Mini project

Data Visualization project

In [5]:
```
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],
    ['VASU','JAVA',27,'DEHIL',40000,3],
    ['BARB','JAVA',30,'CHE',10000,4],
    ['CHERRY','C',35,'PUNE',15000,5],
    ['DIPAN','DA',38,'MUMBAI',20000,6],
    ['ESWAR','DS',40,'HYD',50000,7],
    ['KRISHNA','BA',25,'VIZ',30000,8],
    ['SAIKUMAR','JAVA',22,'HYD',50000,9],
    ['MADHU','C++',28,'BNG',60000,10]
]
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],
 ['VASU', 'JAVA', 27, 'DEHIL', 40000, 3],
 ['BARB', 'JAVA', 30, 'CHE', 10000, 4],
 ['CHERRY', 'C', 35, 'PUNE', 15000, 5],
 ['DIPAN', 'DA', 38, 'MUMBAI', 20000, 6],
 ['ESWAR', 'DS', 40, 'HYD', 50000, 7],
 ['KRISHNA', 'BA', 25, 'VIZ', 30000, 8],
 ['SAIKUMAR', 'JAVA', 22, 'HYD', 50000, 9],
 ['MADHU', 'C++', 28, 'BNG', 60000, 10]]
```

In [7]:
```
import os
os.getcwd()
```

Out[7]:
```
'C:\\Users\\harisai\\Downloads\\4th, 5th - Basic Python\\4th, 5th - Basic Pytho
n'
```

In [8]:
```
emp = pd.read_excel(r'C:\\Users\\harisai\\Downloads\\4th, 5th - Basic Python\\4t
emp
```

Out[8]:

	NAME	DOMAIN	AGE	LOCATION	SALARY	EXP
0	ALEX	TESTING	25	BNG	5000	2
1	VASU	JAVA	27	DEHIL	40000	3
2	BARB	JAVA	30	CHE	10000	4
3	CHERRY	C	35	PUNE	15000	5
4	DIPAN	DA	38	MUMBAI	20000	6
5	ESWAR	DS	40	HYD	50000	7
6	KRISHNA	BA	25	VIZ	30000	8
7	SAIKUMAR	JAVA	22	HYD	50000	9
8	MADHU	C++	28	BNG	60000	10

In [9]:
```
emp.shape
```

Out[9]:
```
(9, 6)
```

In [10]:
```
emp.columns
```

Out[10]:
```
Index(['NAME', 'DOMAIN', 'AGE', 'LOCATION', 'SALARY', 'EXP'], dtype='object')
```

In [11]:
```
len(emp.columns)
```

Out[11]:
```
6
```

In [12]:
```
len(emp)
```

Out[12]:
```
9
```

In [13]:
```
emp
```

Out[13]:

	NAME	DOMAIN	AGE	LOCATION	SALARY	EXP
0	ALEX	TESTING	25	BNG	5000	2
1	VASU	JAVA	27	DEHIL	40000	3
2	BARB	JAVA	30	CHE	10000	4
3	CHERRY	C	35	PUNE	15000	5
4	DIPAN	DA	38	MUMBAI	20000	6
5	ESWAR	DS	40	HYD	50000	7
6	KRISHNA	BA	25	VIZ	30000	8
7	SAIKUMAR	JAVA	22	HYD	50000	9
8	MADHU	C++	28	BNG	60000	10

In [14]:
```
emp['SALARY']
```

Out[14]:
```
0     5000
1    40000
2    10000
3    15000
4    20000
5    50000
6    30000
7    50000
8    60000
Name: SALARY, dtype: int64
```

In [16]:
```
emp[['AGE','LOCATION','NAME']]
```

Out[16]:

	AGE	LOCATION	NAME
0	25	BNG	ALEX
1	27	DEHIL	VASU
2	30	CHE	BARB
3	35	PUNE	CHERRY
4	38	MUMBAI	DIPAN
5	40	HYD	ESWAR
6	25	VIZ	KRISHNA
7	22	HYD	SAIKUMAR
8	28	BNG	MADHU

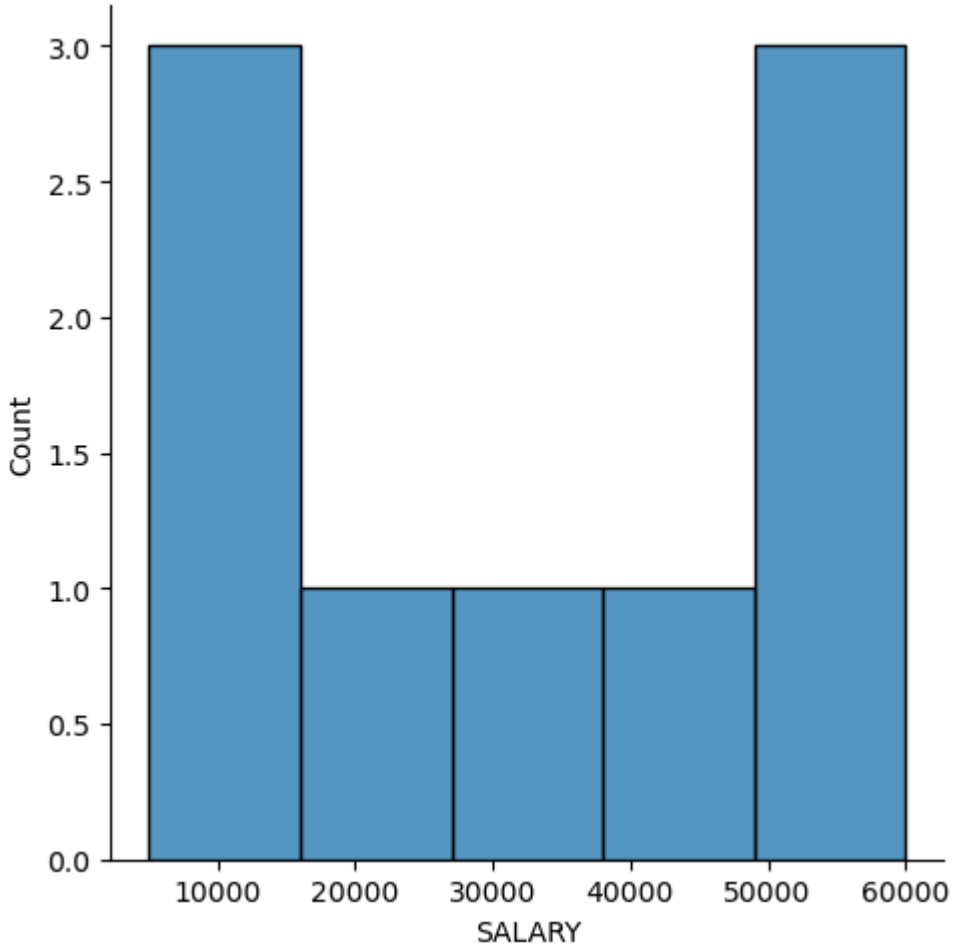
In [17]:
```
emp[['SALARY','EXP']]
```

Out[17]:

	SALARY	EXP
0	5000	2
1	40000	3
2	10000	4
3	15000	5
4	20000	6
5	50000	7
6	30000	8
7	50000	9
8	60000	10

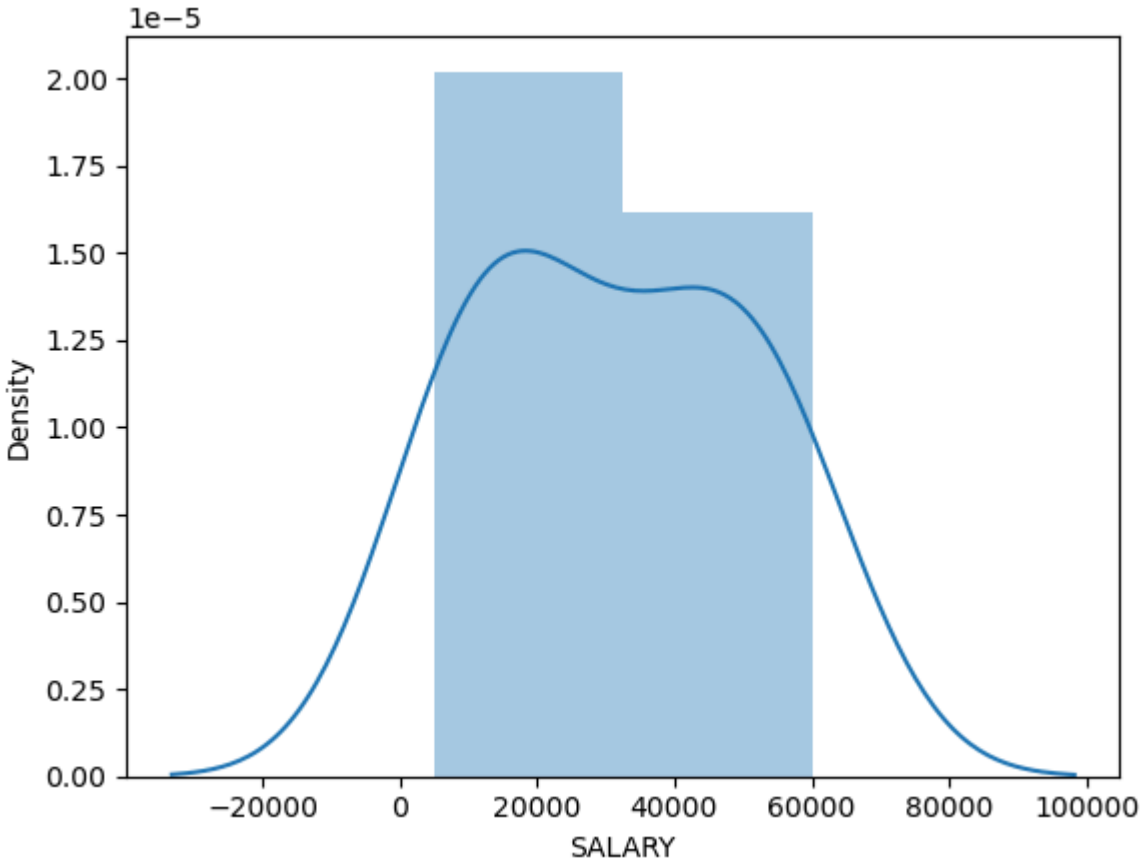
In [18]:
```
import numpy as np #ND ARRAY
import matplotlib.pyplot as plt # VISUALIZATION
import seaborn as sns # STATICES VISUALIZATION
```

