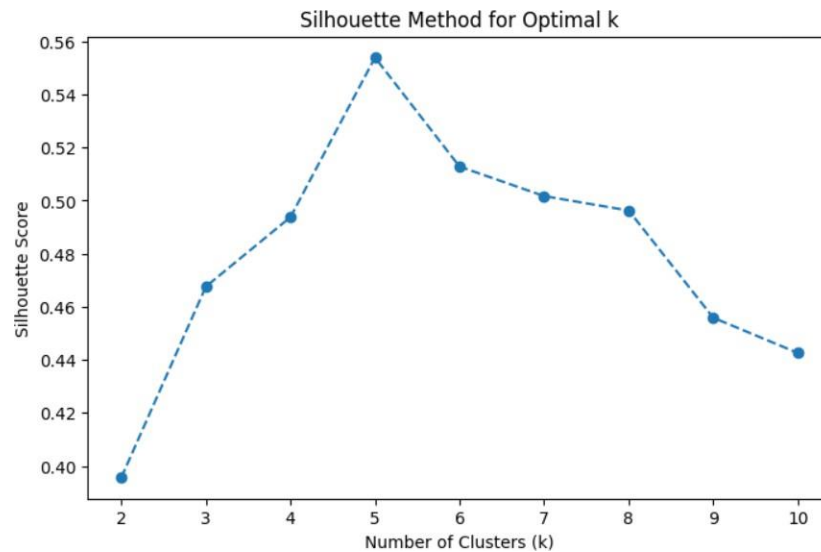


## ASSIGNMENT NO 5

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
df = pd.read_csv(r'Datasets\Mall_Customers.csv')
df.head()
```

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

```
X = df.iloc[:, [3, 4]].values
silhouette_scores = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters=k, init="k-means++", random_state=42)
    labels = kmeans.fit_predict(X)
    score = silhouette_score(X, labels)
    silhouette_scores.append(score)
plt.figure(figsize=(8, 5))
plt.plot(range(2, 11), silhouette_scores, marker='o', linestyle='--')
plt.xlabel("Number of Clusters (k)")
plt.ylabel("Silhouette Score")
plt.title("Silhouette Method for Optimal k")
plt.show()
```



```

optimal_k = np.argmax(silhouette_scores) + 2

kmeans = KMeans(n_clusters=optimal_k, init="k-means++", random_state=42)

y_kmeans = kmeans.fit_predict(X)

plt.figure(figsize=(8, 5))

colors = ['red', 'blue', 'green', 'grey', 'magenta', 'orange', 'purple', 'brown', 'pink']

for i in range(optimal_k):

    plt.scatter(X[y_kmeans == i, 0], X[y_kmeans == i, 1],

                s=100, c=colors[i], label=f'Cluster {i+1}')

# Plot the centroids

plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1],

            s=500, c='black', marker='*', label='Centroids')

plt.xlabel("Annual Income (k$)")

plt.ylabel("Spending Score (1-100)")

plt.title(f"Mall Customer Segmentation (k={optimal_k})")

```

```
plt.legend()
```

```
plt.show()
```

