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

In [5]:
```
df = pd.read_csv(r"C:\Users\Roshan Ramdas Kate\Downloads\iris.csv")
```

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

```
     sepal.length  sepal.width  petal.length  petal.width     variety
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]
```

In [7]:
```
df.head()
```

Out[7]:

	sepal.length	sepal.width	petal.length	petal.width	variety
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

In [8]:
```
df.head(10)
```

Out[8]:

	sepal.length	sepal.width	petal.length	petal.width	variety
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
5	5.4	3.9	1.7	0.4	Setosa
6	4.6	3.4	1.4	0.3	Setosa
7	5.0	3.4	1.5	0.2	Setosa
8	4.4	2.9	1.4	0.2	Setosa
9	4.9	3.1	1.5	0.1	Setosa

In [9]:
```
df.tail()
```

Out[9]:

	sepal.length	sepal.width	petal.length	petal.width	variety
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

In [10]:
```
df.sample(10)
```

Out[10]:

	sepal.length	sepal.width	petal.length	petal.width	variety
125	7.2	3.2	6.0	1.8	Virginica
93	5.0	2.3	3.3	1.0	Versicolor
81	5.5	2.4	3.7	1.0	Versicolor
112	6.8	3.0	5.5	2.1	Virginica
11	4.8	3.4	1.6	0.2	Setosa
75	6.6	3.0	4.4	1.4	Versicolor
120	6.9	3.2	5.7	2.3	Virginica
7	5.0	3.4	1.5	0.2	Setosa
101	5.8	2.7	5.1	1.9	Virginica
94	5.6	2.7	4.2	1.3	Versicolor

In [11]:
```
df.iloc[5]
```

Out[11]:
```
sepal.length       5.4
sepal.width        3.9
petal.length       1.7
petal.width        0.4
variety         Setosa
Name: 5, dtype: object
```

In [20]:
```
result=df.loc[df["variety"]=="Setosa"]
```

In [21]:
```
print(result)
```

	sepal.length	sepal.width	petal.length	petal.width	variety
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
5	5.4	3.9	1.7	0.4	Setosa
6	4.6	3.4	1.4	0.3	Setosa
7	5.0	3.4	1.5	0.2	Setosa
8	4.4	2.9	1.4	0.2	Setosa
9	4.9	3.1	1.5	0.1	Setosa
10	5.4	3.7	1.5	0.2	Setosa
11	4.8	3.4	1.6	0.2	Setosa
12	4.8	3.0	1.4	0.1	Setosa
13	4.3	3.0	1.1	0.1	Setosa
14	5.8	4.0	1.2	0.2	Setosa
15	5.7	4.4	1.5	0.4	Setosa
16	5.4	3.9	1.3	0.4	Setosa
17	5.1	3.5	1.4	0.3	Setosa
18	5.7	3.8	1.7	0.3	Setosa
19	5.1	3.8	1.5	0.3	Setosa
20	5.4	3.4	1.7	0.2	Setosa
21	5.1	3.7	1.5	0.4	Setosa
22	4.6	3.6	1.0	0.2	Setosa
23	5.1	3.3	1.7	0.5	Setosa
24	4.8	3.4	1.9	0.2	Setosa
25	5.0	3.0	1.6	0.2	Setosa
26	5.0	3.4	1.6	0.4	Setosa
27	5.2	3.5	1.5	0.2	Setosa
28	5.2	3.4	1.4	0.2	Setosa
29	4.7	3.2	1.6	0.2	Setosa
30	4.8	3.1	1.6	0.2	Setosa
31	5.4	3.4	1.5	0.4	Setosa
32	5.2	4.1	1.5	0.1	Setosa
33	5.5	4.2	1.4	0.2	Setosa
34	4.9	3.1	1.5	0.2	Setosa
35	5.0	3.2	1.2	0.2	Setosa
36	5.5	3.5	1.3	0.2	Setosa
37	4.9	3.6	1.4	0.1	Setosa
38	4.4	3.0	1.3	0.2	Setosa
39	5.1	3.4	1.5	0.2	Setosa
40	5.0	3.5	1.3	0.3	Setosa
41	4.5	2.3	1.3	0.3	Setosa
42	4.4	3.2	1.3	0.2	Setosa
43	5.0	3.5	1.6	0.6	Setosa
44	5.1	3.8	1.9	0.4	Setosa
45	4.8	3.0	1.4	0.3	Setosa
46	5.1	3.8	1.6	0.2	Setosa
47	4.6	3.2	1.4	0.2	Setosa
48	5.3	3.7	1.5	0.2	Setosa
49	5.0	3.3	1.4	0.2	Setosa

In [22]:
```
df.describe()
```

Out[22]:

	sepal.length	sepal.width	petal.length	petal.width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
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

In [23]:
```
sum = pd["sepal.length"].sum()
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_19916\3090335039.py in <module>
----> 1 sum = pd["sepal.length"].sum()

TypeError: 'module' object is not subscriptable
```

In [24]:
```
sum = df["sepal.length"].sum()
```

In [25]:
```
print(sum)
```

```
876.5
```

In [26]:
```
mean=df["sepal.length"].mean()
```

In [27]:
```
print(mean)
```

```
5.843333333333335
```

In [28]:
```
median= df["sepal.length"].median()
```

In [29]:
```
print(median)
```

```
5.8
```

In [30]:
```
data=df.columns
```

In [31]:
```
print(data)
```

```
Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
       'variety'],
      dtype='object')
```

In [32]:
```
sd = df[10:21]
```

In [33]:
```
print(sd)
```

```
    sepal.length  sepal.width  petal.length  petal.width variety
10           5.4          3.7           1.5          0.2  Setosa
11           4.8          3.4           1.6          0.2  Setosa
12           4.8          3.0           1.4          0.1  Setosa
13           4.3          3.0           1.1          0.1  Setosa
14           5.8          4.0           1.2          0.2  Setosa
15           5.7          4.4           1.5          0.4  Setosa
16           5.4          3.9           1.3          0.4  Setosa
17           5.1          3.5           1.4          0.3  Setosa
18           5.7          3.8           1.7          0.3  Setosa
19           5.1          3.8           1.5          0.3  Setosa
20           5.4          3.4           1.7          0.2  Setosa
```

In [34]:
```
sd.head()
```

Out[34]:

	sepal.length	sepal.width	petal.length	petal.width	variety
10	5.4	3.7	1.5	0.2	Setosa
11	4.8	3.4	1.6	0.2	Setosa
12	4.8	3.0	1.4	0.1	Setosa
13	4.3	3.0	1.1	0.1	Setosa
14	5.8	4.0	1.2	0.2	Setosa

In [35]:
```
col = df.columns
```

In [36]:
```
print(col)
```

```
Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
       'variety'],
      dtype='object')
```

In [38]:
```
nd = col[0:4]
```

In [39]:
```
print(nd)
```

```
Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width'], dtype='objec
t')
```

In [41]:
```
datal=df[col]
```

In [45]:
```
print(df)
```

```
     sepal.length  sepal.width  petal.length  petal.width    variety
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]
```

In [46]:
```
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   variety       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

In [47]:
```
df.isnull()
```

Out[47]:

	sepal.length	sepal.width	petal.length	petal.width	variety
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 [48]:
```
df.isnull().sum(axis = 1)
```

Out[48]:
```
0      0
1      0
2      0
3      0
4      0
      ..
145    0
146    0
147    0
148    0
149    0
Length: 150, dtype: int64
```

In [49]:
```
df.isnull().sum(axis = 0)
```

Out[49]:
```
sepal.length    0
sepal.width     0
petal.length    0
petal.width     0
variety         0
dtype: int64
```

In [50]:
```
df.dtypes
```

Out[50]:
```
sepal.length    float64
sepal.width     float64
petal.length    float64
petal.width     float64
variety          object
dtype: object
```

In [51]:
```
df['sepal.length']=df['sepal.length'].astype("int")
```

In [52]:
```
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    int32
 1   sepal.width   150 non-null    float64
 2   petal.length  150 non-null    float64
 3   petal.width   150 non-null    float64
 4   variety       150 non-null    object
dtypes: float64(3), int32(1), object(1)
memory usage: 5.4+ KB
```

In [53]:
```
!pip install sklearn
```

