```
Lab Assignment 1

AIM:-Perform the following operations using Python on any open source
dataset (e.g., data.csv)
1. Import all the required Python Libraries.
2. Locate an open source data from the web (e.g.,
https://www.kaggle.com). Provide a clear
 description of the data and its source (i.e., URL of the web site).
3. Load the Dataset into pandas dataframe.
4. Data Preprocessing: check for missing values in the data using
pandas isnull(), describe()
function to get some initial statistics. Provide variable
descriptions. Types of variables etc.
Check the dimensions of the data frame.
5. Data Formatting and Data Normalization: Summarize the types of
variables by checking
the data types (i.e., character, numeric, integer, factor, and
logical) of the variables in the
data set. If variables are not in the correct data type, apply proper
type conversions.
6. Turn categorical variables into quantitative variables in Python.
In addition to the codes and outputs, explain every operation that you
do in the above steps and
explain everything that you do to import/read/scrape the data set.

csv_url =
'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.d
ata'

import pandas as pd

iris = pd.read_csv(csv_url, header = None)

col_names
=['Sepal_Length','Sepal_Width','Petal_Length','Petal_Width','Species']

iris = pd.read_csv(csv_url, names = col_names)

df1=df=iris

iris.head(8)
```

	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
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa

```
iris.tail()
```

```
     Sepal_Length  Sepal_Width  Petal_Length  Petal_Width
Species
145           6.7          3.0           5.2          2.3  Iris-
virginica
146           6.3          2.5           5.0          1.9  Iris-
virginica
147           6.5          3.0           5.2          2.0  Iris-
virginica
148           6.2          3.4           5.4          2.3  Iris-
virginica
149           5.9          3.0           5.1          1.8  Iris-
virginica
```

```
iris.index
```

```
RangeIndex(start=0, stop=150, step=1)
```

```
iris.columns
```

```
Index(['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width',
       'Species'],
      dtype='object')
```

```
iris.shape
```

```
(150, 5)
```

```
iris.dtypes
```

```
 Sepal_Leng       floa
 th                t64
 Sepal_Widt       floa
 h                 t64
 Petal_Leng       floa
dtype: object
```

```
iris.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
```

```
iris.columns.values
```

```
array(['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width',
       'Species'], dtype=object)
```

```
iris.iloc[5]
```

```
Sepal_Length            5.4
Sepal_Width             3.9
Petal_Length            1.7
Petal_Width             0.4
Species         Iris-setosa
Name: 5, dtype: object
```

```
iris[47:51]
```

```
    Sepal_Length  Sepal_Width  Petal_Length  Petal_Width
Species
47             4.6          3.2           1.4          0.2       Iris-
setosa
48             5.3          3.7           1.5          0.2       Iris-
setosa
49             5.0          3.3           1.4          0.2       Iris-
setosa
50             7.0          3.2           4.7          1.4  Iris-
versicolor
```

```
iris.loc[:,["Sepal_Length","Sepal_Width"]]
```

```
     Sepal_Length  Sepal_Width
0             5.1          3.5
1             4.9          3.0
2             4.7          3.2
3             4.6          3.1
4             5.0          3.6
..            ...          ...
145           6.7          3.0
146           6.3          2.5
147           6.5          3.0
148           6.2          3.4
149           5.9          3.0

[150 rows x 2 columns]
```

```
cols_2_4=iris.columns[2:4]
iris[cols_2_4]
```

```
     Petal_Length  Petal_Width
0             1.4          0.2
1             1.4          0.2
2             1.3          0.2
3             1.5          0.2
```

```
4                5.4                0.2
..               ...                ...
145              5.2                2.3
146              5.0                1.9
147              5.2                2.0
148              5.4                2.3
149              5.1                1.8

[150 rows x 2 columns]
```

```
iris.isnull().any()
```

```
 Sepal_Leng       Fa
 th               ls
 Sepal_Widt       e
 h                Fa
 Petal Leng       ls
dtype: bool
```

```
iris.isnull().sum()
```

```
Sepal_Length    0
Sepal_Width     0
Petal_Length    0
Petal_Width     0
Species         0
dtype: int64
```

```
df=iris
df['petal Length(cm)']=iris['Petal_Length'].astype("int")

df1=df
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-
virginica
146           6.3          2.5           5.0          1.9  Iris-
```

```
virginica
147            6.5           3.0            5.2           2.0  Iris-
virginica
148            6.2           3.4            5.4           2.3  Iris-
virginica
149            5.9           3.0            5.1           1.8  Iris-
virginica
```

```
     petal Length(cm)
0                   1
1                   1
2                   1
3                   1
4                   1
..                ...
145                 5
146                 5
147                 5
148                 5
149                 5

[150 rows x 6 columns]
```

```
from sklearn import preprocessing
min_max_scaler = preprocessing.MinMaxScaler()

X=iris.iloc[:,:4]
X
```

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

[150 rows x 4 columns]
```

```
X_scaled = min_max_scaler.fit_transform(X)

df_normalized = pd.DataFrame(X_scaled)
df_normalized
```

```
              0         1         2         3
0      0.222222  0.625000  0.067797  0.041667
```

```
1     0.166667  0.416667  0.067797  0.041667
2     0.111111  0.500000  0.050847  0.041667
3     0.083333  0.458333  0.084746  0.041667
4     0.194444  0.666667  0.067797  0.041667
..         ...       ...       ...       ...
145   0.666667  0.416667  0.711864  0.916667
146   0.555556  0.208333  0.677966  0.750000
147   0.611111  0.416667  0.711864  0.791667
148   0.527778  0.583333  0.745763  0.916667
149   0.444444  0.416667  0.694915  0.708333

[150 rows x 4 columns]
```

```
df2=df
df2['Species'].unique()
```

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

```
df_normalized = pd.DataFrame(X_scaled)
df_normalized
```

```
             0         1         2         3
0     0.222222  0.625000  0.067797  0.041667
1     0.166667  0.416667  0.067797  0.041667
2     0.111111  0.500000  0.050847  0.041667
3     0.083333  0.458333  0.084746  0.041667
4     0.194444  0.666667  0.067797  0.041667
..         ...       ...       ...       ...
145   0.666667  0.416667  0.711864  0.916667
146   0.555556  0.208333  0.677966  0.750000
147   0.611111  0.416667  0.711864  0.791667
148   0.527778  0.583333  0.745763  0.916667
149   0.444444  0.416667  0.694915  0.708333

[150 rows x 4 columns]
```

```
df2=df
df2['Species'].unique()

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

from sklearn import preprocessing
enc = preprocessing.OneHotEncoder()

features_df=df2.drop(columns=['Species'])
features_df

     Sepal_Length  Sepal_Width  Petal_Length  Petal_Width  petal
Length(cm)
```

```
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
5
146           6.3           2.5            5.0            1.9
5
147           6.5           3.0            5.2            2.0
5
148           6.2           3.4            5.4            2.3
5
149           5.9           3.0            5.1            1.8
5

[150 rows x 5 columns]
```

```
enc_df=(enc.fit_transform(df2[['Species']])).toarray()

enc_df = pd.DataFrame(enc_df, columns = ['Iris-Setosa','Iris-
Versicolor','Iris-Virginica'])

df_encode = features_df.join(enc_df)
df_encode
```

```
     Sepal_Length   Sepal_Width   Petal_Length   Petal_Width   petal
Length(cm) \
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
..            ...           ... 3.         ... 5.         ... 2.
...
145           6.7             0              2              3
5
146           6.3           2.5            5.0            1.9
```

```
5
147             6.5             3.0             5.2             2.0
5
148             6.2             3.4             5.4             2.3
5
149             5.9             3.0             5.1             1.8
5

     Iris-Setosa  Iris-Versicolor  Iris-Virginica
0            1.0              0.0             0.0
1            1.0              0.0             0.0
2            1.0              0.0             0.0
3            1.0              0.0             0.0
4            1.0              0.0             0.0
..           ...              ...             ...
145          0.0              0.0             1.0
146          0.0              0.0             1.0
147          0.0              0.0             1.0
148          0.0              0.0             1.0
149          0.0              0.0             1.0

[150 rows x 8 columns]

Name - Sushant Jawale

Roll no. - 13209
```