

```
In [13]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
```

```
In [16]: data = pd.read_csv("Mall_Customers.csv")
```

```
In [17]: data.head()
```

Out[17]:

	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

```
In [18]: data.drop_duplicates(inplace=True)
data.dropna(inplace=True)
data.drop(['CustomerID'], axis=1, inplace=True)
```

```
In [19]: data.describe()
```

Out[19]:

	Age	Annual Income (k\$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000
mean	38.850000	60.560000	50.200000
std	13.969007	26.264721	25.823522
min	18.000000	15.000000	1.000000
25%	28.750000	41.500000	34.750000
50%	36.000000	61.500000	50.000000
75%	49.000000	78.000000	73.000000
max	70.000000	137.000000	99.000000

```
In [20]: sns.pairplot(data)
```

Out[20]: <seaborn.axisgrid.PairGrid at 0x20a4313ad00>

```
In [21]: sns.boxplot(data=data)
```

```
Out[21]: <AxesSubplot:xlabel='Spending Score (1-100)', ylabel='Count'>
```

```
In [24]: scaler = StandardScaler()
```

```
In [27]: wcss = []  
         for i in range(1, 1):  
             kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)  
             kmeans.fit(scaler)  
             wcss.append(kmeans.inertia_)
```

```
In [29]: plt.plot(range(1, -1), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

