In [7]:

```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
import warnings
warnings.simplefilter(action='ignore')
```

In [8]:

```
df = pd.read_csv(r"C:\Users\HP\OneDrive\Documents\Income.csv")
df
```

Out[8]:

	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

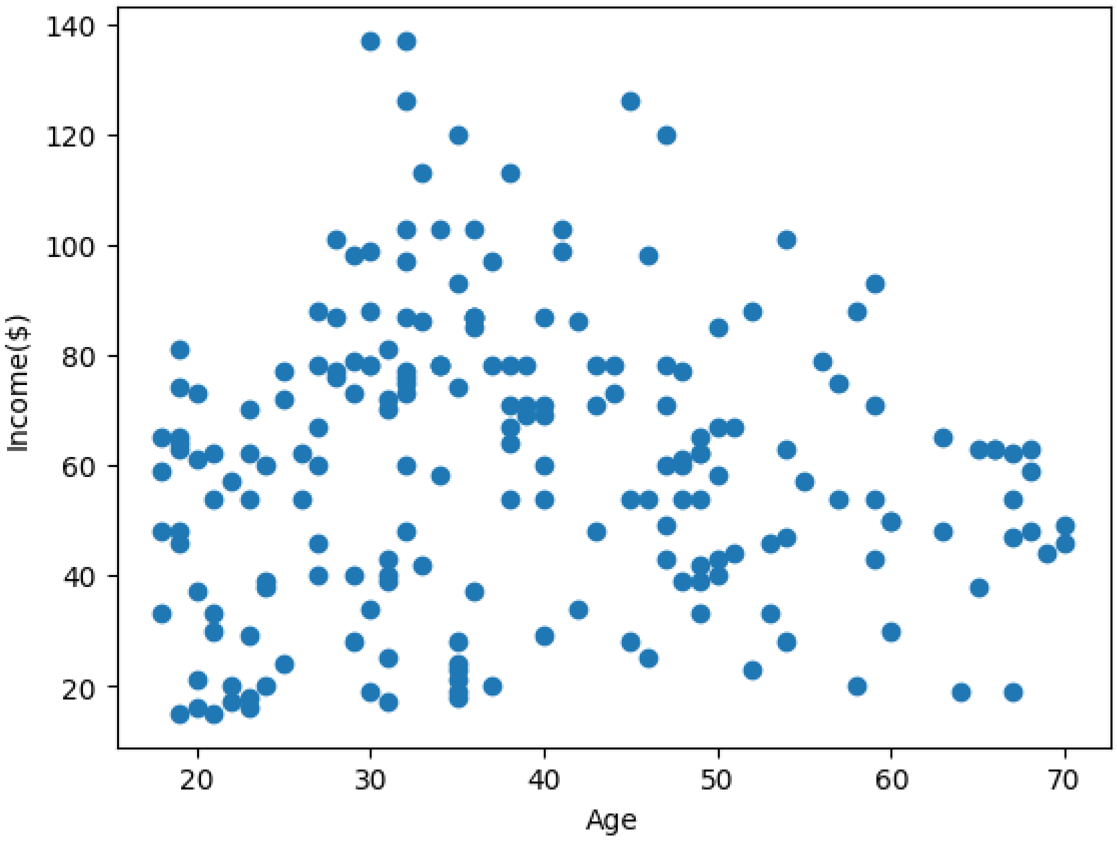
200 rows × 3 columns

In [9]:

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

Out[9]:

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



In [10]:

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

In [11]:

```
km = KMeans()
km
```

Out[11]:

```
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 [12]:

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

Out[12]:

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

In [13]:

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

Out[13]:

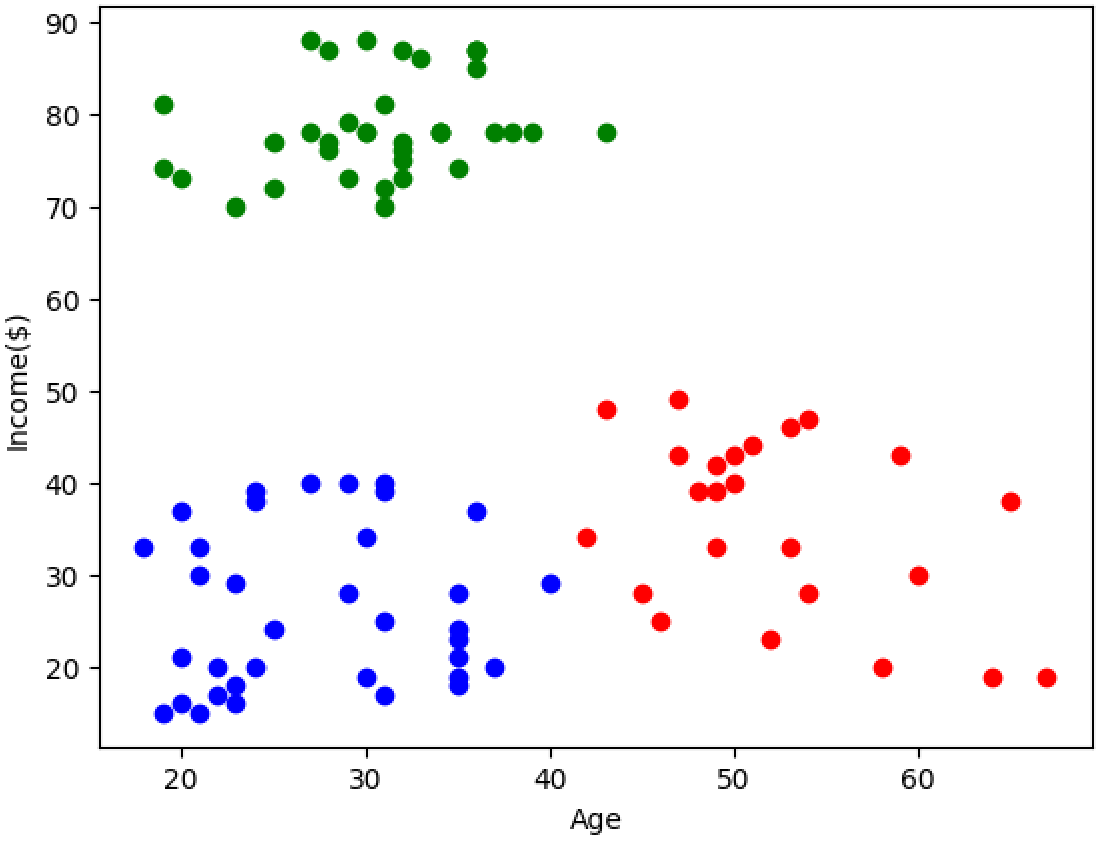
	Gender	Age	Income($)	Cluster
0	Male	19	15	2
1	Male	21	15	2
2	Female	20	16	2
3	Female	23	16	2
4	Female	31	17	2

In [14]:

```
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[14]:

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



In [15]:

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

In [16]:

```
scaler = MinMaxScaler()
```

In [17]:

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

Out[17]:

	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 [18]:

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

Out[18]:

	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 [19]:

```
km =KMeans()
```

In [21]:

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

Out[21]:

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

In [23]:

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

Out[23]:

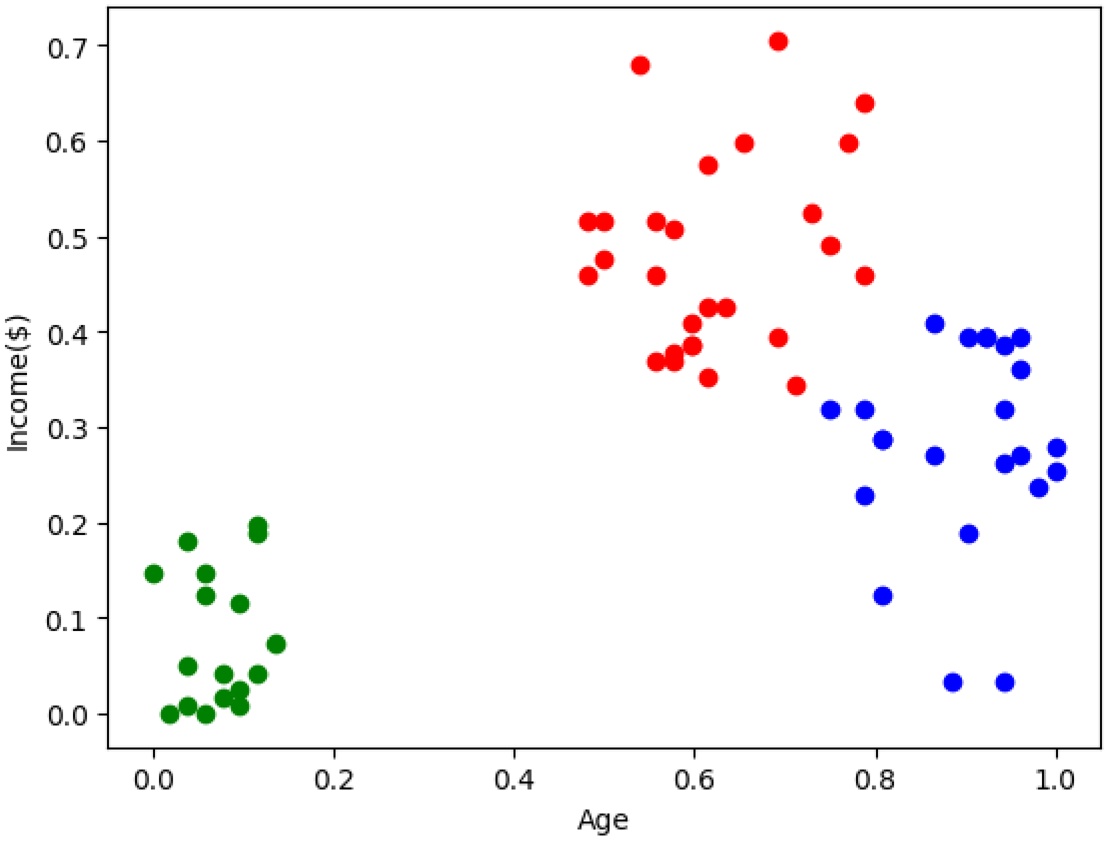
	Gender	Age	Income($)	Cluster	New Cluster
0	Male	0.019231	0.000000	2	1
1	Male	0.057692	0.000000	2	1
2	Female	0.038462	0.008197	2	1
3	Female	0.096154	0.008197	2	1
4	Female	0.250000	0.016393	2	3

In [24]:

```
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[24]:

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



In [25]:

```
km.cluster_centers_
```

Out[25]:

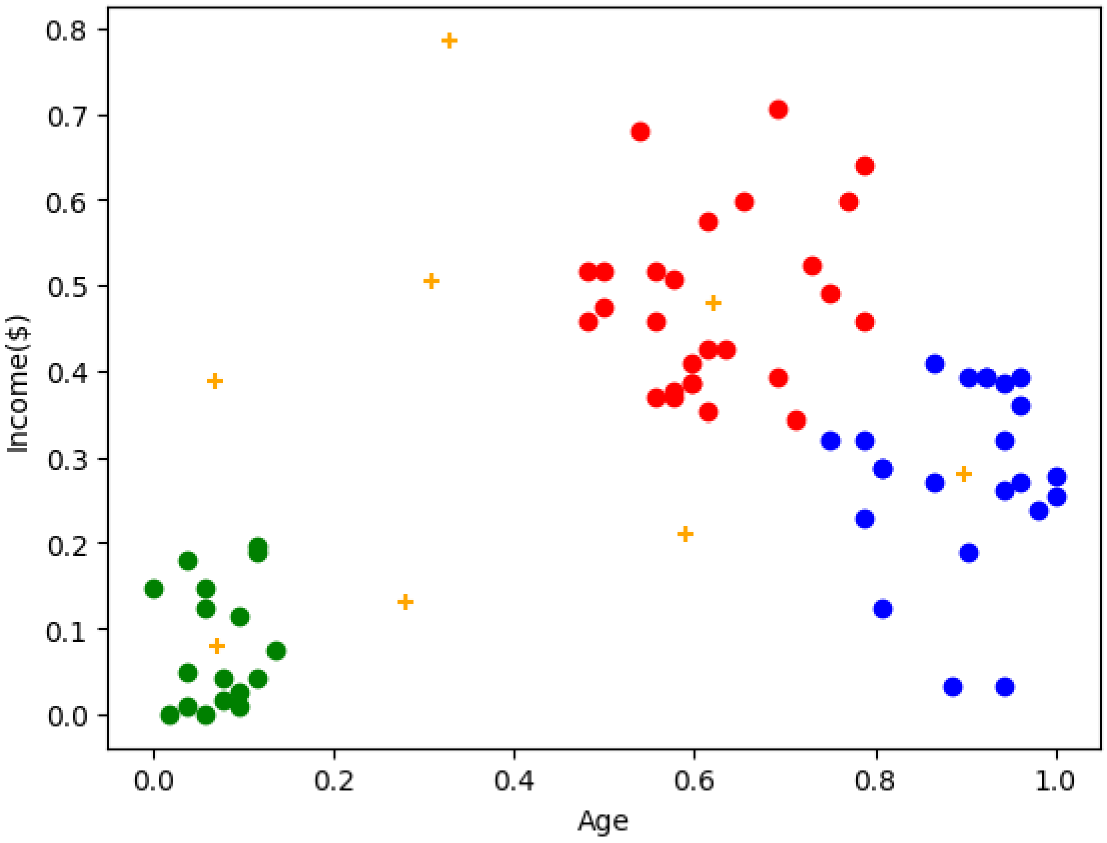
```
array([[0.62037037, 0.47996357],
       [0.07239819, 0.08003857],
       [0.89799331, 0.28011404],
       [0.27884615, 0.13040238],
       [0.30944056, 0.50428465],
       [0.06923077, 0.38786885],
       [0.58974359, 0.20969945],
       [0.32905983, 0.78551913]])
```

In [26]:

```
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[26]:

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



In [28]:

```
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
```

Out[28]:

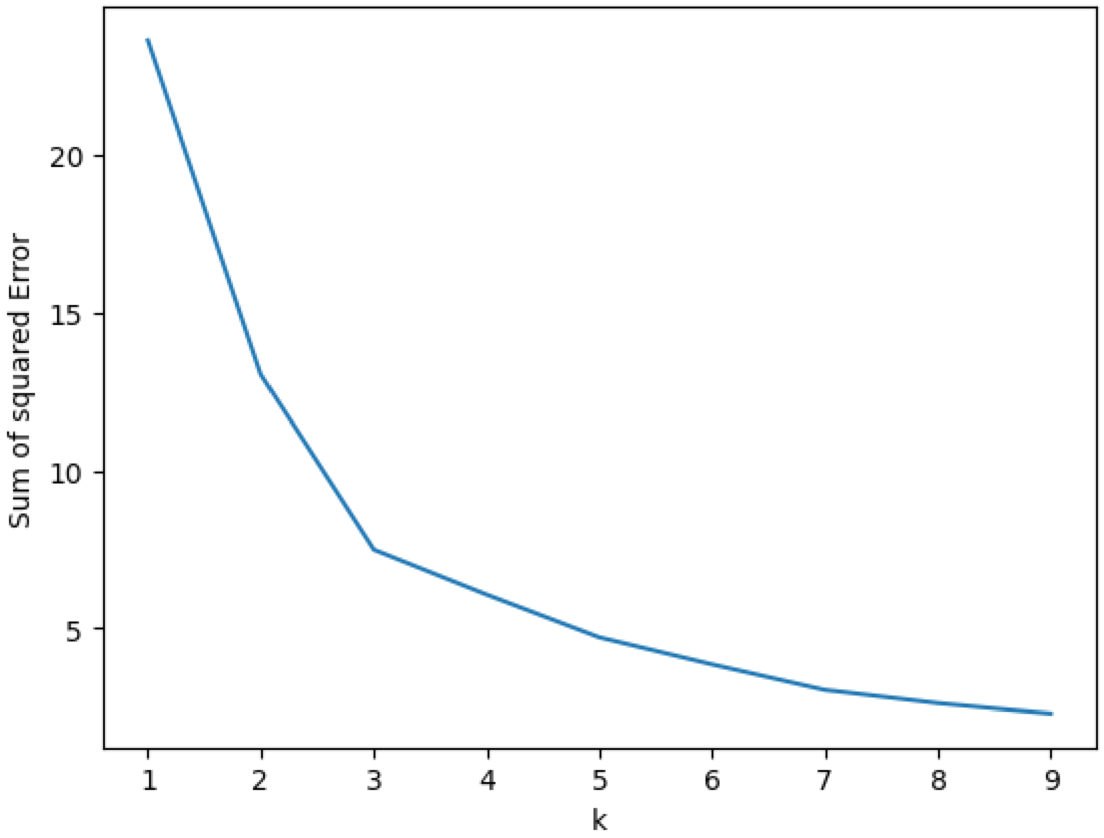
```
[23.58390615036361,
 13.028938428018284,
 7.493024843304989,
 6.072884728742553,
 4.7195043964081105,
 3.8651257592912627,
 3.058061107078988,
 2.642693946921809,
 2.3135720353543285]
```

In [29]:

```
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("Sum of squared Error")
```

Out[29]:

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



In [ ]: