

In [1]:

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 %matplotlib inline
```

In [2]:

```
1 df=pd.read_csv(r"C:\Users\91955\Downloads\Income.csv")
2 df
```

Out[2]:

	Gender	Age	Income(\$)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17
...
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

200 rows × 3 columns

In [3]:

```
1 df.head()
```

Out[3]:

	Gender	Age	Income(\$)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17

In [4]:

```
1 df.tail()
```

Out[4]:

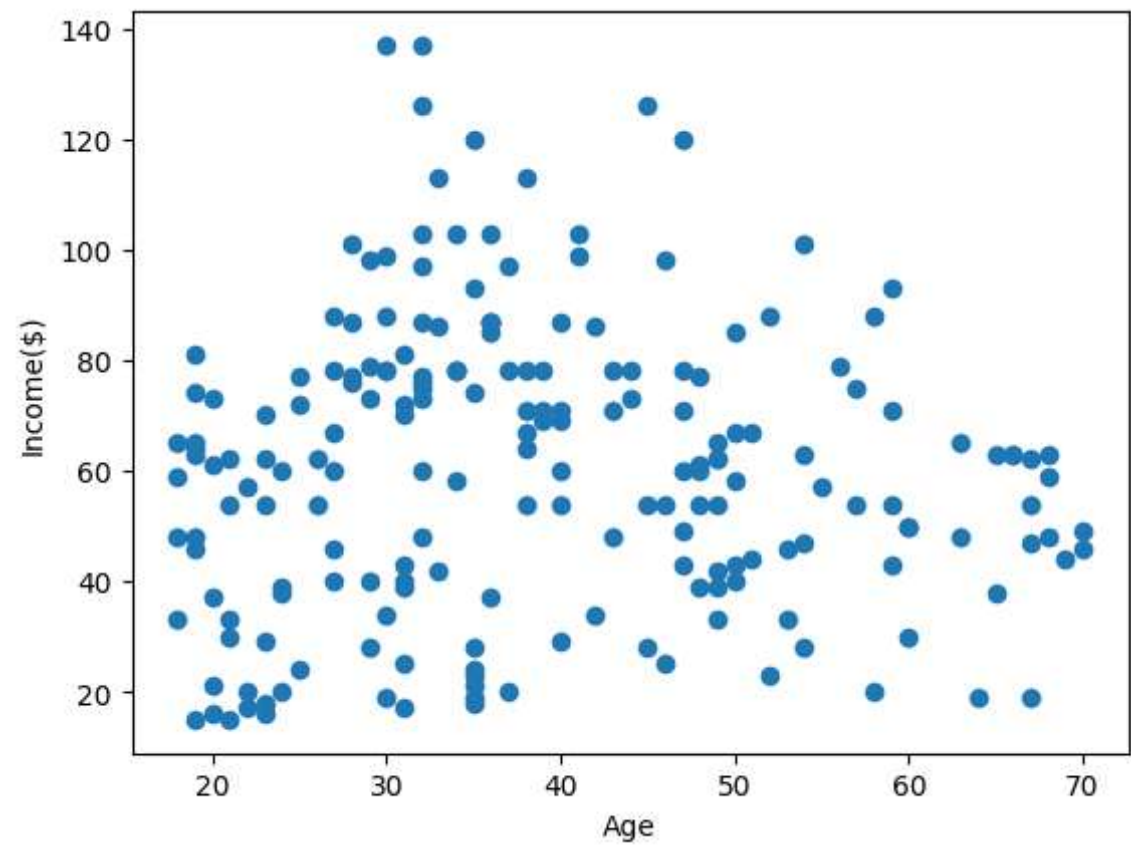
	Gender	Age	Income(\$)
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

In [5]:

```
1 plt.scatter(df["Age"],df["Income($)"])
2 plt.xlabel("Age")
3 plt.ylabel("Income($)")
```

Out[5]:

Text(0, 0.5, 'Income(\$)')



```
1 from sklearn.cluster import KMeans
2 km=KMeans()
3 km
```

- ▼ KMeans

```
1 y_predicted=km.fit_predict(df[["Age", "Income($)"]])
2 y_predicted
```

[illegible]

```
1 df["cluster"]=y_predicted
2 df.head()
```

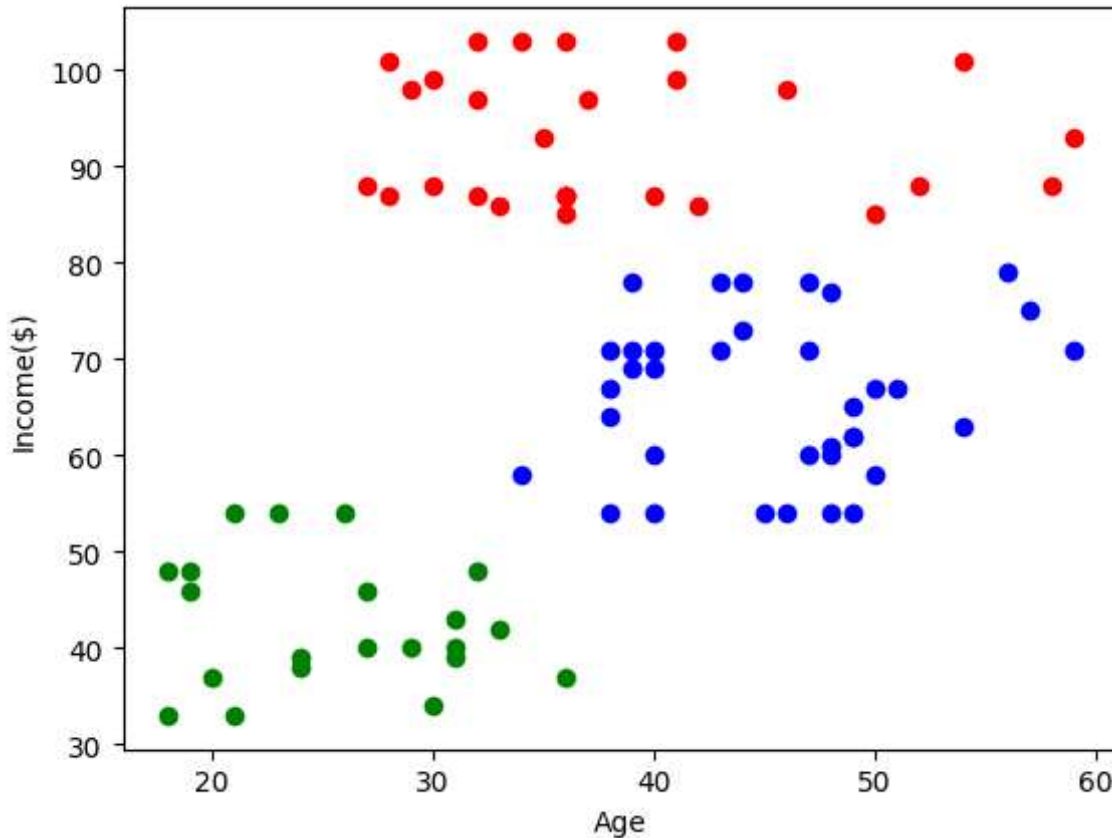
	Gender	Age	Income(\$)	cluster
0	Male	19	15	7
1	Male	21	15	7
2	Female	20	16	7
3	Female	23	16	7
4	Female	31	17	7

In [9]:

```
1 df1=df[df.cluster==0]
2 df2=df[df.cluster==1]
3 df3=df[df.cluster==2]
4 plt.scatter(df1["Age"],df1["Income($)"],color="red")
5 plt.scatter(df2["Age"],df2["Income($)"],color="green")
6 plt.scatter(df3["Age"],df3["Income($)"],color="blue")
7 plt.xlabel("Age")
8 plt.ylabel("Income($)")
```

Out[9]:

Text(0, 0.5, 'Income(\$)')



In [10]:

```
1 from sklearn.preprocessing import MinMaxScaler
2 scaler=MinMaxScaler()
3 scaler.fit(df[["Income($)"]])
4 df["Income($)"]=scaler.transform(df[["Income($)"]])
5 df.head()
```

Out[10]:

	Gender	Age	Income(\$)	cluster
0	Male	19	0.000000	7
1	Male	21	0.000000	7
2	Female	20	0.008197	7
3	Female	23	0.008197	7
4	Female	31	0.016393	7

In [11]:

```
1 scaler.fit(df[["Age"]])
2 df["Age"]=scaler.transform(df[["Age"]])
3 df.head()
```

Out[11]:

	Gender	Age	Income(\$)	cluster
0	Male	0.019231	0.000000	7
1	Male	0.057692	0.000000	7
2	Female	0.038462	0.008197	7
3	Female	0.096154	0.008197	7
4	Female	0.250000	0.016393	7

In [12]:

```
1 km=KMeans()
```

In [13]:

```
1 y_predicted=km.fit_predict(df[["Age", "Income($)"]])
2 y_predicted
```

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\s
klearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init`
` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicit
ly to suppress the warning
warnings.warn(

Out[13]:

```
array([2, 2, 2, 2, 6, 2, 6, 2, 3, 6, 3, 6, 1, 2, 6, 2, 6, 2, 1, 6, 6, 2,  
       1, 6, 1, 6, 1, 6, 6, 2, 3, 2, 1, 2, 1, 2, 1, 6, 6, 2, 3, 2, 1, 6,  
       1, 2, 1, 6, 6, 6, 1, 6, 6, 3, 1, 1, 1, 3, 6, 1, 3, 7, 3, 1, 3, 7,  
       1, 3, 7, 6, 3, 1, 3, 3, 3, 7, 1, 1, 7, 1, 3, 0, 3, 1, 7, 1, 1, 7,  
       0, 1, 3, 7, 4, 0, 0, 7, 4, 7, 4, 7, 7, 4, 3, 7, 4, 7, 3, 4, 3, 3,  
       3, 7, 0, 7, 7, 7, 3, 4, 4, 4, 7, 0, 0, 0, 7, 0, 4, 0, 4, 0, 4, 0,  
       7, 0, 7, 0, 4, 0, 7, 0, 4, 0, 0, 0, 7, 0, 4, 0, 0, 0, 4, 0, 4, 0,  
       4, 0, 0, 0, 0, 0, 4, 0, 7, 0, 4, 0, 4, 0, 0, 0, 0, 0, 0, 4, 0, 0,  
       4, 0, 4, 0, 5, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,  
       5, 5])
```

In [14]:

```
1 df["New Cluster"]=y_predicted
2 df.head()
```

Out[14]:

	Gender	Age	Income(\$)	cluster	New Cluster
0	Male	0.019231	0.000000	7	2
1	Male	0.057692	0.000000	7	2
2	Female	0.038462	0.008197	7	2
3	Female	0.096154	0.008197	7	2
4	Female	0.250000	0.016393	7	6

In [15]:

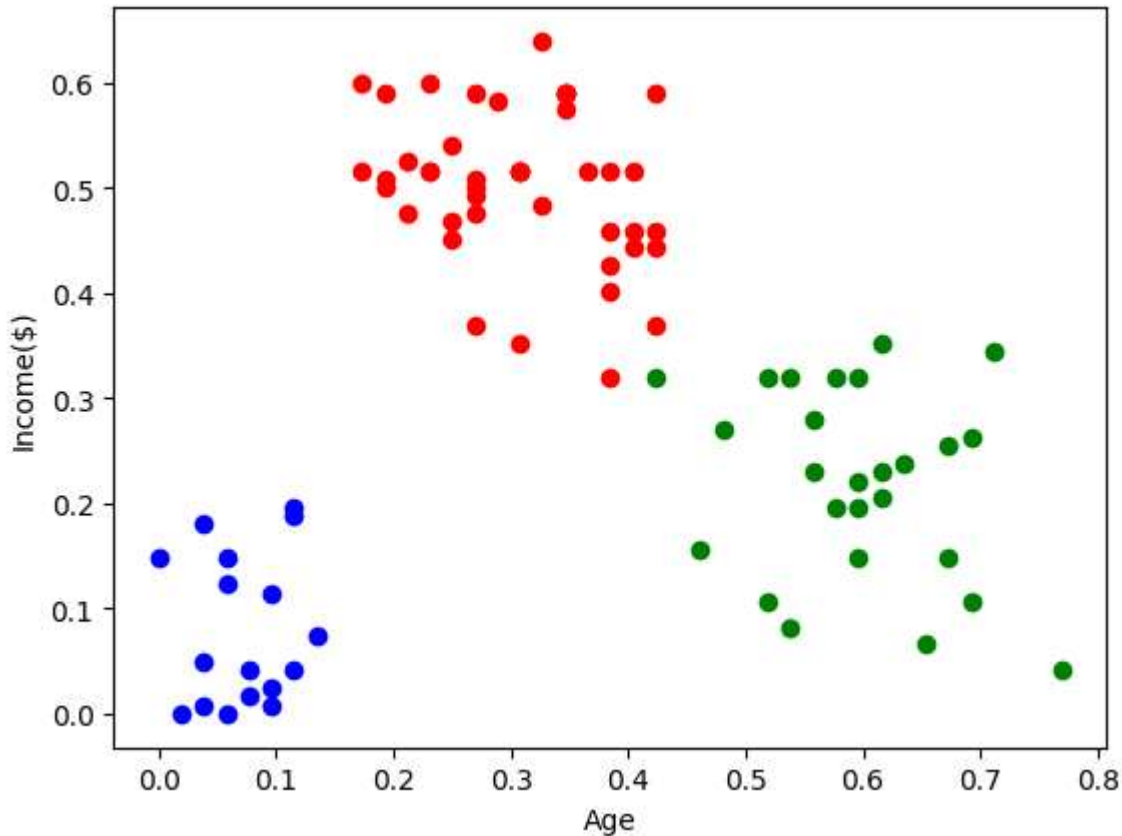
```

1 df1=df[df["New Cluster"]==0]
2 df2=df[df["New Cluster"]==1]
3 df3=df[df["New Cluster"]==2]
4 plt.scatter(df1["Age"],df1["Income($)"],color="red")
5 plt.scatter(df2["Age"],df2["Income($)"],color="green")
6 plt.scatter(df3["Age"],df3["Income($)"],color="blue")
7 plt.xlabel("Age")
8 plt.ylabel("Income($)")

```

Out[15]:

Text(0, 0.5, 'Income(\$)')



In [16]:

```

1 km.cluster_centers_

```

Out[16]:

```

array([[0.3059034 , 0.50247808],
       [0.5954142 , 0.2203657 ],
       [0.07239819, 0.08003857],
       [0.89799331, 0.28011404],
       [0.61094675, 0.49401009],
       [0.32905983, 0.78551913],
       [0.27884615, 0.13040238],
       [0.06923077, 0.38786885]])

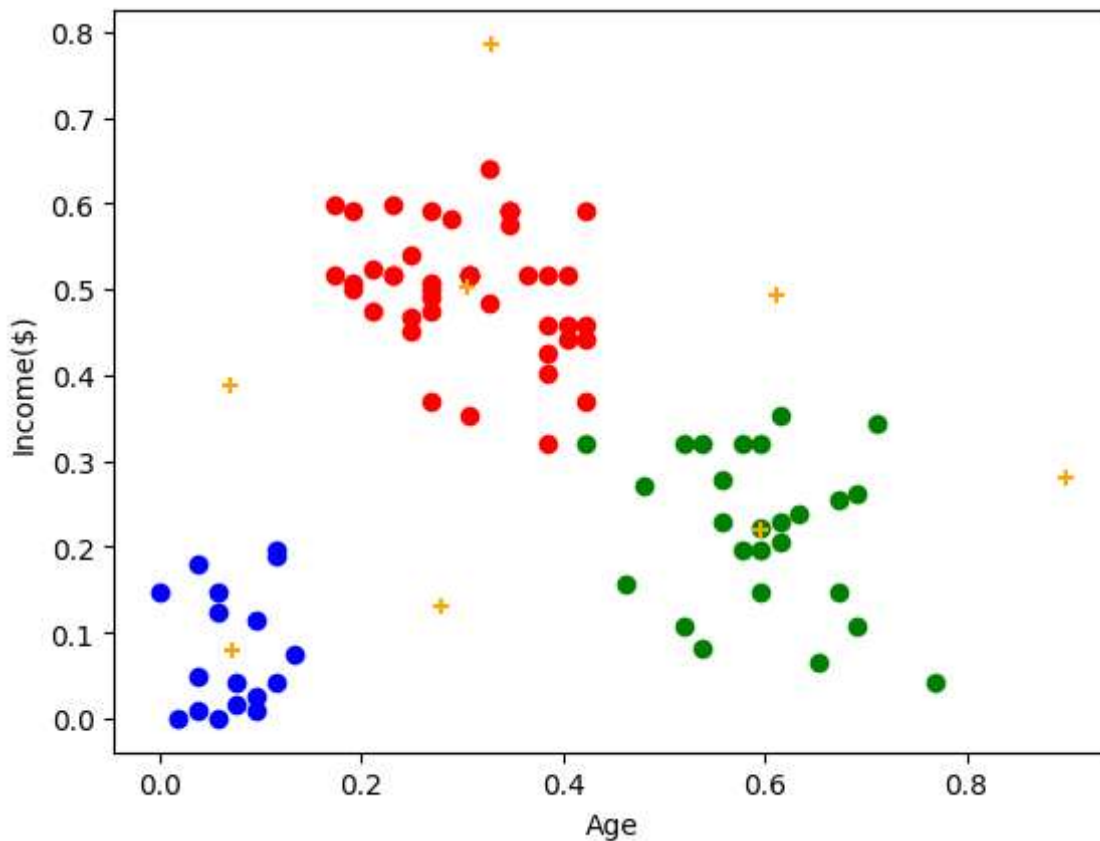
```

In [17]:

```
1 df1=df[df["New Cluster"]==0]
2 df2=df[df["New Cluster"]==1]
3 df3=df[df["New Cluster"]==2]
4 plt.scatter(df1["Age"],df1["Income($)"],color="red")
5 plt.scatter(df2["Age"],df2["Income($)"],color="green")
6 plt.scatter(df3["Age"],df3["Income($)"],color="blue")
7 plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color="orange",marker=
8 plt.xlabel("Age")
9 plt.ylabel("Income($)")
```

Out[17]:

Text(0, 0.5, 'Income(\$)')



In [18]:

```
1 k_rng=range(1,10)
2 sse=[]
```


In [19]:

```

1 for k in k_rng:
2     km=KMeans(n_clusters=k)
3     km.fit(df[["Age", "Income($)"]])
4     sse.append(km.inertia_)
5     #km.inertia_ will give you the value of sum of square error
6     print(sse)
7     plt.plot(k_rng, sse)
8     plt.xlabel("K")
9     plt.ylabel("Sum of Squared Error")

```

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning

warnings.warn(

[23.583906150363603, 13.028938428018286, 7.492107868586012, 6.0728847287425545, 4.713416604872824, 3.854282629100963, 3.058084466878064, 2.645864014245702, 2.3135720353543285]

```
C:\Users\91955\AppData\Local\Programs\Python\Python310\lib\site-packages\s  
klearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init  
` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicit  
ly to suppress the warning  
warnings.warn(
```

Out[19]:

Text(0, 0.5, 'Sum of Squared Error')

