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

In [2]:
```
df=pd.read_csv(r"C:\Users\Mastan Reddy\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]:
```
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]:
```
df.tail()
```

Out[4]:

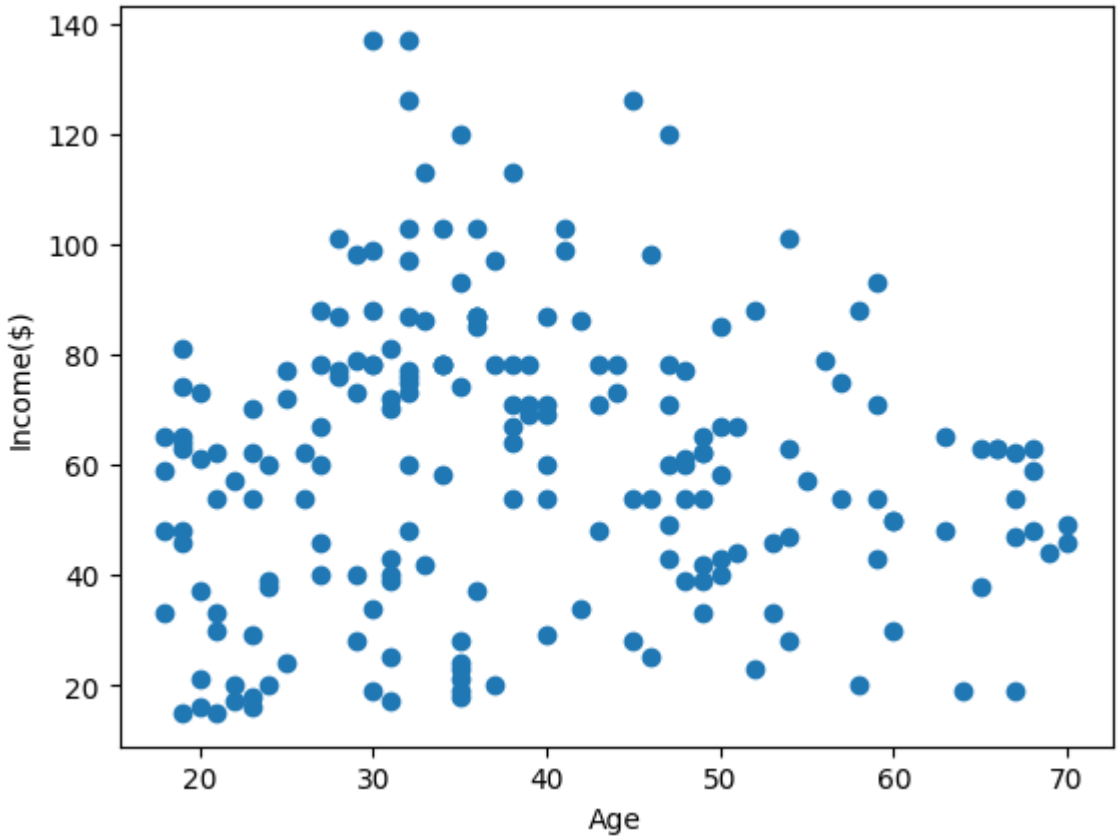
	Gender	Age	Income($)
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

In [5]:
```
df.shape
```

Out[5]:
```
(200, 3)
```

In [6]:
```
plt.scatter(df["Age"],df["Income($)"])
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[6]:
```
Text(0, 0.5, 'Income($)')
```



In [7]:
```
from sklearn.cluster import KMeans
```

In [8]:
```
kM=KMeans()
kM
```

Out[8]:
```
KMeans()
```

In [9]:
```
y_predicted = kM.fit_predict(df[["Age","Income($)"]])
y_predicted
```

Out[9]:
```
array([1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 1, 1, 2, 1, 1, 1,
       2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 4, 2, 4, 2, 4,
       2, 4, 2, 4, 4, 4, 2, 4, 4, 2, 2, 2, 2, 7, 4, 2, 7, 4, 7, 2, 7, 4,
       2, 7, 4, 4, 7, 2, 7, 7, 7, 4, 0, 0, 4, 0, 7, 4, 7, 0, 4, 0, 7, 4,
       4, 0, 7, 4, 0, 0, 4, 4, 0, 4, 0, 4, 4, 0, 7, 4, 0, 4, 7, 0, 7, 7,
       7, 4, 0, 4, 4, 4, 7, 0, 0, 0, 6, 0, 0, 0, 6, 6, 0, 0, 0, 6, 0, 0,
       6, 6, 6, 6, 0, 6, 6, 6, 0, 6, 6, 6, 6, 6, 0, 6, 6, 6, 6, 6, 0, 6,
       0, 6, 6, 6, 6, 6, 0, 6, 6, 6, 5, 6, 5, 6, 5, 6, 5, 6, 5, 5, 5, 6,
       5, 6, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 3, 3, 3, 3, 3,
       3, 3])
```

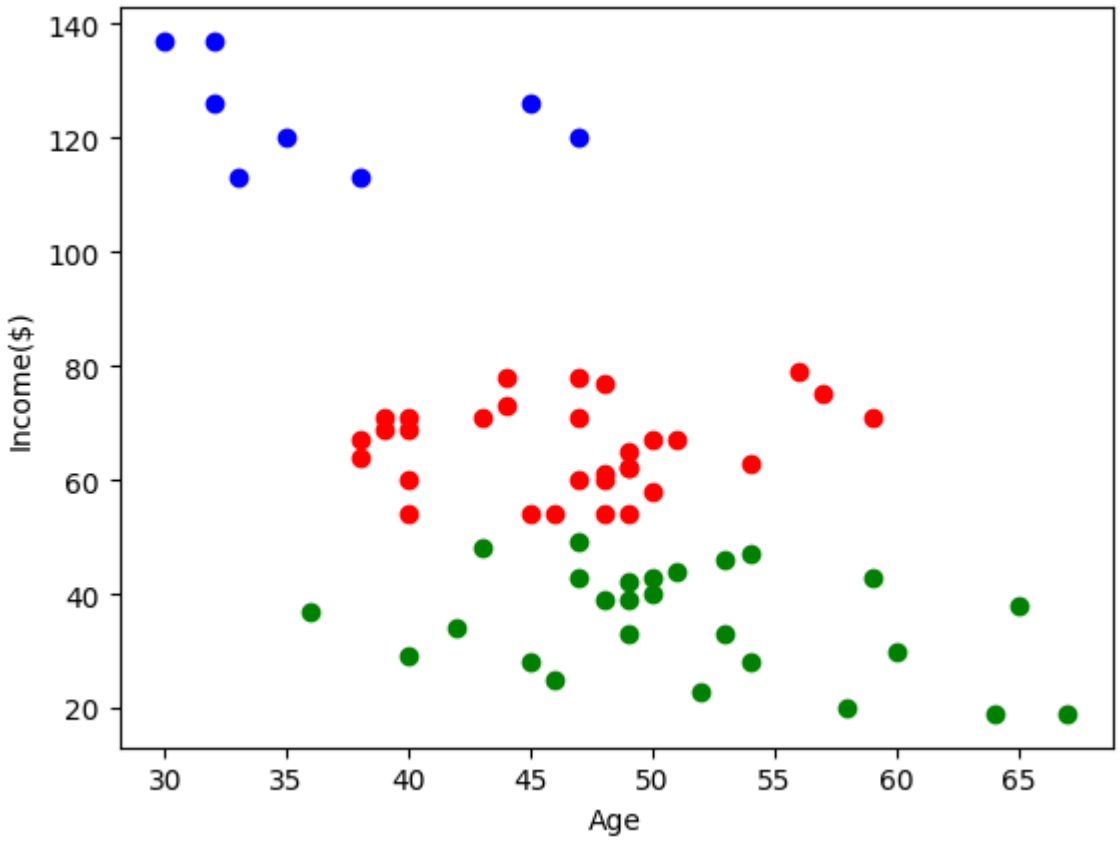
In [10]:
```
df["cluster"]=y_predicted
df.head()
```

Out[10]:

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

In [11]:
```
df1 = df[df.cluster == 0]
df2 = df[df.cluster == 2]
df3 = df[df.cluster == 3]
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[11]:
```
Text(0, 0.5, 'Income($)')
```



In [12]:
```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["Income($)"]])
df["Income($)"]=scaler.transform(df[["Income($)"]])
df.head()
```

Out[12]:

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

In [13]:
```
scaler.fit(df[["Age"]])
df["Age"]=scaler.transform(df[["Age"]])
df.head()
```

Out[13]:

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

In [14]:
```
km=KMeans()
```

In [15]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

Out[15]:
```
array([3, 3, 3, 3, 0, 3, 0, 3, 4, 0, 4, 0, 7, 3, 0, 3, 0, 3, 7, 0, 0, 3,
       7, 0, 7, 0, 7, 0, 0, 3, 4, 3, 7, 3, 7, 3, 7, 0, 0, 3, 4, 3, 7, 0,
       7, 3, 7, 0, 0, 0, 7, 0, 0, 4, 7, 7, 7, 4, 6, 7, 4, 6, 4, 7, 4, 6,
       7, 4, 6, 0, 4, 7, 4, 4, 4, 6, 7, 7, 6, 7, 4, 5, 4, 7, 6, 7, 7, 6,
       5, 7, 4, 6, 1, 5, 5, 6, 1, 6, 1, 6, 6, 1, 4, 6, 1, 6, 4, 1, 4, 4,
       4, 6, 5, 6, 6, 6, 4, 1, 1, 1, 6, 5, 5, 5, 6, 5, 1, 5, 1, 5, 1, 5,
       6, 5, 6, 5, 1, 5, 6, 5, 1, 5, 5, 5, 6, 5, 1, 5, 5, 5, 1, 5, 1, 5,
       1, 5, 5, 5, 5, 5, 1, 5, 6, 5, 1, 5, 1, 5, 5, 5, 5, 5, 5, 5, 1, 5,
       1, 5, 1, 2, 2, 2, 1, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2])
```

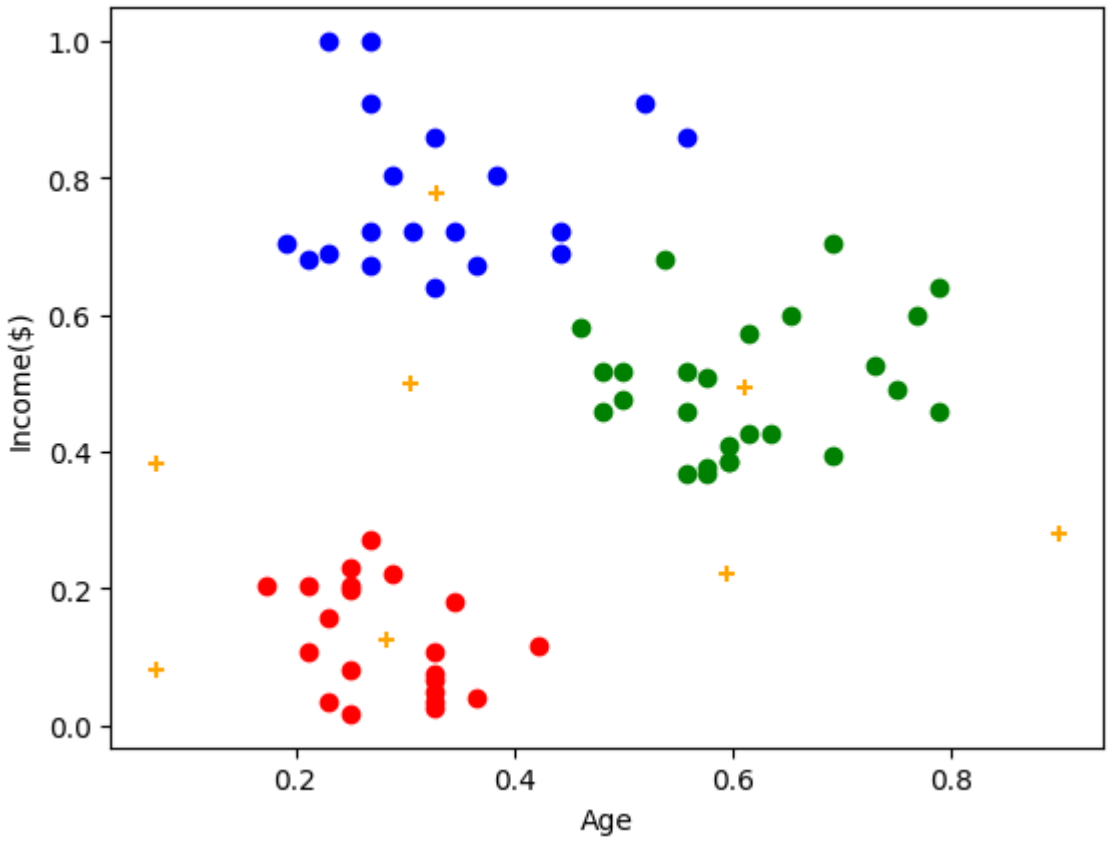
In [16]:
```
df["New Cluster"] = y_predicted
df.head()
```

Out[16]:

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

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="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="orange",m
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[17]:
```
Text(0, 0.5, 'Income($)')
```



In [18]:
```
km.cluster_centers_
```

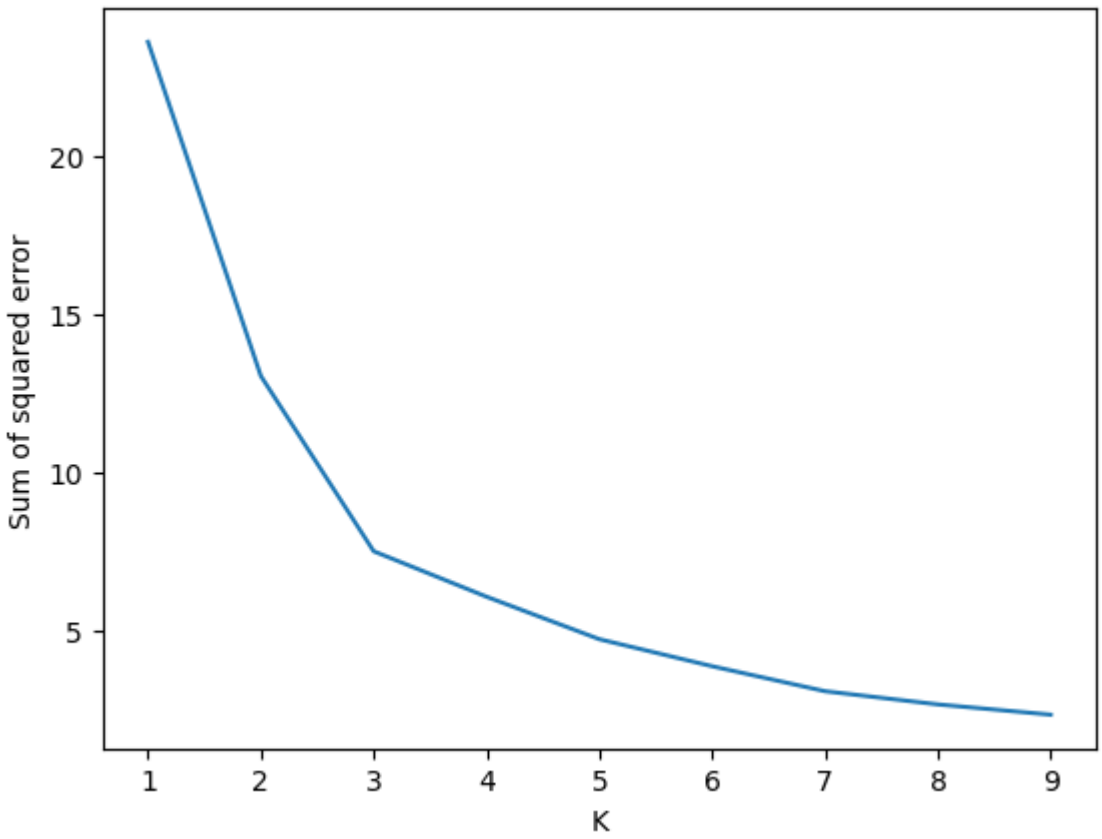
Out[18]:
```
array([[0.28388278, 0.1245121 ],
       [0.61094675, 0.49401009],
       [0.32894737, 0.77782571],
       [0.07239819, 0.08003857],
       [0.89799331, 0.28011404],
       [0.30540293, 0.49921936],
       [0.07322485, 0.38272383],
       [0.5954142 , 0.2203657 ]])
```

In [19]:
```
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_)
    print(sse)
```

```
[23.583906150363607]
[23.583906150363607, 13.028938428018286]
[23.583906150363607, 13.028938428018286, 7.49210786858601]
[23.583906150363607, 13.028938428018286, 7.49210786858601, 6.055697956177862
6]
[23.583906150363607, 13.028938428018286, 7.49210786858601, 6.055697956177862
6, 4.713416604872824]
[23.583906150363607, 13.028938428018286, 7.49210786858601, 6.055697956177862
6, 4.713416604872824, 3.8591087368887145]
[23.583906150363607, 13.028938428018286, 7.49210786858601, 6.055697956177862
6, 4.713416604872824, 3.8591087368887145, 3.0701267041545526]
[23.583906150363607, 13.028938428018286, 7.49210786858601, 6.055697956177862
6, 4.713416604872824, 3.8591087368887145, 3.0701267041545526, 2.6535016786291
08]
[23.583906150363607, 13.028938428018286, 7.49210786858601, 6.055697956177862
6, 4.713416604872824, 3.8591087368887145, 3.0701267041545526, 2.6535016786291
08, 2.329176554466517]
```

In [20]:
```
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of squared error")
```

Out[20]: Text(0, 0.5, 'Sum of squared error')



In [ ]: