

```
In [23]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn.cluster import KMeans

import warnings
warnings.filterwarnings("ignore")
```

```
In [10]: import pandas as pd

df = pd.read_excel("customer.xlsx")
print(df.head())

   Income  Spendings
0      80       24.0
1      85       25.5
2      87       26.1
3      75       22.5
4      78       23.4
```

```
In [11]: df.describe()
```

	Income	Spendings
count	17.000000	17.000000
mean	55.470588	23.370588
std	24.711125	2.786074
min	25.000000	17.500000
25%	33.000000	21.900000
50%	43.000000	23.400000
75%	79.000000	25.500000
max	87.000000	28.000000

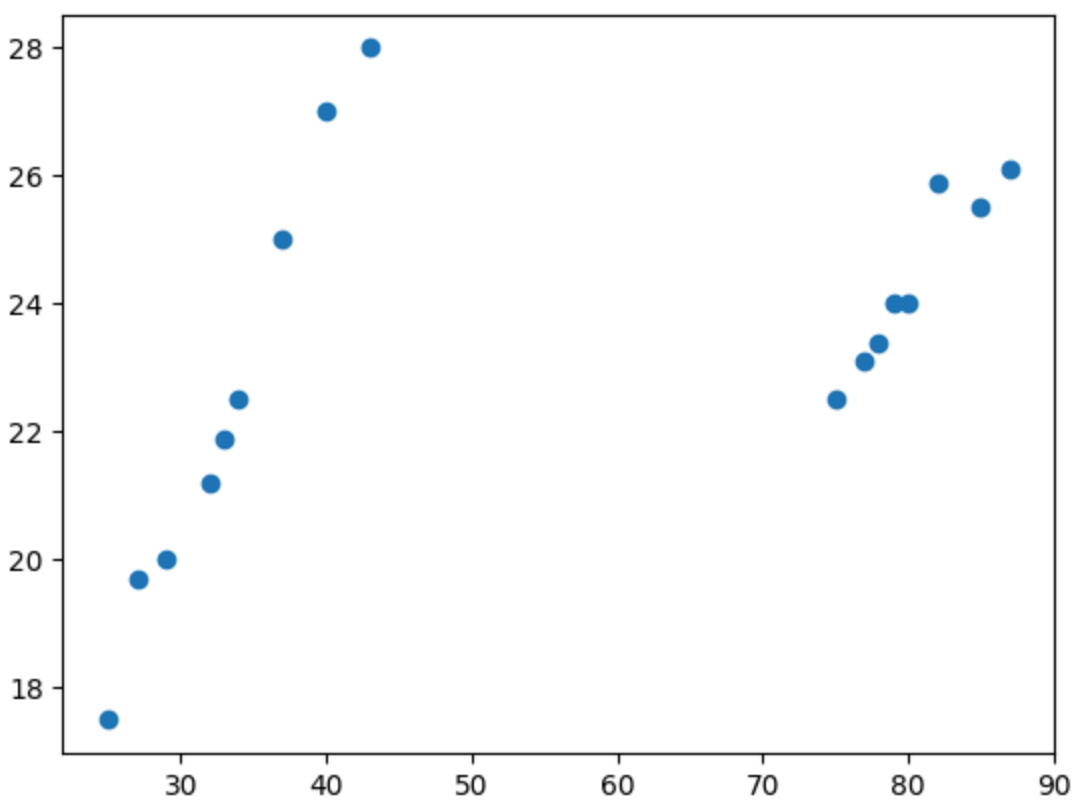
```
In [12]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17 entries, 0 to 16
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype  
---  --
0   Income  17 non-null        int64  
1   Spendings 17 non-null        float64
dtypes: float64(1), int64(1)
memory usage: 404.0 bytes
```

```
In [13]: df.shape
```

```
Out[13]: (17, 2)
```

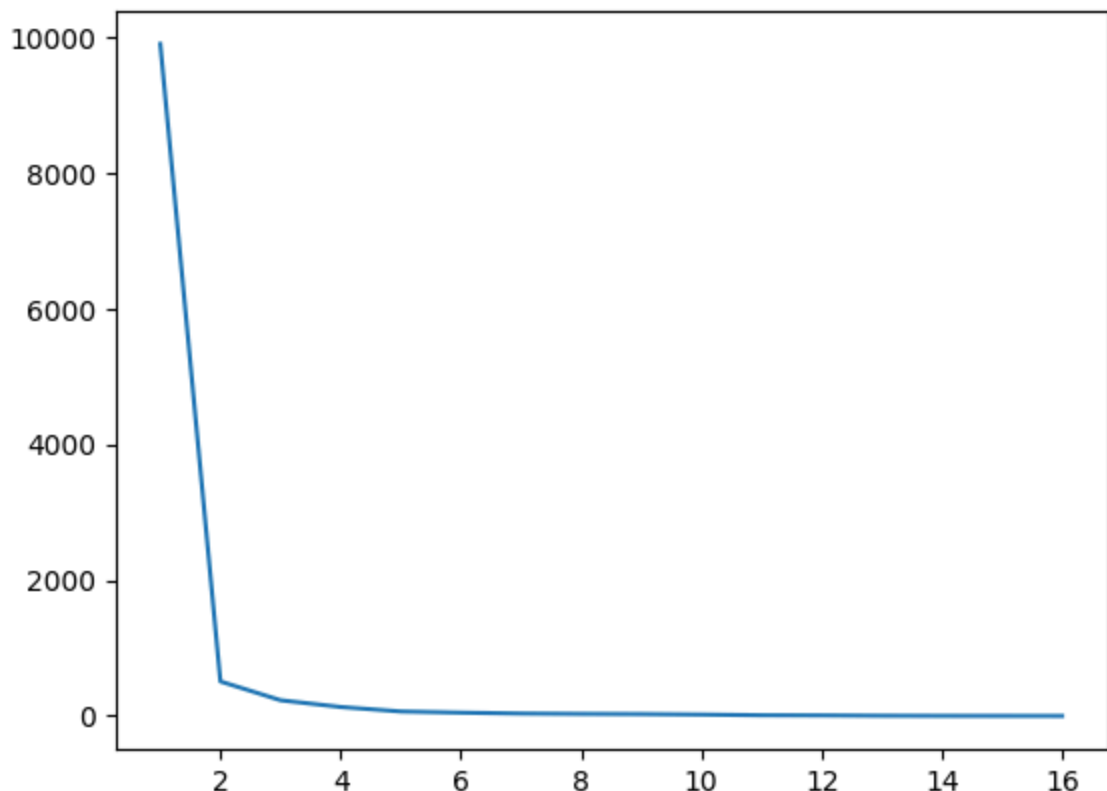
```
In [19]: plt.scatter(df["Income"],df["Spendings"])
plt.show()
```



```
In [26]: individual_clustering_score=[]

for i in range(1,17):
    kmeans=KMeans(n_clusters=i , init="random",random_state=42)
    kmeans.fit(df)
    individual_clustering_score.append(kmeans.inertia_)

plt.plot(range(1,17),individual_clustering_score)
plt.show()
```



```
In [27]: individual_clustering_score
```

```
Out[27]: [9894.430588235295,
508.66375,
231.78375,
132.27333333333334,
66.51333333333334,
48.846666666666664,
36.23083333333333,
30.86666666666666,
27.689833333333335,
20.618333333333332,
10.936666666666667,
8.77,
3.8916666666666666,
1.7249999999999999,
1.2449999999999994,
0.5]
```

```
In [28]: kmeans=KMeans(n_clusters=2,random_state=42)
kmeans.fit(df)
pred=kmeans.predict(df)
print(pred)

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

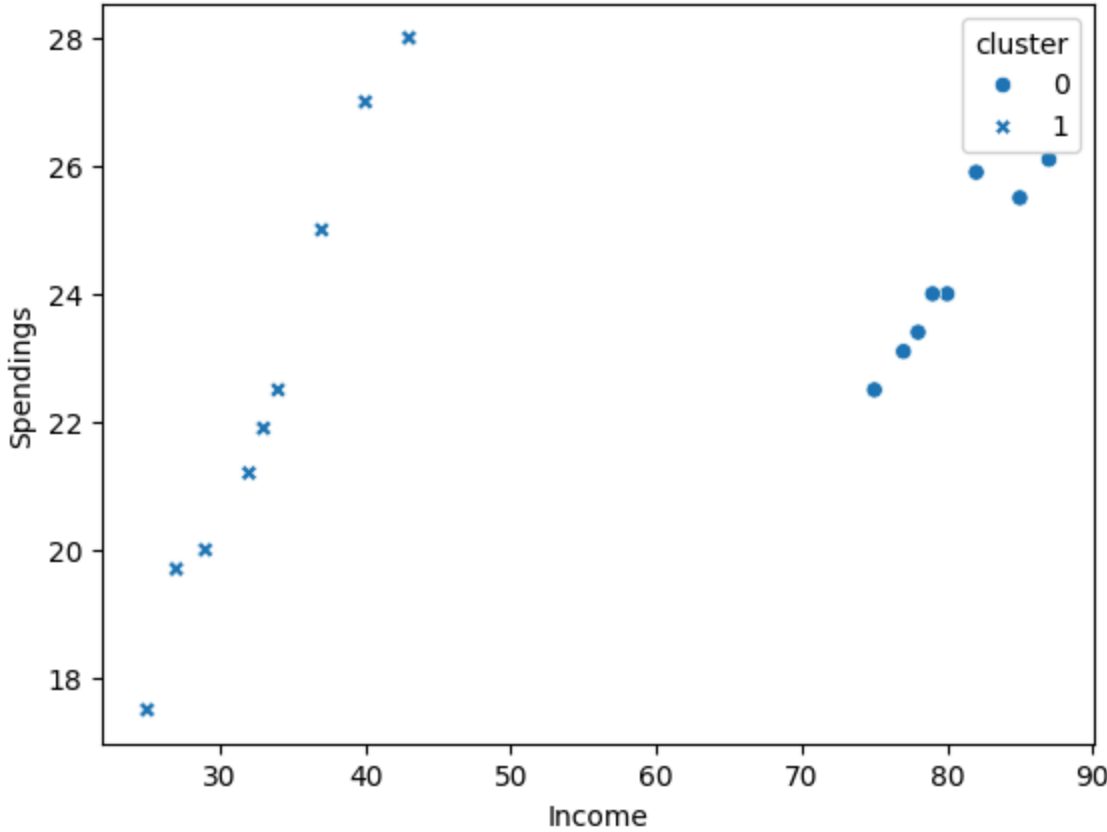
```
In [29]: df["cluster"]=pred
```

```
In [30]: df
```

	Income	Spendings	cluster
0	80	24.0	0
1	85	25.5	0
2	87	26.1	0
3	75	22.5	0
4	78	23.4	0
5	77	23.1	0
6	82	25.9	0
7	79	24.0	0
8	25	17.5	1
9	27	19.7	1
10	29	20.0	1
11	32	21.2	1
12	33	21.9	1
13	34	22.5	1
14	37	25.0	1
15	40	27.0	1
16	43	28.0	1

```
In [33]: sns.scatterplot(data=df,x="Income",y="Spendings",style="cluster")
```

```
Out[33]: <Axes: xlabel='Income', ylabel='Spendings'>
```



```
In [35]: clustering=DBSCAN(eps=3,min_samples=2)
predd=clustering.fit_predict(df)
df["cluster"]=predd
df
```

	Income	Spendings	cluster
0	80	24.0	0
1	85	25.5	1
2	87	26.1	1
3	75	22.5	0
4	78	23.4	0
5	77	23.1	0
6	82	25.9	0
7	79	24.0	0
8	25	17.5	2
9	27	19.7	2
10	29	20.0	2
11	32	21.2	3
12	33	21.9	3
13	34	22.5	3
14	37	25.0	-1
15	40	27.0	-1
16	43	28.0	-1

```
In [38]: sns.scatterplot(data=df,x="Income",y="Spendings",style="cluster")
```

```
Out[38]: <Axes: xlabel='Income', ylabel='Spendings'>
```

