In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
from sklearn.cluster import KMeans
from sklearn import datasets
```

In [4]:
```
df=pd.read_csv("shopping-data.csv")
df
```

Out[4]:

	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

In [5]:
```
df.head()
```

Out[5]:

	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

In [6]:
```
df.tail()
```

Out[6]:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
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

In [14]:
```
df1=df.iloc[:,2:5]
df1
```

Out[14]:

	Age	Annual Income (k$)	Spending Score (1-100)
0	19	15	39
1	21	15	81
2	20	16	6
3	23	16	77
4	31	17	40
...	...	...	...
195	35	120	79
196	45	126	28
197	32	126	74
198	32	137	18
199	30	137	83

200 rows × 3 columns

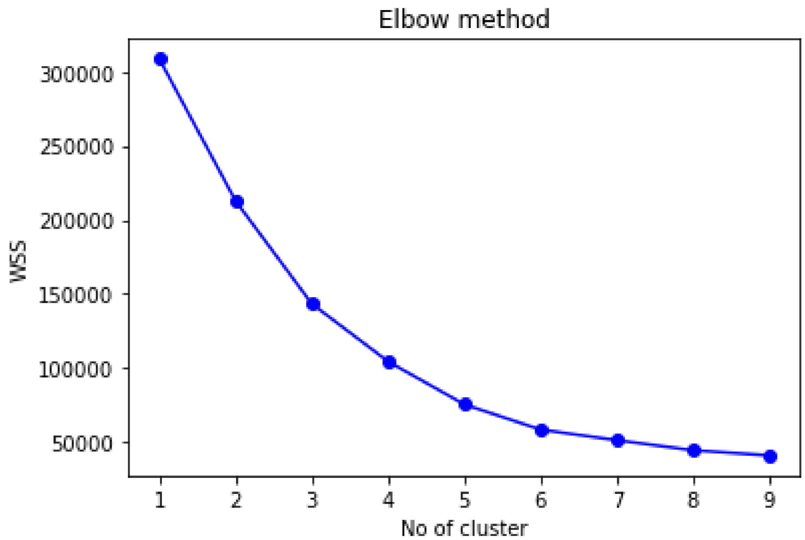
In [15]:
```
distortion=[]
k=range(1,10)
```

In [16]:
```
for i in k:
    Kmeanmodel=KMeans(n_clusters=i,max_iter=25)
    Kmeanmodel.fit(df1)
    distortion.append(Kmeanmodel.inertia_)
```

```
C:\Users\admin\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:881: Us
erWarning: KMeans is known to have a memory leak on Windows with MKL, when th
ere are less chunks than available threads. You can avoid it by setting the e
nvironment variable OMP_NUM_THREADS=1.
  warnings.warn(
```

In [17]:
```
plt.plot(figure=[16.8])
plt.plot(k,distortion,'bo-')
plt.xlabel("No of cluster")
plt.ylabel("WSS")
plt.title("Elbow method")
```

Out[17]: Text(0.5, 1.0, 'Elbow method')



In [18]:
```
final_model=KMeans(n_clusters=7,max_iter=25)
final_model.fit(df1)
```

Out[18]: KMeans(max_iter=25, n_clusters=7)

In [20]:
```
final_model.fit(df1)
```

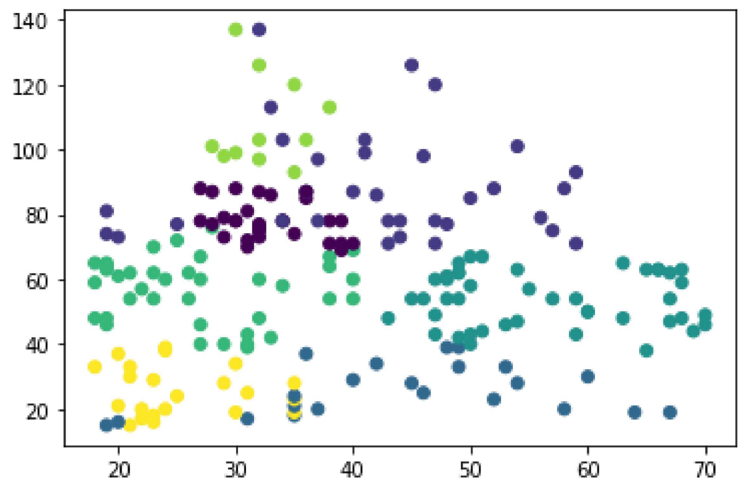
Out[20]: KMeans(max_iter=25, n_clusters=7)

In [21]:
```
final_model.cluster_centers_
final_model.labels_
```

Out[21]:
```
array([2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6,
       2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 2, 6, 3, 6, 2, 4,
       2, 6, 3, 4, 4, 4, 3, 4, 4, 3, 3, 3, 3, 3, 4, 3, 3, 4, 3, 3, 3, 4,
       3, 3, 4, 4, 3, 3, 3, 3, 3, 4, 3, 4, 4, 3, 3, 4, 3, 3, 4, 3, 3, 4,
       4, 3, 3, 4, 3, 4, 4, 4, 3, 4, 3, 4, 4, 3, 3, 4, 3, 4, 3, 3, 3, 3,
       3, 4, 4, 4, 4, 4, 3, 3, 3, 3, 4, 4, 4, 0, 4, 0, 1, 0, 1, 0, 1, 0,
       4, 0, 1, 0, 1, 0, 1, 0, 1, 0, 4, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,
       1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0,
       1, 0, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5, 1, 5,
       1, 5])
```

In [22]:
```
plt.scatter(df1["Age"],df1["Annual Income (k$)"],c=final_model.labels_)
```

Out[22]: <matplotlib.collections.PathCollection at 0x1d1bc7b1f40>



In [23]:
```
final_model=KMeans(n_clusters=6,max_iter=25)
final_model.fit(df1)
```

Out[23]: KMeans(max_iter=25, n_clusters=6)

In [24]:
```
KMeans(max_iter=25, n_clusters=6)
```

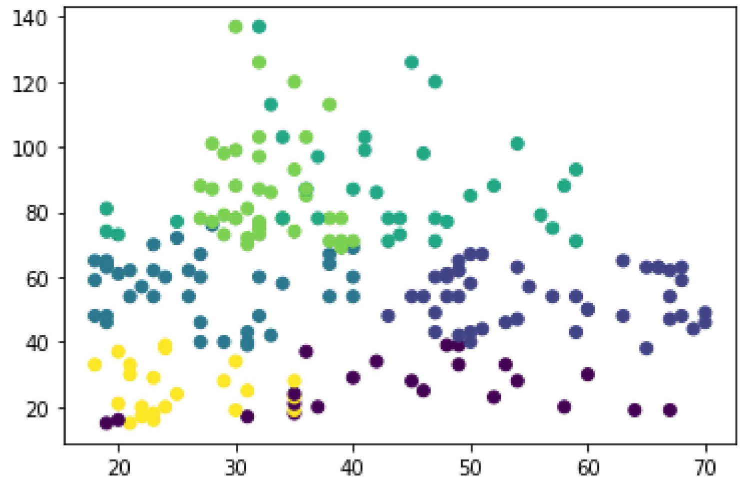
Out[24]: KMeans(max_iter=25, n_clusters=6)

In [25]:
```
final_model.cluster_centers_
final_model.labels_
```

Out[25]:
```
array([0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5,
       0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 0, 5, 1, 5, 0, 2,
       0, 5, 1, 2, 2, 2, 1, 2, 2, 1, 1, 1, 1, 1, 2, 1, 1, 2, 1, 1, 1, 2,
       1, 1, 2, 2, 1, 1, 1, 1, 1, 2, 1, 2, 2, 1, 1, 2, 1, 1, 2, 1, 1, 2,
       2, 1, 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 2, 1, 1, 2, 1, 2, 1, 1, 1, 1,
       1, 2, 2, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 4, 2, 4, 3, 4, 3, 4, 3, 4,
       2, 4, 3, 4, 3, 4, 3, 4, 3, 4, 2, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4,
       3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4,
       3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4,
       3, 4])
```

In [26]:
```
plt.scatter(df1["Age"],df1["Annual Income (k$)"],c=final_model.labels_)
```

Out[26]:
```
<matplotlib.collections.PathCollection at 0x1d1bc827e50>
```



In [ ]: