Importing basic libraries

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv('Salary_Dataset_with_Extra_Features.csv')
```

```
# getting first five rows of dataset
df.head()
```

	Rating	Company Name	Job Title	Salary	Salaries Reported	Location	Employment Status
0	3.8	Sasken	Android Developer	400000	3	Bangalore	Full Time
1	4.5	Advanced Millennium Technologies	Android Developer	400000	3	Bangalore	Full Time
2	4.0	Unacademy	Android Developer	1000000	3	Bangalore	Full Time

```
df.shape
```

```
(190, 8)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 190 entries, 0 to 189
Data columns (total 8 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   Rating             190 non-null    float64
 1   Company Name       190 non-null    object 
 2   Job Title          190 non-null    object 
 3   Salary             190 non-null    int64  
 4   Salaries Reported  190 non-null    int64  
 5   Location           190 non-null    object 
 6   Employment Status  190 non-null    object 
 7   Job Roles          190 non-null    object 
dtypes: float64(1), int64(2), object(5)
memory usage: 12.0+ KB
```

```
df.describe()
```

	Rating	Salary	Salaries Reported
count	190.000000	1.900000e+02	190.000000
mean	3.859474	6.049053e+05	4.952632
std	0.541983	6.882517e+05	26.043774
min	1.000000	2.400000e+04	1.000000
25%	3.600000	2.160000e+05	1.000000
50%	3.900000	4.000000e+05	1.000000
75%	4.100000	7.000000e+05	2.000000

```
df.columns
```

```
Index(['Rating', 'Company Name', 'Job Title', 'Salary', 'Salaries Reported',
       'Location', 'Employment Status', 'Job Roles'],
      dtype='object')
```

```
# which company has maximum number of employess

df['Company Name'].value_counts()
```

```
Tata Consultancy Services    3
Accenture                    3
Infosys                      3
Amazon                       2
Tech Mahindra                2
                            ..
Spandana Sphoorty            1
Bussan Auto Finance          1
Z5X                          1
Axis Bank                    1
DesignBids                   1
Name: Company Name, Length: 178, dtype: int64
```

```
# maximum employees works as which job title
df['Job Title'].value_counts()
```

```
Android Developer                                   41
Front End Developer                                 16
Junior Java Developer                               15
Senior Database Administrator                       14
IOS Software Developer                               9
Java Developer                                       8
Python/Django Developer                              7
Test Engineer                                        7
Backend Process                                      6
Software Engineer (Ios Developer)                    6
Web Developer                                        5
Tester                                               4
SDE-2 Backend                                        3
Ios Software Engineer                                3
Lead Software Development Engineer In Test (SDET)    3
Backend Developer                                    3
Senior Manager Software Development Engineering      2
Software Tester                                      2
```

```
Front End Developer - Intern                          2
Manual Test Engineer                                  2
Android Developer - Intern                            2
Web Developer - Intern                                2
Senior Database Administrator Contractor              2
Backend Executive                                     2
Backend Developer - Intern                            1
Java Backend Web Developer                            1
Junior Backend Engineer                               1
Jr Backend Developer - Intern                         1
Backend Lead Developer                                1
Senior Software Development Engineer In Test          1
Software Quality Assurance Engineer and Testers       1
Lead Backend Engineer - Intern                        1
Backend Developer, NodeJs - Intern                    1
Backend Operations                                    1
Lead Developer BackEnd                                1
Backend Support                                       1
Assistant Manager Backend                             1
Senior Software Development Engineer Lead             1
Backend Support Executive                             1
Node Js Backend Developer                             1
Backend Associate                                     1
IOS Software Developer - Intern                       1
Operations Backend                                    1
Front End Developer - Contractor                      1
Chargeback Backend                                    1
Backend Engineer Head                                 1
Senior Database Administrator - Contractor            1
Web Developer Contractor                              1
Name: Job Title, dtype: int64
```

