**Assignment No 3**

**Code:**

# Assignment number 3:

# Load the dataset

# Display basic information

# Display null values

# Fill the null values

# Display overall statistical information

# Display groupwise statistical information

#---------------------------------------------------------------------------------------

import pandas as pd

import numpy as np

df = pd.read\_csv('Employee\_Salary.csv')

#---------------------------------------------------------------------------------------

# Display basic information

print('Information of Dataset:\n', df.info)

print('Shape of Dataset (row x column): ', df.shape)

print('Columns Name: ', df.columns)

print('Total elements in dataset:', df.size)

print('Datatype of attributes (columns):', df.dtypes)

print('First 5 rows:\n', df.head().T)

print('Last 5 rows:\n',df.tail().T)

print('Any 5 rows:\n',df.sample(5).T)

#---------------------------------------------------------------------------------------

# Display Null values

print('Total Number of Null Values in Dataset:', df.isna().sum())

#---------------------------------------------------------------------------------------

# Fill the missing values

df['Gender'].fillna(df['Gender'].mode()[0], inplace=True)

df['Experience'].fillna(df['Experience'].mean(), inplace=True)

print('Total Number of Null Values in Dataset:', df.isna().sum())

#---------------------------------------------------------------------------------------

# Display Overall Statistical information

print('Statistical information of Numerical Columns: \n',df.describe())

#---------------------------------------------------------------------------------------

# groupwise statistical information

print('Groupwise Statistical Summary....')

print('\n-------------------------- Experience -----------------------\n')

print(df['Experience'].groupby(df['Gender']).describe())

print('\n-------------------------- Age -----------------------\n')

print(df['Age'].groupby(df['Gender']).describe())

print('\n-------------------------- Salary -----------------------\n')

print(df['Salary'].groupby(df['Gender']).describe())

#---------------------------------------------------------------------------------------

df = pd.read\_csv('iris.csv')

df = df.drop('Id', axis=1)

df.columns = ('SL', 'SW', 'PL', 'PW', 'Species')

print(df.head().T)

# Display Statistical information

print('Statistical information of Numerical Columns: \n',df.describe())

print('Groupwise Statistical Summary....')

print('\n-------------------------- Sepal Length -----------------------\n')

print(df['SL'].groupby(df['Species']).describe())

print('\n-------------------------- Sepal Width -----------------------\n')

print(df['SW'].groupby(df['Species']).describe())

print('\n-------------------------- Petal Length -----------------------\n')

print(df['PL'].groupby(df['Species']).describe())

print('\n-------------------------- Petal Width -----------------------\n')

print(df['SW'].groupby(df['Species']).describe())

**Output:**

Information of Dataset:

 <bound method DataFrame.info of     ID  Experience  Age  Gender    Salary

0    1         5.0   28  Female    250000

1    2         1.0   21    Male     50000

2    3         3.0   23  Female    170000

3    4         2.0   22    Male     25000

4    5         1.0   17    Male     10000

5    6        25.0   62    Male   5001000

6    7        19.0   54  Female    800000

7    8         2.0   21  Female      9000

8    9        10.0   36  Female     61500

9   10        15.0   54  Female    650000

10  11         4.0   26  Female    250000

11  12         6.0   29    Male   1400000

12  13        14.0   39    Male   6000050

13  14         NaN   40    Male    220100

14  15         2.0   23    Male      7500

15  16         4.0   27     NaN     87000

16  17        10.0   34  Female    930000

17  18        15.0   54  Female   7900000

18  19         2.0   21    Male     15000

19  20        10.0   36    Male    330000

20  21        15.0   54    Male   6570000

21  22         4.0   26    Male     25000

22  23         5.0   29    Male   6845000

23  24         1.0   21  Female      6000

24  25         4.0   23  Female      8900

25  26         3.0   22  Female     20000

26  27         1.0   18    Male      3000

27  28        27.0   62  Female  10000000

28  29        19.0   54  Female   5000000

29  30         2.0   21  Female      6100

30  31        10.0   34    Male     80000

31  32        15.0   54    Male    900000

32  33        20.0   55  Female   1540000

33  34        19.0   53  Female   9300000

34  35        16.0   49    Male   7600000>

Shape of Dataset (row x column):  (35, 5)

Columns Name:  Index(['ID', 'Experience', 'Age', 'Gender', 'Salary'], dtype='object')

Total elements in dataset: 175

Datatype of attributes (columns): ID              int64

Experience    float64

Age             int64

Gender         object

Salary          int64

dtype: object

First 5 rows:

                  0      1       2      3      4

ID               1      2       3      4      5

Experience     5.0    1.0     3.0    2.0    1.0

Age             28     21      23     22     17

Gender      Female   Male  Female   Male   Male

Salary      250000  50000  170000  25000  10000

Last 5 rows:

                30      31       32       33       34

ID             31      32       33       34       35

Experience   10.0    15.0     20.0     19.0     16.0

Age            34      54       55       53       49

Gender       Male    Male   Female   Female     Male

Salary      80000  900000  1540000  9300000  7600000

Any 5 rows:

                15      23    26      0       29

ID             16      24    27       1      30

Experience    4.0     1.0   1.0     5.0     2.0

Age            27      21    18      28      21

Gender        NaN  Female  Male  Female  Female

Salary      87000    6000  3000  250000    6100

Total Number of Null Values in Dataset: ID            0

Experience    1

Age           0

Gender        1

Salary        0

dtype: int64

Total Number of Null Values in Dataset: ID            0

Experience    1

Age           0

Gender        1

Salary        0

dtype: int64

Statistical information of Numerical Columns:

               ID  Experience        Age        Salary

count  35.000000   34.000000  35.000000  3.500000e+01

mean   18.000000    9.147059  35.485714  2.059147e+06

std    10.246951    7.659940  14.643552  3.170124e+06

min     1.000000    1.000000  17.000000  3.000000e+03

25%     9.500000    2.250000  22.500000  2.250000e+04

50%    18.000000    5.500000  29.000000  2.500000e+05

75%    26.500000   15.000000  53.500000  3.270000e+06

max    35.000000   27.000000  62.000000  1.000000e+07

Groupwise Statistical Summary....

-------------------------- Experience -----------------------

        count       mean       std  min  25%   50%    75%   max

Gender

Female   17.0  10.470588  8.224336  1.0  3.0  10.0  19.00  27.0

Male     16.0   8.062500  7.206189  1.0  2.0   5.5  14.25  25.0

-------------------------- Age -----------------------

        count       mean        std   min   25%   50%   75%   max

Gender

Female   17.0  37.705882  15.711320  21.0  23.0  34.0  54.0  62.0

Male     17.0  33.764706  13.997899  17.0  22.0  29.0  40.0  62.0

-------------------------- Salary -----------------------

        count          mean           std  ...       50%        75%         max

Gender                                     ...

Female   17.0  2.170676e+06  3.520086e+06  ...  250000.0  1540000.0  10000000.0

Male     17.0  2.063626e+06  2.950974e+06  ...  220100.0  5001000.0   7600000.0

[2 rows x 8 columns]

              0       1       2       3       4

SL          5.1     4.9     4.7     4.6     5.0

SW          3.5     3.0     3.2     3.1     3.6

PL          1.4     1.4     1.3     1.5     1.4

PW          0.2     0.2     0.2     0.2     0.2

Species  Setosa  Setosa  Setosa  Setosa  Setosa

Statistical information of Numerical Columns:

                SL          SW          PL          PW

count  148.000000  150.000000  150.000000  148.000000

mean     5.856757    3.054000    3.758667    1.197973

std      0.825459    0.433594    1.764420    0.760278

min      4.300000    2.000000    1.000000    0.100000

25%      5.100000    2.800000    1.600000    0.300000

50%      5.800000    3.000000    4.350000    1.300000

75%      6.400000    3.300000    5.100000    1.800000

max      7.900000    4.400000    6.900000    2.500000

Groupwise Statistical Summary....

-------------------------- Sepal Length -----------------------

            count    mean       std  min    25%  50%  75%  max

Species

Setosa       48.0  5.0125  0.358261  4.3  4.800  5.0  5.2  5.8

Versicolor   50.0  5.9360  0.516171  4.9  5.600  5.9  6.3  7.0

Virginica    50.0  6.5880  0.635880  4.9  6.225  6.5  6.9  7.9

-------------------------- Sepal Width -----------------------

            count   mean       std  min    25%  50%    75%  max

Species

Setosa       50.0  3.418  0.381024  2.3  3.125  3.4  3.675  4.4

Versicolor   50.0  2.770  0.313798  2.0  2.525  2.8  3.000  3.4

Virginica    50.0  2.974  0.322497  2.2  2.800  3.0  3.175  3.8

-------------------------- Petal Length -----------------------

            count   mean       std  min  25%   50%    75%  max

Species

Setosa       50.0  1.464  0.173511  1.0  1.4  1.50  1.575  1.9

Versicolor   50.0  4.260  0.469911  3.0  4.0  4.35  4.600  5.1

Virginica    50.0  5.552  0.551895  4.5  5.1  5.55  5.875  6.9

-------------------------- Petal Width -----------------------

            count   mean       std  min    25%  50%    75%  max

Species

Setosa       50.0  3.418  0.381024  2.3  3.125  3.4  3.675  4.4

Versicolor   50.0  2.770  0.313798  2.0  2.525  2.8  3.000  3.4

Virginica    50.0  2.974  0.322497  2.2  2.800  3.0  3.175  3.8