

In [3]:

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

In [5]:

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

Out[5]:

	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

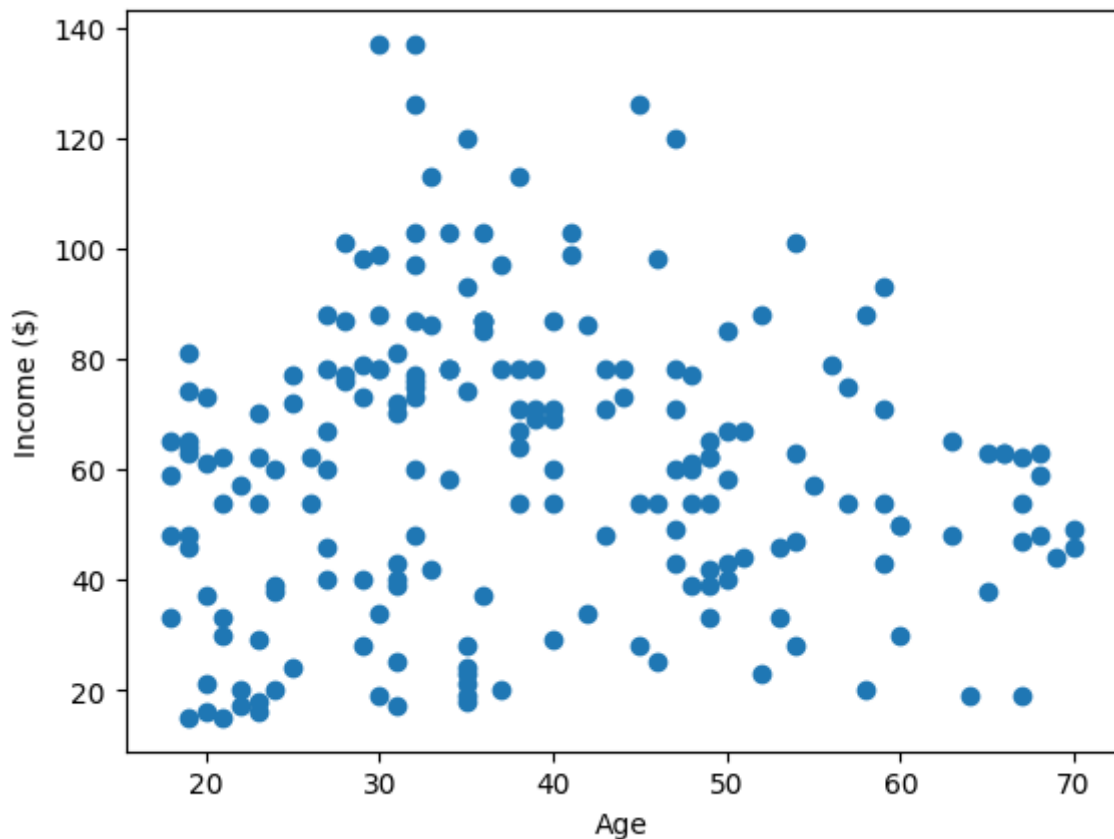
200 rows × 3 columns

In [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 a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [9]:

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

C:\Users\raja\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(
```

Out[9]:

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

In [10]:

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

Out[10]:

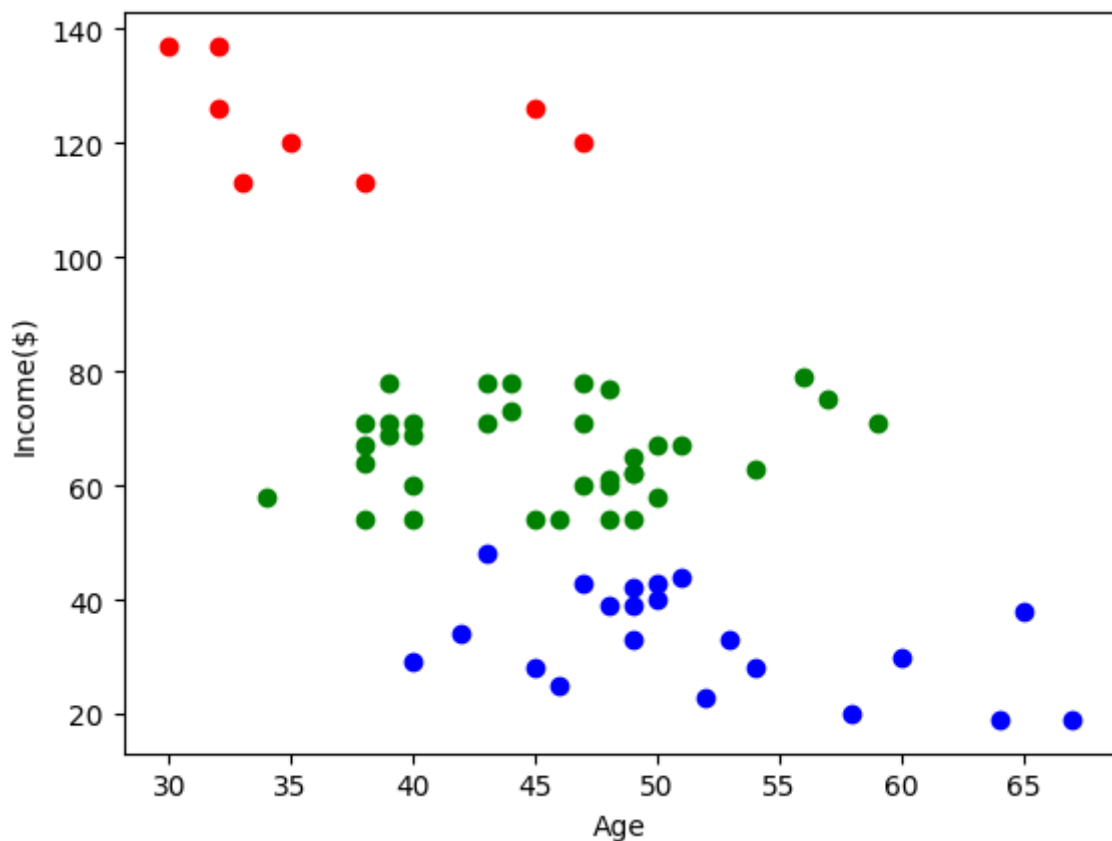
	Gender	Age	Income(\$)	Cluster
0	Male	19	15	7
1	Male	21	15	7
2	Female	20	16	7
3	Female	23	16	7
4	Female	31	17	7

In [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
```

In [13]:

```
scaler=MinMaxScaler()
```

In [14]:

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

Out[14]:

	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 [15]:

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

Out[15]:

	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 [17]:

```
km=KMeans()
```

In [18]:

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

C:\Users\raja\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(
```

Out[18]:

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

In [19]:

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

Out[19]:

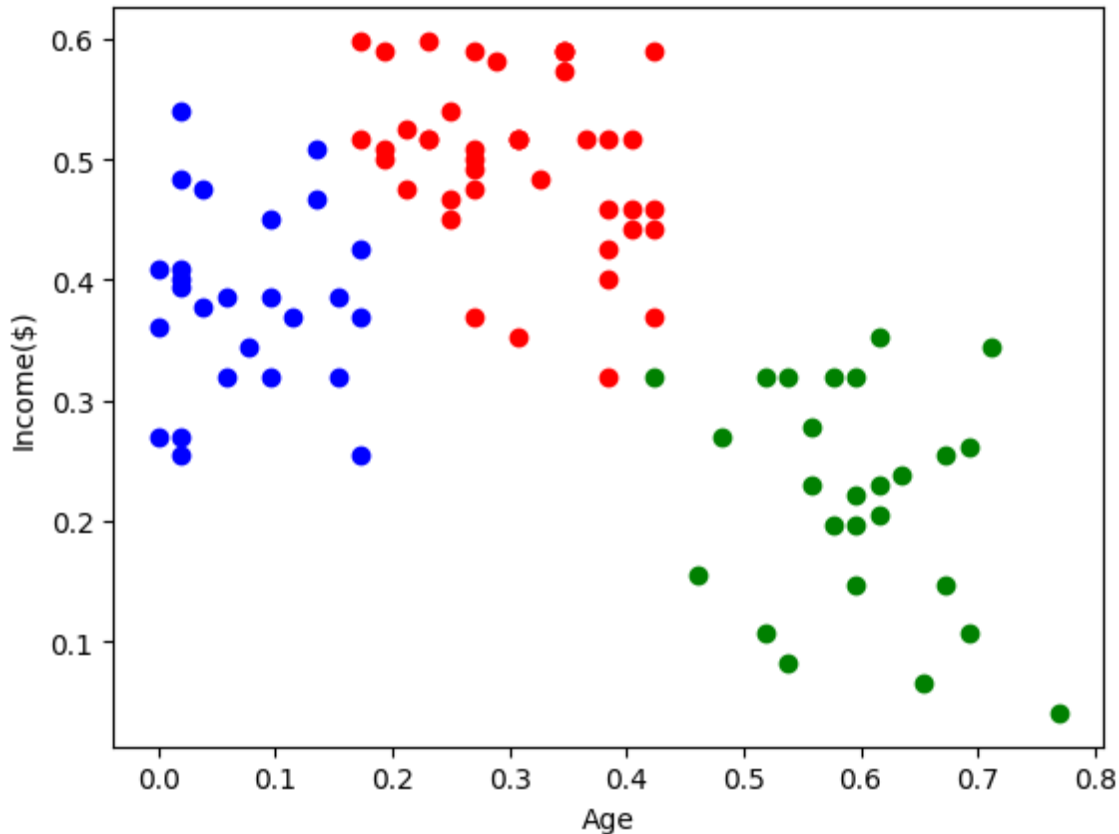
	Gender	Age	Income(\$)	Cluster	New cluster
0	Male	0.019231	0.000000	7	6
1	Male	0.057692	0.000000	7	6
2	Female	0.038462	0.008197	7	6
3	Female	0.096154	0.008197	7	6
4	Female	0.250000	0.016393	7	7

In [21]:

```
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[21]:

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



In [22]:

```
km.cluster_centers_
```

Out[22]:

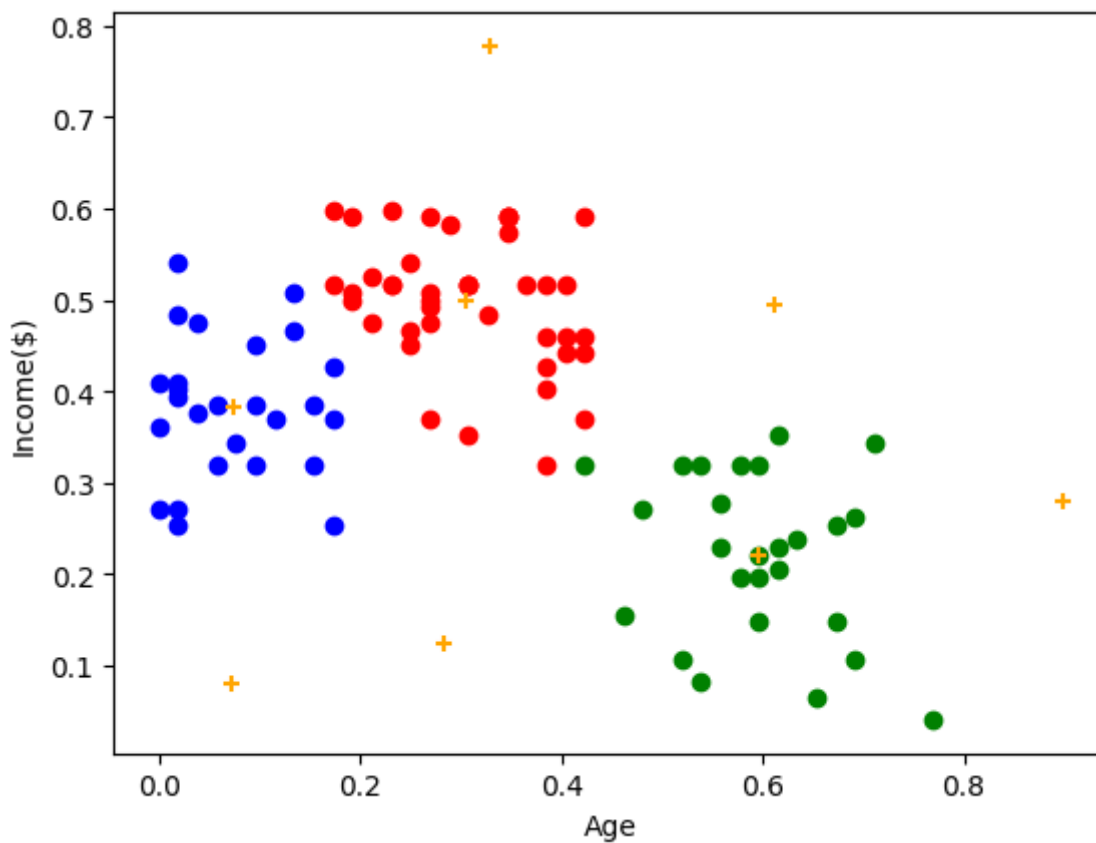
```
array([[0.30540293, 0.49921936],
       [0.5954142 , 0.2203657 ],
       [0.07322485, 0.38272383],
       [0.89799331, 0.28011404],
       [0.32894737, 0.77782571],
       [0.61094675, 0.49401009],
       [0.07239819, 0.08003857],
       [0.28388278, 0.1245121 ]])
```

In [25]:

```
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[25]:

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



In [32]:

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


In [33]:

```

for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[["Age", "Income($)"]])
    sse.append(km.inertia_)
#km.inertia_ will give you the value of sum of square errorprint(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")

```

C:\Users\raja\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\raja\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\raja\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\raja\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\raja\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\raja\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\raja\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\raja\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\raja\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(
```

Out[33]:

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