```
df['Job Title'].unique()
```

```
array(['Android Developer', 'Android Developer - Intern',
       'Backend Process', 'SDE-2 Backend', 'Junior Backend Engineer',
       'Jr Backend Developer - Intern', 'Backend Lead Developer',
       'Backend Executive', 'Lead Backend Engineer - Intern',
       'Backend Developer, NodeJs - Intern', 'Backend Operations',
       'Lead Developer BackEnd', 'Java Backend Web Developer',
       'Backend Support', 'Backend Support Executive',
       'Node Js Backend Developer', 'Backend Associate',
       'Operations Backend', 'Chargeback Backend',
       'Backend Engineer Head', 'Assistant Manager Backend',
       'Backend Developer - Intern', 'Backend Developer',
       'Senior Database Administrator',
       'Senior Database Administrator Contractor',
       'Senior Database Administrator - Contractor',
       'Front End Developer', 'Front End Developer - Intern',
       'Front End Developer - Contractor', 'IOS Software Developer',
       'IOS Software Developer - Intern', 'Ios Software Engineer',
       'Software Engineer (Ios Developer)', 'Junior Java Developer',
       'Senior Software Development Engineer Lead',
       'Lead Software Development Engineer In Test (SDET)',
       'Senior Manager Software Development Engineering',
       'Python/Django Developer', 'Java Developer', 'Software Tester',
       'Manual Test Engineer', 'Tester',
       'Software Quality Assurance Engineer and Testers',
       'Senior Software Development Engineer In Test', 'Test Engineer',
```

```
       'Web Developer', 'Web Developer - Intern',
       'Web Developer Contractor'], dtype=object)
```

```
# maximum employees works in which location
df['Location'].value_counts()
```

```
Bangalore    101
Hyderabad     52
New Delhi     19
Pune          10
Chennai        8
Name: Location, dtype: int64
```

```
df['Job Roles'].value_counts()
```

```
Android     43
Backend     31
Java        23
Frontend    19
IOS         19
Database    17
Testing     13
SDE         10
Web          8
Python       7
Name: Job Roles, dtype: int64
```

```
# distribution of employees on the basis of employment type

df['Employment Status'].value_counts()
```

```
Full Time     174
Intern         11
Contractor      5
Name: Employment Status, dtype: int64
```

```
#finding and dropping null values
df.isnull()
```

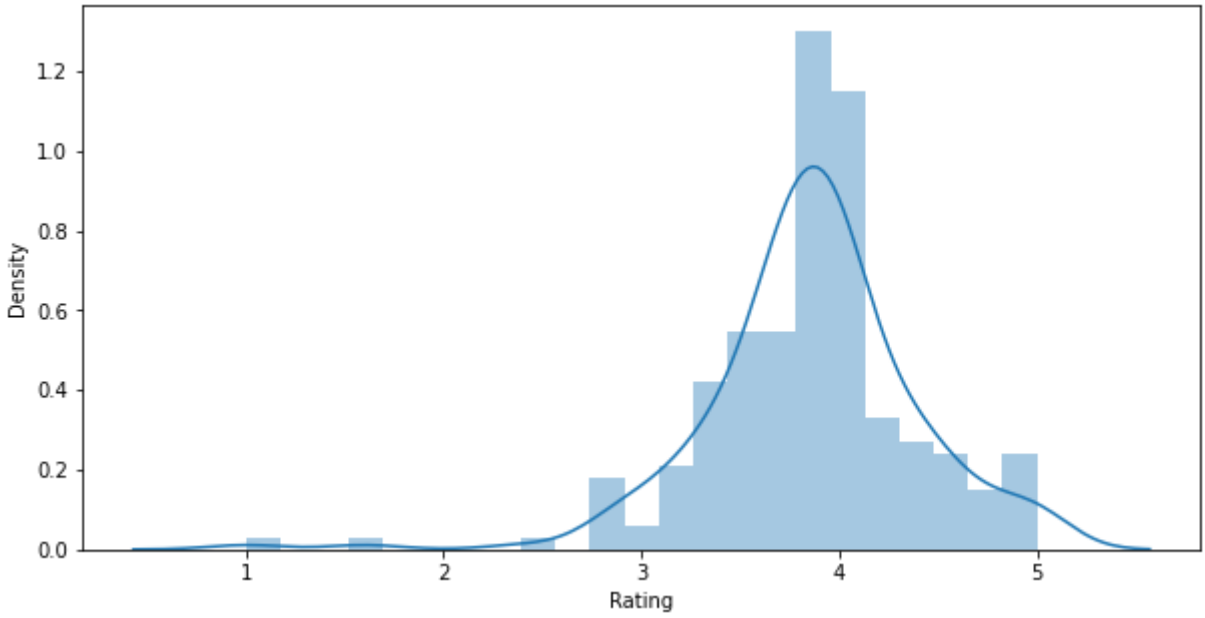
	Rating	Company Name	Job Title	Salary	Salaries Reported	Location	Employment Status
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False

Visualizing insights of Features

165	False	False	False	False	False	False	False

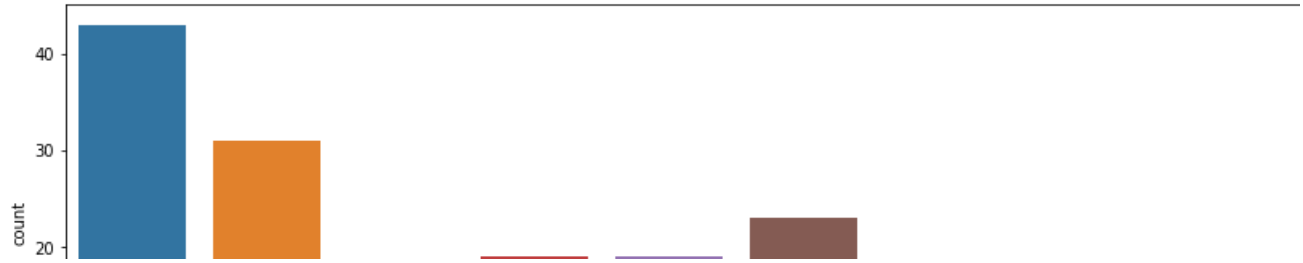
```
# checking distribution of rating feature
plt.figure(figsize=(10,5))
sns.distplot(df['Rating'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning:
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f219c090>
```



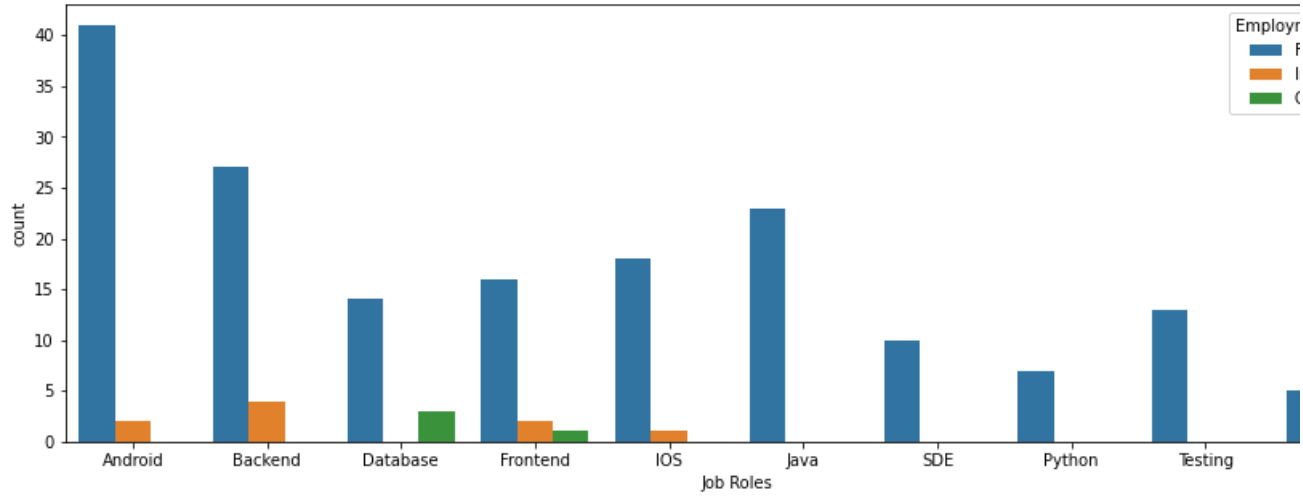

```
plt.figure(figsize=(15,5))
sns.countplot(df['Job Roles'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f2008d90>
```




```
plt.figure(figsize=(15,5))
sns.countplot(df['Job Roles'],hue=df['Employment Status'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f1acfe50>
```

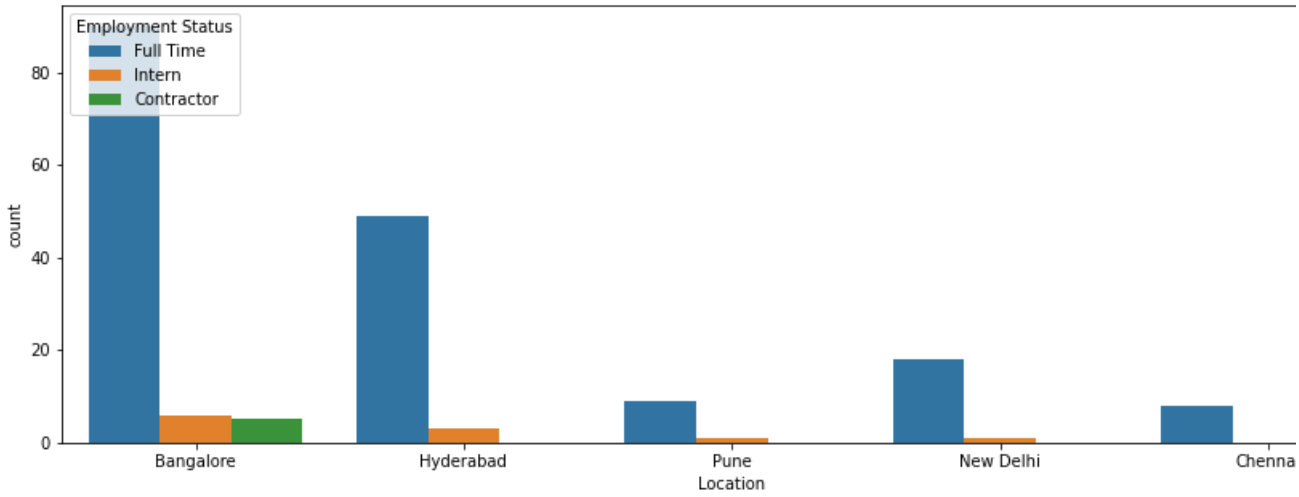



```
plt.figure(figsize=(15,5))
sns.countplot(df['Location'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f19b07d0>
```

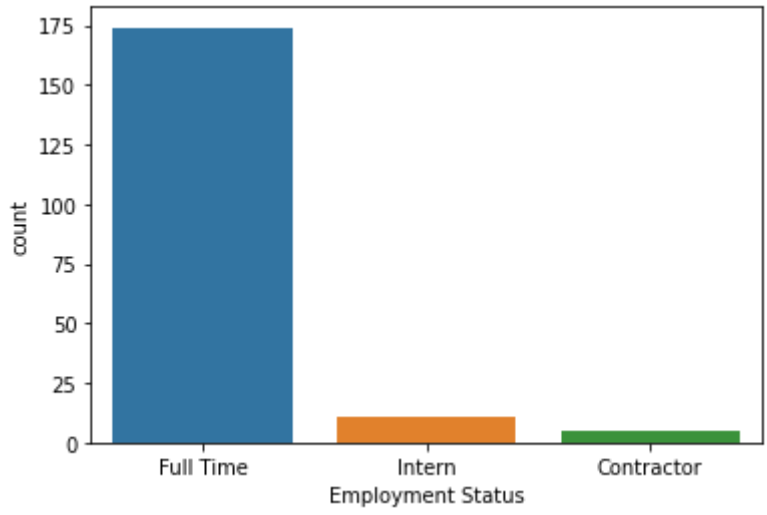

```
plt.figure(figsize=(15,5))
sns.countplot(df['Location'],hue=df['Employment Status'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f1937410>
```




```
sns.countplot(df['Employment Status'])
```

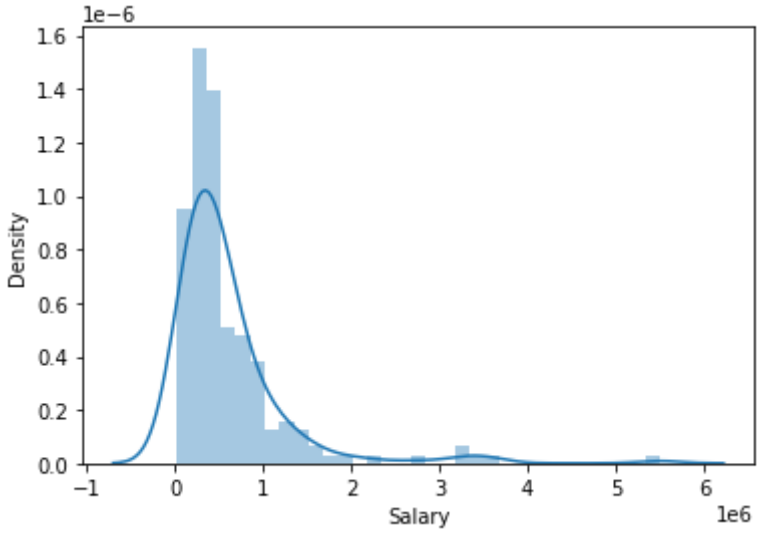
```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass
  FutureWarning
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f18d7a50>
```



People are more interested in getting Full time job as compared to intern , contrater and trainee.

```
sns.distplot(df['Salary'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning:
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f17e6390>
```




```
df['Salaries Reported'].value_counts()
```

```
1      130
2       42
3       11
331      1
83       1
70       1
60       1
52       1
50       1
48       1
Name: Salaries Reported, dtype: int64
```

```
df.drop('Salaries Reported',axis=1,inplace=True)
```

```
df.columns
```

```
Index(['Rating', 'Company Name', 'Job Title', 'Salary', 'Location',
       'Employment Status', 'Job Roles'],
      dtype='object')
```

```
# Top 20 companies with 5.0 ratings.
```

```
df[['Company Name','Rating']].sort_values('Rating',ascending=False).head(20)
```

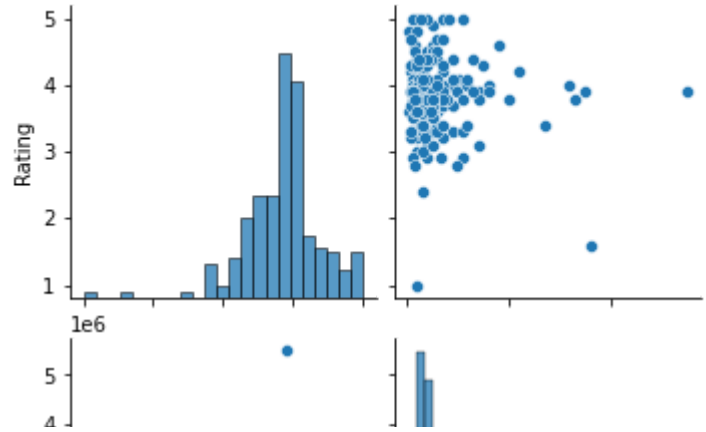
	Company Name	Rating
107	Tokhimo	5.0
30	powerplay app	5.0
113	Aquad Software Solutions (India)	5.0
57	TechPranee	5.0
35	Vispara Technosoft	5.0
182	Getzites Technologies	5.0
92	MAXC Studio	5.0
27	Retranz Infolabs	4.9
79	Liventus	4.8
13	Retail Pulse	4.8
70	Capital Float	4.7
72	Velotio	4.7
131	GHSL Technologies	4.7
32	Masai School	4.6
174	Gainsight	4.6
58	Fusion Tech	4.6
105	Havstruck	4.5

```
highest_salary_job = df.nlargest(5,['Salary'])
highest_salary_job
```

	Rating	Company Name	Job Title	Salary	Location	Employment Status	
149	3.9	Veritas Technologies	Senior Manager Software Development Engineering	5500000	Pune	Full Time	
68	1.6	Z5X	Backend Engineer Head	3600000	Hyderabad	Full Time	I
148	3.9	NortonLifeLock	Senior Manager Software Development Engineering	3500000	Pune	Full Time	
49	3.8	Amazon	SDE-2 Backend	3300000	Hvderabad	Full Time	I

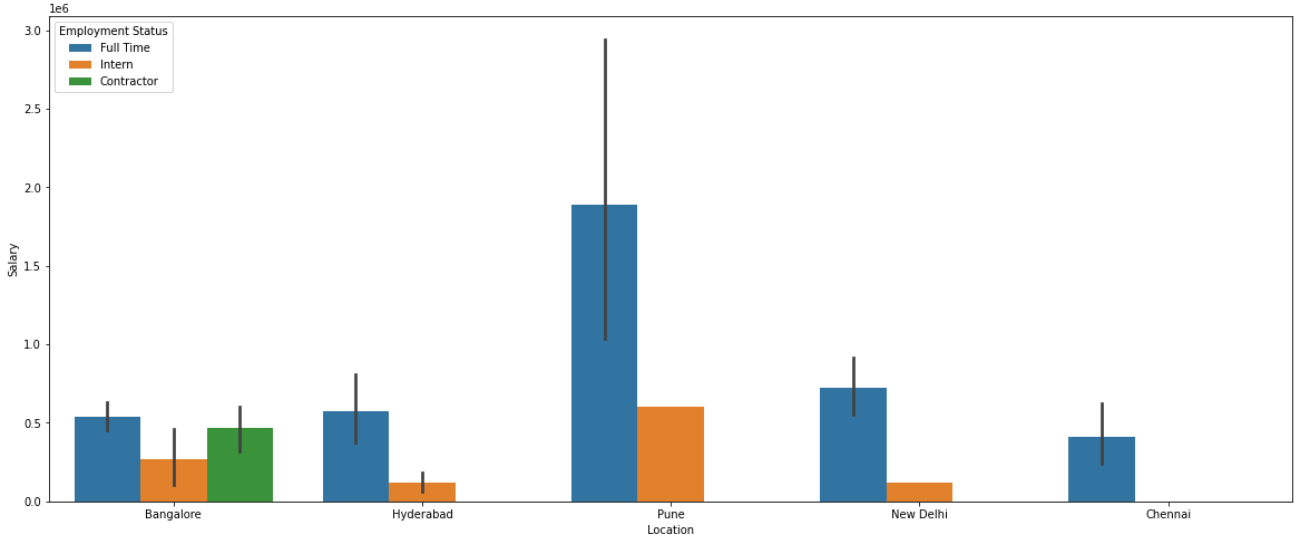
```
sns.pairplot(df)
```

```
<seaborn.axisgrid.PairGrid at 0x7f77f1748150>
```



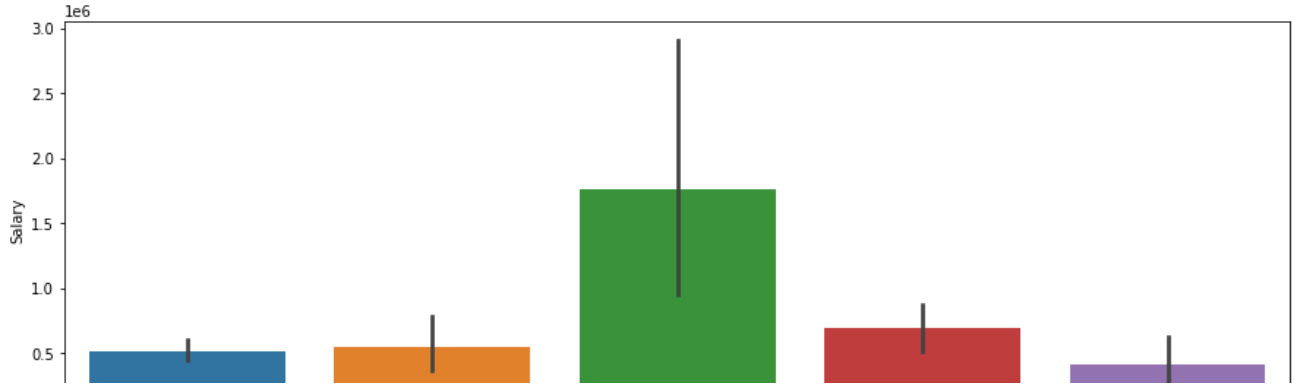

```
plt.figure(figsize=(20,8))
sns.barplot(data=df,x='Location',y='Salary',hue=df['Employment Status'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f168f910>
```



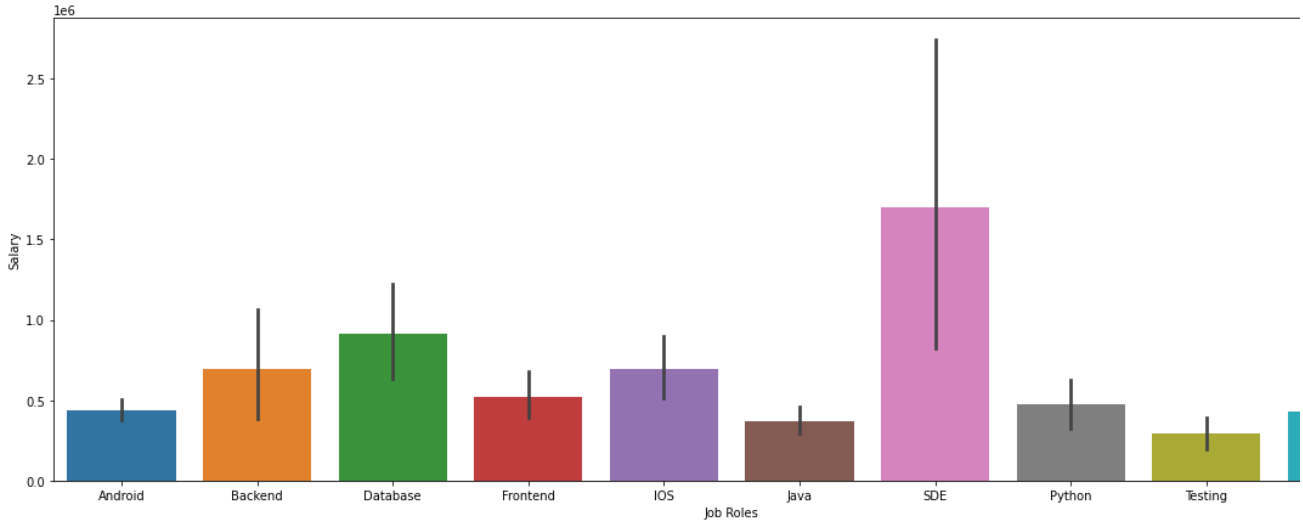

```
plt.figure(figsize=(15,5))
sns.barplot(data=df,x='Location',y='Salary')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f1f32390>
```




```
plt.figure(figsize=(20,7))
sns.barplot(data=df,x='Job Roles',y='Salary')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f77f19d54d0>
```



Instead maximum people work as SDE but they can't pay much by companies. We can see Database job role payed more rather than other job roles.

#