In [4]:
```
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 [6]:
```
data
```

Out[6]:
```
[['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 [8]:
```
import os
os.getcwd()
```

Out[8]:
```
'C:\\Users\\swati\\OneDrive\\Documents\\Data_Analyst\\nov\\NONTECH PROJECT'
```

In [16]:
```
emp = pd.read_excel(r'C:\\Users\\swati\\OneDrive\\Documents\\Data_Analyst\\nov\\
emp
```

Out[16]:

	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 [18]:
```
emp.shape
```

Out[18]:
```
(5, 6)
```

In [20]:
```
emp.columns
```

Out[20]:
```
Index(['NAME', 'DOMAIN', 'AGE', 'LOCATION', 'SALARY', 'EXP'], dtype='object')
```

In [22]:
```
len(emp.columns)
```

Out[22]: 6

In [24]:
```
len(emp)
```

Out[24]: 5

In [26]:
```
emp
```

Out[26]:

	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 [28]:
```
emp[SALARY]
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[28], line 1
----> 1 emp[SALARY]

NameError: name 'SALARY' is not defined
```

In [30]:
```
emp['SALARY']
```

Out[30]:
```
0     5000
1    10000
2    15000
3    20000
4    50000
Name: SALARY, dtype: int64
```

In [32]:
```
emp[['SALARY', 'EXP']]
```

Out[32]:

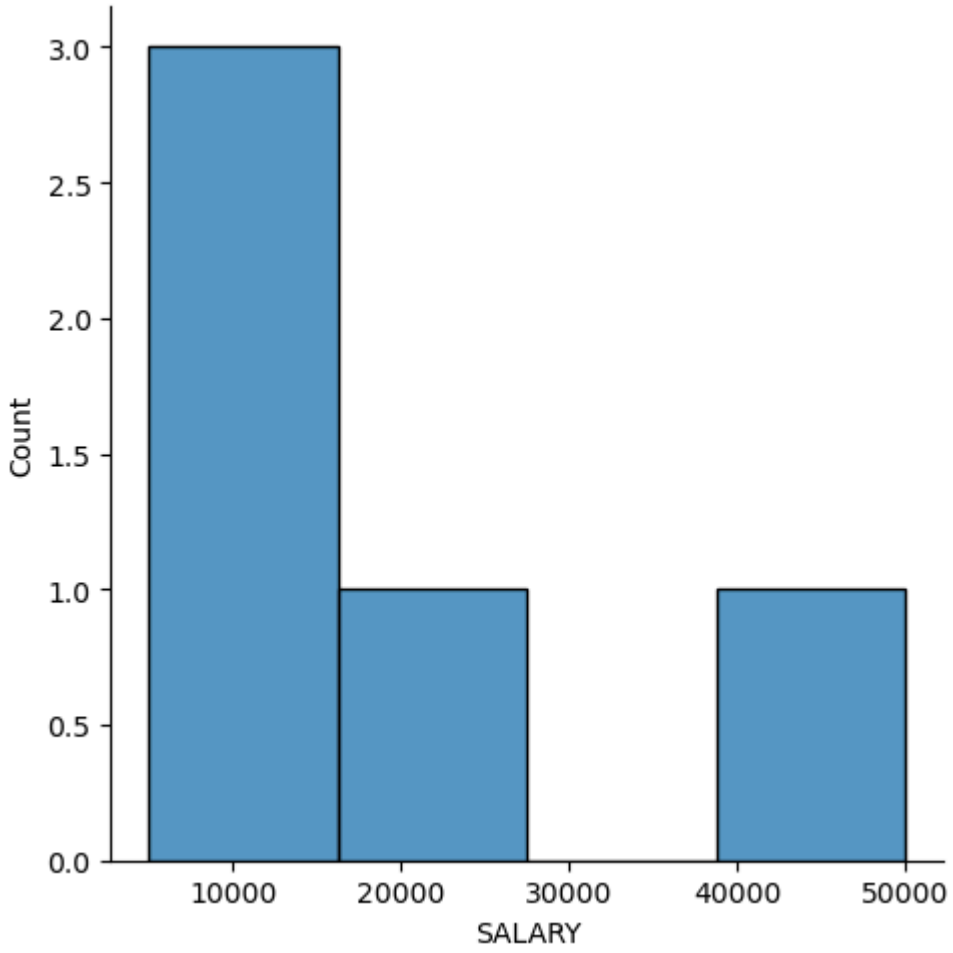
	SALARY	EXP
0	5000	2
1	10000	3
2	15000	4
3	20000	5
4	50000	6

In [34]:
```
emp.shape
```

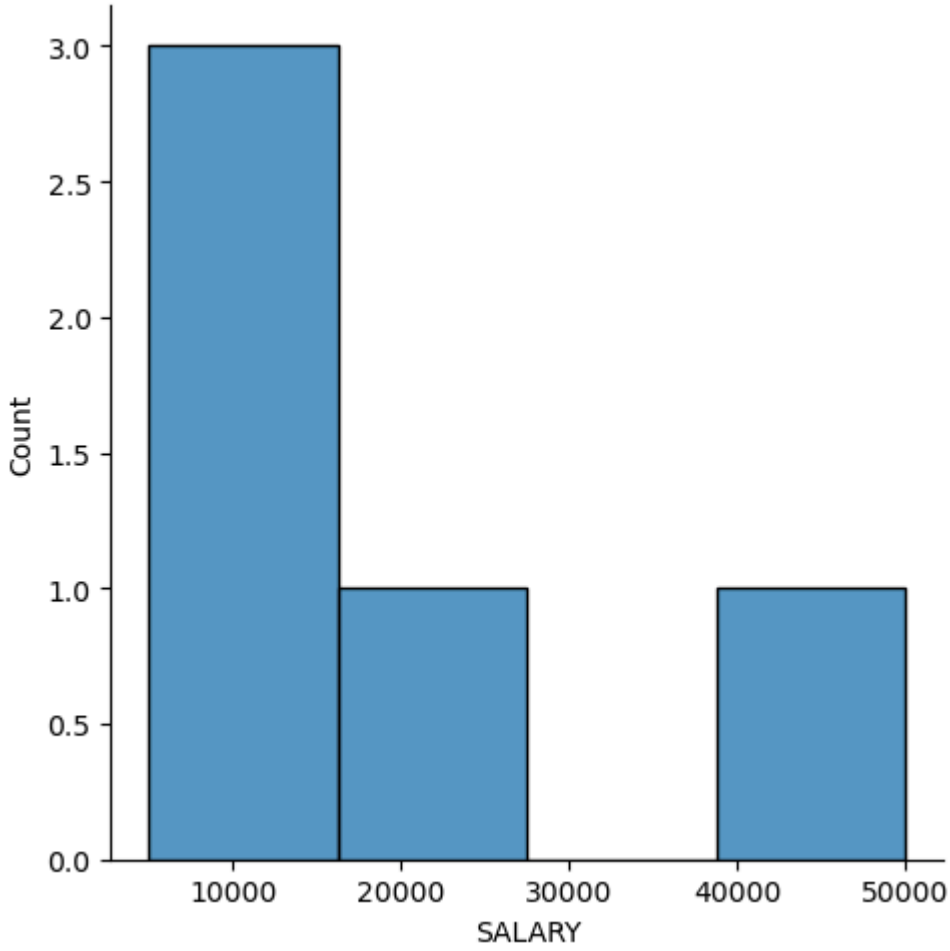
Out[34]: (5, 6)

In [38]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [40]:
```
vis1 = sns.displot(emp['SALARY'])
```

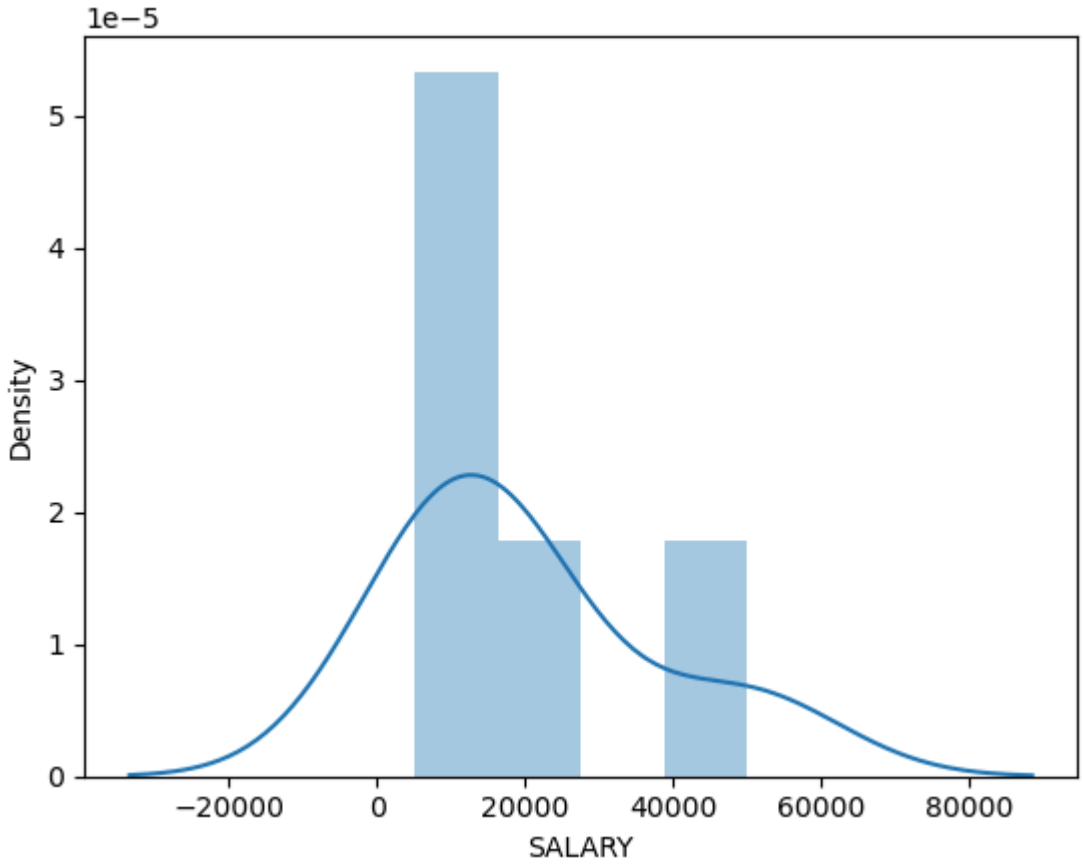


In [40]:
```
vis1 = sns.displot(emp['SALARY'])
```



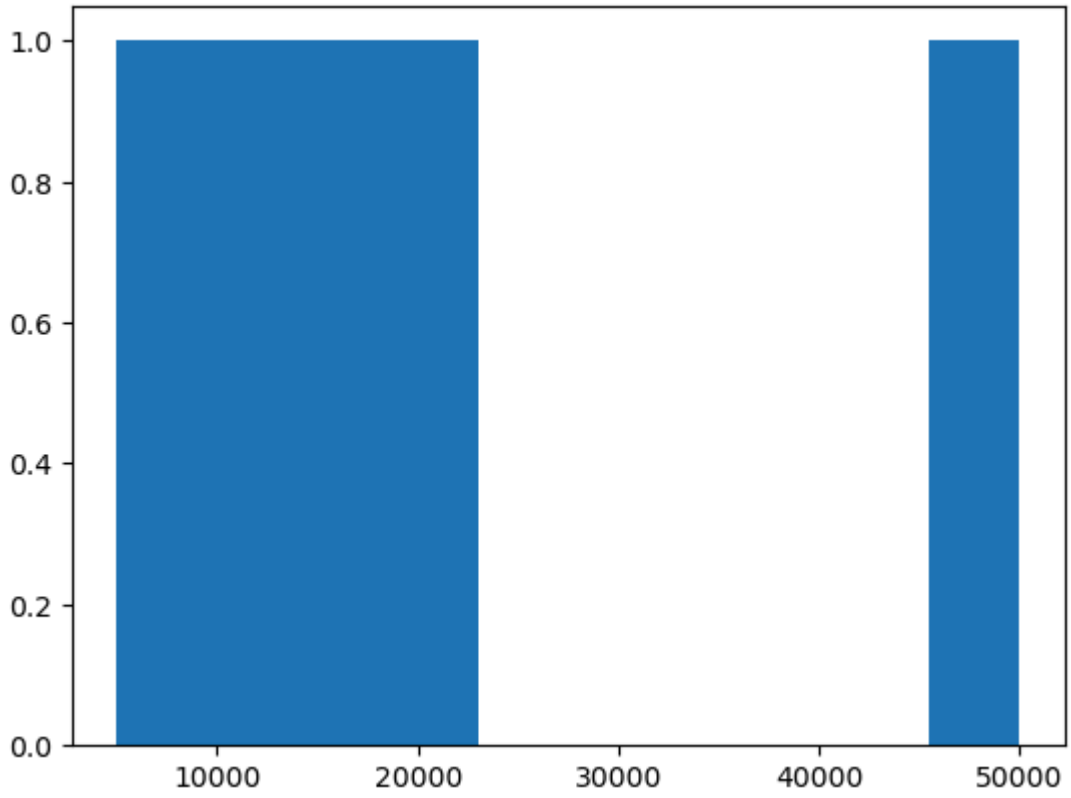
In [42]:
```
vis2 = sns.distplot(emp['SALARY'])
```

```
C:\Users\swati\AppData\Local\Temp\ipykernel_21220\826855712.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'])
```



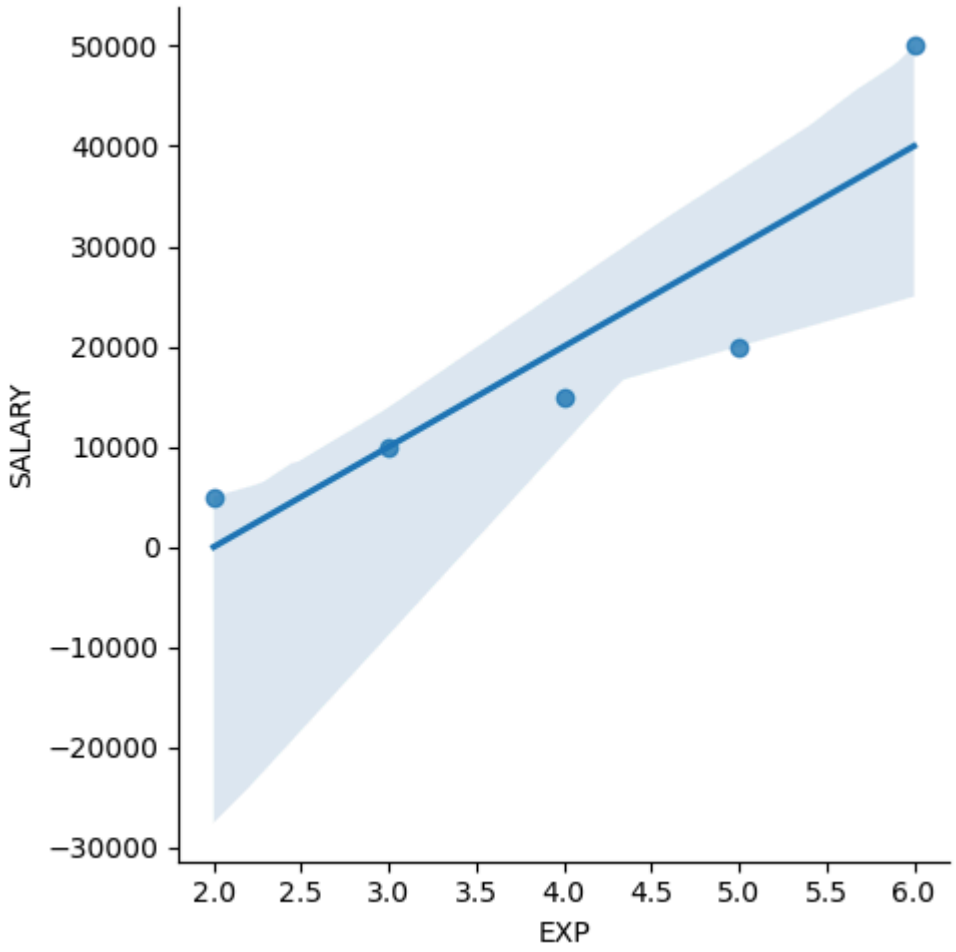
- plot the graph using 1 variable --univariate anaysis
- plot the graph using 2 variable--bivariate analysis
- plot the graph many variable--multivariate anaysis

In [48]:
```
vis3 = plt.hist(emp['SALARY'])
```

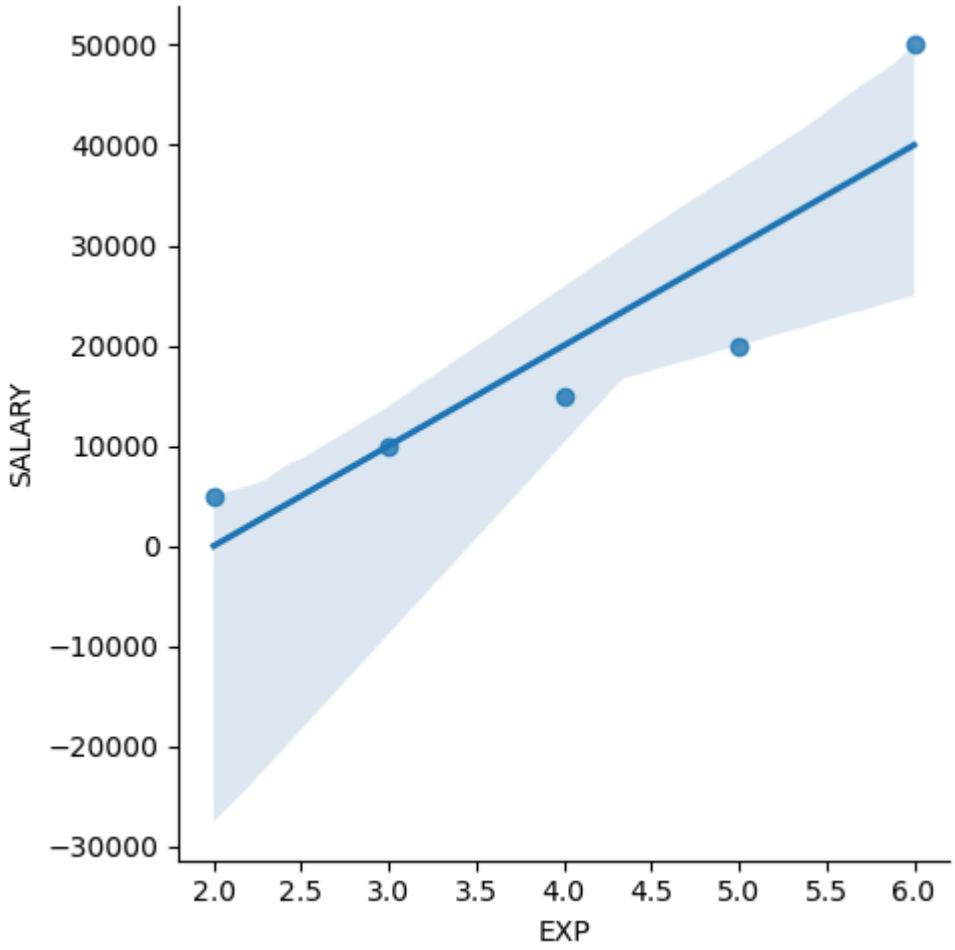


In [ ]:
```
plt.rcparams['figure.figsize'] = 10,3
```

In [54]:
```
vis5 = sns.lmplot(data=emp, x = 'EXP', y = 'SALARY')
```

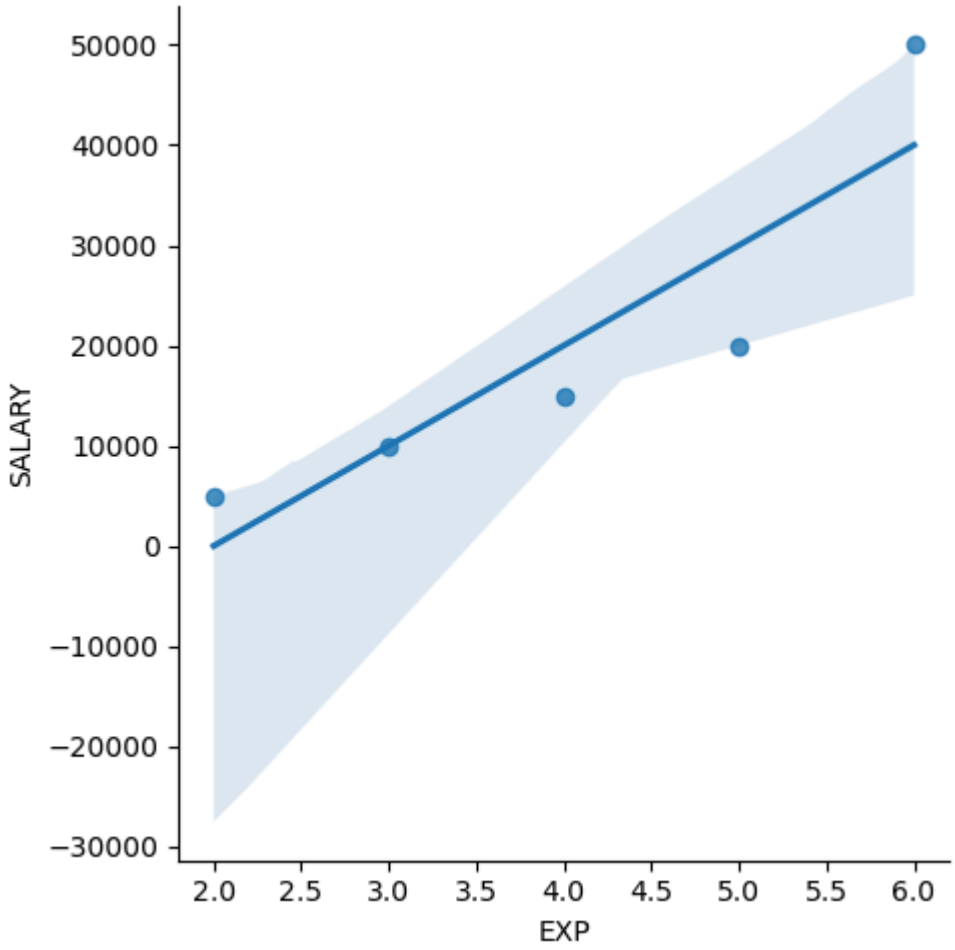


In [54]:
```
vis5 = sns.lmplot(data=emp, x = 'EXP', y = 'SALARY')
```

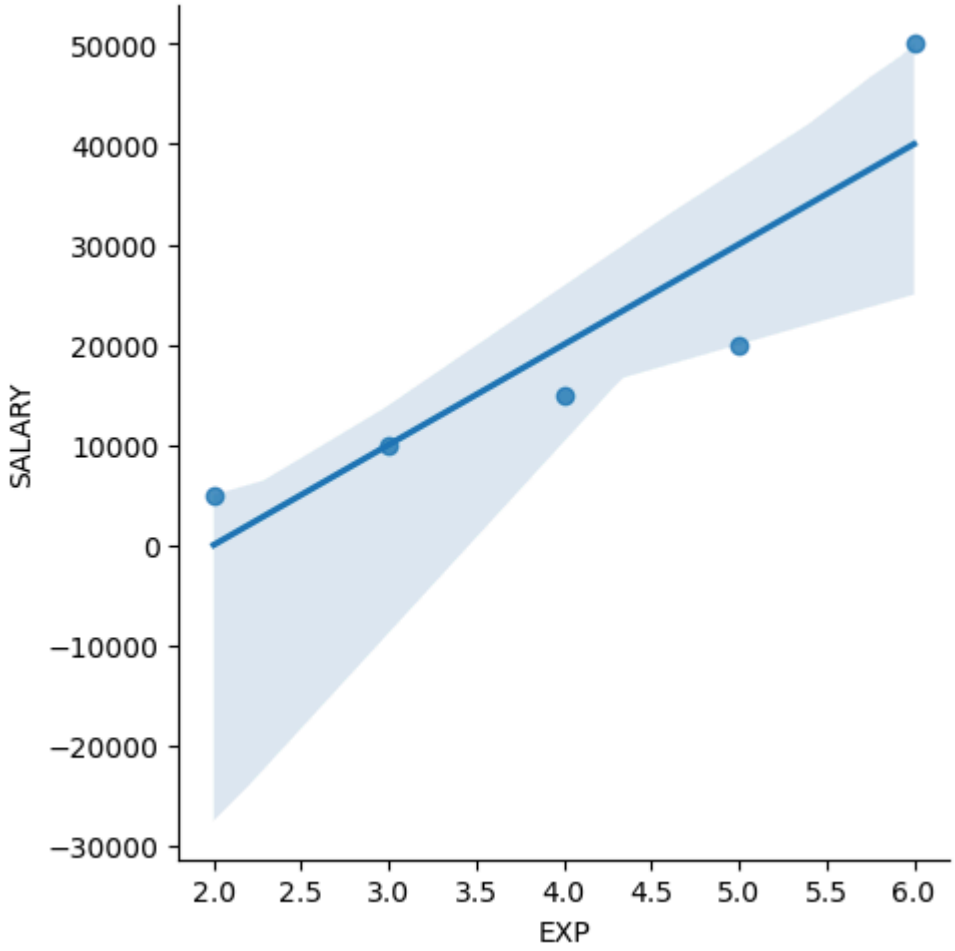


In [ ]:

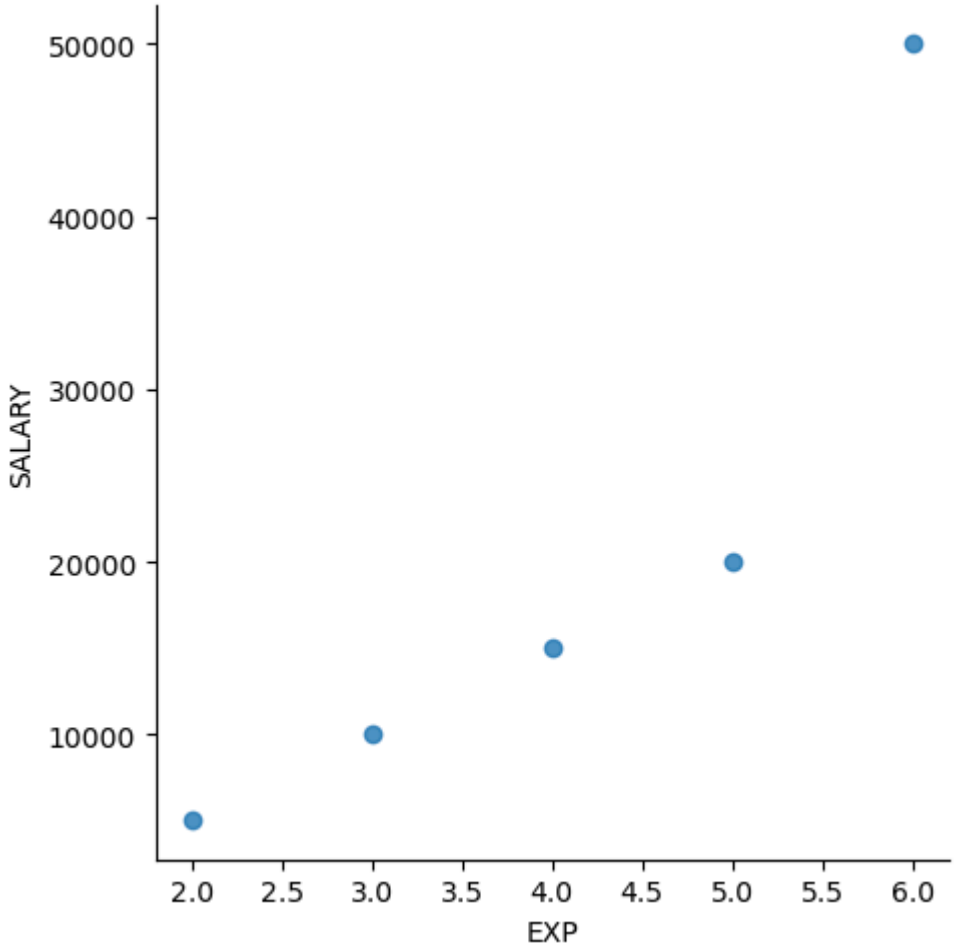
In [54]:
```
vis5 = sns.lmplot(data=emp, x = 'EXP', y = 'SALARY')
```



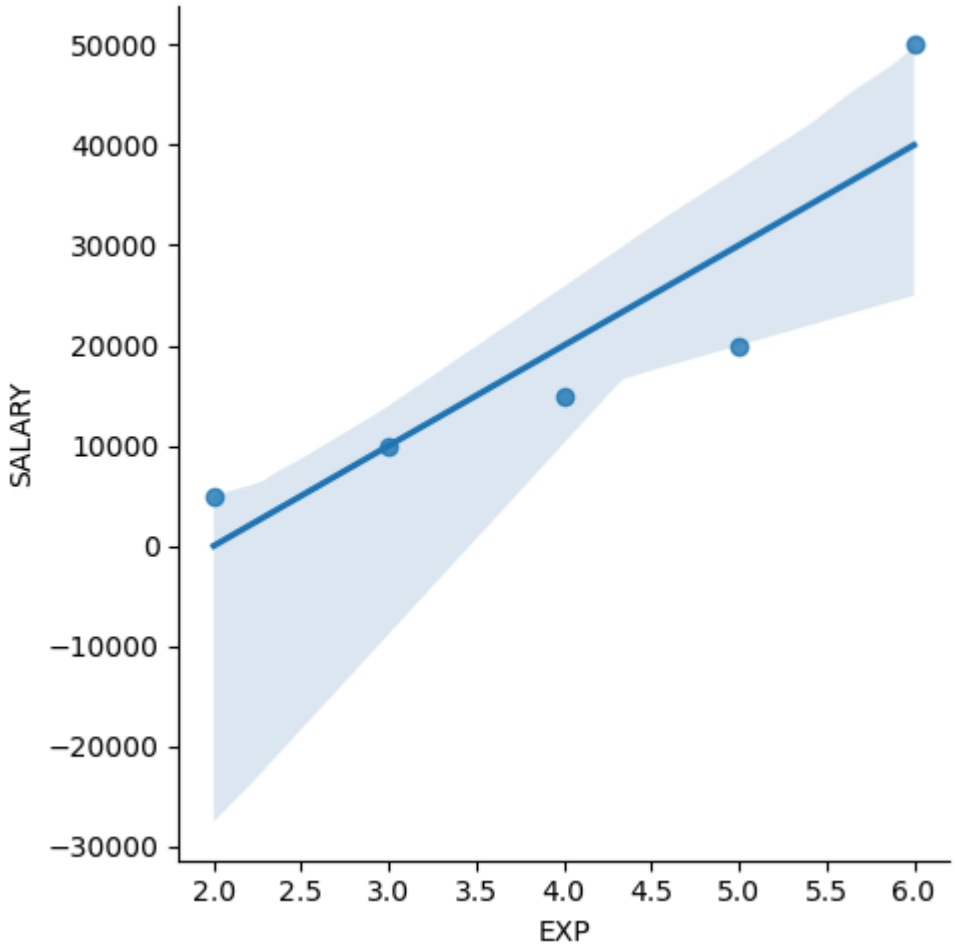
In [58]:
```
vis5 = sns.lmplot(data=emp, x = 'EXP', y = 'SALARY')
```



In [72]:
```
vis5 = sns.lmplot(data=emp, x = 'EXP', y = 'SALARY', fit_reg = False)
```



In [76]:
```
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 [ ]:

In [ ]:

In [ ]: