

3

April 15, 2024

```
[1]: import pandas as pd
```

```
[2]: import numpy as np
```

```
[3]: data = pd.read_csv('toy_dataset.csv')
```

```
[4]: data.head()
```

```
[4]:
```

	Number	City	Gender	Age	Income	Illness
0	1	Dallas	Male	41	40367.0	No
1	2	Dallas	Male	54	45084.0	No
2	3	Dallas	Male	42	52483.0	No
3	4	Dallas	Male	40	40941.0	No
4	5	Dallas	Male	46	50289.0	No

```
[5]: data.tail()
```

```
[5]:
```

	Number	City	Gender	Age	Income	Illness
149995	149996	Austin	Male	48	93669.0	No
149996	149997	Austin	Male	25	96748.0	No
149997	149998	Austin	Male	26	111885.0	No
149998	149999	Austin	Male	25	111878.0	No
149999	150000	Austin	Female	37	87251.0	No

```
[6]: data.columns
```

```
[6]: Index(['Number', 'City', 'Gender', 'Age', 'Income', 'Illness'], dtype='object')
```

```
[7]: data1 = data.iloc[0:51, 3:5]
```

```
[8]: data1
```

```
[8]:
```

	Age	Income
0	41	40367.0
1	54	45084.0
2	42	52483.0
3	40	40941.0

4	46	50289.0
5	36	50786.0
6	32	33155.0
7	39	30914.0
8	51	68667.0
9	30	50082.0
10	48	41524.0
11	47	54777.0
12	46	62749.0
13	42	50894.0
14	61	38429.0
15	43	34074.0
16	27	50398.0
17	38	46373.0
18	47	51137.0
19	35	23688.0
20	57	17378.0
21	33	45919.0
22	33	23001.0
23	27	34292.0
24	58	55190.0
25	64	26169.0
26	58	57322.0
27	44	61704.0
28	34	53619.0
29	45	47421.0
30	44	40353.0
31	39	28125.0
32	55	42630.0
33	27	56645.0
34	63	41946.0
35	41	50312.0
36	64	47872.0
37	41	29538.0
38	61	39881.0
39	59	48518.0
40	26	16168.0
41	41	68522.0
42	47	50750.0
43	58	49614.0
44	33	56169.0
45	30	40661.0
46	51	53730.0
47	45	34613.0
48	38	35249.0
49	56	52218.0
50	55	47702.0

```
[9]: data1.mean()
```

```
[9]: Age          44.549020  
     Income      44510.627451  
     dtype: float64
```

```
[10]: data1.median()
```

```
[10]: Age          44.0  
     Income      47421.0  
     dtype: float64
```

```
[11]: data1.min()
```

```
[11]: Age          26.0  
     Income      16168.0  
     dtype: float64
```

```
[12]: data1.max()
```

```
[12]: Age          64.0  
     Income      68667.0  
     dtype: float64
```

```
[13]: data1.std()
```

```
[13]: Age          10.826474  
     Income      12028.903774  
     dtype: float64
```

```
[14]: data1.var()
```

```
[14]: Age          1.172125e+02  
     Income      1.446945e+08  
     dtype: float64
```

```
[18]: import pandas as pd
```

```
[19]: pwd
```

```
[19]: 'C:\\Users\\Tej\\Downloads'
```

```
[20]: cd E:/
```

```
E:\
```

```
[21]: data1 = pd.read_csv('iris.csv')
```

```
[22]: data1.head()
```

```
[22]:   sepallength  sepalwidth  petallength  petalwidth      class
0         5.1         3.5         1.4         0.2  Iris-setosa
1         4.9         3.0         1.4         0.2  Iris-setosa
2         4.7         3.2         1.3         0.2  Iris-setosa
3         4.6         3.1         1.5         0.2  Iris-setosa
4         5.0         3.6         1.4         0.2  Iris-setosa
```

```
[23]: setosa = data1['class'] == 'Iris-setosa'
```

```
[24]: print(data1[setosa].describe())
```

	sepallength	sepalwidth	petallength	petalwidth
count	50.00000	50.000000	50.000000	50.00000
mean	5.00600	3.418000	1.464000	0.24400
std	0.35249	0.381024	0.173511	0.10721
min	4.30000	2.300000	1.000000	0.10000
25%	4.80000	3.125000	1.400000	0.20000
50%	5.00000	3.400000	1.500000	0.20000
75%	5.20000	3.675000	1.575000	0.30000
max	5.80000	4.400000	1.900000	0.60000

```
[25]: versicolor = data1['class'] == 'Iris-versicolor'
```

```
[26]: print(data1[versicolor].describe())
```

	sepallength	sepalwidth	petallength	petalwidth
count	50.000000	50.000000	50.000000	50.000000
mean	5.936000	2.770000	4.260000	1.326000
std	0.516171	0.313798	0.469911	0.197753
min	4.900000	2.000000	3.000000	1.000000
25%	5.600000	2.525000	4.000000	1.200000
50%	5.900000	2.800000	4.350000	1.300000
75%	6.300000	3.000000	4.600000	1.500000
max	7.000000	3.400000	5.100000	1.800000

```
[27]: virginica = data1['class'] == 'Iris-virginica'
```

```
[28]: print(data1[virginica].describe())
```

	sepallength	sepalwidth	petallength	petalwidth
count	50.00000	50.000000	50.000000	50.00000
mean	6.58800	2.974000	5.552000	2.02600
std	0.63588	0.322497	0.551895	0.27465
min	4.90000	2.200000	4.500000	1.40000
25%	6.22500	2.800000	5.100000	1.80000
50%	6.50000	3.000000	5.550000	2.00000

75%	6.90000	3.175000	5.875000	2.30000
max	7.90000	3.800000	6.900000	2.50000

```
[29]: setosa.mean()
```

```
[29]: 0.3333333333333333
```

```
[30]: versicolor.mean()
```

```
[30]: 0.3333333333333333
```

```
[31]: virginica.mean()
```

```
[31]: 0.3333333333333333
```

```
[32]: setosa.std()
```

```
[32]: 0.4729837698404022
```

```
[33]: versicolor.std()
```

```
[33]: 0.47298376984040214
```

```
[34]: virginica.std()
```

```
[34]: 0.4729837698404021
```