

In [105]:

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files
under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved
as output when you create a version using "Save & Run All"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of
the current session
```

```
/kaggle/input/iris/Iris.csv
/kaggle/input/iris/database.sqlite
```

In [1]:

```
#importing the external libraries
import pandas as pd
data = pd.read_csv('/kaggle/input/iris/Iris.csv')
data.head(10)
```

Out[1]:

	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
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa

In [107]:

```
data
```

Out[107]:

	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	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

150 rows × 6 columns

In [108]:

```
data.tail()
```

Out[108]:

	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

In [109]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   Id              150 non-null   int64
 1   SepalLengthCm   150 non-null   float64
 2   SepalWidthCm    150 non-null   float64
 3   PetalLengthCm   150 non-null   float64
 4   PetalWidthCm    150 non-null   float64
 5   Species         150 non-null   object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

In [110]:

```
data.shape
```

Out[110]:

(150, 6)

In [111]:

```
data.columns
```

Out[111]:

```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
```

In [112]:

```
data[['Species', 'SepalLengthCm']]
```

Out[112]:

	Species	SepalLengthCm
0	Iris-setosa	5.1
1	Iris-setosa	4.9
2	Iris-setosa	4.7
3	Iris-setosa	4.6
4	Iris-setosa	5.0
...	...	...
145	Iris-virginica	6.7
146	Iris-virginica	6.3
147	Iris-virginica	6.5
148	Iris-virginica	6.2
149	Iris-virginica	5.9

150 rows x 2 columns

In [113]:

```
data.index
```

Out[113]:

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

In [114]:

```
data.SepalLengthCm
```

Out[114]:

```
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: SepalLengthCm, Length: 150, dtype: float64
```

In [115]:

```
data.SepalLengthCm!=4
```

Out[115]:

```
0      True
1      True
2      True
3      True
4      True
...
145     True
146     True
147     True
148     True
149     True
Name: SepalLengthCm, Length: 150, dtype: bool
```

In [116]:

```
data.describe()
```

Out[116]:

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

In [117]:

```
data.loc[140:150, ['SepalLengthCm', 'SepalWidthCm', 'PetalWidthCm']]
```

Out[117]:

	SepalLengthCm	SepalWidthCm	PetalWidthCm
140	6.7	3.1	2.4
141	6.9	3.1	2.3
142	5.8	2.7	1.9
143	6.8	3.2	2.3
144	6.7	3.3	2.5
145	6.7	3.0	2.3
146	6.3	2.5	1.9
147	6.5	3.0	2.0
148	6.2	3.4	2.3
149	5.9	3.0	1.8

In [118]:

```
data.columns
```

Out[118]:

Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',  
 'Species'],  
 dtype='object')

In [119]:

```
#Get the last row of the DataFrame:  
data.iloc[-1]
```

Out[119]:

Id 150  
SepalLengthCm 5.9  
SepalWidthCm 3.0  
PetalLengthCm 5.1  
PetalWidthCm 1.8  
Species Iris-virginica  
Name: 149, dtype: object

In [120]:

```
# Get every 3rd row of the DataFrame:  
data.iloc[::3]
```

Out[120]:

Out[120]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
15	16	5.7	4.4	1.5	0.4	Iris-setosa
18	19	5.7	3.8	1.7	0.3	Iris-setosa
21	22	5.1	3.7	1.5	0.4	Iris-setosa
24	25	4.8	3.4	1.9	0.2	Iris-setosa
27	28	5.2	3.5	1.5	0.2	Iris-setosa
30	31	4.8	3.1	1.6	0.2	Iris-setosa
33	34	5.5	4.2	1.4	0.2	Iris-setosa
36	37	5.5	3.5	1.3	0.2	Iris-setosa
39	40	5.1	3.4	1.5	0.2	Iris-setosa
42	43	4.4	3.2	1.3	0.2	Iris-setosa
45	46	4.8	3.0	1.4	0.3	Iris-setosa
48	49	5.3	3.7	1.5	0.2	Iris-setosa
51	52	6.4	3.2	4.5	1.5	Iris-versicolor
54	55	6.5	2.8	4.6	1.5	Iris-versicolor
57	58	4.9	2.4	3.3	1.0	Iris-versicolor
60	61	5.0	2.0	3.5	1.0	Iris-versicolor
63	64	6.1	2.9	4.7	1.4	Iris-versicolor
66	67	5.6	3.0	4.5	1.5	Iris-versicolor
69	70	5.6	2.5	3.9	1.1	Iris-versicolor
72	73	6.3	2.5	4.9	1.5	Iris-versicolor
75	76	6.6	3.0	4.4	1.4	Iris-versicolor
78	79	6.0	2.9	4.5	1.5	Iris-versicolor
81	82	5.5	2.4	3.7	1.0	Iris-versicolor
84	85	5.4	3.0	4.5	1.5	Iris-versicolor
87	88	6.3	2.3	4.4	1.3	Iris-versicolor
90	91	5.5	2.6	4.4	1.2	Iris-versicolor
93	94	5.0	2.3	3.3	1.0	Iris-versicolor
96	97	5.7	2.9	4.2	1.3	Iris-versicolor
99	100	5.7	2.8	4.1	1.3	Iris-versicolor
102	103	7.1	3.0	5.9	2.1	Iris-virginica
105	106	7.6	3.0	6.6	2.1	Iris-virginica
108	109	6.7	2.5	5.8	1.8	Iris-virginica
111	112	6.4	2.7	5.3	1.9	Iris-virginica
114	115	5.8	2.8	5.1	2.4	Iris-virginica
117	118	7.7	3.8	6.7	2.2	Iris-virginica
120	121	6.9	3.2	5.7	2.3	Iris-virginica
123	124	6.3	2.7	4.9	1.8	Iris-virginica
126	127	6.2	2.8	4.8	1.8	Iris-virginica

129	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
130	130	7.2	3.0	5.8	1.6	Iris-virginica
132	133	6.4	2.8	5.6	2.2	Iris-virginica
135	136	7.7	3.0	6.1	2.3	Iris-virginica
138	139	6.0	3.0	4.8	1.8	Iris-virginica
141	142	6.9	3.1	5.1	2.3	Iris-virginica
144	145	6.7	3.3	5.7	2.5	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica

In [121]:

```
#get the rows index labels '0', '5' and '10':
data.iloc[['0', '5', '10']]
```

Out[121]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa

In [122]:

```
data.iloc[4]
```

Out[122]:

Id 5  
SepalLengthCm 5.0  
SepalWidthCm 3.6  
PetalLengthCm 1.4  
PetalWidthCm 0.2  
Species Iris-setosa  
Name: 4, dtype: object

In [123]:

```
data.iloc[4:10,2:5]
```

Out[123]:

	SepalWidthCm	PetalLengthCm	PetalWidthCm
4	3.6	1.4	0.2
5	3.9	1.7	0.4
6	3.4	1.4	0.3
7	3.4	1.5	0.2
8	2.9	1.4	0.2
9	3.1	1.5	0.1

In [124]:

```
data.iloc[:10,:3]
```

Out[124]:

	Id	SepalLengthCm	SepalWidthCm
0	1	5.1	3.5
1	2	4.9	3.0
2	3	4.7	3.2

3	Id	SepalLengthCm	PetalLengthCm	SepalWidthCm
4	5		5.0	3.6
5	6		5.4	3.9
6	7		4.6	3.4
7	8		5.0	3.4
8	9		4.4	2.9
9	10		4.9	3.1

In [125]:

```
# Get the columns 'sepalength', 'petallength', and 'petalwidth':
data.loc[10:50,['SepalLengthCm','PetalLengthCm','PetalWidthCm']]
```

Out[125]:

	SepalLengthCm	PetalLengthCm	PetalWidthCm
10	5.4	1.5	0.2
11	4.8	1.6	0.2
12	4.8	1.4	0.1
13	4.3	1.1	0.1
14	5.8	1.2	0.2
15	5.7	1.5	0.4
16	5.4	1.3	0.4
17	5.1	1.4	0.3
18	5.7	1.7	0.3
19	5.1	1.5	0.3
20	5.4	1.7	0.2
21	5.1	1.5	0.4
22	4.6	1.0	0.2
23	5.1	1.7	0.5
24	4.8	1.9	0.2
25	5.0	1.6	0.2
26	5.0	1.6	0.4
27	5.2	1.5	0.2
28	5.2	1.4	0.2
29	4.7	1.6	0.2
30	4.8	1.6	0.2
31	5.4	1.5	0.4
32	5.2	1.5	0.1
33	5.5	1.4	0.2
34	4.9	1.5	0.1
35	5.0	1.2	0.2
36	5.5	1.3	0.2
37	4.9	1.5	0.1
38	4.4	1.3	0.2
39	5.1	1.5	0.2
40	5.0	1.3	0.3
41	4.5	1.3	0.3

42	SepalLengthCm	PetalLengthCm	PetalWidthCm
43	5.0	1.6	0.6
44	5.1	1.9	0.4
45	4.8	1.4	0.3
46	5.1	1.6	0.2
47	4.6	1.4	0.2
48	5.3	1.5	0.2
49	5.0	1.4	0.2
50	7.0	4.7	1.4

In [126]:

```
# Slice the DataFrame to only include the rows with index labels '30' to '40':
data.loc['30':'40']
```

Out[126]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
30	31	4.8	3.1	1.6	0.2	Iris-setosa
31	32	5.4	3.4	1.5	0.4	Iris-setosa
32	33	5.2	4.1	1.5	0.1	Iris-setosa
33	34	5.5	4.2	1.4	0.2	Iris-setosa
34	35	4.9	3.1	1.5	0.1	Iris-setosa
35	36	5.0	3.2	1.2	0.2	Iris-setosa
36	37	5.5	3.5	1.3	0.2	Iris-setosa
37	38	4.9	3.1	1.5	0.1	Iris-setosa
38	39	4.4	3.0	1.3	0.2	Iris-setosa
39	40	5.1	3.4	1.5	0.2	Iris-setosa
40	41	5.0	3.5	1.3	0.3	Iris-setosa

In [127]:

```
# Basic Statistics:
#Get the mean value of 'sepal_length':
data['SepalLengthCm'].mean()
```

Out[127]:

5.8433333333333334

In [128]:

```
#Get the median value of 'sepal_width':
data['SepalWidthCm'].median()
```

Out[128]:

3.0

In [129]:

```
# Get the mode value of 'species':
data['Species'].mode()
```

Out[129]:

0 Iris-setosa
1 Iris-versicolor
2 Iris-virginica
Name: Species, dtype: object



In [130]:

```
data['Species'].mode()[0]
```

Out[130]:

'Iris-setosa'

In [131]:

```
# Get the standard deviation of 'petal_length':
data['PetalLengthCm'].std()
```

Out[131]:

1.7644204199522626

In [132]:

```
# Calculate the correlation between 'sepal_length' and 'sepal_width':
data[['SepalLengthCm', 'SepalWidthCm']].corr()
```

Out[132]:

	SepalLengthCm	SepalWidthCm
SepalLengthCm	1.000000	-0.109369
SepalWidthCm	-0.109369	1.000000

In [133]:

```
#Conditional Selection:
#Filter rows where 'sepal_length' is greater than 6:
data[data['SepalLengthCm'] > 6]
```

Out[133]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
50	51	7.0	3.2	4.7	1.4	Iris-versicolor
51	52	6.4	3.2	4.5	1.5	Iris-versicolor
52	53	6.9	3.1	4.9	1.5	Iris-versicolor
54	55	6.5	2.8	4.6	1.5	Iris-versicolor
56	57	6.3	3.3	4.7	1.6	Iris-versicolor
...	...	...	...	...	...	...
144	145	6.7	3.3	5.7	2.5	Iris-virginica
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

61 rows x 6 columns

In [134]:

```
data[data['SepalLengthCm'] > 6]['SepalLengthCm']
```

Out[134]:

50 7.0  
51 6.4  
52 6.9  
54 6.5  
56 6.3

...  
144 6.7  
145 6.7  
146 6.3  
147 6.5  
148 6.2  
Name: SepalLengthCm, Length: 61, dtype: float64

In [135]:

```
#Filter rows where 'species' is 'setosa':  
data[data['Species'] == 'setosa']
```

Out[135]:

	<b>Id</b>	<b>SepalLengthCm</b>	<b>SepalWidthCm</b>	<b>PetalLengthCm</b>	<b>PetalWidthCm</b>	<b>Species</b>
--	-----------	----------------------	---------------------	----------------------	---------------------	----------------

In [136]:

```
data[data['Species']=='Iris-setosa']
```

Out[136]:

	<b>Id</b>	<b>SepalLengthCm</b>	<b>SepalWidthCm</b>	<b>PetalLengthCm</b>	<b>PetalWidthCm</b>	<b>Species</b>
<b>0</b>	<b>1</b>	<b>5.1</b>	<b>3.5</b>	<b>1.4</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>1</b>	<b>2</b>	<b>4.9</b>	<b>3.0</b>	<b>1.4</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>2</b>	<b>3</b>	<b>4.7</b>	<b>3.2</b>	<b>1.3</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>3</b>	<b>4</b>	<b>4.6</b>	<b>3.1</b>	<b>1.5</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>4</b>	<b>5</b>	<b>5.0</b>	<b>3.6</b>	<b>1.4</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>5</b>	<b>6</b>	<b>5.4</b>	<b>3.9</b>	<b>1.7</b>	<b>0.4</b>	<b>Iris-setosa</b>
<b>6</b>	<b>7</b>	<b>4.6</b>	<b>3.4</b>	<b>1.4</b>	<b>0.3</b>	<b>Iris-setosa</b>
<b>7</b>	<b>8</b>	<b>5.0</b>	<b>3.4</b>	<b>1.5</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>8</b>	<b>9</b>	<b>4.4</b>	<b>2.9</b>	<b>1.4</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>9</b>	<b>10</b>	<b>4.9</b>	<b>3.1</b>	<b>1.5</b>	<b>0.1</b>	<b>Iris-setosa</b>
<b>10</b>	<b>11</b>	<b>5.4</b>	<b>3.7</b>	<b>1.5</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>11</b>	<b>12</b>	<b>4.8</b>	<b>3.4</b>	<b>1.6</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>12</b>	<b>13</b>	<b>4.8</b>	<b>3.0</b>	<b>1.4</b>	<b>0.1</b>	<b>Iris-setosa</b>
<b>13</b>	<b>14</b>	<b>4.3</b>	<b>3.0</b>	<b>1.1</b>	<b>0.1</b>	<b>Iris-setosa</b>
<b>14</b>	<b>15</b>	<b>5.8</b>	<b>4.0</b>	<b>1.2</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>15</b>	<b>16</b>	<b>5.7</b>	<b>4.4</b>	<b>1.5</b>	<b>0.4</b>	<b>Iris-setosa</b>
<b>16</b>	<b>17</b>	<b>5.4</b>	<b>3.9</b>	<b>1.3</b>	<b>0.4</b>	<b>Iris-setosa</b>
<b>17</b>	<b>18</b>	<b>5.1</b>	<b>3.5</b>	<b>1.4</b>	<b>0.3</b>	<b>Iris-setosa</b>
<b>18</b>	<b>19</b>	<b>5.7</b>	<b>3.8</b>	<b>1.7</b>	<b>0.3</b>	<b>Iris-setosa</b>
<b>19</b>	<b>20</b>	<b>5.1</b>	<b>3.8</b>	<b>1.5</b>	<b>0.3</b>	<b>Iris-setosa</b>
<b>20</b>	<b>21</b>	<b>5.4</b>	<b>3.4</b>	<b>1.7</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>21</b>	<b>22</b>	<b>5.1</b>	<b>3.7</b>	<b>1.5</b>	<b>0.4</b>	<b>Iris-setosa</b>
<b>22</b>	<b>23</b>	<b>4.6</b>	<b>3.6</b>	<b>1.0</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>23</b>	<b>24</b>	<b>5.1</b>	<b>3.3</b>	<b>1.7</b>	<b>0.5</b>	<b>Iris-setosa</b>
<b>24</b>	<b>25</b>	<b>4.8</b>	<b>3.4</b>	<b>1.9</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>25</b>	<b>26</b>	<b>5.0</b>	<b>3.0</b>	<b>1.6</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>26</b>	<b>27</b>	<b>5.0</b>	<b>3.4</b>	<b>1.6</b>	<b>0.4</b>	<b>Iris-setosa</b>
<b>27</b>	<b>28</b>	<b>5.2</b>	<b>3.5</b>	<b>1.5</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>28</b>	<b>29</b>	<b>5.2</b>	<b>3.4</b>	<b>1.4</b>	<b>0.2</b>	<b>Iris-setosa</b>

29	30	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Iris-species
30	31	4.8	3.1	1.6	0.2	Iris-setosa
31	32	5.4	3.4	1.5	0.4	Iris-setosa
32	33	5.2	4.1	1.5	0.1	Iris-setosa
33	34	5.5	4.2	1.4	0.2	Iris-setosa
34	35	4.9	3.1	1.5	0.1	Iris-setosa
35	36	5.0	3.2	1.2	0.2	Iris-setosa
36	37	5.5	3.5	1.3	0.2	Iris-setosa
37	38	4.9	3.1	1.5	0.1	Iris-setosa
38	39	4.4	3.0	1.3	0.2	Iris-setosa
39	40	5.1	3.4	1.5	0.2	Iris-setosa
40	41	5.0	3.5	1.3	0.3	Iris-setosa
41	42	4.5	2.3	1.3	0.3	Iris-setosa
42	43	4.4	3.2	1.3	0.2	Iris-setosa
43	44	5.0	3.5	1.6	0.6	Iris-setosa
44	45	5.1	3.8	1.9	0.4	Iris-setosa
45	46	4.8	3.0	1.4	0.3	Iris-setosa
46	47	5.1	3.8	1.6	0.2	Iris-setosa
47	48	4.6	3.2	1.4	0.2	Iris-setosa
48	49	5.3	3.7	1.5	0.2	Iris-setosa
49	50	5.0	3.3	1.4	0.2	Iris-setosa

In [137]:

```
#Filter rows where 'petal_width' is between 1 and 2:
data[(data['PetalWidthCm'] >= 1) & (data['PetalWidthCm'] <= 2)]
```

Out[137]:

	<b>Id</b>	<b>SepalLengthCm</b>	<b>SepalWidthCm</b>	<b>PetalLengthCm</b>	<b>PetalWidthCm</b>	<b>Species</b>
	<b>50</b>	<b>51</b>	<b>7.0</b>	<b>3.2</b>	<b>4.7</b>	<b>1.4</b> Iris-versicolor
	<b>51</b>	<b>52</b>	<b>6.4</b>	<b>3.2</b>	<b>4.5</b>	<b>1.5</b> Iris-versicolor
	<b>52</b>	<b>53</b>	<b>6.9</b>	<b>3.1</b>	<b>4.9</b>	<b>1.5</b> Iris-versicolor
	<b>53</b>	<b>54</b>	<b>5.5</b>	<b>2.3</b>	<b>4.0</b>	<b>1.3</b> Iris-versicolor
	<b>54</b>	<b>55</b>	<b>6.5</b>	<b>2.8</b>	<b>4.6</b>	<b>1.5</b> Iris-versicolor
	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>	<b>...</b>
	<b>138</b>	<b>139</b>	<b>6.0</b>	<b>3.0</b>	<b>4.8</b>	<b>1.8</b> Iris-virginica
	<b>142</b>	<b>143</b>	<b>5.8</b>	<b>2.7</b>	<b>5.1</b>	<b>1.9</b> Iris-virginica
	<b>146</b>	<b>147</b>	<b>6.3</b>	<b>2.5</b>	<b>5.0</b>	<b>1.9</b> Iris-virginica
	<b>147</b>	<b>148</b>	<b>6.5</b>	<b>3.0</b>	<b>5.2</b>	<b>2.0</b> Iris-virginica
	<b>149</b>	<b>150</b>	<b>5.9</b>	<b>3.0</b>	<b>5.1</b>	<b>1.8</b> Iris-virginica

77 rows x 6 columns

In [138]:

```
# Filter rows where 'sepal_width' is greater than 4 or 'petal_width' is less than 1:
data[(data['SepalWidthCm'] > 4) | (data['PetalWidthCm'] < 1)]
```

Out[138]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
--	----	---------------	--------------	---------------	--------------	---------

Id	Sepal		Petal		Species
	LengthCm	WidthCm	LengthCm	WidthCm	
0	1	5.1	3.5	1.4	Iris-setosa
1	2	4.9	3.0	1.4	Iris-setosa
2	3	4.7	3.2	1.3	Iris-setosa
3	4	4.6	3.1	1.5	Iris-setosa
4	5	5.0	3.6	1.4	Iris-setosa
5	6	5.4	3.9	1.7	Iris-setosa
6	7	4.6	3.4	1.4	Iris-setosa
7	8	5.0	3.4	1.5	Iris-setosa
8	9	4.4	2.9	1.4	Iris-setosa
9	10	4.9	3.1	1.5	Iris-setosa
10	11	5.4	3.7	1.5	Iris-setosa
11	12	4.8	3.4	1.6	Iris-setosa
12	13	4.8	3.0	1.4	Iris-setosa
13	14	4.3	3.0	1.1	Iris-setosa
14	15	5.8	4.0	1.2	Iris-setosa
15	16	5.7	4.4	1.5	Iris-setosa
16	17	5.4	3.9	1.3	Iris-setosa
17	18	5.1	3.5	1.4	Iris-setosa
18	19	5.7	3.8	1.7	Iris-setosa
19	20	5.1	3.8	1.5	Iris-setosa
20	21	5.4	3.4	1.7	Iris-setosa
21	22	5.1	3.7	1.5	Iris-setosa
22	23	4.6	3.6	1.0	Iris-setosa
23	24	5.1	3.3	1.7	Iris-setosa
24	25	4.8	3.4	1.9	Iris-setosa
25	26	5.0	3.0	1.6	Iris-setosa
26	27	5.0	3.4	1.6	Iris-setosa
27	28	5.2	3.5	1.5	Iris-setosa
28	29	5.2	3.4	1.4	Iris-setosa
29	30	4.7	3.2	1.6	Iris-setosa
30	31	4.8	3.1	1.6	Iris-setosa
31	32	5.4	3.4	1.5	Iris-setosa
32	33	5.2	4.1	1.5	Iris-setosa
33	34	5.5	4.2	1.4	Iris-setosa
34	35	4.9	3.1	1.5	Iris-setosa
35	36	5.0	3.2	1.2	Iris-setosa
36	37	5.5	3.5	1.3	Iris-setosa
37	38	4.9	3.1	1.5	Iris-setosa
38	39	4.4	3.0	1.3	Iris-setosa
39	40	5.1	3.4	1.5	Iris-setosa
40	41	5.0	3.5	1.3	Iris-setosa
41	42	4.5	2.3	1.3	Iris-setosa
42	43	4.4	3.2	1.3	Iris-setosa
43	44	5.0	3.5	1.6	Iris-setosa
44	45	5.1	3.8	1.0	Iris-setosa

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
45	4.8	3.0	1.4	0.3	Iris-setosa
46	5.1	3.8	1.6	0.2	Iris-setosa
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

In [195]:

```
# Data Cleaning:
#Replace missing values in 'petal_width' with the median value:
data1= pd.DataFrame(data['PetalWidthCm'].fillna(data['PetalWidthCm'].median()))
data1
```

Out[195]:

PetalWidthCm	
0	0.2
2	0.2
3	0.2
5	0.4
11	0.2
13	0.1
14	0.2
22	0.2
24	0.2
50	1.4
51	1.5
52	1.5
53	1.3
54	1.5
57	1.0
59	1.4
60	1.0
61	1.5
64	1.3
65	1.4
67	1.0
70	1.8
74	1.3
77	1.7
80	1.1
81	1.0
83	1.6
98	1.1
100	2.5
101	1.9
102	2.1
103	1.8

<b>104</b>	<b>PetalWidthCm</b>	<b>2.2</b>
<b>105</b>		<b>2.1</b>
<b>106</b>		<b>1.7</b>
<b>107</b>		<b>1.8</b>
<b>109</b>		<b>2.5</b>
<b>111</b>		<b>1.9</b>
<b>112</b>		<b>2.1</b>
<b>113</b>		<b>2.0</b>
<b>117</b>		<b>2.2</b>
<b>118</b>		<b>2.3</b>
<b>120</b>		<b>2.3</b>
<b>121</b>		<b>2.0</b>
<b>126</b>		<b>1.8</b>
<b>131</b>		<b>2.0</b>
<b>139</b>		<b>2.1</b>
<b>145</b>		<b>2.3</b>

In [196]:

```
#To check the missing values
missing_values=data.isnull().sum()
missing_values
```

Out[196]:

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

In [197]:

```
#Count null values in each column:
null_counts = data.isnull().sum()
null_counts
```

Out[197]:

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

In [142]:

```
#Remove duplicate rows based on the 'species' column:
data.drop_duplicates(subset='Species')
```

Out[142]:

	<b>Id</b>	<b>SepalLengthCm</b>	<b>SepalWidthCm</b>	<b>PetalLengthCm</b>	<b>PetalWidthCm</b>	<b>Species</b>
<b>0</b>	<b>1</b>	<b>5.1</b>	<b>3.5</b>	<b>1.4</b>	<b>0.2</b>	<b>Iris-setosa</b>
<b>50</b>	<b>51</b>	<b>7.0</b>	<b>3.2</b>	<b>4.7</b>	<b>1.4</b>	<b>Iris-versicolor</b>
<b>100</b>	<b>101</b>	<b>6.3</b>	<b>3.3</b>	<b>6.0</b>	<b>2.5</b>	<b>Iris-virginica</b>

In [143]:

```
# to find the duplicates
duplicates=data.duplicated()
print(duplicates.sum())
```

0

In [144]:

```
#Remove duplicates
data.drop_duplicates()
```

Out[144]:

	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 x 6 columns

In [145]:

```
data1.isnull().sum()
```

Out[145]:

```
PetalWidthCm    0
dtype: int64
```

In [146]:

```
#Remove rows where 'sepal_width' is less than or equal to 0:
data[data['SepalWidthCm'] > 0].reset_index(drop=True)
```

Out[146]:

	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  
Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Species

150 rows x 6 columns

In [147]:

```
#Remove rows where 'sepal_width' is less than or equal to 0:  
data[data['SepalWidthCm'] > 0].reset_index(drop=True)
```

Out[147]:

	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 x 6 columns

In [148]:

data

Out[148]:

	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 x 6 columns

In [149]:

```
#Convert 'species' column values to uppercase:  
data['Species'].str.upper()
```

Out[149]:

0 IRIS-SETOSA  
1 IRIS-SETOSA  
2 IRIS-SETOSA  
3 IRIS-SETOSA



```
3         IRIS-SETOSA
4         IRIS-SETOSA
5         ...
145        IRIS-VIRGINICA
146        IRIS-VIRGINICA
147        IRIS-VIRGINICA
148        IRIS-VIRGINICA
149        IRIS-VIRGINICA
Name: Species, Length: 150, dtype: object
```

In [198]:

```
# Sorting:
#Sort the DataFrame by 'sepal_length' in ascending order:
data.sort_values(by='SepalLengthCm',inplace = True)
data
```

Out[198]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
22	23	4.6	3.6	1.0	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
24	25	4.8	3.4	1.9	0.2	Iris-setosa
106	107	4.9	2.5	4.5	1.7	Iris-virginica
57	58	4.9	2.4	3.3	1.0	Iris-versicolor
60	61	5.0	2.0	3.5	1.0	Iris-versicolor
98	99	5.1	2.5	3.0	1.1	Iris-versicolor
0	1	5.1	3.5	1.4	0.2	Iris-setosa
59	60	5.2	2.7	3.9	1.4	Iris-versicolor
5	6	5.4	3.9	1.7	0.4	Iris-setosa
53	54	5.5	2.3	4.0	1.3	Iris-versicolor
81	82	5.5	2.4	3.7	1.0	Iris-versicolor
80	81	5.5	2.4	3.8	1.1	Iris-versicolor
121	122	5.6	2.8	4.9	2.0	Iris-virginica
64	65	5.6	2.9	3.6	1.3	Iris-versicolor
113	114	5.7	2.5	5.0	2.0	Iris-virginica
101	102	5.8	2.7	5.1	1.9	Iris-virginica
14	15	5.8	4.0	1.2	0.2	Iris-setosa
67	68	5.8	2.7	4.1	1.0	Iris-versicolor
70	71	5.9	3.2	4.8	1.8	Iris-versicolor
61	62	5.9	3.0	4.2	1.5	Iris-versicolor
83	84	6.0	2.7	5.1	1.6	Iris-versicolor
126	127	6.2	2.8	4.8	1.8	Iris-virginica
103	104	6.3	2.9	5.6	1.8	Iris-virginica
100	101	6.3	3.3	6.0	2.5	Iris-virginica
111	112	6.4	2.7	5.3	1.9	Iris-virginica
51	52	6.4	3.2	4.5	1.5	Iris-versicolor
74	75	6.4	2.9	4.3	1.3	Iris-versicolor
54	55	6.5	2.8	4.6	1.5	Iris-versicolor

104	105	6.5	3.0	5.8	2.2	Iris-virginica
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	
77	78	6.7	3.0	5.0	1.7	Iris-versicolor
65	66	6.7	3.1	4.4	1.4	Iris-versicolor
145	146	6.7	3.0	5.2	2.3	Iris-virginica
112	113	6.8	3.0	5.5	2.1	Iris-virginica
139	140	6.9	3.1	5.4	2.1	Iris-virginica
120	121	6.9	3.2	5.7	2.3	Iris-virginica
52	53	6.9	3.1	4.9	1.5	Iris-versicolor
50	51	7.0	3.2	4.7	1.4	Iris-versicolor
102	103	7.1	3.0	5.9	2.1	Iris-virginica
109	110	7.2	3.6	6.1	2.5	Iris-virginica
107	108	7.3	2.9	6.3	1.8	Iris-virginica
105	106	7.6	3.0	6.6	2.1	Iris-virginica
117	118	7.7	3.8	6.7	2.2	Iris-virginica
118	119	7.7	2.6	6.9	2.3	Iris-virginica
131	132	7.9	3.8	6.4	2.0	Iris-virginica

In [199]:

```
# Sort the DataFrame by 'petal_width' in descending order:
data.sort_values(by='PetalWidthCm', ascending=False,inplace=True)
data
```

Out[199]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
100	101	6.3	3.3	6.0	2.5	Iris-virginica
109	110	7.2	3.6	6.1	2.5	Iris-virginica
118	119	7.7	2.6	6.9	2.3	Iris-virginica
120	121	6.9	3.2	5.7	2.3	Iris-virginica
145	146	6.7	3.0	5.2	2.3	Iris-virginica
104	105	6.5	3.0	5.8	2.2	Iris-virginica
117	118	7.7	3.8	6.7	2.2	Iris-virginica
112	113	6.8	3.0	5.5	2.1	Iris-virginica
105	106	7.6	3.0	6.6	2.1	Iris-virginica
102	103	7.1	3.0	5.9	2.1	Iris-virginica
139	140	6.9	3.1	5.4	2.1	Iris-virginica
113	114	5.7	2.5	5.0	2.0	Iris-virginica
121	122	5.6	2.8	4.9	2.0	Iris-virginica
131	132	7.9	3.8	6.4	2.0	Iris-virginica
111	112	6.4	2.7	5.3	1.9	Iris-virginica
101	102	5.8	2.7	5.1	1.9	Iris-virginica
70	71	5.9	3.2	4.8	1.8	Iris-versicolor
107	108	7.3	2.9	6.3	1.8	Iris-virginica
103	104	6.3	2.9	5.6	1.8	Iris-virginica
126	127	6.2	2.8	4.8	1.8	Iris-virginica
77	78	6.7	3.0	5.0	1.7	Iris-versicolor
106	107	4.9	2.5	4.5	1.7	Iris-virginica
52	53	6.9	3.1	4.9	1.5	Iris-versicolor

83	84	SepalLengthCm	6.0	2.7	5.1	1.6	Iris-versicolor
Id							Species
61	62		5.9	3.0	4.2	1.5	Iris-versicolor
52	53		6.9	3.1	4.9	1.5	Iris-versicolor
51	52		6.4	3.2	4.5	1.5	Iris-versicolor
54	55		6.5	2.8	4.6	1.5	Iris-versicolor
65	66		6.7	3.1	4.4	1.4	Iris-versicolor
59	60		5.2	2.7	3.9	1.4	Iris-versicolor
50	51		7.0	3.2	4.7	1.4	Iris-versicolor
64	65		5.6	2.9	3.6	1.3	Iris-versicolor
74	75		6.4	2.9	4.3	1.3	Iris-versicolor
53	54		5.5	2.3	4.0	1.3	Iris-versicolor
80	81		5.5	2.4	3.8	1.1	Iris-versicolor
98	99		5.1	2.5	3.0	1.1	Iris-versicolor
67	68		5.8	2.7	4.1	1.0	Iris-versicolor
57	58		4.9	2.4	3.3	1.0	Iris-versicolor
81	82		5.5	2.4	3.7	1.0	Iris-versicolor
60	61		5.0	2.0	3.5	1.0	Iris-versicolor
5	6		5.4	3.9	1.7	0.4	Iris-setosa
0	1		5.1	3.5	1.4	0.2	Iris-setosa
3	4		4.6	3.1	1.5	0.2	Iris-setosa
14	15		5.8	4.0	1.2	0.2	Iris-setosa
24	25		4.8	3.4	1.9	0.2	Iris-setosa
11	12		4.8	3.4	1.6	0.2	Iris-setosa
2	3		4.7	3.2	1.3	0.2	Iris-setosa
22	23		4.6	3.6	1.0	0.2	Iris-setosa
13	14		4.3	3.0	1.1	0.1	Iris-setosa

In [200]:

```
#Sort the DataFrame by 'petal_width' and then 'sepal_width', both in ascending order:
data.sort_values(by=['PetalWidthCm', 'SepalWidthCm'],inplace=True)
data
```

Out[200]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
24	25	4.8	3.4	1.9	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
0	1	5.1	3.5	1.4	0.2	Iris-setosa
22	23	4.6	3.6	1.0	0.2	Iris-setosa
14	15	5.8	4.0	1.2	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
60	61	5.0	2.0	3.5	1.0	Iris-versicolor
57	58	4.9	2.4	3.3	1.0	Iris-versicolor
81	82	5.5	2.4	3.7	1.0	Iris-versicolor
67	68	5.8	2.7	4.1	1.0	Iris-versicolor

80	Id 81	SepalLengthCm 5.5	SepalWidthCm 2.4	PetalLengthCm 3.8	PetalWidthCm 1.1	Species Iris-versicolor
98	99	5.1	2.5	3.0	1.1	Iris-versicolor
53	54	5.5	2.3	4.0	1.3	Iris-versicolor
64	65	5.6	2.9	3.6	1.3	Iris-versicolor
74	75	6.4	2.9	4.3	1.3	Iris-versicolor
59	60	5.2	2.7	3.9	1.4	Iris-versicolor
65	66	6.7	3.1	4.4	1.4	Iris-versicolor
50	51	7.0	3.2	4.7	1.4	Iris-versicolor
54	55	6.5	2.8	4.6	1.5	Iris-versicolor
61	62	5.9	3.0	4.2	1.5	Iris-versicolor
52	53	6.9	3.1	4.9	1.5	Iris-versicolor
51	52	6.4	3.2	4.5	1.5	Iris-versicolor
83	84	6.0	2.7	5.1	1.6	Iris-versicolor
106	107	4.9	2.5	4.5	1.7	Iris-virginica
77	78	6.7	3.0	5.0	1.7	Iris-versicolor
126	127	6.2	2.8	4.8	1.8	Iris-virginica
107	108	7.3	2.9	6.3	1.8	Iris-virginica
103	104	6.3	2.9	5.6	1.8	Iris-virginica
70	71	5.9	3.2	4.8	1.8	Iris-versicolor
111	112	6.4	2.7	5.3	1.9	Iris-virginica
101	102	5.8	2.7	5.1	1.9	Iris-virginica
113	114	5.7	2.5	5.0	2.0	Iris-virginica
121	122	5.6	2.8	4.9	2.0	Iris-virginica
131	132	7.9	3.8	6.4	2.0	Iris-virginica
112	113	6.8	3.0	5.5	2.1	Iris-virginica
105	106	7.6	3.0	6.6	2.1	Iris-virginica
102	103	7.1	3.0	5.9	2.1	Iris-virginica
139	140	6.9	3.1	5.4	2.1	Iris-virginica
104	105	6.5	3.0	5.8	2.2	Iris-virginica
117	118	7.7	3.8	6.7	2.2	Iris-virginica
118	119	7.7	2.6	6.9	2.3	Iris-virginica
145	146	6.7	3.0	5.2	2.3	Iris-virginica
120	121	6.9	3.2	5.7	2.3	Iris-virginica
100	101	6.3	3.3	6.0	2.5	Iris-virginica
109	110	7.2	3.6	6.1	2.5	Iris-virginica

In [153]:

```
# Get the memory usage of the DataFrame:
mem_usage = data.memory_usage()
print("Memory usage of each column:")
print(mem_usage)
```

Memory usage of each column:

```
Index      1200
Id          1200
SepalLengthCm  1200
SepalWidthCm  1200
PetalLengthCm  1200
PetalWidthCm  1200
Species     1200
```

```
Species          1200  
dtype: int64
```

In [154]:

```
# Total memory usage  
total_mem_usage = data.memory_usage().sum()  
print(f"\nTotal memory usage: {total_mem_usage} bytes")
```

Total memory usage: 8400 bytes

In [155]:

```
# Deep memory usage of each column  
deep_mem_usage = data.memory_usage(deep=True)  
print("\nDeep memory usage of each column:")  
print(deep_mem_usage)
```

Deep memory usage of each column:

```
Index          1200  
Id             1200  
SepalLengthCm  1200  
SepalWidthCm   1200  
PetalLengthCm  1200  
PetalWidthCm   1200  
Species        10550  
dtype: int64
```

In [156]:

```
# Total deep memory usage  
total_deep_mem_usage = deep_mem_usage.sum()  
print(f"\nTotal deep memory usage: {total_deep_mem_usage} bytes")
```

Total deep memory usage: 17750 bytes

In [157]:

```
#Grouping and Aggregation:  
#Group by 'species' and calculate the mean 'sepal_length' for each species:  
data.groupby('Species')['SepalLengthCm'].mean()
```

Out[157]:

```
Species  
Iris-setosa      5.006  
Iris-versicolor  5.936  
Iris-virginica   6.588  
Name: SepalLengthCm, dtype: float64
```

In [158]:

```
#Group by 'species' and calculate the standard deviation of 'sepal_width' for each species  
:  
data.groupby('Species')['SepalWidthCm'].std()
```

Out[158]:

```
Species  
Iris-setosa      0.381024  
Iris-versicolor  0.313798  
Iris-virginica   0.322497  
Name: SepalWidthCm, dtype: float64
```

In [159]:

```
#Group by 'species' and calculate the minimum, maximum, and mean of 'petal_length':  
data.groupby('Species')['PetalLengthCm'].agg(['min', 'max', 'mean'])
```

Out[159]:

**min max mean**

Species	min	max	mean
Iris-setosa	1.0	1.9	1.464
Iris-versicolor	3.0	5.1	4.260
Iris-virginica	4.5	6.9	5.552

In [160]:

```
#Calculate the total count of each species:
data['Species'].value_counts()
```

Out[160]:

Species
Iris-setosa 50
Iris-versicolor 50
Iris-virginica 50
Name: count, dtype: int64

In [161]:

```
#Calculate the percentage of each species in the dataset:
data['Species'].value_counts(normalize=True)
```

Out[161]:

Species
Iris-setosa 0.333333
Iris-versicolor 0.333333
Iris-virginica 0.333333
Name: proportion, dtype: float64

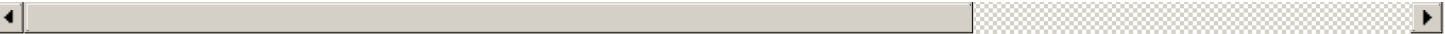
In [2]:

```
#Merging and Joining:
#Import another dataset and merge the two DataFrames based on a common column:
import pandas as pd
data2 = pd.read_csv('/kaggle/input/iris/Iris.csv')
merged_data = pd.merge(data, data2, on='Id')
merged_data
```

Out[2]:

	Id	SepalLengthCm_x	SepalWidthCm_x	PetalLengthCm_x	PetalWidthCm_x	Species_x	SepalLengthCm_y	SepalWidthCr
0	1	5.1	3.5	1.4	0.2	Iris-setosa	5.1	
1	2	4.9	3.0	1.4	0.2	Iris-setosa	4.9	
2	3	4.7	3.2	1.3	0.2	Iris-setosa	4.7	
3	4	4.6	3.1	1.5	0.2	Iris-setosa	4.6	
4	5	5.0	3.6	1.4	0.2	Iris-setosa	5.0	
...	...	...	...	...	...	...	...	
145	146	6.7	3.0	5.2	2.3	Iris-virginica	6.7	
146	147	6.3	2.5	5.0	1.9	Iris-virginica	6.3	
147	148	6.5	3.0	5.2	2.0	Iris-virginica	6.5	
148	149	6.2	3.4	5.4	2.3	Iris-virginica	6.2	
149	150	5.9	3.0	5.1	1.8	Iris-virginica	5.9	

Id SepalLengthCm\_x SepalWidthCm\_x PetalLengthCm\_x PetalWidthCm\_x Species\_x SepalLengthCm\_y SepalWidthCm\_y  
150 rows x 11 columns



In [201]:

```
#Perform an inner join on two DataFrames using a specific column as the key:
inner_join_data = pd.merge(data, data2, on='Species', how='inner')
inner_join_data
```

Out[201]:

	Id_x	SepalLengthCm_x	SepalWidthCm_x	PetalLengthCm_x	PetalWidthCm_x	Species	Id_y	SepalLengthCm_y	SepalWidthCm_y
0	14	4.3	3.0	1.1	0.1	Iris-setosa	1	5.1	4.9
1	14	4.3	3.0	1.1	0.1	Iris-setosa	2	4.9	4.7
2	14	4.3	3.0	1.1	0.1	Iris-setosa	3	4.7	4.6
3	14	4.3	3.0	1.1	0.1	Iris-setosa	4	4.6	4.5
4	14	4.3	3.0	1.1	0.1	Iris-setosa	5	5.0	4.4
...	...	...	...	...	...	...	...	...	...
2395	110	7.2	3.6	6.1	2.5	Iris-virginica	146	6.7	3.4
2396	110	7.2	3.6	6.1	2.5	Iris-virginica	147	6.3	3.5
2397	110	7.2	3.6	6.1	2.5	Iris-virginica	148	6.5	3.9
2398	110	7.2	3.6	6.1	2.5	Iris-virginica	149	6.2	3.6
2399	110	7.2	3.6	6.1	2.5	Iris-virginica	150	5.9	3.8

2400 rows x 11 columns



In [202]:

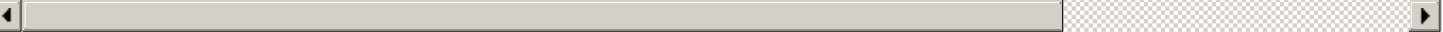
```
#Perform a left join on two DataFrames using a specific column as the key:
left_join_data = pd.merge(data, data2, on='PetalLengthCm', how='left')
left_join_data
```

Out[202]:

	Id_x	SepalLengthCm_x	SepalWidthCm_x	PetalLengthCm	PetalWidthCm_x	Species_x	Id_y	SepalLengthCm_y	SepalWidthCm_y
0	14	4.3	3.0	1.1	0.1	Iris-setosa	14	4.3	3.0
1	4	4.6	3.1	1.5	0.2	Iris-setosa	4	4.6	3.1
2	4	4.6	3.1	1.5	0.2	Iris-setosa	8	5.0	3.4
3	4	4.6	3.1	1.5	0.2	Iris-setosa	10	4.9	3.6
4	4	4.6	3.1	1.5	0.2	Iris-setosa	11	5.4	3.9
...	...	...	...	...	...	...	...	...	...
174	101	6.3	3.3	6.0	2.5	Iris-virginica	101	6.3	3.3

175	Id_x	SepalLengthCm_x	SepalWidthCm_x	PetalLengthCm_x	PetalWidthCm_x	Species_x	Id_y	SepalLengthCm_y	SepalWidthCm_y
	101	6.3	3.3	6.0	2.5	Iris-virginica	126	7.2	3.3
176	110	7.2	3.6	6.1	2.5	Iris-virginica	110	7.2	3.6
177	110	7.2	3.6	6.1	2.5	Iris-virginica	131	7.4	3.6
178	110	7.2	3.6	6.1	2.5	Iris-virginica	136	7.7	3.6

179 rows x 11 columns



In [203]:

```
# Perform a right join on two DataFrames using a specific column as the key:
right_join_data = pd.merge(data, data2, on='SepalLengthCm', how='right')
right_join_data
```

Out[203]:

	Id_x	SepalLengthCm	SepalWidthCm_x	PetalLengthCm_x	PetalWidthCm_x	Species_x	Id_y	SepalWidthCm_y	PetalLengthCm_y
0	1.0	5.1	3.5	1.4	0.2	Iris-setosa	1	3.5	1.4
1	99.0	5.1	2.5	3.0	1.1	Iris-versicolor	1	3.5	3.0
2	58.0	4.9	2.4	3.3	1.0	Iris-versicolor	2	3.0	3.3
3	107.0	4.9	2.5	4.5	1.7	Iris-virginica	2	3.0	4.5
4	3.0	4.7	3.2	1.3	0.2	Iris-setosa	3	3.2	1.3
...	...	...	...	...	...	...	...	...	...
262	55.0	6.5	2.8	4.6	1.5	Iris-versicolor	148	3.0	4.6
263	105.0	6.5	3.0	5.8	2.2	Iris-virginica	148	3.0	5.8
264	127.0	6.2	2.8	4.8	1.8	Iris-virginica	149	3.4	4.8
265	62.0	5.9	3.0	4.2	1.5	Iris-versicolor	150	3.0	4.2
266	71.0	5.9	3.2	4.8	1.8	Iris-versicolor	150	3.0	4.8

267 rows x 11 columns



In [204]:

```
#Pivoting:
#Create a pivot table with 'species' as the index, 'petal_length' as columns, and 'sepal_width' as values:
pivot_df = pd.pivot_table(data, index='Species', columns='PetalLengthCm', values='SepalWidthCm')
pivot_df
```

Out[204]:

PetalLengthCm	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.9	3.0	...	5.7	5.8	5.9	6.0	6.1	6.3	6.4	6.6
Species																			
Iris-setosa	3.6	3.0	4.0	3.2	3.5	3.1	3.4	3.9	3.4	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Iris-versicolor	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.5	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN



PetalLengthCm	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.9	3.0	...	5.7	5.8	5.9	6.0	6.1	6.3	6.4	6.6
Iris-virginica	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	3.2	3.0	3.0	3.3	3.6	2.9	3.8	3.0
Species																			

3 rows x 43 columns



In [205]:

```
#Create a pivot table with 'species' as the index and the mean of 'petal_width' as values
:
pivot_df = pd.pivot_table(data, index='Species', values='PetalWidthCm', aggfunc='mean')
pivot_df
```

Out[205]:

PetalWidthCm	
Species	
Iris-setosa	0.211111
Iris-versicolor	1.336842
Iris-virginica	2.080000

In [169]:

```
#String Manipulation:
#Extract the first character of the 'species' column:
data['Species'].str[0]
```

Out[169]:

```
12      I
13      I
34      I
9       I
37      I
..
140     I
136     I
144     I
100     I
109     I
Name: Species, Length: 150, dtype: object
```

In [206]:

```
#Convert all values in the 'species' column to lowercase:
data['Species'] = data['Species'].str.lower()
data['Species']
```

Out[206]:

```
13      iris-setosa
3       iris-setosa
2       iris-setosa
24      iris-setosa
11      iris-setosa
0       iris-setosa
22      iris-setosa
14      iris-setosa
5       iris-setosa
60      iris-versicolor
57      iris-versicolor
81      iris-versicolor
67      iris-versicolor
80      iris-versicolor
98      iris-versicolor
53      iris-versicolor
64      iris-versicolor
74      iris-versicolor
59      iris-versicolor
```

```
65      iris-versicolor
50      iris-versicolor
54      iris-versicolor
61      iris-versicolor
52      iris-versicolor
51      iris-versicolor
83      iris-versicolor
106     iris-virginica
77      iris-versicolor
126     iris-virginica
107     iris-virginica
103     iris-virginica
70      iris-versicolor
111     iris-virginica
101     iris-virginica
113     iris-virginica
121     iris-virginica
131     iris-virginica
112     iris-virginica
105     iris-virginica
102     iris-virginica
139     iris-virginica
104     iris-virginica
117     iris-virginica
118     iris-virginica
145     iris-virginica
120     iris-virginica
100     iris-virginica
109     iris-virginica
Name: Species, dtype: object
```

In [171]:

```
#Count the number of characters in each 'species' column value:
data['Species'].str.len()
```

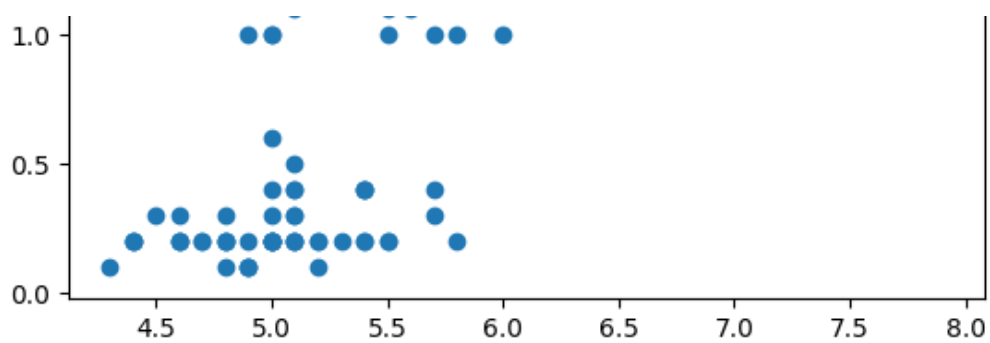
Out[171]:

```
12      11
13      11
34      11
9        11
37      11
..
140     14
136     14
144     14
100     14
109     14
Name: Species, Length: 150, dtype: int64
```

In [172]:

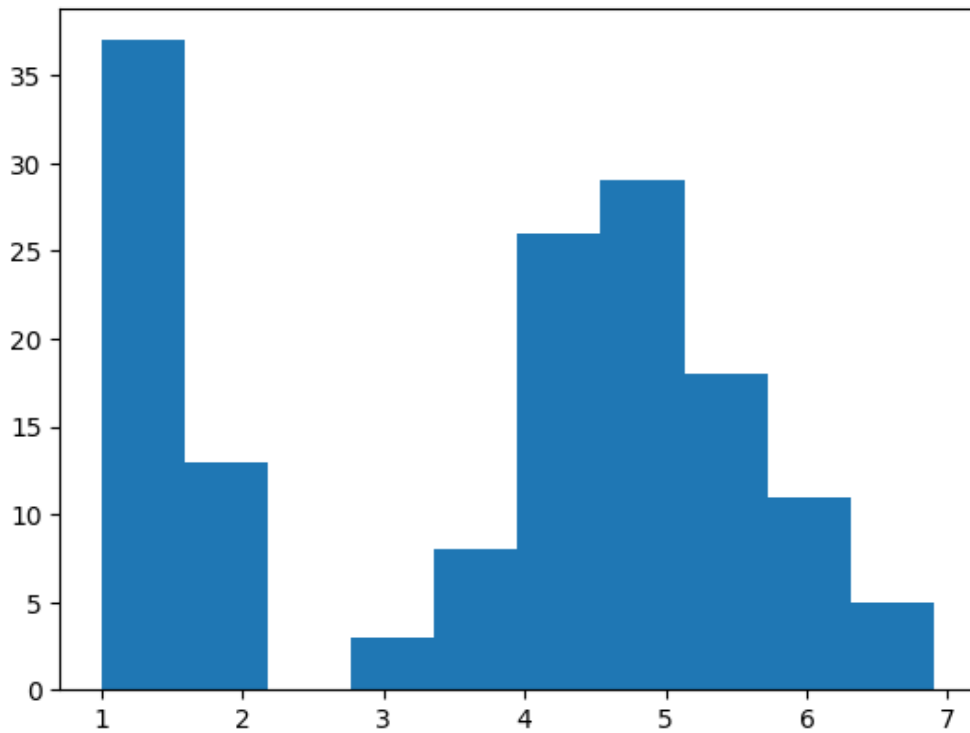
```
#Data Visualization (using matplotlib and seaborn):
#Create a scatter plot of 'sepal_length' versus 'petal_width':
import matplotlib.pyplot as plt
plt.scatter(data['SepalLengthCm'], data['PetalWidthCm'])
plt.show()
```





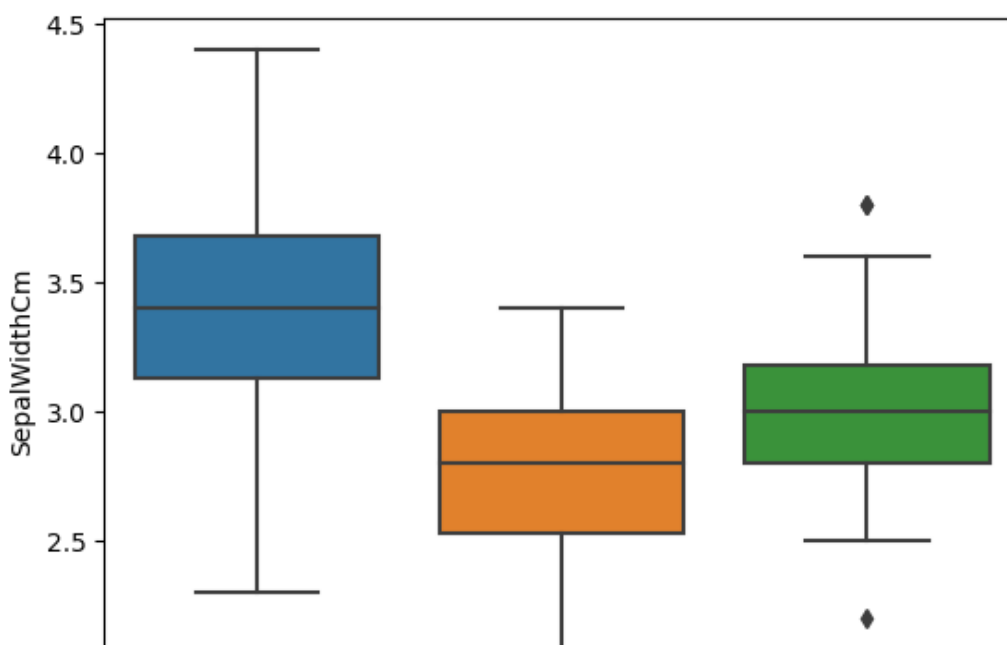
In [173]:

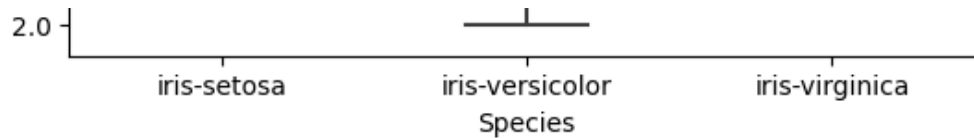
```
#Create a histogram of the 'petal_length' column:
plt.hist(data['PetalLengthCm'])
plt.show()
```



In [174]:

```
#Create a box plot of the 'sepal_width' column for each species:
import seaborn as sns
sns.boxplot(x='Species', y='SepalWidthCm', data=data)
plt.show()
```





In [207]:

```
#Missing Data Handling:
#Remove rows with missing values in any column:
data.dropna(inplace=True)
data
```

Out[207]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	iris-setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa
22	23	4.6	3.6	1.0	0.2	iris-setosa
14	15	5.8	4.0	1.2	0.2	iris-setosa
5	6	5.4	3.9	1.7	0.4	iris-setosa
60	61	5.0	2.0	3.5	1.0	iris-versicolor
57	58	4.9	2.4	3.3	1.0	iris-versicolor
81	82	5.5	2.4	3.7	1.0	iris-versicolor
67	68	5.8	2.7	4.1	1.0	iris-versicolor
80	81	5.5	2.4	3.8	1.1	iris-versicolor
98	99	5.1	2.5	3.0	1.1	iris-versicolor
53	54	5.5	2.3	4.0	1.3	iris-versicolor
64	65	5.6	2.9	3.6	1.3	iris-versicolor
74	75	6.4	2.9	4.3	1.3	iris-versicolor
59	60	5.2	2.7	3.9	1.4	iris-versicolor
65	66	6.7	3.1	4.4	1.4	iris-versicolor
50	51	7.0	3.2	4.7	1.4	iris-versicolor
54	55	6.5	2.8	4.6	1.5	iris-versicolor
61	62	5.9	3.0	4.2	1.5	iris-versicolor
52	53	6.9	3.1	4.9	1.5	iris-versicolor
51	52	6.4	3.2	4.5	1.5	iris-versicolor
83	84	6.0	2.7	5.1	1.6	iris-versicolor
106	107	4.9	2.5	4.5	1.7	iris-virginica
77	78	6.7	3.0	5.0	1.7	iris-versicolor
126	127	6.2	2.8	4.8	1.8	iris-virginica
107	108	7.3	2.9	6.3	1.8	iris-virginica
103	104	6.3	2.9	5.6	1.8	iris-virginica
70	71	5.9	3.2	4.8	1.8	iris-versicolor
111	112	6.4	2.7	5.3	1.9	iris-virginica
101	102	5.8	2.7	5.1	1.9	iris-virginica
113	114	5.7	2.5	5.0	2.0	iris-virginica

		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
121	122	5.6	2.8	4.9	2.0	iris-virginica
131	132	7.9	3.8	6.4	2.0	iris-virginica
112	113	6.8	3.0	5.5	2.1	iris-virginica
105	106	7.6	3.0	6.6	2.1	iris-virginica
102	103	7.1	3.0	5.9	2.1	iris-virginica
139	140	6.9	3.1	5.4	2.1	iris-virginica
104	105	6.5	3.0	5.8	2.2	iris-virginica
117	118	7.7	3.8	6.7	2.2	iris-virginica
118	119	7.7	2.6	6.9	2.3	iris-virginica
145	146	6.7	3.0	5.2	2.3	iris-virginica
120	121	6.9	3.2	5.7	2.3	iris-virginica
100	101	6.3	3.3	6.0	2.5	iris-virginica
109	110	7.2	3.6	6.1	2.5	iris-virginica

1. When `inplace=True` is used, the original `DataFrame` is modified directly. If you want to keep the original `DataFrame` unchanged and create a new `DataFrame` without the missing values, you can omit `inplace=True` and assign the result to a new variable: `df_cleaned = df.dropna()`

2.Drop Specific Rows or Columns: `df.dropna(axis=1, inplace=True)`

3.Use the `thresh` parameter to specify a minimum number of non-null values required to keep a row or column: Keeps rows with at least 4 non-null values `df.dropna(thresh=4, inplace=True)`

In [208]:

```
# Replace missing values with the column mean:
data.fillna(data.select_dtypes(include=[float, int]).mean(), inplace=True)
data
```

Out[208]:

		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	iris-setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa
22	23	4.6	3.6	1.0	0.2	iris-setosa
14	15	5.8	4.0	1.2	0.2	iris-setosa
5	6	5.4	3.9	1.7	0.4	iris-setosa
60	61	5.0	2.0	3.5	1.0	iris-versicolor
57	58	4.9	2.4	3.3	1.0	iris-versicolor
81	82	5.5	2.4	3.7	1.0	iris-versicolor
67	68	5.8	2.7	4.1	1.0	iris-versicolor
80	81	5.5	2.4	3.8	1.1	iris-versicolor
98	99	5.1	2.5	3.0	1.1	iris-versicolor
53	54	5.5	2.3	4.0	1.3	iris-versicolor
64	65	5.6	2.9	3.6	1.3	iris-versicolor
74	75	6.4	2.9	4.3	1.3	iris-versicolor

59	60	5.2	2.7	3.9	1.4	iris-versicolor
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	
65	66	6.7	3.1	4.4	1.4	iris-versicolor
50	51	7.0	3.2	4.7	1.4	iris-versicolor
54	55	6.5	2.8	4.6	1.5	iris-versicolor
61	62	5.9	3.0	4.2	1.5	iris-versicolor
52	53	6.9	3.1	4.9	1.5	iris-versicolor
51	52	6.4	3.2	4.5	1.5	iris-versicolor
83	84	6.0	2.7	5.1	1.6	iris-versicolor
106	107	4.9	2.5	4.5	1.7	iris-virginica
77	78	6.7	3.0	5.0	1.7	iris-versicolor
126	127	6.2	2.8	4.8	1.8	iris-virginica
107	108	7.3	2.9	6.3	1.8	iris-virginica
103	104	6.3	2.9	5.6	1.8	iris-virginica
70	71	5.9	3.2	4.8	1.8	iris-versicolor
111	112	6.4	2.7	5.3	1.9	iris-virginica
101	102	5.8	2.7	5.1	1.9	iris-virginica
113	114	5.7	2.5	5.0	2.0	iris-virginica
121	122	5.6	2.8	4.9	2.0	iris-virginica
131	132	7.9	3.8	6.4	2.0	iris-virginica
112	113	6.8	3.0	5.5	2.1	iris-virginica
105	106	7.6	3.0	6.6	2.1	iris-virginica
102	103	7.1	3.0	5.9	2.1	iris-virginica
139	140	6.9	3.1	5.4	2.1	iris-virginica
104	105	6.5	3.0	5.8	2.2	iris-virginica
117	118	7.7	3.8	6.7	2.2	iris-virginica
118	119	7.7	2.6	6.9	2.3	iris-virginica
145	146	6.7	3.0	5.2	2.3	iris-virginica
120	121	6.9	3.2	5.7	2.3	iris-virginica
100	101	6.3	3.3	6.0	2.5	iris-virginica
109	110	7.2	3.6	6.1	2.5	iris-virginica

In [209]:

```
#Replace missing values with the column median:
data.fillna(data.select_dtypes(include=[float,int]).median(), inplace=True)
data
```

Out[209]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	iris-setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa
22	23	4.6	3.6	1.0	0.2	iris-setosa
14	15	5.8	4.0	1.2	0.2	iris-setosa
5	6	5.4	3.9	1.7	0.4	iris-setosa

60	Id 61	SepalLengthCm 5.0	SepalWidthCm 2.0	PetalLengthCm 3.5	PetalWidthCm 1.0	Species iris-versicolor
57	58	4.9	2.4	3.3	1.0	iris-versicolor
81	82	5.5	2.4	3.7	1.0	iris-versicolor
67	68	5.8	2.7	4.1	1.0	iris-versicolor
80	81	5.5	2.4	3.8	1.1	iris-versicolor
98	99	5.1	2.5	3.0	1.1	iris-versicolor
53	54	5.5	2.3	4.0	1.3	iris-versicolor
64	65	5.6	2.9	3.6	1.3	iris-versicolor
74	75	6.4	2.9	4.3	1.3	iris-versicolor
59	60	5.2	2.7	3.9	1.4	iris-versicolor
65	66	6.7	3.1	4.4	1.4	iris-versicolor
50	51	7.0	3.2	4.7	1.4	iris-versicolor
54	55	6.5	2.8	4.6	1.5	iris-versicolor
61	62	5.9	3.0	4.2	1.5	iris-versicolor
52	53	6.9	3.1	4.9	1.5	iris-versicolor
51	52	6.4	3.2	4.5	1.5	iris-versicolor
83	84	6.0	2.7	5.1	1.6	iris-versicolor
106	107	4.9	2.5	4.5	1.7	iris-virginica
77	78	6.7	3.0	5.0	1.7	iris-versicolor
126	127	6.2	2.8	4.8	1.8	iris-virginica
107	108	7.3	2.9	6.3	1.8	iris-virginica
103	104	6.3	2.9	5.6	1.8	iris-virginica
70	71	5.9	3.2	4.8	1.8	iris-versicolor
111	112	6.4	2.7	5.3	1.9	iris-virginica
101	102	5.8	2.7	5.1	1.9	iris-virginica
113	114	5.7	2.5	5.0	2.0	iris-virginica
121	122	5.6	2.8	4.9	2.0	iris-virginica
131	132	7.9	3.8	6.4	2.0	iris-virginica
112	113	6.8	3.0	5.5	2.1	iris-virginica
105	106	7.6	3.0	6.6	2.1	iris-virginica
102	103	7.1	3.0	5.9	2.1	iris-virginica
139	140	6.9	3.1	5.4	2.1	iris-virginica
104	105	6.5	3.0	5.8	2.2	iris-virginica
117	118	7.7	3.8	6.7	2.2	iris-virginica
118	119	7.7	2.6	6.9	2.3	iris-virginica
145	146	6.7	3.0	5.2	2.3	iris-virginica
120	121	6.9	3.2	5.7	2.3	iris-virginica
100	101	6.3	3.3	6.0	2.5	iris-virginica
109	110	7.2	3.6	6.1	2.5	iris-virginica

In [210]:

```
#Assign a specific value to missing values in a column:
data['SepalWidthCm'].fillna('abc')
data
```

Out[210]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	iris-setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa
22	23	4.6	3.6	1.0	0.2	iris-setosa
14	15	5.8	4.0	1.2	0.2	iris-setosa
5	6	5.4	3.9	1.7	0.4	iris-setosa
60	61	5.0	2.0	3.5	1.0	iris-versicolor
57	58	4.9	2.4	3.3	1.0	iris-versicolor
81	82	5.5	2.4	3.7	1.0	iris-versicolor
67	68	5.8	2.7	4.1	1.0	iris-versicolor
80	81	5.5	2.4	3.8	1.1	iris-versicolor
98	99	5.1	2.5	3.0	1.1	iris-versicolor
53	54	5.5	2.3	4.0	1.3	iris-versicolor
64	65	5.6	2.9	3.6	1.3	iris-versicolor
74	75	6.4	2.9	4.3	1.3	iris-versicolor
59	60	5.2	2.7	3.9	1.4	iris-versicolor
65	66	6.7	3.1	4.4	1.4	iris-versicolor
50	51	7.0	3.2	4.7	1.4	iris-versicolor
54	55	6.5	2.8	4.6	1.5	iris-versicolor
61	62	5.9	3.0	4.2	1.5	iris-versicolor
52	53	6.9	3.1	4.9	1.5	iris-versicolor
51	52	6.4	3.2	4.5	1.5	iris-versicolor
83	84	6.0	2.7	5.1	1.6	iris-versicolor
106	107	4.9	2.5	4.5	1.7	iris-virginica
77	78	6.7	3.0	5.0	1.7	iris-versicolor
126	127	6.2	2.8	4.8	1.8	iris-virginica
107	108	7.3	2.9	6.3	1.8	iris-virginica
103	104	6.3	2.9	5.6	1.8	iris-virginica
70	71	5.9	3.2	4.8	1.8	iris-versicolor
111	112	6.4	2.7	5.3	1.9	iris-virginica
101	102	5.8	2.7	5.1	1.9	iris-virginica
113	114	5.7	2.5	5.0	2.0	iris-virginica
121	122	5.6	2.8	4.9	2.0	iris-virginica
131	132	7.9	3.8	6.4	2.0	iris-virginica
112	113	6.8	3.0	5.5	2.1	iris-virginica
105	106	7.6	3.0	6.6	2.1	iris-virginica
102	103	7.1	3.0	5.9	2.1	iris-virginica
139	140	6.9	3.1	5.4	2.1	iris-virginica
104	105	6.5	3.0	5.8	2.2	iris-virginica
117	118	7.7	3.8	6.7	2.2	iris-virginica
118	119	7.7	2.6	6.9	2.3	iris-virginica



145	146	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
120	121	6.9	3.2	5.7	2.3	iris-virginica
100	101	6.3	3.3	6.0	2.5	iris-virginica
109	110	7.2	3.6	6.1	2.5	iris-virginica

In [211]:

```
#Outlier Detection and Handling:
#Identify and remove outliers in the 'sepal_length' column:
Q1 = data['SepalLengthCm'].quantile(0.25)
Q3 = data['SepalLengthCm'].quantile(0.75)
IQR = Q3- Q1
data = data[~((data['SepalLengthCm'] < (Q1- 1.5 * IQR)) | (data['SepalLengthCm'] > (Q3 + 1.5 * IQR)))]
data
```

Out[211]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	iris-setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa
22	23	4.6	3.6	1.0	0.2	iris-setosa
14	15	5.8	4.0	1.2	0.2	iris-setosa
5	6	5.4	3.9	1.7	0.4	iris-setosa
60	61	5.0	2.0	3.5	1.0	iris-versicolor
57	58	4.9	2.4	3.3	1.0	iris-versicolor
81	82	5.5	2.4	3.7	1.0	iris-versicolor
67	68	5.8	2.7	4.1	1.0	iris-versicolor
80	81	5.5	2.4	3.8	1.1	iris-versicolor
98	99	5.1	2.5	3.0	1.1	iris-versicolor
53	54	5.5	2.3	4.0	1.3	iris-versicolor
64	65	5.6	2.9	3.6	1.3	iris-versicolor
74	75	6.4	2.9	4.3	1.3	iris-versicolor
59	60	5.2	2.7	3.9	1.4	iris-versicolor
65	66	6.7	3.1	4.4	1.4	iris-versicolor
50	51	7.0	3.2	4.7	1.4	iris-versicolor
54	55	6.5	2.8	4.6	1.5	iris-versicolor
61	62	5.9	3.0	4.2	1.5	iris-versicolor
52	53	6.9	3.1	4.9	1.5	iris-versicolor
51	52	6.4	3.2	4.5	1.5	iris-versicolor
83	84	6.0	2.7	5.1	1.6	iris-versicolor
106	107	4.9	2.5	4.5	1.7	iris-virginica
77	78	6.7	3.0	5.0	1.7	iris-versicolor
126	127	6.2	2.8	4.8	1.8	iris-virginica
107	108	7.3	2.9	6.3	1.8	iris-virginica
103	104	6.3	2.9	5.6	1.8	iris-virginica

70	71	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
111	112	6.4	2.7	5.3	1.9	iris-virginica
101	102	5.8	2.7	5.1	1.9	iris-virginica
113	114	5.7	2.5	5.0	2.0	iris-virginica
121	122	5.6	2.8	4.9	2.0	iris-virginica
131	132	7.9	3.8	6.4	2.0	iris-virginica
112	113	6.8	3.0	5.5	2.1	iris-virginica
105	106	7.6	3.0	6.6	2.1	iris-virginica
102	103	7.1	3.0	5.9	2.1	iris-virginica
139	140	6.9	3.1	5.4	2.1	iris-virginica
104	105	6.5	3.0	5.8	2.2	iris-virginica
117	118	7.7	3.8	6.7	2.2	iris-virginica
118	119	7.7	2.6	6.9	2.3	iris-virginica
145	146	6.7	3.0	5.2	2.3	iris-virginica
120	121	6.9	3.2	5.7	2.3	iris-virginica
100	101	6.3	3.3	6.0	2.5	iris-virginica
109	110	7.2	3.6	6.1	2.5	iris-virginica

In [212]:

```
#Identify and replace outliers in the 'petal_width' column with a threshold value:
threshold = 0.2
data.loc[data['PetalWidthCm'] > threshold, 'PetalWidthCm'] = threshold
data
```

Out[212]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
13	14	4.3	3.0	1.1	0.1	iris-setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa
22	23	4.6	3.6	1.0	0.2	iris-setosa
14	15	5.8	4.0	1.2	0.2	iris-setosa
5	6	5.4	3.9	1.7	0.2	iris-setosa
60	61	5.0	2.0	3.5	0.2	iris-versicolor
57	58	4.9	2.4	3.3	0.2	iris-versicolor
81	82	5.5	2.4	3.7	0.2	iris-versicolor
67	68	5.8	2.7	4.1	0.2	iris-versicolor
80	81	5.5	2.4	3.8	0.2	iris-versicolor
98	99	5.1	2.5	3.0	0.2	iris-versicolor
53	54	5.5	2.3	4.0	0.2	iris-versicolor
64	65	5.6	2.9	3.6	0.2	iris-versicolor
74	75	6.4	2.9	4.3	0.2	iris-versicolor
59	60	5.2	2.7	3.9	0.2	iris-versicolor
65	66	6.7	3.1	4.4	0.2	iris-versicolor
50	51	7.0	3.2	4.7	0.2	iris-versicolor

50	51	7.0	3.2	4.7	0.2	iris-versicolor
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	
54	55	6.5	2.8	4.6	0.2	iris-versicolor
61	62	5.9	3.0	4.2	0.2	iris-versicolor
52	53	6.9	3.1	4.9	0.2	iris-versicolor
51	52	6.4	3.2	4.5	0.2	iris-versicolor
83	84	6.0	2.7	5.1	0.2	iris-versicolor
106	107	4.9	2.5	4.5	0.2	iris-virginica
77	78	6.7	3.0	5.0	0.2	iris-versicolor
126	127	6.2	2.8	4.8	0.2	iris-virginica
107	108	7.3	2.9	6.3	0.2	iris-virginica
103	104	6.3	2.9	5.6	0.2	iris-virginica
70	71	5.9	3.2	4.8	0.2	iris-versicolor
111	112	6.4	2.7	5.3	0.2	iris-virginica
101	102	5.8	2.7	5.1	0.2	iris-virginica
113	114	5.7	2.5	5.0	0.2	iris-virginica
121	122	5.6	2.8	4.9	0.2	iris-virginica
131	132	7.9	3.8	6.4	0.2	iris-virginica
112	113	6.8	3.0	5.5	0.2	iris-virginica
105	106	7.6	3.0	6.6	0.2	iris-virginica
102	103	7.1	3.0	5.9	0.2	iris-virginica
139	140	6.9	3.1	5.4	0.2	iris-virginica
104	105	6.5	3.0	5.8	0.2	iris-virginica
117	118	7.7	3.8	6.7	0.2	iris-virginica
118	119	7.7	2.6	6.9	0.2	iris-virginica
145	146	6.7	3.0	5.2	0.2	iris-virginica
120	121	6.9	3.2	5.7	0.2	iris-virginica
100	101	6.3	3.3	6.0	0.2	iris-virginica
109	110	7.2	3.6	6.1	0.2	iris-virginica

In [181]:

```
#Data Transformation:
#Create a new column that is the sum of 'sepal_length' and 'petal_length':
data['sum_length'] = data['SepalLengthCm'] + data['PetalLengthCm']
print(data['sum_length'])
```

```
12      6.2
13      5.4
34      6.4
9       6.4
37      6.4
...
140     12.3
136     11.9
144     12.4
100     12.3
109     13.3
Name: sum_length, Length: 150, dtype: float64
```

In [182]:

```
#Apply the log transformation to the 'sepal_width' column:
data['SepalWidthCm'] = np.log(data['SepalWidthCm'])
print(data['SepalWidthCm'])
```

```
12      1.098612
```

```
13      1.098612
34      1.131402
9       1.131402
37      1.131402
...
140     1.131402
136     1.223775
144     1.193922
100     1.193922
109     1.280934
Name: SepalWidthCm, Length: 150, dtype: float64
```

In [183]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 150 entries, 12 to 109
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   Id               150 non-null   int64
1   SepalLengthCm    150 non-null   float64
2   SepalWidthCm     150 non-null   float64
3   PetalLengthCm    150 non-null   float64
4   PetalWidthCm     150 non-null   float64
5   Species          150 non-null   object
6   sum_length       150 non-null   float64
dtypes: float64(5), int64(1), object(1)
memory usage: 9.4+ KB
```

In [184]:

```
#String Manipulation (continued):
#Convert a column containing string representations of numbers to numeric values:
data['SepalLengthCm'] = pd.to_numeric(data['SepalLengthCm'])
data['SepalLengthCm']
```

Out[184]:

```
12      4.8
13      4.3
34      4.9
9       4.9
37      4.9
...
140     6.7
136     6.3
144     6.7
100     6.3
109     7.2
Name: SepalLengthCm, Length: 150, dtype: float64
```

In [213]:

```
# Split a column containing multiple space-separated values into multiple columns:
data[['Dataset_Name','Species_Name']] = data['Species'].str.split('-', expand=True)
data
```

Out[213]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
13	14	4.3	3.0	1.1	0.1	iris-setosa	iris	setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa	iris	setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa	iris	setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa	iris	setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa	iris	setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa	iris	setosa

22	21	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
14	15	5.8	4.0	1.2	0.2	iris-setosa	iris	setosa
5	6	5.4	3.9	1.7	0.2	iris-setosa	iris	setosa
60	61	5.0	2.0	3.5	0.2	iris-versicolor	iris	versicolor
57	58	4.9	2.4	3.3	0.2	iris-versicolor	iris	versicolor
81	82	5.5	2.4	3.7	0.2	iris-versicolor	iris	versicolor
67	68	5.8	2.7	4.1	0.2	iris-versicolor	iris	versicolor
80	81	5.5	2.4	3.8	0.2	iris-versicolor	iris	versicolor
98	99	5.1	2.5	3.0	0.2	iris-versicolor	iris	versicolor
53	54	5.5	2.3	4.0	0.2	iris-versicolor	iris	versicolor
64	65	5.6	2.9	3.6	0.2	iris-versicolor	iris	versicolor
74	75	6.4	2.9	4.3	0.2	iris-versicolor	iris	versicolor
59	60	5.2	2.7	3.9	0.2	iris-versicolor	iris	versicolor
65	66	6.7	3.1	4.4	0.2	iris-versicolor	iris	versicolor
50	51	7.0	3.2	4.7	0.2	iris-versicolor	iris	versicolor
54	55	6.5	2.8	4.6	0.2	iris-versicolor	iris	versicolor
61	62	5.9	3.0	4.2	0.2	iris-versicolor	iris	versicolor
52	53	6.9	3.1	4.9	0.2	iris-versicolor	iris	versicolor
51	52	6.4	3.2	4.5	0.2	iris-versicolor	iris	versicolor
83	84	6.0	2.7	5.1	0.2	iris-versicolor	iris	versicolor
106	107	4.9	2.5	4.5	0.2	iris-virginica	iris	virginica
77	78	6.7	3.0	5.0	0.2	iris-versicolor	iris	versicolor
126	127	6.2	2.8	4.8	0.2	iris-virginica	iris	virginica
107	108	7.3	2.9	6.3	0.2	iris-virginica	iris	virginica
103	104	6.3	2.9	5.6	0.2	iris-virginica	iris	virginica
70	71	5.9	3.2	4.8	0.2	iris-versicolor	iris	versicolor
111	112	6.4	2.7	5.3	0.2	iris-virginica	iris	virginica
101	102	5.8	2.7	5.1	0.2	iris-virginica	iris	virginica
113	114	5.7	2.5	5.0	0.2	iris-virginica	iris	virginica
121	122	5.6	2.8	4.9	0.2	iris-virginica	iris	virginica
131	132	7.9	3.8	6.4	0.2	iris-virginica	iris	virginica
112	113	6.8	3.0	5.5	0.2	iris-virginica	iris	virginica
105	106	7.6	3.0	6.6	0.2	iris-virginica	iris	virginica
102	103	7.1	3.0	5.9	0.2	iris-virginica	iris	virginica
139	140	6.9	3.1	5.4	0.2	iris-virginica	iris	virginica
104	105	6.5	3.0	5.8	0.2	iris-virginica	iris	virginica
117	118	7.7	3.8	6.7	0.2	iris-virginica	iris	virginica
118	119	7.7	2.6	6.9	0.2	iris-virginica	iris	virginica
145	146	6.7	3.0	5.2	0.2	iris-virginica	iris	virginica
120	121	6.9	3.2	5.7	0.2	iris-virginica	iris	virginica
100	101	6.3	3.3	6.0	0.2	iris-virginica	iris	virginica
109	110	7.2	3.6	6.1	0.2	iris-virginica	iris	virginica

In [243]:

```
# #Extract the domain from a column containing email addresses:
```

```
data['Name'] = data['Species'].str.split('-').str[1]
data
```

Out[243]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
13	14	4.3	3.0	1.1	0.1	IRISSETOSA	iris	setosa
3	4	4.6	3.1	1.5	0.2	IRISSETOSA	iris	setosa
2	3	4.7	3.2	1.3	0.2	IRISSETOSA	iris	setosa
24	25	4.8	3.4	1.9	0.2	IRISSETOSA	iris	setosa
11	12	4.8	3.4	1.6	0.2	IRISSETOSA	iris	setosa
0	1	5.1	3.5	1.4	0.2	IRISSETOSA	iris	setosa
22	23	4.6	3.6	1.0	0.2	IRISSETOSA	iris	setosa
14	15	5.8	4.0	1.2	0.2	IRISSETOSA	iris	setosa
5	6	5.4	3.9	1.7	0.2	IRISSETOSA	iris	setosa
60	61	5.0	2.0	3.5	0.2	IRISVERSICOLOR	iris	versicolor
57	58	4.9	2.4	3.3	0.2	IRISVERSICOLOR	iris	versicolor
81	82	5.5	2.4	3.7	0.2	IRISVERSICOLOR	iris	versicolor
67	68	5.8	2.7	4.1	0.2	IRISVERSICOLOR	iris	versicolor
80	81	5.5	2.4	3.8	0.2	IRISVERSICOLOR	iris	versicolor
98	99	5.1	2.5	3.0	0.2	IRISVERSICOLOR	iris	versicolor
53	54	5.5	2.3	4.0	0.2	IRISVERSICOLOR	iris	versicolor
64	65	5.6	2.9	3.6	0.2	IRISVERSICOLOR	iris	versicolor
74	75	6.4	2.9	4.3	0.2	IRISVERSICOLOR	iris	versicolor
59	60	5.2	2.7	3.9	0.2	IRISVERSICOLOR	iris	versicolor
65	66	6.7	3.1	4.4	0.2	IRISVERSICOLOR	iris	versicolor
50	51	7.0	3.2	4.7	0.2	IRISVERSICOLOR	iris	versicolor
54	55	6.5	2.8	4.6	0.2	IRISVERSICOLOR	iris	versicolor
61	62	5.9	3.0	4.2	0.2	IRISVERSICOLOR	iris	versicolor
52	53	6.9	3.1	4.9	0.2	IRISVERSICOLOR	iris	versicolor
51	52	6.4	3.2	4.5	0.2	IRISVERSICOLOR	iris	versicolor
83	84	6.0	2.7	5.1	0.2	IRISVERSICOLOR	iris	versicolor
106	107	4.9	2.5	4.5	0.2	IRISVIRGINICA	iris	virginica
77	78	6.7	3.0	5.0	0.2	IRISVERSICOLOR	iris	versicolor
126	127	6.2	2.8	4.8	0.2	IRISVIRGINICA	iris	virginica
107	108	7.3	2.9	6.3	0.2	IRISVIRGINICA	iris	virginica
103	104	6.3	2.9	5.6	0.2	IRISVIRGINICA	iris	virginica
70	71	5.9	3.2	4.8	0.2	IRISVERSICOLOR	iris	versicolor
111	112	6.4	2.7	5.3	0.2	IRISVIRGINICA	iris	virginica
101	102	5.8	2.7	5.1	0.2	IRISVIRGINICA	iris	virginica
113	114	5.7	2.5	5.0	0.2	IRISVIRGINICA	iris	virginica
121	122	5.6	2.8	4.9	0.2	IRISVIRGINICA	iris	virginica
131	132	7.9	3.8	6.4	0.2	IRISVIRGINICA	iris	virginica
112	113	6.8	3.0	5.5	0.2	IRISVIRGINICA	iris	virginica
105	106	7.6	3.0	6.6	0.2	IRISVIRGINICA	iris	virginica
102	103	7.1	3.0	5.9	0.2	IRISVIRGINICA	iris	virginica
139	140	6.9	3.1	5.4	0.2	IRISVIRGINICA	iris	virginica

104	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
117	118	7.7	3.8	6.7	0.2	IRISVIRGINICA	iris	virginica
118	119	7.7	2.6	6.9	0.2	IRISVIRGINICA	iris	virginica
145	146	6.7	3.0	5.2	0.2	IRISVIRGINICA	iris	virginica
120	121	6.9	3.2	5.7	0.2	IRISVIRGINICA	iris	virginica
100	101	6.3	3.3	6.0	0.2	IRISVIRGINICA	iris	virginica
109	110	7.2	3.6	6.1	0.2	IRISVIRGINICA	iris	virginica

In [187]:

```
# Data Duplication:
# Identify and remove duplicate rows:
data.drop_duplicates
data
```

Out[187]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	sum_length	Dataset_Name	Species_Nam
12	13	4.8	1.098612	1.4	0.1	iris-setosa	6.2	iris	setos
13	14	4.3	1.098612	1.1	0.1	iris-setosa	5.4	iris	setos
34	35	4.9	1.131402	1.5	0.1	iris-setosa	6.4	iris	setos
9	10	4.9	1.131402	1.5	0.1	iris-setosa	6.4	iris	setos
37	38	4.9	1.131402	1.5	0.1	iris-setosa	6.4	iris	setos
...	...	...	...	...	...	...	...	...	.
140	141	6.7	1.131402	5.6	0.2	iris-virginica	12.3	iris	virginic
136	137	6.3	1.223775	5.6	0.2	iris-virginica	11.9	iris	virginic
144	145	6.7	1.193922	5.7	0.2	iris-virginica	12.4	iris	virginic
100	101	6.3	1.193922	6.0	0.2	iris-virginica	12.3	iris	virginic
109	110	7.2	1.280934	6.1	0.2	iris-virginica	13.3	iris	virginic

150 rows x 10 columns

In [188]:

```
# Identify and remove duplicate rows based on a subset of columns:
data.drop_duplicates(subset=['SepalLengthCm', 'PetalLengthCm'])
```

Out[188]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	sum_length	Dataset_Name	Species_Nam
12	13	4.8	1.098612	1.4	0.1	iris-setosa	6.2	iris	setos
13	14	4.3	1.098612	1.1	0.1	iris-setosa	5.4	iris	setos
34	35	4.9	1.131402	1.5	0.1	iris-setosa	6.4	iris	setos

32	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	sum_length	Dataset_Name	Species_Name
	33	5.2	1.410987	1.5	0.1	iris-setosa	6.7	iris	setosa
8	9	4.4	1.064711	1.4	0.2	iris-setosa	5.8	iris	setosa
...	...	...	...	...	...	...	...	...	...
120	121	6.9	1.163151	5.7	0.2	iris-virginica	12.6	iris	virginica
148	149	6.2	1.223775	5.4	0.2	iris-virginica	11.6	iris	virginica
140	141	6.7	1.131402	5.6	0.2	iris-virginica	12.3	iris	virginica
100	101	6.3	1.193922	6.0	0.2	iris-virginica	12.3	iris	virginica
109	110	7.2	1.280934	6.1	0.2	iris-virginica	13.3	iris	virginica

123 rows x 10 columns



In [214]:

```
# Merging and Joining (continued):
# Perform a full outer join on two DataFrames:
full_outer_join_data = pd.merge(data, data2, on='Id', how='outer')
full_outer_join_data
```

Out[214]:

	Id	SepalLengthCm_x	SepalWidthCm_x	PetalLengthCm_x	PetalWidthCm_x	Species_x	Dataset_Name	Species_Name
0	1	5.1	3.5	1.4	0.2	iris-setosa	iris	setosa
1	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	3	4.7	3.2	1.3	0.2	iris-setosa	iris	setosa
3	4	4.6	3.1	1.5	0.2	iris-setosa	iris	setosa
4	5	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	0.2	iris-virginica	iris	virginica
146	147	NaN	NaN	NaN	NaN	NaN	NaN	NaN
147	148	NaN	NaN	NaN	NaN	NaN	NaN	NaN
148	149	NaN	NaN	NaN	NaN	NaN	NaN	NaN
149	150	NaN	NaN	NaN	NaN	NaN	NaN	NaN

150 rows x 13 columns



In [215]:

```
# Perform a cross join on two DataFrames:
cross_join_data = pd.merge(data, data2, how='outer')
cross_join_data
```

Out[215]:



	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
0	1	5.1	3.5	1.4	0.2	Iris-setosa	NaN	NaN
1	1	5.1	3.5	1.4	0.2	iris-setosa	iris	setosa
2	2	4.9	3.0	1.4	0.2	Iris-setosa	NaN	NaN
3	3	4.7	3.2	1.3	0.2	Iris-setosa	NaN	NaN
4	3	4.7	3.2	1.3	0.2	iris-setosa	iris	setosa
...	...	...	...	...	...	...	...	...
193	146	6.7	3.0	5.2	2.3	Iris-virginica	NaN	NaN
194	147	6.3	2.5	5.0	1.9	Iris-virginica	NaN	NaN
195	148	6.5	3.0	5.2	2.0	Iris-virginica	NaN	NaN
196	149	6.2	3.4	5.4	2.3	Iris-virginica	NaN	NaN
197	150	5.9	3.0	5.1	1.8	Iris-virginica	NaN	NaN

198 rows x 8 columns

In [191]:

```
# Grouping and Aggregation (continued):
# Group by 'species' and calculate the sum, mean, and standard deviation of 'petal_width'
:
data.groupby('Species')['PetalWidthCm'].agg(['sum', 'mean', 'std'])
```

Out[191]:

	sum	mean	std
Species			
iris-setosa	9.4	0.188	0.032826
iris-versicolor	10.0	0.200	0.000000
iris-virginica	10.0	0.200	0.000000

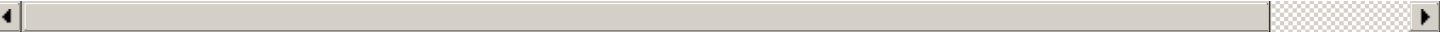
In [216]:

```
# Reshaping Data:
# Convert a long-form DataFrame to wide-form:
data.drop_duplicates(subset=['Species', 'PetalLengthCm'], inplace=True)
numeric_cols = data.select_dtypes(include=[float, int]).columns
numeric_cols = numeric_cols[numeric_cols != 'PetalLengthCm'] # Exclude 'PetalLengthCm'
agg_data = data.groupby(['Species', 'PetalLengthCm'])[numeric_cols].mean().reset_index()
wide_data = agg_data.pivot(index='Species', columns='PetalLengthCm', values='SepalWidthCm')
wide_data
```

Out[216]:

PetalLengthCm	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.9	3.0	...	5.7	5.8	5.9	6.0	6.1	6.3	6.4	6.6
Species																			
iris-setosa	3.6	3.0	4.0	3.2	3.5	3.1	3.4	3.9	3.4	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
iris-versicolor	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2.5	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
iris-virginica	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	3.2	3.0	3.0	3.3	3.6	2.9	3.8	3.0

3 rows x 43 columns



In [193]:

```
import pandas as pd
```

```
data = pd.read_csv('/kaggle/input/iris/Iris.csv')
```

In [4]:

```
# Dropping duplicates and selecting numeric columns
data.drop_duplicates(subset=['Species', 'PetalLengthCm'], inplace=True)
numeric_cols = data.select_dtypes(include=[float, int]).columns
numeric_cols = numeric_cols[numeric_cols != 'PetalLengthCm'] # Exclude 'PetalLengthCm'

# Aggregating data by Species and PetalLengthCm
agg_data = data.groupby(['Species', 'PetalLengthCm'])[numeric_cols].mean().reset_index()

# Pivoting the data to wide-form
wide_data = agg_data.pivot(index='Species', columns='PetalLengthCm', values='SepalWidthCm')

# Resetting index to make 'Species' a column
wide_data.reset_index(inplace=True)
print(wide_data)

# Converting wide-form to long-form
long_data = wide_data.melt(id_vars='Species', var_name='PetalLengthCm', value_name='SepalWidthCm')
long_data
```

```
PetalLengthCm      Species  1.0  1.1  1.2  1.3  1.4  1.5  1.6  1.7  1.9  \
0                Iris-setosa  3.6  3.0  4.0  3.2  3.5  3.1  3.4  3.9  3.4
1                Iris-versicolor  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
2                Iris-virginica  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
```

```
PetalLengthCm  ...  5.7  5.8  5.9  6.0  6.1  6.3  6.4  6.6  6.7  6.9
0              ...  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
1              ...  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN  NaN
2              ...  3.2  3.0  3.0  3.3  3.6  2.9  3.8  3.0  3.8  2.6
```

[3 rows x 44 columns]

Out[4]:

	Species	PetalLengthCm	SepalWidthCm
0	Iris-setosa	1.0	3.6
1	Iris-versicolor	1.0	NaN
2	Iris-virginica	1.0	NaN
3	Iris-setosa	1.1	3.0
4	Iris-versicolor	1.1	NaN
...	...	...	...
124	Iris-versicolor	6.7	NaN
125	Iris-virginica	6.7	3.8
126	Iris-setosa	6.9	NaN
127	Iris-versicolor	6.9	NaN
128	Iris-virginica	6.9	2.6

129 rows x 3 columns

In [218]:

```
# Handling Missing Data:
# Remove columns that have more than 50% missing values:
data.dropna(axis=1, thresh=0.5)
```

Out[218]:

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
123	4.4	4.2	3.0	1.1	0.1	Iris-setosa	Iris-setosa

13	14	4.5	3.0	1.1	0.1	iris-setosa	iris	setosa
Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name	
3	4	4.6	3.1	1.5	0.2	iris-setosa	iris	setosa
2	3	4.7	3.2	1.3	0.2	iris-setosa	iris	setosa
24	25	4.8	3.4	1.9	0.2	iris-setosa	iris	setosa
11	12	4.8	3.4	1.6	0.2	iris-setosa	iris	setosa
0	1	5.1	3.5	1.4	0.2	iris-setosa	iris	setosa
22	23	4.6	3.6	1.0	0.2	iris-setosa	iris	setosa
14	15	5.8	4.0	1.2	0.2	iris-setosa	iris	setosa
5	6	5.4	3.9	1.7	0.2	iris-setosa	iris	setosa
60	61	5.0	2.0	3.5	0.2	iris-versicolor	iris	versicolor
57	58	4.9	2.4	3.3	0.2	iris-versicolor	iris	versicolor
81	82	5.5	2.4	3.7	0.2	iris-versicolor	iris	versicolor
67	68	5.8	2.7	4.1	0.2	iris-versicolor	iris	versicolor
80	81	5.5	2.4	3.8	0.2	iris-versicolor	iris	versicolor
98	99	5.1	2.5	3.0	0.2	iris-versicolor	iris	versicolor
53	54	5.5	2.3	4.0	0.2	iris-versicolor	iris	versicolor
64	65	5.6	2.9	3.6	0.2	iris-versicolor	iris	versicolor
74	75	6.4	2.9	4.3	0.2	iris-versicolor	iris	versicolor
59	60	5.2	2.7	3.9	0.2	iris-versicolor	iris	versicolor
65	66	6.7	3.1	4.4	0.2	iris-versicolor	iris	versicolor
50	51	7.0	3.2	4.7	0.2	iris-versicolor	iris	versicolor
54	55	6.5	2.8	4.6	0.2	iris-versicolor	iris	versicolor
61	62	5.9	3.0	4.2	0.2	iris-versicolor	iris	versicolor
52	53	6.9	3.1	4.9	0.2	iris-versicolor	iris	versicolor
51	52	6.4	3.2	4.5	0.2	iris-versicolor	iris	versicolor
83	84	6.0	2.7	5.1	0.2	iris-versicolor	iris	versicolor
106	107	4.9	2.5	4.5	0.2	iris-virginica	iris	virginica
77	78	6.7	3.0	5.0	0.2	iris-versicolor	iris	versicolor
126	127	6.2	2.8	4.8	0.2	iris-virginica	iris	virginica
107	108	7.3	2.9	6.3	0.2	iris-virginica	iris	virginica
103	104	6.3	2.9	5.6	0.2	iris-virginica	iris	virginica
70	71	5.9	3.2	4.8	0.2	iris-versicolor	iris	versicolor
111	112	6.4	2.7	5.3	0.2	iris-virginica	iris	virginica
101	102	5.8	2.7	5.1	0.2	iris-virginica	iris	virginica
113	114	5.7	2.5	5.0	0.2	iris-virginica	iris	virginica
121	122	5.6	2.8	4.9	0.2	iris-virginica	iris	virginica
131	132	7.9	3.8	6.4	0.2	iris-virginica	iris	virginica
112	113	6.8	3.0	5.5	0.2	iris-virginica	iris	virginica
105	106	7.6	3.0	6.6	0.2	iris-virginica	iris	virginica
102	103	7.1	3.0	5.9	0.2	iris-virginica	iris	virginica
139	140	6.9	3.1	5.4	0.2	iris-virginica	iris	virginica
104	105	6.5	3.0	5.8	0.2	iris-virginica	iris	virginica
117	118	7.7	3.8	6.7	0.2	iris-virginica	iris	virginica
118	119	7.7	2.6	6.9	0.2	iris-virginica	iris	virginica
145	146	6.7	3.0	5.2	0.2	iris-virginica	iris	virginica
120	121	6.0	2.2	5.7	0.2	iris-virginica	iris	virginica

Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
100	6.3	3.3	6.0	0.2	iris-virginica	iris	virginica
109	7.2	3.6	6.1	0.2	iris-virginica	iris	virginica

In [219]:

```
# Replace missing values in a column with the mode of the column:
data['SepalWidthCm'].fillna(data['SepalWidthCm'].mode()[0])
```

Out[219]:

```
13      3.0
3       3.1
2       3.2
24      3.4
11      3.4
0       3.5
22      3.6
14      4.0
5       3.9
60      2.0
57      2.4
81      2.4
67      2.7
80      2.4
98      2.5
53      2.3
64      2.9
74      2.9
59      2.7
65      3.1
50      3.2
54      2.8
61      3.0
52      3.1
51      3.2
83      2.7
106     2.5
77      3.0
126     2.8
107     2.9
103     2.9
70      3.2
111     2.7
101     2.7
113     2.5
121     2.8
131     3.8
112     3.0
105     3.0
102     3.0
139     3.1
104     3.0
117     3.8
118     2.6
145     3.0
120     3.2
100     3.3
109     3.6
```

Name: SepalWidthCm, dtype: float64

In [220]:

```
# Assuming 'data' is your DataFrame
missing_values = data.isna().sum()
print("Missing values in each column:")
print(missing_values)

# Alternatively, you can use isnull()
# missing_values = data.isnull().sum()
```

```
missing_values = data.isna().sum()
print(missing_values)
```

Missing values in each column:

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

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

In [221]:

```
# Impute missing values in a column using mean values from another column:
data['SepalLengthCm'].fillna(data['SepalWidthCm'].mean())
```

Out[221]:

```
13      4.3
3       4.6
2       4.7
24      4.8
11      4.8
0       5.1
22      4.6
14      5.8
5       5.4
60      5.0
57      4.9
81      5.5
67      5.8
80      5.5
98      5.1
53      5.5
64      5.6
74      6.4
59      5.2
65      6.7
50      7.0
54      6.5
61      5.9
52      6.9
51      6.4
83      6.0
106     4.9
77      6.7
126     6.2
107     7.3
103     6.3
70      5.9
111     6.4
101     5.8
113     5.7
121     5.6
131     7.9
112     6.8
105     7.6
102     7.1
139     6.9
104     6.5
117     7.7
```

```
117      7.7
118      7.7
145      6.7
120      6.9
100      6.3
109      7.2
Name: SepalLengthCm, dtype: float64
```

In [222]:

```
# Interpolate missing values in a time series data:
data['SepalLengthCm'].interpolate()
```

Out[222]:

```
13      4.3
3       4.6
2       4.7
24      4.8
11      4.8
0       5.1
22      4.6
14      5.8
5       5.4
60      5.0
57      4.9
81      5.5
67      5.8
80      5.5
98      5.1
53      5.5
64      5.6
74      6.4
59      5.2
65      6.7
50      7.0
54      6.5
61      5.9
52      6.9
51      6.4
83      6.0
106     4.9
77      6.7
126     6.2
107     7.3
103     6.3
70      5.9
111     6.4
101     5.8
113     5.7
121     5.6
131     7.9
112     6.8
105     7.6
102     7.1
139     6.9
104     6.5
117     7.7
118     7.7
145     6.7
120     6.9
100     6.3
109     7.2
Name: SepalLengthCm, dtype: float64
```

In [223]:

```
# String and Text Processing:
# Convert a column of text to uppercase:
data['Species'] = data['Species'].str.upper()
print(data['Species'])
```

```

3          IRIS-SETOSA
2          IRIS-SETOSA
24         IRIS-SETOSA
11         IRIS-SETOSA
0          IRIS-SETOSA
22         IRIS-SETOSA
14         IRIS-SETOSA
5          IRIS-SETOSA
60         IRIS-VERSICOLOR
57         IRIS-VERSICOLOR
81         IRIS-VERSICOLOR
67         IRIS-VERSICOLOR
80         IRIS-VERSICOLOR
98         IRIS-VERSICOLOR
53         IRIS-VERSICOLOR
64         IRIS-VERSICOLOR
74         IRIS-VERSICOLOR
59         IRIS-VERSICOLOR
65         IRIS-VERSICOLOR
50         IRIS-VERSICOLOR
54         IRIS-VERSICOLOR
61         IRIS-VERSICOLOR
52         IRIS-VERSICOLOR
51         IRIS-VERSICOLOR
83         IRIS-VERSICOLOR
106        IRIS-VIRGINICA
77         IRIS-VERSICOLOR
126        IRIS-VIRGINICA
107        IRIS-VIRGINICA
103        IRIS-VIRGINICA
70         IRIS-VERSICOLOR
111        IRIS-VIRGINICA
101        IRIS-VIRGINICA
113        IRIS-VIRGINICA
121        IRIS-VIRGINICA
131        IRIS-VIRGINICA
112        IRIS-VIRGINICA
105        IRIS-VIRGINICA
102        IRIS-VIRGINICA
139        IRIS-VIRGINICA
104        IRIS-VIRGINICA
117        IRIS-VIRGINICA
118        IRIS-VIRGINICA
145        IRIS-VIRGINICA
120        IRIS-VIRGINICA
100        IRIS-VIRGINICA
109        IRIS-VIRGINICA
Name: Species, dtype: object

```

In [224]:

```

# Remove punctuation from a column of text:
import string
data['Species'] = data['Species'].str.translate(str.maketrans('', '', string.punctuation))
print(data['Species'])

```

```

13         IRISSETOSA
3          IRISSETOSA
2          IRISSETOSA
24         IRISSETOSA
11         IRISSETOSA
0          IRISSETOSA
22         IRISSETOSA
14         IRISSETOSA
5          IRISSETOSA
60         IRISVERSICOLOR
57         IRISVERSICOLOR
81         IRISVERSICOLOR
67         IRISVERSICOLOR
80         IRISVERSICOLOR
98         IRISVERSICOLOR

```

```
53      IRISVERSICOLOR
64      IRISVERSICOLOR
74      IRISVERSICOLOR
59      IRISVERSICOLOR
65      IRISVERSICOLOR
50      IRISVERSICOLOR
54      IRISVERSICOLOR
61      IRISVERSICOLOR
52      IRISVERSICOLOR
51      IRISVERSICOLOR
83      IRISVERSICOLOR
106      IRISVIRGINICA
77      IRISVERSICOLOR
126      IRISVIRGINICA
107      IRISVIRGINICA
103      IRISVIRGINICA
70      IRISVERSICOLOR
111      IRISVIRGINICA
101      IRISVIRGINICA
113      IRISVIRGINICA
121      IRISVIRGINICA
131      IRISVIRGINICA
112      IRISVIRGINICA
105      IRISVIRGINICA
102      IRISVIRGINICA
139      IRISVIRGINICA
104      IRISVIRGINICA
117      IRISVIRGINICA
118      IRISVIRGINICA
145      IRISVIRGINICA
120      IRISVIRGINICA
100      IRISVIRGINICA
109      IRISVIRGINICA
Name: Species, dtype: object
```

In [225]:

```
data
```

Out[225]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name	
	13	14	4.3	3.0	1.1	0.1	IRISSETOSA	iris	setosa
	3	4	4.6	3.1	1.5	0.2	IRISSETOSA	iris	setosa
	2	3	4.7	3.2	1.3	0.2	IRISSETOSA	iris	setosa
	24	25	4.8	3.4	1.9	0.2	IRISSETOSA	iris	setosa
	11	12	4.8	3.4	1.6	0.2	IRISSETOSA	iris	setosa
	0	1	5.1	3.5	1.4	0.2	IRISSETOSA	iris	setosa
	22	23	4.6	3.6	1.0	0.2	IRISSETOSA	iris	setosa
	14	15	5.8	4.0	1.2	0.2	IRISSETOSA	iris	setosa
	5	6	5.4	3.9	1.7	0.2	IRISSETOSA	iris	setosa
	60	61	5.0	2.0	3.5	0.2	IRISVERSICOLOR	iris	versicolor
	57	58	4.9	2.4	3.3	0.2	IRISVERSICOLOR	iris	versicolor
	81	82	5.5	2.4	3.7	0.2	IRISVERSICOLOR	iris	versicolor
	67	68	5.8	2.7	4.1	0.2	IRISVERSICOLOR	iris	versicolor
	80	81	5.5	2.4	3.8	0.2	IRISVERSICOLOR	iris	versicolor
	98	99	5.1	2.5	3.0	0.2	IRISVERSICOLOR	iris	versicolor
	53	54	5.5	2.3	4.0	0.2	IRISVERSICOLOR	iris	versicolor
	64	65	5.6	2.9	3.6	0.2	IRISVERSICOLOR	iris	versicolor
	74	75	6.4	2.9	4.3	0.2	IRISVERSICOLOR	iris	versicolor



59	69	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	IRISVERSICOLOR	Species	Dataset_Name	Species_Name
65	66	6.7	3.1	4.4	0.2	IRISVERSICOLOR		iris	versicolor
50	51	7.0	3.2	4.7	0.2	IRISVERSICOLOR		iris	versicolor
54	55	6.5	2.8	4.6	0.2	IRISVERSICOLOR		iris	versicolor
61	62	5.9	3.0	4.2	0.2	IRISVERSICOLOR		iris	versicolor
52	53	6.9	3.1	4.9	0.2	IRISVERSICOLOR		iris	versicolor
51	52	6.4	3.2	4.5	0.2	IRISVERSICOLOR		iris	versicolor
83	84	6.0	2.7	5.1	0.2	IRISVERSICOLOR		iris	versicolor
106	107	4.9	2.5	4.5	0.2	IRISVIRGINICA		iris	virginica
77	78	6.7	3.0	5.0	0.2	IRISVERSICOLOR		iris	versicolor
126	127	6.2	2.8	4.8	0.2	IRISVIRGINICA		iris	virginica
107	108	7.3	2.9	6.3	0.2	IRISVIRGINICA		iris	virginica
103	104	6.3	2.9	5.6	0.2	IRISVIRGINICA		iris	virginica
70	71	5.9	3.2	4.8	0.2	IRISVERSICOLOR		iris	versicolor
111	112	6.4	2.7	5.3	0.2	IRISVIRGINICA		iris	virginica
101	102	5.8	2.7	5.1	0.2	IRISVIRGINICA		iris	virginica
113	114	5.7	2.5	5.0	0.2	IRISVIRGINICA		iris	virginica
121	122	5.6	2.8	4.9	0.2	IRISVIRGINICA		iris	virginica
131	132	7.9	3.8	6.4	0.2	IRISVIRGINICA		iris	virginica
112	113	6.8	3.0	5.5	0.2	IRISVIRGINICA		iris	virginica
105	106	7.6	3.0	6.6	0.2	IRISVIRGINICA		iris	virginica
102	103	7.1	3.0	5.9	0.2	IRISVIRGINICA		iris	virginica
139	140	6.9	3.1	5.4	0.2	IRISVIRGINICA		iris	virginica
104	105	6.5	3.0	5.8	0.2	IRISVIRGINICA		iris	virginica
117	118	7.7	3.8	6.7	0.2	IRISVIRGINICA		iris	virginica
118	119	7.7	2.6	6.9	0.2	IRISVIRGINICA		iris	virginica
145	146	6.7	3.0	5.2	0.2	IRISVIRGINICA		iris	virginica
120	121	6.9	3.2	5.7	0.2	IRISVIRGINICA		iris	virginica
100	101	6.3	3.3	6.0	0.2	IRISVIRGINICA		iris	virginica
109	110	7.2	3.6	6.1	0.2	IRISVIRGINICA		iris	virginica

In [226]:

```
# Tokenize a column of text into individual words:
data['tokenized_text'] = data['Species'].str.split()
print(data['tokenized_text'])
```

```
13      [IRISSETOSA]
3       [IRISSETOSA]
2       [IRISSETOSA]
24      [IRISSETOSA]
11      [IRISSETOSA]
0       [IRISSETOSA]
22      [IRISSETOSA]
14      [IRISSETOSA]
5       [IRISSETOSA]
60      [IRISVERSICOLOR]
57      [IRISVERSICOLOR]
81      [IRISVERSICOLOR]
67      [IRISVERSICOLOR]
80      [IRISVERSICOLOR]
98      [IRISVERSICOLOR]
53      [IRISVERSICOLOR]
64      [IRISVERSICOLOR]
```

```
74      [IRISVERSICOLOR]
59      [IRISVERSICOLOR]
65      [IRISVERSICOLOR]
50      [IRISVERSICOLOR]
54      [IRISVERSICOLOR]
61      [IRISVERSICOLOR]
52      [IRISVERSICOLOR]
51      [IRISVERSICOLOR]
83      [IRISVERSICOLOR]
106     [IRISVIRGINICA]
77      [IRISVERSICOLOR]
126     [IRISVIRGINICA]
107     [IRISVIRGINICA]
103     [IRISVIRGINICA]
70      [IRISVERSICOLOR]
111     [IRISVIRGINICA]
101     [IRISVIRGINICA]
113     [IRISVIRGINICA]
121     [IRISVIRGINICA]
131     [IRISVIRGINICA]
112     [IRISVIRGINICA]
105     [IRISVIRGINICA]
102     [IRISVIRGINICA]
139     [IRISVIRGINICA]
104     [IRISVIRGINICA]
117     [IRISVIRGINICA]
118     [IRISVIRGINICA]
145     [IRISVIRGINICA]
120     [IRISVIRGINICA]
100     [IRISVIRGINICA]
109     [IRISVIRGINICA]
Name: tokenized_text, dtype: object
```

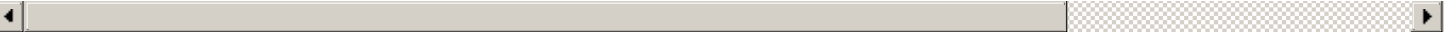
In [227]:

```
# Handling Categorical Data:
# Convert a categorical column to numerical values:
data['category_column'] = data['Species_Name'].astype('category').cat.codes
data
```

Out[227]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
13	14	4.3	3.0	1.1	0.1	IRISSETOSA	iris	setosa
3	4	4.6	3.1	1.5	0.2	IRISSETOSA	iris	setosa
2	3	4.7	3.2	1.3	0.2	IRISSETOSA	iris	setosa
24	25	4.8	3.4	1.9	0.2	IRISSETOSA	iris	setosa
11	12	4.8	3.4	1.6	0.2	IRISSETOSA	iris	setosa
0	1	5.1	3.5	1.4	0.2	IRISSETOSA	iris	setosa
22	23	4.6	3.6	1.0	0.2	IRISSETOSA	iris	setosa
14	15	5.8	4.0	1.2	0.2	IRISSETOSA	iris	setosa
5	6	5.4	3.9	1.7	0.2	IRISSETOSA	iris	setosa
60	61	5.0	2.0	3.5	0.2	IRISVERSICOLOR	iris	versicolor
57	58	4.9	2.4	3.3	0.2	IRISVERSICOLOR	iris	versicolor
81	82	5.5	2.4	3.7	0.2	IRISVERSICOLOR	iris	versicolor
67	68	5.8	2.7	4.1	0.2	IRISVERSICOLOR	iris	versicolor
80	81	5.5	2.4	3.8	0.2	IRISVERSICOLOR	iris	versicolor
98	99	5.1	2.5	3.0	0.2	IRISVERSICOLOR	iris	versicolor
53	54	5.5	2.3	4.0	0.2	IRISVERSICOLOR	iris	versicolor
64	65	5.6	2.9	3.6	0.2	IRISVERSICOLOR	iris	versicolor
74	75	6.4	2.9	4.3	0.2	IRISVERSICOLOR	iris	versicolor

59	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
60		5.2	2.7	3.9	0.2	IRISVERSICOLOR	iris	versicolor
65	66	6.7	3.1	4.4	0.2	IRISVERSICOLOR	iris	versicolor
50	51	7.0	3.2	4.7	0.2	IRISVERSICOLOR	iris	versicolor
54	55	6.5	2.8	4.6	0.2	IRISVERSICOLOR	iris	versicolor
61	62	5.9	3.0	4.2	0.2	IRISVERSICOLOR	iris	versicolor
52	53	6.9	3.1	4.9	0.2	IRISVERSICOLOR	iris	versicolor
51	52	6.4	3.2	4.5	0.2	IRISVERSICOLOR	iris	versicolor
83	84	6.0	2.7	5.1	0.2	IRISVERSICOLOR	iris	versicolor
106	107	4.9	2.5	4.5	0.2	IRISVIRGINICA	iris	virginica
77	78	6.7	3.0	5.0	0.2	IRISVERSICOLOR	iris	versicolor
126	127	6.2	2.8	4.8	0.2	IRISVIRGINICA	iris	virginica
107	108	7.3	2.9	6.3	0.2	IRISVIRGINICA	iris	virginica
103	104	6.3	2.9	5.6	0.2	IRISVIRGINICA	iris	virginica
70	71	5.9	3.2	4.8	0.2	IRISVERSICOLOR	iris	versicolor
111	112	6.4	2.7	5.3	0.2	IRISVIRGINICA	iris	virginica
101	102	5.8	2.7	5.1	0.2	IRISVIRGINICA	iris	virginica
113	114	5.7	2.5	5.0	0.2	IRISVIRGINICA	iris	virginica
121	122	5.6	2.8	4.9	0.2	IRISVIRGINICA	iris	virginica
131	132	7.9	3.8	6.4	0.2	IRISVIRGINICA	iris	virginica
112	113	6.8	3.0	5.5	0.2	IRISVIRGINICA	iris	virginica
105	106	7.6	3.0	6.6	0.2	IRISVIRGINICA	iris	virginica
102	103	7.1	3.0	5.9	0.2	IRISVIRGINICA	iris	virginica
139	140	6.9	3.1	5.4	0.2	IRISVIRGINICA	iris	virginica
104	105	6.5	3.0	5.8	0.2	IRISVIRGINICA	iris	virginica
117	118	7.7	3.8	6.7	0.2	IRISVIRGINICA	iris	virginica
118	119	7.7	2.6	6.9	0.2	IRISVIRGINICA	iris	virginica
145	146	6.7	3.0	5.2	0.2	IRISVIRGINICA	iris	virginica
120	121	6.9	3.2	5.7	0.2	IRISVIRGINICA	iris	virginica
100	101	6.3	3.3	6.0	0.2	IRISVIRGINICA	iris	virginica
109	110	7.2	3.6	6.1	0.2	IRISVIRGINICA	iris	virginica



In [228]:

```
# Rename categories in a categorical column:
data['Species_Name'] = data['Species_Name'].astype('category')
new_categories = ['Species_1', 'Species_2', 'Species_3']
data['Species_Type'] = data['Species_Name'].cat.rename_categories(new_categories)
data
```

Out[228]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
13	14	4.3	3.0	1.1	0.1	IRISSETOSA	iris	setosa
3	4	4.6	3.1	1.5	0.2	IRISSETOSA	iris	setosa
2	3	4.7	3.2	1.3	0.2	IRISSETOSA	iris	setosa
24	25	4.8	3.4	1.9	0.2	IRISSETOSA	iris	setosa
11	12	4.8	3.4	1.6	0.2	IRISSETOSA	iris	setosa
0	1	5.1	3.5	1.4	0.2	IRISSETOSA	iris	setosa

22	14	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	IRISSpecies	Dataset_Name	Species_Name
14	15	5.8	4.0	1.2	0.2	IRISSETOSA	iris	setosa
5	6	5.4	3.9	1.7	0.2	IRISSETOSA	iris	setosa
60	61	5.0	2.0	3.5	0.2	IRISVERSICOLOR	iris	versicolor
57	58	4.9	2.4	3.3	0.2	IRISVERSICOLOR	iris	versicolor
81	82	5.5	2.4	3.7	0.2	IRISVERSICOLOR	iris	versicolor
67	68	5.8	2.7	4.1	0.2	IRISVERSICOLOR	iris	versicolor
80	81	5.5	2.4	3.8	0.2	IRISVERSICOLOR	iris	versicolor
98	99	5.1	2.5	3.0	0.2	IRISVERSICOLOR	iris	versicolor
53	54	5.5	2.3	4.0	0.2	IRISVERSICOLOR	iris	versicolor
64	65	5.6	2.9	3.6	0.2	IRISVERSICOLOR	iris	versicolor
74	75	6.4	2.9	4.3	0.2	IRISVERSICOLOR	iris	versicolor
59	60	5.2	2.7	3.9	0.2	IRISVERSICOLOR	iris	versicolor
65	66	6.7	3.1	4.4	0.2	IRISVERSICOLOR	iris	versicolor
50	51	7.0	3.2	4.7	0.2	IRISVERSICOLOR	iris	versicolor
54	55	6.5	2.8	4.6	0.2	IRISVERSICOLOR	iris	versicolor
61	62	5.9	3.0	4.2	0.2	IRISVERSICOLOR	iris	versicolor
52	53	6.9	3.1	4.9	0.2	IRISVERSICOLOR	iris	versicolor
51	52	6.4	3.2	4.5	0.2	IRISVERSICOLOR	iris	versicolor
83	84	6.0	2.7	5.1	0.2	IRISVERSICOLOR	iris	versicolor
106	107	4.9	2.5	4.5	0.2	IRISVIRGINICA	iris	virginica
77	78	6.7	3.0	5.0	0.2	IRISVERSICOLOR	iris	versicolor
126	127	6.2	2.8	4.8	0.2	IRISVIRGINICA	iris	virginica
107	108	7.3	2.9	6.3	0.2	IRISVIRGINICA	iris	virginica
103	104	6.3	2.9	5.6	0.2	IRISVIRGINICA	iris	virginica
70	71	5.9	3.2	4.8	0.2	IRISVERSICOLOR	iris	versicolor
111	112	6.4	2.7	5.3	0.2	IRISVIRGINICA	iris	virginica
101	102	5.8	2.7	5.1	0.2	IRISVIRGINICA	iris	virginica
113	114	5.7	2.5	5.0	0.2	IRISVIRGINICA	iris	virginica
121	122	5.6	2.8	4.9	0.2	IRISVIRGINICA	iris	virginica
131	132	7.9	3.8	6.4	0.2	IRISVIRGINICA	iris	virginica
112	113	6.8	3.0	5.5	0.2	IRISVIRGINICA	iris	virginica
105	106	7.6	3.0	6.6	0.2	IRISVIRGINICA	iris	virginica
102	103	7.1	3.0	5.9	0.2	IRISVIRGINICA	iris	virginica
139	140	6.9	3.1	5.4	0.2	IRISVIRGINICA	iris	virginica
104	105	6.5	3.0	5.8	0.2	IRISVIRGINICA	iris	virginica
117	118	7.7	3.8	6.7	0.2	IRISVIRGINICA	iris	virginica
118	119	7.7	2.6	6.9	0.2	IRISVIRGINICA	iris	virginica
145	146	6.7	3.0	5.2	0.2	IRISVIRGINICA	iris	virginica
120	121	6.9	3.2	5.7	0.2	IRISVIRGINICA	iris	virginica
100	101	6.3	3.3	6.0	0.2	IRISVIRGINICA	iris	virginica
109	110	7.2	3.6	6.1	0.2	IRISVIRGINICA	iris	virginica



In [251]:

```
# Remove categories that have less than 2% representation in a categorical column:
```

```

# Remove categories that have less than 2% representation in a categorical column.
# Convert 'Name' column to a categorical type
data['Name'] = data['Name'].astype('category')
# Calculate the percentage representation of each category
category_counts = data['Name'].value_counts(normalize=True) * 100
print("Category Representation (%):")
print(category_counts)

```

```

Category Representation (%):
Series([], Name: proportion, dtype: float64)

```

In [252]:

```

# Identify categories with less than 2% representation
threshold = 2
categories_to_remove = category_counts[category_counts < threshold].index
print(categories_to_remove)

```

```

CategoricalIndex([], categories=[], ordered=False, dtype='category', name='Name')

```

In [255]:

```

# Remove categories with less than 2% representation
data['Name'] = data['Name'].cat.remove_categories(categories_to_remove)
print(data['Name'])

```

Out[255]:

```

13      NaN
3       NaN
2       NaN
24      NaN
11      NaN
0       NaN
22      NaN
14      NaN
5       NaN
60      NaN
57      NaN
81      NaN
67      NaN
80      NaN
98      NaN
53      NaN
64      NaN
74      NaN
59      NaN
65      NaN
50      NaN
54      NaN
61      NaN
52      NaN
51      NaN
83      NaN
106     NaN
77      NaN
126     NaN
107     NaN
103     NaN
70      NaN
111     NaN
101     NaN
113     NaN
121     NaN
131     NaN
112     NaN
105     NaN
102     NaN
139     NaN
104     NaN
117     NaN
118     NaN
145     NaN

```

120 NaN  
100 NaN  
109 NaN  
Name: Name, dtype: category  
Categories (0, float64): []

In [256]:

```
# Convert a numerical column to its rank:
data['Rank'] = data['SepalLengthCm'].rank()
data['Rank_Average'] = data['SepalLengthCm'].rank(method='average')
data['Rank_Min'] = data['SepalLengthCm'].rank(method='min')
data['Rank_Max'] = data['SepalLengthCm'].rank(method='max')
data['Rank_First'] = data['SepalLengthCm'].rank(method='first')
data['Rank_Dense'] = data['SepalLengthCm'].rank(method='dense')
data
```

Out[256]:

Id		SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Dataset_Name	Species_Name
13	14	4.3	3.0	1.1	0.1	IRISSETOSA	iris	setosa
3	4	4.6	3.1	1.5	0.2	IRISSETOSA	iris	setosa
2	3	4.7	3.2	1.3	0.2	IRISSETOSA	iris	setosa
24	25	4.8	3.4	1.9	0.2	IRISSETOSA	iris	setosa
11	12	4.8	3.4	1.6	0.2	IRISSETOSA	iris	setosa
0	1	5.1	3.5	1.4	0.2	IRISSETOSA	iris	setosa
22	23	4.6	3.6	1.0	0.2	IRISSETOSA	iris	setosa
14	15	5.8	4.0	1.2	0.2	IRISSETOSA	iris	setosa
5	6	5.4	3.9	1.7	0.2	IRISSETOSA	iris	setosa
60	61	5.0	2.0	3.5	0.2	IRISVERSICOLOR	iris	versicolor []
57	58	4.9	2.4	3.3	0.2	IRISVERSICOLOR	iris	versicolor []
81	82	5.5	2.4	3.7	0.2	IRISVERSICOLOR	iris	versicolor []
67	68	5.8	2.7	4.1	0.2	IRISVERSICOLOR	iris	versicolor []
80	81	5.5	2.4	3.8	0.2	IRISVERSICOLOR	iris	versicolor []
98	99	5.1	2.5	3.0	0.2	IRISVERSICOLOR	iris	versicolor []
53	54	5.5	2.3	4.0	0.2	IRISVERSICOLOR	iris	versicolor []
64	65	5.6	2.9	3.6	0.2	IRISVERSICOLOR	iris	versicolor []
74	75	6.4	2.9	4.3	0.2	IRISVERSICOLOR	iris	versicolor []
59	60	5.2	2.7	3.9	0.2	IRISVERSICOLOR	iris	versicolor []
65	66	6.7	3.1	4.4	0.2	IRISVERSICOLOR	iris	versicolor []
50	51	7.0	3.2	4.7	0.2	IRISVERSICOLOR	iris	versicolor []
54	55	6.5	2.8	4.6	0.2	IRISVERSICOLOR	iris	versicolor []
61	62	5.9	3.0	4.2	0.2	IRISVERSICOLOR	iris	versicolor []
52	53	6.9	3.1	4.9	0.2	IRISVERSICOLOR	iris	versicolor []
51	52	6.4	3.2	4.5	0.2	IRISVERSICOLOR	iris	versicolor []
83	84	6.0	2.7	5.1	0.2	IRISVERSICOLOR	iris	versicolor []
106	107	4.9	2.5	4.5	0.2	IRISVIRGINICA	iris	virginica
77	78	6.7	3.0	5.0	0.2	IRISVERSICOLOR	iris	versicolor []
126	127	6.2	2.8	4.8	0.2	IRISVIRGINICA	iris	virginica
107	108	7.3	2.9	6.3	0.2	IRISVIRGINICA	iris	virginica
103	104	6.3	2.9	5.6	0.2	IRISVIRGINICA	iris	virginica
70	71	5.9	3.2	4.8	0.2	IRISVERSICOLOR	iris	versicolor []

111	112	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	IRISVIRGINICA	Dataset_Name	Species	Y_Value
101	102	5.8	2.7	5.1	0.2	IRISVIRGINICA	iris	virginica	
113	114	5.7	2.5	5.0	0.2	IRISVIRGINICA	iris	virginica	
121	122	5.6	2.8	4.9	0.2	IRISVIRGINICA	iris	virginica	
131	132	7.9	3.8	6.4	0.2	IRISVIRGINICA	iris	virginica	
112	113	6.8	3.0	5.5	0.2	IRISVIRGINICA	iris	virginica	
105	106	7.6	3.0	6.6	0.2	IRISVIRGINICA	iris	virginica	
102	103	7.1	3.0	5.9	0.2	IRISVIRGINICA	iris	virginica	
139	140	6.9	3.1	5.4	0.2	IRISVIRGINICA	iris	virginica	
104	105	6.5	3.0	5.8	0.2	IRISVIRGINICA	iris	virginica	
117	118	7.7	3.8	6.7	0.2	IRISVIRGINICA	iris	virginica	
118	119	7.7	2.6	6.9	0.2	IRISVIRGINICA	iris	virginica	
145	146	6.7	3.0	5.2	0.2	IRISVIRGINICA	iris	virginica	
120	121	6.9	3.2	5.7	0.2	IRISVIRGINICA	iris	virginica	
100	101	6.3	3.3	6.0	0.2	IRISVIRGINICA	iris	virginica	
109	110	7.2	3.6	6.1	0.2	IRISVIRGINICA	iris	virginica	

In [257]:

```
#Lbraries for plotting (Data VIZ)
import matplotlib.pyplot as plt
import seaborn as sns
```

In [258]:

```
df1=pd.read_csv('/kaggle/input/iris/Iris.csv')
df1
```

Out[258]:

	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

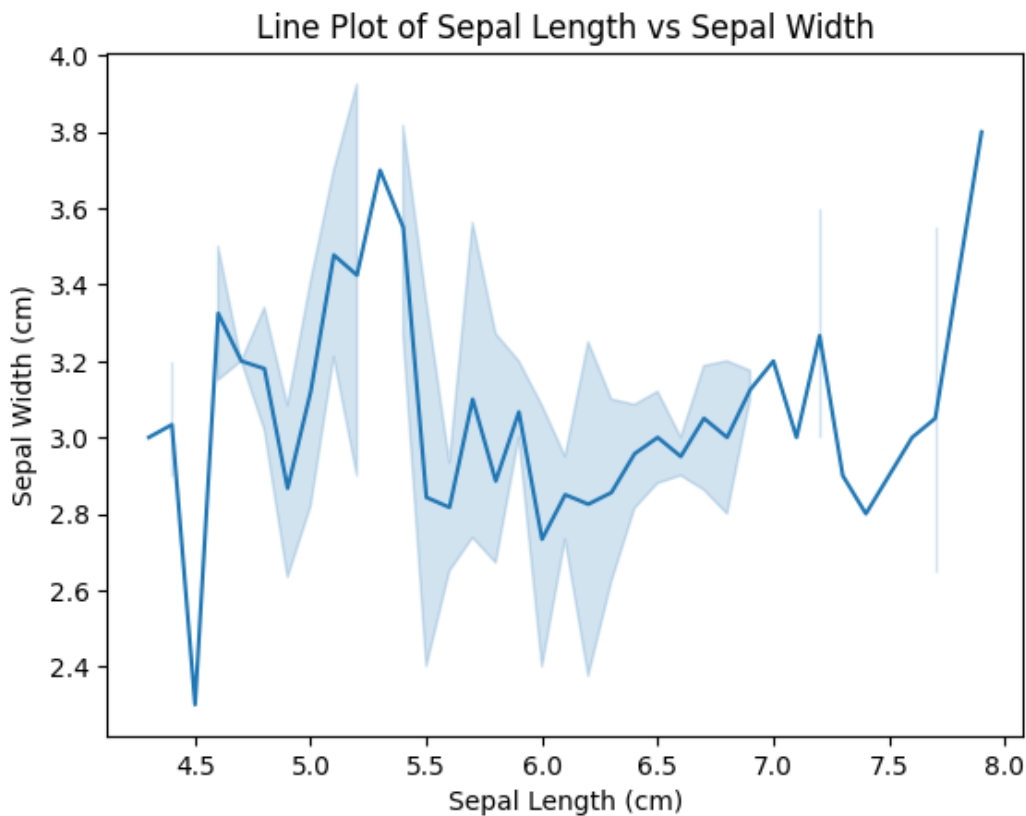
150 rows x 6 columns

In [308]:

```
# Check for inf values in the data
data.replace([np.inf, -np.inf], np.nan, inplace=True)

# Create the line plot
sns.lineplot(x='SepalLengthCm', y='SepalWidthCm', data=data)
plt.xlabel('Sepal Length (cm)')
```

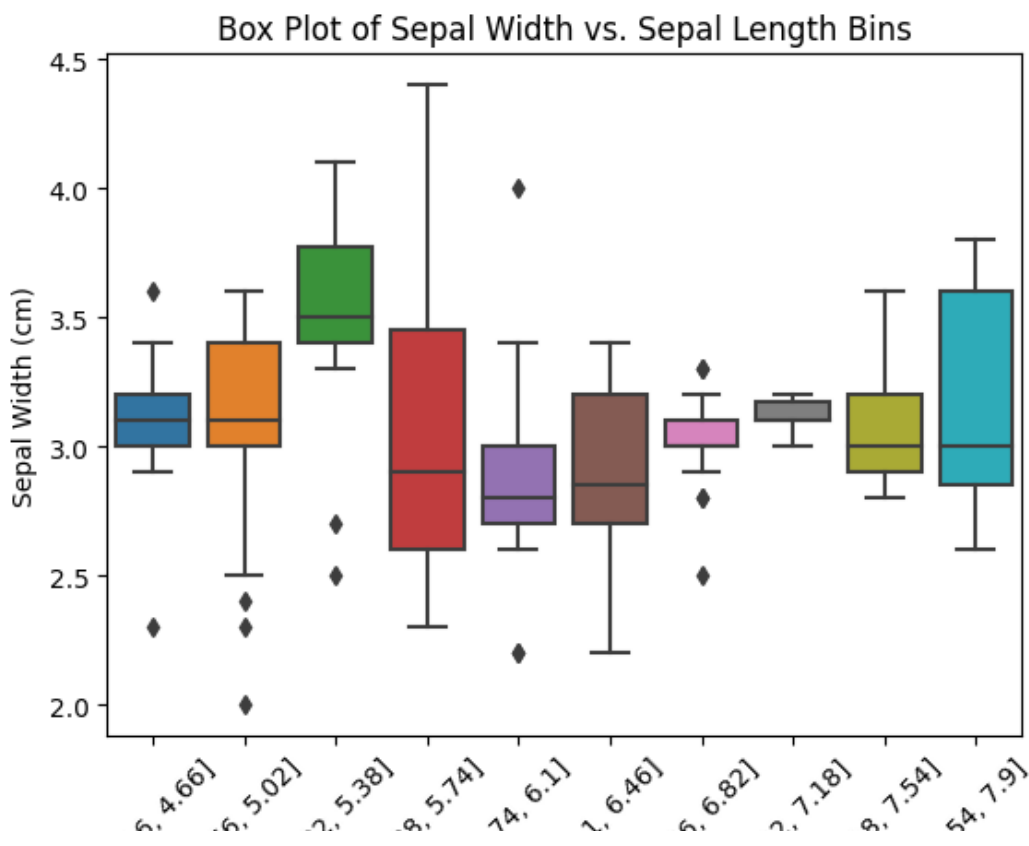
```
plt.ylabel('Sepal Width (cm)')
plt.title('Line Plot of Sepal Length vs Sepal Width')
plt.show()
```



In [309]:

```
# Bin SepalLengthCm into categories
bins = pd.cut(data['SepalLengthCm'], bins=10)

# Create the box plot
sns.boxplot(x=bins, y='SepalWidthCm', data=data)
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Sepal Width (cm)')
plt.title('Box Plot of Sepal Width vs. Sepal Length Bins')
plt.xticks(rotation=45)
plt.show()
```





Sepal Length (cm)

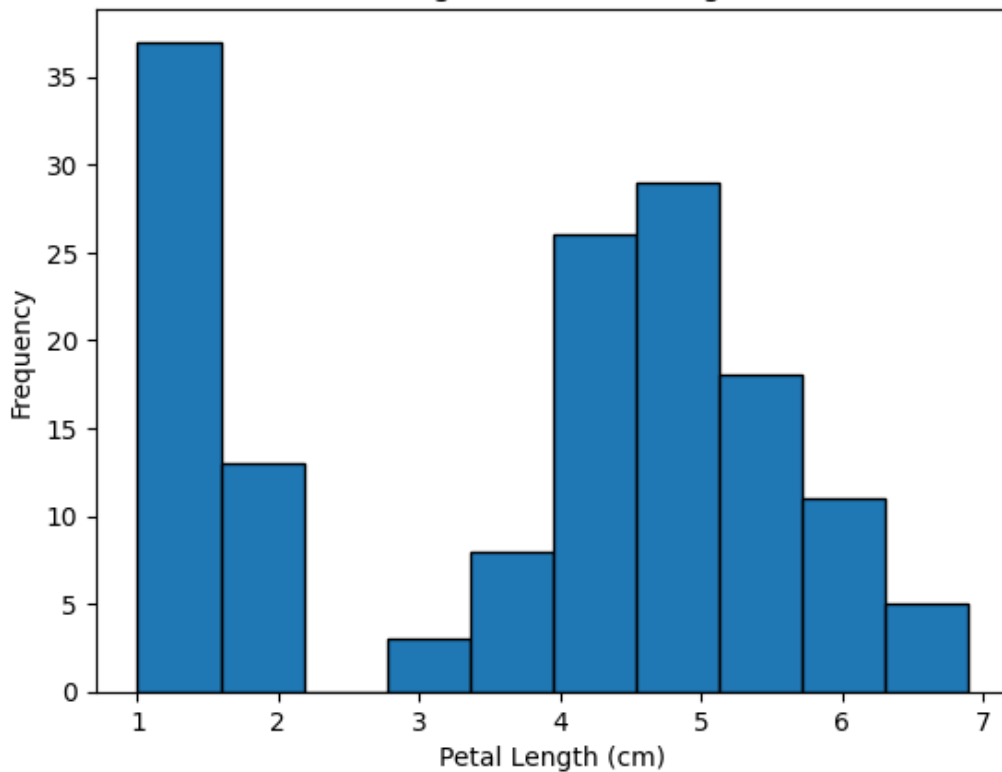
In [310]:

```
# Create the histogram
plt.hist(data['PetalLengthCm'], bins=10, edgecolor='black')

# Adding labels and title for better understanding
plt.xlabel('Petal Length (cm)')
plt.ylabel('Frequency')
plt.title('Histogram of Petal Length')

# Display the plot
plt.show()
```

Histogram of Petal Length



In [306]:

```
import matplotlib.pyplot as plt

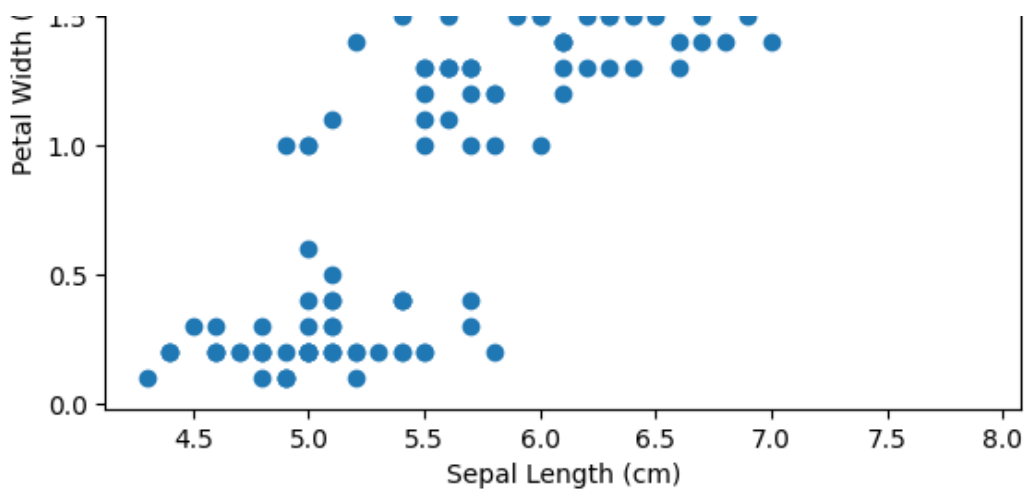
# Create the scatter plot
plt.scatter(data['SepalLengthCm'], data['PetalWidthCm'])

# Adding labels and title for better understanding
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Petal Width (cm)')
plt.title('Scatter Plot of Sepal Length vs Petal Width')

# Show the plot
plt.show()
```

Scatter Plot of Sepal Length vs Petal Width





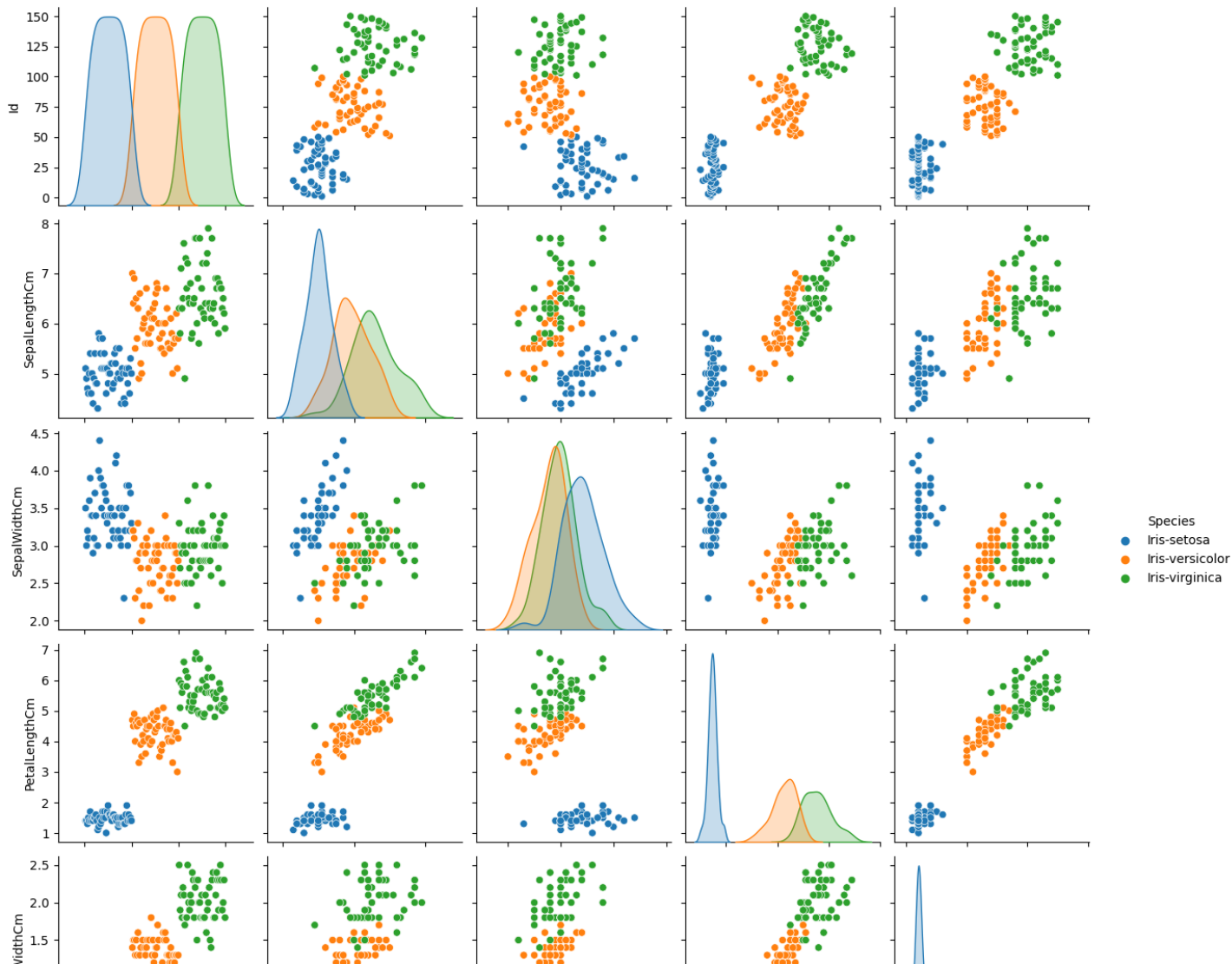
In [303]:

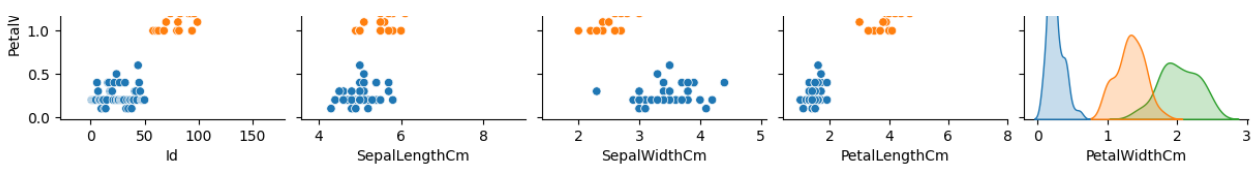
```
# Import necessary libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings

# Load the Iris dataset
data = pd.read_csv('/kaggle/input/iris/Iris.csv')

# Suppress specific warnings
warnings.filterwarnings("ignore", category=FutureWarning, module="pandas")
warnings.filterwarnings("ignore", category=FutureWarning, module="seaborn")

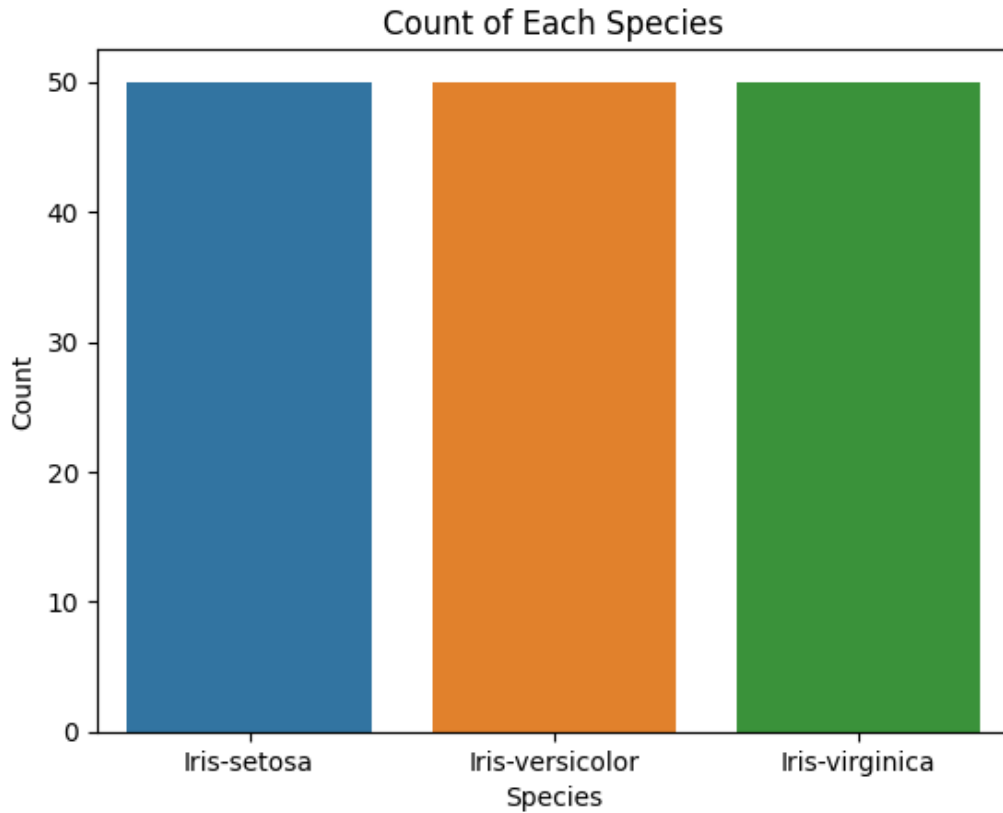
# Plot the pairplot
sns.pairplot(data, hue='Species')
plt.show()
```





In [305]:

```
# Plotting the count plot for Species
sns.countplot(x='Species', data=data)
plt.title('Count of Each Species')
plt.xlabel('Species')
plt.ylabel('Count')
plt.show()
```



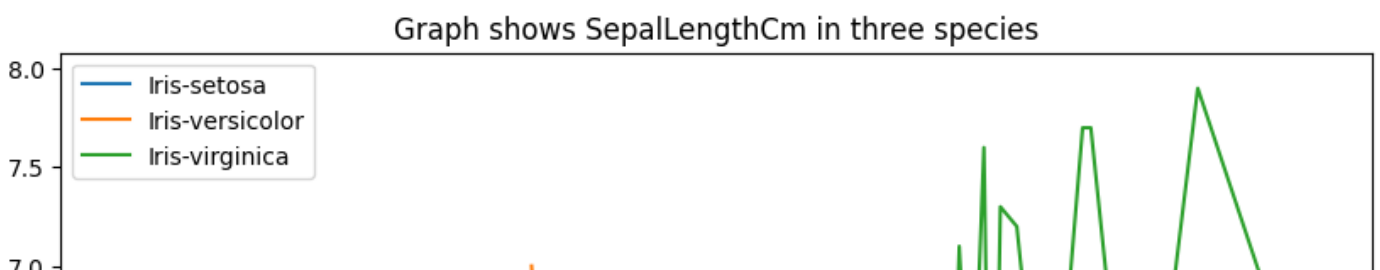
In [6]:

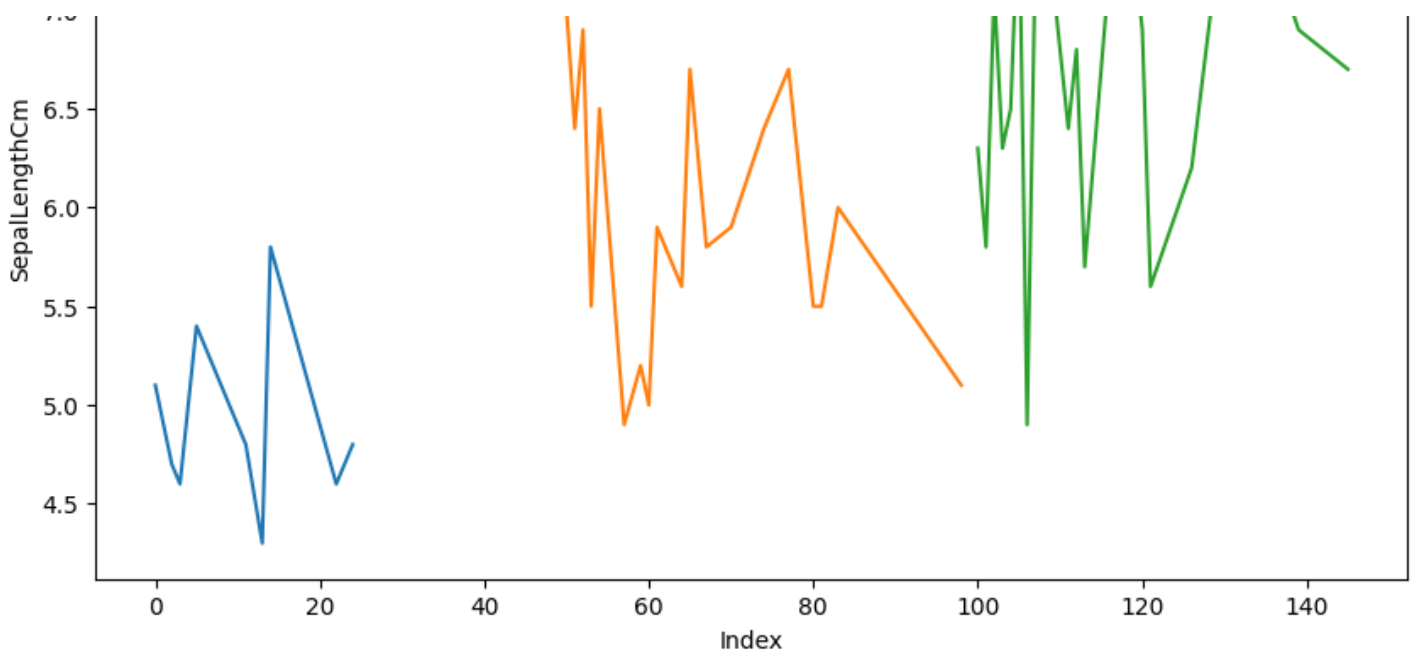
```
import matplotlib.pyplot as plt
# Plotting SepalLengthCm for each species
plt.figure(figsize=(10, 6))

# Plotting data for each species
for species in data['Species'].unique():
    subset = data[data['Species'] == species]
    plt.plot(subset['SepalLengthCm'], label=species)

# Adding title and labels
plt.title('Graph shows SepalLengthCm in three species')
plt.xlabel('Index')
plt.ylabel('SepalLengthCm')
plt.legend()

# Displaying the plot
plt.show()
```

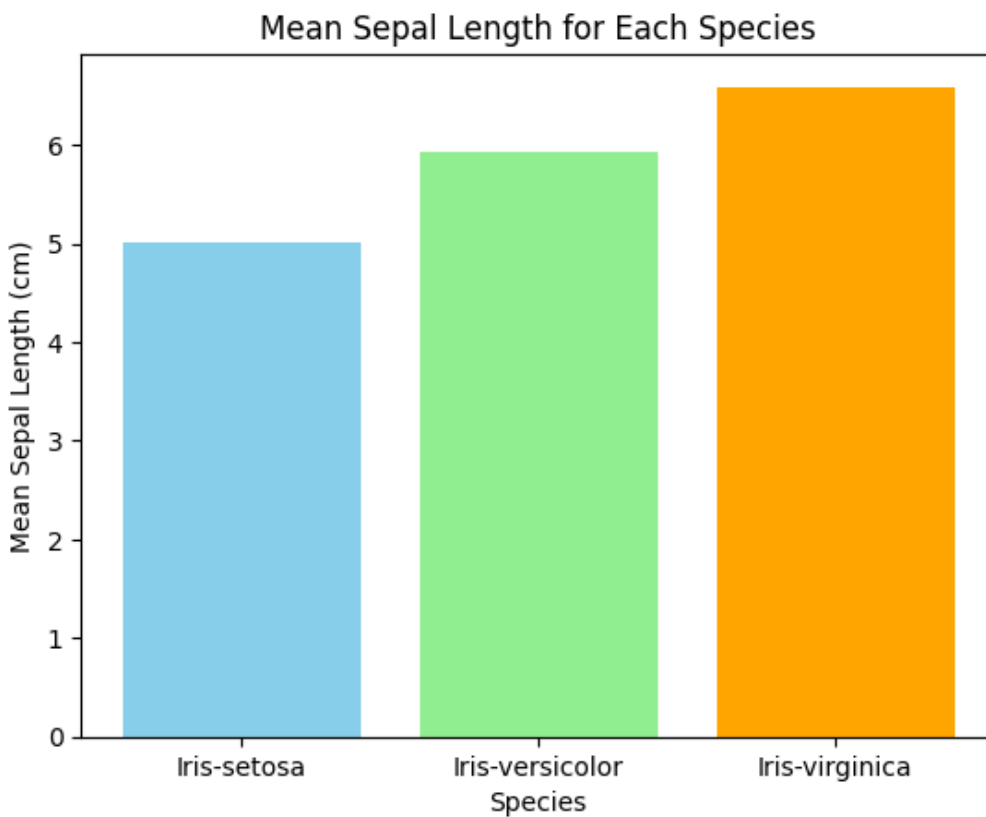




In [317]:

```
# Calculate the mean SepalLengthCm for each species
data_plot = data.groupby('Species')['SepalLengthCm'].mean().reset_index()

# Plot the data
plt.bar(data_plot['Species'], data_plot['SepalLengthCm'], color=['skyblue', 'lightgreen',
'orange'])
plt.xlabel('Species')
plt.ylabel('Mean Sepal Length (cm)')
plt.title('Mean Sepal Length for Each Species')
plt.show()
```



In [318]:

```
# Print the mean SepalLengthCm for each species
print(data.groupby('Species')['SepalLengthCm'].mean())

# Select the size of the plot
plt.figure(figsize=(5, 5))
```

```
# Add the title
plt.title('Average Sepal Length by Species')

# Create a bar plot
sns.barplot(x='Species', y='SepalLengthCm', data=data, ci=None)

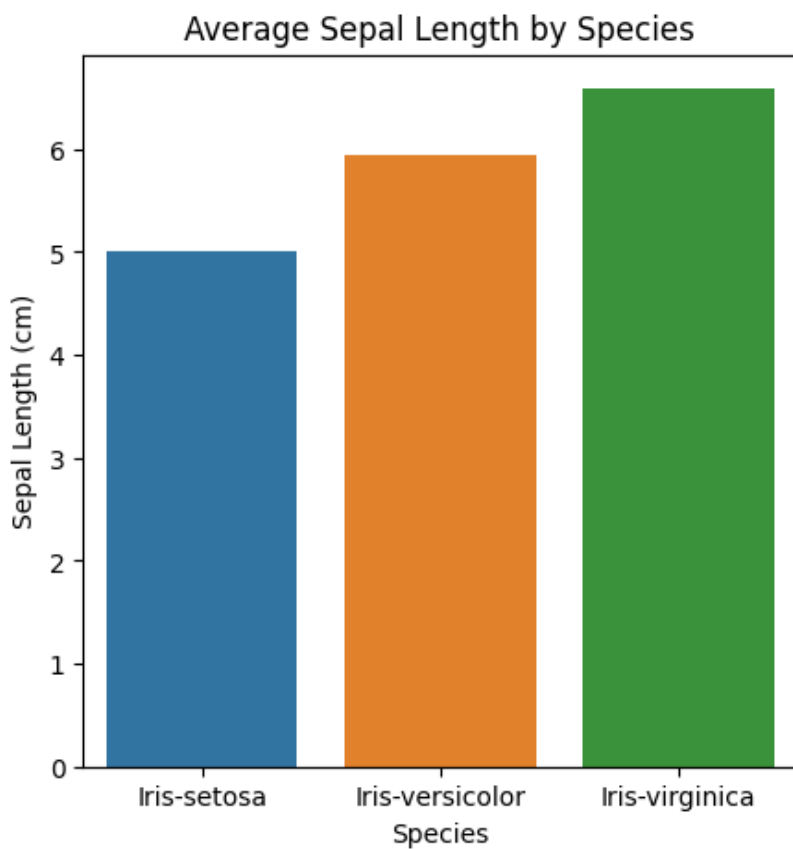
# Add the labels
plt.xlabel('Species')
plt.ylabel('Sepal Length (cm)')
plt.show()
```

```
Species
Iris-setosa      5.006
Iris-versicolor  5.936
Iris-virginica   6.588
Name: SepalLengthCm, dtype: float64
```

/tmp/ipykernel\_33/1831826488.py:11: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

```
sns.barplot(x='Species', y='SepalLengthCm', data=data, ci=None)
```



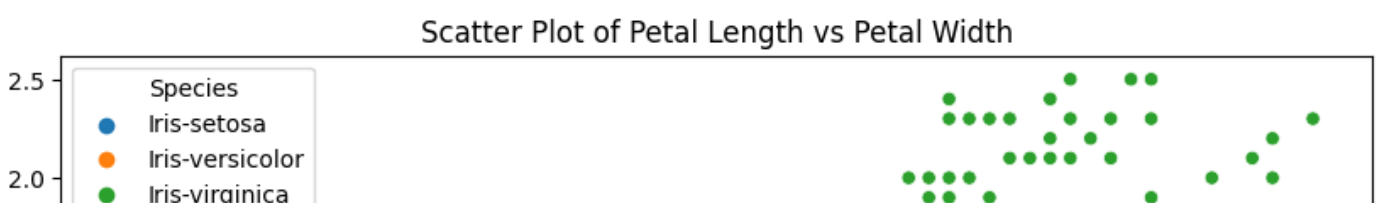
In [320]:

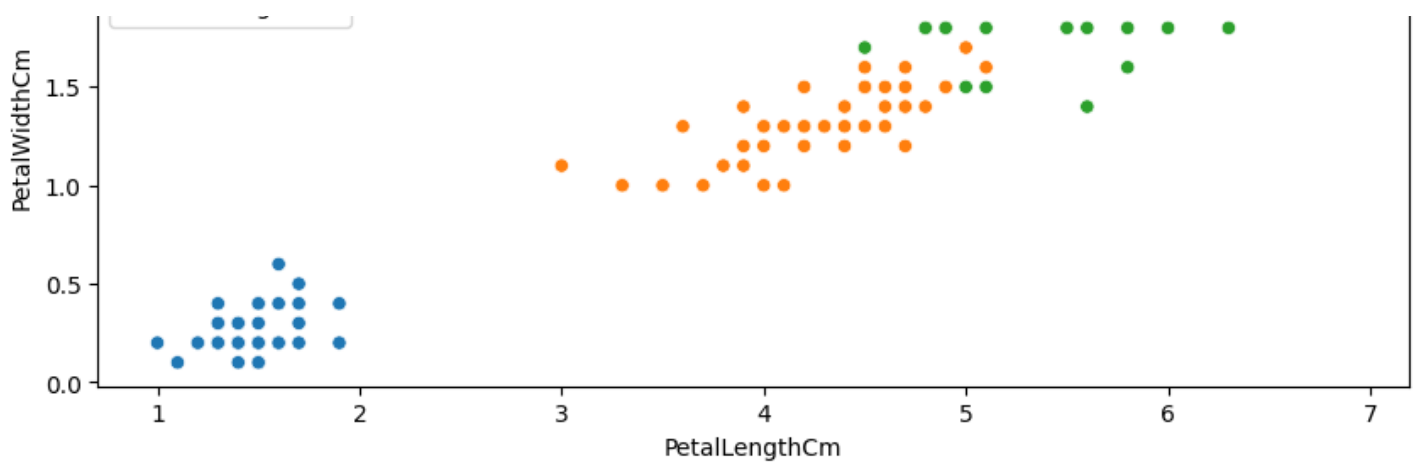
```
# Set the size of the plot
plt.figure(figsize=(10, 4))

# Create the scatter plot with Seaborn
sns.scatterplot(x=data['PetalLengthCm'], y=data['PetalWidthCm'], hue=data['Species'])

# Add title
plt.title('Scatter Plot of Petal Length vs Petal Width')

# Show the plot
plt.show()
```





In [321]:

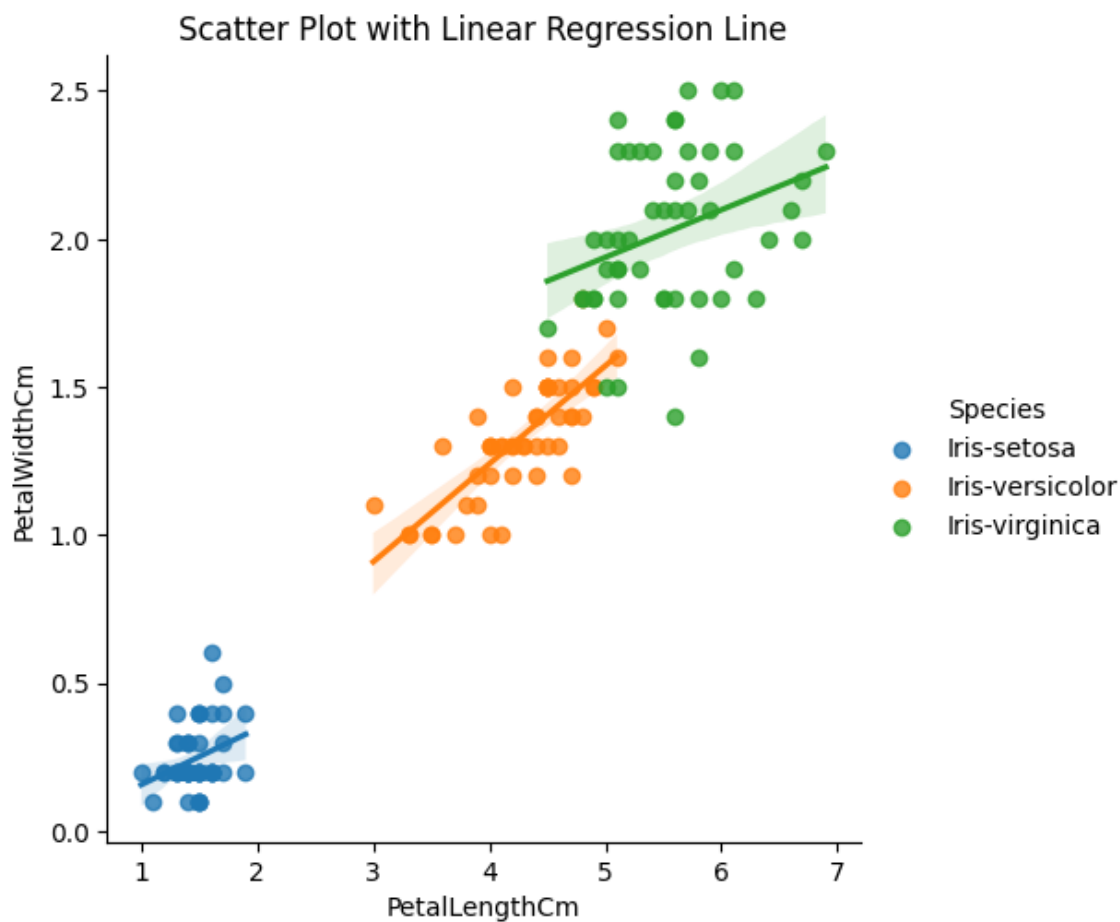
```
# Set the size of the plot
plt.figure(figsize=(10, 7))

# Create the lmplot with Seaborn
sns.lmplot(x='PetalLengthCm', y='PetalWidthCm', hue='Species', data=data)

# Add title
plt.title('Scatter Plot with Linear Regression Line')

# Show the plot
plt.show()
```

<Figure size 1000x700 with 0 Axes>



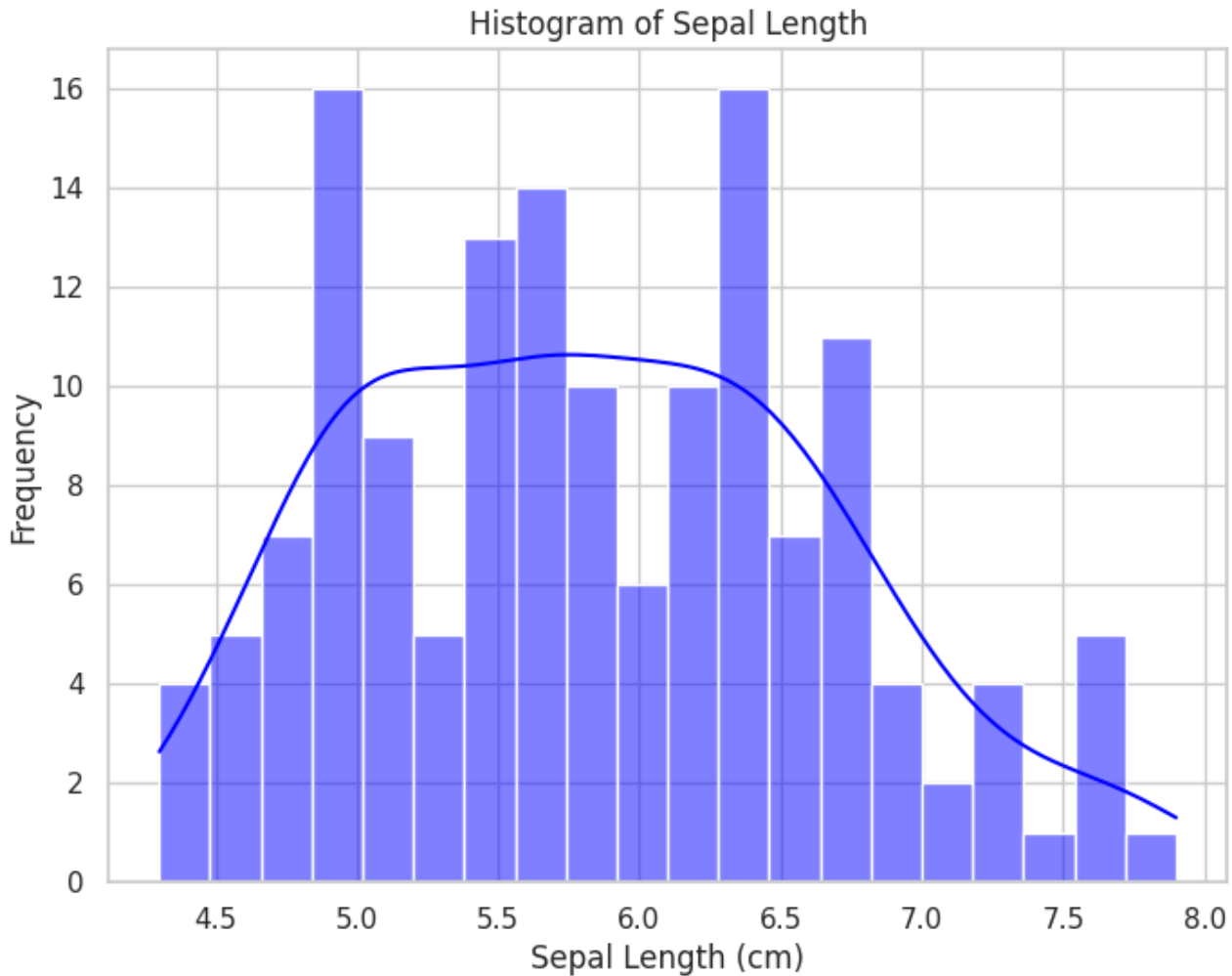
In [323]:

```
# Set the style and color palette (optional)
sns.set(style="whitegrid")

# Create the histogram using histplot
plt.figure(figsize=(8, 6))
sns.histplot(data['SepalLengthCm'], kde=True, color='blue', bins=20)
```

```
# Add labels and title
plt.xlabel('Sepal Length (cm)')
plt.ylabel('Frequency')
plt.title('Histogram of Sepal Length')
```

```
# Show the plot
plt.show()
```



In [292]:

```
df=pd.read_csv('/kaggle/input/iris/Iris.csv')
```

In [325]:

```
# Filter the data for each species
setosa = data[data['Species'] == 'Iris-setosa']
versicolor = data[data['Species'] == 'Iris-versicolor']
virginica = data[data['Species'] == 'Iris-virginica']

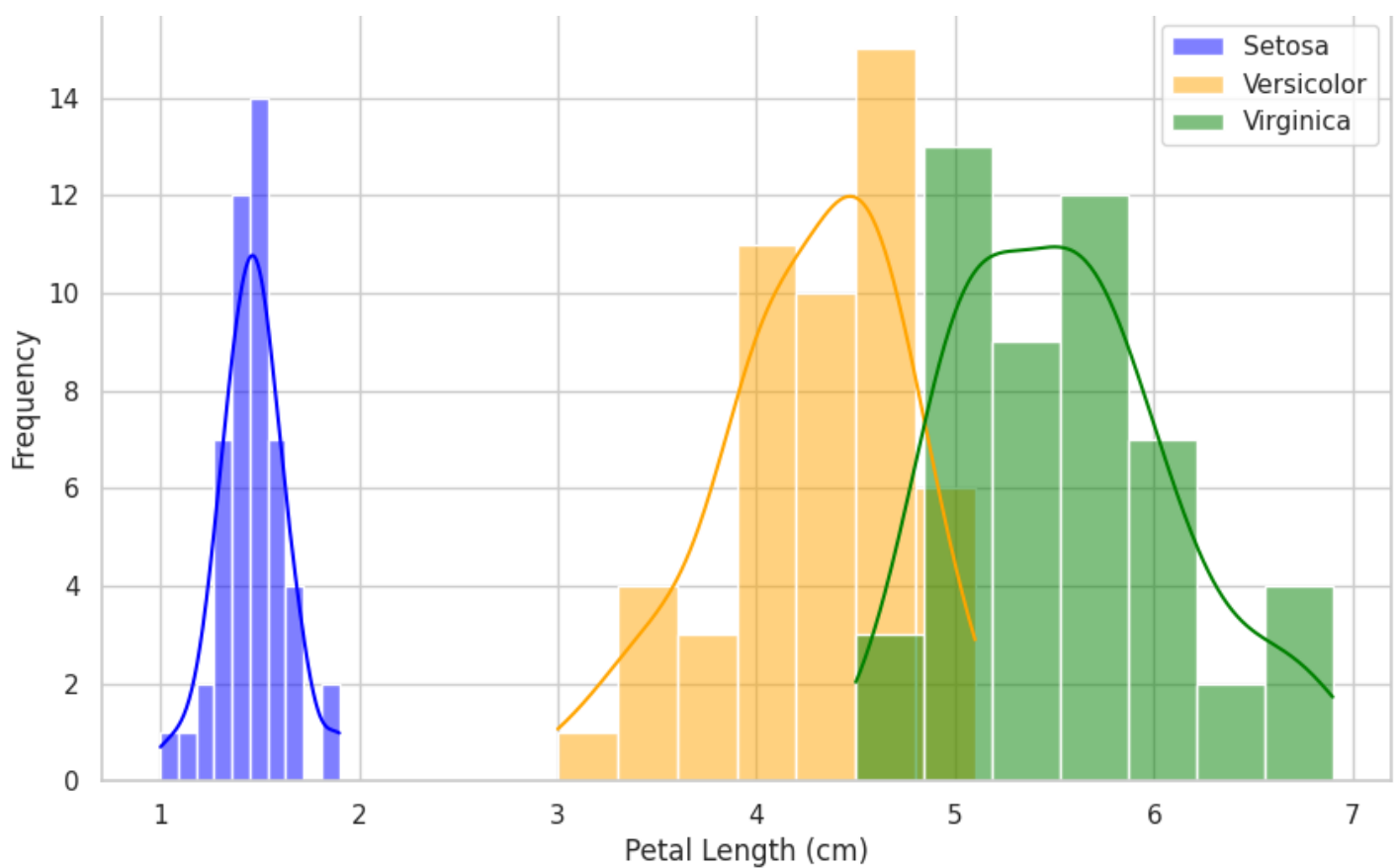
# Create histograms for each species
plt.figure(figsize=(10, 6))

sns.histplot(setosa['PetalLengthCm'], color='blue', label='Setosa', kde=True)
sns.histplot(versicolor['PetalLengthCm'], color='orange', label='Versicolor', kde=True)
sns.histplot(virginica['PetalLengthCm'], color='green', label='Virginica', kde=True)

# Add title and legend
plt.title('Histogram of Petal Length by Species')
plt.xlabel('Petal Length (cm)')
plt.ylabel('Frequency')
plt.legend()

# Show the plot
plt.show()
```

Histogram of Petal Length by Species



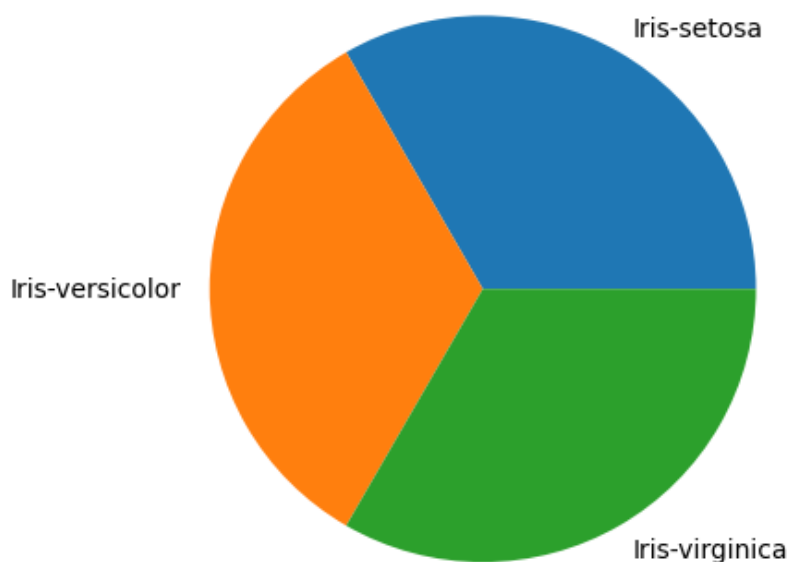
In [300]:

```
# pie chart
import matplotlib.pyplot as plt
labels = 'Iris-setosa', 'Iris-versicolor', 'Iris-virginica'
sizes = [50, 50, 50]

fig, ax = plt.subplots()
ax.pie(sizes, labels=labels)
```

Out[300]:

```
([<matplotlib.patches.Wedge at 0x7a9c971576d0>,
 <matplotlib.patches.Wedge at 0x7a9c9130beb0>,
 <matplotlib.patches.Wedge at 0x7a9c971564d0>],
 [Text(0.5499999702695115, 0.9526279613277875, 'Iris-setosa'),
 Text(-1.0999999999999954, -1.0298943258065002e-07, 'Iris-versicolor'),
 Text(0.5500001486524352, -0.9526278583383436, 'Iris-virginica')])
```





In [328]:

```
# Create a pie chart
fig, ax = plt.subplots()
ax.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=90)
ax.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.

plt.title('Distribution of Iris Species')

plt.show()
```

