

```
import numpy as np
import pandas as pd
```

```
df = pd.read_csv("/content/Employee_Salary.csv")
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   ID           35 non-null    int64
1   Experience   34 non-null    float64
2   Age          35 non-null    int64
3   Gender       34 non-null    object
4   Salary       35 non-null    int64
dtypes: float64(1), int64(3), object(1)
memory usage: 1.5+ KB
Information of Dataset:
None
```

```
print("Shape of Dataset: ", df.shape)
```

```
Shape of Dataset: (35, 5)
```

```
print("Columns name: ", df.columns)
```

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

```
print("Total elements: ", df.size)
```

```
Total elements: 175
```

```
print("Datatype of attribute: ", df.dtypes)
```

```
Datatype of attribute: ID           int64
Experience   float64
Age          int64
Gender       object
Salary       int64
dtype: object
```

```
print("First 5 rows: ", df.head())
```

```
First 5 rows:      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
```

```
print("Last 5 rows: ", df.tail())
```

```
Last 5 rows:      ID  Experience  Age  Gender  Salary
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
```

```
print("Total no of null values:", df.isna().sum())
```

```
Total no of null values: ID          0
Experience      1
Age            0
Gender         1
Salary        0
dtype: int64
```

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

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

```
df.isna().sum()
```

```
ID          0
Experience   0
Age          0
Gender       0
Salary      0
dtype: int64
```

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

```
Statistical information of Numerical Columns:
      ID  Experience      Age      Salary
count  35.000000  35.000000  35.000000  3.500000e+01
mean   18.000000   9.147059  35.485714  2.059147e+06
std    10.246951   7.546454  14.643552  3.170124e+06
min     1.000000   1.000000  17.000000  3.000000e+03
25%     9.500000   2.500000  22.500000  2.250000e+04
50%    18.000000   6.000000  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
```

## ✓ Group wise statistical summary

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

	count	mean	std	min	25%	50%	75%	max
Gender								
Female	18.0	10.111111	8.123234	1.0	3.25	7.5	18.0	27.0
Male	17.0	8.126298	6.982320	1.0	2.00	6.0	14.0	25.0

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

	count	mean	std	min	25%	50%	75%	max
Gender								
Female	18.0	37.111111	15.449686	21.0	23.0	31.0	54.0	62.0
Male	17.0	33.764706	13.997899	17.0	22.0	29.0	40.0	62.0

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

	count	mean	std	min	25%	50%	\
Gender							
Female	18.0	2.054917e+06	3.450120e+06	6000.0	30375.0	250000.0	
Male	17.0	2.063626e+06	2.950974e+06	3000.0	25000.0	220100.0	

  

	75%	max
Gender		
Female	1387500.0	10000000.0
Male	5001000.0	7600000.0

```
df = pd.read_csv("/content/iris.csv")
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Id               150 non-null   int64
1   SepalLength     148 non-null   float64
2   SepalWidth      150 non-null   float64
3   PetalLength     150 non-null   float64
4   PetalWidth      148 non-null   float64
5   Species         150 non-null   object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

```
df.head()
```

Next steps:

Id	SepalLength	SepalWidth	PetalLength	PetalWidth	Species	
	Generate code with df		View recommended plots			
0	1	5.1	3.5	1.4	0.2	Setosa

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

	SepalLength	SepalWidth	PetalLength	PetalWidth	Species	
0	5.1	3.5	1.4	0.2	Setosa	
1	4.9	3.0	1.4	0.2	Setosa	
2	4.7	3.2	1.3	0.2	Setosa	
3	4.6	3.1	1.5	0.2	Setosa	
4	5.0	3.6	1.4	0.2	Setosa	
...	...	...	...	...	...	
145	6.7	3.0	5.2	2.3	Virginica	
146	6.3	2.5	5.0	1.9	Virginica	
147	6.5	3.0	5.2	2.0	Virginica	
148	6.2	3.4	5.4	2.3	Virginica	
149	5.9	3.0	5.1	1.8	Virginica	

150 rows × 5 columns

Start coding or [generate](#) with AI.