```
import pandas as pd
```

DataSet for Experiment No.01

Read Data Set

```
df= pd.read_csv('iris.csv')

pd.read_csv('iris.csv')
```

```
     sepal_length  sepal_width  petal_length  petal_width    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 x 5 columns]
```

```
df.describe()
```

```
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
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
```

Preprocessing

```
df.shape

(150, 5)

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
```

```
df.info
```

```
<bound method DataFrame.info of      sepal_length  sepal_width
petal_length  petal_width    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 x 5 columns]>
```

```
df.dtypes
```

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

```
df.ndim
```

```
2
```

```
df.head
```

```
<bound method NDFrame.head of      sepal_length  sepal_width
petal_length  petal_width    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 x 5 columns]>
```

```
df.head()
```

```
   sepal_length  sepal_width  petal_length  petal_width 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
```

```
df.tail
```

```
<bound method NDFrame.tail of      sepal_length  sepal_width
petal_length  petal_width    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 x 5 columns]>
```

```
df.tail()
```

```
     sepal_length  sepal_width  petal_length  petal_width    species
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
```

```
df.notnull()
```

```
     sepal_length  sepal_width  petal_length  petal_width  species
0            True         True          True         True     True
1            True         True          True         True     True
2            True         True          True         True     True
3            True         True          True         True     True
4            True         True          True         True     True
..            ...          ...           ...          ...      ...
145          True         True          True         True     True
```

```
146          True         True          True         True     True
147          True         True          True         True     True
148          True         True          True         True     True
149          True         True          True         True     True

[150 rows x 5 columns]
```

```
df.isnull()
```

```
     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]
```

```
df.isnull
```

```
<bound method DataFrame.isnull of      sepal_length  sepal_width
petal_length  petal_width    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 x 5 columns]>
```

```
df.sum()
```

```
sepal_length                                                876.5
sepal_width                                                 458.1
petal_length                                                563.8
petal_width                                                 179.8
species         setosasetosasetosasetosasetosasetosasetosaseto...
dtype: object
```

```
df.max()
```

```
sepal_length          7.9
sepal_width           4.4
petal_length          6.9
petal_width           2.5
species         virginica
dtype: object
```

```
df.min()
```

```
sepal_length       4.3
sepal_width        2.0
petal_length       1.0
petal_width        0.1
species         setosa
dtype: object
```

```
df.min()
```

```
sepal_length       4.3
sepal_width        2.0
petal_length       1.0
petal_width        0.1
species         setosa
dtype: object
```

```
df.count()
```

```
sepal_length    150
sepal_width     150
petal_length    150
petal_width     150
species         150
dtype: int64
```

```
df.sort_index()
```

	sepal_length	sepal_width	petal_length	petal_width	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 x 5 columns]
```

```
df.sort_values
```

```
<bound method DataFrame.sort_values of      sepal_length  sepal_width
petal_length  petal_width    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 x 5 columns]>
```

```
df['sepal_length'].sort_values()
```

```
13     4.3
42     4.4
38     4.4
8      4.4
41     4.5
      ...
122    7.7
118    7.7
117    7.7
135    7.7
131    7.9
Name: sepal_length, Length: 150, dtype: float64
```

```
df['sepal_length'].sort_index()
```

```
0      5.1
1      4.9
2      4.7
3      4.6
4      5.0
      ...
145    6.7
146    6.3
147    6.5
148    6.2
149    5.9
Name: sepal_length, Length: 150, dtype: float64
```

```
df['sepal_length'].sort_index()
```

```
0      5.1
1      4.9
2      4.7
3      4.6
4      5.0
      ...
145    6.7
146    6.3
147    6.5
148    6.2
149    5.9
Name: sepal_length, Length: 150, dtype: float64
```

```
df['sepal_length'].mean()
```

```
5.843333333333334
```

```
df['sepal_length'].describe()
```

```
count    150.000000
mean       5.843333
std        0.828066
min        4.300000
25%        5.100000
50%        5.800000
75%        6.400000
max        7.900000
Name: sepal_length, dtype: float64
```

```
df['sepal_length'].sum()
```

```
876.5
```

```
df['sepal_length'].mean()
```

```
5.843333333333334
```

```
df['sepal_length'].max()
```

```
7.9
```

```
df['sepal_length'].min()
```

```
4.3
```

```
df['sepal_length'].count()
```

```
150
```

```
df['species'].value_counts()
```

```
species
setosa        50
versicolor    50
virginica     50
Name: count, dtype: int64
```

```
df['sepal_length'].value_counts()
```

```
sepal_length
5.0    10
5.1     9
6.3     9
5.7     8
6.7     8
5.8     7
5.5     7
6.4     7
4.9     6
5.4     6
6.1     6
6.0     6
5.6     6
4.8     5
6.5     5
6.2     4
7.7     4
6.9     4
4.6     4
5.2     4
5.9     3
4.4     3
7.2     3
6.8     3
6.6     2
4.7     2
7.6     1
7.4     1
7.3     1
7.0     1
7.1     1
5.3     1
4.3     1
4.5     1
7.9     1
Name: count, dtype: int64
```

#Data Fromating and Normalization

```
df['species'].replace(['setosa','varsicolor','virginica'],[1,2,3])
```

```
0      1
1      1
2      1
3      1
4      1
      ..
145    3
146    3
147    3
148    3
149    3
Name: species, Length: 150, dtype: object
```

```
df.value_counts()
```

```
sepal_length  sepal_width  petal_length  petal_width  species
4.9           3.1          1.5           0.1          setosa        3
5.8           2.7          5.1           1.9          virginica     2
              4.0          1.2           0.2          setosa        1
5.9           3.0          4.2           1.5          versicolor    1
6.2           3.4          5.4           2.3          virginica     1
                                                                   ..
5.5           2.3          4.0           1.3          versicolor    1
              2.4          3.7           1.0          versicolor    1
                           3.8           1.1          versicolor    1
              2.5          4.0           1.3          versicolor    1
7.9           3.8          6.4           2.0          virginica     1
Name: count, Length: 147, dtype: int64
```

```
df['species'].value_counts()
```

```
species
setosa        50
versicolor    50
virginica     50
Name: count, dtype: int64
```

```
df
```

```
     sepal_length  sepal_width  petal_length  petal_width    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 x 5 columns]
```

```
df['species'].replace(['setosa','varsicolor','virginica'],[1,2,3])
```

```
0      1
1      1
2      1
3      1
4      1
      ..
145    3
146    3
147    3
148    3
149    3
Name: species, Length: 150, dtype: object
```

```
df['species'].replace(['setosa','versicolor','virginica'],
[1,2,3],inplace=True)
```

```
df
```

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

```
[150 rows x 5 columns]
```

```
df['species'].value_counts()
```

```
species
1    50
2    50
3    50
Name: count, dtype: int64
```

```
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    int64
dtypes: float64(4), int64(1)
memory usage: 6.0 KB
```

```
df.info
```

```
<bound method DataFrame.info of      sepal_length  sepal_width
petal_length  petal_width  species
0             5.1          3.5           1.4          0.2        1
1             4.9          3.0           1.4          0.2        1
2             4.7          3.2           1.3          0.2        1
3             4.6          3.1           1.5          0.2        1
4             5.0          3.6           1.4          0.2        1
..            ...          ...           ...          ...      ...
145           6.7          3.0           5.2          2.3        3
146           6.3          2.5           5.0          1.9        3
147           6.5          3.0           5.2          2.0        3
148           6.2          3.4           5.4          2.3        3
149           5.9          3.0           5.1          1.8        3

[150 rows x 5 columns]>
```

```
df['species'].astype
```

```
<bound method NDFrame.astype of 0      1
1      1
2      1
3      1
4      1
      ..
145    3
146    3
147    3
148    3
149    3
Name: species, Length: 150, dtype: int64>
```

```
df.astype(float)
```

```
     sepal_length  sepal_width  petal_length  petal_width  species
0             5.1          3.5           1.4          0.2      1.0
1             4.9          3.0           1.4          0.2      1.0
2             4.7          3.2           1.3          0.2      1.0
3             4.6          3.1           1.5          0.2      1.0
4             5.0          3.6           1.4          0.2      1.0
```

```
..            ...          ...           ...          ...
145           6.7          3.0           5.2          2.3      3.0
146           6.3          2.5           5.0          1.9      3.0
147           6.5          3.0           5.2          2.0      3.0
148           6.2          3.4           5.4          2.3      3.0
149           5.9          3.0           5.1          1.8      3.0

[150 rows x 5 columns]
```

```
df
```

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

[150 rows x 5 columns]
```

```
df=pd.read_csv('iris.csv')
```

```
df.describe()
```

```
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
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
```

#Categarical-> Normalization

from sklearn.preprocessing

```
import numpy as np
from sklearn import preprocessing

import LabelEncoder as le

-----------------------------------------------------------------------
-----
```

```
ModuleNotFoundError                       Traceback (most recent call last)
Cell In[212], line 1
----> 1 import LabelEncoder as le

ModuleNotFoundError: No module named 'LabelEncoder'
```