```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

saldf=pd.read_csv(r"C:\my pythonfiles\Salary_EDA.csv")
saldf.head()
```

```
    Age  Gender Education Level          Job Title  Years of
Experience  \
0  32.0    Male      Bachelor's  Software Engineer
5.0
1  28.0  Female        Master's       Data Analyst
3.0
2  45.0    Male             PhD     Senior Manager
15.0
3  36.0  Female      Bachelor's    Sales Associate
7.0
4  36.0  Female      Bachelor's    Sales Associate
7.0

     Salary
0   90000.0
1   65000.0
2  150000.0
3   60000.0
4   60000.0
```

```
saldf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 375 entries, 0 to 374
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Age                  373 non-null    float64
 1   Gender               371 non-null    object
 2   Education Level      372 non-null    object
 3   Job Title            370 non-null    object
 4   Years of Experience  373 non-null    float64
 5   Salary               372 non-null    float64
dtypes: float64(3), object(3)
memory usage: 17.7+ KB
```

1. Age, Years of Experience, Salary are in float type
2. Gender, Education Level, Job Title are in object type
3. Null values exists
4. Total 6-features and 375 rows are observed

```
saldf.isnull().sum()
```

```
Age                    2
Gender                 4
Education Level        3
Job Title              5
Years of Experience    2
Salary                 3
dtype: int64
```

1. Null values are found in all 6 features

```
saldf.dropna(inplace=True)
saldf
```

```
      Age  Gender Education Level                         Job Title  \
0    32.0    Male      Bachelor's              Software Engineer
1    28.0  Female        Master's                   Data Analyst
2    45.0    Male             PhD                 Senior Manager
3    36.0  Female      Bachelor's                Sales Associate
4    36.0  Female      Bachelor's                Sales Associate
..    ...     ...             ...                            ...
370  35.0  Female      Bachelor's       Senior Marketing Analyst
371  43.0    Male        Master's         Director of Operations
372  29.0  Female      Bachelor's         Junior Project Manager
373  34.0    Male      Bachelor's  Senior Operations Coordinator
374  44.0  Female             PhD        Senior Business Analyst

     Years of Experience    Salary
0                    5.0   90000.0
1                    3.0   65000.0
2                   15.0  150000.0
3                    7.0   60000.0
4                    7.0   60000.0
..                   ...       ...
370                  8.0   85000.0
371                 19.0  170000.0
372                  2.0   40000.0
373                  7.0   90000.0
374                 15.0  150000.0

[366 rows x 6 columns]
```

```
saldf.isnull().sum()
```

```
Age                    0
Gender                 0
Education Level        0
Job Title              0
Years of Experience    0
Salary                 0
dtype: int64
```

1. Rows containing null values are deleted using "dropna" function
2. 366 rows are obtained after dropping all the null values.

```
saldf.describe(include='all')
```

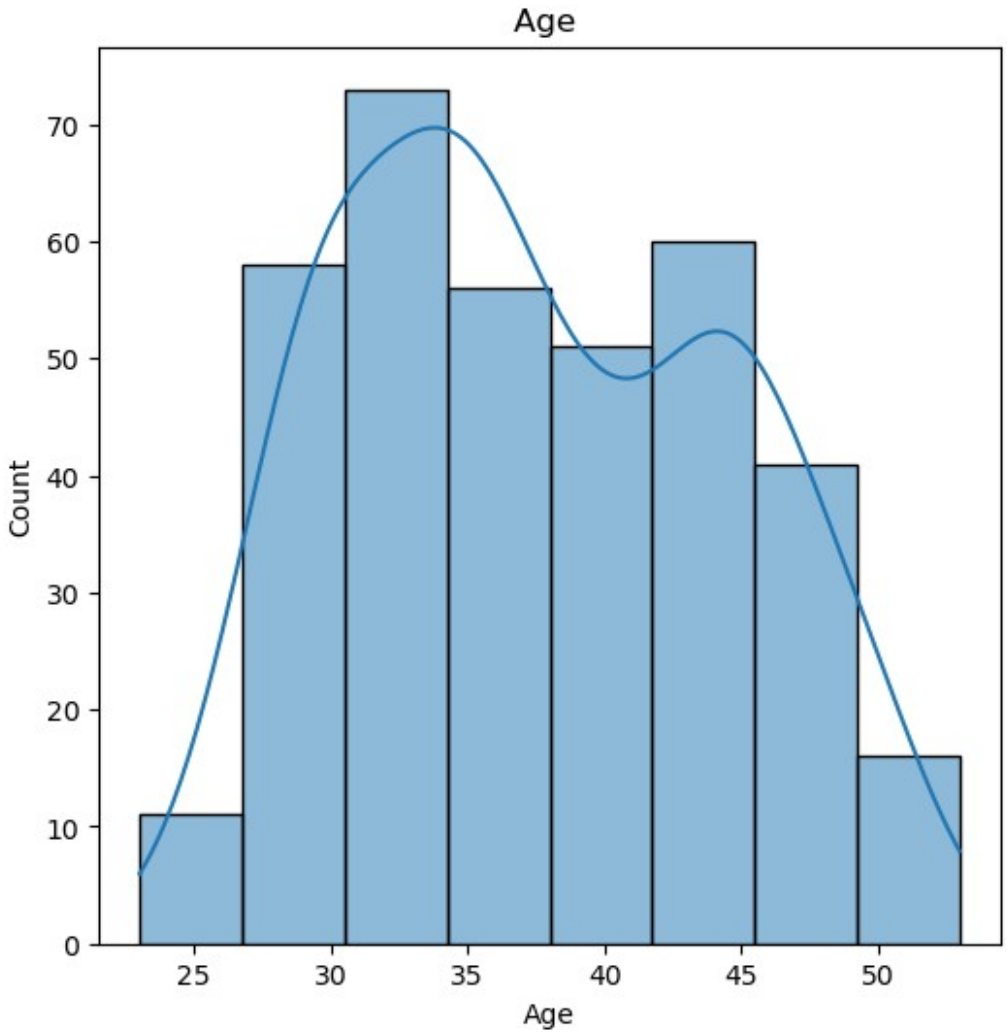
```
                Age Gender Education Level              Job Title  \
count   366.000000    366             366                    366   
unique         NaN      2               3                    169   
top            NaN   Male      Bachelor's  Director of Marketing   
freq           NaN    189             220                     12   
mean     37.459016    NaN             NaN                    NaN   
std       6.962303    NaN             NaN                    NaN   
min      23.000000    NaN             NaN                    NaN   
25%      32.000000    NaN             NaN                    NaN   
50%      36.000000    NaN             NaN                    NaN   
75%      44.000000    NaN             NaN                    NaN   
max      53.000000    NaN             NaN                    NaN   

        Years of Experience         Salary
count            366.000000     366.000000
unique                  NaN            NaN
top                     NaN            NaN
freq                    NaN            NaN
mean              10.045082  100492.759563
std                6.517102   48013.732434
min                0.000000     350.000000
25%                4.000000   56250.000000
50%                9.000000   95000.000000
75%               15.000000  140000.000000
max               25.000000  250000.000000
```

1. Total rows are counted amd given as count of all 6 features.
2. Average age is 37.45,minimum age of getting job is 23,maximum age of getting job is 53 -The range of age who getting job is between 32 - 44 -few entries from 50s
3. Two unique values are found in gender -male,female -We can observe that males are slightly dominating the females.
4. Majority of employees are recruited on there Bachelor's degree.
5. If we see of job title the Director of Marketing are leading with 12 posts
6. Average expirience in job is 10 years,minimum expirience in job is 0,maximum expirience in job is 25 -The range of expirience in job is between 32 - 44
7. Average salary is 100492 ,minimum salary is 350,maximum salary is 250000 The range of salary lies between 350 - 250000

```
plt.figure(figsize=(6,6))
sns.histplot(saldf['Age'],kde=True,bins=8)
plt.title('Age ')
plt.show()
```

```
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
```

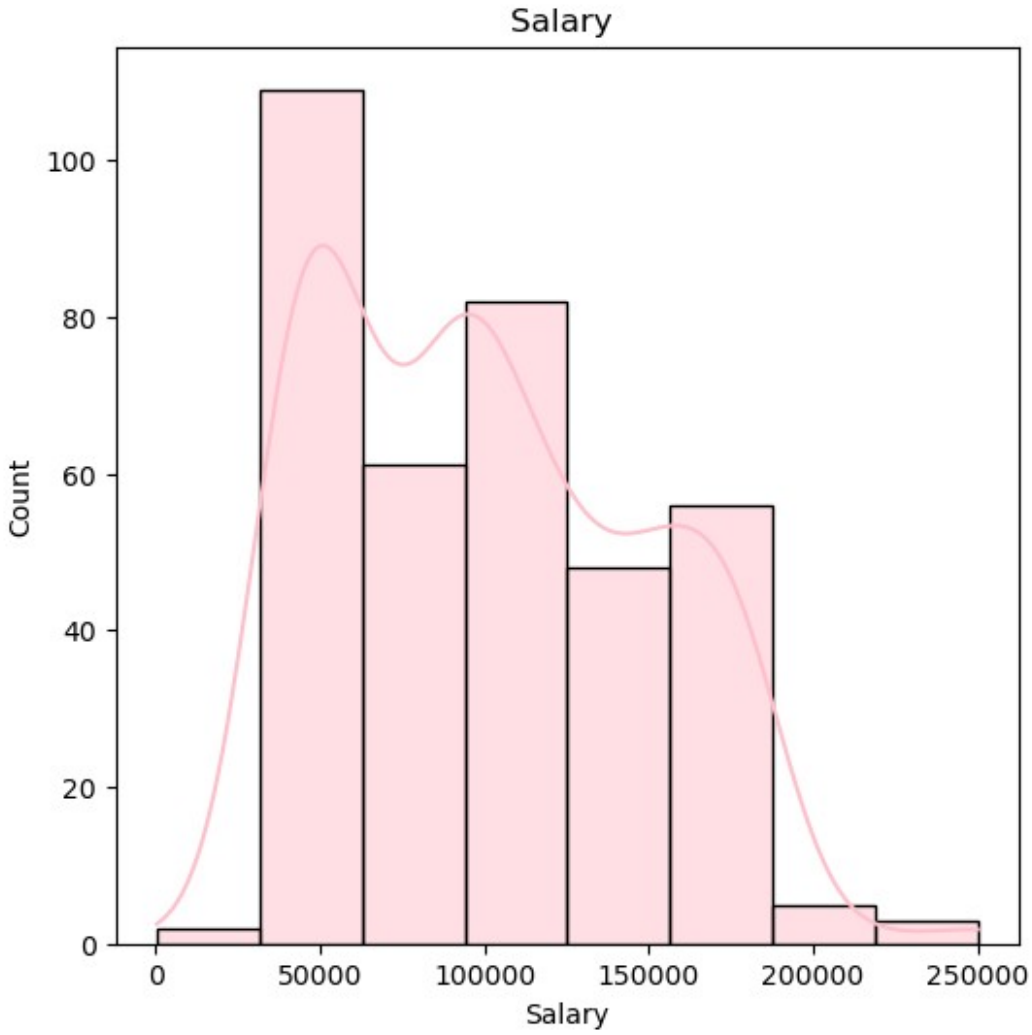
```
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



1. Average age lies between 30-35
2. minimum age of getting job is 23,
3. maximum age of getting job is 53 -The range of age who getting job is between 32 - 44 - few entries from 50s

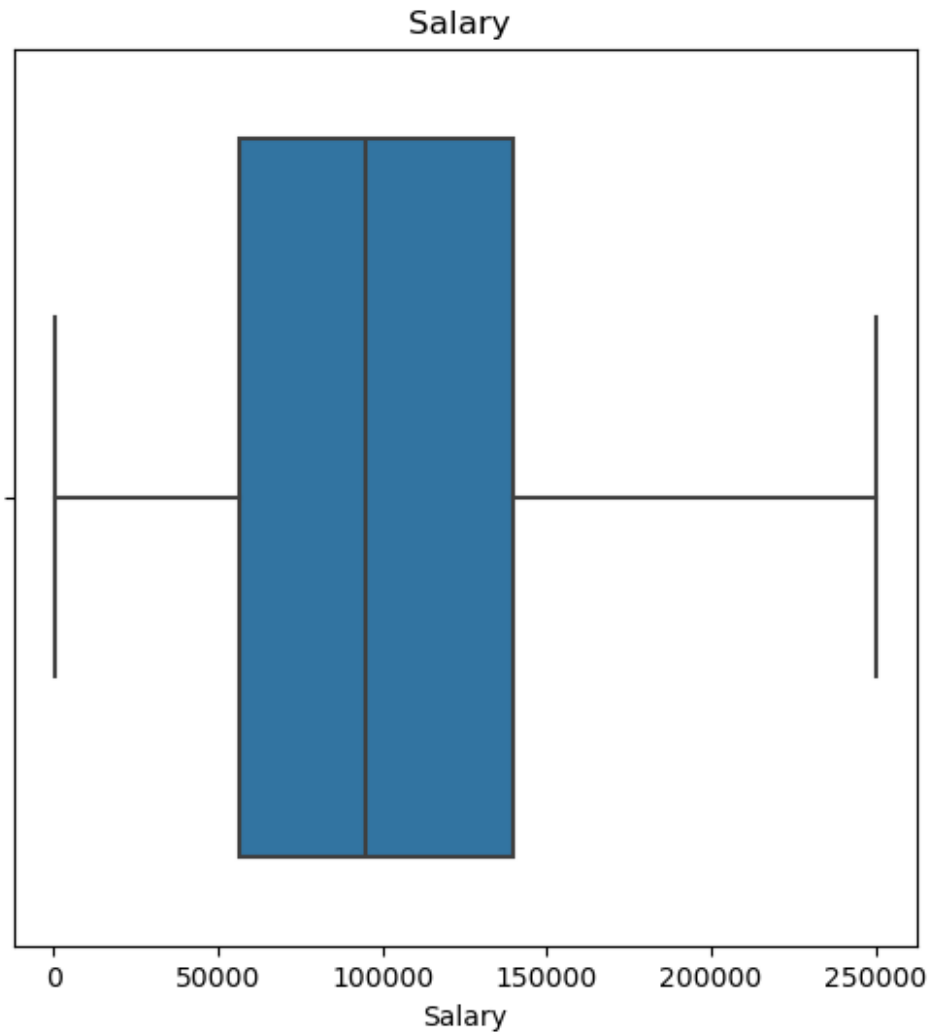
```
analyse the dist of sal using histo

plt.figure(figsize=(6,6))
sns.histplot(saldf['Salary'],kde=True,bins=8,color='pink')
plt.title('Salary ')
plt.show()

C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
```

```
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

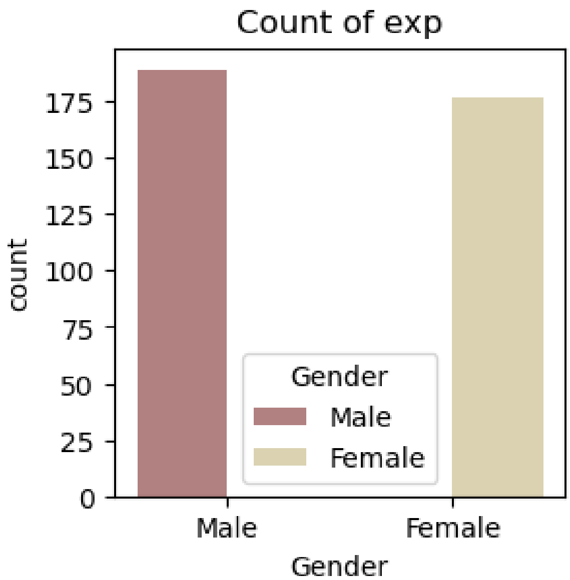


1. The maximum

```
analyse salary distribution using the boxplot

  Cell In[37], line 1
    analyse salary distribution using the boxplot
            ^
SyntaxError: invalid syntax


plt.figure(figsize=(6,6))
sns.boxplot(x=saldf['Salary'])
plt.title('Salary ')
plt.show()
```




```
find the corelation matrix


draw count plot for feature grnder

plt.figure(figsize=(3,3))
sns.countplot(x=saldf['Gender'],palette='pink',hue=saldf['Gender'])
plt.title('Count of exp')

Text(0.5, 1.0, 'Count of exp')
```

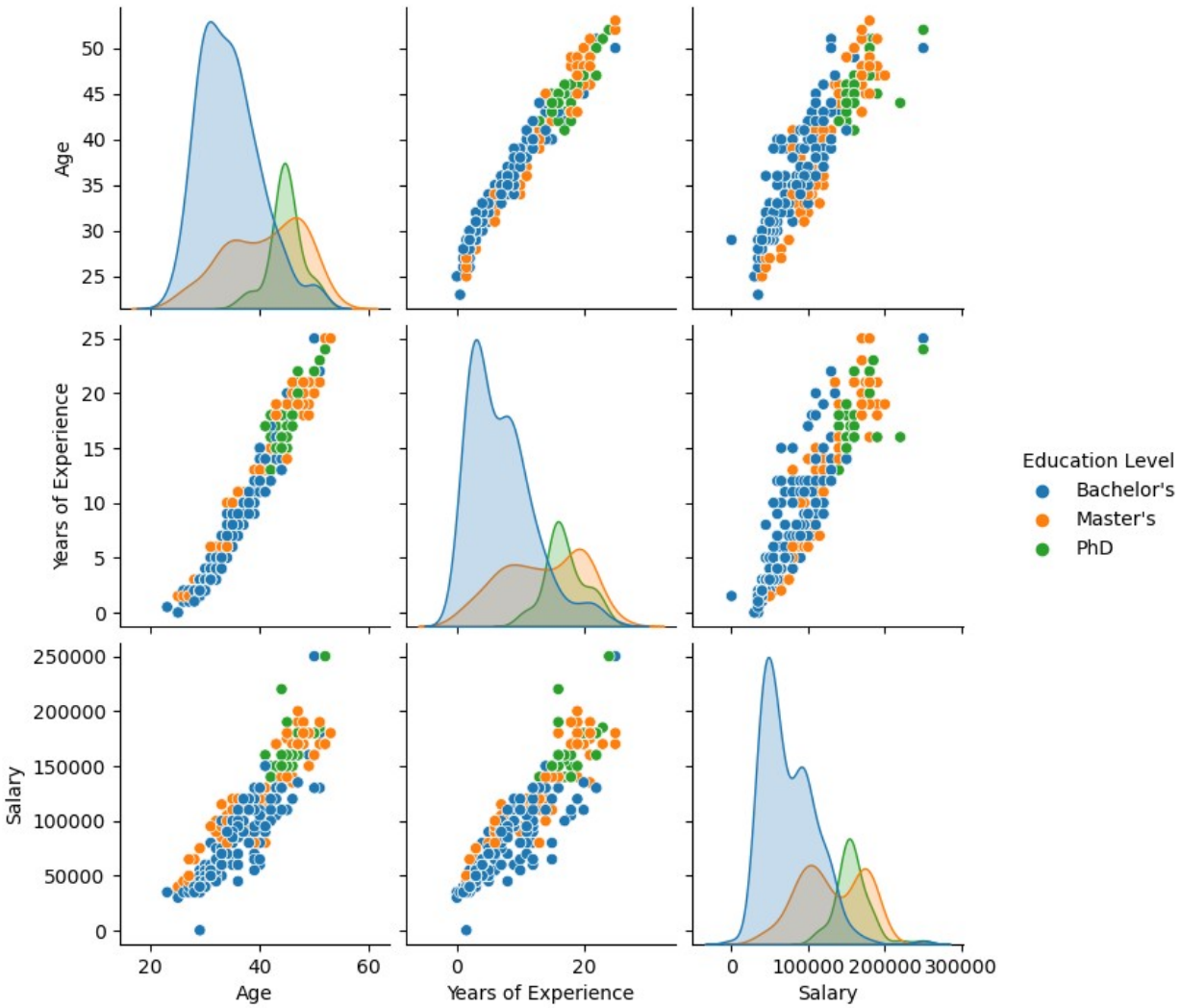



```
     Education Level
plt.figure(figsize=(3,3))
sns.countplot(x=saldf['Education
Level'],palette='k',hue=saldf['Education Level'])
plt.title('Count of exp')
```

```
  Cell In[46], line 1
    Education Level
              ^
SyntaxError: invalid syntax
```

```
sns.pairplot(saldf,hue='Education Level')
```

```
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):
C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1119:
FutureWarning: use_inf_as_na option is deprecated and will be removed
in a future version. Convert inf values to NaN before operating
instead.
  with pd.option_context('mode.use_inf_as_na', True):

<seaborn.axisgrid.PairGrid at 0x26117bb84d0>
```



graph has been plotted in every corner phd and bacholers have more salery bacholers have more experince bachelors has more mejority and with more salary AND EXPERIENCE WE OBSERVE THAT WHEN AGE INCREASES YEARS OF EXP INCREASE the peak sal are given to bacholer degree people bacholers are consistants in the job sal is also effected by the years of experience

```
gour edcn level and find avg sal in every catogaryfind avg salary on
that dataset
filter data set in which exp is more than 20 years and find avg sal on
dataset

saldf.groupby('Education Level')['Salary'].mean()
```

```
Education Level
Bachelor's     74683.409091
Master's      129473.684211
PhD           157843.137255
Name: Salary, dtype: float64
```

```
filt_exp=(saldf[saldf['Years of Experience']>20])['Salary'].mean()
filt_exp
```

```
175892.85714285713
```

```
filt_exp=(saldf[saldf['Years of Experience']>20])
filt_exp
```

```
      Age  Gender Education Level                         Job Title  \
19   51.0    Male      Bachelor's                  Sales Director
30   50.0    Male      Bachelor's                             CEO
39   49.0    Male      Bachelor's                 Sales Executive
50   51.0  Female      Bachelor's         Customer Service Manager
60   51.0  Female        Master's           Director of Operations
63   47.0    Male             PhD            Senior Data Scientist
76   50.0  Female      Bachelor's              Operations Manager
83   52.0    Male             PhD         Chief Technology Officer
88   46.0    Male        Master's           Senior Project Manager
93   52.0  Female        Master's         Senior Marketing Manager
96   47.0    Male             PhD               Research Scientist
112  50.0  Female      Bachelor's             Supply Chain Analyst
121  53.0    Male        Master's            Director of Marketing
130  50.0  Female        Master's           Director of Operations
133  47.0    Male             PhD        Senior Research Scientist
158  51.0  Female             PhD      Director of Human Resources
161  48.0    Male        Master's   Director of Product Management
188  50.0  Female             PhD  Director of Sales and Marketing
200  49.0  Female        Master's        Director of Human Capital
217  50.0  Female             PhD           Director of Operations
229  51.0  Female        Master's      Director of Human Resources
246  49.0  Female        Master's            Director of Marketing
258  50.0  Female             PhD           Director of Operations
279  49.0  Female        Master's           Director of Operations
291  49.0    Male             PhD           Director of Operations
306  49.0  Female        Master's            Director of Marketing
329  48.0    Male        Master's            Director of Marketing
353  48.0    Male        Master's            Director of Marketing

     Years of Experience    Salary
19                  22.0  180000.0
30                  25.0  250000.0
39                  21.0  160000.0
50                  22.0  130000.0
60                  23.0  170000.0
63                  21.0  180000.0
76                  22.0  160000.0
83                  24.0  250000.0
88                  21.0  135000.0
93                  25.0  170000.0
96                  22.0  160000.0
```

```
112                  22.0  130000.0
121                  25.0  180000.0
130                  21.0  160000.0
133                  22.0  160000.0
158                  23.0  185000.0
161                  21.0  175000.0
188                  22.0  180000.0
200                  21.0  180000.0
217                  22.0  180000.0
229                  21.0  190000.0
246                  21.0  180000.0
258                  22.0  180000.0
279                  21.0  180000.0
291                  21.0  180000.0
306                  21.0  180000.0
329                  21.0  180000.0
353                  21.0  180000.0
```

```
fmaster=saldf[(saldf['Gender'] =='Female')&(saldf['Education
Level']=="Master's")]
fmaster
```

```
      Age  Gender Education Level                         Job
Title  \
1    28.0  Female        Master's                  Data Analyst

6    42.0  Female        Master's                Product Manager

13   40.0  Female        Master's                Project Manager

16   33.0  Female        Master's               Marketing Manager

26   37.0  Female        Master's                Software Manager

40   34.0  Female        Master's                    UX Designer

47   45.0  Female        Master's           Director of Marketing

56   27.0  Female        Master's                  UX Researcher

60   51.0  Female        Master's         Director of Operations

65   38.0  Female        Master's      Digital Marketing Manager

69   49.0  Female        Master's        Senior Financial Analyst

74   42.0  Female        Master's               Creative Director

78   48.0  Female        Master's        Human Resources Director

81   41.0  Female        Master's                  Data Analyst
```

85	34.0	Female	Master's	Financial Advisor
93	52.0	Female	Master's	Senior Marketing Manager
98	38.0	Female	Master's	Public Relations Manager
102	49.0	Female	Master's	Senior HR Manager
104	39.0	Female	Master's	Senior Project Coordinator
108	41.0	Female	Master's	Senior Marketing Analyst
110	42.0	Female	Master's	Senior Graphic Designer
126	37.0	Female	Master's	Senior HR Generalist
130	50.0	Female	Master's	Director of Operations
132	40.0	Female	Master's	Senior Training Specialist
139	43.0	Female	Master's	Senior Product Marketing Manager
141	41.0	Female	Master's	Senior Marketing Manager
144	25.0	Female	Master's	Junior Marketing Specialist
154	37.0	Female	Master's	Senior Marketing Analyst
164	41.0	Female	Master's	Senior Human Resources Manager
170	50.0	Female	Master's	Director of Finance
174	26.0	Female	Master's	Junior Data Scientist
186	33.0	Female	Master's	Senior Financial Analyst
194	40.0	Female	Master's	Senior Human Resources Specialist
200	49.0	Female	Master's	Director of Human Capital
203	27.0	Female	Master's	Junior UX Designer
215	34.0	Female	Master's	Senior Financial Advisor
223	42.0	Female	Master's	Senior Human Resources Manager
229	51.0	Female	Master's	Director of Human Resources
232	27.0	Female	Master's	Junior Research Scientist
246	49.0	Female	Master's	Director of Marketing

```
256  34.0  Female        Master's            Senior Financial Advisor
267  44.0  Female        Master's                Senior HR Specialist
273  47.0  Female        Master's               Director of Marketing
279  49.0  Female        Master's              Director of Operations
294  45.0  Female        Master's                   Senior HR Manager
300  48.0  Female        Master's                      Director of HR
306  49.0  Female        Master's               Director of Marketing
316  34.0  Female        Master's            Senior Financial Advisor
320  45.0  Female        Master's            Senior Marketing Manager
```

	Years of Experience	Salary
1	3.0	65000.0
6	12.0	120000.0
13	14.0	130000.0
16	7.0	90000.0
26	11.0	110000.0
40	5.0	80000.0
47	16.0	180000.0
56	2.0	65000.0
60	23.0	170000.0
65	10.0	90000.0
69	18.0	150000.0
74	14.0	120000.0
78	20.0	180000.0
81	13.0	80000.0
85	10.0	95000.0
93	25.0	170000.0
98	10.0	90000.0
102	19.0	150000.0
104	13.0	80000.0
108	14.0	100000.0
110	15.0	110000.0
126	9.0	95000.0
130	21.0	160000.0
132	12.0	100000.0
139	14.0	120000.0
141	13.0	110000.0
144	1.5	40000.0
154	9.0	95000.0
164	13.0	120000.0

```
170                  20.0  180000.0
174                   1.5   45000.0
186                   6.0   95000.0
194                  13.0  120000.0
200                  21.0  180000.0
203                   1.5   45000.0
215                   6.0  100000.0
223                  13.0  140000.0
229                  21.0  190000.0
232                   1.5   50000.0
246                  21.0  180000.0
256                   6.0  100000.0
267                  15.0  140000.0
273                  20.0  180000.0
279                  21.0  180000.0
294                  14.0  140000.0
300                  20.0  180000.0
306                  21.0  180000.0
316                   6.0   80000.0
320                  16.0  160000.0
```

```
fmaster['Salary'].mean()
```

```
121020.40816326531
```

```
saldf.groupby('Education Level').agg({'Age':['count','mean']})
```

	Age	
	count	mean
Education Level		
Bachelor's	220	34.368182
Master's	95	40.715789
PhD	51	44.725490