In [19]:
```
vis1 = sns.displot(emp['SALARY'])
```

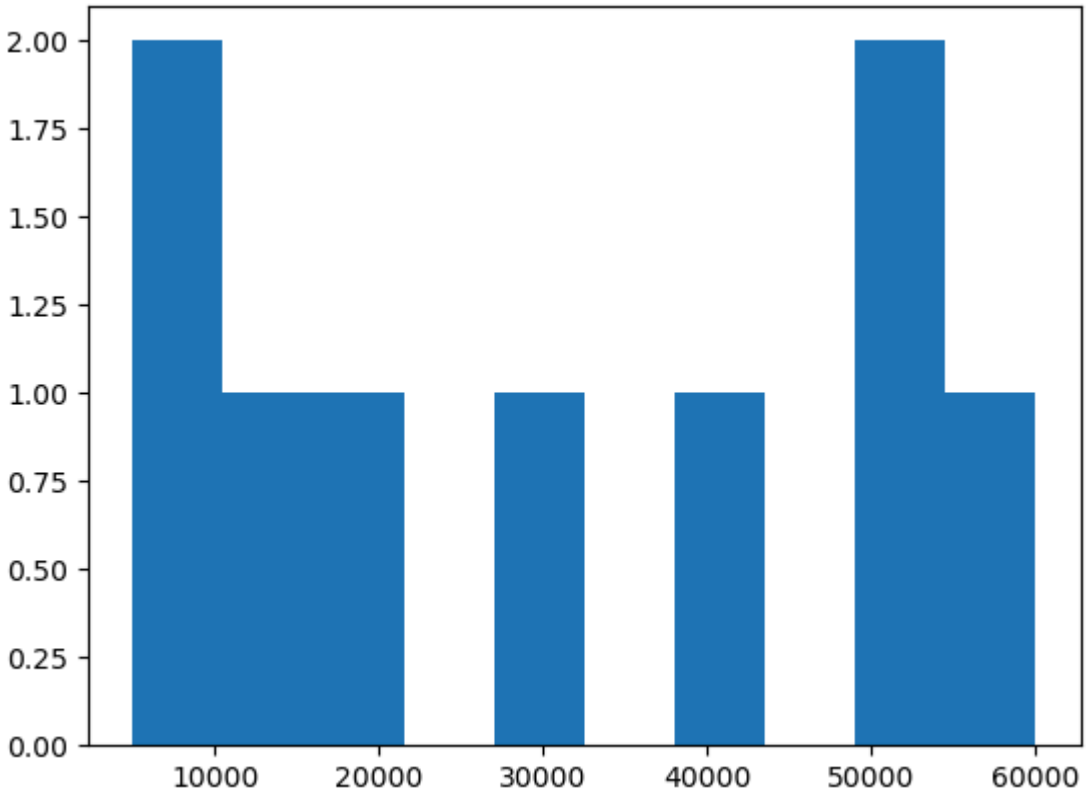


In [20]:
```
vis2 = sns.distplot(emp['SALARY'])
```

```
C:\Users\harisai\AppData\Local\Temp\ipykernel_8380\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'])
```

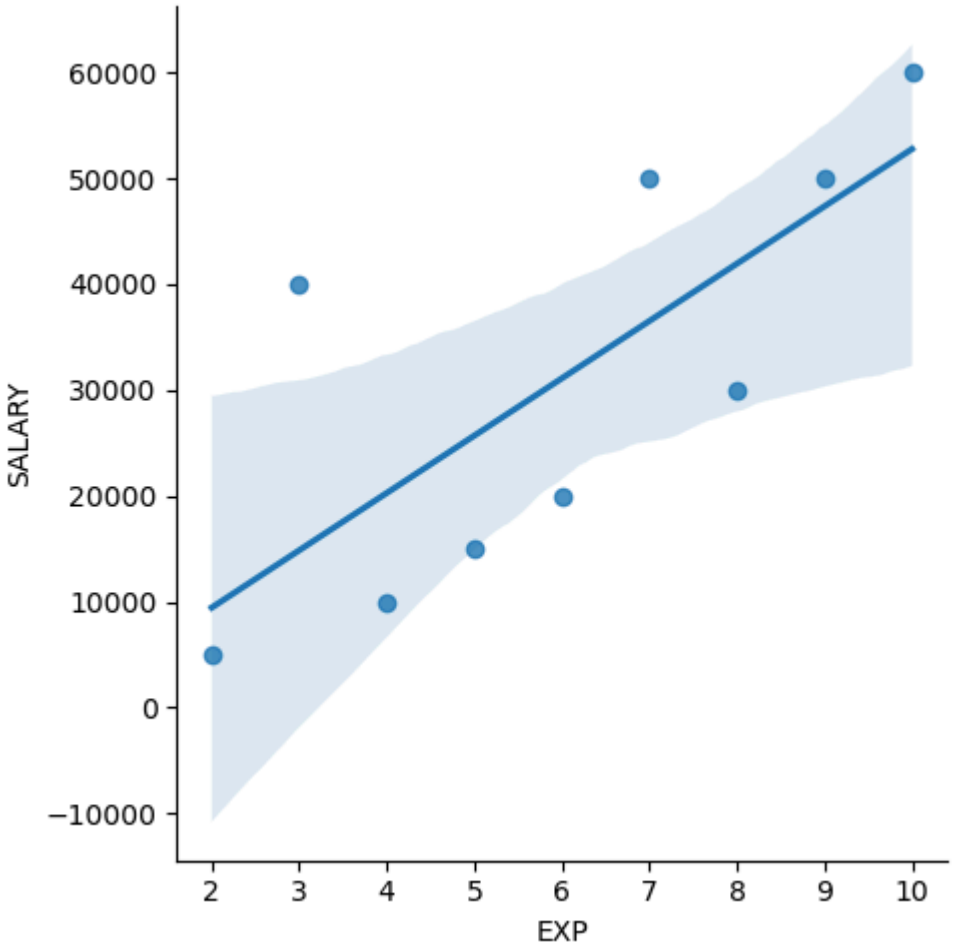


In [22]:
```
vis3 = plt.hist(emp['SALARY'])
```

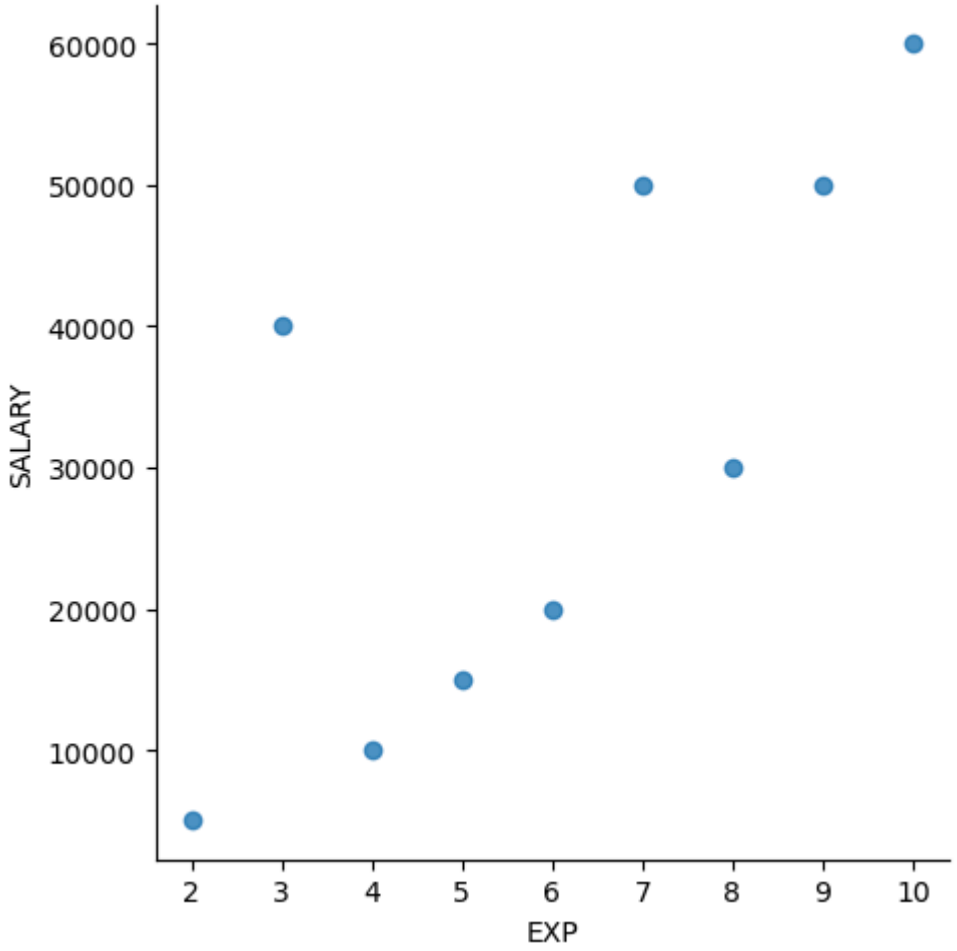


In [23]:
```
plt.rcParams['figure.figsize'] = 5,1
```

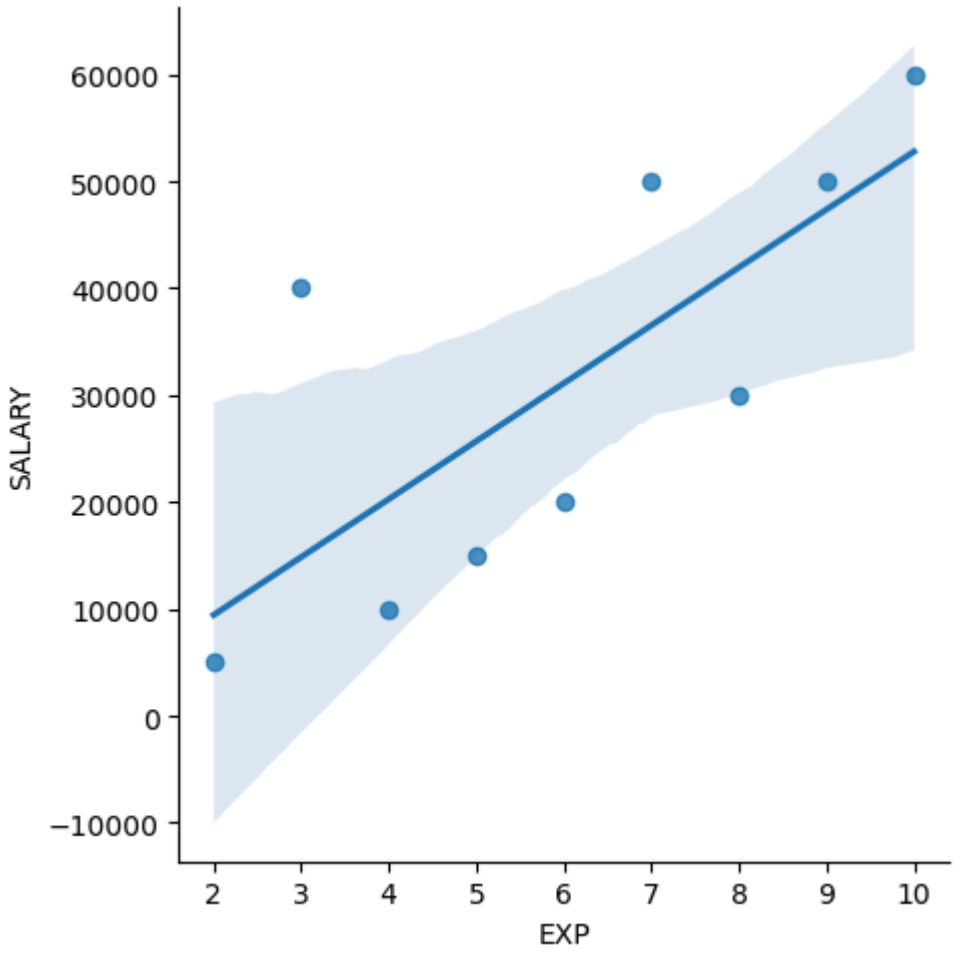
In [24]:
```
vis5 = sns.lmplot(data=emp, x ='EXP', y = 'SALARY')
```



In [25]:
```
vis5 = sns.lmplot(data=emp, x ='EXP', y = 'SALARY',fit_reg = False)
```



In [26]:
```
vis5 = sns.lmplot(data=emp, x ='EXP', y = 'SALARY',fit_reg = True)
```



In [ ]:
```
# mini project completed
```