Program to implement k-means clustering technique using any standard dataset available in the public domain

Dataset: College_Data.csv

```
1 import numpy as np
2 import pandas as pd
```

```
1 dataset = pd.read_csv('Mall_Customers.csv')
```

```
1 dataset
```

	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
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

```
1 from sklearn.cluster import KMeans
2
3 kmeans=KMeans(n_clusters=2)
4
5 kmeans.fit(dataset.drop('Genre',axis=1))
6
7 kmeans.cluster_centers_
```

```
    array([[ 50.        ,  39.94949495,  39.34343434,  49.93939394,
             0.4040404 ],
```

```
      [150.        ,  37.77227723,  81.35643564,  50.45544554,
```

```python
1  def converter(Genre):
2    if Genre=="Male":
3      return 1
4    else:
5      return 0
```

```python
1  dataset['cluster']=dataset['Genre'].apply(converter)
2  dataset.head()
```

```python
1  from sklearn.metrics import accuracy_score
```

```python
1  print("Accuracy Score:",accuracy_score(dataset['cluster'],kmeans.labels_))
```

```
    Accuracy Score: 0.535
```

✓ 0s   completed at 12:41 PM