In [1]:
```python
import pandas as pd
```

In [3]:
```python
import numpy as nm
```

In [ ]:
```python
#Locate open source data from the web
```

In [4]:
```python
csv_url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris
```

In [ ]:
```python
#Provide a clear description of the data and its source
```

In [5]:
```python
iris = pd.read_csv(csv_url, header = None)
```

In [6]:
```python
col_names = ['Sepal_Length','Sepal_Width','Petal_Length','Petal_Width','S
```

In [7]:
```python
iris = pd.read_csv(csv_url, names = col_names)
```

In [ ]:
```python
#Load the Dataset into the pandas data frame.
```

In [8]:
```python
dataset= pd.DataFrame(iris)
```

In [9]:
```python
dataset
```

Out[9]:

	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

150 rows × 5 columns

In [ ]:
```python
#Data Preprocessing
```

In [10]:
```python
dataset.head(n=5)
```

Out[10]:

	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 [11]:
```python
dataset.tail(n=5)
```

Out[11]:

	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

In [12]:
```python
dataset.index
```

Out[12]: RangeIndex(start=0, stop=150, step=1)

In [13]:
```python
dataset.columns
```

Out[13]: Index(['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width',
       'Species'],
      dtype='object')

In [14]:
```python
dataset.shape
```

Out[14]: (150, 5)

In [15]:
```python
dataset.dtypes
```

Out[15]: Sepal_Length   float64
        Sepal_Width   float64
        Petal_Length   float64
        Petal_Width   float64
        Species   object
        dtype: object

In [16]:
```python
dataset.columns.values
```

Out[16]: array(['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width',
           'Species'], dtype=object)

In [17]:
```python
dataset.describe(include='all')
```

Out[17]:

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
count	150.000000	150.000000	150.000000	150.000000	150
unique	NaN	NaN	NaN	NaN	3
top	NaN	NaN	NaN	NaN	Iris-setosa
freq	NaN	NaN	NaN	NaN	50
mean	5.843333	3.054000	3.758667	1.198667	NaN
std	0.828066	0.433594	1.764420	0.763161	NaN
min	4.300000	2.000000	1.000000	0.100000	NaN
25%	5.100000	2.800000	1.600000	0.300000	NaN
50%	5.800000	3.000000	4.350000	1.300000	NaN
75%	6.400000	3.300000	5.100000	1.800000	NaN
max	7.900000	4.400000	6.900000	2.500000	NaN

In [19]:
```python
dataset['Sepal_Length']
```

Out[19]:
```
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
```

In [20]:
```python
dataset.sort_index(axis=1,
ascending=False)
```

Out[20]:

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

150 rows × 5 columns

In [22]: `dataset.sort_values(by="Sepal_Length")`

Out[22]:

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
13	4.3	3.0	1.1	0.1	Iris-setosa
42	4.4	3.2	1.3	0.2	Iris-setosa
38	4.4	3.0	1.3	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
41	4.5	2.3	1.3	0.3	Iris-setosa
...	...	...	...	...	...
122	7.7	2.8	6.7	2.0	Iris-virginica
118	7.7	2.6	6.9	2.3	Iris-virginica
117	7.7	3.8	6.7	2.2	Iris-virginica
135	7.7	3.0	6.1	2.3	Iris-virginica
131	7.9	3.8	6.4	2.0	Iris-virginica

150 rows × 5 columns

In [23]: `dataset.iloc[5]`

Out[23]:
```
Sepal_Length              5.4
Sepal_Width               3.9
Petal_Length              1.7
Petal_Width               0.4
Species           Iris-setosa
Name: 5, dtype: object
```

In [24]: `dataset[0:3]`

Out[24]:

	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

In [25]: `dataset.loc[:, ["Sepal_Length", "Sepal_Width"]]`

Out[25]:

	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 × 2 columns

In [27]:
```python
dataset.iloc[:5, :]
```

Out[27]:

	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 [29]:
```python
dataset.isnull()
```

Out[29]:

	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 × 5 columns

In [ ]:
```python
# Data Formatting
```

In [30]:
```python
dataset.dtypes
```

Out[30]:
```
Sepal_Length     float64
Sepal_Width      float64
Petal_Length     float64
Petal_Width      float64
Species           object
dtype: object
```

In [33]:
```python
dataset['Petal_Length']= dataset['Petal_Length'].astype("int")
```

In [34]:
```python
dataset.dtypes
```

Out[34]:
```
Sepal_Length     float64
Sepal_Width      float64
Petal_Length       int64
Petal_Width      float64
Species           object
dtype: object
```

In [ ]:
```python
#Data Normalization:
```

In [36]:
```python
from sklearn import preprocessing
```

In [38]:
```python
dataset.head()
```

Out[38]:

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
0	5.1	3.5	1	0.2	Iris-setosa
1	4.9	3.0	1	0.2	Iris-setosa
2	4.7	3.2	1	0.2	Iris-setosa
3	4.6	3.1	1	0.2	Iris-setosa
4	5.0	3.6	1	0.2	Iris-setosa

In [39]:
```python
min_max_scaler = preprocessing.MinMaxScaler()
```

In [41]:
```python
x=dataset.iloc[:,:4]
```

In [42]:
```python
x_scaled = min_max_scaler.fit_transform(x)
```

In [44]:
```python
df_normalized = pd.DataFrame(x_scaled)
```

In [45]:
```python
df_normalized
```

Out[45]:

	0	1	2	3
0	0.222222	0.625000	0.0	0.041667
1	0.166667	0.416667	0.0	0.041667
2	0.111111	0.500000	0.0	0.041667
3	0.083333	0.458333	0.0	0.041667
4	0.194444	0.666667	0.0	0.041667
...	...	...	...	...
145	0.666667	0.416667	0.8	0.916667
146	0.555556	0.208333	0.8	0.750000
147	0.611111	0.416667	0.8	0.791667
148	0.527778	0.583333	0.8	0.916667
149	0.444444	0.416667	0.8	0.708333

150 rows × 4 columns

In [ ]:
```python
#Turn categorical variables into quantitative variables
```

In [48]:
```python
dataset['Species'].unique()
```

Out[48]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)

In [49]:
```python
label_encoder = preprocessing.LabelEncoder()
```

In [51]:
```python
dataset['Species']= label_encoder.fit_transform(dataset['Species'])
```

In [52]:
```python
dataset['Species'].unique()
```

Out[52]: array([0, 1, 2])

In [53]:
```python
dataset
```

Out[53]:

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
0	5.1	3.5	1	0.2	0
1	4.9	3.0	1	0.2	0
2	4.7	3.2	1	0.2	0
3	4.6	3.1	1	0.2	0
4	5.0	3.6	1	0.2	0
...	...	...	...	...	...
145	6.7	3.0	5	2.3	2
146	6.3	2.5	5	1.9	2
147	6.5	3.0	5	2.0	2
148	6.2	3.4	5	2.3	2
149	5.9	3.0	5	1.8	2

150 rows × 5 columns

In [ ]: