

# SALARY DATASET

```
1 import pandas as pd
2 dataset1 =pd.read_csv("salary.csv")
```

In [2]: 1 dataset1.head()

Out [2]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

In [3]: 1 dataset1.tail()

Out [3]:

	YearsExperience	Salary
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

In [4]: 1 dataset1.head(10)

Out [4]:

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0
5	2.9	56642.0
6	3.0	60150.0
7	3.2	54445.0
8	3.2	64445.0
9	3.7	57189.0

```
In [5]: 1 dataset1.tail(10)
```

Out [5]:

	YearsExperience	Salary
20	6.8	91738.0
21	7.1	98273.0
22	7.9	101302.0
23	8.2	113812.0
24	8.7	109431.0
25	9.0	105582.0
26	9.5	116969.0
27	9.6	112635.0
28	10.3	122391.0
29	10.5	121872.0

```
In [6]: 1 import pandas as pd
2 crime_df = pd.read_csv('salary.csv')
3 null_values_last_5_columns = crime_df.iloc[:, -5:].isnull().sum()
4 print(null_values_last_5_columns)
```

```
YearsExperience    0
Salary            0
dtype: int64
```

```
In [7]: 1 dataset1.isnull().sum()
```

Out [7]:

```
YearsExperience    0
Salary            0
dtype: int64
```

```
In [8]: 1 dataset1[dataset1.YearsExperience.isnull()]
```

Out [8]:

YearsExperience	Salary
-----------------	--------

```
In [9]: 1 dataset1.shape
```

Out [9]: (30, 2)

```
In [10]: 1 dataset1['YearsExperience'].value_counts()
```

```
Out[10]: YearsExperience
3.2      2
4.0      2
1.1      1
5.3      1
10.3     1
9.6      1
9.5      1
9.0      1
8.7      1
8.2      1
7.9      1
7.1      1
6.8      1
6.0      1
5.9      1
5.1      1
1.3      1
4.9      1
4.5      1
4.1      1
3.9      1
3.7      1
3.0      1
2.9      1
2.2      1
2.0      1
1.5      1
10.5     1
Name: count, dtype: int64
```

```
In [11]: 1 import pandas as pd
          2 crime_df = pd.read_csv('salary.csv')
          3 for col in crime_df.columns:
          4     print(f"Value counts for column {col}:")
          5     print(crime_df[col].value_counts())
          6     print()
```

Value counts for column YearsExperience:

YearsExperience

3.2	2
4.0	2
1.1	1
5.3	1
10.3	1
9.6	1
9.5	1
9.0	1
8.7	1
8.2	1
7.9	1
7.1	1
6.8	1
6.0	1
5.9	1
5.1	1
1.3	1
4.9	1
4.5	1
4.1	1
3.9	1
3.7	1
3.0	1
2.9	1
2.2	1
2.0	1
1.5	1
10.5	1

Name: count, dtype: int64

Value counts for column Salary:

Salary

39343.0	1
46205.0	1
122391.0	1
112635.0	1
116969.0	1
105582.0	1
109431.0	1
113812.0	1
101302.0	1
98273.0	1
91738.0	1
93940.0	1
81363.0	1
83088.0	1
66029.0	1
67938.0	1
61111.0	1
57081.0	1
56957.0	1
55794.0	1
63218.0	1
57189.0	1
64445.0	1
54445.0	1
60150.0	1
56642.0	1
39891.0	1

```
43525.0    1
37731.0    1
121872.0   1
Name: count, dtype: int64
```

```
In [12]: 1 dataset_length=len(dataset1)
          2 dataset_length
```

```
Out[12]: 30
```

```
In [13]: 1 dataset_col=len(dataset1.columns)
          2 dataset_col
```

```
Out[13]: 2
```

```
In [14]: 1 dataset1.describe()
```

```
Out[14]:
```

	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

```
In [15]: 1 dataset1.YearsExperience.describe()
```

```
Out[15]: count    30.000000
mean         5.313333
std          2.837888
min          1.100000
25%          3.200000
50%          4.700000
75%          7.700000
max          10.500000
Name: YearsExperience, dtype: float64
```

```
In [16]: 1 dataset1.skew()
```

```
Out[16]: YearsExperience    0.37956
Salary                    0.35412
dtype: float64
```

```
In [17]: 1 dataset1.var()
```

```
Out[17]: YearsExperience    8.053609e+00
Salary                    7.515510e+08
dtype: float64
```

```
In [18]: 1 dataset1.kurtosis()
```

```
Out[18]: YearsExperience    -1.012212  
Salary                  -1.295421  
dtype: float64
```

```
In [19]: 1 print(dataset1.dtypes)
```

```
YearsExperience    float64  
Salary            float64  
dtype: object
```

```
In [ ]: 1
```