seaborn

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

mydata = { "Name" : ["Ram","Sam","Joe","Asha"],
           "Age" : [23,22,26,47],
           "salary" : [15000,20000,32000,49000],
           "Exp" : [2,7,3,10]
         }

df = pd.DataFrame(mydata)
df.head()
```

```
    Name  Age  salary  Exp
0   Ram   23   15000    2
1   Sam   22   20000    7
2   Joe   26   32000    3
3   Asha  47   49000   10
```

histogram

```python
plt.figure(figsize = (6,5))
sns.histplot(df["salary"],kde = True, bins =2)
plt.title("Distribution of 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):
```

Distribution of salary

1. postitve skew,large salary value detected
2. majarity is between 15000 to 30000
3. average salary s 20k
4. no otliners detected

correlation Matrix(heatmap):

```python
#step 1: filter numerical data
ndf = df.select_dtypes(include = ["number"])
ndf.head()

    Age   salary   Exp
0    23    15000     2
1    22    20000     7
2    26    32000     3
3    47    49000    10

plt.figure(figsize = (6,5))
sns.heatmap(ndf.corr(),cmap = "cividis",annot = True)
plt.title("correation matrix between sal ,age, exp")
plt.show()
```
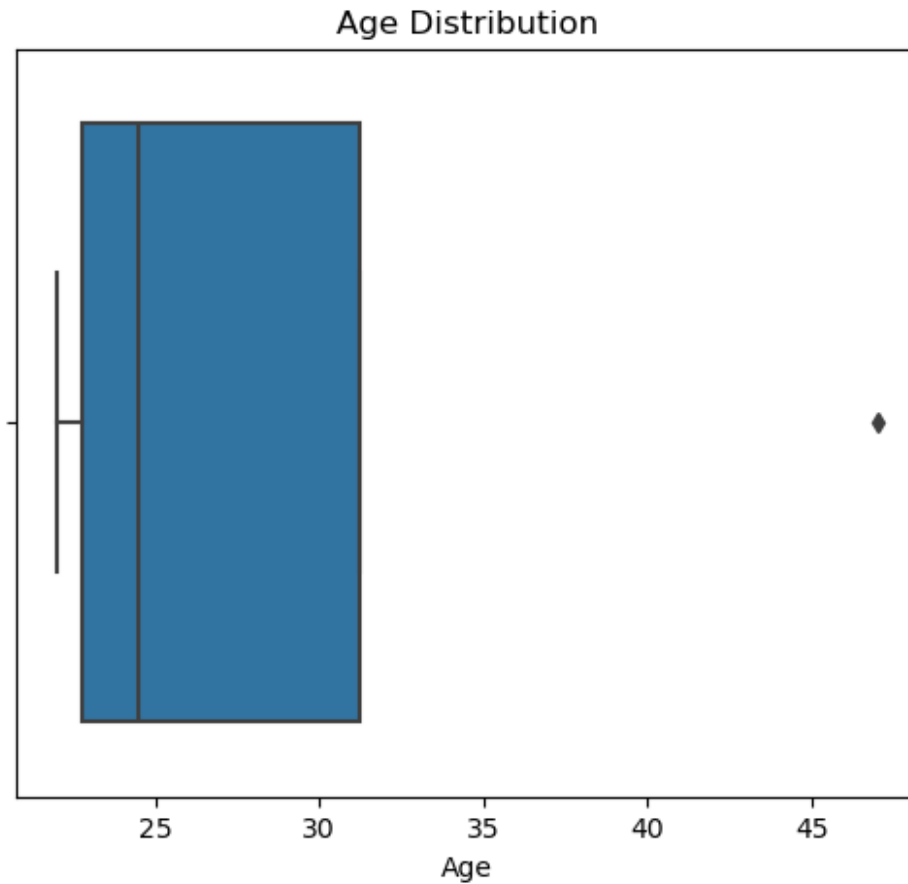
correation matrix between sal ,age, exp
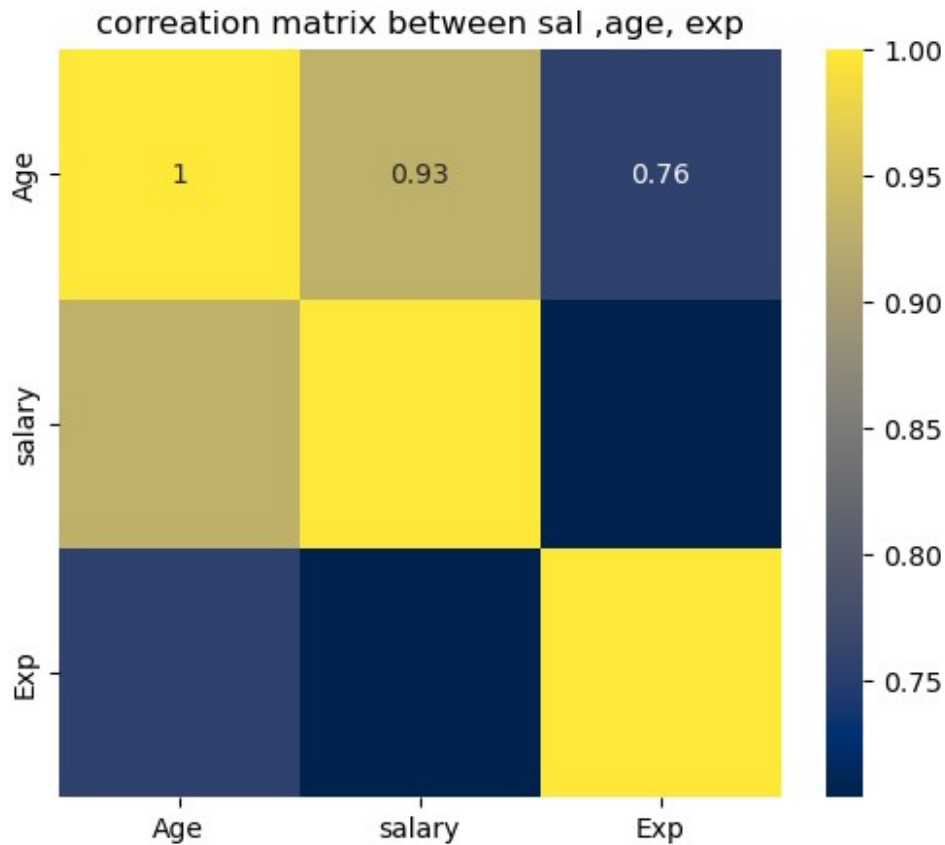
1.it shows the correlation btw age sal and exp 2.there is a postivie correlation btw age and sal 3.the correlation btw age and exp is less

box plot

```python
plt.figure(figsize = (6,5))
sns.boxplot(x=df["Age"])
plt.title("Age Distribution")
plt.show()
```

Age Distribution

1. abnormal oytlier is around 45
2. the average age is around 25

Find the Outliers in the following data : temp = [21, 47, 39,22 , 31, 33, 29, 26, 27, 25, 49, 46] using a box plot

```
temp : [21, 47, 39,22 , 31, 33, 29, 26, 27, 25, 49, 46]
plt.figure(figsize = (6,5))
sns.boxplot(x=temp)
plt.title(" Temp Distribution")
plt.show()
```

Temp Distribution
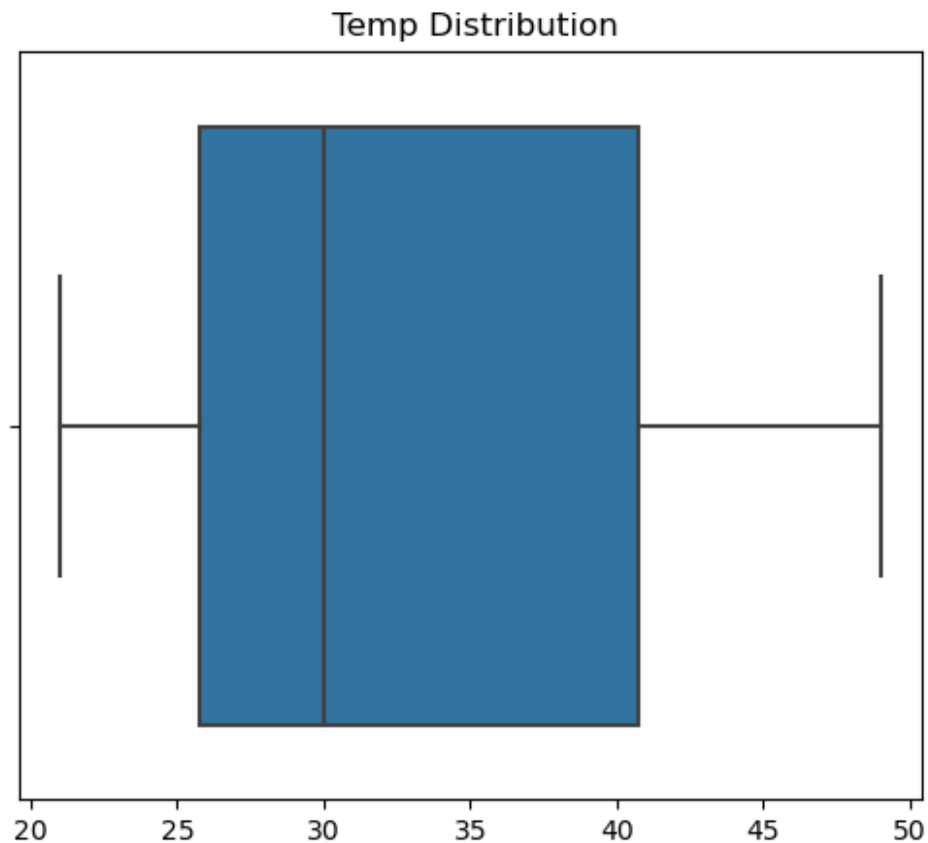
the average is around 30 no abnormal
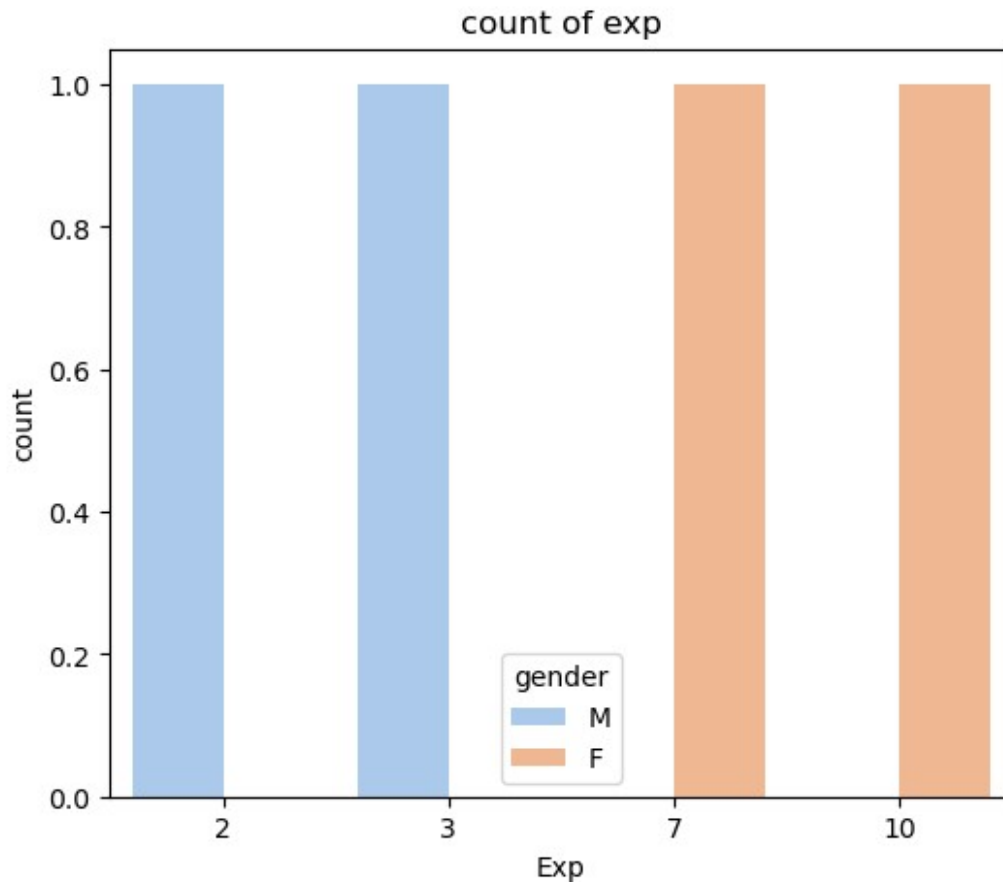
countplot

```
mydata1 = { "Name" : ["Ram","Sam","Joe","Asha"],
            "Age" : [23,22,26,47],
            "salary" : [15000,20000,32000,49000],
            "Exp" : [2,7,3,10],
            "gender" :['M','F','M','F']
          }
df1 = pd.DataFrame(mydata1)
df1.head()

    Name  Age  salary  Exp gender
0    Ram   23   15000    2      M
1    Sam   22   20000    7      F
2    Joe   26   32000    3      M
3   Asha   47   49000   10      F

plt.figure(figsize = (6,5))
sns.countplot(x=df1["Exp"],palette = "pastel", hue = df1["gender"])
plt.title("count of exp")

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

count of exp

pair plot

```
sns.pairplot(df1,hue='gender')

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 0x2671db629d0>
```
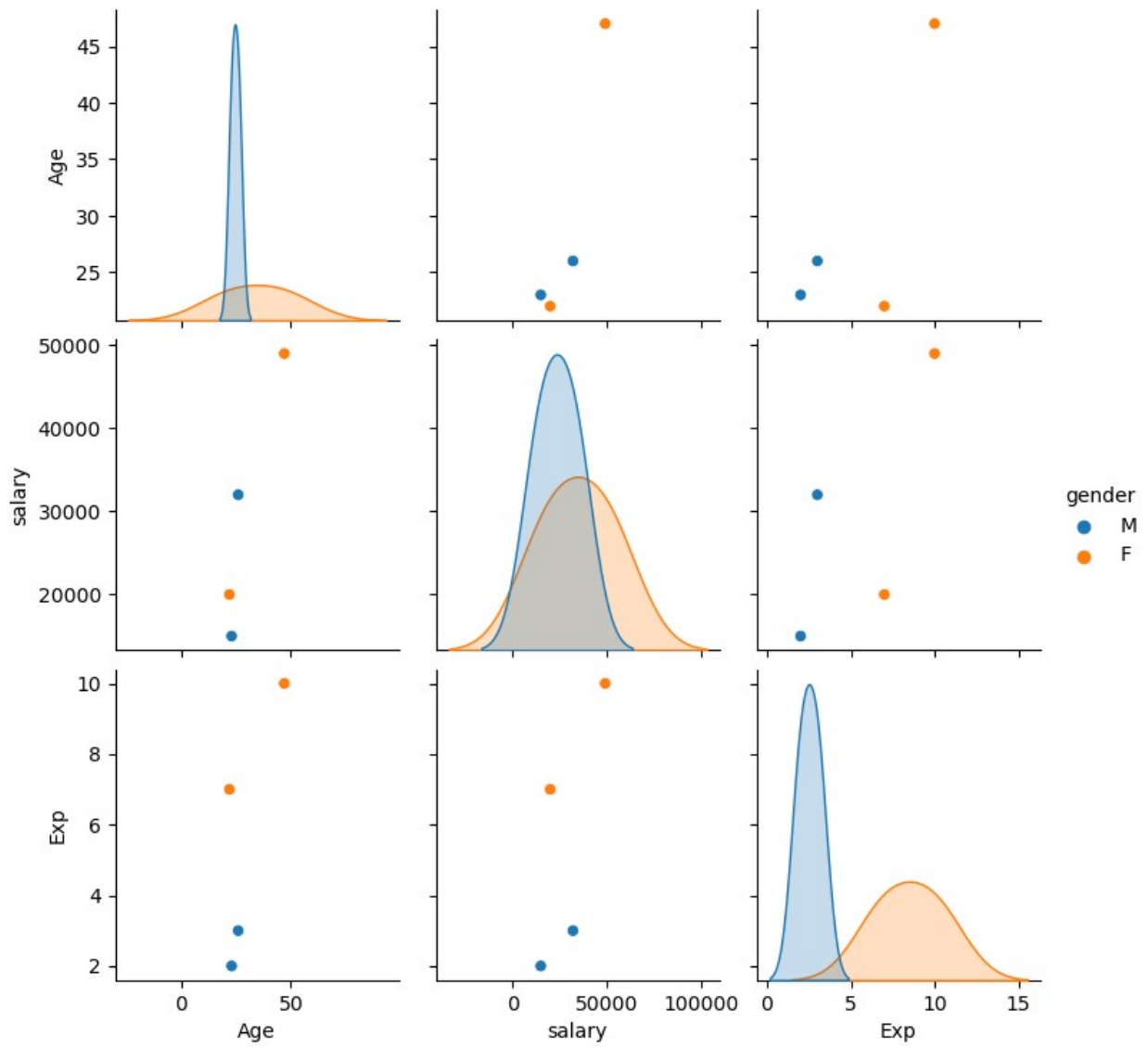
1.