Assignment No:01

1. Import all the required Python Libraries.

```
import pandas as pd
```

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).

```
!pip install -q kaggle
```

```
from google.colab import files
```

```
files.upload()
```

Choose Files Iris.csv

- **Iris.csv**(text/csv) - 5107 bytes, last modified: 3/31/2024 - 100% done

```
Saving Iris.csv to Iris.csv
{'Iris.csv':
b'Id,SepalLengthCm,SepalWidthCm,PetalLengthCm,PetalWidthCm,Species\n1,5.1,3.5,1.4,0.2,Iris-
setosa\n2,4.9,3.0,1.4,0.2,Iris-setosa\n3,4.7,3.2,1.3,0.2,Iris-setosa\n4,4.6,3.1,1.5,0.2,Iris-
setosa\n5,5.0,3.6,1.4,0.2,Iris-setosa\n6,5.4,3.9,1.7,0.4,Iris-setosa\n7,4.6,3.4,1.4,0.3,Iris-
setosa\n8,5.0,3.4,1.5,0.2,Iris-setosa\n9,4.4,2.9,1.4,0.2,Iris-setosa\n10,4.9,3.1,1.5,0.1,Iris-
setosa\n11,5.4,3.7,1.5,0.2,Iris-setosa\n12,4.8,3.4,1.6,0.2,Iris-
setosa\n13,4.8,3.0,1.4,0.1,Iris-setosa\n14,4.3,3.0,1.1,0.1,Iris-
setosa\n15,5.8,4.0,1.2,0.2,Iris-setosa\n16,5.7,4.4,1.5,0.4,Iris-
setosa\n17,5.4,3.9,1.3,0.4,Iris-setosa\n18,5.1,3.5,1.4,0.3,Iris-
setosa\n19,5.7,3.8,1.7,0.3,Iris-setosa\n20,5.1,3.8,1.5,0.3,Iris-
setosa\n21,5.4,3.4,1.7,0.2,Iris-setosa\n22,5.1,3.7,1.5,0.4,Iris-
setosa\n23,4.6,3.6,1.0,0.2,Iris-setosa\n24,5.1,3.3,1.7,0.5,Iris-
setosa\n25,4.8,3.4,1.9,0.2,Iris-setosa\n26,5.0,3.0,1.6,0.2,Iris-
setosa\n27,5.0,3.4,1.6,0.4,Iris-setosa\n28,5.2,3.5,1.5,0.2,Iris-
setosa\n29,5.2,3.4,1.4,0.2,Iris-setosa\n30,4.7,3.2,1.6,0.2,Iris-
setosa\n31,4.8,3.1,1.6,0.2,Iris-setosa\n32,5.4,3.4,1.5,0.4,Iris-
setosa\n33,5.2,4.1,1.5,0.1,Iris-setosa\n34,5.5,4.2,1.4,0.2,Iris-
setosa\n35,4.9,3.1,1.5,0.1,Iris-setosa\n36,5.0,3.2,1.2,0.2,Iris-
setosa\n37,5.5,3.5,1.3,0.2,Iris-setosa\n38,4.9,3.1,1.5,0.1,Iris-
setosa\n39,4.4,3.0,1.3,0.2,Iris-setosa\n40,5.1,3.4,1.5,0.2,Iris-
setosa\n41,5.0,3.5,1.3,0.3,Iris-setosa\n42,4.5,2.3,1.3,0.3,Iris-
setosa\n43,4.4,3.2,1.3,0.2,Iris-setosa\n44,5.0,3.5,1.6,0.6,Iris-
setosa\n45,5.1,3.8,1.9,0.4,Iris-setosa\n46,4.8,3.0,1.4,0.3,Iris-
setosa\n47,5.1,3.8,1.6,0.2,Iris-setosa\n48,4.6,3.2,1.4,0.2,Iris-
setosa\n49,5.3,3.7,1.5,0.2,Iris-setosa\n50,5.0,3.3,1.4,0.2,Iris-
setosa\n51,7.0,3.2,4.7,1.4,Iris-versicolor\n52,6.4,3.2,4.5,1.5,Iris-
versicolor\n53,6.9,3.1,4.9,1.5,Iris-versicolor\n54,5.5,2.3,4.0,1.3,Iris-
versicolor\n55,6.5,2.8,4.6,1.5,Iris-versicolor\n56,5.7,2.8,4.5,1.3,Iris-
versicolor\n57,6.3,3.3,4.7,1.6,Iris-versicolor\n58,4.9,2.4,3.3,1.0,Iris-
versicolor\n59,6.6,2.9,4.6,1.3,Iris-versicolor\n60,5.2,2.7,3.9,1.4,Iris-
versicolor\n61,5.0,2.0,3.5,1.0,Iris-versicolor\n62,5.9,3.0,4.2,1.5,Iris-
```

3.Load the Dataset into pandas dataframe.

```
iris=pd.read_csv("Iris.csv")
```

```
iris
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Next steps: Generate code with iris View recommended plots

4. Data Preprocessing:

```
iris.head()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

Next steps: Generate code with iris View recommended plots

```
iris.tail()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

```
iris.describe(include="all")
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
count	150.000000	150.000000	150.000000	150.000000	150.000000	150
unique	NaN	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	NaN	Iris-setosa
freq	NaN	NaN	NaN	NaN	NaN	50
mean	75.500000	5.843333	3.054000	3.758667	1.198667	NaN
std	43.445368	0.828066	0.433594	1.764420	0.763161	NaN
min	1.000000	4.300000	2.000000	1.000000	0.100000	NaN
25%	38.250000	5.100000	2.800000	1.600000	0.300000	NaN
50%	75.500000	5.800000	3.000000	4.350000	1.300000	NaN
75%	112.750000	6.400000	3.300000	5.100000	1.800000	NaN
max	150.000000	7.900000	4.400000	6.900000	2.500000	NaN

```
iris.shape
```

```
(150, 6)
```

```
iris[0:3]
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa

```
iris.loc[0:2]
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa

```
iris.loc[0:4,'Id':'PetalWidthCm']
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
0	1	5.1	3.5	1.4	0.2
1	2	4.9	3.0	1.4	0.2
2	3	4.7	3.2	1.3	0.2
3	4	4.6	3.1	1.5	0.2
4	5	5.0	3.6	1.4	0.2

```
iris.iloc[1:3]
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa

```
iris.iloc[2:6,2:6]
```

	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
2	3.2	1.3	0.2	Iris-setosa
3	3.1	1.5	0.2	Iris-setosa
4	3.6	1.4	0.2	Iris-setosa
5	3.9	1.7	0.4	Iris-setosa

```
iris.isnull()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
145	False	False	False	False	False	False
146	False	False	False	False	False	False
147	False	False	False	False	False	False
148	False	False	False	False	False	False
149	False	False	False	False	False	False

150 rows × 6 columns

```
iris.isna()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
145	False	False	False	False	False	False
146	False	False	False	False	False	False
147	False	False	False	False	False	False
148	False	False	False	False	False	False
149	False	False	False	False	False	False

150 rows × 6 columns

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

```
Id               False
SepalLengthCm    False
SepalWidthCm     False
PetalLengthCm    False
PetalWidthCm     False
Species          False
dtype: bool
```

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

```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

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

```
0
```

5. Data Formatting and Data Normalization:

Data Formatting:

```
iris.dtypes
```

```
Id                 int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

```
iris.SepalLengthCm=iris.SepalLengthCm.astype("int")
```

```
iris.dtypes
```

```
Id                 int64
SepalLengthCm      int64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

Data Normalization:

```
from sklearn import preprocessing
```

```
iris.head()
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5	3.5	1.4	0.2	Iris-setosa
1	2	4	3.0	1.4	0.2	Iris-setosa
2	3	4	3.2	1.3	0.2	Iris-setosa
3	4	4	3.1	1.5	0.2	Iris-setosa
4	5	5	3.6	1.4	0.2	Iris-setosa

Next steps: Generate code with iris View recommended plots

```
min_max_scaler = preprocessing.MinMaxScaler()
```

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

```
x
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm
0	1	5	3.5	1.4
1	2	4	3.0	1.4
2	3	4	3.2	1.3
3	4	4	3.1	1.5
4	5	5	3.6	1.4
...	...	...	...	...
145	146	6	3.0	5.2
146	147	6	2.5	5.0
147	148	6	3.0	5.2
148	149	6	3.4	5.4
149	150	5	3.0	5.1

150 rows × 4 columns

Next steps: Generate code with x View recommended plots

```
x_scaled = min_max_scaler.fit_transform(x)
```

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

```
df_normalized
```

	0	1	2	3
0	0.000000	0.333333	0.625000	0.067797
1	0.006711	0.000000	0.416667	0.067797
2	0.013423	0.000000	0.500000	0.050847
3	0.020134	0.000000	0.458333	0.084746
4	0.026846	0.333333	0.666667	0.067797
...	...	...	...	...
145	0.973154	0.666667	0.416667	0.711864
146	0.979866	0.666667	0.208333	0.677966
147	0.986577	0.666667	0.416667	0.711864
148	0.993289	0.666667	0.583333	0.745763
149	1.000000	0.333333	0.416667	0.694915

150 rows × 4 columns

Next steps: Generate code with df_normalized View recommended plots

6. Turn categorical variables into quantitative variables in Python.

```
iris
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5	3.5	1.4	0.2	Iris-setosa
1	2	4	3.0	1.4	0.2	Iris-setosa
2	3	4	3.2	1.3	0.2	Iris-setosa
3	4	4	3.1	1.5	0.2	Iris-setosa
4	5	5	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6	3.0	5.2	2.3	Iris-virginica
146	147	6	2.5	5.0	1.9	Iris-virginica
147	148	6	3.0	5.2	2.0	Iris-virginica
148	149	6	3.4	5.4	2.3	Iris-virginica
149	150	5	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Next steps: Generate code with iris View recommended plots

i. Replace method

```
iris['Species'].unique()
```

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

```
iris['Species'].replace(['Iris-setosa','Iris-versicolor', 'Iris-virginica'],[0,1,2], inplace=True)
```

```
iris['Species'].unique()
```

```
array([0, 1, 2])
```

```
iris
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5	3.5	1.4	0.2	0
1	2	4	3.0	1.4	0.2	0
2	3	4	3.2	1.3	0.2	0
3	4	4	3.1	1.5	0.2	0
4	5	5	3.6	1.4	0.2	0
...	...	...	...	...	...	...
145	146	6	3.0	5.2	2.3	2
146	147	6	2.5	5.0	1.9	2
147	148	6	3.0	5.2	2.0	2
148	149	6	3.4	5.4	2.3	2
149	150	5	3.0	5.1	1.8	2

150 rows × 6 columns

Next steps: Generate code with iris View recommended plots

```
iris.dtypes
```

```
Id                 int64
SepalLengthCm      int64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species            int64
dtype: object
```

```
iris['Species'].replace([0,1,2],['Iris-setosa','Iris-versicolor', 'Iris-virginica'], inplace=True)
```

```
iris['Species'].unique()
```

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

```
iris
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5	3.5	1.4	0.2	Iris-setosa
1	2	4	3.0	1.4	0.2	Iris-setosa
2	3	4	3.2	1.3	0.2	Iris-setosa
3	4	4	3.1	1.5	0.2	Iris-setosa
4	5	5	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6	3.0	5.2	2.3	Iris-virginica
146	147	6	2.5	5.0	1.9	Iris-virginica
147	148	6	3.0	5.2	2.0	Iris-virginica
148	149	6	3.4	5.4	2.3	Iris-virginica
149	150	5	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Next steps: Generate code with iris View recommended plots

```
iris.dtypes
```

```
Id                 int64
SepalLengthCm      int64
SepalWidthCm     float64
```

```
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

ii. Label Encoding

```
iris['Species'].unique()
```

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

```
label_encoder = preprocessing.LabelEncoder()
```

```
iris['Species']= label_encoder.fit_transform(iris['Species'])
```

```
iris['Species'].unique()
```

```
array([0, 1, 2])
```

```
iris
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5	3.5	1.4	0.2	0
1	2	4	3.0	1.4	0.2	0
2	3	4	3.2	1.3	0.2	0
3	4	4	3.1	1.5	0.2	0
4	5	5	3.6	1.4	0.2	0
...	...	...	...	...	...	...
145	146	6	3.0	5.2	2.3	2
146	147	6	2.5	5.0	1.9	2
147	148	6	3.0	5.2	2.0	2
148	149	6	3.4	5.4	2.3	2
149	150	5	3.0	5.1	1.8	2

150 rows × 6 columns