DATA VISUALIZATION USING PYTHON(MODULE-MATPLOTLIB):

DATASET: https://www.kaggle.com/datasets/amirmahdiabbootalebi/salary-by-job-title-and-country:

```
import pandas as pd
df=pd.read_csv("/Salary.csv")
```

df

```
       Age  Gender  Education Level             Job Title  \
0     32.0    Male                1      Software Engineer
1     28.0  Female                2           Data Analyst
2     45.0    Male                3                Manager
3     36.0  Female                1        Sales Associate
4     52.0    Male                2               Director
...    ...     ...              ...                    ...
6679  49.0  Female                3  Director of Marketing
6680  32.0    Male                0        Sales Associate
6681  30.0  Female                1      Financial Manager
6682  46.0    Male                2      Marketing Manager
6683  26.0  Female                0        Sales Executive

      Years of Experience    Salary    Country        Race  Senior
0                     5.0   90000.0         UK       White       0
1                     3.0   65000.0        USA    Hispanic       0
2                    15.0  150000.0     Canada       White       1
3                     7.0   60000.0        USA    Hispanic       0
4                    20.0  200000.0        USA       Asian       0
...                   ...       ...        ...         ...     ...
6679                 20.0  200000.0         UK       Mixed       0
6680                  3.0   50000.0  Australia  Australian       0
6681                  4.0   55000.0      China     Chinese       0
6682                 14.0  140000.0      China      Korean       0
6683                  1.0   35000.0     Canada       Black       0

[6684 rows x 9 columns]
```

#BASIC ANALYSIS

df.columns

```
Index(['Age', 'Gender', 'Education Level', 'Job Title', 'Years of
Experience','Salary', 'Country', 'Race', 'Senior'],
dtype='object')
```

df.size

```
60156
```

df.shape

```
(6684, 9)
```

df.describe()

```
               Age  Education Level  Years of Experience         Salary  \
count  6684.000000      6684.000000          6684.000000    6684.000000
mean     33.610563         1.622382             8.077723  115307.175194
std       7.595994         0.880474             6.029305   52806.810881
min      21.000000         0.000000             0.000000     350.000000
25%      28.000000         1.000000             3.000000   70000.000000
50%      32.000000         1.000000             7.000000  115000.000000
75%      38.000000         2.000000            12.000000  160000.000000
max      62.000000         3.000000            34.000000  250000.000000

            Senior
count  6684.000000
mean      0.143477
std       0.350585
min       0.000000
25%       0.000000
50%       0.000000
75%       0.000000
max       1.000000
```

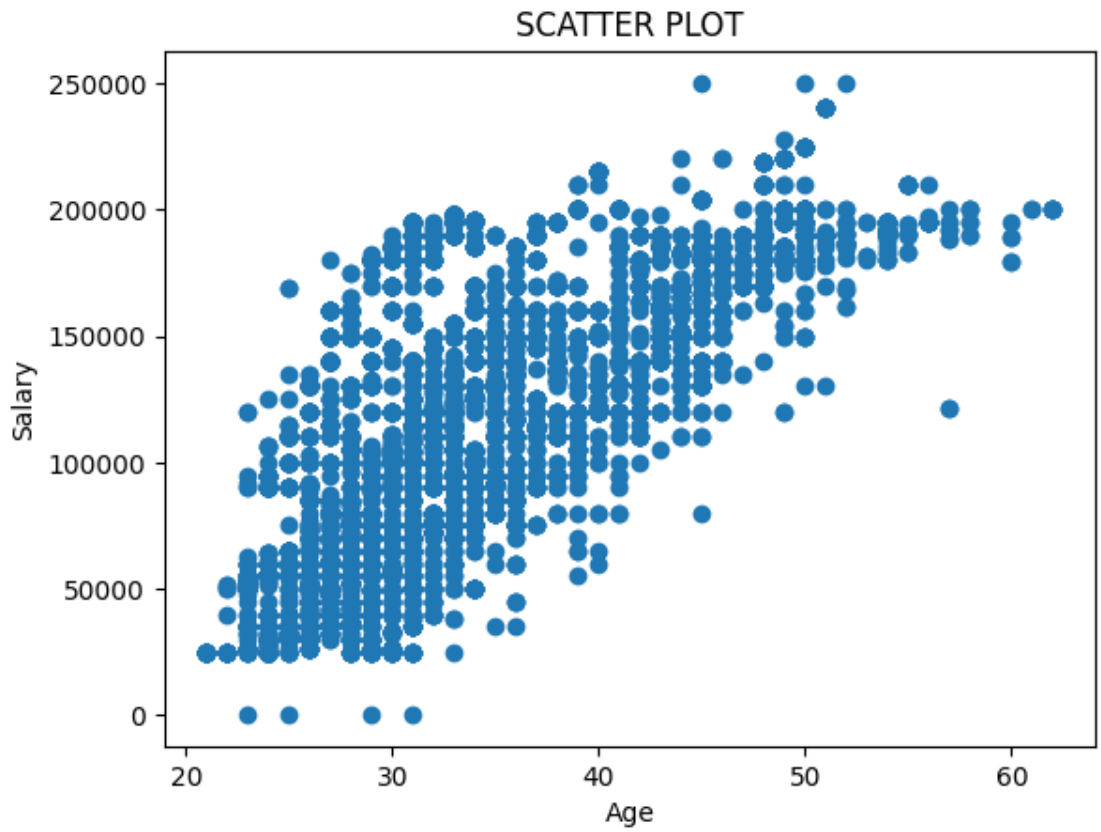
df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6684 entries, 0 to 6683
Data columns (total 9 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Age                  6684 non-null   float64
 1   Gender               6684 non-null   object 
 2   Education Level      6684 non-null   int64  
 3   Job Title            6684 non-null   object 
 4   Years of Experience  6684 non-null   float64
 5   Salary               6684 non-null   float64
 6   Country              6684 non-null   object 
 7   Race                 6684 non-null   object 
 8   Senior               6684 non-null   int64  
dtypes: float64(3), int64(2), object(4)
memory usage: 470.1+ KB
```

df.isnull().sum()

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

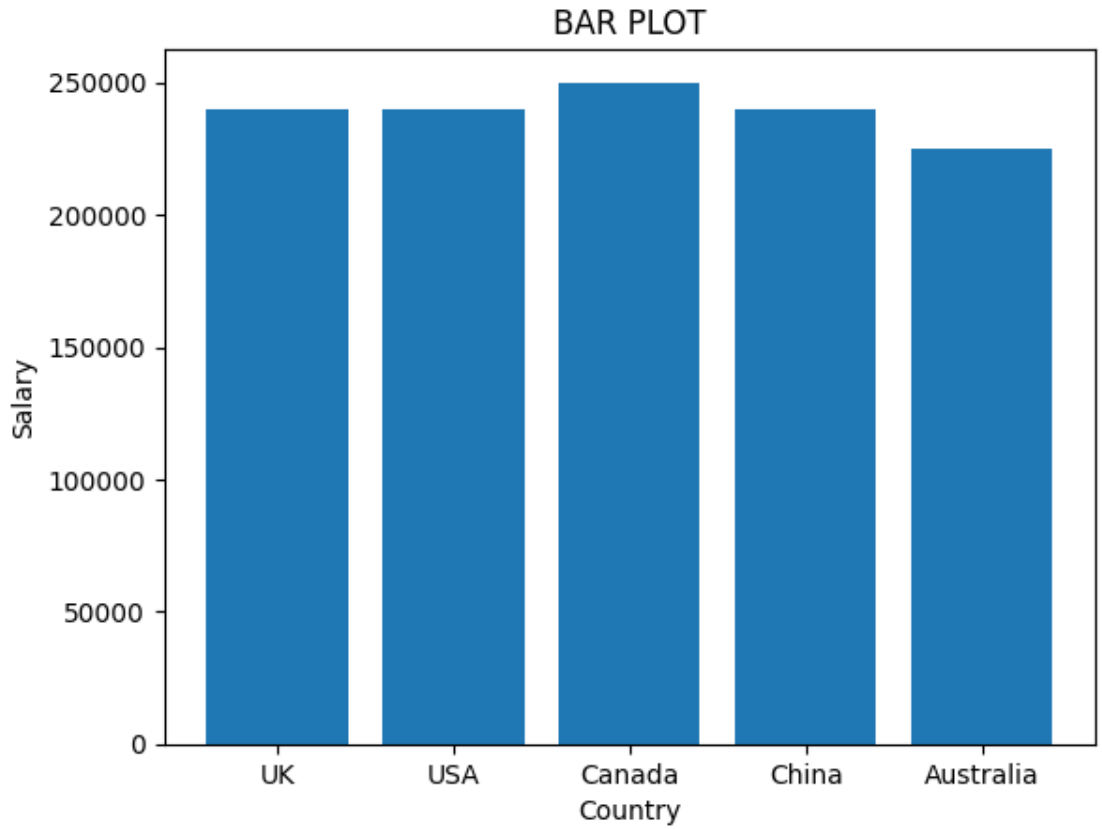
```
import matplotlib.pyplot as plt

