

In [2]:

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

In [3]:

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

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 |
| ... | ... | ... | ... |
| 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 [4]:

```
df.head()
```

Out[4]:

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

In [5]:

```
df.tail()
```

Out[5]:

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

In [6]:

```
df.shape
```

Out[6]:

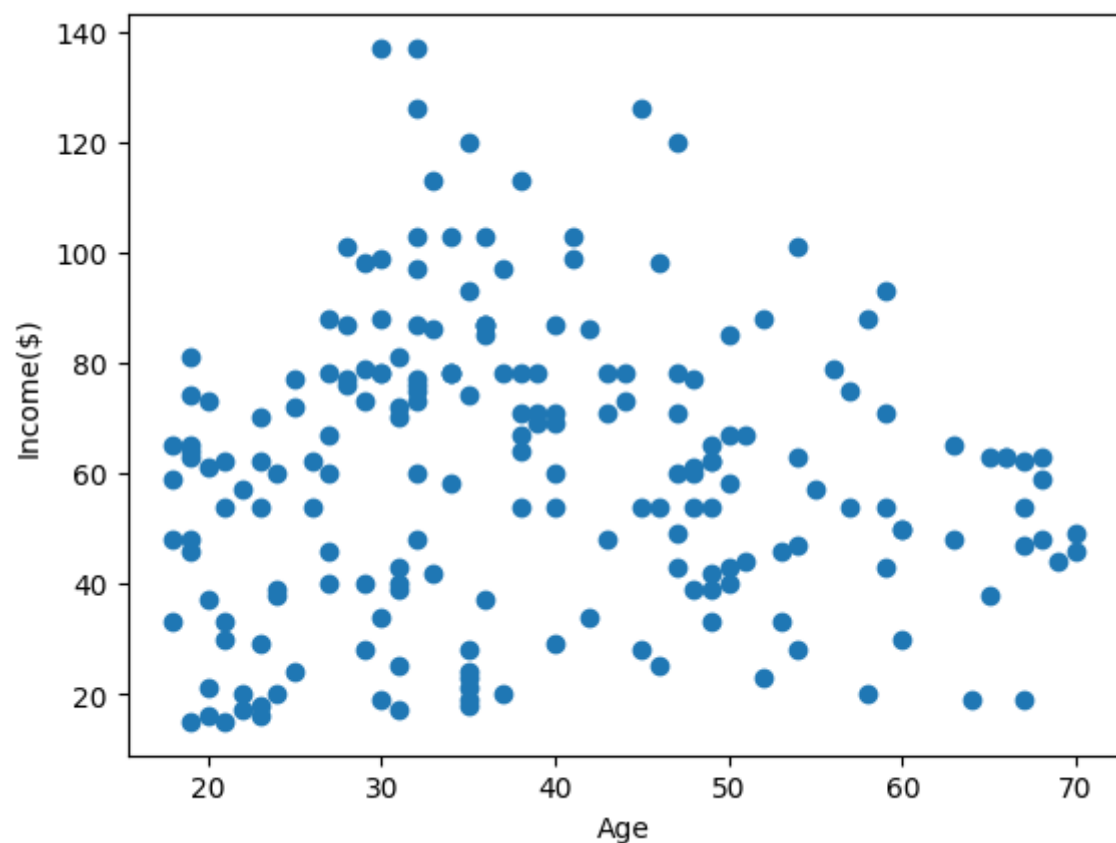
```
(200, 3)
```

In [7]:

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

Out[7]:

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



In [8]:

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

In [9]:

```
kM=KMeans()  
kM
```

Out[9]:

▼ KMeans
KMeans()

In [10]:

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

C:\Users\monim\AppData\Local\Programs\Python\Python310\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[10]:

```
array([7, 7, 7, 7, 7, 7, 7, 7, 1, 7, 1, 7, 1, 7, 7, 7, 7, 7, 1, 7, 7, 7,  
       1, 7, 1, 7, 1, 7, 1, 7, 1, 7, 1, 3, 1, 3, 1, 3, 3, 3, 1, 3, 1, 3,  
       1, 3, 1, 3, 3, 3, 1, 3, 3, 1, 1, 1, 1, 4, 3, 1, 4, 3, 4, 1, 4, 3,  
       1, 4, 3, 3, 4, 1, 4, 4, 4, 3, 6, 6, 3, 6, 4, 6, 4, 6, 3, 6, 4, 3,  
       6, 6, 4, 2, 6, 6, 2, 2, 6, 2, 6, 2, 2, 6, 4, 2, 6, 2, 4, 6, 4, 4,  
       4, 2, 6, 2, 2, 2, 4, 6, 6, 6, 2, 6, 6, 6, 2, 2, 6, 6, 6, 6, 6, 6,  
       2, 2, 2, 2, 6, 2, 2, 2, 6, 2, 2, 2, 2, 2, 6, 2, 2, 2, 6, 2, 6, 2,  
       6, 2, 2, 2, 2, 2, 6, 2, 2, 2, 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 [11]:

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

Out[11]:

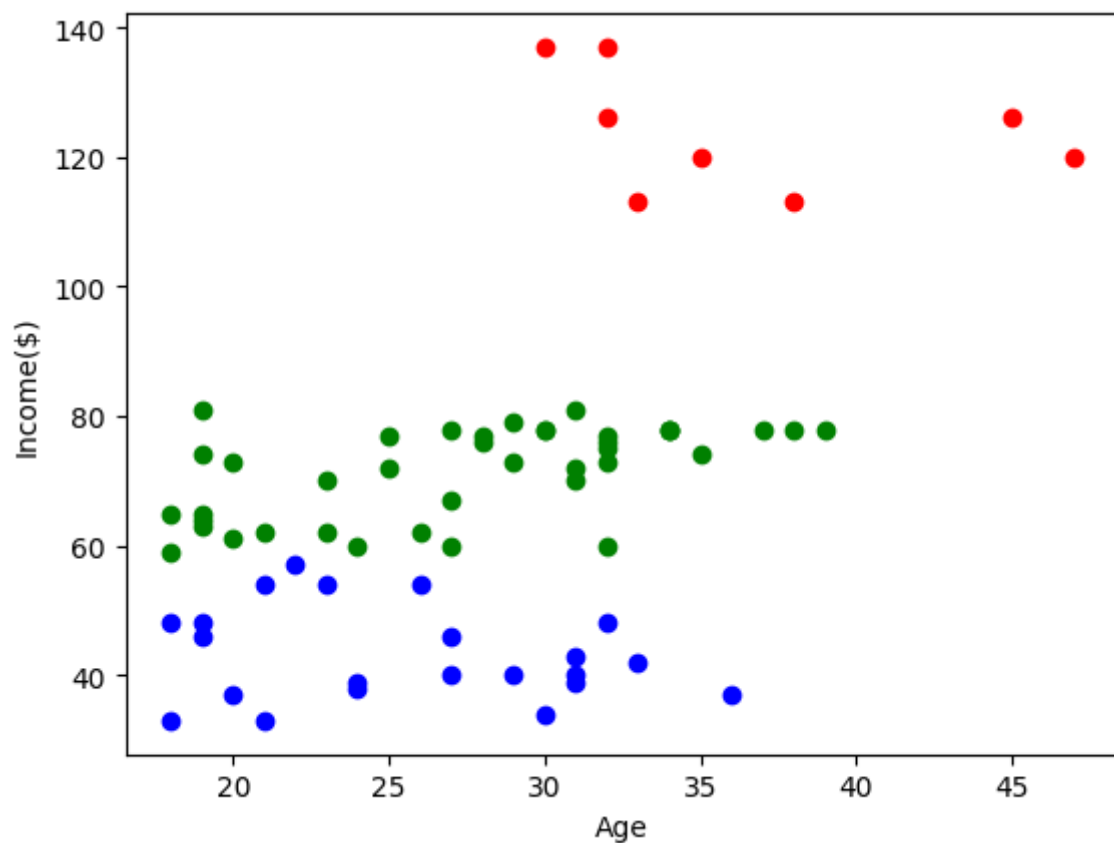
| | Gender | Age | Income(\$) | cluster |
|---|--------|-----|------------|---------|
| 0 | Male | 19 | 15 | 7 |
| 1 | Male | 21 | 15 | 7 |
| 2 | Female | 20 | 16 | 7 |
| 3 | Female | 23 | 16 | 7 |
| 4 | Female | 31 | 17 | 7 |

In [12]:

```
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[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 | 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 [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 | 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 [15]:

```
km=KMeans()
```

In [16]:

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

C:\Users\monim\AppData\Local\Programs\Python\Python310\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([6, 6, 6, 6, 0, 6, 0, 6, 2, 0, 2, 0, 7, 6, 0, 6, 0, 6, 7, 0, 0, 6,  
       7, 0, 7, 0, 7, 0, 0, 6, 2, 6, 7, 6, 7, 6, 7, 0, 0, 6, 2, 6, 7, 0,  
       7, 6, 7, 0, 0, 0, 7, 0, 0, 2, 7, 7, 7, 2, 3, 7, 2, 3, 2, 7, 2, 3,  
       7, 2, 3, 0, 2, 7, 2, 2, 2, 3, 7, 7, 3, 7, 2, 1, 2, 7, 3, 7, 4, 3,  
       1, 4, 2, 3, 4, 1, 1, 3, 4, 3, 4, 3, 3, 4, 2, 3, 4, 3, 2, 4, 2, 2,  
       2, 3, 1, 3, 3, 3, 2, 4, 4, 4, 3, 1, 1, 1, 3, 1, 4, 1, 4, 1, 4, 1,  
       3, 1, 3, 1, 4, 1, 3, 1, 4, 1, 1, 1, 3, 1, 4, 1, 1, 1, 4, 1, 4, 1,  
       4, 1, 1, 1, 1, 1, 4, 1, 3, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 4, 1,  
       4, 1, 4, 1, 5, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5,  
       5, 5])
```

In [17]:

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

Out[17]:

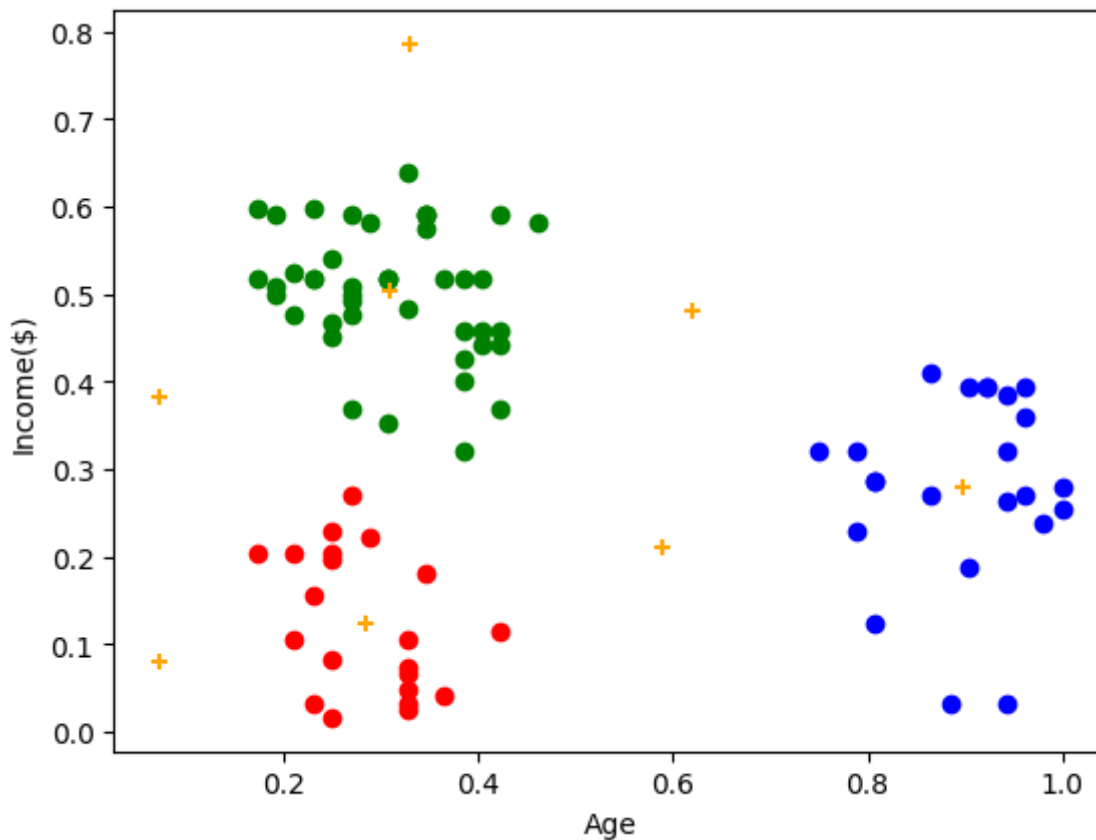
| | Gender | Age | Income(\$) | cluster | New Cluster |
|---|--------|----------|------------|---------|-------------|
| 0 | Male | 0.019231 | 0.000000 | 7 | 6 |
| 1 | Male | 0.057692 | 0.000000 | 7 | 6 |
| 2 | Female | 0.038462 | 0.008197 | 7 | 6 |
| 3 | Female | 0.096154 | 0.008197 | 7 | 6 |
| 4 |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

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($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],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[18]:

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



In [19]:

km.cluster_centers_

Out[19]:

```
array([[0.28388278, 0.1245121 ],
       [0.30944056, 0.50428465],
       [0.89799331, 0.28011404],
       [0.07322485, 0.38272383],
       [0.62037037, 0.47996357],
       [0.32905983, 0.78551913],
       [0.07239819, 0.08003857],
       [0.58974359, 0.20969945]])
```


In [24]:

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

In [25]:

```

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")

```

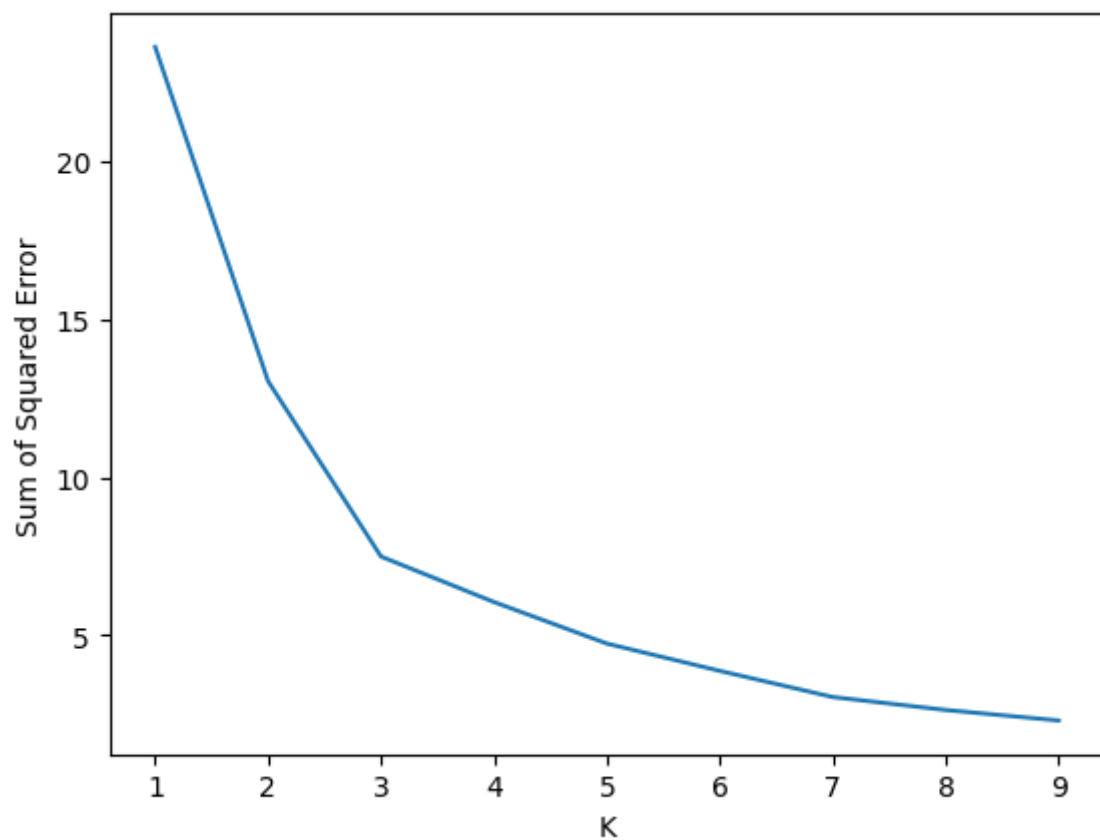
```

C:\Users\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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\monim\AppData\Local\Programs\Python\Python310\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[25]:

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



In []: