In [1]:
```
import pandas as pd
import numpy as np #importing libraries
import random
```

In [3]:
```
df=pd.read_csv("iris.csv")#read csv file
df
```

Out[3]:

	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 × 5 columns

In [5]:
```
df.info() #showing 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
```

In [48]:
```
df.describe()#perform all predefined operations
```

Out[48]:

	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

In [7]:
```
df.head(20) #shows first 20 values
```

Out[7]:

	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
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

In [8]:
```
df.tail(20)#shows last 20 records
```

Out[8]:

	sepal_length	sepal_width	petal_length	petal_width	species
130	7.4	2.8	6.1	1.9	virginica
131	7.9	3.8	6.4	2.0	virginica
132	6.4	2.8	5.6	2.2	virginica
133	6.3	2.8	5.1	1.5	virginica
134	6.1	2.6	5.6	1.4	virginica
135	7.7	3.0	6.1	2.3	virginica
136	6.3	3.4	5.6	2.4	virginica
137	6.4	3.1	5.5	1.8	virginica
138	6.0	3.0	4.8	1.8	virginica
139	6.9	3.1	5.4	2.1	virginica
140	6.7	3.1	5.6	2.4	virginica
141	6.9	3.1	5.1	2.3	virginica
142	5.8	2.7	5.1	1.9	virginica
143	6.8	3.2	5.9	2.3	virginica
144	6.7	3.3	5.7	2.5	virginica
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 [9]:
```
df.columns#shows all columns
```

Out[9]:
```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'],
      dtype='object')
```

In [11]:
```
df['species'].value_counts() #it counts number of values present in columns
```

Out[11]:
```
setosa        50
versicolor    50
virginica     50
Name: species, dtype: int64
```

In [18]:
```
df['petal_length'].mean()#mean
```

Out[18]:
```
3.7586666666666693
```

In [14]:
```
df['sepal_length'].mean()
```

Out[14]:
```
5.843333333333335
```

In [15]:
```
df['petal_length'].mean()
```

Out[15]:
```
3.7586666666666693
```

In [16]:
```
df['petal_width'].mean()
```

Out[16]:
```
1.1986666666666672
```

In [19]:
```
df['petal_length'].median()#median
```

Out[19]:
```
4.35
```

In [20]:
```
df['sepal_length'].median()
```

Out[20]:
```
5.8
```

In [21]:
```
df['petal_length'].median()
```

Out[21]:
```
4.35
```

In [22]:
```
df['petal_width'].median()
```

Out[22]:
```
1.3
```

In [23]:
```
df['petal_length'].max()#maximum value
```

Out[23]:
```
6.9
```

In [24]:
```
df['sepal_length'].max()
```

Out[24]:
```
7.9
```

In [25]:
```
df['petal_length'].max()
```

Out[25]:
```
6.9
```

In [26]:
```
df['petal_width'].max()
```

Out[26]:
```
2.5
```

In [27]:
```
df['petal_length'].min()#minimun value
```

Out[27]:
```
1.0
```

In [28]:
```
df['sepal_length'].min()
```

Out[28]: 4.3

In [29]:
```
df['petal_length'].min()
```

Out[29]: 1.0

In [30]:
```
df['petal_width'].min()
```

Out[30]: 0.1

In [31]:
```
df['petal_length'].std()#standard deviation
```

Out[31]: 1.7644204199522617

In [32]:
```
df['sepal_length'].std()
```

Out[32]: 0.8280661279778629

In [33]:
```
df['petal_length'].std()
```

Out[33]: 1.7644204199522617

In [34]:
```
df['petal_width'].std()
```

Out[34]: 0.7631607417008414

In [35]:
```
df.replace('virginica', '0', inplace=True)#replace verginica by 0
```

In [36]:
```
df
```

Out[36]:

	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	0
146	6.3	2.5	5.0	1.9	0
147	6.5	3.0	5.2	2.0	0
148	6.2	3.4	5.4	2.3	0
149	5.9	3.0	5.1	1.8	0

150 rows × 5 columns

In [37]:
```
df.replace('setosa', '1', inplace=True)#replace setosaa by 1
```

In [38]:
```
df
```

Out[38]:

	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	0
146	6.3	2.5	5.0	1.9	0
147	6.5	3.0	5.2	2.0	0
148	6.2	3.4	5.4	2.3	0
149	5.9	3.0	5.1	1.8	0

150 rows × 5 columns

In [39]:
```
df.replace('versicolor', '2', inplace=True) #replace versicolor by 2
```

In [40]:
```
df
```

Out[40]:

	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	0
146	6.3	2.5	5.0	1.9	0
147	6.5	3.0	5.2	2.0	0
148	6.2	3.4	5.4	2.3	0
149	5.9	3.0	5.1	1.8	0

150 rows × 5 columns

In [42]:
```
df['species'].value_counts()#it shows value counts of specific columns
```

Out[42]:
```
1    50
2    50
0    50
Name: species, dtype: int64
```

In [43]:
```
df.groupby('species').count()#group by count in species column
```

Out[43]:

species	sepal_length	sepal_width	petal_length	petal_width
0	50	50	50	50
1	50	50	50	50
2	50	50	50	50

In [44]:
```
df.groupby('species').size()#group by c size in species column
```

Out[44]:
```
species
0    50
1    50
2    50
dtype: int64
```

In [45]:
```
df.groupby('species').mean()#group by mean in species column
```

Out[45]:

species	sepal_length	sepal_width	petal_length	petal_width
0	6.588	2.974	5.552	2.026
1	5.006	3.418	1.464	0.244
2	5.936	2.770	4.260	1.326

In [46]:
```
df.groupby('species').median()#group by median in species column
```

Out[46]:

species	sepal_length	sepal_width	petal_length	petal_width
0	6.5	3.0	5.55	2.0
1	5.0	3.4	1.50	0.2
2	5.9	2.8	4.35	1.3

In [47]:
```
df.groupby('species').std()#group by standard deviation in species column
```

Out[47]:

species	sepal_length	sepal_width	petal_length	petal_width
0	0.635880	0.322497	0.551895	0.274650
1	0.352490	0.381024	0.173511	0.107210
2	0.516171	0.313798	0.469911	0.197753

In [52]:
```
df.groupby('species').quantile(0.25) #shows percentage of columns
```

Out[52]:

species	sepal_length	sepal_width	petal_length	petal_width
0	6.225	2.800	5.1	1.8
1	4.800	3.125	1.4	0.2
2	5.600	2.525	4.0	1.2

In [53]:
```
df.groupby('species').quantile(0.50)
```

Out[53]:

species	sepal_length	sepal_width	petal_length	petal_width
0	6.5	3.0	5.55	2.0
1	5.0	3.4	1.50	0.2
2	5.9	2.8	4.35	1.3

In [54]:
```
df.groupby('species').quantile(0.75)
```

Out[54]:

species	sepal_length	sepal_width	petal_length	petal_width
0	6.9	3.175	5.875	2.3
1	5.2	3.675	1.575	0.3
2	6.3	3.000	4.600	1.5

In [ ]:
```

```