```
Defaulting to user installation because normal site-packages is not writeable
Collecting sklearn
  Downloading sklearn-0.0.post5.tar.gz (3.7 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Building wheels for collected packages: sklearn
  Building wheel for sklearn (setup.py): started
  Building wheel for sklearn (setup.py): finished with status 'done'
  Created wheel for sklearn: filename=sklearn-0.0.post5-py3-none-any.whl size=2951
sha256=f3d9fef0479987a82fe79be04e268afd2482bb7f058f8b3a89bbe4549f98b284
  Stored in directory: c:\users\roshan ramdas kate\appdata\local\pip\cache\wheels
\36\49\c9\2374f1dee1b599effabf63d948635e6608f62d0ccde027b7e2
Successfully built sklearn
Installing collected packages: sklearn
Successfully installed sklearn-0.0.post5
```

In [54]:
```
import sklearn as sk
```

In [55]:
```
from sklearn import preprocessing
```

In [56]:
```
minmax=preprocessing.MinMaxScaler()
```

In [57]:
```
x=df.iloc[:,:4]
```

In [58]:
```
print(x)
```

```
     sepal.length  sepal.width  petal.length  petal.width
0               5          3.5           1.4          0.2
1               4          3.0           1.4          0.2
2               4          3.2           1.3          0.2
3               4          3.1           1.5          0.2
4               5          3.6           1.4          0.2
..            ...          ...           ...          ...
145             6          3.0           5.2          2.3
146             6          2.5           5.0          1.9
147             6          3.0           5.2          2.0
148             6          3.4           5.4          2.3
149             5          3.0           5.1          1.8

[150 rows x 4 columns]
```

In [59]:
```
x_scaled = minmax.fit_transform(x)
```

In [60]:
```
df_normalized = pd.DataFrame(x_scaled)
```

In [61]:
```
df_normalized.head()
```

Out[61]:

	0	1	2	3
0	0.333333	0.625000	0.067797	0.041667
1	0.000000	0.416667	0.067797	0.041667
2	0.000000	0.500000	0.050847	0.041667
3	0.000000	0.458333	0.084746	0.041667
4	0.333333	0.666667	0.067797	0.041667

In [63]:
```
df['variety'].unique()
```

Out[63]:
```
array(['Setosa', 'Versicolor', 'Virginica'], dtype=object)
```

In [64]:
```
dflabel=preprocessing.LabelEncoder()
```

In [65]:
```
df['class']= dflabel.fit_transform(df['variety'])
```

In [66]:
```
print(df)
```

```
     sepal.length  sepal.width  petal.length  petal.width    variety  class
0               5          3.5           1.4          0.2     Setosa      0
1               4          3.0           1.4          0.2     Setosa      0
2               4          3.2           1.3          0.2     Setosa      0
3               4          3.1           1.5          0.2     Setosa      0
4               5          3.6           1.4          0.2     Setosa      0
..            ...          ...           ...          ...        ...    ...
145             6          3.0           5.2          2.3  Virginica      2
146             6          2.5           5.0          1.9  Virginica      2
147             6          3.0           5.2          2.0  Virginica      2
148             6          3.4           5.4          2.3  Virginica      2
149             5          3.0           5.1          1.8  Virginica      2

[150 rows x 6 columns]
```

In [67]:
```
df['class'].unique()
```

Out[67]:
```
array([0, 1, 2])
```

In [ ]: