

In [1]: `import pandas as pd`

In [2]: `df=pd.read_csv('IRIS.csv')`
`print(df.head())`

| | sepal_length | sepal_width | petal_length | petal_width | species |
|---|--------------|-------------|--------------|-------------|-------------|
| 0 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 1 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| 2 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 3 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 4 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |

In [3]: `df.describe()`

Out[3]:

| | sepal_length | sepal_width | petal_length | petal_width |
|-------|--------------|-------------|--------------|-------------|
| count | 150.000000 | 147.000000 | 148.000000 | 150.000000 |
| mean | 5.843333 | 3.051701 | 3.785811 | 1.198667 |
| std | 0.828066 | 0.437044 | 1.760566 | 0.763161 |
| min | 4.300000 | 2.000000 | 1.000000 | 0.100000 |
| 25% | 5.100000 | 2.800000 | 1.575000 | 0.300000 |
| 50% | 5.800000 | 3.000000 | 4.400000 | 1.300000 |
| 75% | 6.400000 | 3.300000 | 5.100000 | 1.800000 |
| max | 7.900000 | 4.400000 | 6.900000 | 2.500000 |

In [4]: `print(df.describe(include=['object']))`

| | species |
|--------|-------------|
| count | 150 |
| unique | 3 |
| top | Iris-setosa |
| freq | 50 |

In [5]: `print(df.dtypes)`

```
sepal_length    float64
sepal_width     float64
petal_length    float64
petal_width     float64
species         object
dtype: object
```

In [6]: `print(df.info())`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   sepal_length    150 non-null   float64
1   sepal_width     147 non-null   float64
2   petal_length    148 non-null   float64
3   petal_width     150 non-null   float64
4   species         150 non-null   object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
None
```

In [7]: `print(df.isnull().sum())`

```
sepal_length    0
sepal_width     3
petal_length    2
petal_width     0
species         0
dtype: int64
```

In [8]: `sepal_width= df.select_dtypes(include=[np.number]).columns`
`df[sepal_width]=df[sepal_width].fillna(df[sepal_width].mean())`

```
-----
-----
NameError                                Traceback (most recent c
all last)
Cell In[8], line 1
----> 1 sepal_width= df.select_dtypes(include=[np.number]).columns
      2 df[sepal_width]=df[sepal_width].fillna(df[sepal_width].mea
n())

NameError: name 'np' is not defined
```

In [9]: `print(df.isnull().sum())`

```
<bound method NDFrame._add_numeric_operations.<locals>.sum of
sepal_length sepal_width petal_length petal_width species
0           False          False          False          False          False
1           False          False          False          False          False
2           False          False          False          False          False
3           False          False          False          False          False
4           False          False          False          False          False
..          ...          ...          ...          ...          ...
145          False          False          False          False          False
146          False          False          False          False          False
147          False          False          False          False          False
148          False          False          False          False          False
149          False          False          False          False          False
```

[150 rows x 5 columns]>

```
-----
-----
TypeError                                Traceback (most recent c
all last)
Cell In[9], line 1
----> 1 print(df.isnull().sum())

TypeError: 'NoneType' object is not callable
```

In [10]: `import numpy as np`

In [11]: `print(df.isnull().sum())`

```
<bound method NDFrame._add_numeric_operations.<locals>.sum of
sepal_length sepal_width petal_length petal_width species
0           False          False          False          False          False
1           False          False          False          False          False
2           False          False          False          False          False
3           False          False          False          False          False
4           False          False          False          False          False
..          ...          ...          ...          ...          ...
145          False          False          False          False          False
146          False          False          False          False          False
147          False          False          False          False          False
148          False          False          False          False          False
149          False          False          False          False          False
```

[150 rows x 5 columns]>

```
-----
-----
TypeError                                Traceback (most recent c
all last)
Cell In[11], line 1
----> 1 print(df.isnull().sum())

TypeError: 'NoneType' object is not callable
```

```
In [12]: print(df.isnull().sum())
```

```
sepal_length    0  
sepal_width     3  
petal_length    2  
petal_width     0  
species         0  
dtype: int64
```

```
In [13]: sepal_width= df.select_dtypes(include=[np.number]).columns  
df[sepal_width]=df[sepal_width].fillna(df[sepal_width].mean())
```

```
In [14]: print(df.isnull().sum())
```

```
sepal_length    0  
sepal_width     0  
petal_length    0  
petal_width     0  
species         0  
dtype: int64
```

```
In [16]: print(df.head(25))
```

| | sepal_length | sepal_width | petal_length | petal_width | spec |
|-----|--------------|-------------|--------------|-------------|----------|
| ies | | | | | |
| 0 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-set |
| osa | | | | | |
| 1 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-set |
| osa | | | | | |
| 2 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-set |
| osa | | | | | |
| 3 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-set |
| osa | | | | | |
| 4 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-set |
| osa | | | | | |
| 5 | 5.4 | 3.9 | 1.7 | 0.4 | Iris-set |
| osa | | | | | |
| 6 | 4.6 | 3.4 | 1.4 | 0.3 | Iris-set |
| osa | | | | | |
| 7 | 5.0 | 3.4 | 1.5 | 0.2 | Iris-set |
| osa | | | | | |
| 8 | 4.4 | 2.9 | 1.4 | 0.2 | Iris-set |
| osa | | | | | |
| 9 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-set |
| osa | | | | | |
| 10 | 5.4 | 3.7 | 1.5 | 0.2 | Iris-set |
| osa | | | | | |
| 11 | 4.8 | 3.4 | 1.6 | 0.2 | Iris-set |
| osa | | | | | |
| 12 | 4.8 | 3.0 | 1.4 | 0.1 | Iris-set |
| osa | | | | | |
| 13 | 4.3 | 3.0 | 1.1 | 0.1 | Iris-set |
| osa | | | | | |
| 14 | 5.8 | 4.0 | 1.2 | 0.2 | Iris-set |
| osa | | | | | |
| 15 | 5.7 | 4.4 | 1.5 | 0.4 | Iris-set |
| osa | | | | | |
| 16 | 5.4 | 3.9 | 1.3 | 0.4 | Iris-set |
| osa | | | | | |
| 17 | 5.1 | 3.5 | 1.4 | 0.3 | Iris-set |
| osa | | | | | |
| 18 | 5.7 | 3.8 | 1.7 | 0.3 | Iris-set |
| osa | | | | | |
| 19 | 5.1 | 3.8 | 1.5 | 0.3 | Iris-set |
| osa | | | | | |
| 20 | 5.4 | 3.4 | 1.7 | 0.2 | Iris-set |
| osa | | | | | |
| 21 | 5.1 | 3.7 | 1.5 | 0.4 | Iris-set |
| osa | | | | | |
| 22 | 4.6 | 3.6 | 1.0 | 0.2 | Iris-set |
| osa | | | | | |
| 23 | 5.1 | 3.3 | 1.7 | 0.5 | Iris-set |
| osa | | | | | |
| 24 | 4.8 | 3.4 | 1.9 | 0.2 | Iris-set |
| osa | | | | | |

In [17]: `print(df.info())`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   sepal_length    150 non-null   float64
1   sepal_width     150 non-null   float64
2   petal_length    150 non-null   float64
3   petal_width     150 non-null   float64
4   species         150 non-null   object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
None
```

In [18]: `print(df.shape)`

```
(150, 5)
```

In [19]: `df['species']=df['species'].astype('category')`

In [20]: `print(df['species'].dtype)`

```
category
```

In [21]: `print(df['species'].cat.categories)`

```
Index(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype
='object')
```

In [22]: `df['normalized_column'] = (df['sepal_length'] - df['sepal_length'].min()) / (df['sepal_length'].max() - df['sepal_length'].min())`

In [23]: `print(df)`

```

      sepal_length  sepal_width  petal_length  petal_width
species \
0          5.1          3.5          1.4          0.2      Iris
-setosa
1          4.9          3.0          1.4          0.2      Iris
-setosa
2          4.7          3.2          1.3          0.2      Iris
-setosa
3          4.6          3.1          1.5          0.2      Iris
-setosa
4          5.0          3.6          1.4          0.2      Iris
-setosa
..          ...          ...          ...          ...
145         6.7          3.0          5.2          2.3  Iris-vi
rginica
146         6.3          2.5          5.0          1.9  Iris-vi
rginica
147         6.5          3.0          5.2          2.0  Iris-vi
rginica
148         6.2          3.4          5.4          2.3  Iris-vi
rginica
149         5.9          3.0          5.1          1.8  Iris-vi
rginica

      normalized_column
0          0.222222
1          0.166667
2          0.111111
3          0.083333
4          0.194444
..          ...
145         0.666667
146         0.555556
147         0.611111
148         0.527778
149         0.444444

[150 rows x 6 columns]
```

In [24]: `from sklearn.preprocessing import LabelEncoder`

In [25]: `df = pd.DataFrame({
 'sepal_length': [5.1, 4.9, 5.0, 5.1],
 'sepal_width': [3.5, 3.0, 3.2, 3.6],
 'petal_length': [1.4, 1.4, 1.3, 1.5],
 'petal_width': [0.2, 0.2, 0.1, 0.2],
 'species': ['Iris-setosa', 'Iris-setosa', 'Iris-versicolor',
'Iris-virginica']
})`

In [26]: `label_encoder = LabelEncoder()`

In [27]: `df['species_encoded'] = label_encoder.fit_transform(df['species'])`

In [28]: `print(df)`

```

      sepal_length  sepal_width  petal_length  petal_width      s
species \
0          5.1          3.5          1.4          0.2  Iris-
setosa
1          4.9          3.0          1.4          0.2  Iris-
setosa
2          5.0          3.2          1.3          0.1  Iris-vers
icolor
3          5.1          3.6          1.5          0.2  Iris-vir
ginica

      species_encoded
0              0
1              0
2              1
3              2

```

In [29]: `from sklearn.preprocessing import MinMaxScaler`

In [30]: `scaler = MinMaxScaler()`

In [31]: `df[['sepal_length', 'petal_length']] = scaler.fit_transform(df[['sepal_length', 'petal_length']])`

In [32]: `print(df)`

```

      sepal_length  sepal_width  petal_length  petal_width      s
species \
0          1.0          3.5          0.5          0.2  Iris-
setosa
1          0.0          3.0          0.5          0.2  Iris-
setosa
2          0.5          3.2          0.0          0.1  Iris-vers
icolor
3          1.0          3.6          1.0          0.2  Iris-vir
ginica

      species_encoded
0              0
1              0
2              1
3              2

```

In []: