In [1]:

```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

In [2]:

```
df=pd.read_csv(r"C:\Users\91720\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

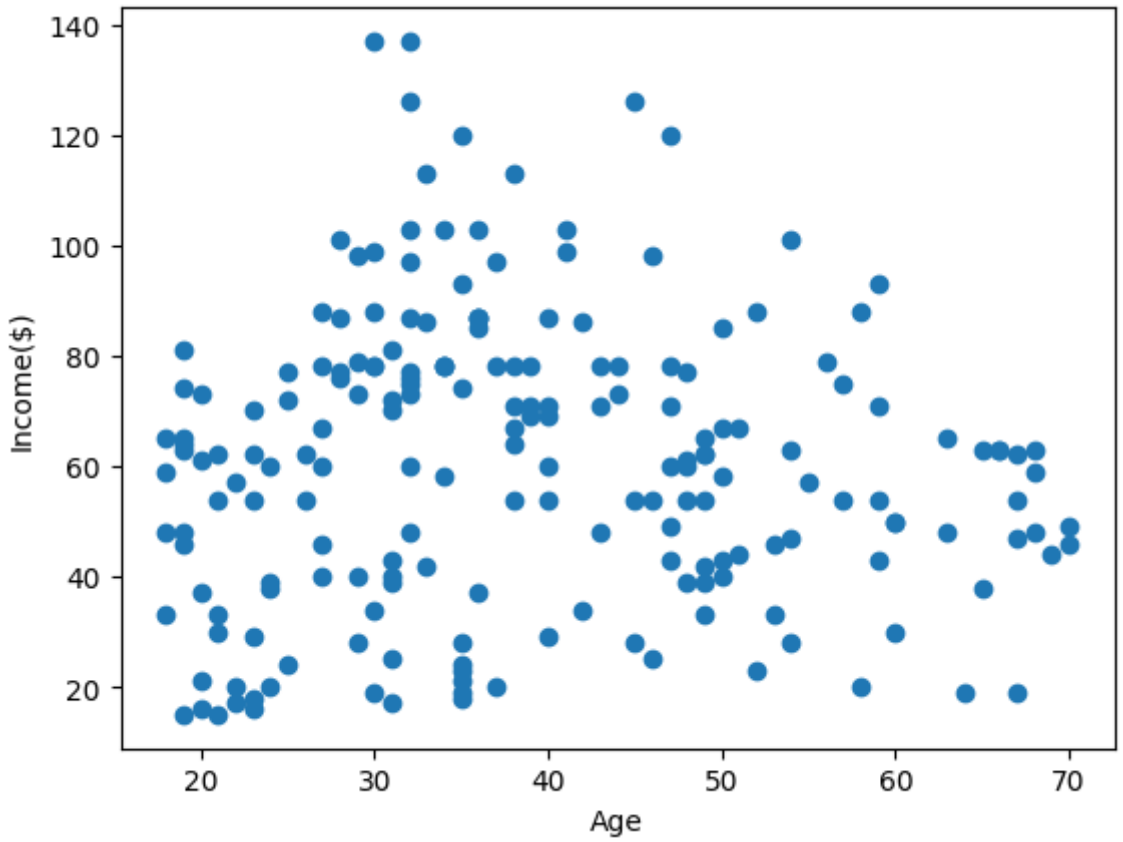
200 rows × 3 columns

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
KM=KMeans()
KM
```

Out[4]:

```
KMeans()
```

In [5]:

```
y_predicted=KM.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\91720\AppData\Local\Programs\Python\Python311\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[5]:

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

In [6]:

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

Out[6]:

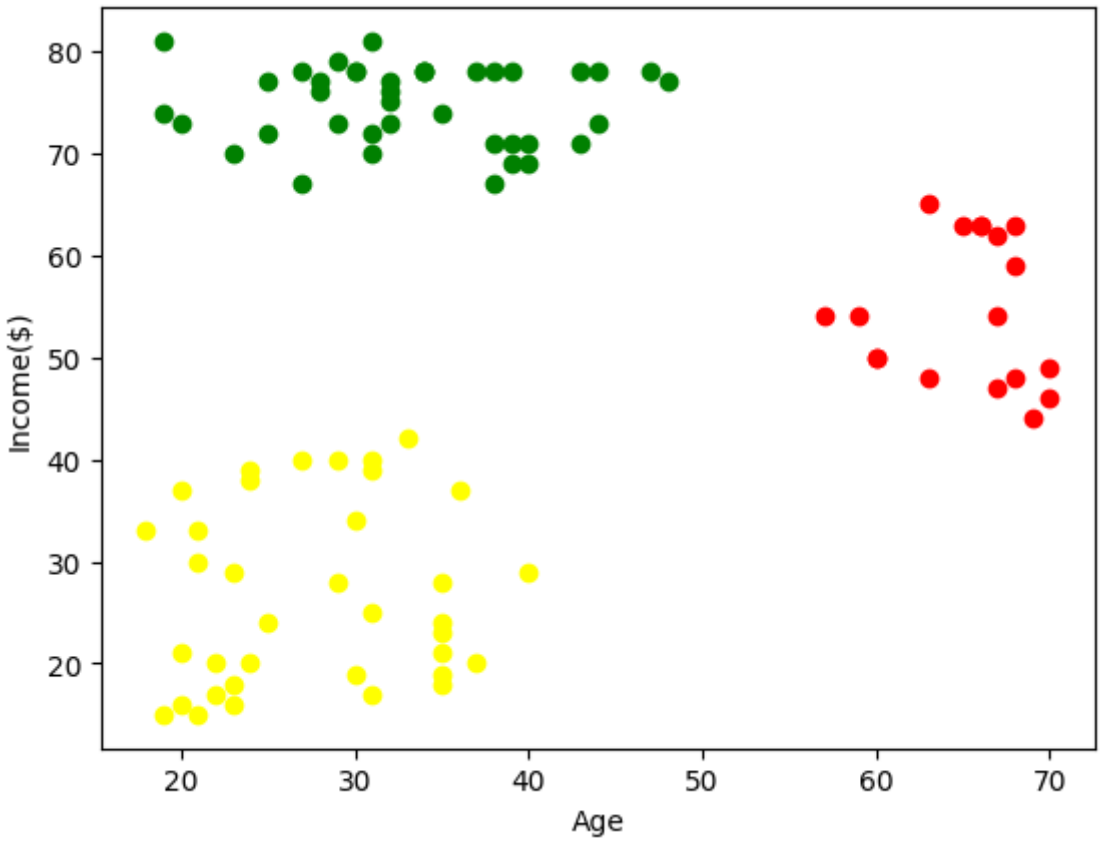
	Gender	Age	Income($)	cluster
0	Male	19	15	2
1	Male	21	15	2
2	Female	20	16	2
3	Female	23	16	2
4	Female	31	17	2

In [7]:

```
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($)"],color="yellow")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[7]:

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



In [8]:

```
from sklearn.preprocessing import MinMaxScaler
```

In [9]:

```
Scaler=MinMaxScaler()
```

In [10]:

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

Out[10]:

	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 [11]:

```
Scaler.fit(df[["Age"]])
df["Age"]=Scaler.transform(df[["Age"]])
df.head()
```

Out[11]:

	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 [12]:

```
KM=KMeans()
```

In [13]:

```
y_predicted=KM.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\91720\AppData\Local\Programs\Python\Python311\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([6, 6, 6, 6, 1, 6, 1, 6, 0, 1, 0, 1, 0, 6, 1, 6, 1, 6, 3, 1, 1, 6,
       3, 1, 3, 1, 3, 1, 1, 6, 0, 6, 3, 6, 3, 6, 3, 1, 1, 6, 0, 6, 3, 1,
       3, 6, 3, 1, 1, 1, 3, 1, 1, 0, 3, 3, 3, 0, 1, 3, 0, 7, 0, 3, 0, 7,
       3, 0, 7, 1, 0, 3, 0, 0, 0, 7, 3, 3, 7, 3, 0, 2, 0, 3, 7, 3, 3, 7,
       2, 3, 0, 7, 3, 2, 2, 7, 3, 7, 3, 7, 7, 3, 0, 7, 3, 7, 0, 5, 0, 0,
       0, 7, 2, 7, 7, 7, 0, 5, 5, 5, 7, 2, 2, 2, 7, 2, 5, 2, 5, 2, 5, 2,
       7, 2, 7, 2, 5, 2, 7, 2, 5, 2, 2, 2, 7, 2, 5, 2, 2, 2, 5, 2, 5, 2,
       5, 2, 2, 2, 2, 2, 5, 2, 7, 2, 5, 2, 5, 2, 2, 2, 2, 2, 2, 2, 5, 2,
       5, 2, 5, 2, 4, 4, 5, 4, 4, 4, 5, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
       4, 4])
```

In [14]:

```
df['New cluster']=y_predicted
df.head()
```

Out[14]:

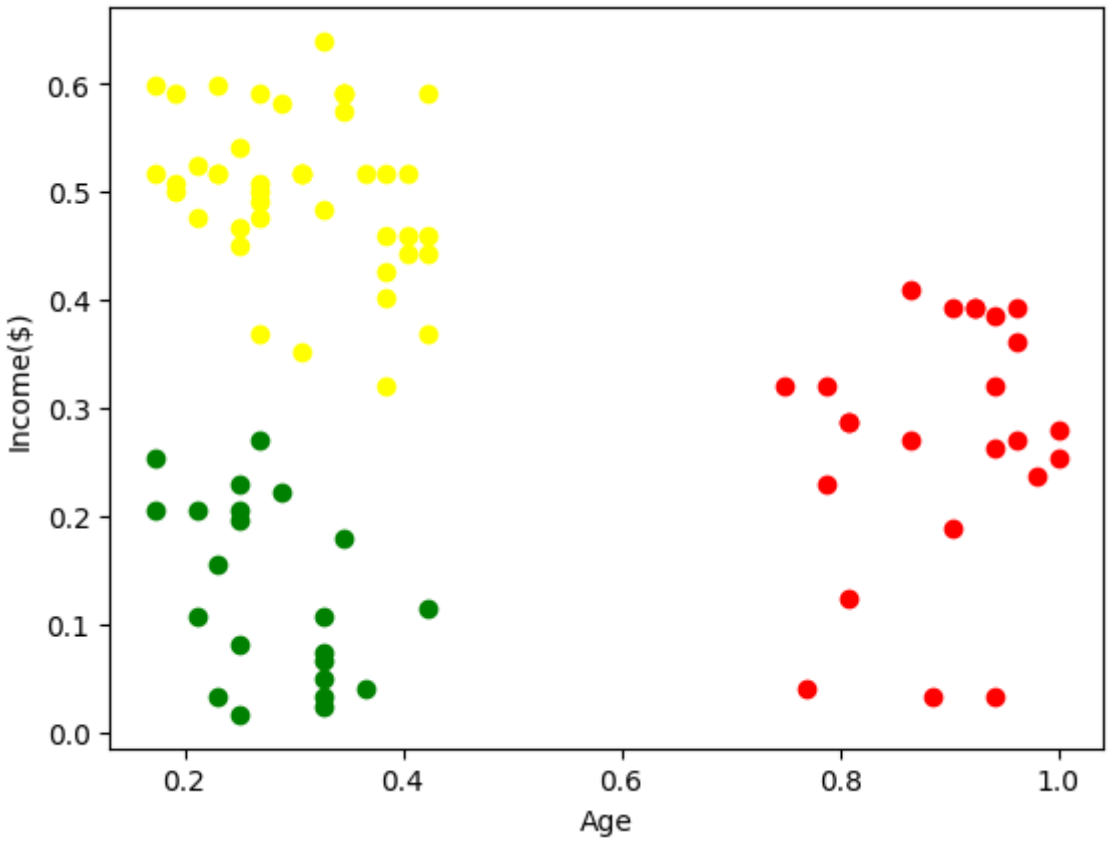
	Gender	Age	Income($)	cluster	New cluster
0	Male	0.019231	0.000000	2	6
1	Male	0.057692	0.000000	2	6
2	Female	0.038462	0.008197	2	6
3	Female	0.096154	0.008197	2	6
4	Female	0.250000	0.016393	2	1

In [15]:

```
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($)"],color="yellow")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[15]:

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



In [16]:

```
KM.cluster_centers_
```

Out[16]:

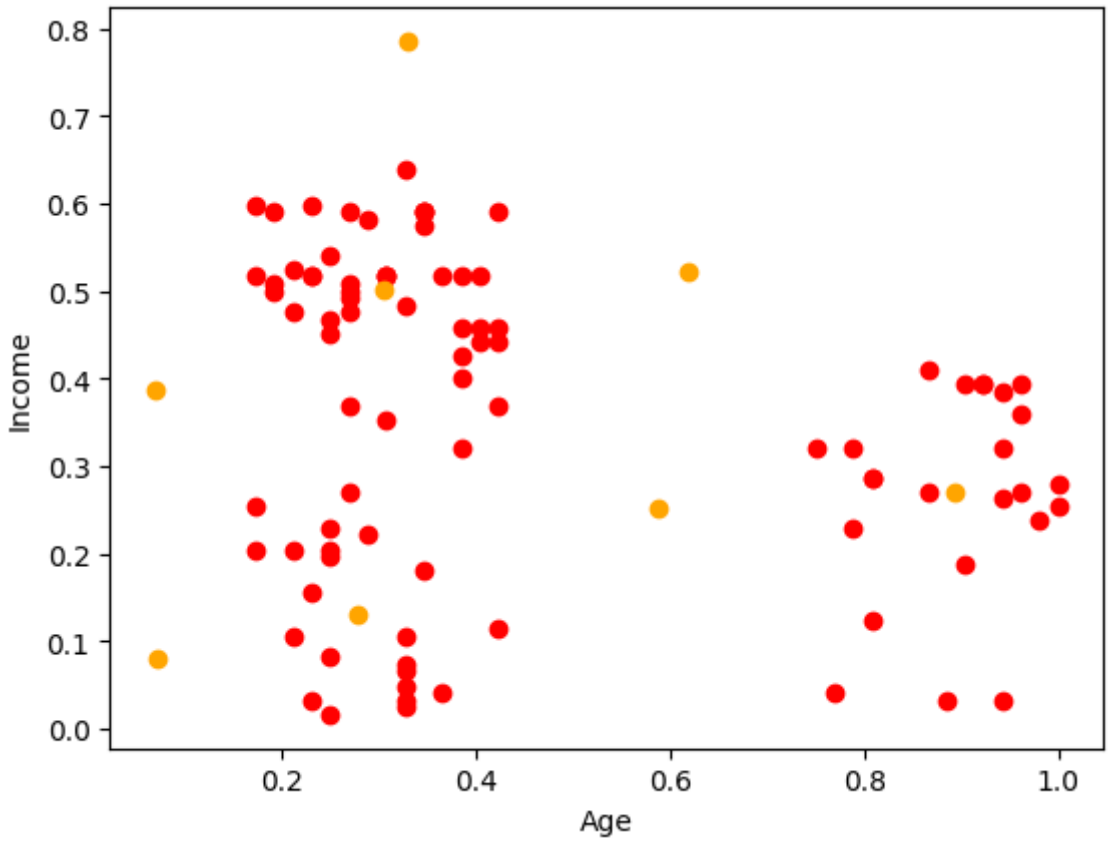
```
array([[0.89262821, 0.27015027],
       [0.27884615, 0.13040238],
       [0.3059034 , 0.50247808],
       [0.58717949, 0.25245902],
       [0.32905983, 0.78551913],
       [0.61813187, 0.52185792],
       [0.07239819, 0.08003857],
       [0.06923077, 0.38786885]])
```

In [17]:

```
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="red")
plt.scatter(df3["Age"],df3["Income($)"],color="red")
plt.scatter(KM.cluster_centers_[:,0],KM.cluster_centers_[:,1],color="orange")
plt.xlabel("Age")
plt.ylabel("Income")
```

Out[17]:

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



In [18]:

```
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\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
C:\Users\91720\AppData\Local\Programs\Python\Python311\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(
```