In [13]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.spatial.distance import cdist
```

In [14]:
```
iris = pd.read_csv("iris.csv",header=None)
```

In [15]:
```
iris.columns=['SepalLength','SepalWidth','PetalLength','PetalWidth','Species']
```

In [16]:
```
X=iris[['SepalLength','SepalWidth','PetalLength','PetalWidth']]
```

In [17]:
```
Y=iris['Species']
```

In [18]:
```
print(X.head())
```

```
   SepalLength  SepalWidth  PetalLength  PetalWidth
0          5.1         3.5          1.4         0.2
1          4.9         3.0          1.4         0.2
2          4.7         3.2          1.3         0.2
3          4.6         3.1          1.5         0.2
4          5.0         3.6          1.4         0.2
```

In [19]:
```
def k_means(X,K):
    centroids_history=[]
    labels_history=[]
    rand_index=np.random.choice(X.shape[0],K)
    centroids=X[rand_index]
    centroids_history.append(centroids)
    while True:
        labels=np.argmin(cdist(X, centroids), axis=1)
        labels_history.append(labels)
        new_centroids=np.array([X[labels==i].mean(axis=0)for i in range(K)])
        centroids_history.append(new_centroids)
        if np.all(centroids==new_centroids):
            break
        centroids=new_centroids
    return centroids,labels,centroids_history,labels_history
```

In [20]:
```
X_mat=X.values
```

In [21]:
```
centroids,labels,centroids_history,label_history=k_means(X_mat,3)
```

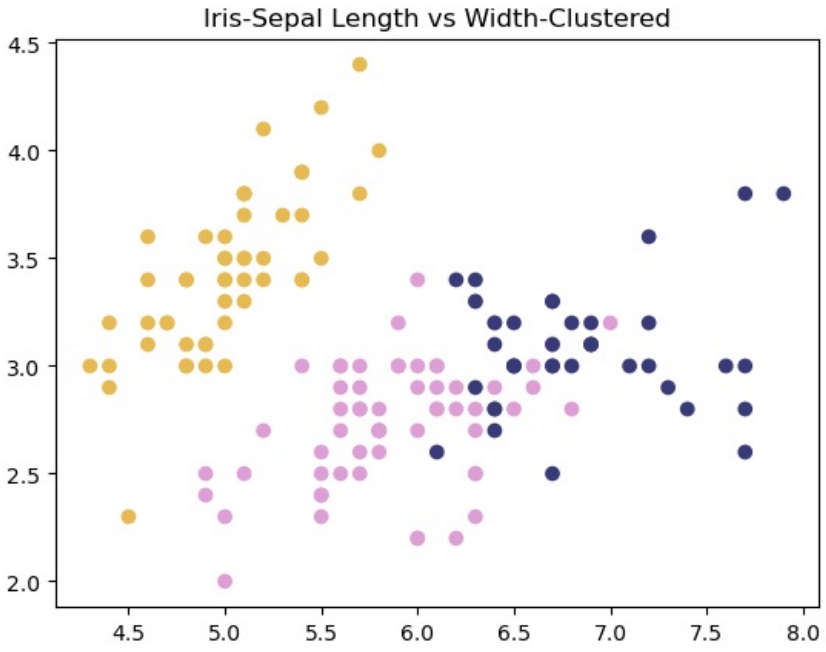
In [22]:
```
print("Samples Labels")
```

```
Samples Labels
```

In [23]:
```
print(labels)
```

```
[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 2 2 2 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 0 2 0 0 0 0 2 0 0 0 0
 0 0 2 2 0 0 0 0 2 0 2 0 2 0 0 2 2 0 0 0 0 0 2 0 0 0 0 2 0 0 0 2 0 0 0 2 0
 0 2]
```

In [28]:
```
plt.scatter(X['SepalLength'],X['SepalWidth'],c=labels,cmap='tab20b')
plt.title('Iris-Sepal Length vs Width-Clustered')
plt.show()
```



In [ ]: