

EXPERIMENT 10

Aim:

To perform customer segmentation using the K-Means clustering algorithm on the Mall Customers dataset.

Algorithm:

1. Import necessary libraries such as NumPy, Pandas, Matplotlib, and Seaborn.
2. Load the dataset using Pandas.
3. Explore the dataset by checking its information and displaying the first few rows.
4. Visualize pairwise relationships between features using Seaborn's pairplot().
5. Select relevant features (Annual Income and Spending Score) for clustering.
6. Apply the K-Means algorithm with a defined number of clusters (k=5).
7. Predict cluster labels and add them to the dataset.
8. Visualize the clusters using scatter plots.
9. Determine the optimal number of clusters using the Elbow method.

Code:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

df = pd.read_csv('Mall_Customers - Mall_Customers.csv')
df.info()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   CustomerID      200 non-null   int64
```

```
1 Gender      200 non-null object
2 Age         200 non-null int64
3 Annual Income (k$)  200 non-null int64
4 Spending Score (1-100) 200 non-null int64
```

```
dtypes: int64(4), object(1)
```

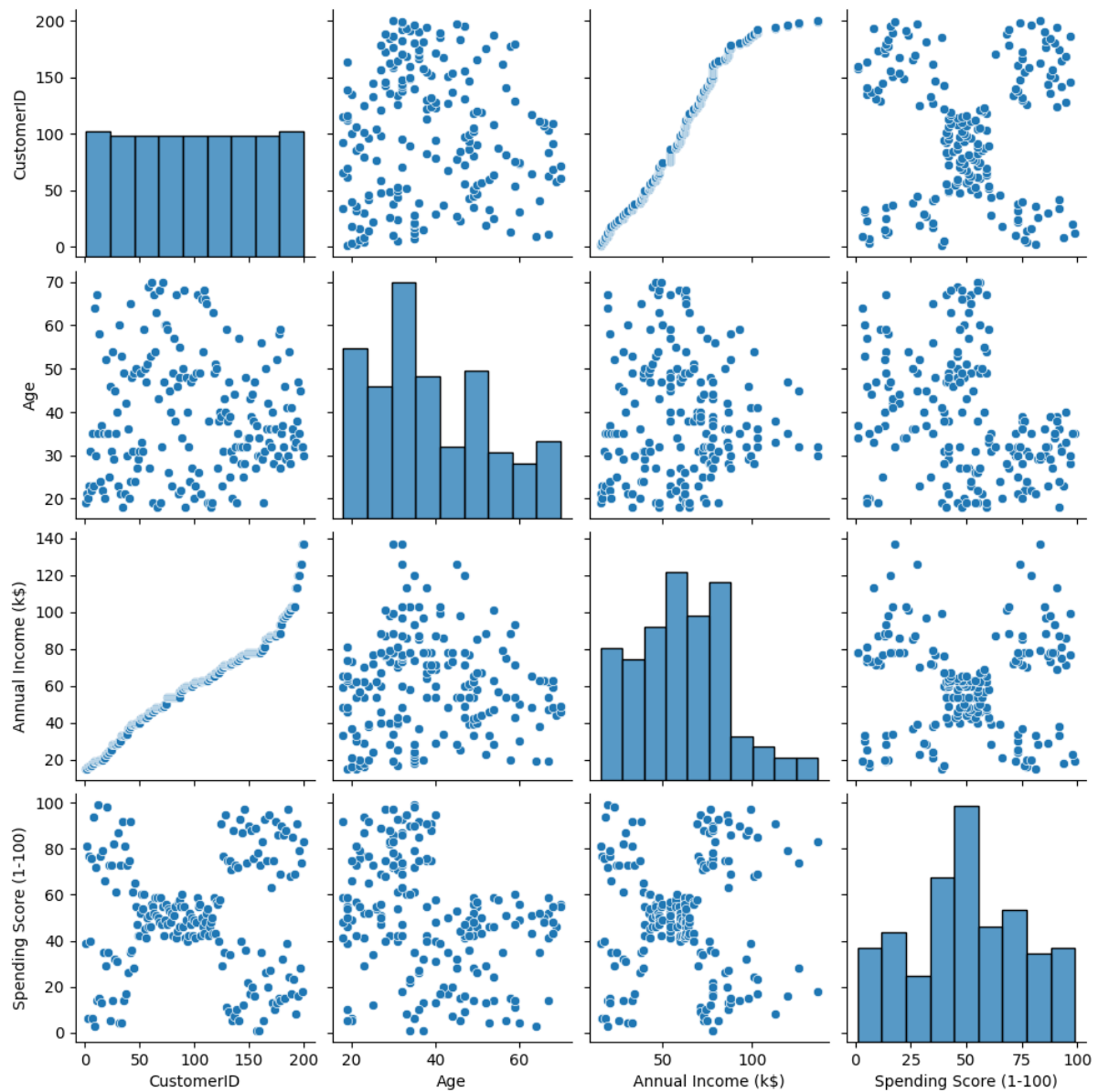
```
memory usage: 7.9+ KB
```

```
df.head()
```

Output:

	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

```
sns.pairplot(df)
```

Output:

```
features = df.iloc[:, [3, 4]].values
```

```
from sklearn.cluster import KMeans
```

```
model = KMeans(n_clusters=5)
```

```
model.fit(features)
```

```
KMeans(n_clusters=5)
```

Output:

```
KMeans(n_clusters=5)
```

```
Final = df.iloc[:, [3, 4]].copy()
```

```
Final['label'] = model.predict(features)
```

```
Final.head()
```

Output:

	Annual Income (k\$)	Spending Score (1-100)	label
0	15	39	2
1	15	81	4
2	16	6	1
3	16	77	4
4	17	40	2

```
sns.set_style("whitegrid")
```

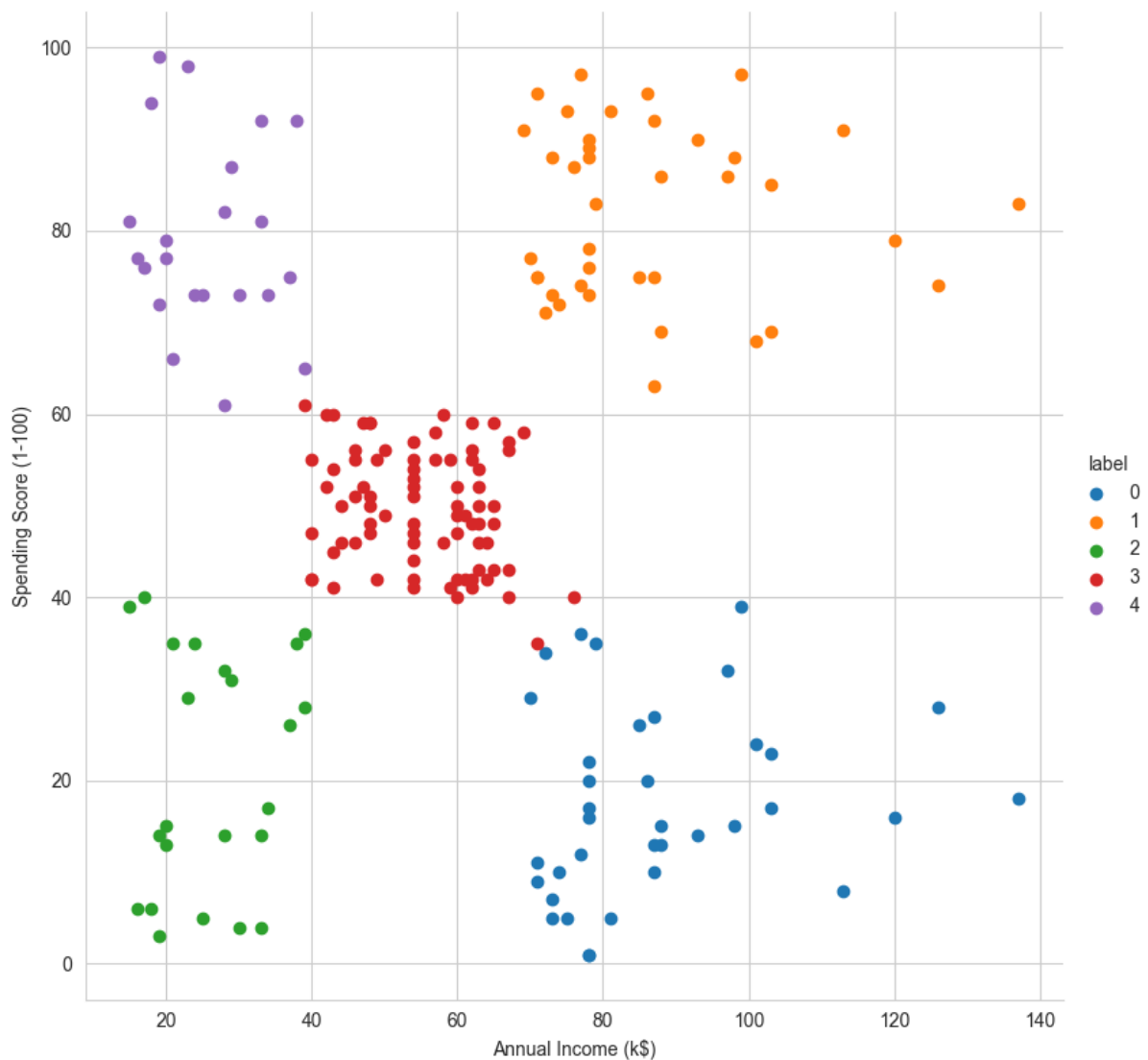
```
sns.FacetGrid(Final, hue="label", height=8) \
```

```
    .map(plt.scatter, "Annual Income (k$)", "Spending Score (1-100)") \
```

```
    .add_legend()
```

```
plt.show()
```

Output:



```
features_el = df.iloc[:, [2, 3, 4]].values
```

```
from sklearn.cluster import KMeans
```

```
wcss = []
```

```
for i in range(1, 10):
```

```
    model = KMeans(n_clusters=i)
```

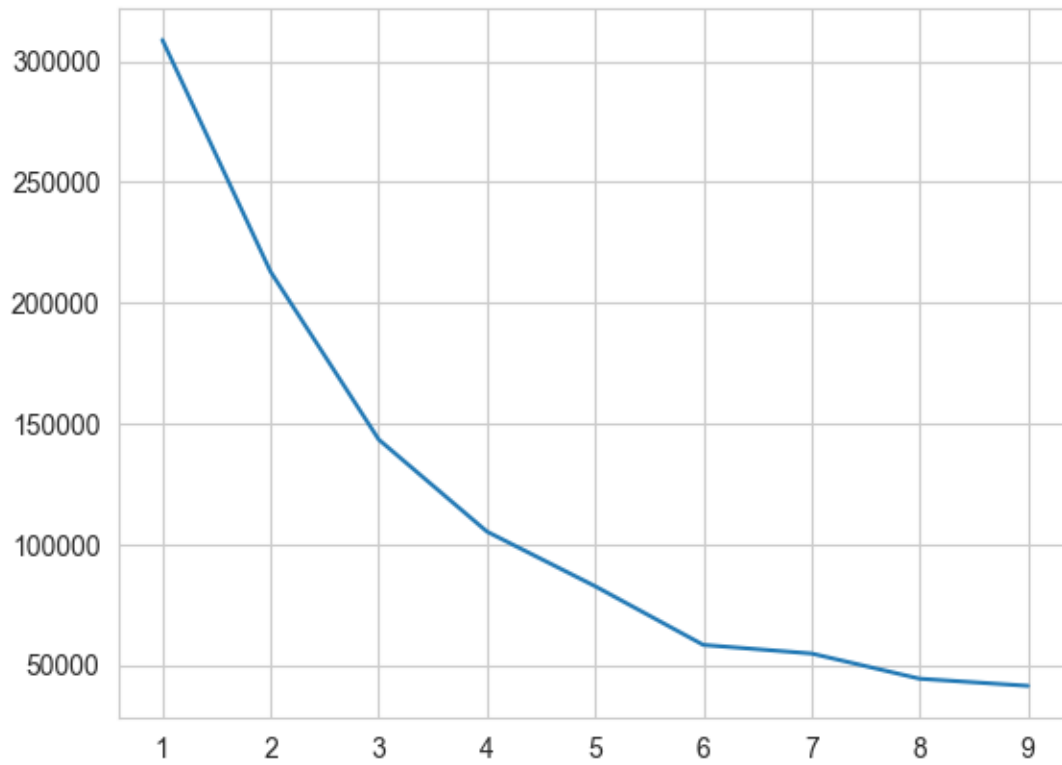
```
    model.fit(features_el)
```

```
    wcss.append(model.inertia_)
```

```
plt.plot(range(1, 10), wcss)
```

Output:

```
[<matplotlib.lines.Line2D at 0x259fb1ede50>]
```



Result:

The K-Means clustering algorithm was successfully applied on the Mall Customers dataset to segment customers into five distinct clusters based on their annual income and spending score.