	0.130540	1.000000	0.401514	0.311640
Pulse	0.112285	0.721413	0.618764	0.659340	0.464819	0.401514	1.000000	0.506310
Stress	0.343152	0.163901	0.368224	0.034355	0.018446	0.311640	0.506310	1.000000

In [9]:
```
from statsmodels.stats.outliers_influence import variance_inflation_factor
```

In [14]:
```
vif_data = pd.DataFrame()
```

In [15]:
```
vif_data["feature"] = df.columns
```

In [17]:
```
vif_data["VIF"] = [variance_inflation_factor(df.values, i)
                   for i in range(len(df.columns))]
```

In [18]:
```
print(vif_data)
```

```
  feature           VIF
0      Pt      5.304970
1      BP  41936.549706
2     Age   3112.352730
3  Weight  22617.025732
4     BSA   1617.891583
5     Dur     14.680941
6   Pulse   1691.763967
7  Stress      7.512732
```

In [27]:
```
lba = np.linalg.eig(corr)[0]
```

In [28]:
```
mcn = max(lba) / min(lba)
mcn
```

Out[28]:
```
1818.2344678567126
```

Since MCN is large than 1000, so the dataset have very series issue.

In [ ]:
```

```