

```
In [2]: import numpy as np
import seaborn as se
import matplotlib.pyplot as plt
import pandas as pd
```

```
In [5]: sp=pd.read_csv('/home/student/Desktop/Employee_Salary_Dataset.csv')
```

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

```
Out[6]:
```

	ID	Experience_Years	Age	Gender	Salary
0	1	5	28	Female	250000
1	2	1	21	Male	50000
2	3	3	23	Female	170000
3	4	2	22	Male	25000
4	5	1	17	Male	10000

```
In [7]: sp.mean()
```

```
Out[7]: ID                1.800000e+01
Experience_Years          9.200000e+00
Age                      3.548571e+01
Salary                   2.059147e+06
dtype: float64
```

```
In [9]: sp.loc[:, 'Age'].mean()
```

```
Out[9]: 35.48571428571429
```

```
In [11... sp.mean(axis=1)[0:4]
```

```
Out[11]: 0    62508.50
1    12506.00
2    42507.25
3     6257.00
dtype: float64
```

```
In [12... sp.median()
```

```
Out[12]: ID                18.0
Experience_Years           6.0
Age                      29.0
Salary                   250000.0
dtype: float64
```

```
In [13... sp.loc[:, 'Age'].median()
```

```
Out[13]: 29.0
```

```
In [14... sp.median(axis=1)[1:7]
```

```
Out[14]: 1    11.5
2    13.0
3    13.0
4    11.0
5    43.5
```

```
6      36.5  
dtype: float64
```

```
In [15...
```

```
sp.mode()
```

```
Out[15]:
```

	ID	Experience_Years	Age	Gender	Salary
0	1	2.0	54.0	Female	25000.0
1	2	NaN	NaN	NaN	250000.0
2	3	NaN	NaN	NaN	NaN
3	4	NaN	NaN	NaN	NaN
4	5	NaN	NaN	NaN	NaN
5	6	NaN	NaN	NaN	NaN
6	7	NaN	NaN	NaN	NaN
7	8	NaN	NaN	NaN	NaN
8	9	NaN	NaN	NaN	NaN
9	10	NaN	NaN	NaN	NaN
10	11	NaN	NaN	NaN	NaN
11	12	NaN	NaN	NaN	NaN
12	13	NaN	NaN	NaN	NaN
13	14	NaN	NaN	NaN	NaN
14	15	NaN	NaN	NaN	NaN
15	16	NaN	NaN	NaN	NaN
16	17	NaN	NaN	NaN	NaN
17	18	NaN	NaN	NaN	NaN
18	19	NaN	NaN	NaN	NaN
19	20	NaN	NaN	NaN	NaN
20	21	NaN	NaN	NaN	NaN
21	22	NaN	NaN	NaN	NaN
22	23	NaN	NaN	NaN	NaN
23	24	NaN	NaN	NaN	NaN
24	25	NaN	NaN	NaN	NaN
25	26	NaN	NaN	NaN	NaN
26	27	NaN	NaN	NaN	NaN
27	28	NaN	NaN	NaN	NaN
28	29	NaN	NaN	NaN	NaN
29	30	NaN	NaN	NaN	NaN
30	31	NaN	NaN	NaN	NaN
31	32	NaN	NaN	NaN	NaN

	ID	Experience_Years	Age	Gender	Salary
32	33		NaN	NaN	NaN
33	34		NaN	NaN	NaN
34	35		NaN	NaN	NaN

```
In [16... sp.loc[:, 'Age'].mode()
```

```
Out[16]: 0    54  
         Name: Age, dtype: int64
```

```
In [17... sp.mode(axis=1)[0:4]
```

```
Out[17]:
```

	0	1	2	3	4
0	1	5.0	28.0	Female	250000.0
1	2	1.0	21.0	Male	50000.0
2	3	NaN	NaN	NaN	NaN
3	4	2.0	22.0	Male	25000.0

```
In [18... sp.min()
```

```
Out[18]: ID                1  
         Experience_Years    1  
         Age                17  
         Gender              Female  
         Salary              3000  
         dtype: object
```

```
In [19... sp.max()
```

```
Out[19]: ID                35  
         Experience_Years    27  
         Age                62  
         Gender              Male  
         Salary             10000000  
         dtype: object
```

```
In [20... sp.loc[:, 'Salary'].max(skipna=False)
```

```
Out[20]: 10000000
```

```
In [21... sp.std()
```

```
Out[21]: ID                1.024695e+01  
         Experience_Years    7.552950e+00  
         Age                1.464355e+01  
         Salary             3.170124e+06  
         dtype: float64
```

```
In [22... sp.loc[:, 'Age'].std()
```

```
Out[22]: 14.643551940884361
```

```
In [23... sp.std(axis=1)[1:6]
```

```
Out[23]: 1    2.499600e+04  
         2    8.499517e+04
```

```
3    1.249534e+04
4    4.996171e+03
5    2.500485e+06
dtype: float64
```

```
In [24... sp.groupby(['Salary'])['Age'].mean()
```

```
Out[24]: Salary
3000      18.0
6000      21.0
6100      21.0
7500      23.0
8900      23.0
9000      21.0
10000     17.0
15000     21.0
20000     22.0
25000     24.0
50000     21.0
61500     36.0
80000     34.0
87000     27.0
170000    23.0
220100    40.0
250000    27.0
330000    36.0
650000    54.0
800000    54.0
900000    54.0
930000    34.0
1400000   29.0
1540000   55.0
5000000   54.0
5001000   62.0
6000050   39.0
6570000   54.0
6845000   29.0
7600000   49.0
7900000   54.0
9300000   53.0
10000000  62.0
Name: Age, dtype: float64
```

```
In [25... sp_u=sp.rename(columns={'Experience_Years': 'Age'}, inplace=False)
(sp_u.groupby(['Salary']).Age.mean())
```

```
Out[25]:
```

	Age	Age
Salary		
3000	1.0	18.0
6000	1.0	21.0
6100	2.0	21.0
7500	2.0	23.0
8900	4.0	23.0
9000	2.0	21.0
10000	1.0	17.0
15000	2.0	21.0

	Age	Age
Salary		
20000	3.0	22.0
25000	3.0	24.0
50000	1.0	21.0
61500	10.0	36.0
80000	10.0	34.0
87000	4.0	27.0
170000	3.0	23.0
220100	11.0	40.0
250000	4.5	27.0
330000	10.0	36.0
650000	15.0	54.0
800000	19.0	54.0
900000	15.0	54.0
930000	10.0	34.0
1400000	6.0	29.0
1540000	20.0	55.0
5000000	19.0	54.0
5001000	25.0	62.0
6000050	14.0	39.0
6570000	15.0	54.0
6845000	5.0	29.0
7600000	16.0	49.0
7900000	15.0	54.0
9300000	19.0	53.0
10000000	27.0	62.0

In [26... `sp.head()`

Out[26]:

	ID	Experience_Years	Age	Gender	Salary
0	1	5	28	Female	250000
1	2	1	21	Male	50000
2	3	3	23	Female	170000
3	4	2	22	Male	25000
4	5	1	17	Male	10000

In [27...]

```
from sklearn import preprocessing
enc=preprocessing.OneHotEncoder()
enc_sp=pd.DataFrame(enc.fit_transform(sp[['Age']]).toarray())
enc_sp
```

Out[27]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
7	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
14	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
18	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
21	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
23	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
24	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
26	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
28	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
29	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
31	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
32	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
33	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
34	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0

In [28...]

```
sp_encode=sp_u.join(enc_sp)  
sp_encode
```

Out[28]:

	ID	Age	Age	Gender	Salary	0	1	2	3	4	...	8	9	10	11	12	13
0	1	5	28	Female	250000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
1	2	1	21	Male	50000	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
2	3	3	23	Female	170000	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0
3	4	2	22	Male	25000	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
4	5	1	17	Male	10000	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
5	6	25	62	Male	5001000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
6	7	19	54	Female	800000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
7	8	2	21	Female	9000	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
8	9	10	36	Female	61500	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0
9	10	15	54	Female	650000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
10	11	4	26	Female	250000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
11	12	6	29	Male	1400000	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.0	0.0	0.0
12	13	14	39	Male	6000050	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	1.0	0.0	0.0
13	14	11	40	Male	220100	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0
14	15	2	23	Male	7500	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0
15	16	4	27	Female	87000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
16	17	10	34	Female	930000	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0
17	18	15	54	Female	7900000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
18	19	2	21	Male	15000	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
19	20	10	36	Male	330000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0
20	21	15	54	Male	6570000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
21	22	4	26	Male	25000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
22	23	5	29	Male	6845000	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.0	0.0	0.0
23	24	1	21	Female	6000	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
24	25	4	23	Female	8900	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0
25	26	3	22	Female	20000	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
26	27	1	18	Male	3000	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
27	28	27	62	Female	10000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
28	29	19	54	Female	5000000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0

	ID	Age	Age	Gender	Salary	0	1	2	3	4	...	8	9	10	11	12	13
29	30	2	21	Female	6100	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
30	31	10	34	Male	80000	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0
31	32	15	54	Male	900000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
32	33	20	55	Female	1540000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
33	34	19	53	Female	9300000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0
34	35	16	49	Male	7600000	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	1.0

35 rows × 23 columns

```
In [29... iris=pd.read_csv('/home/student/Desktop/Iris.csv')
```

```
In [30... iris.head()
```

```
Out[30]:
```

	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 [31... irisSet=(iris['Species']=='Iris-setosa')
print('Iris-setosa')
print(iris[irisSet].describe())
```

```
Iris-setosa
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	50.00000	50.00000	50.00000	50.00000	50.00000
mean	25.50000	5.00600	3.41800	1.46400	0.24400
std	14.57738	0.35249	0.38102	0.17351	0.10721
min	1.00000	4.30000	2.30000	1.00000	0.10000
25%	13.25000	4.80000	3.12500	1.40000	0.20000
50%	25.50000	5.00000	3.40000	1.50000	0.20000
75%	37.75000	5.20000	3.67500	1.57500	0.30000
max	50.00000	5.80000	4.40000	1.90000	0.60000

```
In [34... irisVer=(iris['SepalWidthCm']=='Iris-setosa')
print('Iris-setosa')
print(iris[irisVer].describe())
```

```
Iris-setosa
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	0.0	0.0	0.0	0.0	0.0



mean	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN

```
In [36... irisVir=(iris['PetalLengthCm']=='Iris-setosa')
print('Iris-setosa')
print(iris[irisVir].describe())
```

Iris-setosa						
	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	
count	0.0	0.0	0.0	0.0	0.0	0.0
mean	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN

```
In [ ]:
```