Machine Learning

Ex: 03 Data Preprocessing

Name: R.Athithraja

Reg.no: 2022503702

1. Handling Missing Values

Code:

```
import pandas as pd

df = pd.read_excel("Salary_Data_miss.xlsx")

print(df.isnull().sum())

df = df.fillna(df.median())

#df = df.dropna()

print("\nHandled Missing Values: \n")

print(df)
```

Output:

```
YearsExperience     0
Salary              2
dtype: int64

Handled Missing Values:

    YearsExperience     Salary
0               1.1   39343.0
1               1.3   46205.0
2               1.5   37731.0
3               2.0   43525.0
4               2.2   39891.0
5               2.9   56642.0
6               3.0   60150.0
7               3.2   54445.0
8               3.2   64445.0
9               3.7   57189.0
10              3.9   63218.0
11              4.0   55794.0
12              4.0   56957.0
13              4.1   65237.0
14              4.5   61111.0
15              4.9   67938.0
16              5.1   66029.0
17              5.3   83088.0
18              5.9   81363.0
19              6.0   93940.0
20              6.8   91738.0
21              7.1   65237.0
22              7.9  101302.0
23              8.2  113812.0
24              8.7  109431.0
25              9.0  105582.0
```

2. Five-Number Summary

Code:

```
summary = df['Salary'].describe()

print(summary)
```

Output:

```
count        30.000000
mean      75173.666667
std       27007.127135
min       37731.000000
25%       56720.750000
50%       65237.000000
75%       99461.500000
max      122391.000000
Name: Salary, dtype: float64
```

3. Box Plot

Code:

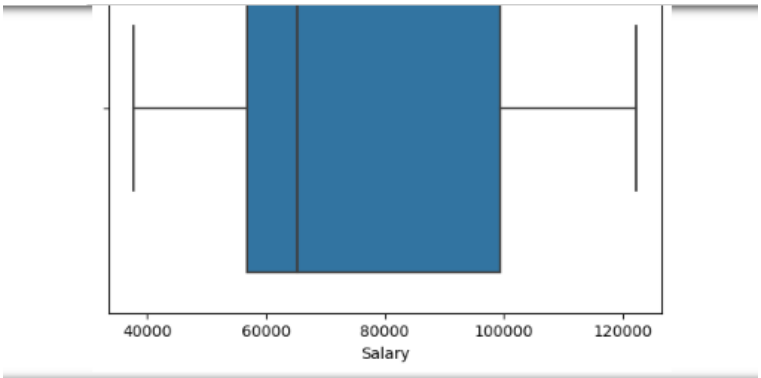
```
import seaborn as sns

import matplotlib.pyplot as plt

sns.boxplot(x=df['Salary'])

plt.show()
```

Output:



4. Correlation Matrix

Code:

```
corr_matrix = df.corr()

print(corr_matrix)
```

output:

```
                 YearsExperience    Salary
YearsExperience         1.000000  0.961987
Salary                  0.961987  1.000000
```

5. Covariance Matrix

Code:

```
cov_matrix = df.cov()

print(cov_matrix)
```

Output:

```
                 YearsExperience        Salary
YearsExperience         8.053609  7.372974e+04
Salary              73729.742529  7.293849e+08
```