

LAB 2 : Data Preprocessing Tools

In [1]:

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

In [3]:

```
dataset = pd.read_csv(r"C:\Users\Admin\Downloads\employee_performance_dataset.csv")
```

In [4]:

```
dataset.head()
```

Out[4]:

	Employee_ID	Age	Department	Job_Role	Experience_Years	Salary	Performance_Rating
0	1001	50	Sales	Engineer	15.0	94895.0	3
1	1002	36	Engineering	Manager	12.0	62307.0	2
2	1003	29	Sales	Executive	29.0	84098.0	3
3	1004	42	Marketing	Executive	18.0	90921.0	1
4	1005	40	Sales	Engineer	16.0	35486.0	2

In [5]:

```
dataset.columns
```

Out[5]:

Index(['Employee_ID', 'Age', 'Department', 'Job_Role', 'Experience_Years', 'Salary', 'Performance_Rating'], dtype='object')

In [6]:

```
dataset.shape
```

Out[6]:

(120, 7)

In [7]:

```
dataset.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120 entries, 0 to 119
Data columns (total 7 columns):
Column Non-Null Count Dtype
--- -
0 Employee_ID 120 non-null int64
1 Age 120 non-null int64
2 Department 120 non-null object
3 Job_Role 120 non-null object
4 Experience_Years 119 non-null float64
5 Salary 119 non-null float64
6 Performance_Rating 120 non-null int64
dtypes: float64(2), int64(3), object(2)
memory usage: 6.7+ KB

In [8]:

```
dataset.describe()
```

Out[8]:

	Employee_ID	Age	Experience_Years	Salary	Performance_Rating
count	120.000000	120.000000	119.000000	119.000000	120.000000
mean	1060.500000	40.016667	18.386555	74173.722689	3.041667
std	34.785054	10.991963	10.073551	26448.766806	1.169614
min	1001.000000	22.000000	0.000000	31062.000000	1.000000
25%	1030.750000	30.000000	11.000000	53162.500000	2.000000
50%	1060.500000	40.500000	19.000000	69298.000000	3.000000
75%	1090.250000	48.250000	27.000000	96121.500000	4.000000
max	1120.000000	59.000000	34.000000	119930.000000	5.000000

In [9]:

```
dataset.describe(include = 'object')
```

Out[9]:

	Department	Job_Role
count	120	120
unique	5	5
top	HR	Executive
freq	32	27

Preprocessing Steps

Step 1 : Divide dataframe into input and output Featutres

In [10]:

```
X = dataset.iloc[:, :-1]
Y = dataset.iloc[:, -1]
```

In [11]:

```
print(X)
```

	Employee_ID	Age	Department	Job_Role	Experience_Years	Salary
0	1001	50	Sales	Engineer	15.0	94895.0
1	1002	36	Engineering	Manager	12.0	62307.0
2	1003	29	Sales	Executive	29.0	84098.0
3	1004	42	Marketing	Executive	18.0	90921.0
4	1005	40	Sales	Engineer	16.0	35486.0
..
115	1116	53	Engineering	Executive	8.0	91529.0
116	1117	53	Engineering	Consultant	6.0	114665.0
117	1118	25	Marketing	Executive	27.0	61921.0
118	1119	51	HR	Engineer	13.0	63307.0
119	1120	58	Marketing	Executive	30.0	57355.0

[120 rows x 6 columns]

In [12]:

```
print(Y)
```

0	3
1	2

```

2      3
3      1
4      2
..
115    4
116    3
117    2
118    3
119    1
Name: Performance_Rating, Length: 120, dtype: int64

```

Step 2: Handle the missing values in Dataset

In [15]:

```
!pip install scikit-learn
```

Collecting scikit-learn

Obtaining dependency information for scikit-learn from https://files.pythonhosted.org/packages/62/27/585859e72e117fe861c2079bcba35591a84f801e21bc1ab85bce6ce60305/scikit_learn-1.6.1-cp312-cp312-win_amd64.whl.metadata

Downloading scikit_learn-1.6.1-cp312-cp312-win_amd64.whl.metadata (15 kB)

Requirement already satisfied: numpy>=1.19.5 in c:\users\admin\appdata\local\programs\python\python312\lib\site-packages (from scikit-learn) (2.1.2)

Collecting scipy>=1.6.0 (from scikit-learn)

Obtaining dependency information for scipy>=1.6.0 from https://files.pythonhosted.org/packages/f5/6f/e6e5aff77ea2a48dd96808bb51d7450875af154ee7cbe72188afb0b37929/scipy-1.15.2-cp312-cp312-win_amd64.whl.metadata

Downloading scipy-1.15.2-cp312-cp312-win_amd64.whl.metadata (60 kB)

----- 0.0/60.8 kB ? eta -:--:--

----- 10.2/60.8 kB ? eta -:--:--

----- 10.2/60.8 kB ? eta -:--:--

----- 60.8/60.8 kB 537.5 kB/s eta 0:00:00

Requirement already satisfied: joblib>=1.2.0 in c:\users\admin\appdata\roaming\python\python312\site-packages (from scikit-learn) (1.4.2)

Collecting threadpoolctl>=3.1.0 (from scikit-learn)

Obtaining dependency information for threadpoolctl>=3.1.0 from <https://files.pythonhosted.org/packages/32/d5/f9a850d79b0851d1d4ef6456097579a9005b31fea68726a4ae5f2d82ddd9/threadpoolctl-3.6.0-py3-none-any.whl.metadata>

Downloading threadpoolctl-3.6.0-py3-none-any.whl.metadata (13 kB)

Downloading scikit_learn-1.6.1-cp312-cp312-win_amd64.whl (11.1 MB)

----- 0.0/11.1 MB ? eta -:--:--

----- 0.0/11.1 MB ? eta -:--:--

----- 0.0/11.1 MB 640.0 kB/s eta 0:00:18

----- 0.0/11.1 MB 393.8 kB/s eta 0:00:29

----- 0.1/11.1 MB 363.1 kB/s eta 0:00:31

----- 0.1/11.1 MB 476.3 kB/s eta 0:00:24

----- 0.1/11.1 MB 476.3 kB/s eta 0:00:24

----- 0.1/11.1 MB 476.3 kB/s eta 0:00:24

----- 0.1/11.1 MB 476.3 kB/s eta 0:00:24

----- 0.1/11.1 MB 476.3 kB/s eta 0:00:24

----- 0.2/11.1 MB 540.4 kB/s eta 0:00:21

----- 0.2/11.1 MB 550.0 kB/s eta 0:00:20

----- 0.2/11.1 MB 464.5 kB/s eta 0:00:24

----- 0.2/11.1 MB 464.5 kB/s eta 0:00:24

----- 0.3/11.1 MB 436.5 kB/s eta 0:00:25

----- 0.3/11.1 MB 453.2 kB/s eta 0:00:24

----- 0.3/11.1 MB 463.3 kB/s eta 0:00:24

----- 0.3/11.1 MB 446.5 kB/s eta 0:00:25

----- 0.3/11.1 MB 446.5 kB/s eta 0:00:25

----- 0.5/11.1 MB 541.6 kB/s eta 0:00:20

----- 0.5/11.1 MB 582.4 kB/s eta 0:00:19

----- 0.6/11.1 MB 609.1 kB/s eta 0:00:18

----- 0.6/11.1 MB 622.0 kB/s eta 0:00:17

----- 0.6/11.1 MB 634.2 kB/s eta 0:00:17

----- 0.7/11.1 MB 654.9 kB/s eta 0:00:16

----- 0.7/11.1 MB 664.7 kB/s eta 0:00:16

----- 0.8/11.1 MB 682.7 kB/s eta 0:00:16

----- 0.8/11.1 MB 699.0 kB/s eta 0:00:15

----- 0.9/11.1 MB 713.1 kB/s eta 0:00:15

----- 0.9/11.1 MB 737.3 kB/s eta 0:00:14

----- 1.0/11.1 MB 733.1 kB/s eta 0:00:14

-----	1.0/11.1	MB	733.1	kB/s	eta	0:00:14
-----	1.0/11.1	MB	755.7	kB/s	eta	0:00:14
-----	1.1/11.1	MB	771.5	kB/s	eta	0:00:13
-----	1.1/11.1	MB	771.5	kB/s	eta	0:00:13
-----	1.1/11.1	MB	757.8	kB/s	eta	0:00:14
-----	1.1/11.1	MB	757.8	kB/s	eta	0:00:14
-----	1.2/11.1	MB	738.8	kB/s	eta	0:00:14
-----	1.2/11.1	MB	749.9	kB/s	eta	0:00:14
-----	1.3/11.1	MB	746.2	kB/s	eta	0:00:14
-----	1.3/11.1	MB	767.3	kB/s	eta	0:00:13
-----	1.4/11.1	MB	752.8	kB/s	eta	0:00:13
-----	1.4/11.1	MB	745.0	kB/s	eta	0:00:14
-----	1.4/11.1	MB	745.0	kB/s	eta	0:00:14
-----	1.4/11.1	MB	731.0	kB/s	eta	0:00:14
-----	1.4/11.1	MB	731.0	kB/s	eta	0:00:14
-----	1.5/11.1	MB	741.6	kB/s	eta	0:00:13
-----	1.5/11.1	MB	740.8	kB/s	eta	0:00:13
-----	1.5/11.1	MB	728.6	kB/s	eta	0:00:14
-----	1.6/11.1	MB	727.0	kB/s	eta	0:00:14
-----	1.6/11.1	MB	729.8	kB/s	eta	0:00:14
-----	1.6/11.1	MB	728.2	kB/s	eta	0:00:13
-----	1.7/11.1	MB	731.4	kB/s	eta	0:00:13
-----	1.7/11.1	MB	743.4	kB/s	eta	0:00:13
-----	1.7/11.1	MB	743.4	kB/s	eta	0:00:13
-----	1.8/11.1	MB	732.3	kB/s	eta	0:00:13
-----	1.8/11.1	MB	726.6	kB/s	eta	0:00:13
-----	1.8/11.1	MB	711.8	kB/s	eta	0:00:14
-----	1.8/11.1	MB	715.2	kB/s	eta	0:00:13
-----	1.9/11.1	MB	710.0	kB/s	eta	0:00:14
-----	1.9/11.1	MB	713.0	kB/s	eta	0:00:13
-----	1.9/11.1	MB	716.0	kB/s	eta	0:00:13
-----	1.9/11.1	MB	716.0	kB/s	eta	0:00:13
-----	2.0/11.1	MB	699.4	kB/s	eta	0:00:14
-----	2.0/11.1	MB	699.4	kB/s	eta	0:00:14
-----	2.0/11.1	MB	699.4	kB/s	eta	0:00:14
-----	2.0/11.1	MB	679.6	kB/s	eta	0:00:14
-----	2.0/11.1	MB	675.8	kB/s	eta	0:00:14
-----	2.0/11.1	MB	668.7	kB/s	eta	0:00:14
-----	2.1/11.1	MB	665.2	kB/s	eta	0:00:14
-----	2.1/11.1	MB	668.3	kB/s	eta	0:00:14
-----	2.1/11.1	MB	664.9	kB/s	eta	0:00:14
-----	2.1/11.1	MB	661.7	kB/s	eta	0:00:14
-----	2.2/11.1	MB	664.8	kB/s	eta	0:00:14
-----	2.2/11.1	MB	664.6	kB/s	eta	0:00:14
-----	2.2/11.1	MB	664.6	kB/s	eta	0:00:14
-----	2.3/11.1	MB	655.3	kB/s	eta	0:00:14
-----	2.3/11.1	MB	664.2	kB/s	eta	0:00:14
-----	2.3/11.1	MB	652.4	kB/s	eta	0:00:14
-----	2.3/11.1	MB	652.4	kB/s	eta	0:00:14
-----	2.4/11.1	MB	652.5	kB/s	eta	0:00:14
-----	2.4/11.1	MB	652.5	kB/s	eta	0:00:14
-----	2.4/11.1	MB	647.0	kB/s	eta	0:00:14
-----	2.4/11.1	MB	644.3	kB/s	eta	0:00:14
-----	2.5/11.1	MB	644.5	kB/s	eta	0:00:14
-----	2.5/11.1	MB	644.7	kB/s	eta	0:00:14
-----	2.5/11.1	MB	650.0	kB/s	eta	0:00:14
-----	2.6/11.1	MB	650.1	kB/s	eta	0:00:14
-----	2.6/11.1	MB	652.7	kB/s	eta	0:00:14
-----	2.6/11.1	MB	655.4	kB/s	eta	0:00:13
-----	2.7/11.1	MB	655.3	kB/s	eta	0:00:13
-----	2.7/11.1	MB	657.8	kB/s	eta	0:00:13
-----	2.7/11.1	MB	657.8	kB/s	eta	0:00:13
-----	2.8/11.1	MB	657.7	kB/s	eta	0:00:13
-----	2.8/11.1	MB	655.4	kB/s	eta	0:00:13
-----	2.8/11.1	MB	652.9	kB/s	eta	0:00:13
-----	2.9/11.1	MB	653.0	kB/s	eta	0:00:13
-----	2.9/11.1	MB	653.0	kB/s	eta	0:00:13
-----	2.9/11.1	MB	653.0	kB/s	eta	0:00:13
-----	3.0/11.1	MB	653.0	kB/s	eta	0:00:13
-----	3.0/11.1	MB	653.0	kB/s	eta	0:00:13
-----	3.0/11.1	MB	657.6	kB/s	eta	0:00:13
-----	3.1/11.1	MB	655.3	kB/s	eta	0:00:13
-----	3.1/11.1	MB	652.1	kB/s	eta	0:00:13

-----	3.1/11.1	MB	653.1	kB/s	eta	0:00:13
-----	3.1/11.1	MB	655.4	kB/s	eta	0:00:13
-----	3.1/11.1	MB	655.4	kB/s	eta	0:00:13
-----	3.1/11.1	MB	655.4	kB/s	eta	0:00:13
-----	3.2/11.1	MB	649.1	kB/s	eta	0:00:13
-----	3.2/11.1	MB	649.0	kB/s	eta	0:00:13
-----	3.2/11.1	MB	647.1	kB/s	eta	0:00:13
-----	3.3/11.1	MB	645.1	kB/s	eta	0:00:13
-----	3.3/11.1	MB	649.3	kB/s	eta	0:00:13
-----	3.3/11.1	MB	645.4	kB/s	eta	0:00:13
-----	3.4/11.1	MB	651.3	kB/s	eta	0:00:12
-----	3.4/11.1	MB	647.5	kB/s	eta	0:00:12
-----	3.4/11.1	MB	653.4	kB/s	eta	0:00:12
-----	3.5/11.1	MB	651.4	kB/s	eta	0:00:12
-----	3.5/11.1	MB	647.6	kB/s	eta	0:00:12
-----	3.5/11.1	MB	647.7	kB/s	eta	0:00:12
-----	3.5/11.1	MB	645.9	kB/s	eta	0:00:12
-----	3.6/11.1	MB	644.2	kB/s	eta	0:00:12
-----	3.6/11.1	MB	644.2	kB/s	eta	0:00:12
-----	3.6/11.1	MB	642.5	kB/s	eta	0:00:12
-----	3.6/11.1	MB	642.6	kB/s	eta	0:00:12
-----	3.7/11.1	MB	644.5	kB/s	eta	0:00:12
-----	3.7/11.1	MB	644.6	kB/s	eta	0:00:12
-----	3.7/11.1	MB	644.7	kB/s	eta	0:00:12
-----	3.8/11.1	MB	646.6	kB/s	eta	0:00:12
-----	3.8/11.1	MB	646.6	kB/s	eta	0:00:12
-----	3.9/11.1	MB	651.8	kB/s	eta	0:00:12
-----	3.9/11.1	MB	653.6	kB/s	eta	0:00:12
-----	3.9/11.1	MB	655.4	kB/s	eta	0:00:11
-----	4.0/11.1	MB	657.0	kB/s	eta	0:00:11
-----	4.0/11.1	MB	658.7	kB/s	eta	0:00:11
-----	4.1/11.1	MB	660.4	kB/s	eta	0:00:11
-----	4.1/11.1	MB	661.9	kB/s	eta	0:00:11
-----	4.1/11.1	MB	663.6	kB/s	eta	0:00:11
-----	4.2/11.1	MB	660.2	kB/s	eta	0:00:11
-----	4.2/11.1	MB	663.5	kB/s	eta	0:00:11
-----	4.2/11.1	MB	663.4	kB/s	eta	0:00:11
-----	4.3/11.1	MB	668.1	kB/s	eta	0:00:11
-----	4.4/11.1	MB	671.2	kB/s	eta	0:00:11
-----	4.4/11.1	MB	672.7	kB/s	eta	0:00:11
-----	4.4/11.1	MB	669.4	kB/s	eta	0:00:11
-----	4.5/11.1	MB	672.3	kB/s	eta	0:00:10
-----	4.5/11.1	MB	670.7	kB/s	eta	0:00:10
-----	4.5/11.1	MB	676.8	kB/s	eta	0:00:10
-----	4.5/11.1	MB	676.8	kB/s	eta	0:00:10
-----	4.6/11.1	MB	674.9	kB/s	eta	0:00:10
-----	4.7/11.1	MB	680.8	kB/s	eta	0:00:10
-----	4.7/11.1	MB	680.6	kB/s	eta	0:00:10
-----	4.8/11.1	MB	683.4	kB/s	eta	0:00:10
-----	4.8/11.1	MB	687.6	kB/s	eta	0:00:10
-----	4.8/11.1	MB	690.5	kB/s	eta	0:00:10
-----	4.9/11.1	MB	690.1	kB/s	eta	0:00:10
-----	4.9/11.1	MB	694.2	kB/s	eta	0:00:09
-----	5.0/11.1	MB	696.8	kB/s	eta	0:00:09
-----	5.0/11.1	MB	698.0	kB/s	eta	0:00:09
-----	5.1/11.1	MB	699.1	kB/s	eta	0:00:09
-----	5.1/11.1	MB	700.3	kB/s	eta	0:00:09
-----	5.2/11.1	MB	704.0	kB/s	eta	0:00:09
-----	5.2/11.1	MB	705.1	kB/s	eta	0:00:09
-----	5.3/11.1	MB	706.2	kB/s	eta	0:00:09
-----	5.3/11.1	MB	711.4	kB/s	eta	0:00:09
-----	5.4/11.1	MB	712.3	kB/s	eta	0:00:09
-----	5.4/11.1	MB	714.7	kB/s	eta	0:00:08
-----	5.5/11.1	MB	718.5	kB/s	eta	0:00:08
-----	5.6/11.1	MB	720.6	kB/s	eta	0:00:08
-----	5.6/11.1	MB	722.8	kB/s	eta	0:00:08
-----	5.7/11.1	MB	725.0	kB/s	eta	0:00:08
-----	5.7/11.1	MB	728.9	kB/s	eta	0:00:08
-----	5.7/11.1	MB	729.6	kB/s	eta	0:00:08
-----	5.8/11.1	MB	727.7	kB/s	eta	0:00:08
-----	5.8/11.1	MB	732.4	kB/s	eta	0:00:08
-----	5.9/11.1	MB	730.9	kB/s	eta	0:00:08
-----	5.9/11.1	MB	732.1	kB/s	eta	0:00:08

5.9/11.1	MB	733.1	kB/s	eta	0:00:08
5.9/11.1	MB	733.1	kB/s	eta	0:00:08
6.0/11.1	MB	734.8	kB/s	eta	0:00:07
6.0/11.1	MB	733.1	kB/s	eta	0:00:07
6.0/11.1	MB	732.4	kB/s	eta	0:00:07
6.1/11.1	MB	735.7	kB/s	eta	0:00:07
6.2/11.1	MB	737.8	kB/s	eta	0:00:07
6.2/11.1	MB	738.8	kB/s	eta	0:00:07
6.2/11.1	MB	735.6	kB/s	eta	0:00:07
6.3/11.1	MB	740.2	kB/s	eta	0:00:07
6.3/11.1	MB	740.9	kB/s	eta	0:00:07
6.3/11.1	MB	739.0	kB/s	eta	0:00:07
6.4/11.1	MB	742.3	kB/s	eta	0:00:07
6.4/11.1	MB	738.1	kB/s	eta	0:00:07
6.5/11.1	MB	741.4	kB/s	eta	0:00:07
6.5/11.1	MB	743.1	kB/s	eta	0:00:07
6.5/11.1	MB	744.1	kB/s	eta	0:00:07
6.6/11.1	MB	739.9	kB/s	eta	0:00:07
6.6/11.1	MB	737.1	kB/s	eta	0:00:07
6.6/11.1	MB	736.8	kB/s	eta	0:00:07
6.6/11.1	MB	736.4	kB/s	eta	0:00:07
6.7/11.1	MB	739.4	kB/s	eta	0:00:06
6.7/11.1	MB	740.0	kB/s	eta	0:00:06
6.8/11.1	MB	739.5	kB/s	eta	0:00:06
6.8/11.1	MB	744.8	kB/s	eta	0:00:06
6.8/11.1	MB	744.7	kB/s	eta	0:00:06
6.9/11.1	MB	746.1	kB/s	eta	0:00:06
7.0/11.1	MB	749.0	kB/s	eta	0:00:06
7.0/11.1	MB	749.9	kB/s	eta	0:00:06
7.1/11.1	MB	748.9	kB/s	eta	0:00:06
7.1/11.1	MB	751.9	kB/s	eta	0:00:06
7.2/11.1	MB	755.6	kB/s	eta	0:00:06
7.2/11.1	MB	757.2	kB/s	eta	0:00:06
7.3/11.1	MB	760.1	kB/s	eta	0:00:06
7.3/11.1	MB	758.4	kB/s	eta	0:00:05
7.4/11.1	MB	761.0	kB/s	eta	0:00:05
7.4/11.1	MB	763.9	kB/s	eta	0:00:05
7.5/11.1	MB	764.2	kB/s	eta	0:00:05
7.5/11.1	MB	768.0	kB/s	eta	0:00:05
7.6/11.1	MB	772.7	kB/s	eta	0:00:05
7.7/11.1	MB	772.1	kB/s	eta	0:00:05
7.7/11.1	MB	776.8	kB/s	eta	0:00:05
7.8/11.1	MB	778.2	kB/s	eta	0:00:05
7.9/11.1	MB	780.7	kB/s	eta	0:00:05
7.9/11.1	MB	780.3	kB/s	eta	0:00:05
7.9/11.1	MB	779.3	kB/s	eta	0:00:05
8.0/11.1	MB	783.3	kB/s	eta	0:00:05
8.0/11.1	MB	784.4	kB/s	eta	0:00:04
8.1/11.1	MB	786.8	kB/s	eta	0:00:04
8.1/11.1	MB	786.2	kB/s	eta	0:00:04
8.2/11.1	MB	792.6	kB/s	eta	0:00:04
8.3/11.1	MB	791.9	kB/s	eta	0:00:04
8.3/11.1	MB	796.2	kB/s	eta	0:00:04
8.4/11.1	MB	799.5	kB/s	eta	0:00:04
8.4/11.1	MB	801.2	kB/s	eta	0:00:04
8.5/11.1	MB	797.2	kB/s	eta	0:00:04
8.5/11.1	MB	799.7	kB/s	eta	0:00:04
8.6/11.1	MB	802.7	kB/s	eta	0:00:04
8.6/11.1	MB	802.5	kB/s	eta	0:00:04
8.6/11.1	MB	800.9	kB/s	eta	0:00:04
8.7/11.1	MB	800.8	kB/s	eta	0:00:04
8.7/11.1	MB	802.4	kB/s	eta	0:00:03
8.8/11.1	MB	806.1	kB/s	eta	0:00:03
8.9/11.1	MB	808.2	kB/s	eta	0:00:03
9.0/11.1	MB	811.3	kB/s	eta	0:00:03
9.0/11.1	MB	815.3	kB/s	eta	0:00:03
9.1/11.1	MB	820.3	kB/s	eta	0:00:03
9.2/11.1	MB	823.1	kB/s	eta	0:00:03
9.3/11.1	MB	827.4	kB/s	eta	0:00:03
9.4/11.1	MB	830.8	kB/s	eta	0:00:03
9.5/11.1	MB	835.5	kB/s	eta	0:00:02
9.6/11.1	MB	842.8	kB/s	eta	0:00:02
9.6/11.1	MB	842.8	kB/s	eta	0:00:02

```
----- 9.6/11.1 MB 842.9 kB/s eta 0:00:02
----- 9.7/11.1 MB 846.9 kB/s eta 0:00:02
----- 9.8/11.1 MB 853.9 kB/s eta 0:00:02
----- 9.9/11.1 MB 853.1 kB/s eta 0:00:02
----- 9.9/11.1 MB 856.7 kB/s eta 0:00:02
----- 10.1/11.1 MB 864.7 kB/s eta 0:00:02
----- 10.1/11.1 MB 867.0 kB/s eta 0:00:02
----- 10.3/11.1 MB 874.1 kB/s eta 0:00:01
----- 10.3/11.1 MB 881.1 kB/s eta 0:00:01
----- 10.5/11.1 MB 908.1 kB/s eta 0:00:01
----- 10.5/11.1 MB 914.4 kB/s eta 0:00:01
----- 10.6/11.1 MB 918.2 kB/s eta 0:00:01
----- 10.8/11.1 MB 923.4 kB/s eta 0:00:01
----- 10.9/11.1 MB 930.0 kB/s eta 0:00:01
----- 11.0/11.1 MB 938.0 kB/s eta 0:00:01
----- 11.1/11.1 MB 939.3 kB/s eta 0:00:01
----- 11.1/11.1 MB 938.5 kB/s eta 0:00:00
```

Downloading scipy-1.15.2-cp312-cp312-win_amd64.whl (40.9 MB)

```
----- 0.0/40.9 MB ? eta -:--:--
----- 0.1/40.9 MB 2.6 MB/s eta 0:00:16
----- 0.2/40.9 MB 3.4 MB/s eta 0:00:12
----- 0.4/40.9 MB 3.0 MB/s eta 0:00:14
----- 0.4/40.9 MB 2.6 MB/s eta 0:00:16
----- 0.5/40.9 MB 2.6 MB/s eta 0:00:16
----- 0.6/40.9 MB 2.6 MB/s eta 0:00:16
----- 0.7/40.9 MB 2.4 MB/s eta 0:00:17
----- 0.8/40.9 MB 2.4 MB/s eta 0:00:17
----- 1.0/40.9 MB 2.5 MB/s eta 0:00:16
----- 1.1/40.9 MB 2.6 MB/s eta 0:00:16
----- 1.2/40.9 MB 2.5 MB/s eta 0:00:16
----- 1.3/40.9 MB 2.4 MB/s eta 0:00:17
----- 1.5/40.9 MB 2.6 MB/s eta 0:00:16
----- 1.6/40.9 MB 2.6 MB/s eta 0:00:15
----- 1.8/40.9 MB 2.7 MB/s eta 0:00:15
----- 1.8/40.9 MB 2.7 MB/s eta 0:00:15
----- 2.1/40.9 MB 2.8 MB/s eta 0:00:15
----- 2.1/40.9 MB 2.6 MB/s eta 0:00:15
----- 2.3/40.9 MB 2.6 MB/s eta 0:00:15
----- 2.3/40.9 MB 2.7 MB/s eta 0:00:15
----- 2.5/40.9 MB 2.6 MB/s eta 0:00:16
----- 2.6/40.9 MB 2.6 MB/s eta 0:00:15
----- 2.6/40.9 MB 2.6 MB/s eta 0:00:15
----- 2.6/40.9 MB 2.6 MB/s eta 0:00:15
----- 2.6/40.9 MB 2.3 MB/s eta 0:00:17
----- 2.7/40.9 MB 2.3 MB/s eta 0:00:17
----- 2.8/40.9 MB 2.3 MB/s eta 0:00:17
----- 2.8/40.9 MB 2.2 MB/s eta 0:00:18
----- 2.9/40.9 MB 2.2 MB/s eta 0:00:18
----- 3.0/40.9 MB 2.2 MB/s eta 0:00:18
----- 3.2/40.9 MB 2.2 MB/s eta 0:00:17
----- 3.3/40.9 MB 2.2 MB/s eta 0:00:17
----- 3.3/40.9 MB 2.2 MB/s eta 0:00:17
----- 3.5/40.9 MB 2.2 MB/s eta 0:00:17
----- 3.7/40.9 MB 2.3 MB/s eta 0:00:17
----- 3.9/40.9 MB 2.4 MB/s eta 0:00:16
----- 3.9/40.9 MB 2.3 MB/s eta 0:00:16
----- 4.0/40.9 MB 2.3 MB/s eta 0:00:17
----- 4.2/40.9 MB 2.4 MB/s eta 0:00:16
----- 4.3/40.9 MB 2.3 MB/s eta 0:00:16
----- 4.5/40.9 MB 2.4 MB/s eta 0:00:16
----- 4.8/40.9 MB 2.4 MB/s eta 0:00:15
----- 4.9/40.9 MB 2.5 MB/s eta 0:00:15
----- 5.1/40.9 MB 2.5 MB/s eta 0:00:15
----- 5.4/40.9 MB 2.6 MB/s eta 0:00:14
----- 5.5/40.9 MB 2.6 MB/s eta 0:00:14
----- 5.7/40.9 MB 2.6 MB/s eta 0:00:14
----- 5.9/40.9 MB 2.7 MB/s eta 0:00:14
----- 6.1/40.9 MB 2.7 MB/s eta 0:00:13
----- 6.3/40.9 MB 2.7 MB/s eta 0:00:13
----- 6.5/40.9 MB 2.8 MB/s eta 0:00:13
----- 6.6/40.9 MB 2.8 MB/s eta 0:00:13
----- 6.8/40.9 MB 2.8 MB/s eta 0:00:13
----- 7.0/40.9 MB 2.8 MB/s eta 0:00:13
```

		7.2/40.9	MB	2.9	MB/s	eta	0:00:12
		7.2/40.9	MB	2.9	MB/s	eta	0:00:12
		7.3/40.9	MB	2.8	MB/s	eta	0:00:12
		7.4/40.9	MB	2.8	MB/s	eta	0:00:13
		7.4/40.9	MB	2.8	MB/s	eta	0:00:13
		7.5/40.9	MB	2.7	MB/s	eta	0:00:13
		7.5/40.9	MB	2.7	MB/s	eta	0:00:13
		7.5/40.9	MB	2.7	MB/s	eta	0:00:13
		7.5/40.9	MB	2.7	MB/s	eta	0:00:13
		8.0/40.9	MB	2.8	MB/s	eta	0:00:12
		8.1/40.9	MB	2.8	MB/s	eta	0:00:12
		8.1/40.9	MB	2.8	MB/s	eta	0:00:12
		8.1/40.9	MB	2.8	MB/s	eta	0:00:12
		8.1/40.9	MB	2.6	MB/s	eta	0:00:13
		8.1/40.9	MB	2.6	MB/s	eta	0:00:13
		8.1/40.9	MB	2.6	MB/s	eta	0:00:13
		8.1/40.9	MB	2.6	MB/s	eta	0:00:13
		8.1/40.9	MB	2.6	MB/s	eta	0:00:13
		8.1/40.9	MB	2.6	MB/s	eta	0:00:13
		8.1/40.9	MB	2.6	MB/s	eta	0:00:13
		8.5/40.9	MB	2.5	MB/s	eta	0:00:14
		8.5/40.9	MB	2.5	MB/s	eta	0:00:14
		8.6/40.9	MB	2.4	MB/s	eta	0:00:14
		8.7/40.9	MB	2.4	MB/s	eta	0:00:14
		8.7/40.9	MB	2.4	MB/s	eta	0:00:14
		8.7/40.9	MB	2.4	MB/s	eta	0:00:14
		8.7/40.9	MB	2.4	MB/s	eta	0:00:14
		8.7/40.9	MB	2.4	MB/s	eta	0:00:14
		8.9/40.9	MB	2.3	MB/s	eta	0:00:14
		8.9/40.9	MB	2.3	MB/s	eta	0:00:14
		9.0/40.9	MB	2.3	MB/s	eta	0:00:14
		9.2/40.9	MB	2.3	MB/s	eta	0:00:14
		9.3/40.9	MB	2.3	MB/s	eta	0:00:14
		9.3/40.9	MB	2.3	MB/s	eta	0:00:14
		9.3/40.9	MB	2.3	MB/s	eta	0:00:14
		9.3/40.9	MB	2.3	MB/s	eta	0:00:14
		9.3/40.9	MB	2.2	MB/s	eta	0:00:15
		9.4/40.9	MB	2.2	MB/s	eta	0:00:15
		9.5/40.9	MB	2.2	MB/s	eta	0:00:15
		9.7/40.9	MB	2.3	MB/s	eta	0:00:14
		9.8/40.9	MB	2.2	MB/s	eta	0:00:14
		9.8/40.9	MB	2.2	MB/s	eta	0:00:14
		10.0/40.9	MB	2.3	MB/s	eta	0:00:14
		10.2/40.9	MB	2.3	MB/s	eta	0:00:14
		10.3/40.9	MB	2.3	MB/s	eta	0:00:14
		10.5/40.9	MB	2.3	MB/s	eta	0:00:14
		10.6/40.9	MB	2.3	MB/s	eta	0:00:14
		10.8/40.9	MB	2.3	MB/s	eta	0:00:14
		10.9/40.9	MB	2.3	MB/s	eta	0:00:13
		11.1/40.9	MB	2.3	MB/s	eta	0:00:13
		11.3/40.9	MB	2.3	MB/s	eta	0:00:13
		11.4/40.9	MB	2.3	MB/s	eta	0:00:13
		11.6/40.9	MB	2.3	MB/s	eta	0:00:13
		11.7/40.9	MB	2.3	MB/s	eta	0:00:13
		11.9/40.9	MB	2.3	MB/s	eta	0:00:13
		12.1/40.9	MB	2.3	MB/s	eta	0:00:13
		12.1/40.9	MB	2.4	MB/s	eta	0:00:13
		12.1/40.9	MB	2.4	MB/s	eta	0:00:13
		12.1/40.9	MB	2.4	MB/s	eta	0:00:13
		12.1/40.9	MB	2.4	MB/s	eta	0:00:13
		12.2/40.9	MB	2.2	MB/s	eta	0:00:13
		12.2/40.9	MB	2.2	MB/s	eta	0:00:13
		12.3/40.9	MB	2.2	MB/s	eta	0:00:13
		12.5/40.9	MB	2.2	MB		

							22.2/40.9	MB	2.5	MB/s	eta	0:00:08
							22.3/40.9	MB	2.4	MB/s	eta	0:00:08
							22.5/40.9	MB	2.6	MB/s	eta	0:00:08
							22.6/40.9	MB	2.6	MB/s	eta	0:00:08
							22.7/40.9	MB	2.6	MB/s	eta	0:00:08
							22.9/40.9	MB	2.6	MB/s	eta	0:00:07
							23.1/40.9	MB	2.7	MB/s	eta	0:00:07
							23.2/40.9	MB	2.6	MB/s	eta	0:00:07
							23.3/40.9	MB	2.7	MB/s	eta	0:00:07
							23.4/40.9	MB	2.7	MB/s	eta	0:00:07
							23.6/40.9	MB	2.7	MB/s	eta	0:00:07
							23.6/40.9	MB	2.7	MB/s	eta	0:00:07
							23.7/40.9	MB	2.7	MB/s	eta	0:00:07
							23.9/40.9	MB	2.7	MB/s	eta	0:00:07
							24.0/40.9	MB	2.7	MB/s	eta	0:00:07
							24.1/40.9	MB	2.7	MB/s	eta	0:00:07
							24.2/40.9	MB	2.7	MB/s	eta	0:00:07
							24.3/40.9	MB	2.7	MB/s	eta	0:00:07
							24.5/40.9	MB	2.7	MB/s	eta	0:00:07
							24.6/40.9	MB	2.7	MB/s	eta	0:00:07
							24.8/40.9	MB	2.7	MB/s	eta	0:00:07
							25.0/40.9	MB	2.7	MB/s	eta	0:00:06
							25.0/40.9	MB	2.7	MB/s	eta	0:00:07
							25.2/40.9	MB	2.7	MB/s	eta	0:00:06
							25.3/40.9	MB	2.6	MB/s	eta	0:00:06
							25.4/40.9	MB	2.6	MB/s	eta	0:00:06
							25.5/40.9	MB	2.6	MB/s	eta	0:00:06
							25.6/40.9	MB	2.6	MB/s	eta	0:00:06
							25.8/40.9	MB	2.7	MB/s	eta	0:00:06
							26.0/40.9	MB	2.7	MB/s	eta	0:00:06
							26.2/40.9	MB	2.7	MB/s	eta	0:00:06
							26.4/40.9	MB	2.7	MB/s	eta	0:00:06
							26.5/40.9	MB	2.7	MB/s	eta	0:00:06
							26.7/40.9	MB	2.7	MB/s	eta	0:00:06
							26.8/40.9	MB	2.7	MB/s	eta	0:00:06
							27.0/40.9	MB	2.7	MB/s	eta	0:00:06
							27.1/40.9	MB	2.7	MB/s	eta	0:00:06
							27.2/40.9	MB	2.7	MB/s	eta	0:00:06
							27.3/40.9	MB	2.7	MB/s	eta	0:00:06
							27.4/40.9	MB	2.7	MB/s	eta	0:00:06
							27.5/40.9	MB	2.6	MB/s	eta	0:00:06
							27.7/40.9	MB	2.7	MB/s	eta	0:00:05
							27.7/40.9	MB	2.6	MB/s	eta	0:00:06
							27.9/40.9	MB	2.7	MB/s	eta	0:00:05
							27.9/40.9	MB	2.6	MB/s	eta	0:00:05
							28.2/40.9	MB	2.7	MB/s	eta	0:00:05
							28.3/40.9	MB	2.7	MB/s	eta	0:00:05
							28.5/40.9	MB	2.7	MB/s	eta	0:00:05

[illegible]

```

----- -- 38.3/40.9 MB 2.2 MB/s eta 0:00:02
----- -- 38.4/40.9 MB 2.2 MB/s eta 0:00:02
----- -- 38.4/40.9 MB 2.2 MB/s eta 0:00:02
----- -- 38.4/40.9 MB 2.2 MB/s eta 0:00:02
----- -- 38.7/40.9 MB 2.2 MB/s eta 0:00:02
----- -- 38.7/40.9 MB 2.1 MB/s eta 0:00:02
----- -- 38.8/40.9 MB 2.2 MB/s eta 0:00:02
----- -- 38.8/40.9 MB 2.2 MB/s eta 0:00:01
----- -- 38.9/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.0/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.1/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.1/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.2/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.3/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.4/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.5/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.7/40.9 MB 2.2 MB/s eta 0:00:01
----- - 39.8/40.9 MB 2.1 MB/s eta 0:00:01
----- - 39.8/40.9 MB 2.1 MB/s eta 0:00:01
----- 39.9/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.0/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.2/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.2/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.2/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.2/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.4/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.5/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.5/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.7/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.7/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.7/40.9 MB 2.0 MB/s eta 0:00:01
----- 40.9/40.9 MB 2.0 MB/s eta 0:00:01
----- 40.9/40.9 MB 2.1 MB/s eta 0:00:01
----- 40.9/40.9 MB 2.0 MB/s eta 0:00:00

```

Downloading threadpoolctl-3.6.0-py3-none-any.whl (18 kB)

Installing collected packages: threadpoolctl, scipy, scikit-learn

Successfully installed scikit-learn-1.6.1 scipy-1.15.2 threadpoolctl-3.6.0

```
[notice] A new release of pip is available: 23.2.1 -> 25.0.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

In [22]:

```

from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values=np.nan, strategy='mean')
imputer.fit(dataset.iloc[:, 4:6])
dataset.iloc[:, 4:6] = imputer.transform(dataset.iloc[:, 4:6])

```

In [23]:

```
print(X)
```

	Employee_ID	Age	Department	Job_Role	Experience_Years	Salary
0	1001	50	Sales	Engineer	15.0	94895.0
1	1002	36	Engineering	Manager	12.0	62307.0
2	1003	29	Sales	Executive	29.0	84098.0
3	1004	42	Marketing	Executive	18.0	90921.0
4	1005	40	Sales	Engineer	16.0	35486.0
..
115	1116	53	Engineering	Executive	8.0	91529.0
116	1117	53	Engineering	Consultant	6.0	114665.0
117	1118	25	Marketing	Executive	27.0	61921.0
118	1119	51	HR	Engineer	13.0	63307.0
119	1120	58	Marketing	Executive	30.0	57355.0

[120 rows x 6 columns]

Step 3 : Encoding categorical data

In [40]:

```
dataset['Department'].value_counts()
```

Out[40]:

```
Department
HR          32
Marketing   30
Engineering 27
Finance     17
Sales       14
Name: count, dtype: int64
```

In [42]:

```
from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import OneHotEncoder

ct = ColumnTransformer(transformers=[('encoder',
                                     OneHotEncoder(),
                                     [0])],
                       remainder='passthrough')
X = np.array(ct.fit_transform(X))
```

In [30]:

```
print(X)
```

```
[[1.0000e+00 0.0000e+00 1.0000e+00 ... 1.5000e+01 9.4895e+04 3.0000e+00]
 [1.0000e+00 0.0000e+00 1.0000e+00 ... 1.2000e+01 6.2307e+04 2.0000e+00]
 [1.0000e+00 0.0000e+00 1.0000e+00 ... 2.9000e+01 8.4098e+04 3.0000e+00]
 ...
 [1.0000e+00 0.0000e+00 0.0000e+00 ... 2.7000e+01 6.1921e+04 2.0000e+00]
 [0.0000e+00 1.0000e+00 1.0000e+00 ... 1.3000e+01 6.3307e+04 3.0000e+00]
 [1.0000e+00 0.0000e+00 0.0000e+00 ... 3.0000e+01 5.7355e+04 1.0000e+00]]
```

In [31]:

```
Y.value_counts()
```

Out[31]:

```
Performance_Rating
3      44
4      29
2      18
1      16
5      13
Name: count, dtype: int64
```

In [32]:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
Y = le.fit_transform(Y)
```

In [33]:

```
print(Y)
```

```
[2 1 2 0 1 2 3 1 1 0 4 2 2 4 0 4 2 1 2 3 2 4 2 3 2 3 2 0 0 1 0 0 3 2 3 4 2
 3 3 2 1 1 1 1 1 4 3 0 1 0 2 2 3 2 3 4 0 3 2 2 2 3 3 3 4 3 4 3 3 2 0 3 2 2
 4 1 2 3 3 2 3 3 0 0 3 2 1 2 2 2 2 2 2 3 1 3 2 4 2 4 3 1 3 2 2 0 2 2 2 4 1
 2 2 0 2 3 2 1 2 0]
```

Step 4 : Splitting Data into Training and Testing

In [35]:

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, Y_train, Y_test = train_test_split(X,
                                                    dataset['Performance_Rating'],
                                                    test_size=0.3,
                                                    random_state=1)
```

In [36]:

```
print(X_train)
```

```
[[0.0000e+00 1.0000e+00 1.0000e+00 ... 4.0000e+00 7.4261e+04 4.0000e+00]
 [1.0000e+00 0.0000e+00 0.0000e+00 ... 2.5000e+01 9.6040e+04 4.0000e+00]
 [1.0000e+00 0.0000e+00 1.0000e+00 ... 3.2000e+01 9.2856e+04 4.0000e+00]
 ...
 [1.0000e+00 0.0000e+00 1.0000e+00 ... 0.0000e+00 5.7192e+04 3.0000e+00]
 [0.0000e+00 1.0000e+00 1.0000e+00 ... 1.9000e+01 4.0975e+04 3.0000e+00]
 [1.0000e+00 0.0000e+00 0.0000e+00 ... 2.0000e+01 6.9081e+04 4.0000e+00]]
```

Step 5 : Feature Scaling

In [37]:

```
from sklearn.preprocessing import StandardScaler

sc = StandardScaler()
X_train[:, 6:] = sc.fit_transform(X_train[:, 6:])
X_test[:, 6:] = sc.transform(X_test[:, 6:])
```

In [38]:

```
print(X_train)
```

```
[[ 0.          1.          1.          ... -1.481062   -0.08579889
  0.84452403]
 [ 1.          0.          0.          ...  0.67920128   0.73191686
  0.84452403]
 [ 1.          0.          1.          ...  1.39928904   0.61237018
  0.84452403]
 ...
 [ 1.          0.          1.          ... -1.89254072  -0.72667267
 -0.01017499]
 [ 0.          1.          1.          ...  0.0619832   -1.33555721
 -0.01017499]
 [ 1.          0.          0.          ...  0.16485288  -0.28028751
  0.84452403]]
```

In []: