2) On the iris dataset, perform KNN algorithm and discuss result

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

In [2]:
```
iris = pd.read_csv("iris.csv")
iris.dtypes
```

Out[2]:
```
Unnamed: 0        int64
Sepal.Length    float64
Sepal.Width     float64
Petal.Length    float64
Petal.Width     float64
Species          object
dtype: object
```

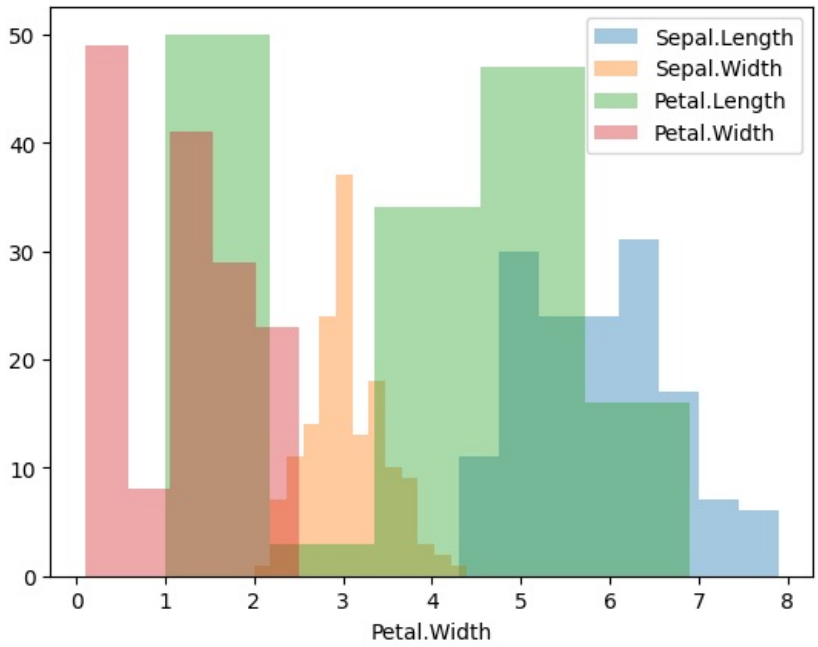
In [3]:
```
iris.columns
```

Out[3]:
```
Index(['Unnamed: 0', 'Sepal.Length', 'Sepal.Width', 'Petal.Length',
       'Petal.Width', 'Species'],
      dtype='object')
```

In [4]:
```
import matplotlib.pyplot as plt # mostly used for visualization purposes
import numpy as np
import seaborn as sns

sns.distplot(iris['Sepal.Length'], kde=False,label='Sepal.Length')
sns.distplot(iris['Sepal.Width'], kde=False,label='Sepal.Width')
sns.distplot(iris['Petal.Length'], kde=False,label='Petal.Length')
sns.distplot(iris['Petal.Width'], kde=False,label='Petal.Width')
plt.legend()
```

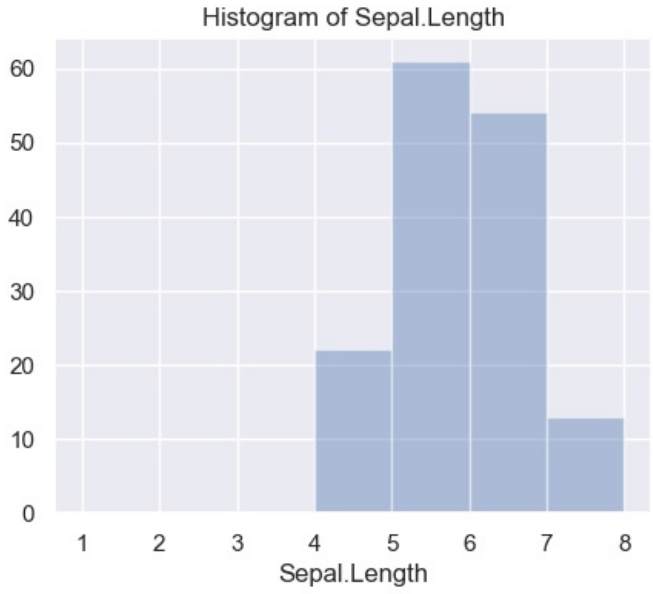
```
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\3255355497.py:5: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris['Sepal.Length'], kde=False,label='Sepal.Length')
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\3255355497.py:6: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris['Sepal.Width'], kde=False,label='Sepal.Width')
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\3255355497.py:7: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris['Petal.Length'], kde=False,label='Petal.Length')
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\3255355497.py:8: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris['Petal.Width'], kde=False,label='Petal.Width')
```

Out[4]:
```
<matplotlib.legend.Legend at 0x184b641ece0>
```



In [5]:
```
# 'Sepal.Length'
bins = [1,2,3,4,5,6,7,8]
plt.figure(figsize=(5,4))
sns.set() # light color background
sns.distplot(iris["Sepal.Length"],bins = bins, kde=False)
plt.xticks(bins) # x-axis (1-8)
plt.title("Histogram of Sepal.Length")
plt.show()
iris["Sepal.Length"].value_counts()
```

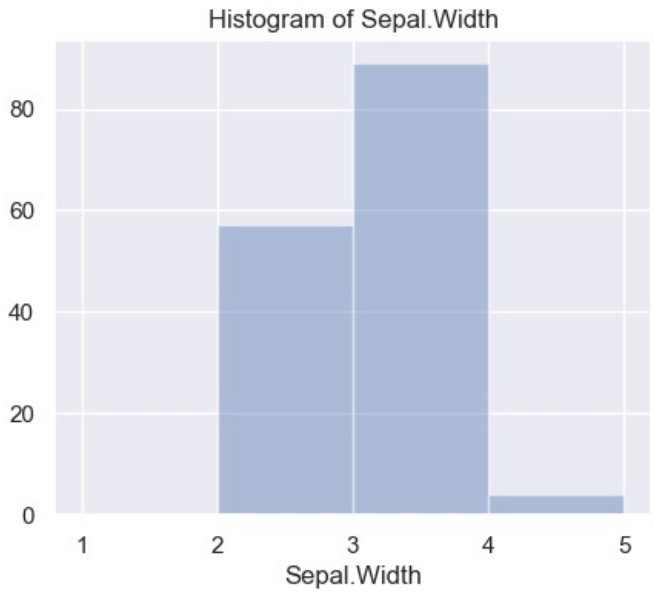
```
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\2154069552.py:5: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris["Sepal.Length"],bins = bins, kde=False)
```



Out[5]:
```
Sepal.Length
5.0    10
5.1     9
6.3     9
5.7     8
6.7     8
5.8     7
5.5     7
6.4     7
4.9     6
5.4     6
6.1     6
6.0     6
5.6     6
4.8     5
6.5     5
6.2     4
7.7     4
6.9     4
4.6     4
5.2     4
5.9     3
4.4     3
7.2     3
6.8     3
6.6     2
4.7     2
7.6     1
7.4     1
7.3     1
7.0     1
7.1     1
5.3     1
4.3     1
4.5     1
7.9     1
Name: count, dtype: int64
```

In [6]:
```
# 'Sepal.Width'
bins = [1,2,3,4,5]
plt.figure(figsize=(5,4))
sns.set() # light color background
sns.distplot(iris["Sepal.Width"],bins = bins, kde=False)
plt.xticks(bins) # x-axis (1-8)
plt.title("Histogram of Sepal.Width")
plt.show()
iris["Sepal.Width"].value_counts()
```

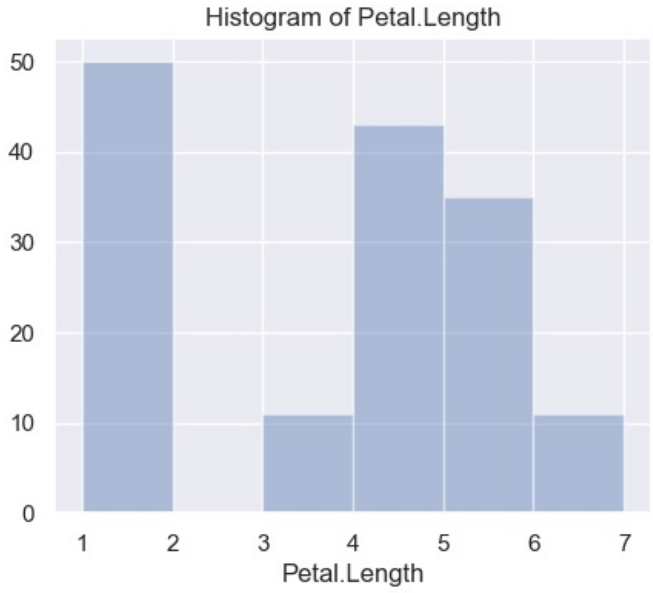
```
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\975662442.py:5: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris["Sepal.Width"],bins = bins, kde=False)
```



Out[6]:
```
Sepal.Width
3.0    26
2.8    14
3.2    13

3.4    12
3.1    11
2.9    10
2.7     9
2.5     8
3.5     6
3.3     6
3.8     6
2.6     5
2.3     4
3.6     4
3.7     3
2.4     3
2.2     3
3.9     2
4.4     1
4.0     1
4.1     1
4.2     1
2.0     1
Name: count, dtype: int64
```

In [7]:
```
# 'Petal.Length'
bins = [1,2,3,4,5,6,7]
plt.figure(figsize=(5,4))
sns.set() # light color background
sns.distplot(iris["Petal.Length"],bins = bins, kde=False)
plt.xticks(bins) # x-axis (1-8)
plt.title("Histogram of Petal.Length")
plt.show()
iris["Petal.Length"].value_counts()
```

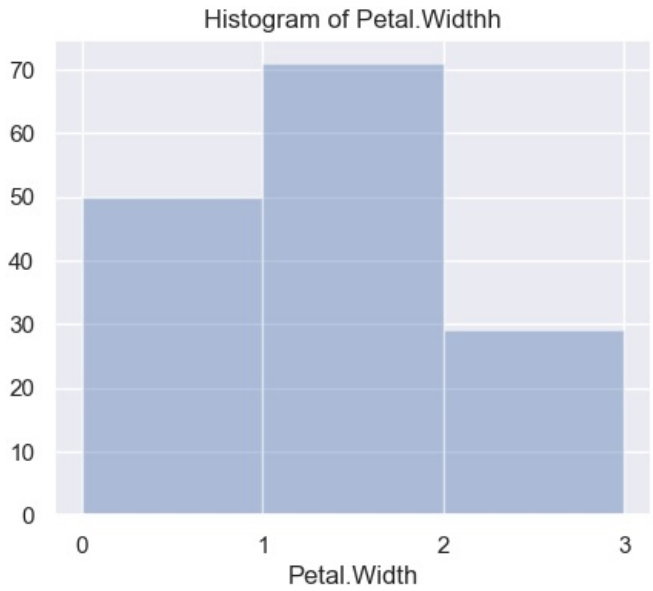
```
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\1909035521.py:5: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris["Petal.Length"],bins = bins, kde=False)
```



Out[7]:
```
Petal.Length
1.4    13
1.5    13
5.1     8
4.5     8
1.6     7
1.3     7
5.6     6
4.7     5
4.9     5
4.0     5
4.2     4
5.0     4
4.4     4
4.8     4
1.7     4

3.9     3
4.6     3
5.7     3
4.1     3
5.5     3
6.1     3
5.8     3
3.3     2
5.4     2
6.7     2
5.3     2
5.9     2
6.0     2
1.2     2
4.3     2
1.9     2
3.5     2
5.2     2
3.0     1
1.1     1
3.7     1
3.8     1
6.6     1
6.3     1
1.0     1
6.9     1
3.6     1
6.4     1
Name: count, dtype: int64
```

In [8]:
```
#'Petal.Width'
bins = [0,1,2,3]
plt.figure(figsize=(5,4))
sns.set() # light color background
sns.distplot(iris["Petal.Width"],bins = bins, kde=False)
plt.xticks(bins) # x-axis (1-8)
plt.title("Histogram of Petal.Widthh")
plt.show()
iris["Petal.Width"].value_counts()
```

```
C:\Users\Mayuri\AppData\Local\Temp\ipykernel_4508\1194133118.py:5: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(iris["Petal.Width"],bins = bins, kde=False)
```



Out[8]:

```
Petal.Width
0.2    29
1.3    13
1.8    12
1.5    12
1.4     8
2.3     8
1.0     7
0.4     7
0.3     7
2.1     6
2.0     6
0.1     5
1.2     5
1.9     5
1.6     4
2.5     3
2.2     3
2.4     3
1.1     3
1.7     2
0.6     1
0.5     1
Name: count, dtype: int64
```

In [14]:

```python
#preparing data with scaling
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler

feature_names = ['Sepal.Length', 'Sepal.Width', 'Petal.Length', 'Petal.Width']
x=iris[feature_names]
y=iris['Species']

x_train, x_test, y_train, y_test = train_test_split(x,y, random_state=0)

print(x_train[:3]) # to check output

scaler = MinMaxScaler()
x_train=scaler.fit_transform(x_train)
x_test= scaler.transform(x_test)
```

```
print("\nAfter scaling\n")
print(x_train[:3]) # to check output
```

```
     Sepal.Length  Sepal.Width  Petal.Length  Petal.Width
61            5.9          3.0           4.2          1.5
92            5.8          2.6           4.0          1.2
112           6.8          3.0           5.5          2.1

After scaling

[[0.44444444 0.41666667 0.53448276 0.58333333]
 [0.41666667 0.25       0.5        0.45833333]
 [0.69444444 0.41666667 0.75862069 0.83333333]]
```

In [17]:

```
from sklearn.neighbors import KNeighborsClassifier

# KNN method
knn = KNeighborsClassifier()
knn.fit(x_train, y_train)

#print score of train data
print('Accuracy of KNN classifier on training set:{:.2f}'
     .format(knn.score(x_train, y_train)))

#print score of test data
print('Accuracy of KNN Classifier on test set:{:.2f}'
     .format(knn.score(x_test, y_test)))
```

```
Accuracy of KNN classifier on training set:0.96
Accuracy of KNN Classifier on test set:0.97
```

In [ ]: