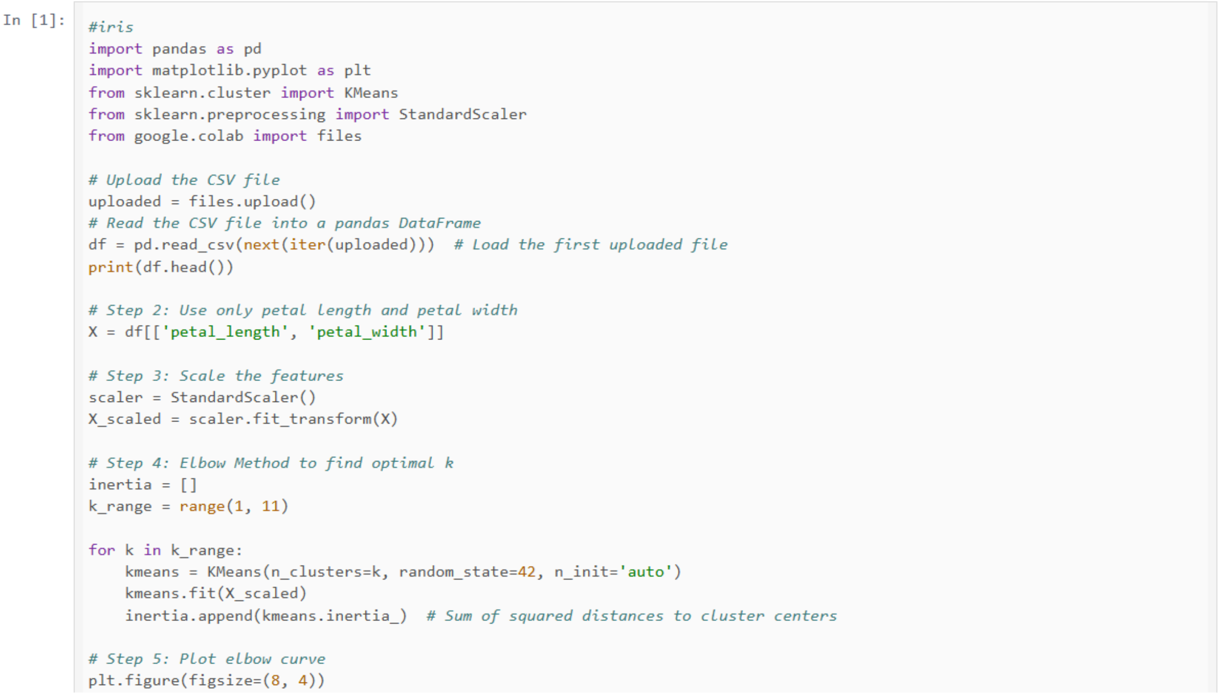
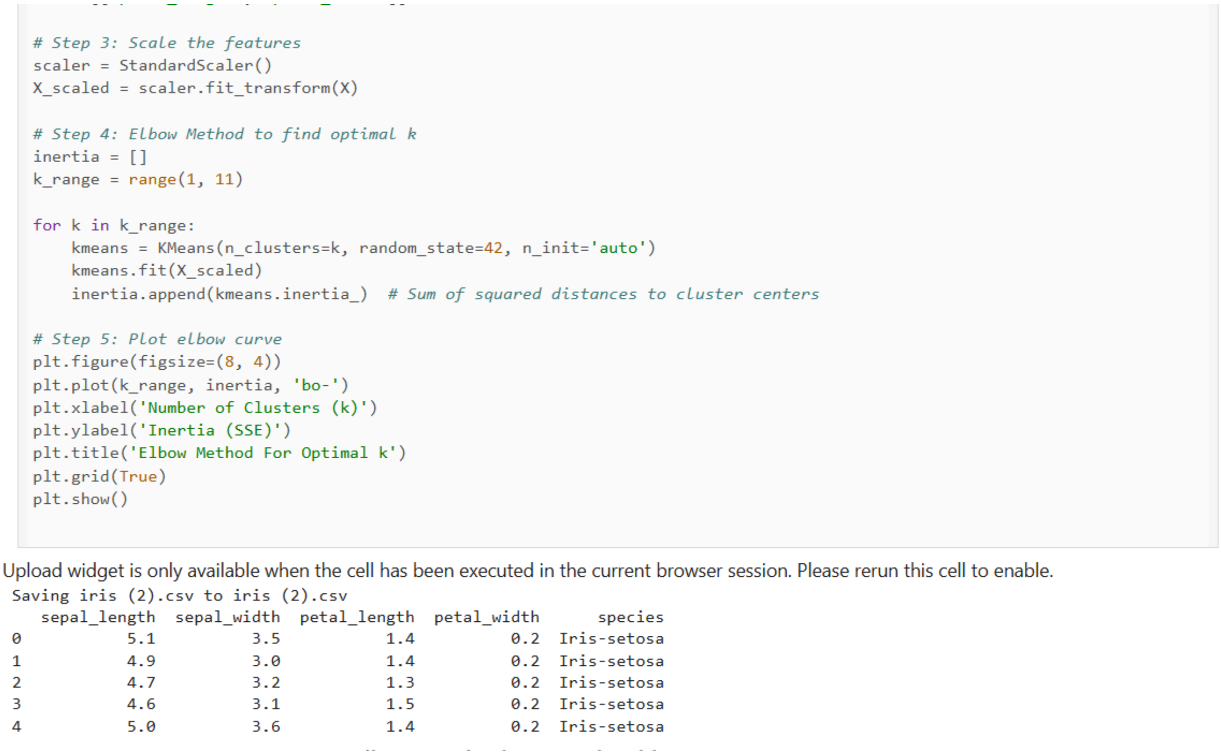
LAB 9
K- MEANS CLUSTURING;



In [1]:

```
#iris
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from google.colab import files

# Upload the CSV file
uploaded = files.upload()
# Read the CSV file into a pandas DataFrame
df = pd.read_csv(next(iter(uploaded)))  # Load the first uploaded file
print(df.head())

# Step 2: Use only petal length and petal width
X = df[['petal_length', 'petal_width']]

# Step 3: Scale the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Step 4: Elbow Method to find optimal k
inertia = []
k_range = range(1, 11)

for k in k_range:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init='auto')
    kmeans.fit(X_scaled)
    inertia.append(kmeans.inertia_)  # Sum of squared distances to cluster centers

# Step 5: Plot elbow curve
plt.figure(figsize=(8, 4))
```

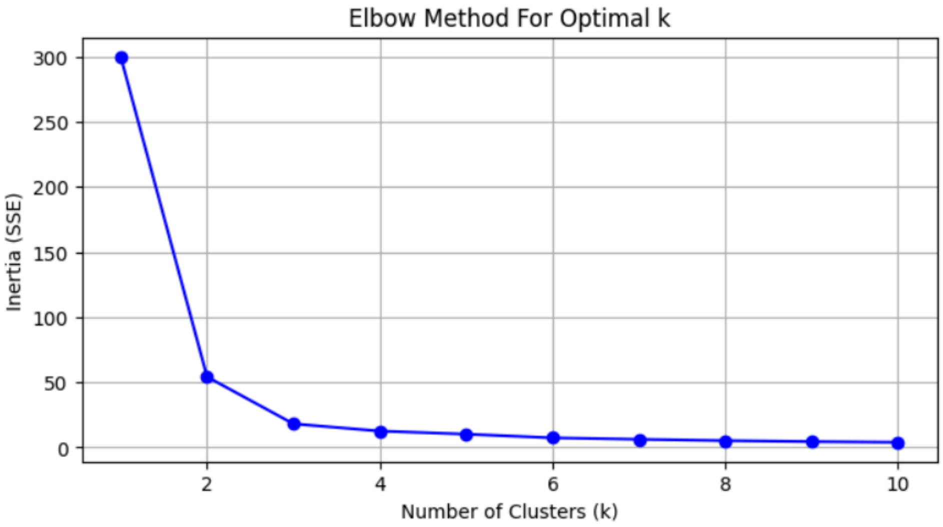


```
# Step 3: Scale the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Step 4: Elbow Method to find optimal k
inertia = []
k_range = range(1, 11)

for k in k_range:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init='auto')
    kmeans.fit(X_scaled)
    inertia.append(kmeans.inertia_)  # Sum of squared distances to cluster centers

# Step 5: Plot elbow curve
plt.figure(figsize=(8, 4))
```

```
plt.plot(k_range, inertia, 'bo-')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Inertia (SSE)')
plt.title('Elbow Method For Optimal k')
plt.grid(True)
plt.show()
```

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving iris (2).csv to iris (2).csv
   sepal_length  sepal_width  petal_length  petal_width      species
0           5.1          3.5           1.4          0.2  Iris-setosa
1           4.9          3.0           1.4          0.2  Iris-setosa
2           4.7          3.2           1.3          0.2  Iris-setosa
3           4.6          3.1           1.5          0.2  Iris-setosa
4           5.0          3.6           1.4          0.2  Iris-setosa
```


Elbow Method For Optimal k
Inertia (SSE)
300
250
200
150
100
50
0
2
4
6
8
10
Number of Clusters (k)

12/5/25
Monday

Lab - 9

Build K-means algorithm to cluster a set of data stored in a. CSV file.

R.No	A	B
R1	1	1
R2	1.5	2
R3	3	4
R4	5	7
R5	3.5	5
R6	4.5	5
R7.	3.5	4.5

cluster centers → (1, 1), (5,7)

Iteration 1

Datapoint	(1,1) cluster 1	(5,7) cluster 2	Assign
1.1	0	$\sqrt{52}$	1
1.5, 2	$\sqrt{1.25}$	$\sqrt{37.25}$	1
3, 4	$\sqrt{13}$	$\sqrt{13}$	1
5, 7	$\sqrt{52}$	0	2
3.5, 5	$\sqrt{22.25}$	$\sqrt{6.25}$	2
4.5, 5	$\sqrt{25.25}$	$\sqrt{4.25}$	2
3.5, 4.5	$\sqrt{18.5}$	$\sqrt{8.5}$	2

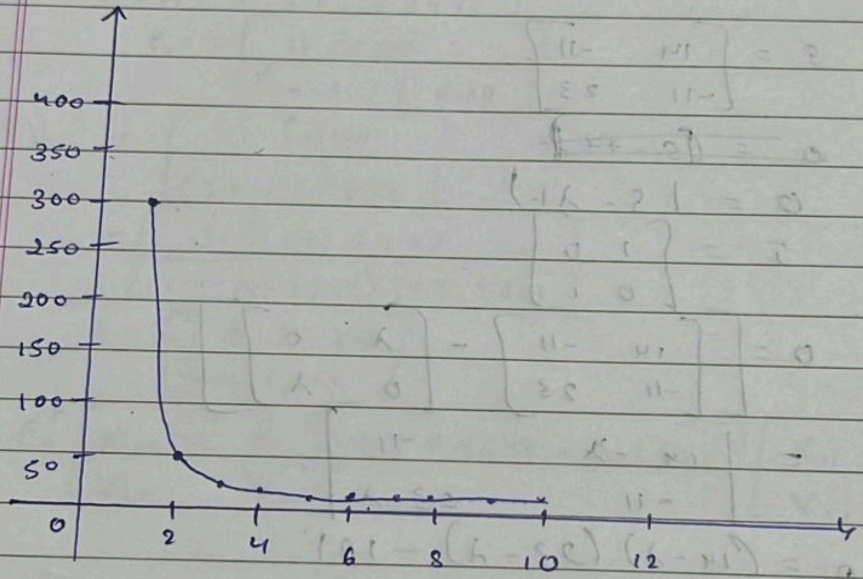
Culter 1	Cluster 2
(1, 1)	(5,7)
(1.5, 2)	(3.5, 5)
(3, 4)	(4.5, 5)
	(3.5, 4.5)

Average (1.83, 2.33) ; (4,125, 5.375)

Iteration 2

Datapoint	Cluster 1 (1.83 , 2.33)	Cluster 2 (4.125, 5.375)	Assign
1.1	1.56	5.36	1
1.5, 2	0.46	4.26	1
3, 4	2.006	1.76	2
5, 7	5.64	1.85	2
3.5, 5	3.149	0.722	2
4.5, 5	3.77	0.53	2
3.5, 4.5	2.73	1.06	2

=> iris.csv



The optimal value for K is 3