

```
In [4]: import pandas as pd
import numpy as np
df=pd.read_csv("/home/ml-lab/Iris.csv")
df.head(5)
```

```
Out[4]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

```
In [5]: df.loc[:, "SepalLengthCm"].mean()
```

```
Out[5]: 5.843333333333335
```

```
In [6]: df.loc[:, "SepalLengthCm"].median()
```

```
Out[6]: 5.8
```

```
In [7]: df.loc[:, "SepalLengthCm"].mode()
```

```
Out[7]: 0    5.0
Name: SepalLengthCm, dtype: float64
```

```
In [8]: df.loc[:, "SepalLengthCm"].std()
```

```
Out[8]: 0.8280661279778629
```

```
In [9]: df.loc[:, "SepalLengthCm"].min()
```

```
Out[9]: 4.3
```

```
In [10]: df.loc[:, "SepalLengthCm"].max()
```

```
Out[10]: 7.9
```

```
In [11]: df.loc[:, "SepalLengthCm"].var()
```

```
Out[11]: 0.6856935123042505
```

```
In [12]: df.loc[:, "SepalLengthCm"].count()
```

```
Out[12]: 150
```

```
In [13]: df.loc[:, "SepalWidthCm"].mean()
```

```
Out[13]: 3.0540000000000007
```

```
In [14]: df.loc[:, "SepalWidthCm"].median()
```

```
Out[14]: 3.0
```

```
In [15]: df.loc[:, "SepalWidthCm"].mode()
```

```
Out[15]: 0      3.0  
         Name: SepalWidthCm, dtype: float64
```

```
In [16]: df.loc[:, "SepalWidthCm"].min()
```

```
Out[16]: 2.0
```

```
In [17]: df.loc[:, "SepalWidthCm"].max()
```

```
Out[17]: 4.4
```

```
In [18]: df.loc[:, "SepalWidthCm"].std()
```

```
Out[18]: 0.4335943113621737
```

```
In [19]: df.loc[:, "SepalWidthCm"].var()
```

```
Out[19]: 0.18800402684563763
```

```
In [20]: df.loc[:, "SepalWidthCm"].count()
```

```
Out[20]: 150
```

```
In [21]: df.loc[:, "PetalLengthCm"].mean()
```

```
Out[21]: 3.75866666666666693
```

```
In [22]: df.loc[:, "PetalLengthCm"].median()
```

```
Out[22]: 4.35
```

```
In [23]: df.loc[:, "PetalLengthCm"].mode()
```

```
Out[23]: 0      1.5  
         Name: PetalLengthCm, dtype: float64
```

```
In [24]: df.loc[:, "PetalLengthCm"].std()
```

```
Out[24]: 1.7644204199522617
```

```
In [25]: df.loc[:, "PetalLengthCm"].min()
```

```
Out[25]: 1.0
```

```
In [26]: df.loc[:, "PetalLengthCm"].max()
```

```
Out[26]: 6.9
```

```
In [27]: df.loc[:, "PetalLengthCm"].var()
```

```
Out[27]: 3.1131794183445156
```

```
In [28]: df.loc[:, "PetalLengthCm"].count()
```

```
Out[28]: 150
```

```
In [29]: df.loc[:, "PetalWidthCm"].mean()
```

```
Out[29]: 1.1986666666666672
```

```
In [30]: df.loc[:, "PetalWidthCm"].median()
```

```
Out[30]: 1.3
```

```
In [31]: df.loc[:, "PetalWidthCm"].mode()
```

```
Out[31]: 0    0.2  
         Name: PetalWidthCm, dtype: float64
```

```
In [32]: df.loc[:, "PetalWidthCm"].std()
```

```
Out[32]: 0.7631607417008414
```

```
In [34]: df.loc[:, "PetalWidthCm"].min()
```

```
Out[34]: 0.1
```

```
In [35]: df.loc[:, "PetalWidthCm"].max()
```

```
Out[35]: 2.5
```

```
In [36]: df.loc[:, "PetalWidthCm"].var()
```

```
Out[36]: 0.5824143176733784
```

```
In [37]: df.loc[:, "PetalWidthCm"].count()
```

```
Out[37]: 150
```

```
In [38]: df.describe()
```

```
Out[38]:
```

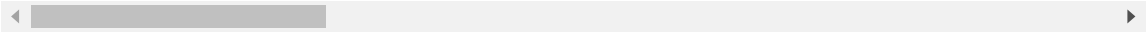
	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

```
In [52]: import pandas as pd
import numpy as np
df1=pd.read_csv("/home/ml-lab/Downloads/HR_Analytics.csv")
df1
```

Out[52]:

	EmpID	Age	AgeGroup	Attrition	BusinessTravel	DailyRate	Department	Distance
0	RM297	18	18-25	Yes	Travel_Rarely	230	Research & Development	
1	RM302	18	18-25	No	Travel_Rarely	812	Sales	
2	RM458	18	18-25	Yes	Travel_Frequently	1306	Sales	
3	RM728	18	18-25	No	Non-Travel	287	Research & Development	
4	RM829	18	18-25	Yes	Non-Travel	247	Research & Development	
...	...	...	...	...	...	...	...	...
1475	RM412	60	55+	No	Travel_Rarely	422	Research & Development	
1476	RM428	60	55+	No	Travel_Frequently	1499	Sales	
1477	RM537	60	55+	No	Travel_Rarely	1179	Sales	
1478	RM880	60	55+	No	Travel_Rarely	696	Sales	
1479	RM1210	60	55+	No	Travel_Rarely	370	Research & Development	

1480 rows × 38 columns



```
In [53]: df1["BusinessTravel"].replace({"Travel_Rarely":1,"Travel_Frequently":0})
df1["Attrition"].replace({"Yes":1,"No":0},inplace=True)
df1
```

Out[53]:

	EmpID	Age	AgeGroup	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome
0	RM297	18	18-25	1	1	230	Research & Development	1
1	RM302	18	18-25	0	1	812	Sales	1
2	RM458	18	18-25	1	0	1306	Sales	1
3	RM728	18	18-25	0	Non-Travel	287	Research & Development	1
4	RM829	18	18-25	1	Non-Travel	247	Research & Development	1
...	...	...	...	...	...	...	...	...
1475	RM412	60	55+	0	1	422	Research & Development	1
1476	RM428	60	55+	0	0	1499	Sales	1
1477	RM537	60	55+	0	1	1179	Sales	1
1478	RM880	60	55+	0	1	696	Sales	1
1479	RM1210	60	55+	0	1	370	Research & Development	1

1480 rows × 38 columns

```
In [51]: df1.describe()
```

Out[51]:

	Age	Attrition	DailyRate	DistanceFromHome	Education	EmployeeCount
count	1480.000000	1480.000000	1480.000000	1480.000000	1480.000000	1480.000000
mean	36.917568	0.160811	801.384459	9.220270	2.910811	1.000000
std	9.128559	0.367481	403.126988	8.131201	1.023796	1.000000
min	18.000000	0.000000	102.000000	1.000000	1.000000	1.000000
25%	30.000000	0.000000	465.000000	2.000000	2.000000	1.000000
50%	36.000000	0.000000	800.000000	7.000000	3.000000	1.000000
75%	43.000000	0.000000	1157.000000	14.000000	4.000000	1.000000
max	60.000000	1.000000	1499.000000	29.000000	5.000000	1.000000

8 rows × 27 columns

```
In [ ]:
```