In [30]:

```python
import pandas as pd
data=pd.read_csv("C:\\Users\\Lenovo\\OneDrive\\Desktop\\pratice_file.csv")
print(type(data))
```

```
<class 'pandas.core.frame.DataFrame'>
```

In [31]:

```python
data.info
```

Out[31]:

```
<bound method DataFrame.info of        NAME  WORK HOUR  ABSENT  SALARY  OVERT
IME  TOTAL
0   Shasank         20     2.0   20000      10.0  22500
1     Myank         60     4.0   60000       6.0  61500
2     Mayur         40     1.0   40000       9.0  42250
3     Monoj         32     NaN   32000      12.0  35000
4   Abhisek         52     NaN   52000      15.0  55750
5     Ayush         15     1.0   15000       3.0  15750
6    Vikram         58     2.0   58000       7.0  59750
7     Tusar         45     NaN   45000      12.0  48000
8    Sourav         67     1.0   67000       NaN  67000
9     Manas         63     NaN   63000       NaN  63000
10  Abhisek         52     NaN   52000      15.0  55750
11   Vikram         58     2.0   58000       7.0  59750
12  Shasank         20     2.0   20000      10.0  22500>
```

In [32]:

```python
data.describe()
```

Out[32]:

	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
count	13.000000	8.000000	13.000000	11.000000	13.000000
mean	44.769231	1.875000	44769.230769	9.636364	46807.692308
std	17.823925	0.991031	17823.925148	3.748939	17532.456167
min	15.000000	1.000000	15000.000000	3.000000	15750.000000
25%	32.000000	1.000000	32000.000000	7.000000	35000.000000
50%	52.000000	2.000000	52000.000000	10.000000	55750.000000
75%	58.000000	2.000000	58000.000000	12.000000	59750.000000
max	67.000000	4.000000	67000.000000	15.000000	67000.000000

In [33]:

```
data=data.drop_duplicates()
data
```

Out[33]:

	NAME	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
0	Shasank	20	2.0	20000	10.0	22500
1	Myank	60	4.0	60000	6.0	61500
2	Mayur	40	1.0	40000	9.0	42250
3	Monoj	32	NaN	32000	12.0	35000
4	Abhisek	52	NaN	52000	15.0	55750
5	Ayush	15	1.0	15000	3.0	15750
6	Vikram	58	2.0	58000	7.0	59750
7	Tusar	45	NaN	45000	12.0	48000
8	Sourav	67	1.0	67000	NaN	67000
9	Manas	63	NaN	63000	NaN	63000

In [34]:

```
data.isnull()
```

Out[34]:

	NAME	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	True	False	False	False
4	False	False	True	False	False	False
5	False	False	False	False	False	False
6	False	False	False	False	False	False
7	False	False	True	False	False	False
8	False	False	False	False	True	False
9	False	False	True	False	True	False

In [35]:

```
data.isnull().sum()
```

Out[35]:

```
NAME         0
WORK HOUR    0
ABSENT       4
SALARY       0
OVERTIME     2
TOTAL        0
dtype: int64
```

In [36]:

```
data.notnull()
```

Out[36]:

```
   NAME  WORK HOUR  ABSENT  SALARY  OVERTIME  TOTAL
0  True       True    True    True      True   True
1  True       True    True    True      True   True
2  True       True    True    True      True   True
3  True       True   False    True      True   True
4  True       True   False    True      True   True
5  True       True    True    True      True   True
6  True       True    True    True      True   True
7  True       True   False    True      True   True
8  True       True    True    True     False   True
9  True       True   False    True     False   True
```

In [37]:

```
data.isnull().sum().sum()
```

Out[37]:

```
6
```

In [38]:

```
data2=data.fillna(value=0)
data2
```

Out[38]:

```
      NAME  WORK HOUR  ABSENT  SALARY  OVERTIME  TOTAL
0  Shasank         20     2.0   20000      10.0  22500
1    Myank         60     4.0   60000       6.0  61500
2    Mayur         40     1.0   40000       9.0  42250
3    Monoj         32     0.0   32000      12.0  35000
4  Abhisek         52     0.0   52000      15.0  55750
5    Ayush         15     1.0   15000       3.0  15750
6   Vikram         58     2.0   58000       7.0  59750
7    Tusar         45     0.0   45000      12.0  48000
8   Sourav         67     1.0   67000       0.0  67000
9    Manas         63     0.0   63000       0.0  63000
```

In [39]:

```python
data3=data.fillna(method='pad')
data3
```

Out[39]:

	NAME	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
0	Shasank	20	2.0	20000	10.0	22500
1	Myank	60	4.0	60000	6.0	61500
2	Mayur	40	1.0	40000	9.0	42250
3	Monoj	32	1.0	32000	12.0	35000
4	Abhisek	52	1.0	52000	15.0	55750
5	Ayush	15	1.0	15000	3.0	15750
6	Vikram	58	2.0	58000	7.0	59750
7	Tusar	45	2.0	45000	12.0	48000
8	Sourav	67	1.0	67000	12.0	67000
9	Manas	63	1.0	63000	12.0	63000

In [40]:

```python
# filling the null value with the next value
data4=data.fillna(method='bfill')
data4
```

Out[40]:

	NAME	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
0	Shasank	20	2.0	20000	10.0	22500
1	Myank	60	4.0	60000	6.0	61500
2	Mayur	40	1.0	40000	9.0	42250
3	Monoj	32	1.0	32000	12.0	35000
4	Abhisek	52	1.0	52000	15.0	55750
5	Ayush	15	1.0	15000	3.0	15750
6	Vikram	58	2.0	58000	7.0	59750
7	Tusar	45	1.0	45000	12.0	48000
8	Sourav	67	1.0	67000	NaN	67000
9	Manas	63	NaN	63000	NaN	63000

In [41]:

```python
import numpy as np
from scipy import stats
```

In [42]:

```python
#detect the outliers using IQR
data2.columns
```

Out[42]:

```
Index(['NAME', 'WORK HOUR', 'ABSENT', 'SALARY', 'OVERTIME', 'TOTAL'], dtype
='object')
```

In [43]:

```
data2.drop(['NAME'],axis=1,inplace=True)
data2
```

Out[43]:

	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
0	20	2.0	20000	10.0	22500
1	60	4.0	60000	6.0	61500
2	40	1.0	40000	9.0	42250
3	32	0.0	32000	12.0	35000
4	52	0.0	52000	15.0	55750
5	15	1.0	15000	3.0	15750
6	58	2.0	58000	7.0	59750
7	45	0.0	45000	12.0	48000
8	67	1.0	67000	0.0	67000
9	63	0.0	63000	0.0	63000

In [44]:

```
Q1=data2.quantile(0.25)
Q3=data2.quantile(0.75)
IQR=Q3-Q1
print(IQR)
```

```
WORK HOUR       25.50
ABSENT           1.75
SALARY       25500.00
OVERTIME         7.75
TOTAL        24250.00
dtype: float64
```

In [45]:

```
data2=data2[~((data2<(Q1-1.5*IQR))|(data2>(Q3+1.5*IQR))).any(axis=1)]
data2
```

Out[45]:

	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
0	20	2.0	20000	10.0	22500
1	60	4.0	60000	6.0	61500
2	40	1.0	40000	9.0	42250
3	32	0.0	32000	12.0	35000
4	52	0.0	52000	15.0	55750
5	15	1.0	15000	3.0	15750
6	58	2.0	58000	7.0	59750
7	45	0.0	45000	12.0	48000
8	67	1.0	67000	0.0	67000
9	63	0.0	63000	0.0	63000

In [46]:

```
data2.describe()
```

Out[46]:

	WORK HOUR	ABSENT	SALARY	OVERTIME	TOTAL
count	10.000000	10.000000	10.00000	10.000000	10.000000
mean	45.200000	1.100000	45200.00000	7.400000	47050.000000
std	18.164679	1.286684	18164.67879	5.168279	17794.271862
min	15.000000	0.000000	15000.00000	0.000000	15750.000000
25%	34.000000	0.000000	34000.00000	3.750000	36812.500000
50%	48.500000	1.000000	48500.00000	8.000000	51875.000000
75%	59.500000	1.750000	59500.00000	11.500000	61062.500000
max	67.000000	4.000000	67000.00000	15.000000	67000.000000

In [ ]:

```

```