In [3]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
dataset= pd.read_csv('F:\linear dataset\salary.csv')
```

In [4]:
```python
dataset
```

Out[4]:

	YearsExperience	Salary
0	1.1	39343
1	1.3	46205
2	1.5	37731
3	2.0	43525
4	2.2	39891
5	2.9	56642
6	3.0	60150
7	3.2	54445
8	3.2	64445
9	3.7	57189
10	3.9	63218
11	4.0	55794
12	4.0	56957
13	4.1	57081
14	4.5	61111
15	4.9	67938
16	5.1	66029
17	5.3	83088
18	5.9	81363
19	6.0	93940
20	6.8	91738
21	7.1	98273
22	7.9	101302
23	8.2	113812
24	8.7	109431
25	9.0	105582
26	9.5	116969
27	9.6	112635
28	10.3	122391
29	10.5	121872
30	11.2	127345
31	11.5	126756
32	12.3	128765
33	12.9	135675
34	13.5	139465

In [8]:
```python
figure,axis=plt.subplots(2,1)
axis[0].scatter(dataset.YearsExperience, dataset.Salary)
axis[1].plot(dataset.YearsExperience,dataset.Salary)
```

Out[8]: [<matplotlib.lines.Line2D at 0x19fa2e5e310>]
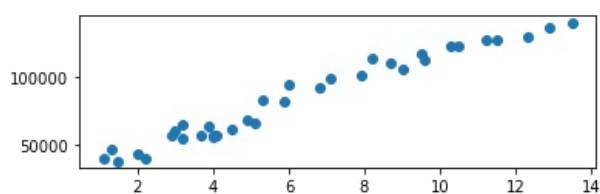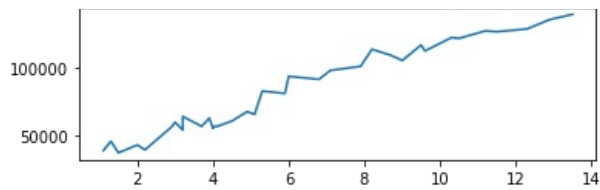
In [9]:
```python
dataset.describe()
```

Out[9]:

	YearsExperience	Salary
count	35.000000	35.000000
mean	6.308571	83945.600000
std	3.618610	32162.673003
min	1.100000	37731.000000
25%	3.450000	57019.000000
50%	5.300000	81363.000000
75%	9.250000	113223.500000
max	13.500000	139465.000000

In [10]:
```python
dataset.skew(axis=0,skipna=True)
```

Out[10]:
```
YearsExperience    0.420163
Salary             0.209005
dtype: float64
```

In [14]:
```python
dataset['Salary'].plot(kind='density')
```

Out[14]: `<AxesSubplot:ylabel='Density'>`
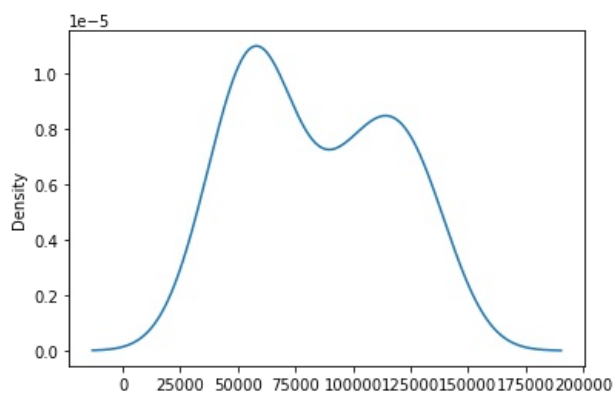
In [15]:
```python
x= dataset.YearsExperience
y= dataset.Salary
```

In [16]:
```python
x
```

Out[16]:
```
0     1.1
1     1.3
2     1.5
3     2.0
4     2.2
5     2.9
6     3.0
7     3.2
8     3.2
9     3.7
10    3.9
11    4.0
12    4.0
13    4.1
```

```
14    4.5
15    4.9
16    5.1
17    5.3
18    5.9
19    6.0
20    6.8
21    7.1
22    7.9
23    8.2
24    8.7
25    9.0
26    9.5
27    9.6
28    10.3
29    10.5
30    11.2
31    11.5
32    12.3
33    12.9
34    13.5
Name: YearsExperience, dtype: float64
```

In [17]:
```python
y
```

Out[17]:
```
0      39343
1      46205
2      37731
3      43525
4      39891
5      56642
6      60150
7      54445
8      64445
9      57189
10     63218
11     55794
12     56957
13     57081
14     61111
15     67938
16     66029
17     83088
18     81363
19     93940
20     91738
21     98273
22    101302
23    113812
24    109431
25    105582
26    116969
27    112635
28    122391
29    121872
30    127345
31    126756
32    128765
33    135675
34    139465
Name: Salary, dtype: int64
```

In [20]:
```python
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test= train_test_split(x,y,test_size=0.2,random_state=0)
```

In [21]:
```python
x_train.describe()
```

Out[21]:
```
count    28.000000
mean      6.121429
std       3.842776
min       1.100000
25%       3.150000
50%       4.900000
75%       9.125000
max      13.500000
Name: YearsExperience, dtype: float64
```

In [ ]: