

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans

dataset=pd.read_csv("./Mall_Customers.csv")
dataset.head()

    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

dataset.isnull().sum()

CustomerID      0
Genre           0
Age             0
Annual Income (k$)  0
Spending Score (1-100)  0
dtype: int64

dataset.dtypes

CustomerID      int64
Genre          object
Age            int64
Annual Income (k$)  int64
Spending Score (1-100)  int64
dtype: object

dataset['Genre']=dataset['Genre'].apply(lambda x: 1 if x=="Male" else
0)
dataset.head()

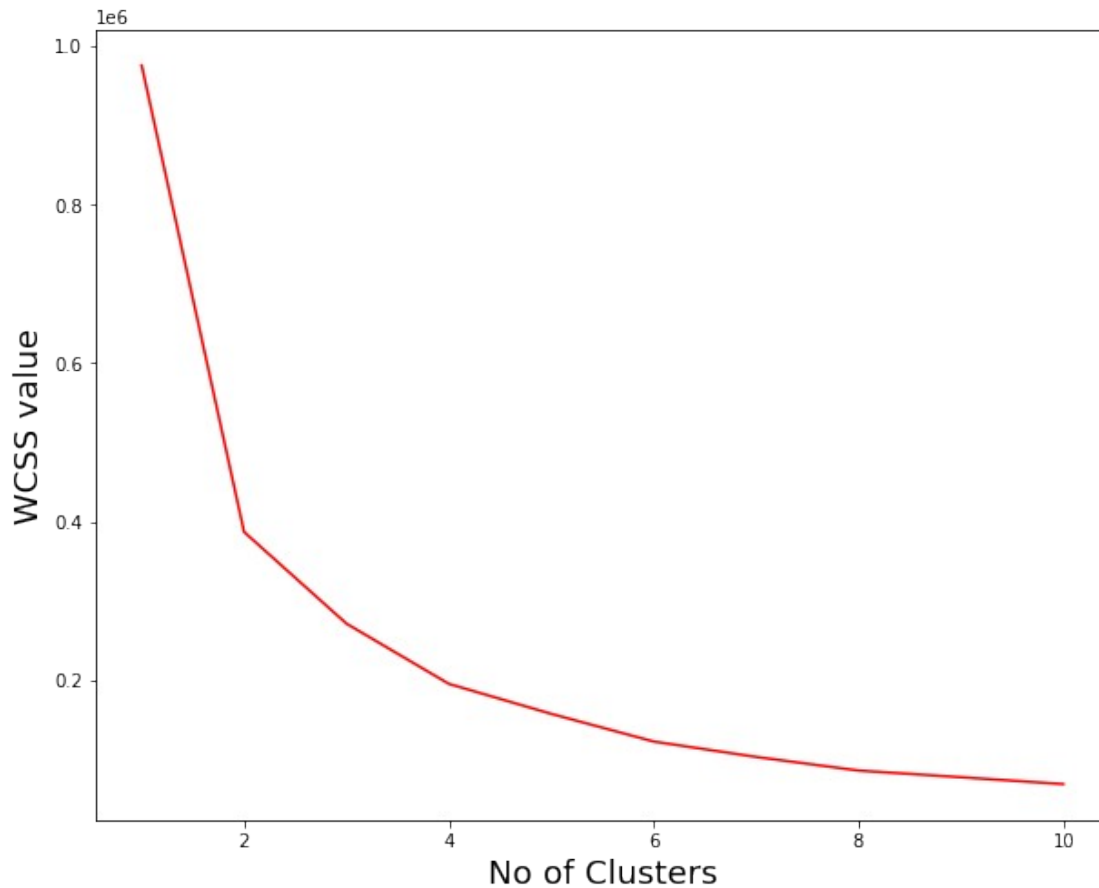
    CustomerID  Genre  Age  Annual Income (k$)  Spending Score (1-100)
0            1     1   19                15                39
1            2     1   21                15                81
2            3     0   20                16                 6
3            4     0   23                16    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

cluster=[]

for n in range(1,11):
    kmeans=KMeans(n_clusters=n).fit(dataset)
    cluster.append(kmeans.inertia_)

```

```
plt.figure(figsize=(10,8))
plt.plot(range(1, 11), cluster, 'r-')
plt.xlabel("No of Clusters",fontsize="18")
plt.ylabel("WCSS value",fontsize="18")
plt.show()
```



#according to the above elbow curve choose the value of K as 4 and create the KMeans model

```
model=KMeans(n_clusters=4).fit(dataset)
dataset['Cluster no']=model.labels_
dataset.tail()
```

	CustomerID	Genre	Age	Annual Income (k\$)	Spending Score (1-
100)	\				
195	196	0	35	120	
79					
196	197	0	45	126	
28					
197	198	1	32	126	
74					
198	199	1	32	137	
18					

```
199          200          1    30          137
83
```

```
Cluster no
195          0
196          3
197          0
198          3
199          0
```

```
plt.figure(figsize=(10,8))
sns.scatterplot(dataset['Annual Income (k$)'], dataset['Spending Score
(1-100)'], hue=dataset['Cluster no'],
palette=sns.color_palette('hls', 4))
plt.show()
```

/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variables as keyword args: x, y.
From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in
an error or misinterpretation.

FutureWarning

