

```
In [10]: import numpy as np
import pandas as pd
```

```
In [11]: df = pd.read_csv("C:/Users/SW20407278/Desktop/Final AI/Hands-On/Pandas/employee.csv")
```

```
In [13]: df.head()
```

```
Out[13]:
```

	Empid	Name	Salary
0	1	Smith	5000
1	2	Jones	6000
2	3	Harry	4000
3	4	Jem	5000
4	5	Sylvia	4000

```
In [14]: type(df)
```

```
Out[14]: pandas.core.frame.DataFrame
```

```
In [15]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 3 columns):
#   Column  Non-Null Count  Dtype
---  -
0   Empid   8 non-null        int64
1   Name    8 non-null        object
2   Salary  8 non-null        int64
dtypes: int64(2), object(1)
memory usage: 320.0+ bytes
```

```
In [16]: df.Salary
```

```
Out[16]:
```

0	5000
1	6000
2	4000
3	5000
4	4000
5	4000
6	5000
7	5000

Name: Salary, dtype: int64

```
In [17]: type(df.Salary)
```

```
Out[17]: pandas.core.series.Series
```

```
In [18]: # Mean
df.Salary.mean()
```

```
Out[18]: 4750.0
```

```
In [19]: # Median
df.Salary.median()
```

```
Out[19]: 5000.0
```

```
In [20]: # Mode
df.Salary.mode()
```

```
Out[20]: 0    5000
Name: Salary, dtype: int64
```

```
In [21]: # Var
df.Salary.var()
```

```
Out[21]: 500000.0
```

```
In [22]: # Std Deviation
df.Salary.std()
```

```
Out[22]: 707.1067811865476
```

```
In [23]: # How to check statistical summary
df.describe()  ## Shows statistical summary for numerical columns
```

```
Out[23]:
```

	Empid	Salary
count	8.000000	8.000000
mean	4.375000	4750.000000
std	2.263846	707.106781
min	1.000000	4000.000000
25%	2.750000	4000.000000
50%	4.500000	5000.000000
75%	6.250000	5000.000000
max	7.000000	6000.000000

```
In [24]: ### How to check statistical summary for all columns
df.describe(include = 'all')
```

Out[24]:

	Empid	Name	Salary
count	8.000000	8	8.000000
unique	NaN	7	NaN
top	NaN	William	NaN
freq	NaN	2	NaN
mean	4.375000	NaN	4750.000000
std	2.263846	NaN	707.106781
min	1.000000	NaN	4000.000000
25%	2.750000	NaN	4000.000000
50%	4.500000	NaN	5000.000000
75%	6.250000	NaN	5000.000000
max	7.000000	NaN	6000.000000

In [25]: *# Conversion of DataFrame into Numpy Array*
empCol = df.columns
emparray = df.values
emparray

Out[25]:

```
array([[1, 'Smith', 5000],
       [2, 'Jones', 6000],
       [3, 'Harry', 4000],
       [4, 'Jem', 5000],
       [5, 'Sylvia', 4000],
       [6, 'Charles', 4000],
       [7, 'William', 5000],
       [7, 'William', 5000]], dtype=object)
```

In [26]: *# Conversion of Numpy Array to DataFrame*

In [27]: employee_DF = pd.DataFrame(emparray, columns = empCol)

In [28]: employee_DF

Out[28]:

	Empid	Name	Salary
0	1	Smith	5000
1	2	Jones	6000
2	3	Harry	4000
3	4	Jem	5000
4	5	Sylvia	4000
5	6	Charles	4000
6	7	William	5000
7	7	William	5000