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

In [2]:
```
df=pd.read_csv(r"C:\Users\my pc\Downloads\Income1.csv")
df
```

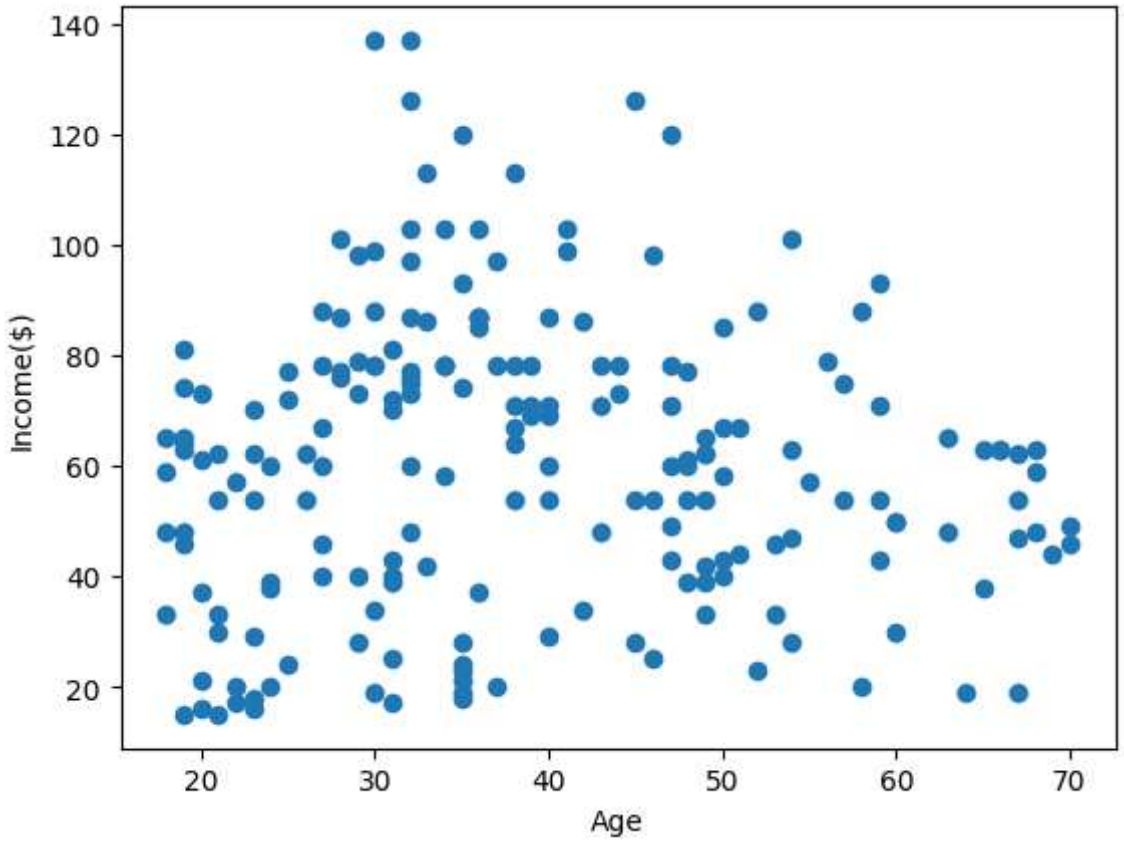
Out[2]:

	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
km=KMeans()
km
```

Out[4]:
```
KMeans()
```

In [5]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
```

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

In [6]:
```
df["cluster"]=y_predicted
```

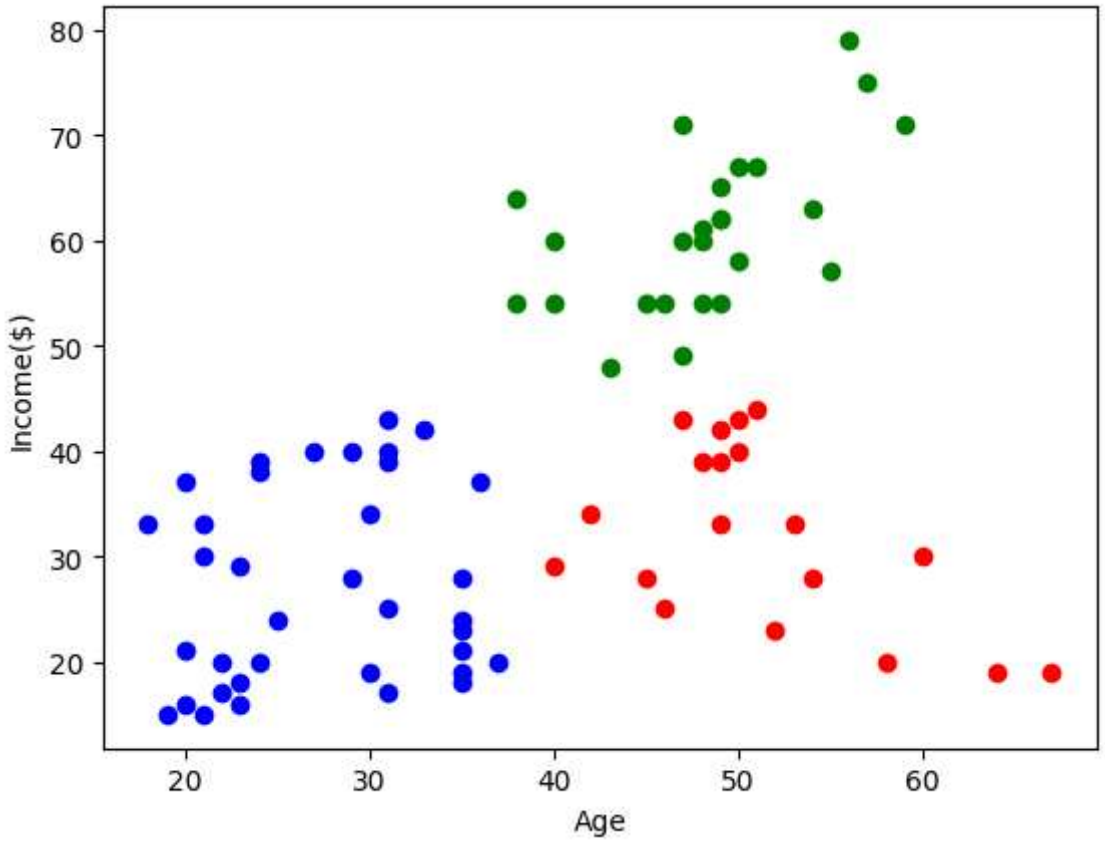
In [7]:
```
df.head()
```

Out[7]:

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

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

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



In [20]:
```
from sklearn.preprocessing import MinMaxScaler
```

In [21]:
```
Scaler=MinMaxScaler()
```

In [22]:
```
Scaler.fit(df[["Income($)"]])
```

Out[22]:
```
MinMaxScaler
MinMaxScaler()
```

In [23]:
```
df["Income($)"]=Scaler.transform(df[["Income($)"]])
df.head()
```

Out[23]:

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

In [24]:
```
Scaler.fit(df[["Age"]])
```

Out[24]:
```
MinMaxScaler
MinMaxScaler()
```

In [25]:
```
df["Age"]=Scaler.transform(df[["Age"]])
df.head()
```

Out[25]:

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

In [26]:
```
km=KMeans()
```

In [27]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
```

Out[27]:
```
array([5, 5, 5, 5, 2, 5, 2, 5, 4, 2, 4, 2, 4, 5, 2, 5, 2, 5, 1, 2, 2, 5,
       1, 2, 1, 2, 1, 2, 2, 5, 4, 5, 1, 5, 1, 5, 1, 2, 2, 5, 4, 5, 1, 2,
       1, 5, 1, 2, 2, 2, 1, 2, 2, 4, 1, 1, 1, 4, 3, 1, 4, 3, 4, 1, 4, 3,
       1, 4, 3, 2, 4, 1, 4, 4, 4, 3, 1, 1, 3, 1, 4, 6, 4, 1, 3, 1, 1, 3,
       6, 1, 4, 3, 1, 6, 6, 3, 1, 3, 1, 3, 3, 1, 4, 3, 1, 3, 4, 7, 4, 4,
       4, 3, 6, 3, 3, 3, 4, 7, 7, 7, 3, 6, 6, 6, 3, 6, 7, 6, 7, 6, 7, 6,
       3, 6, 3, 6, 7, 6, 3, 6, 7, 6, 6, 6, 3, 6, 7, 6, 6, 6, 7, 6, 7, 6,
       7, 6, 6, 6, 6, 6, 7, 6, 3, 6, 7, 6, 7, 6, 6, 6, 6, 6, 6, 6, 7, 6,
       7, 6, 7, 6, 0, 0, 7, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0])
```

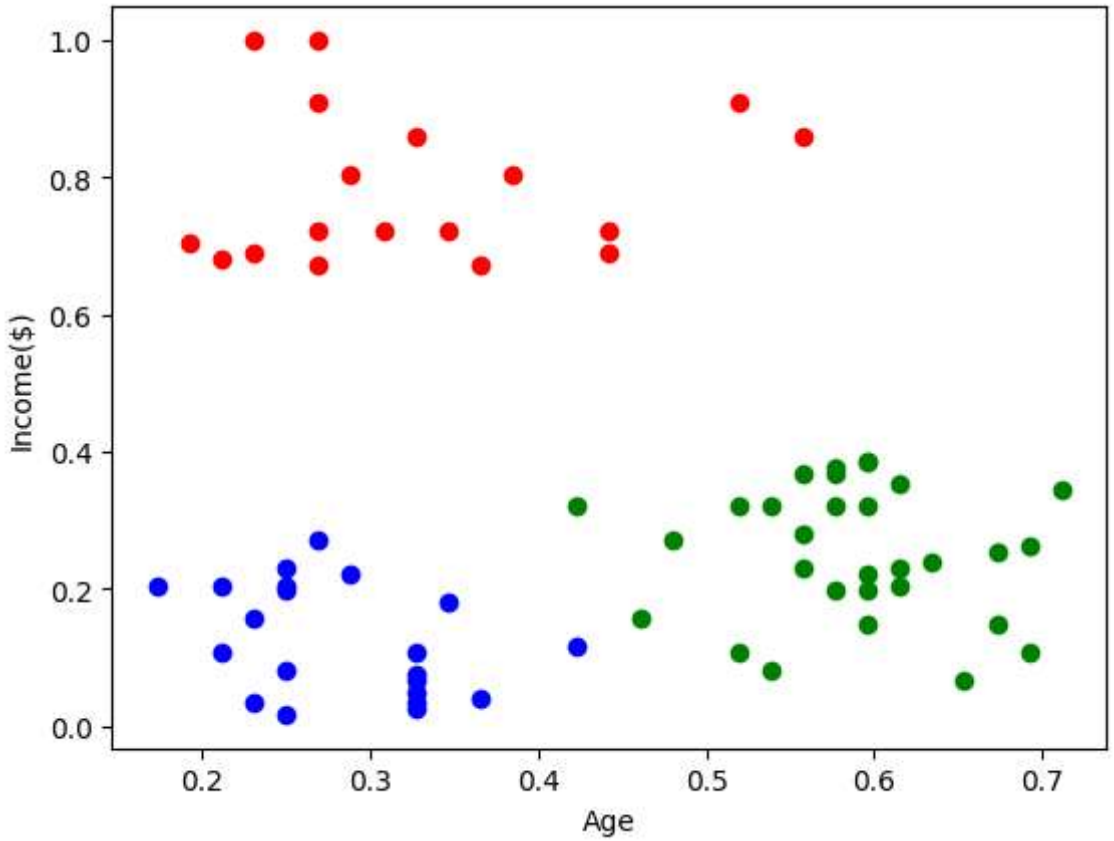
In [28]:
```
df["New Cluster"]=y_predicted
df.head()
```

Out[28]:

	Gender	Age	Income($)	cluster	New Cluster
0	Male	0.019231	0.000000	2	5
1	Male	0.057692	0.000000	2	5
2	Female	0.038462	0.008197	2	5
3	Female	0.096154	0.008197	2	5
4	Female	0.250000	0.016393	2	2

In [31]:
```
df1=df[df["New Cluster"]==0]
df2=df[df["New Cluster"]==1]
df3=df[df["New Cluster"]==2]
plt.scatter(df1["Age"],df1["Income($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[31]:
```
Text(0, 0.5, 'Income($)')
```

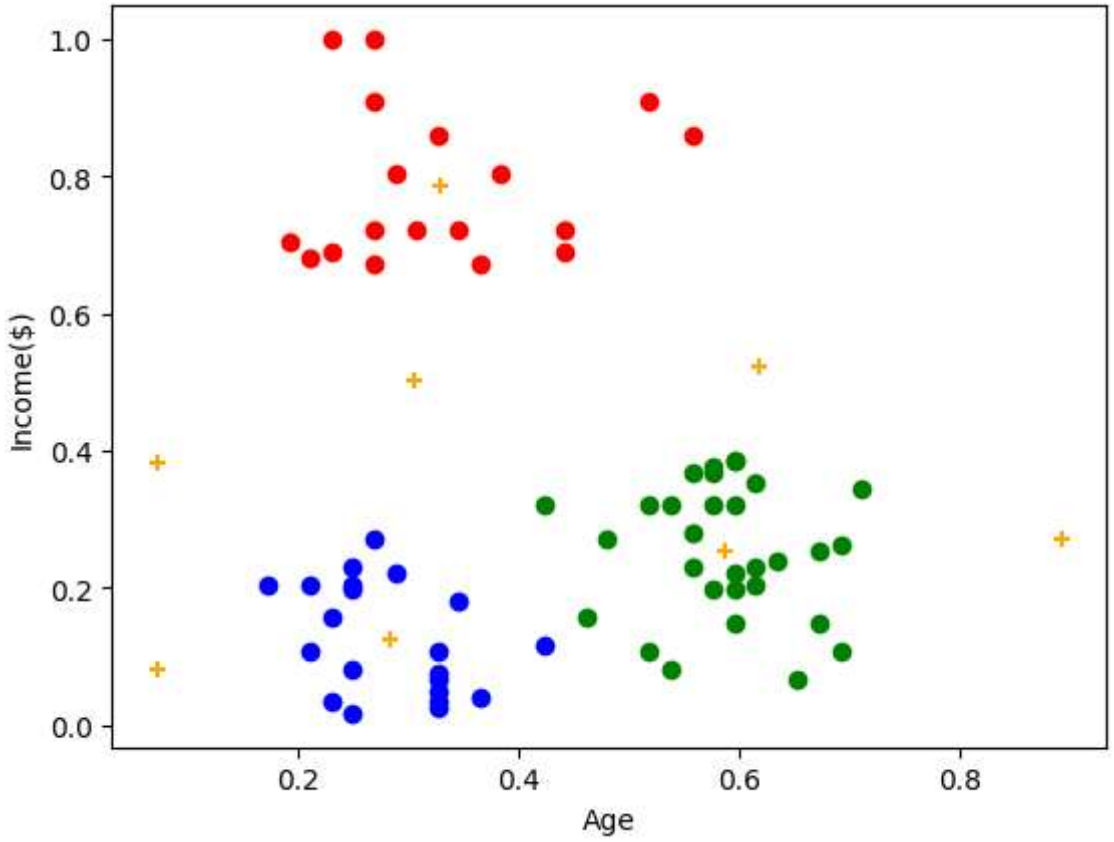


In [32]:
```
km.cluster_centers_
```

Out[32]:
```
array([[0.32905983, 0.78551913],
       [0.58717949, 0.25245902],
       [0.28388278, 0.1245121 ],
       [0.07322485, 0.38272383],
       [0.89262821, 0.27015027],
       [0.07239819, 0.08003857],
       [0.3059034 , 0.50247808],
       [0.61813187, 0.52185792]])
```

In [33]:
```
df1=df[df["New Cluster"]==0]
df2=df[df["New Cluster"]==1]
df3=df[df["New Cluster"]==2]
plt.scatter(df1["Age"],df1["Income($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],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[33]:
```
Text(0, 0.5, 'Income($)')
```



In [38]:
```
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_)
```

```
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
C:\Users\my pc\AppData\Local\Programs\Python\Python310\lib\site-packages\sklearn\cluster\_kmeans.py:870: Fut
ureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` ex
plicitly to suppress the warning
  warnings.warn(
```

In [39]:
```
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("Sum of Squared Error")
```

Out[39]: Text(0, 0.5, 'Sum of Squared Error')