#SCATTER PLOT
plt.scatter('Age','Salary',data=df)
plt.xlabel('Age')
plt.ylabel('Salary')
plt.title('SCATTER PLOT')
plt.show()
```



INTERPRETATION:

From the age 40-50 have the highest salary.

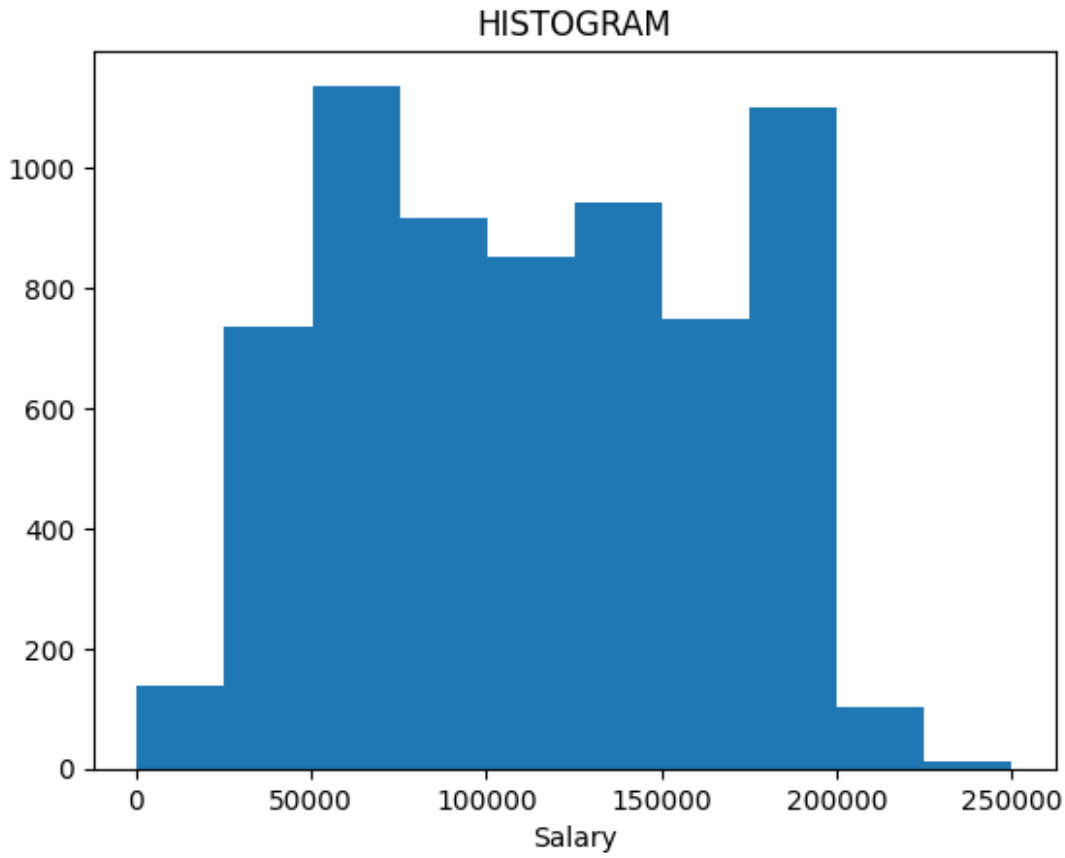
```
#BAR PLOT
plt.bar('Country','Salary',data=df)
plt.xlabel('Country')
plt.ylabel('Salary')
plt.title('BAR PLOT')
plt.show()
```



INTERPRETATION:

The Country Canada earn more than other countries.

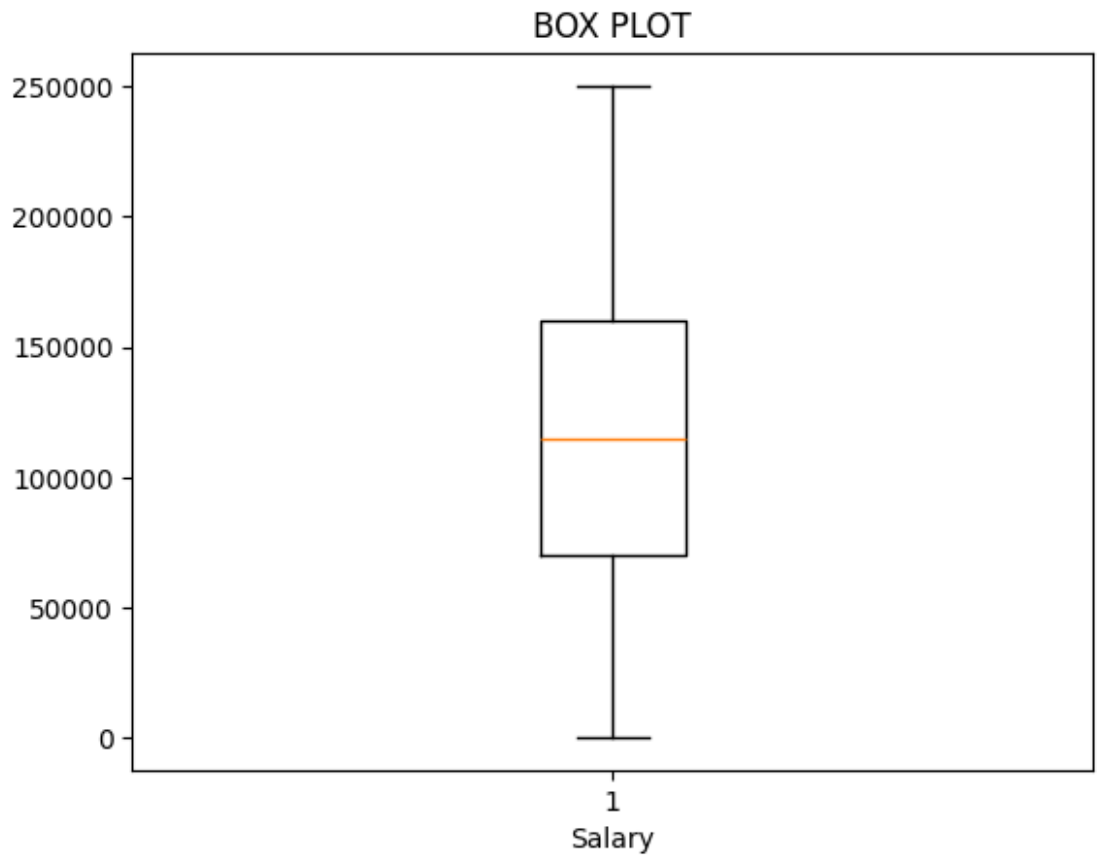
```
#HISTOGRAM
plt.hist('Salary',data=df,bins=10)
plt.xlabel('Salary')
plt.title('HISTOGRAM')
plt.show()
```



INTERPRETATION:

The employees who are all earn the salary range from 2lakhs-2.5lakhs are less in count.

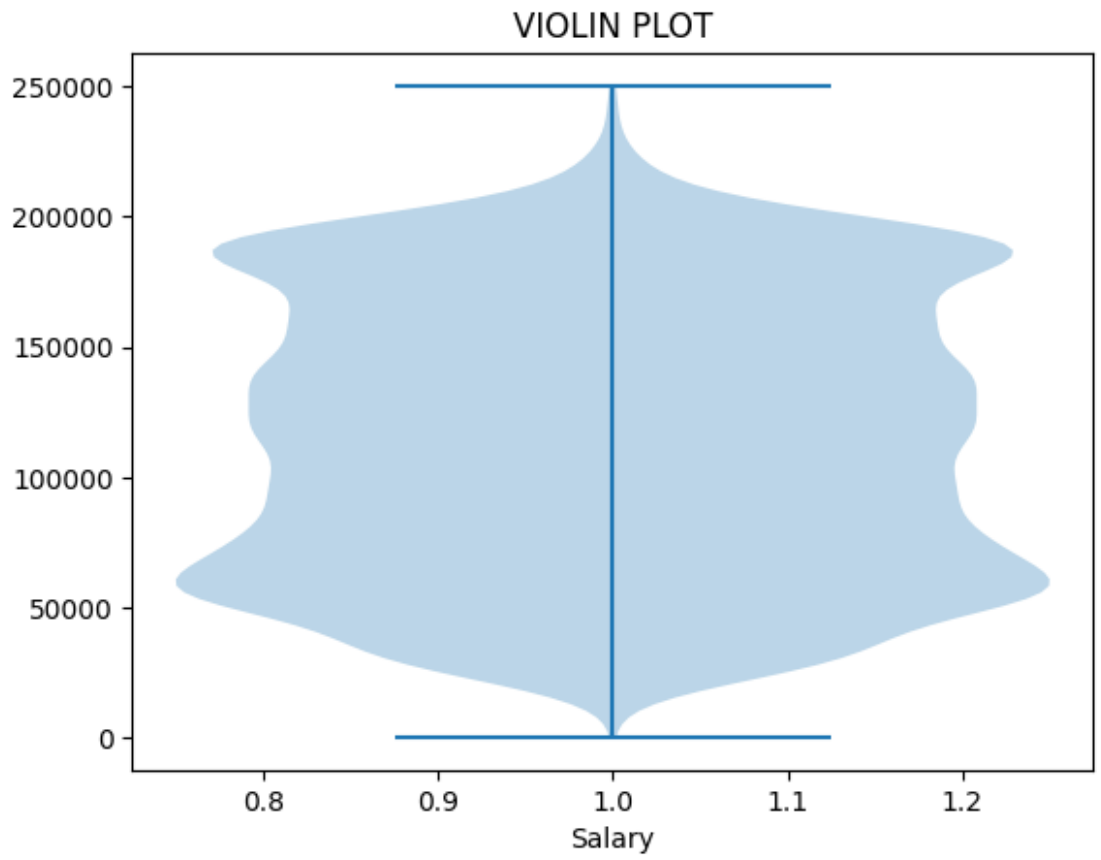
```
#BOXPLOT
plt.boxplot('Salary',data=df)
plt.xlabel('Salary')
plt.title('BOX PLOT')
plt.show()
```



INTERPRETATION:

The Average range of salary in this dataset is from 75000-1.5lakhs.

```
#VIOLIN PLOT
plt.violinplot('Salary',data=df)
plt.xlabel('Salary')
plt.title('VIOLIN PLOT')
plt.show()
```



INTERPRETATION:

The maximum salary range is from 51000-1.9lakhs per month.

DATA VISUALIZATION USING PYTHON(MODULE-SEABORN)::

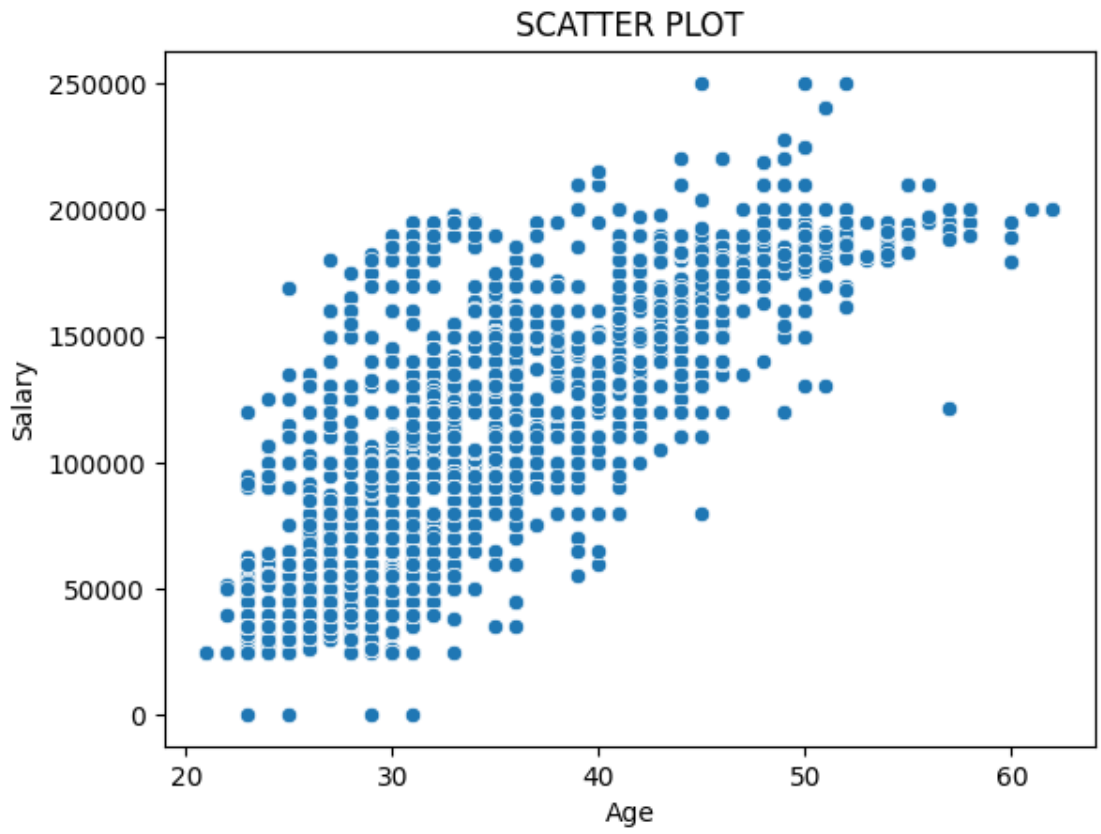
DATASET:

https://www.kaggle.com/datasets/amirmahdiabbootalebi/salary-by-job-title-and-country

```
#Import the library:
import pandas as pd
df=pd.read_csv("Salary.csv")
df
```

```
       Age  Gender  Education Level             Job Title  \
0     32.0    Male                1      Software Engineer
1     28.0  Female                2           Data Analyst
2     45.0    Male                3                Manager
3     36.0  Female                1        Sales Associate
4     52.0    Male                2               Director
...    ...     ...              ...                    ...
6679  49.0  Female                3  Director of Marketing
6680  32.0    Male                0        Sales Associate
6681  30.0  Female                1      Financial Manager
6682  46.0    Male                2      Marketing Manager
6683  26.0  Female                0        Sales Executive

      Years of Experience    Salary    Country        Race  Senior
0                     5.0   90000.0         UK       White       0
1                     3.0   65000.0        USA    Hispanic       0
2                    15.0  150000.0     Canada       White       1
3                     7.0   60000.0        USA    Hispanic       0
4                    20.0  200000.0        USA       Asian       0
...                   ...       ...        ...         ...     ...
6679                 20.0  200000.0         UK       Mixed       0
6680                  3.0   50000.0  Australia  Australian       0
6681                  4.0   55000.0      China     Chinese       0
6682                 14.0  140000.0      China      Korean       0
6683                  1.0   35000.0     Canada       Black       0

[6684 rows x 9 columns]
```

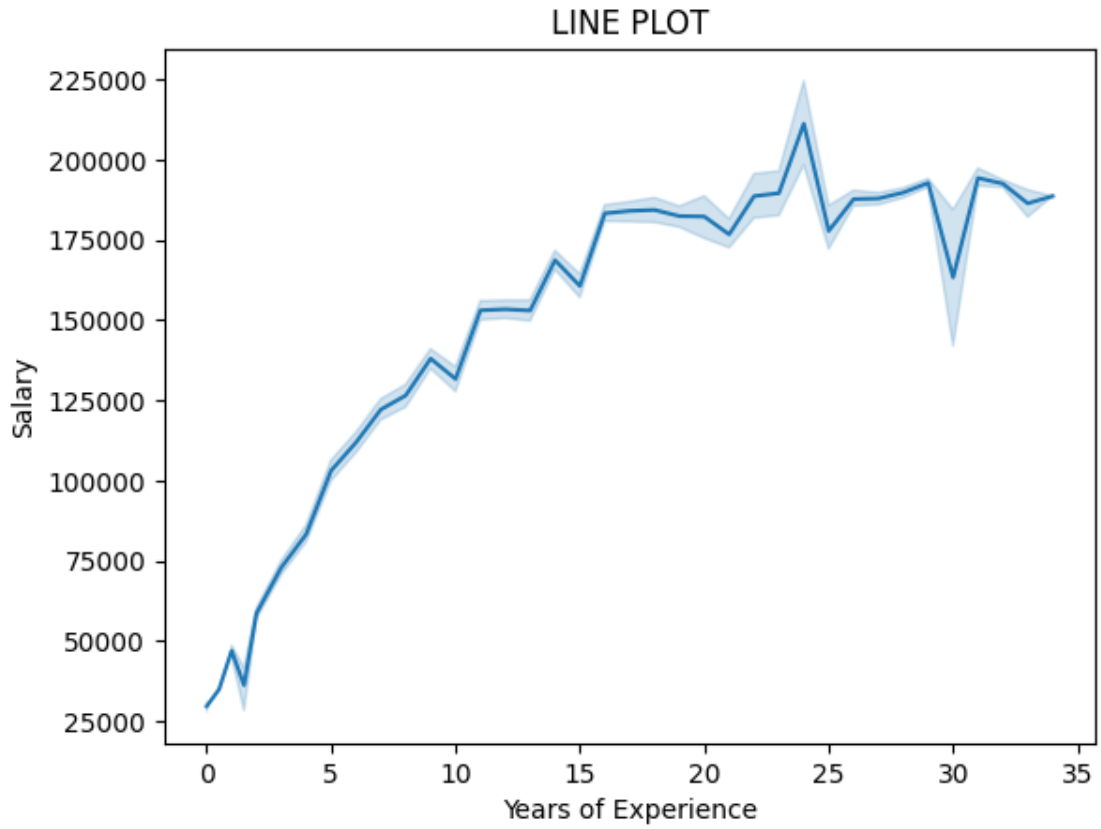
```
#import the required module:
import seaborn as sns
import matplotlib.pyplot as plt

#SCATTER PLOT
sns.scatterplot(x='Age',y='Salary',data=df)
plt.title('SCATTER PLOT')
plt.show()
```



INTERPRETATION:

From the age 40-50 have the highest salary.

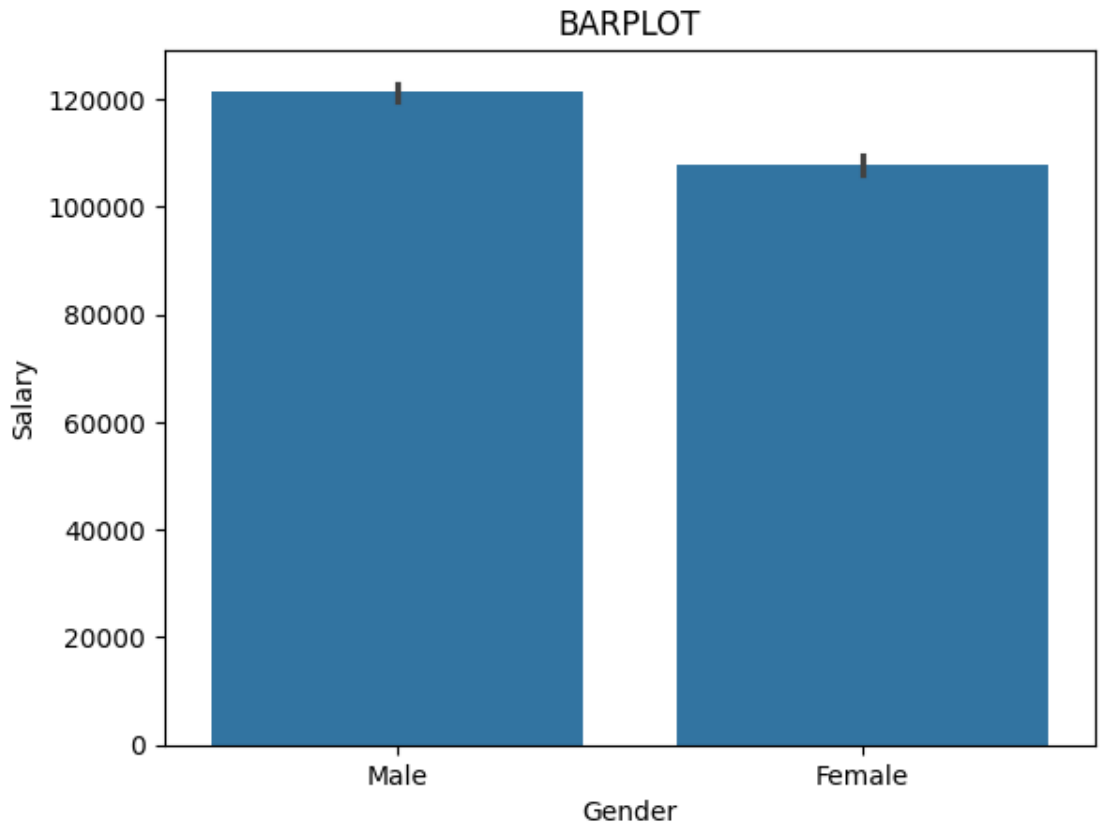
```
#LINE PLOT
sns.lineplot(x='Years of Experience',y='Salary',data=df)
plt.title('LINE PLOT')
plt.show()
```



INTERPRETATION:

Only the employees in the experience of 20-25 years have their salary around 2,25,000 rupees.

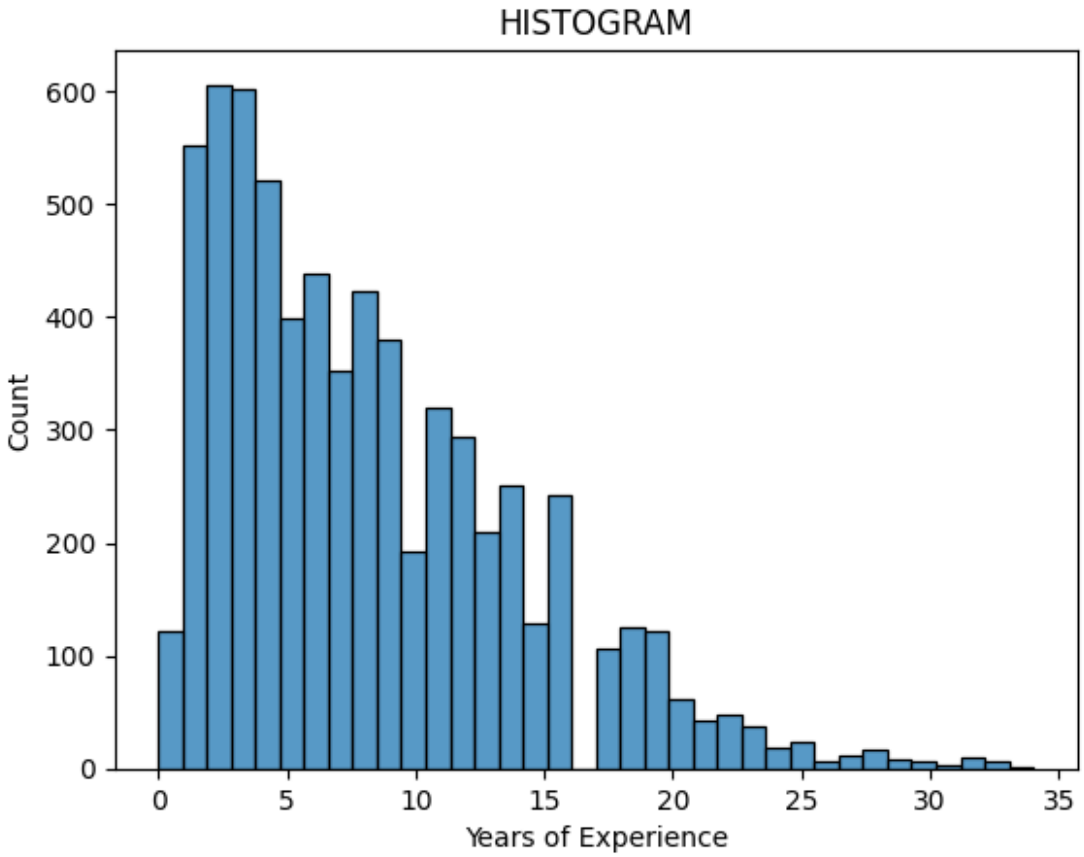
```
#BAR PLOT
sns.barplot(x='Gender',y='Salary',data=df)
plt.title('BARPLOT')
plt.show()
```



INTERPRETATION:

Male employees earn more than Female employees.

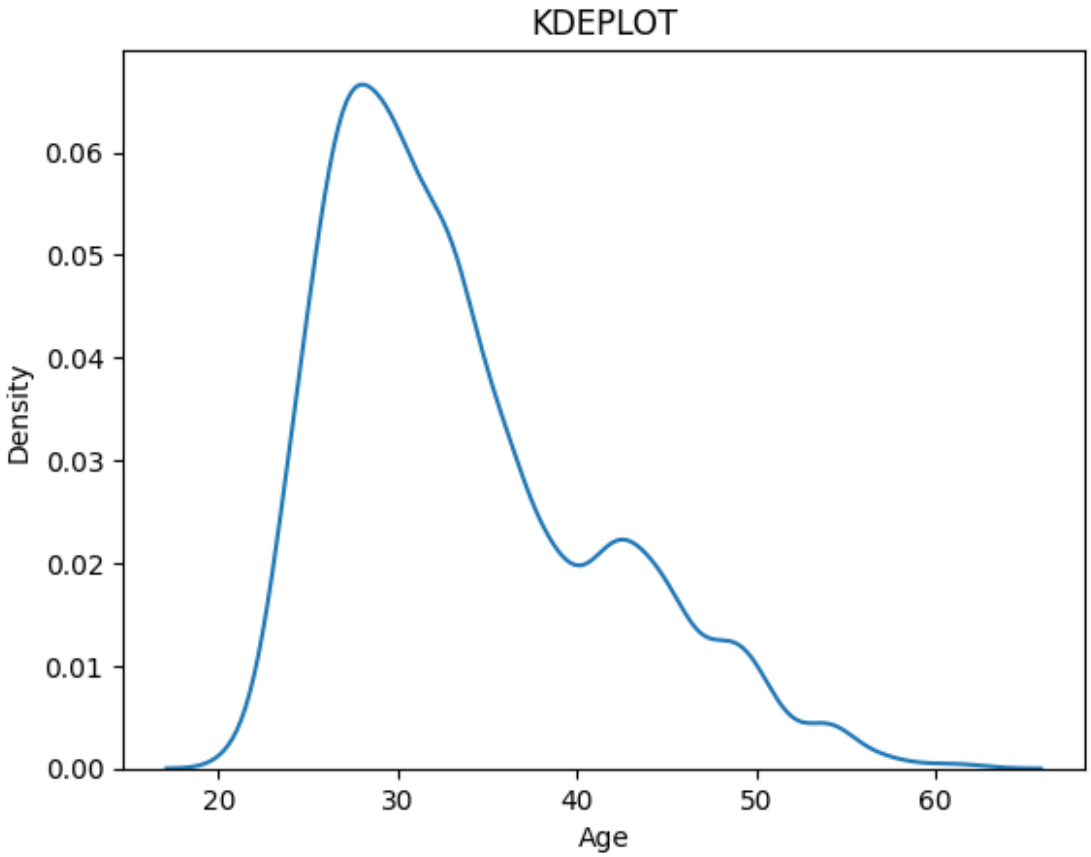
```
#HISTOGRAM
sns.histplot(x='Years of Experience',data=df)
plt.title('HISTOGRAM')
plt.show()
```



INTERPRETATION:

The employee whose experience from 2-5 years has the maximum count of 600.

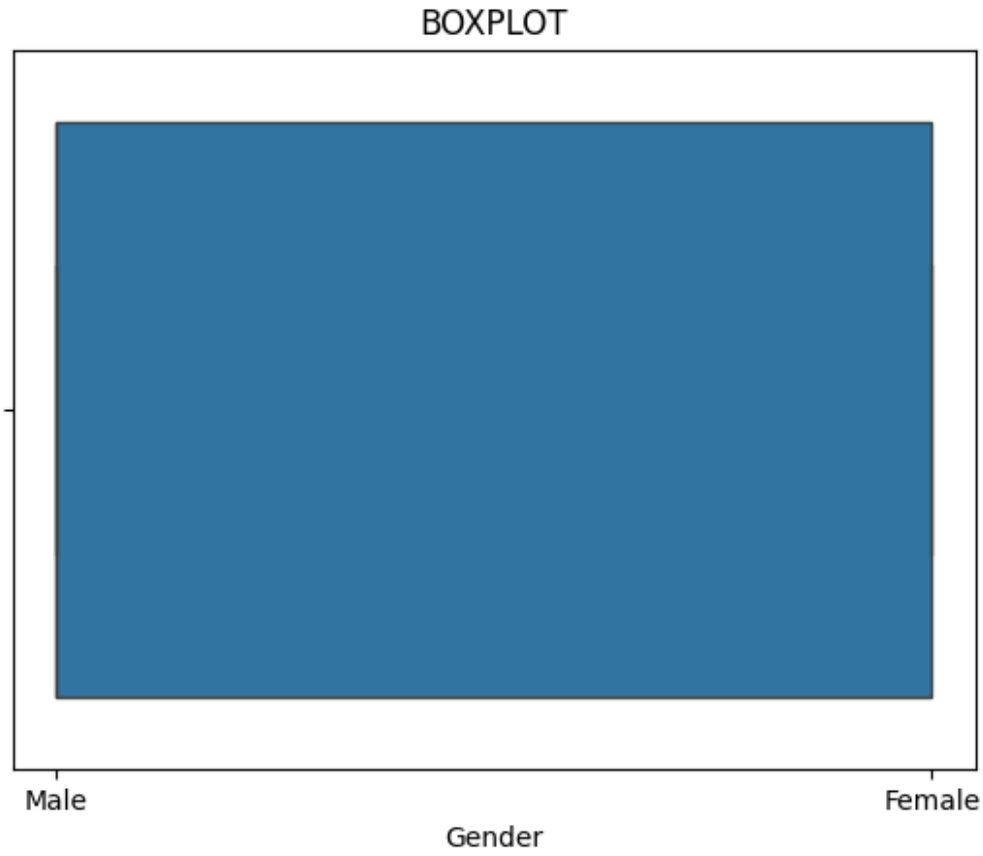
```
#KDE PLOT
sns.kdeplot(x='Age',data=df)
plt.title('KDEPLOT')
plt.show()
```



INTERPRETATION:

Maximum number of employees present at the Age of 20-30.

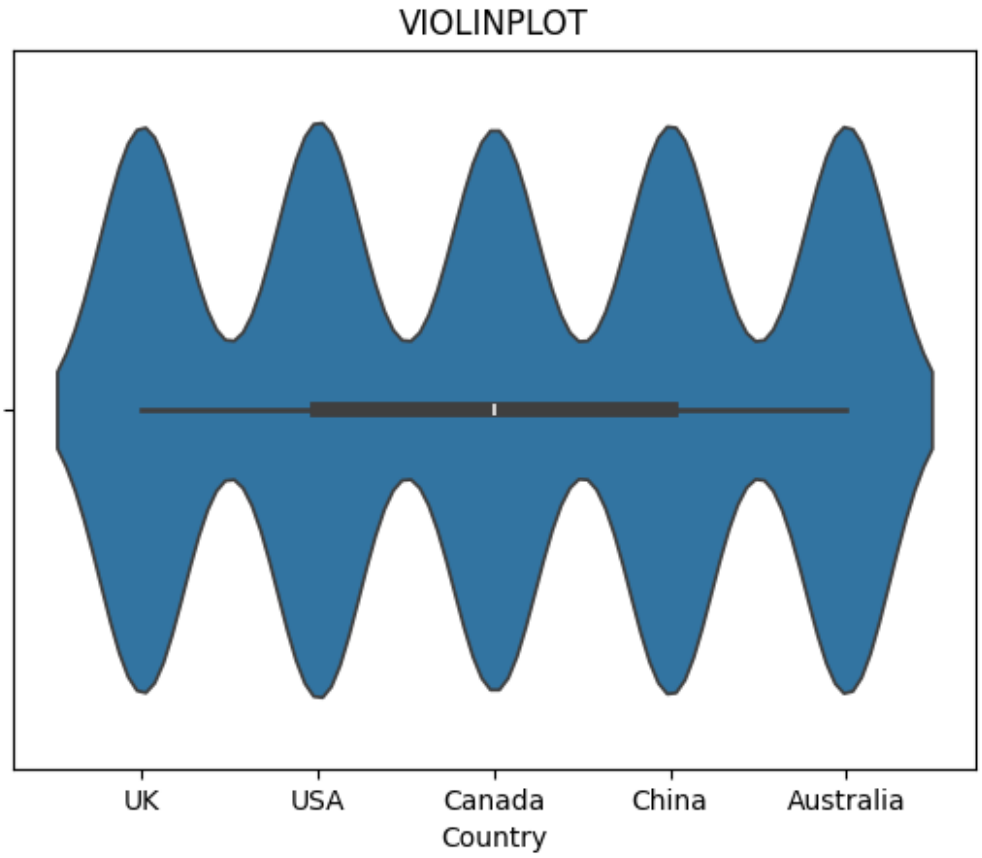
```
#BOX PLOT
sns.boxplot(x='Gender',data=df)
plt.title('BOXPLOT')
plt.show()
```



INTERPRETATION:

Equality of Male and Female are same.

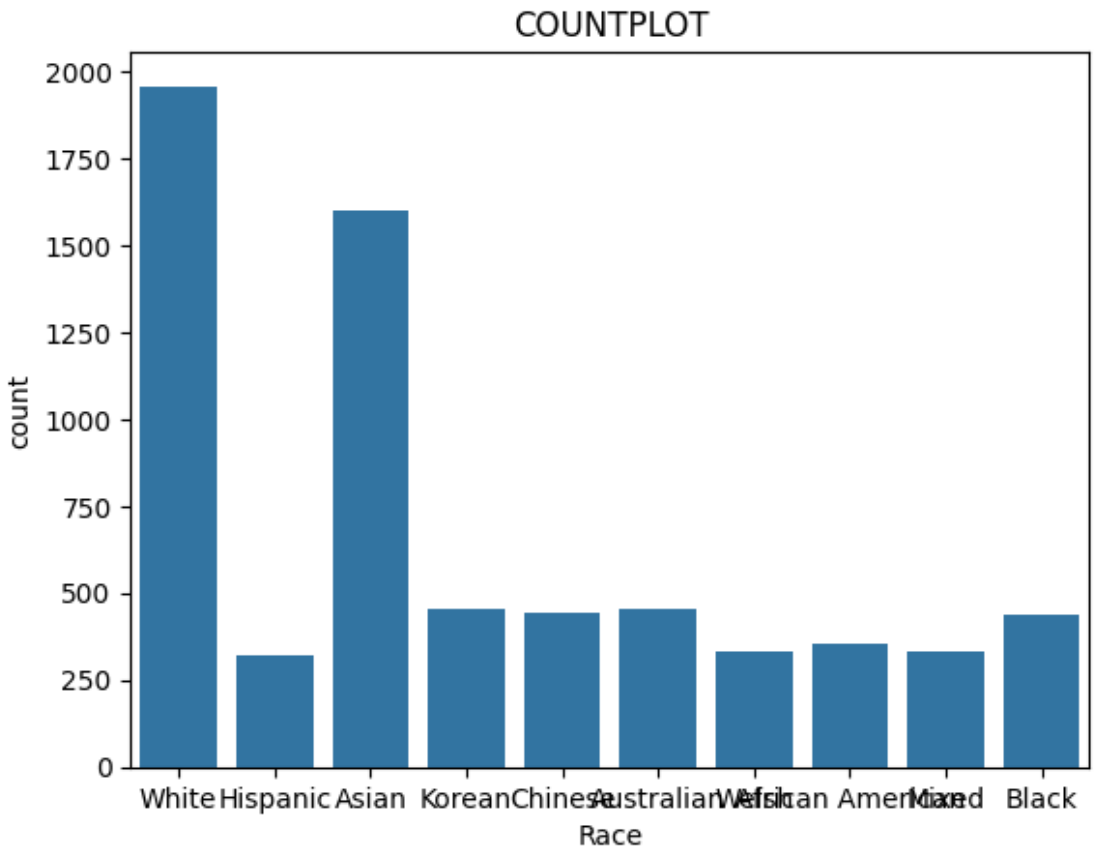
```
#VIOLIN PLOT
sns.violinplot(x='Country',data=df)
plt.title('VIOLINPLOT')
plt.show()
```



INTERPRETATION:

The average countries in the dataset are USA,CANADA and CHINA. bold text

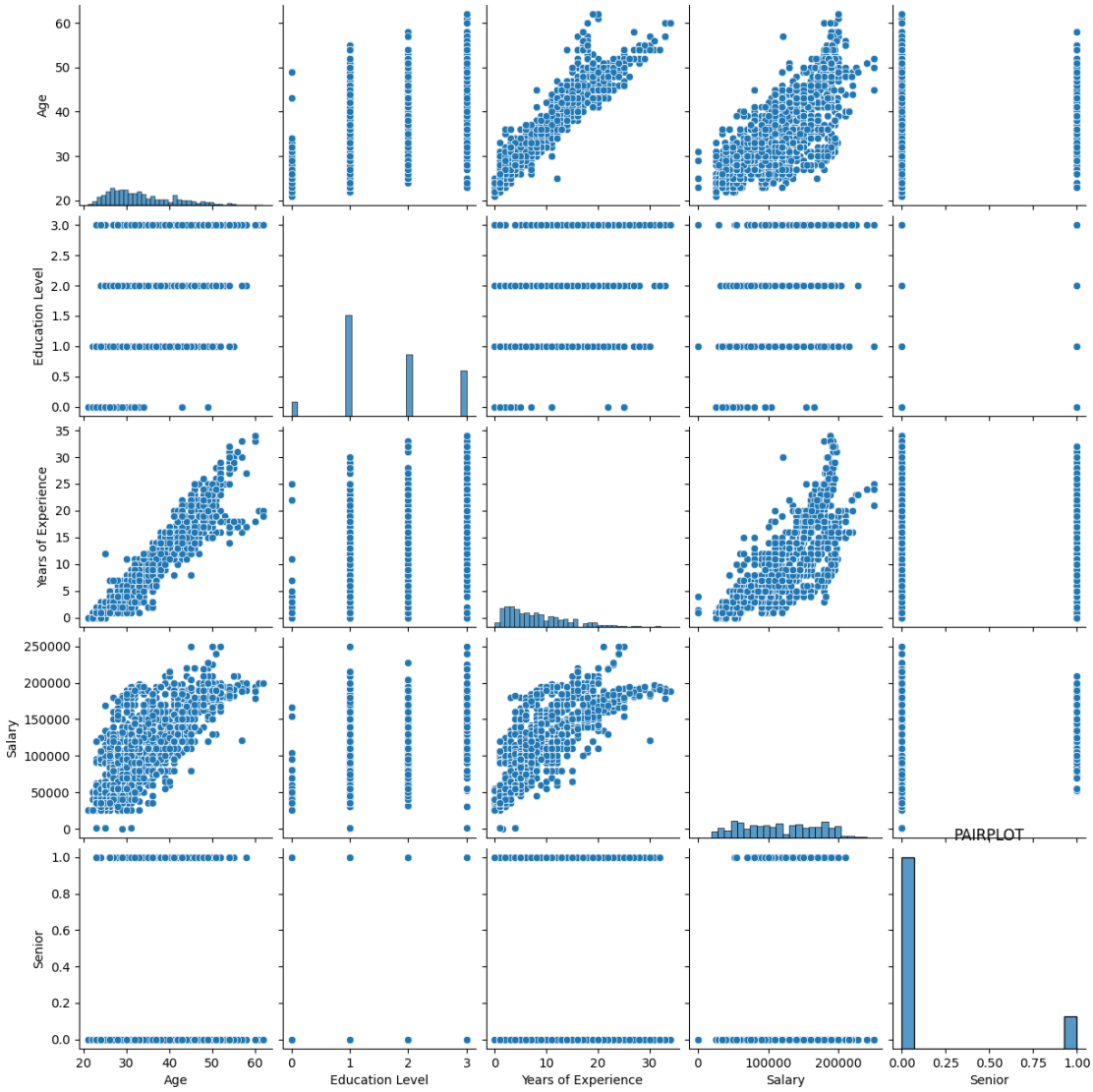
```
#COUNT PLOT
sns.countplot(x='Race',data=df)
plt.title('COUNTPLOT')
plt.show()
```



INTERPRETATION:

The presence of white people employees are more than others.

```
#PAIR PLOT
sns.pairplot(data=df)
plt.title('PAIRPLOT')
plt.show()
```



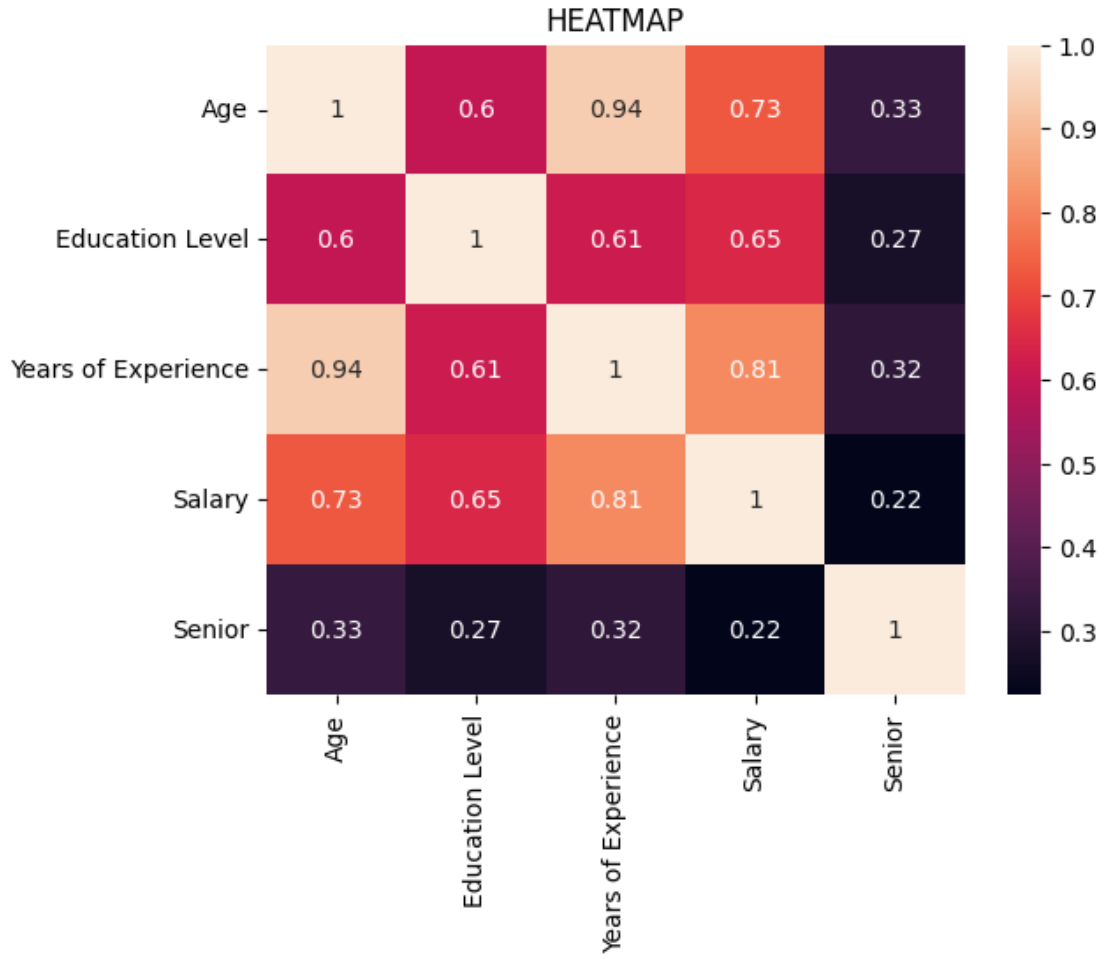
INTERPRETATION:

Here we can able to see the various camparison of charts using the pairplot function.

```
#corr()supports numeric values
df1=df.drop(['Gender','Job Title','Country','Race'],axis=1)
df1
```

```
       Age  Education Level  Years of Experience    Salary  Senior
0     32.0                1                  5.0   90000.0       0
1     28.0                2                  3.0   65000.0       0
2     45.0                3                 15.0  150000.0       1
3     36.0                1                  7.0   60000.0       0
4     52.0                2                 20.0  200000.0       0
...    ...              ...                  ...       ...     ...
6679  49.0                3                 20.0  200000.0       0
6680  32.0                0                  3.0   50000.0       0
6681  30.0                1                  4.0   55000.0       0
6682  46.0                2                 14.0  140000.0       0
6683  26.0                0                  1.0   35000.0       0

[6684 rows x 5 columns]
```

```
#HEAT MAP
sns.heatmap(df1.corr(),annot=True)
plt.title('HEATMAP')
plt.show()
```



INTERPRETATION:

- Years of experience increases,when the age is increases and also the salary is increases when the years of experience increases.