In [2]:
```
import openpyxl
import pandas as pd

workbook = openpyxl.Workbook()
sheet = workbook.active

data = [
    ['NAME', 'DOMAIN', 'AGE', 'LOCATION', 'SALARY', 'EXP'],
    ['ALEX', 'TESTING', 25, 'BNG', 5000, 2],
    ['BARB', 'JAVA', 30, 'CHE', 10000, 3],
    ['CHERRY', 'C', 35, 'PUNE',15000, 4],
    ['DIPAN', 'DA', 38,'MUMBAI', 20000, 5],
    ['ESWAR', 'DS',40, 'HYD', 50000, 6]
]

for row in data:
    sheet.append(row)

workbook.save('data.xlsx')
```

In [3]:
```
data
```

Out[3]:
```
[['NAME', 'DOMAIN', 'AGE', 'LOCATION', 'SALARY', 'EXP'],
 ['ALEX', 'TESTING', 25, 'BNG', 5000, 2],
 ['BARB', 'JAVA', 30, 'CHE', 10000, 3],
 ['CHERRY', 'C', 35, 'PUNE', 15000, 4],
 ['DIPAN', 'DA', 38, 'MUMBAI', 20000, 5],
 ['ESWAR', 'DS', 40, 'HYD', 50000, 6]]
```

In [5]:
```
import os
os.getcwd()
```

Out[5]:
```
'C:\\Users\\YASH'
```

In [6]:
```
emp=pd.read_excel(r'C:\\Users\\YASH\data.xlsx')
emp
```

Out[6]:

| | NAME | DOMAIN | AGE | LOCATION | SALARY | EXP |
|---|---|---|---|---|---|---|
| **0** | ALEX | TESTING | 25 | BNG | 5000 | 2 |
| **1** | BARB | JAVA | 30 | CHE | 10000 | 3 |
| **2** | CHERRY | C | 35 | PUNE | 15000 | 4 |
| **3** | DIPAN | DA | 38 | MUMBAI | 20000 | 5 |
| **4** | ESWAR | DS | 40 | HYD | 50000 | 6 |

In [7]:
```
emp.shape
```

Out[7]:
```
(5, 6)
```

In [8]:
```
emp.columns
```

Out[8]:
```
Index(['NAME', 'DOMAIN', 'AGE', 'LOCATION', 'SALARY', 'EXP'], dtype='object')
```

In [9]:
```
len(emp.columns)
```

Out[9]:
```
6
```

In [10]:
```
len(emp)
```

Out[10]:
```
5
```

In [11]:
```
emp
```

Out[11]:

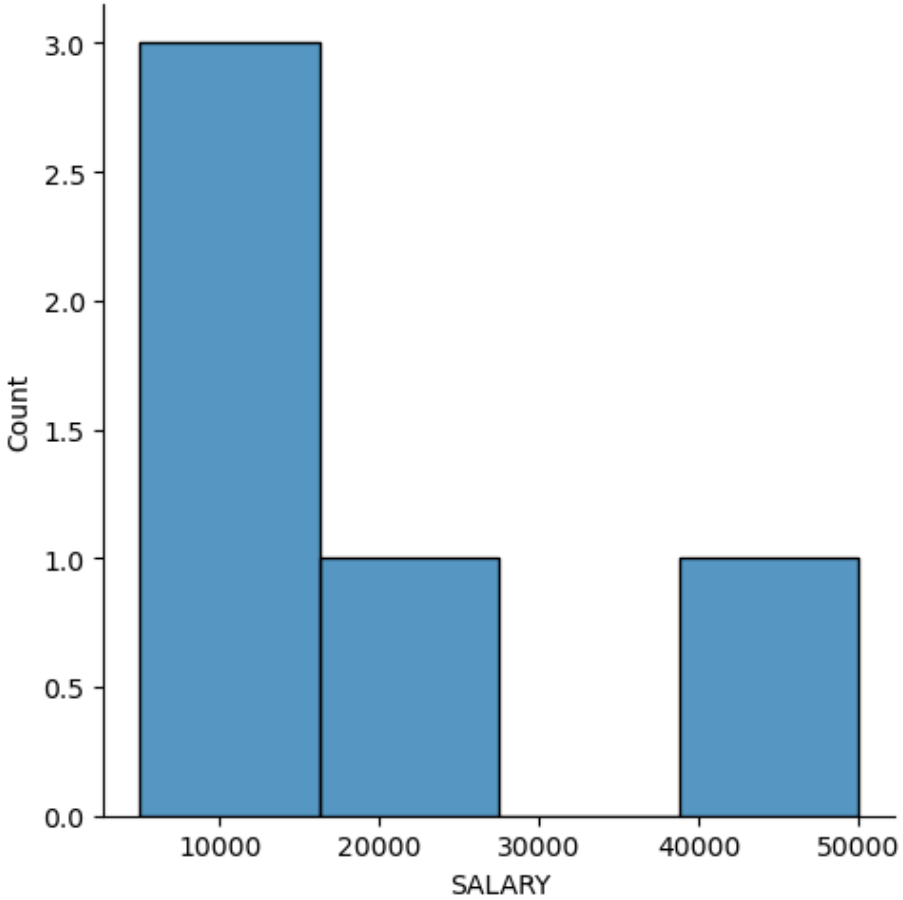
| | NAME | DOMAIN | AGE | LOCATION | SALARY | EXP |
|---|---|---|---|---|---|---|
| **0** | ALEX | TESTING | 25 | BNG | 5000 | 2 |
| **1** | BARB | JAVA | 30 | CHE | 10000 | 3 |
| **2** | CHERRY | C | 35 | PUNE | 15000 | 4 |
| **3** | DIPAN | DA | 38 | MUMBAI | 20000 | 5 |
| **4** | ESWAR | DS | 40 | HYD | 50000 | 6 |

In [12]:
```
emp['SALARY']
```

Out[12]:
```
0     5000
1    10000
2    15000
3    20000
4    50000
Name: SALARY, dtype: int64
```

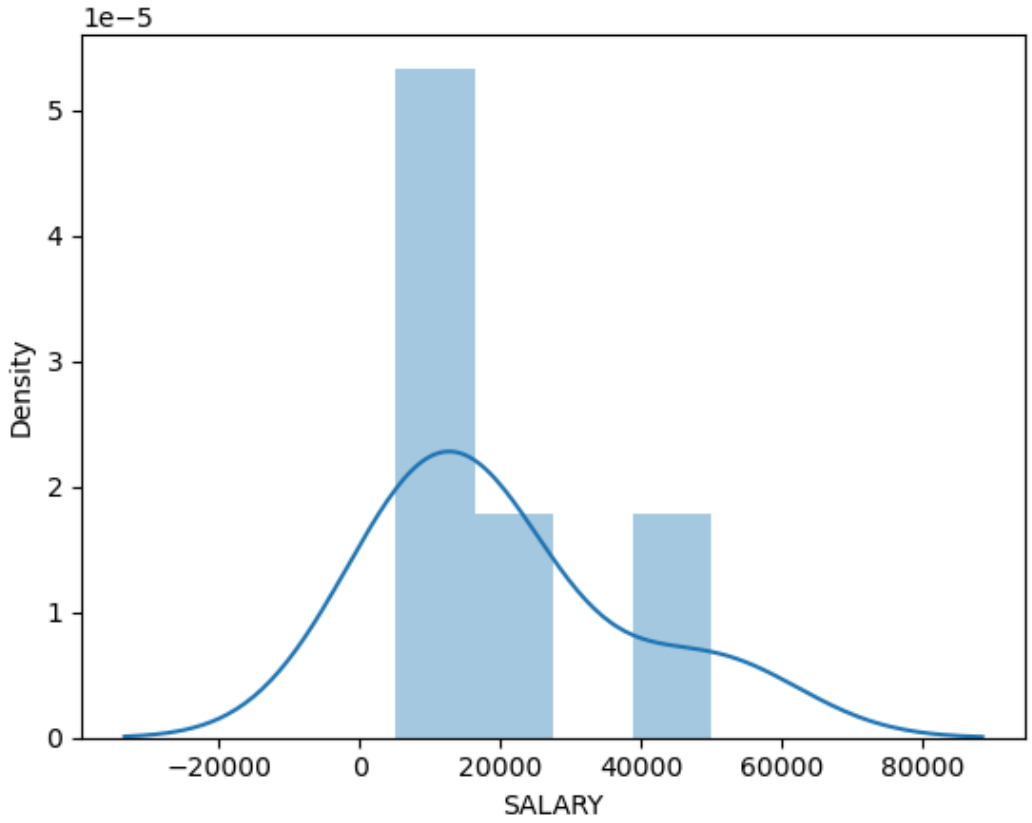
In [13]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [14]:
```
vis1=sns.displot(emp['SALARY'])
```

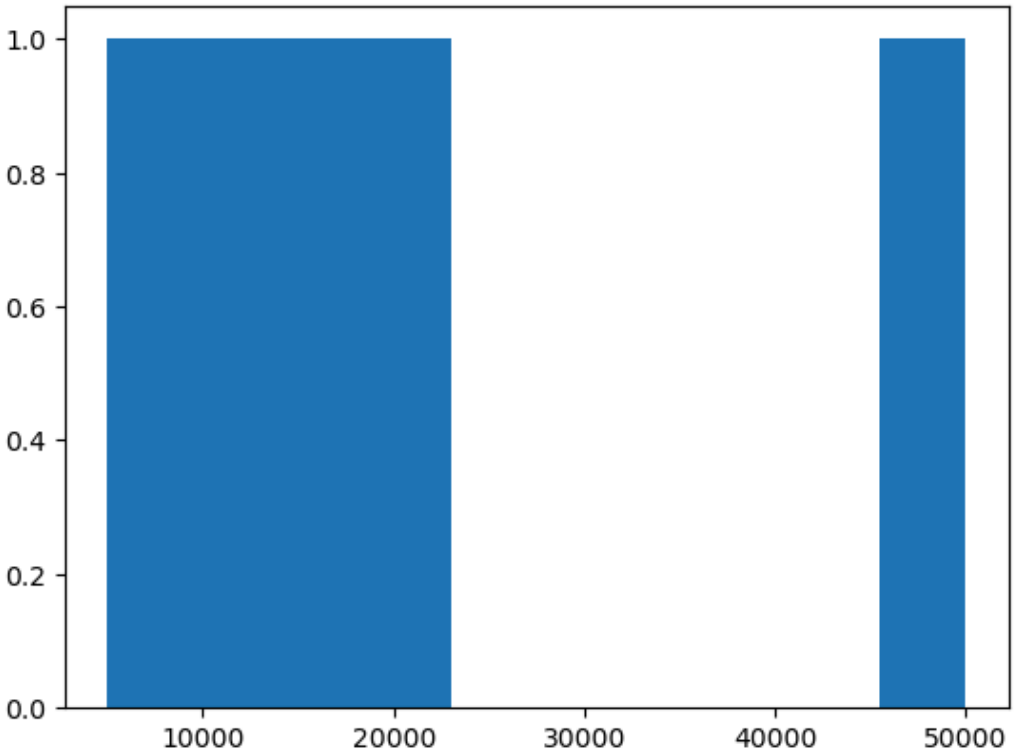



In [15]:
```
vis2=sns.distplot(emp['SALARY'])
```

```
C:\Users\YASH\AppData\Local\Temp\ipykernel_4736\375839575.py:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  vis2=sns.distplot(emp['SALARY'])
```

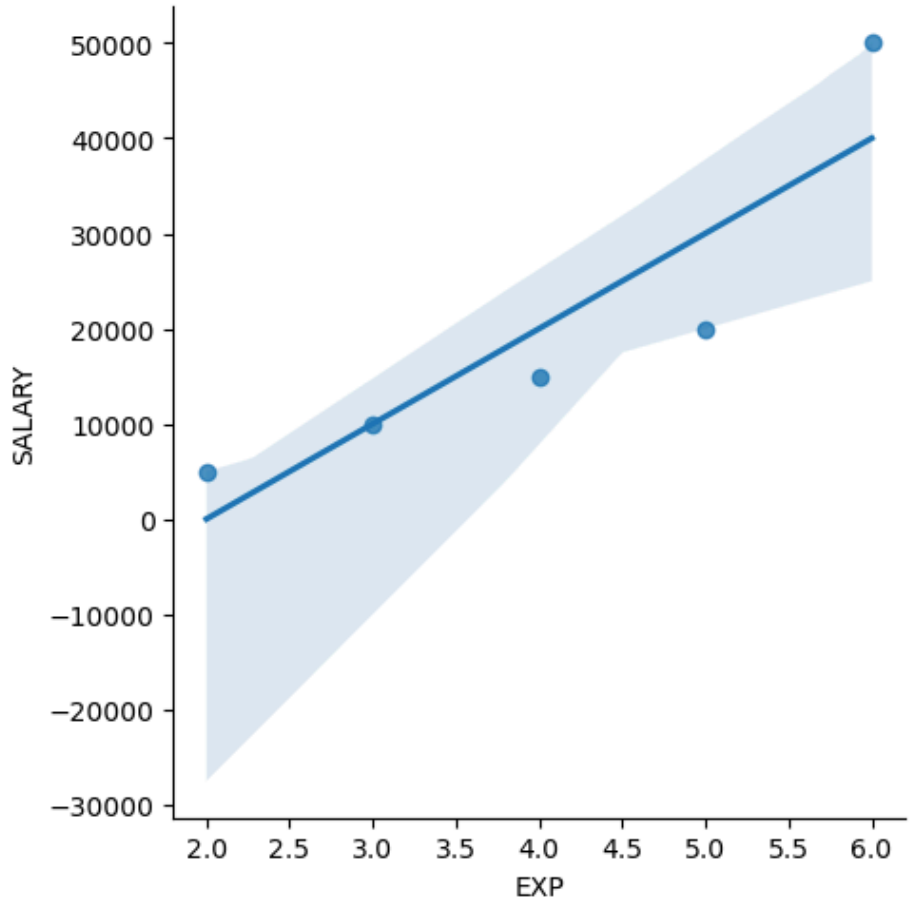



In [16]:
```
vis3=plt.hist(emp['SALARY'])
```




In [17]:
```
plt.rcParams['figure.figsize']=5,1
```

In [18]:
```
vis4=sns.lmplot(data=emp,x='EXP',y='SALARY')
```

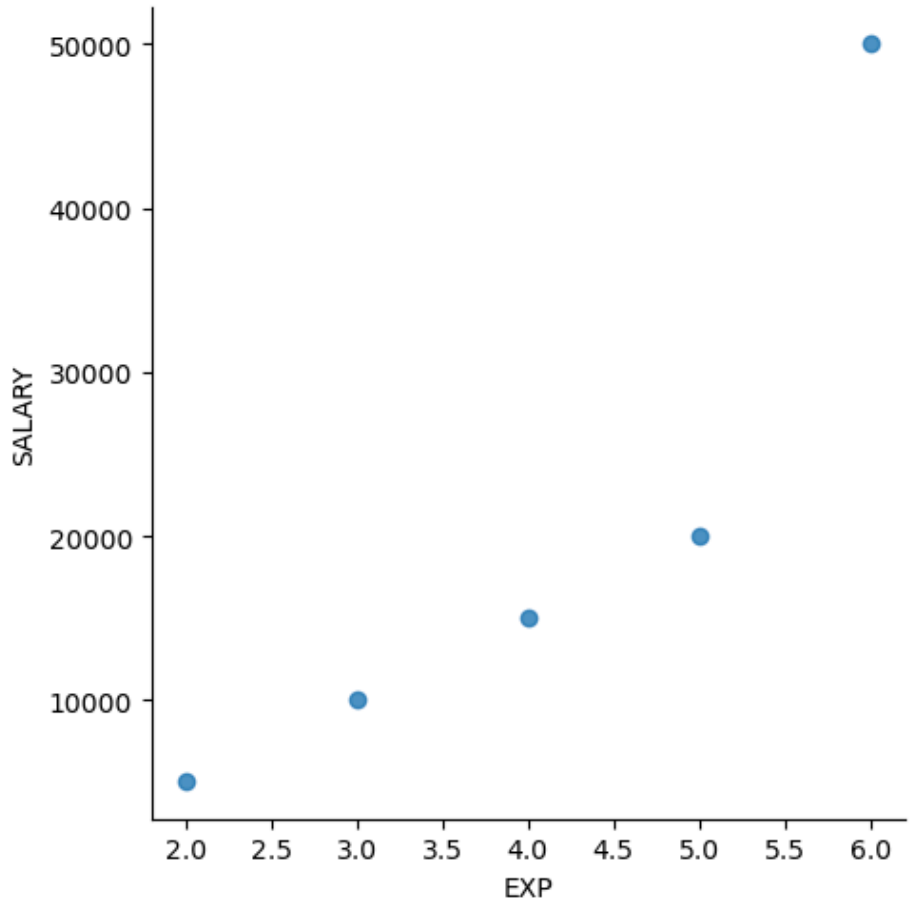

50000
40000
30000
20000
10000
0
−10000
−20000
−30000
SALARY
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
EXP

In [19]:
```
emp
```

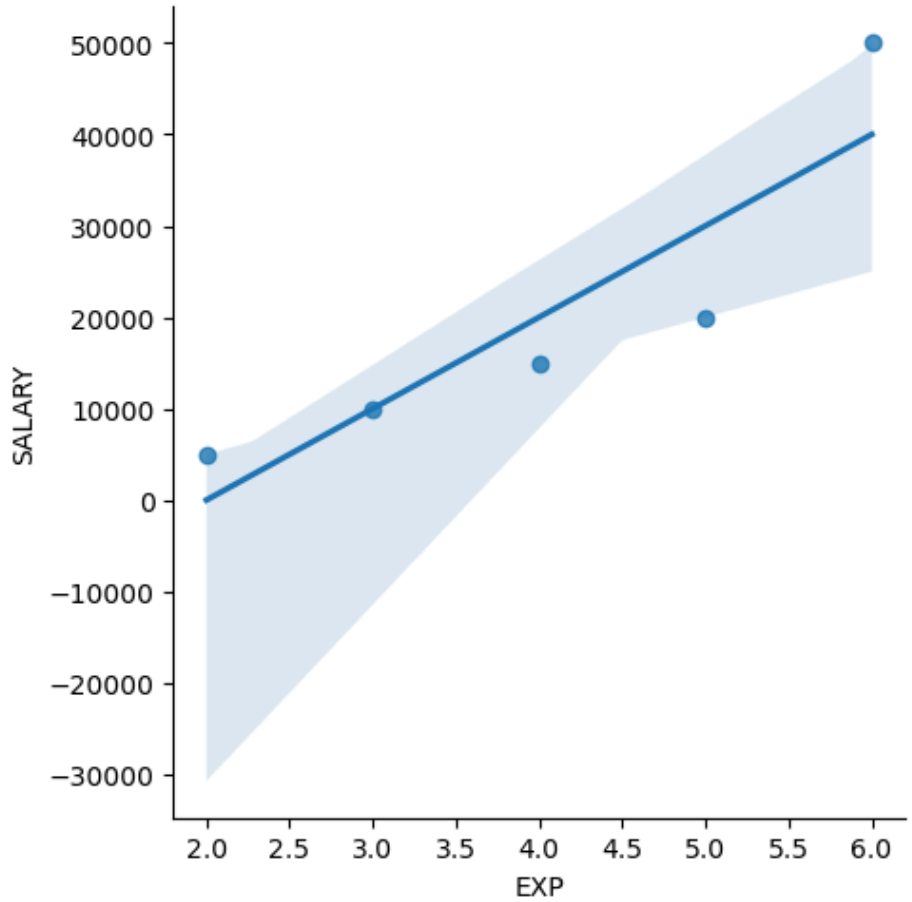
Out[19]:

| | NAME | DOMAIN | AGE | LOCATION | SALARY | EXP |
|---|---|---|---|---|---|---|
| **0** | ALEX | TESTING | 25 | BNG | 5000 | 2 |
| **1** | BARB | JAVA | 30 | CHE | 10000 | 3 |
| **2** | CHERRY | C | 35 | PUNE | 15000 | 4 |
| **3** | DIPAN | DA | 38 | MUMBAI | 20000 | 5 |
| **4** | ESWAR | DS | 40 | HYD | 50000 | 6 |

In [21]:
```
vis5=sns.lmplot(data=emp,x='EXP',y='SALARY',fit_reg=False)
```




In [22]:
```
vis6=sns.lmplot(data=emp,x='EXP',y='SALARY',fit_reg=True)
```


50000
40000
30000
20000
10000
0
−10000
−20000
−30000
SALARY
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
EXP

In [ ]: