```
import pandas as pd
from pandas import read_csv
import numpy as np
import matplotlib.pyplot as plt
```

```
from google.colab import drive
drive.mount("/content/drive")
```

```
Mounted at /content/drive
```

```
data=read_csv("/content/drive/My Drive/dataset/HR.csv")
print(data)
```

```
    umur jenis kelamin pendidikan  lama kerja      gaji
0     25             p         S1           1   5000000
1     24             p         D3           1   4800000
2     26             p         S1           2   6000000
3     27             l         S1           1   5500000
4     30             p         S1           6   7000000
5     25             l         D3           1   6000000
6     32             l         S2           3  18000000
7     23             l         D3           1   4000000
8     27             l         S1           4  65000000
9     28             p         S1           5   7000000
10    31             p         S1           6   9000000
11    32             p         D3           4   9200000
12    35             l         S1           7  10000000
13    29             p         S1           4   8000000
14    29             l         D3           5   8200000
15    30             l         S2           3  12000000
16    25             p         S1           1   7000000
17    29             l         S1           5  75000000
18    31             l         D3           6  98000000
```

```
data['umur']
```

```
0     25
1     24
2     26
3     27
4     30
5     25
6     32
7     23
8     27
9     28
10    31
11    32
12    35
13    29
14    29
15    30
16    25
17    29
18    31
Name: umur, dtype: int64
```

```
data['lama kerja']
```

```
0     1
1     1
2     2
3     1
4     6
5     1
6     3
7     1
8     4
9     5
10    6
11    4
12    7
13    4
14    5
15    3
16    1
17    5
```

```
18    6
Name: lama kerja, dtype: int64
```

```
data ['gaji']
```

```
0      5000000
1      4800000
2      6000000
3      5500000
4      7000000
5      6000000
6     18000000
7      4000000
8     65000000
9      7000000
10     9000000
11     9200000
12    10000000
13     8000000
14     8200000
15    12000000
16     7000000
17    75000000
18    98000000
Name: gaji, dtype: int64
```

```
data ['pendidikan']
```

```
0     S1
1     D3
2     S1
3     S1
4     S1
5     D3
6     S2
7     D3
8     S1
9     S1
10    S1
11    D3
12    S1
13    S1
14    D3
15    S2
16    S1
17    S1
18    D3
Name: pendidikan, dtype: object
```

```
data.describe()
```

	umur	lama kerja	gaji
count	19.000000	19.000000	1.900000e+01
mean	28.315789	3.473684	1.919474e+07
std	3.180717	2.091475	2.751954e+07
min	23.000000	1.000000	4.000000e+06
25%	25.500000	1.000000	6.000000e+06
50%	29.000000	4.000000	8.000000e+06
75%	30.500000	5.000000	1.100000e+07
max	35.000000	7.000000	9.800000e+07

```
data.describe(include=['object'])
```

	jenis kelamin	pendidikan
count	19	19
unique	2	3
top	l	S1
freq	10	11

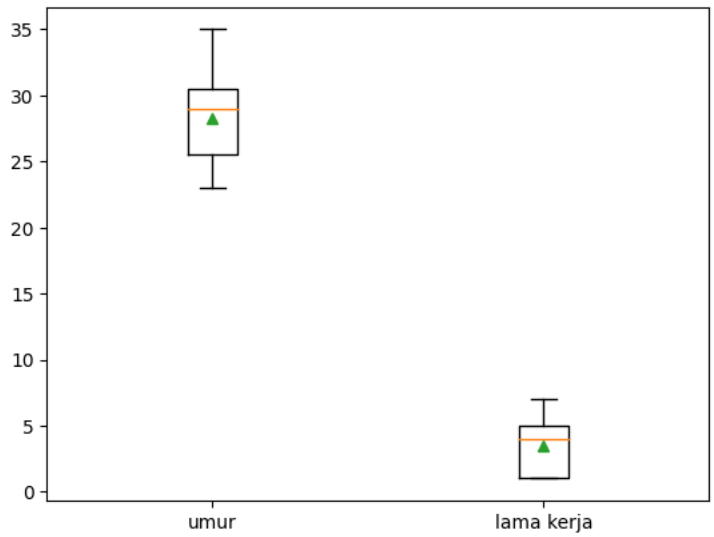
```
data['jenis kelamin'].value_counts()
```

```
l    10
p     9
Name: jenis kelamin, dtype: int64
```

```
data['pendidikan'].value_counts()
```

```
S1    11
D3     6
S2     2
Name: pendidikan, dtype: int64
```

```
umur, lamakerja = data['umur'],data['lama kerja']

plt.boxplot([umur, lamakerja], labels=['umur','lama kerja'], showmeans=True);
```



Inferntal

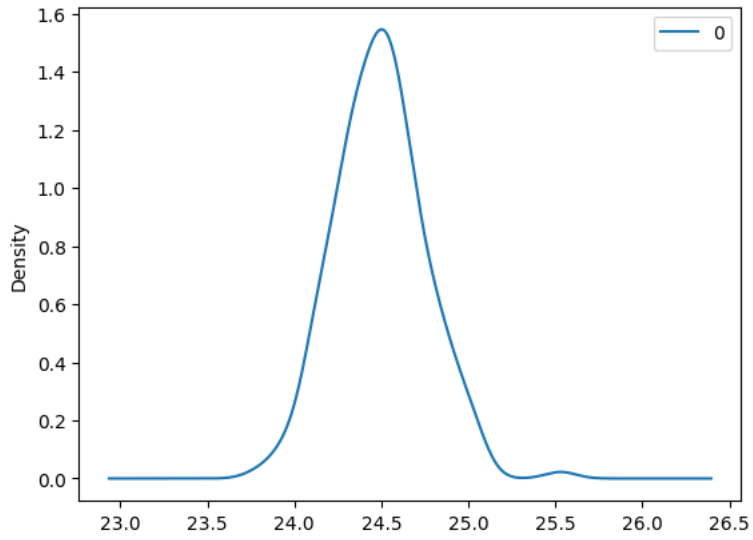
```
population = np.random.randint(20,30,1000)
```

```
np.random.seed(10)

estimates=[]

for x in range(200):
  sample = np.random.choice(a=population, size=100)
  estimates.append (sample.mean())


np.mean(population)
```

```
24.458
```

```
pd.DataFrame(estimates).plot(kind="density")
```

```
<Axes: ylabel='Density'>
```




```
import scipy.stats as stats

z_critical =stats.norm.ppf(q=0.9)


t_critical =stats.t.ppf(q=0.9, df=24)


margin_of_error = z_critical * (np.std(estimates)/np.sqrt(200))


np.mean(estimates) - margin_of_error
```

```
24.470233463780538
```