

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
```

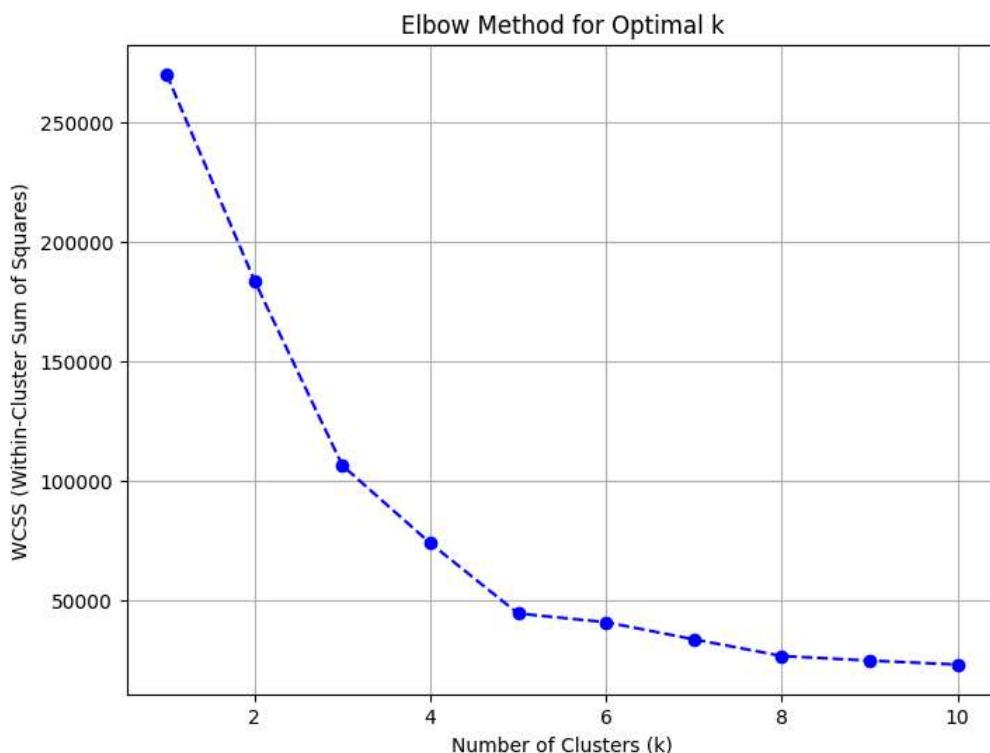
```
In [2]: df = pd.read_csv('Mall_Customers.csv')
df.head()
```

```
Out[2]:   CustomerID  Genre  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
```

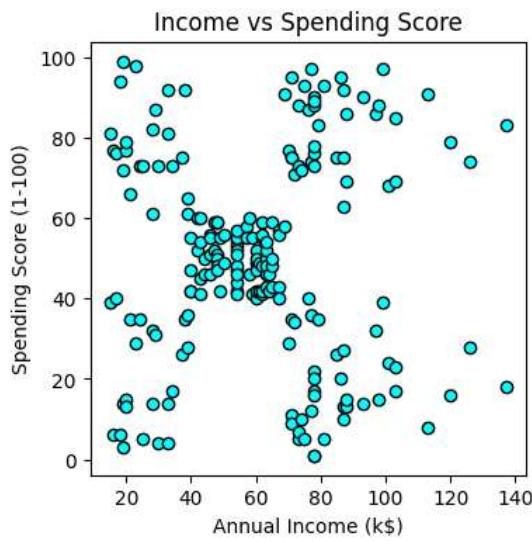
```
In [3]: X = df[['Annual Income (k$)', 'Spending Score (1-100)']]

wcss = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
```

```
In [4]: plt.figure(figsize=(8, 6))
plt.plot(range(1, 11), wcss, marker='o', linestyle='--', color='b')
plt.title('Elbow Method for Optimal k')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('WCSS (Within-Cluster Sum of Squares)')
plt.grid(True)
plt.show()
```



```
In [5]: plt.figure(figsize=(4,4))
plt.scatter(df['Annual Income (k$)'], df['Spending Score (1-100)'], c='cyan', edgecolor='k')
plt.title('Income vs Spending Score')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.show()
```



```
In [6]: kmeans = KMeans(n_clusters=5, init='k-means++', random_state=42)
y_kmeans = kmeans.fit_predict(X)
```

```
In [7]: df['Cluster'] = y_kmeans
```

```
In [9]: plt.figure(figsize=(4, 4))
colors = ['red', 'blue', 'green', 'cyan', 'magenta']
for i in range(5):
    plt.scatter(
        X.values[y_kmeans == i, 0],
        X.values[y_kmeans == i, 1],
        s=80, c=colors[i],
        label=f'Cluster {i+1}'
    )
```

