practical-3-dsbda

May 4, 2025

```python
[ ]: #Practicle 3
     import pandas as pd
     import numpy as np
```

```python
[ ]: df = pd.read_csv("Employee_Salary_Dataset.csv")
     df.head()
```

```
[ ]:    ID  Experience_Years  Age  Gender  Salary
     0   1                 5   28  Female  250000
     1   2                 1   21    Male   50000
     2   3                 3   23  Female  170000
     3   4                 2   22    Male   25000
     4   5                 1   17    Male   10000
```

```python
[ ]: grouped = df.groupby('Gender')['Salary']

     summery_stats= grouped.agg(['mean','median','min','max','std'])

     print("\nSummery statistics of Salary grouped by Gender :")
     print(summery_stats)
```

```
Summery statistics of Salary grouped by Gender :
                mean    median   min       max           std
Gender
Female  2.054917e+06  250000.0  6000  10000000  3.450120e+06
Male    2.063626e+06  220100.0  3000   7600000  2.950974e+06
```

```python
[ ]: grouped1 = df.groupby('Experience_Years')['Salary']
     salary_list_by_exp = grouped1.apply(list)
     print("Salary list grouped by Experience years : ")
     print(salary_list_by_exp)
```

```
Salary list grouped by Experience years :
Experience_Years
1            [50000, 10000, 6000, 3000]
2         [25000, 9000, 7500, 15000, 6100]
3                      [170000, 20000]
```

```
4                [250000, 87000, 25000, 8900]
5                    [250000, 6845000]
6                              [1400000]
10          [61500, 930000, 330000, 80000]
11                             [220100]
14                             [6000050]
15     [650000, 7900000, 6570000, 900000]
16                             [7600000]
19             [800000, 5000000, 9300000]
20                             [1540000]
25                             [5001000]
27                            [10000000]
Name: Salary, dtype: object
```

```python
iris = pd.read_csv("Iris.csv")
```

```python
iris.columns
```

```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
```

```python
print("Species in dataset : ",iris['Species'].unique())
```

```
Species in dataset :  ['Iris-setosa' 'Iris-versicolor' 'Iris-virginica']
```

```python
group_iris = iris.groupby('Species')

for species, data in group_iris:
    print(f"\nDescriptive statistics of {species}")
    print(data.describe())
```

```
Descriptive statistics of Iris-setosa
             Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
count  50.00000       50.00000     50.000000      50.000000      50.00000
mean   25.50000        5.00600      3.418000       1.464000       0.24400
std    14.57738        0.35249      0.381024       0.173511       0.10721
min     1.00000        4.30000      2.300000       1.000000       0.10000
25%    13.25000        4.80000      3.125000       1.400000       0.20000
50%    25.50000        5.00000      3.400000       1.500000       0.20000
75%    37.75000        5.20000      3.675000       1.575000       0.30000
max    50.00000        5.80000      4.400000       1.900000       0.60000

Descriptive statistics of Iris-versicolor
             Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
count  50.00000      50.000000     50.000000      50.000000     50.000000
mean   75.50000       5.936000      2.770000       4.260000      1.326000
```

	std	14.57738	0.516171	0.313798	0.469911	0.197753
min	51.00000	4.900000	2.000000	3.000000	1.000000
25%	63.25000	5.600000	2.525000	4.000000	1.200000
50%	75.50000	5.900000	2.800000	4.350000	1.300000
75%	87.75000	6.300000	3.000000	4.600000	1.500000
max	100.00000	7.000000	3.400000	5.100000	1.800000

Descriptive statistics of Iris-virginica

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	50.00000	50.00000	50.000000	50.000000	50.00000
mean	125.50000	6.58800	2.974000	5.552000	2.02600
std	14.57738	0.63588	0.322497	0.551895	0.27465
min	101.00000	4.90000	2.200000	4.500000	1.40000
25%	113.25000	6.22500	2.800000	5.100000	1.80000
50%	125.50000	6.50000	3.000000	5.550000	2.00000
75%	137.75000	6.90000	3.175000	5.875000	2.30000
max	150.00000	7.90000	3.800000	6.900000	2.50000

[ ]: