

In [1]:

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

In [2]:

```
df=pd.read_csv(r"C:\Users\smb06\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]:

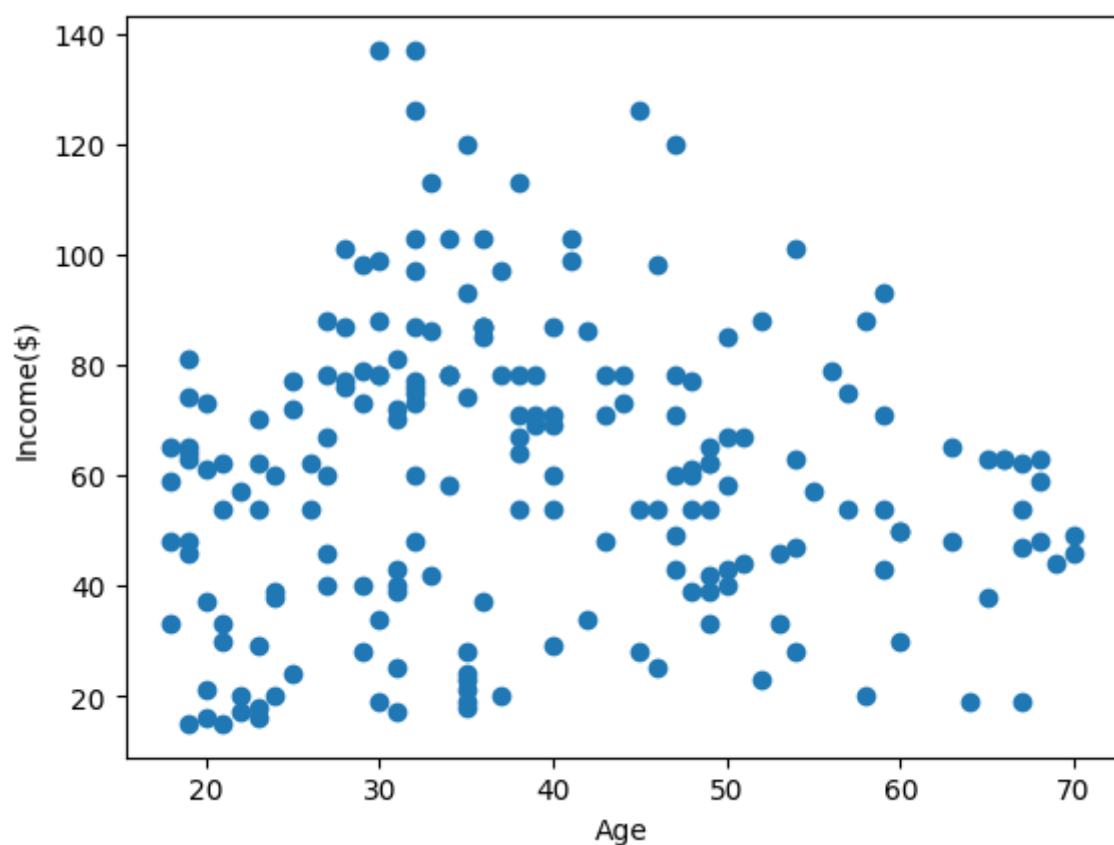
```
(200, 3)
```

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  
KMeans()
```

In [9]:

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

C:\Users\smb06\AppData\Local\Programs\Python\Python311\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([2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 1, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2,  
       1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 1, 2, 1, 2, 1, 2, 2, 2, 1, 2, 1, 2,  
       1, 2, 1, 2, 2, 2, 1, 2, 2, 1, 1, 1, 1, 7, 3, 1, 7, 3, 7, 1, 7, 3,  
       1, 7, 3, 3, 7, 1, 7, 7, 7, 3, 4, 4, 3, 4, 7, 4, 7, 4, 3, 4, 7, 3,  
       3, 4, 7, 3, 4, 4, 3, 3, 4, 3, 4, 3, 3, 4, 7, 3, 4, 3, 7, 4, 7, 7,  
       7, 3, 4, 3, 3, 3, 7, 4, 4, 4, 3, 4, 4, 4, 3, 0, 4, 4, 4, 4, 4, 4,  
       0, 0, 0, 0, 4, 0, 0, 0, 4, 0, 0, 0, 0, 0, 4, 0, 0, 0, 0, 0, 0,  
       4, 0, 0, 0, 0, 0, 4, 0, 0, 0, 5, 0, 5, 0, 0, 0, 5, 0, 0, 0, 5, 0,  
       5, 0, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6,  
       6, 6])
```

In [10]:

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

Out[10]:

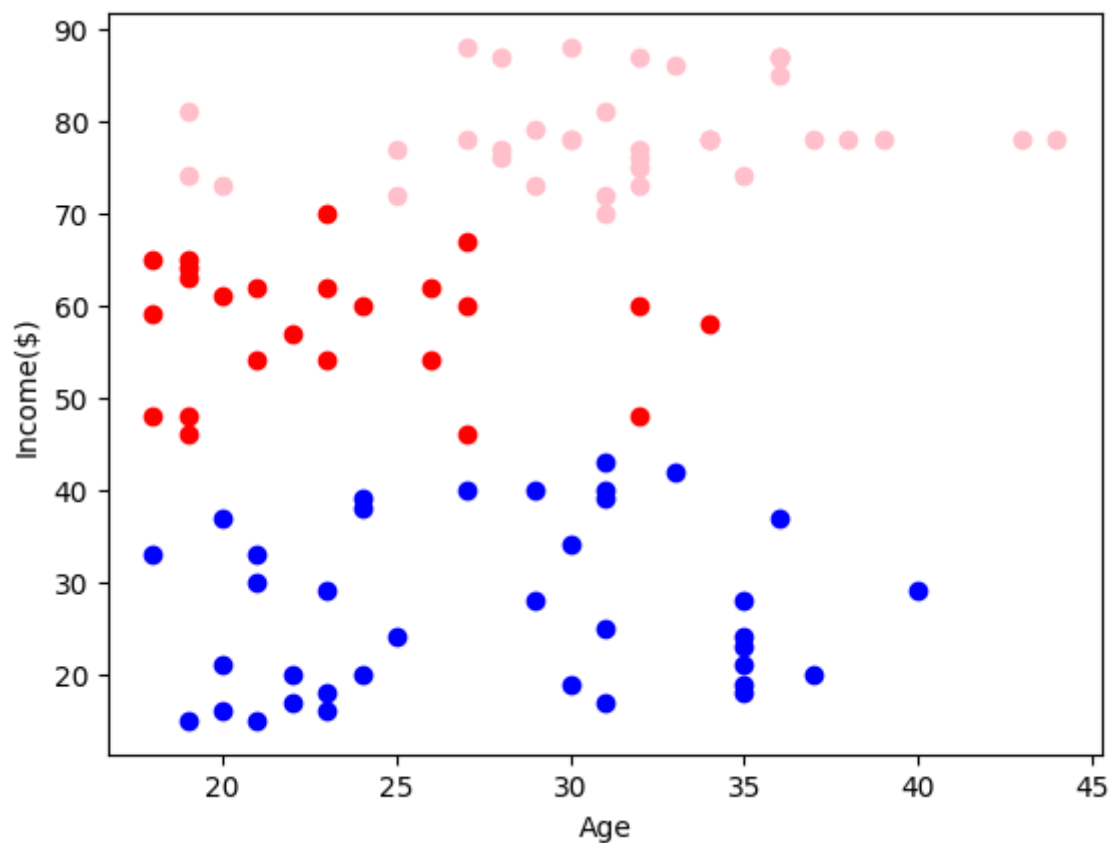
	Gender	Age	Income(\$)	cluster
0	Male	19	15	2
1	Male	21	15	2
2	Female	20	16	2
3	Female	23	16	2
4	Female	31	17	2

In [12]:

```
df1 = df[df.cluster == 0]
df2 = df[df.cluster == 2]
df3 = df[df.cluster == 3]
plt.scatter(df1["Age"],df1["Income($)"],color="pink")
plt.scatter(df2["Age"],df2["Income($)"],color="blue")
plt.scatter(df3["Age"],df3["Income($)"],color="red")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[12]:

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



In [13]:

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

Out[13]:

	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 [14]:

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

Out[14]:

	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 [15]:

```
km=KMeans()
```

In [16]:

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

C:\Users\smb06\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[16]:

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

In [17]:

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

Out[17]:

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

In [18]:

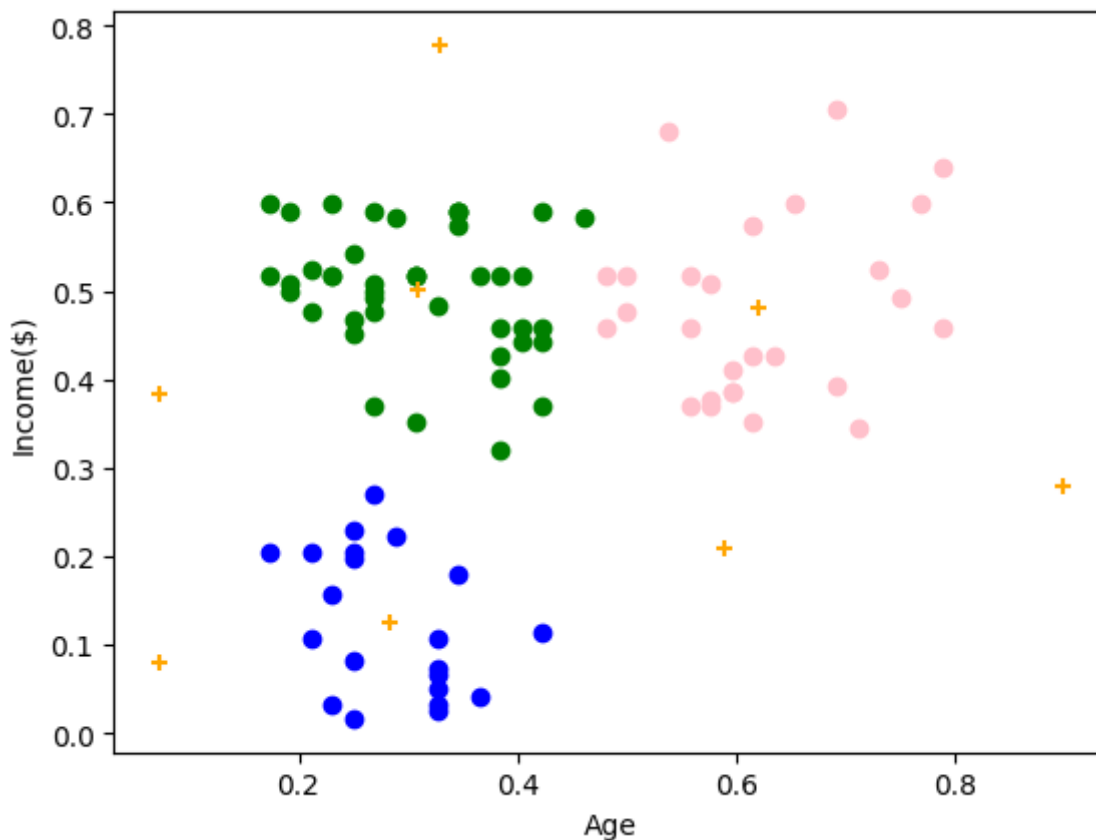
```

df1=df[df["New Cluster"]==0]
df2=df[df["New Cluster"]==1]
df3=df[df["New Cluster"]==2]
plt.scatter(df1["Age"],df1["Income($)"],color="pink")
plt.scatter(df2["Age"],df2["Income($)"],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[18]:

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



In [19]:

km.cluster_centers_

Out[19]:

```

array([[0.62037037, 0.47996357],
       [0.30903399, 0.50114373],
       [0.28388278, 0.1245121 ],
       [0.07322485, 0.38272383],
       [0.89799331, 0.28011404],
       [0.32894737, 0.77782571],
       [0.58974359, 0.20969945],
       [0.07239819, 0.08003857]])

```

In [21]:

```

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)

```

C:\Users\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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\smb06\AppData\Local\Programs\Python\Python311\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(

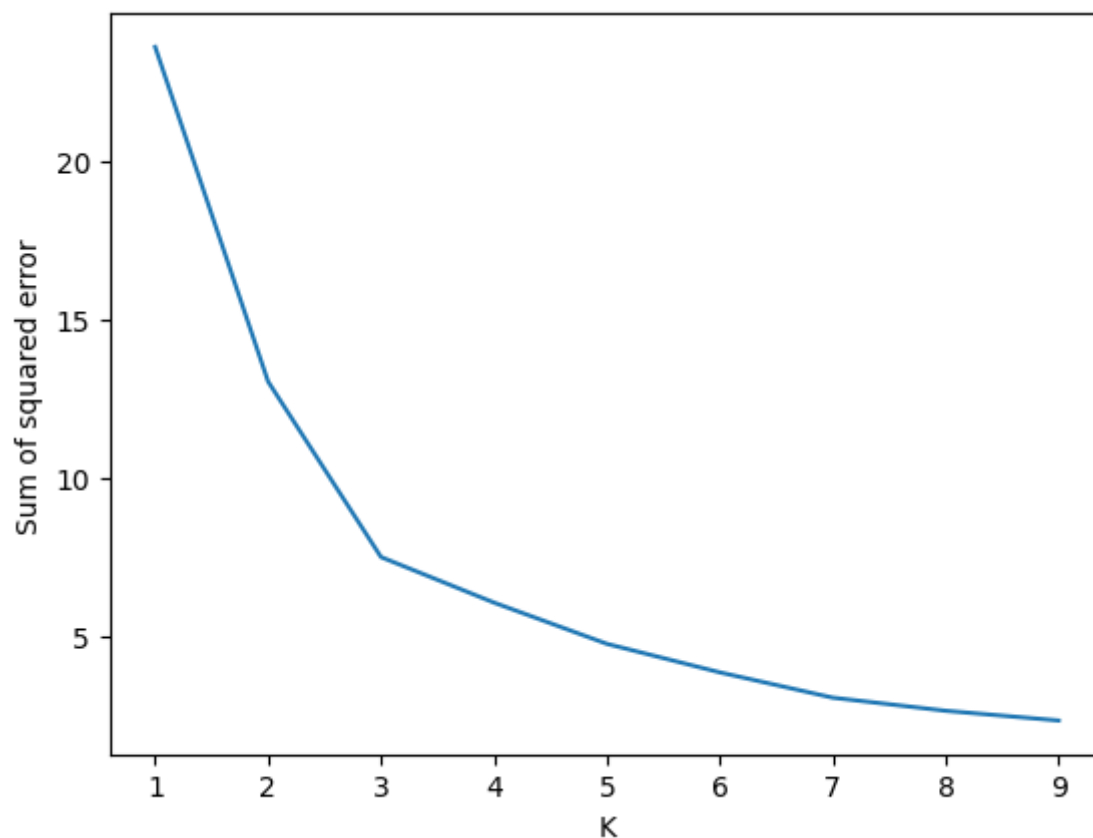
[23.583906150363603, 13.028938428018286, 7.492113413237458, 6.055824667599623, 4.75432557053689, 3.854458807699178, 3.055986211920202, 2.642520343536072, 2.335709659878675]

In [22]:

```
plt.plot(k_rng,sse)  
plt.xlabel("K")  
plt.ylabel("Sum of squared error")
```

Out[22]:

Text(0, 0.5, 'Sum of squared error')



In []: