In [1]:
```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

In [2]:
```
df=pd.read_csv(r"C:\Users\pucha\Downloads\Income.csv")
df
```

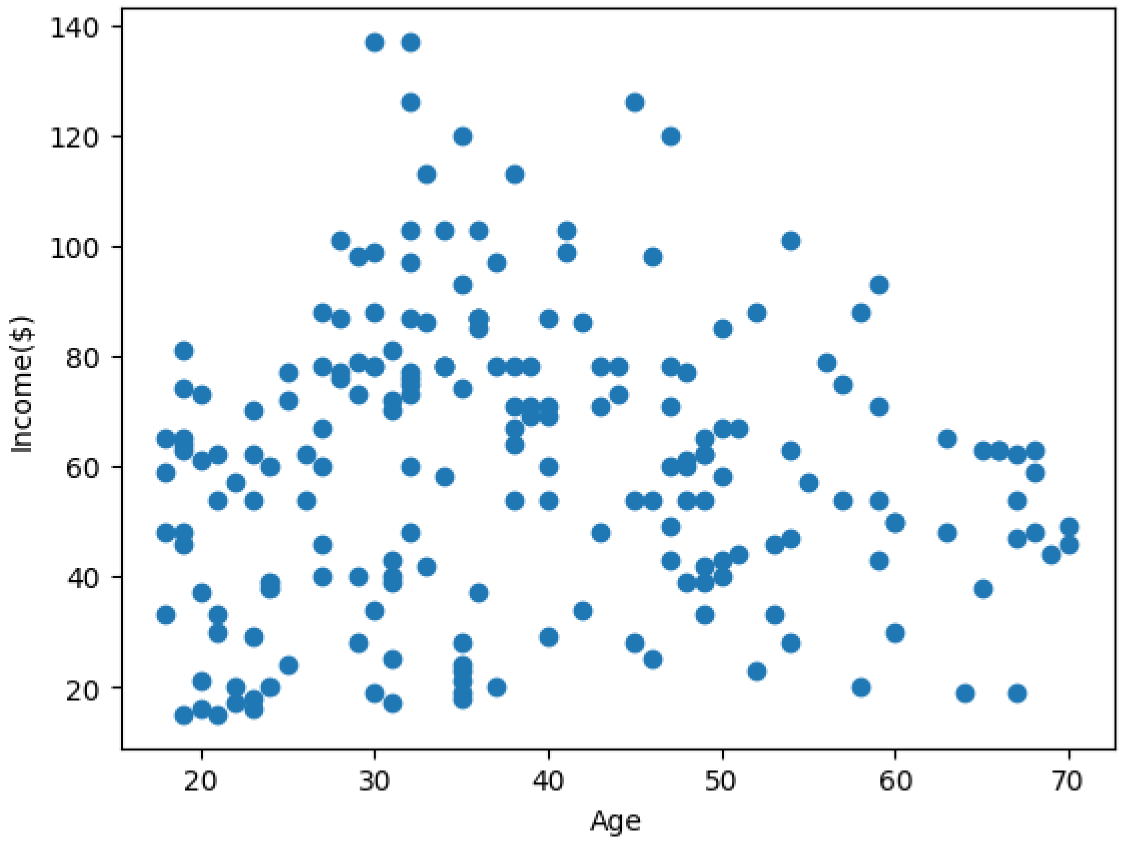
Out[2]:
```
     Gender  Age  Income($)
  0    Male   19         15
  1    Male   21         15
  2  Female   20         16
  3  Female   23         16
  4  Female   31         17
...     ...  ...        ...
195  Female   35        120
196  Female   45        126
197    Male   32        126
198    Male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

200 rows × 3 columns
```

In [3]:

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

Out[3]:

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



In [4]:

```
from sklearn.cluster import KMeans
```

In [5]:
```
km=KMeans()
km
```

Out[5]:
```
▾ KMeans

KMeans()
```

In [6]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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(
```

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

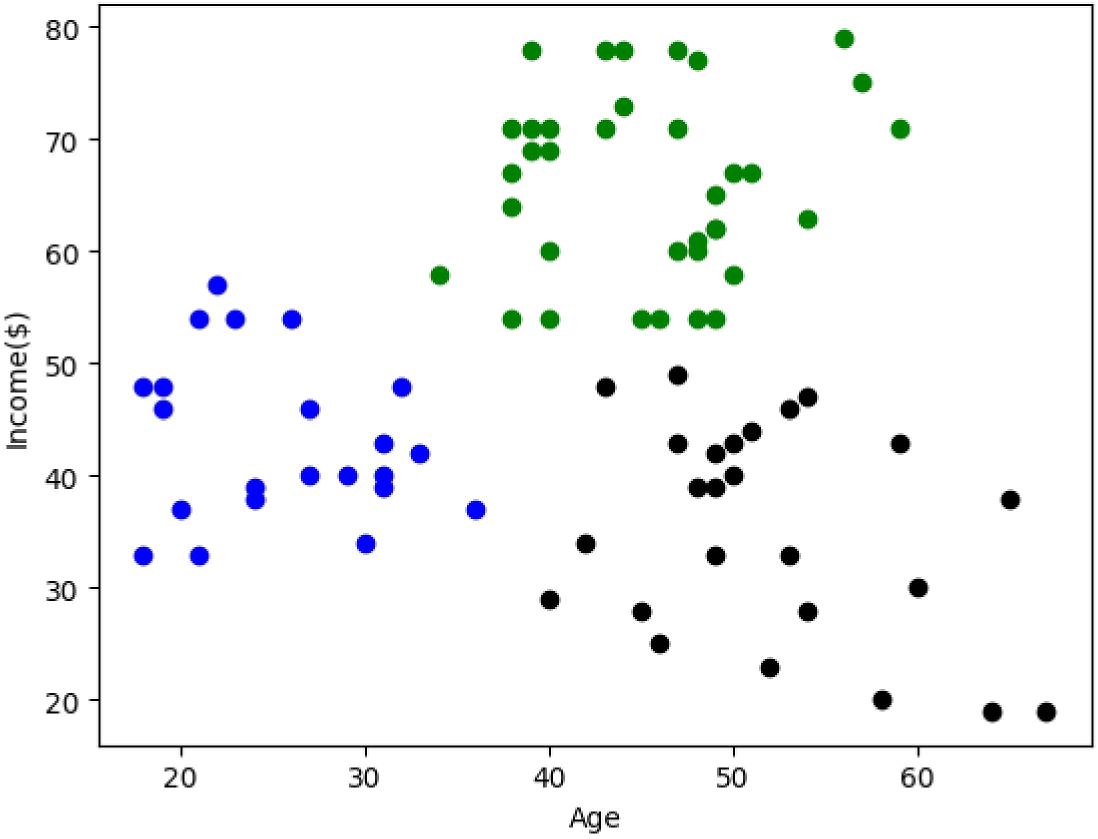
In [7]:
```
df["cluster"]=y_predicted
df.head()
```

Out[7]:

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

```
In [8]: df1=df[df.cluster==0]
        df2=df[df.cluster==1]
        df3=df[df.cluster==2]
        plt.scatter(df1["Age"],df1["Income($)"],color="black")
        plt.scatter(df2["Age"],df2["Income($)"],color="green")
        plt.scatter(df3["Age"],df3["Income($)"],color="blue")
        plt.xlabel("Age")
        plt.ylabel("Income($)")
```

```
Out[8]: Text(0, 0.5, 'Income($)')
```



In [9]:
```
from sklearn.preprocessing import MinMaxScaler
```

In [10]:
```
Scaler=MinMaxScaler()
```

In [11]:
```
Scaler.fit(df[["Income($)"]])
df["Income($)"]=Scaler.transform(df[["Income($)"]])
df.head()
```

Out[11]:

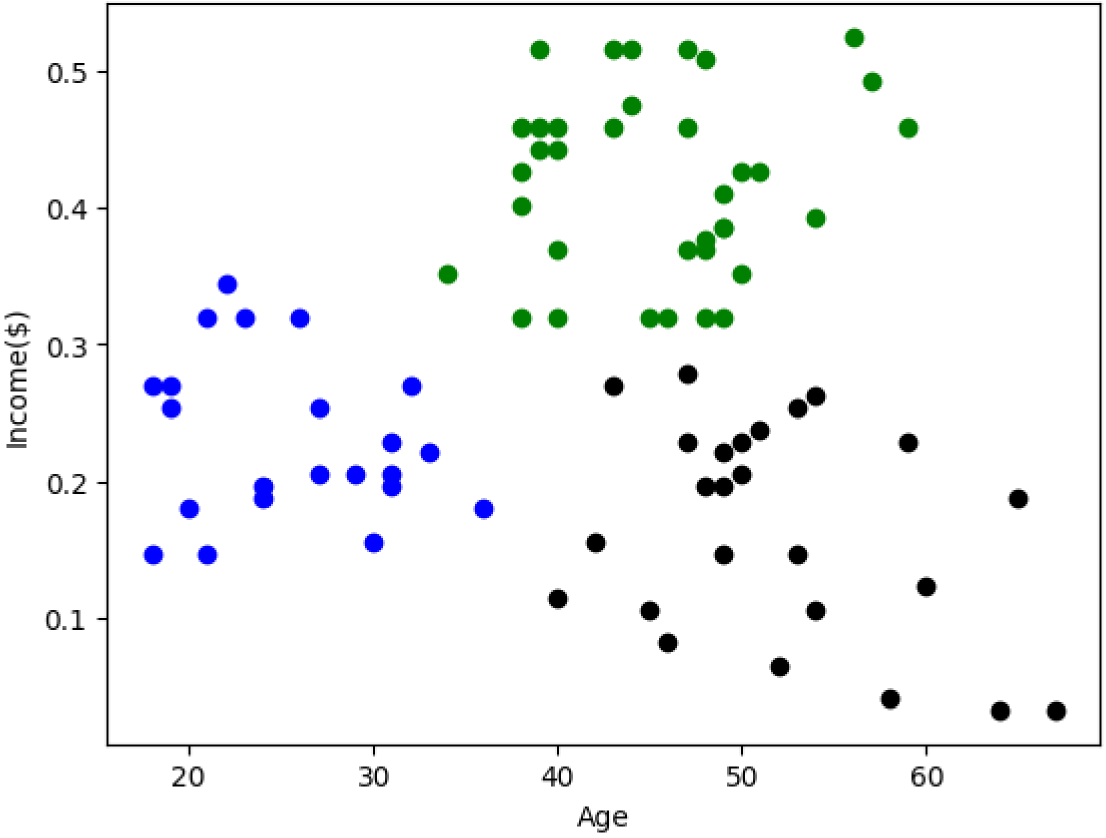
	Gender	Age	Income($)	cluster
0	Male	19	0.000000	4
1	Male	21	0.000000	4
2	Female	20	0.008197	4
3	Female	23	0.008197	4
4	Female	31	0.016393	4

In [12]:

```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="black")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[12]:

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

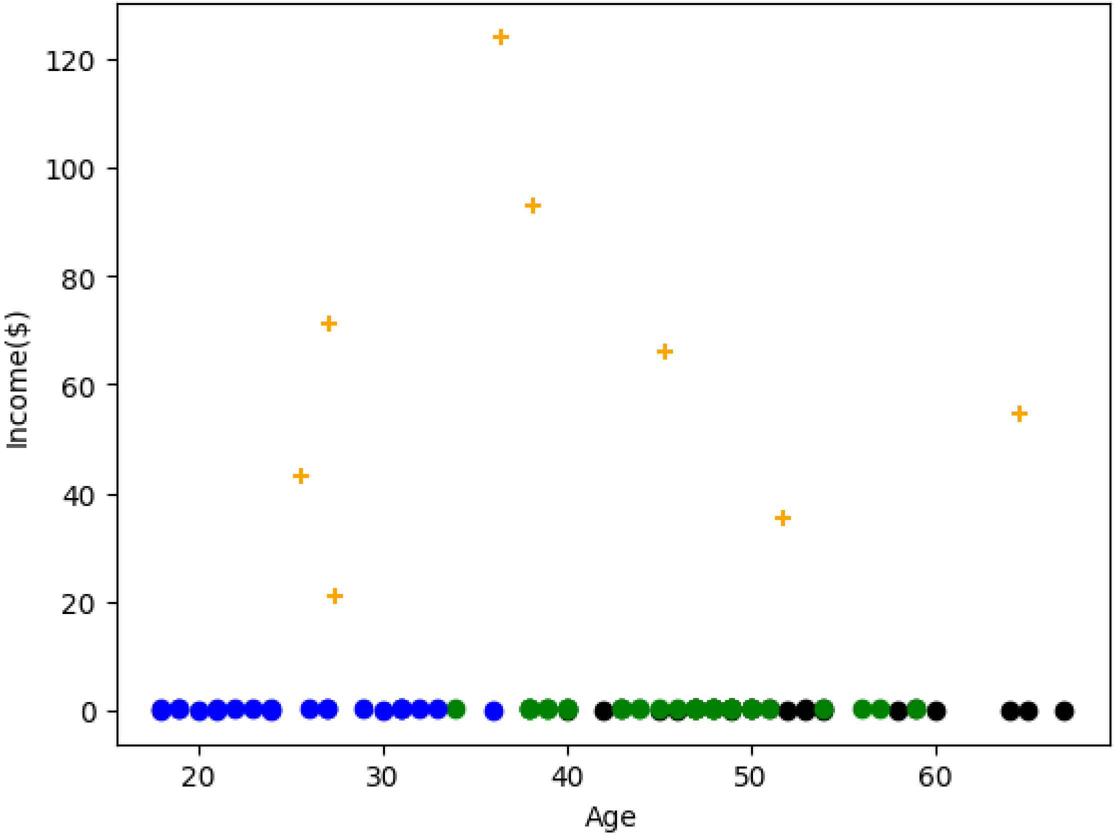


In [13]:
```
km.cluster_centers_
```

Out[13]:
```
array([[ 51.8       ,  35.28      ],
       [ 45.38888889,  66.05555556],
       [ 25.54545455,  43.18181818],
       [ 27.12820513,  71.38461538],
       [ 27.43478261,  21.        ],
       [ 38.21428571,  93.        ],
       [ 36.5       , 124.        ],
       [ 64.63157895,  54.68421053]])
```

In [14]:
```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="black")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="Orange",marker="+")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

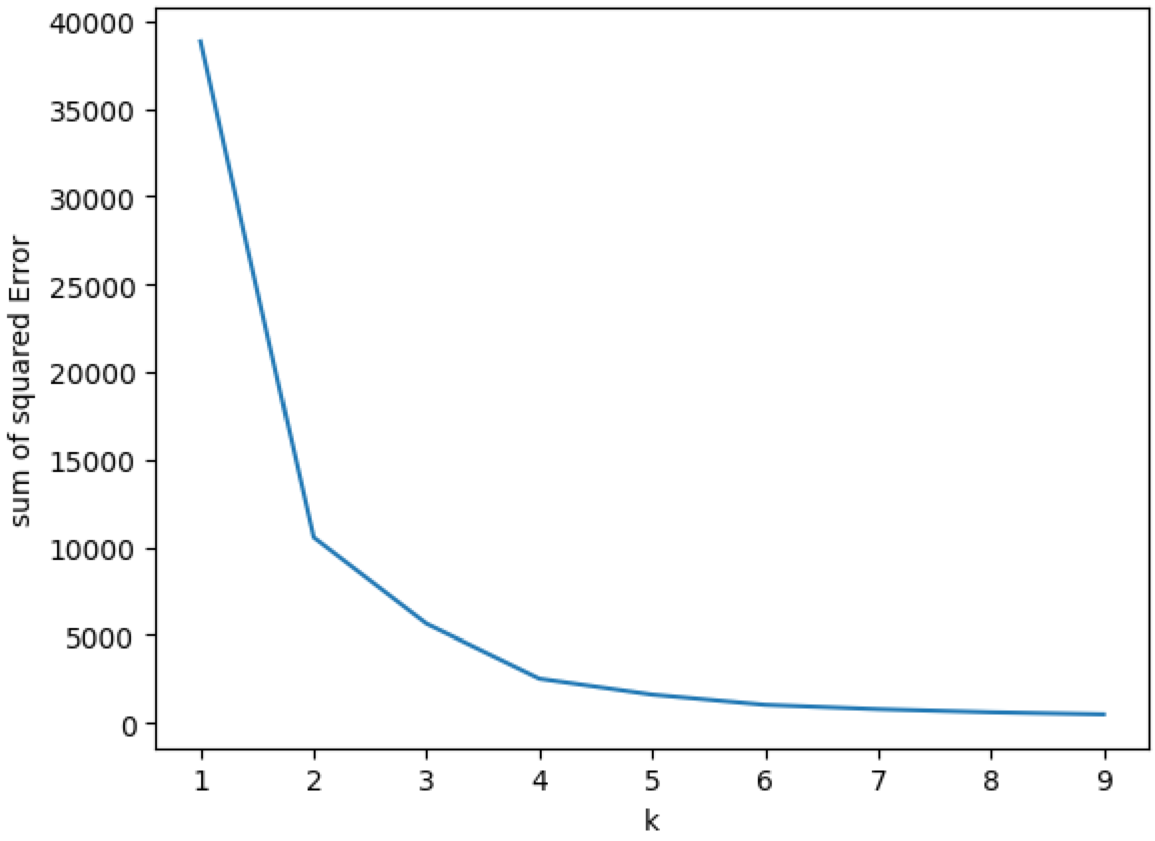
Out[14]:
```
Text(0, 0.5, 'Income($)')
```



In [15]:

```
k_rng=range(1,10)
sse=[]
for k in k_rng:
 km=KMeans(n_clusters=k)
 km.fit(df[["Age","Income($)"]])
 sse.append(km.inertia_)
 sse
```

```
C:\Users\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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\pucha\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureW
arning: 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(
In [16]: plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("sum of squared Error")

Out[16]: Text(0, 0.5, 'sum of squared Error')
```



```
In [ ]:
```