

```
In [2]: import numpy as np
import pandas as pd
import matplotlib as matp
import seaborn as sns
%matplotlib inline
```

```
In [3]: iris = pd.read_csv("IRIS.csv")
```

```
In [5]: iris.shape
```

```
Out[5]: (150, 5)
```

```
In [6]: iris.head()
```

```
Out[6]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
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

```
In [7]: iris.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   sepal_length    150 non-null    float64
1   sepal_width     150 non-null    float64
2   petal_length    150 non-null    float64
3   petal_width     150 non-null    float64
4   species         150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

```
In [8]: iris.describe()
```

```
Out[8]:
```

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
In [9]: iris.groupby('species').sepal_length.agg(['min', 'max', 'mean', 'median', 'std'])
```

Out[9]:

	min	max	mean	median	std
species					
Iris-setosa	4.3	5.8	5.006	5.0	0.352490
Iris-versicolor	4.9	7.0	5.936	5.9	0.516171
Iris-virginica	4.9	7.9	6.588	6.5	0.635880

```
In [10]: iris.groupby('species').sepal_width.agg(['min', 'max', 'mean', 'median', 'std'])
```

Out[10]:

	min	max	mean	median	std
species					
Iris-setosa	2.3	4.4	3.418	3.4	0.381024
Iris-versicolor	2.0	3.4	2.770	2.8	0.313798
Iris-virginica	2.2	3.8	2.974	3.0	0.322497

```
In [11]: iris.groupby('species').petal_length.agg(['min', 'max', 'mean', 'median', 'std'])
```

Out[11]:

	min	max	mean	median	std
species					
Iris-setosa	1.0	1.9	1.464	1.50	0.173511
Iris-versicolor	3.0	5.1	4.260	4.35	0.469911
Iris-virginica	4.5	6.9	5.552	5.55	0.551895

```
In [12]: iris.groupby('species').petal_width.agg(['min', 'max', 'mean', 'median', 'std'])
```

Out[12]:

	min	max	mean	median	std
species					
Iris-setosa	0.1	0.6	0.244	0.2	0.107210
Iris-versicolor	1.0	1.8	1.326	1.3	0.197753
Iris-virginica	1.4	2.5	2.026	2.0	0.274650

```
In [13]: iris.groupby('species').sepal_length.agg(['min', 'max', 'mean', 'median', 'std', 'var'])
```

```
Out[13]:
```

	min	max	mean	median	std	var
species						
Iris-setosa	4.3	5.8	5.006	5.0	0.352490	0.124249
Iris-versicolor	4.9	7.0	5.936	5.9	0.516171	0.266433
Iris-virginica	4.9	7.9	6.588	6.5	0.635880	0.404343

```
In [14]: iris.groupby('species').sepal_length.min()
```

```
Out[14]: species
Iris-setosa      4.3
Iris-versicolor  4.9
Iris-virginica   4.9
Name: sepal_length, dtype: float64
```

```
In [15]: iris.groupby('species').sepal_length.max()
```

```
Out[15]: species
Iris-setosa      5.8
Iris-versicolor  7.0
Iris-virginica   7.9
Name: sepal_length, dtype: float64
```

```
In [16]: iris.groupby('species').sepal_length.mean()
```

```
Out[16]: species
Iris-setosa      5.006
Iris-versicolor  5.936
Iris-virginica   6.588
Name: sepal_length, dtype: float64
```

```
In [17]: iris.groupby('species').sepal_length.median()
```

```
Out[17]: species
Iris-setosa      5.0
Iris-versicolor  5.9
Iris-virginica   6.5
Name: sepal_length, dtype: float64
```

```
In [19]: iris.groupby('species').sepal_length.std()
```

```
Out[19]: species
Iris-setosa      0.352490
Iris-versicolor  0.516171
Iris-virginica   0.635880
Name: sepal_length, dtype: float64
```

```
In [20]: iris.groupby('species').sepal_length.var()
```

```
Out[20]: species
Iris-setosa      0.124249
Iris-versicolor  0.266433
Iris-virginica   0.404343
Name: sepal_length, dtype: float64
```

In []:

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib as matp
%matplotlib inline
```

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

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

Out[3]:

	YearsExperience	Age	Salary
0	1.1	21.0	39343
1	1.3	21.5	46205
2	1.5	21.7	37731
3	2.0	22.0	43525
4	2.2	22.2	39891
5	2.9	23.0	56642
6	3.0	23.0	60150
7	3.2	23.3	54445
8	3.2	23.3	64445
9	3.7	23.6	57189
10	3.9	23.9	63218
11	4.0	24.0	55794
12	4.0	24.0	56957
13	4.1	24.0	57081
14	4.5	25.0	61111
15	4.9	25.0	67938
16	5.1	26.0	66029
17	5.3	27.0	83088
18	5.9	28.0	81363
19	6.0	29.0	93940
20	6.8	30.0	91738
21	7.1	30.0	98273
22	7.9	31.0	101302
23	8.2	32.0	113812
24	8.7	33.0	109431
25	9.0	34.0	105582
26	9.5	35.0	116969
27	9.6	36.0	112635
28	10.3	37.0	122391
29	10.5	38.0	121872

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

Out[4]:

	YearsExperience	Age	Salary
count	30.000000	30.000000	30.000000
mean	5.313333	27.216667	76003.000000
std	2.837888	5.161267	27414.429785
min	1.100000	21.000000	37731.000000
25%	3.200000	23.300000	56720.750000
50%	4.700000	25.000000	65237.000000
75%	7.700000	30.750000	100544.750000
max	10.500000	38.000000	122391.000000

```
In [5]: print(df['Salary'].mean())
```

76003.0

```
In [6]: print(df['Salary'].mode())
```

```
0      37731
1      39343
2      39891
3      43525
4      46205
5      54445
6      55794
7      56642
8      56957
9      57081
10     57189
11     60150
12     61111
13     63218
14     64445
15     66029
16     67938
17     81363
18     83088
19     91738
20     93940
21     98273
22    101302
23    105582
24    109431
25    112635
26    113812
27    116969
28    121872
29    122391
```

Name: Salary, dtype: int64

```
In [7]: print(df['Salary'].median())
```

65237.0

```
In [8]: print("Standard deviation of Salary is - ", df['Salary'].std())
```

Standard deviation of Salary is - 27414.4297845823

```
In [ ]:
```

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js