

In [19]:

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

In [20]:

```
df=pd.read_csv(r"C:\Users\shaha\OneDrive\Desktop\Excel\Income.csv")
df
```

Out[20]:

	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 [21]:

```
df.head()
```

Out[21]:

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

In [22]:

```
df.tail()
```

Out[22]:

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

In [23]:

```
df.shape
```

Out[23]:

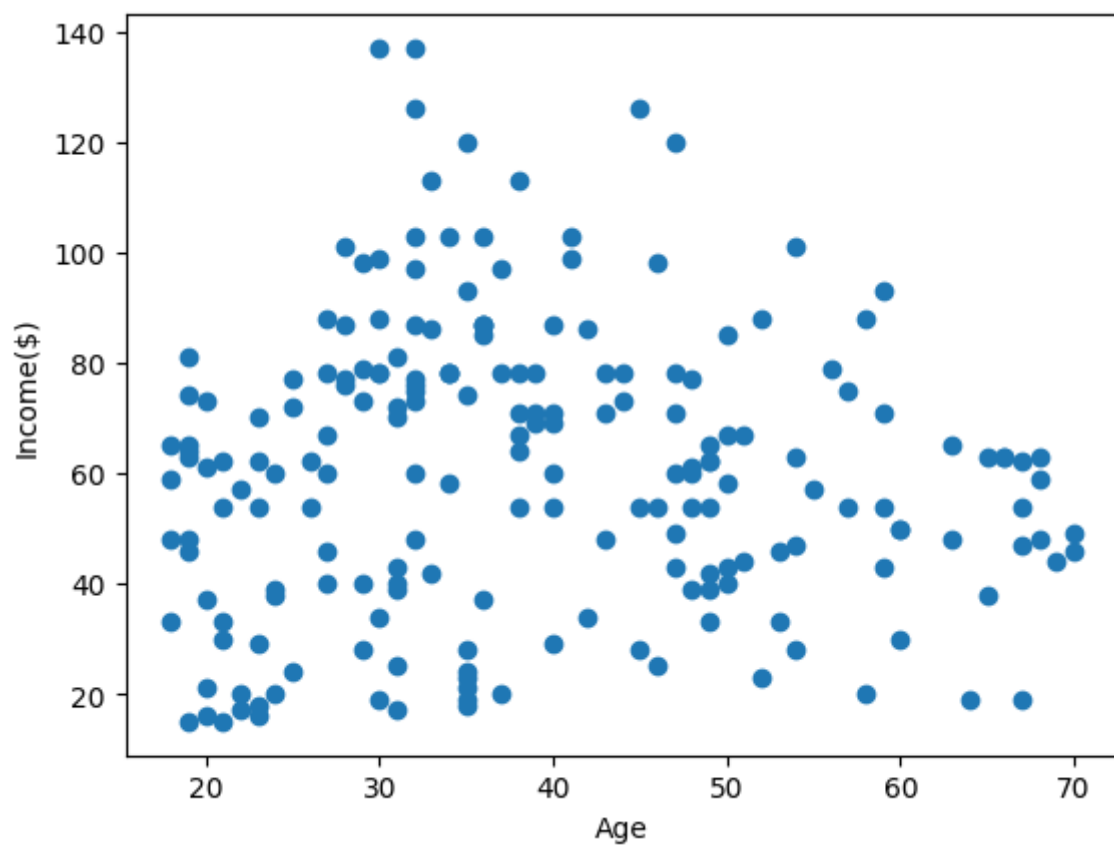
(200, 3)

In [24]:

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

Out[24]:

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



In [31]:

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

In [33]:

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

Out[33]:

▼ KMeans

KMeans()

In [34]:

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

C:\Users\shaha\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[34]:

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

In [37]:

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

Out[37]:

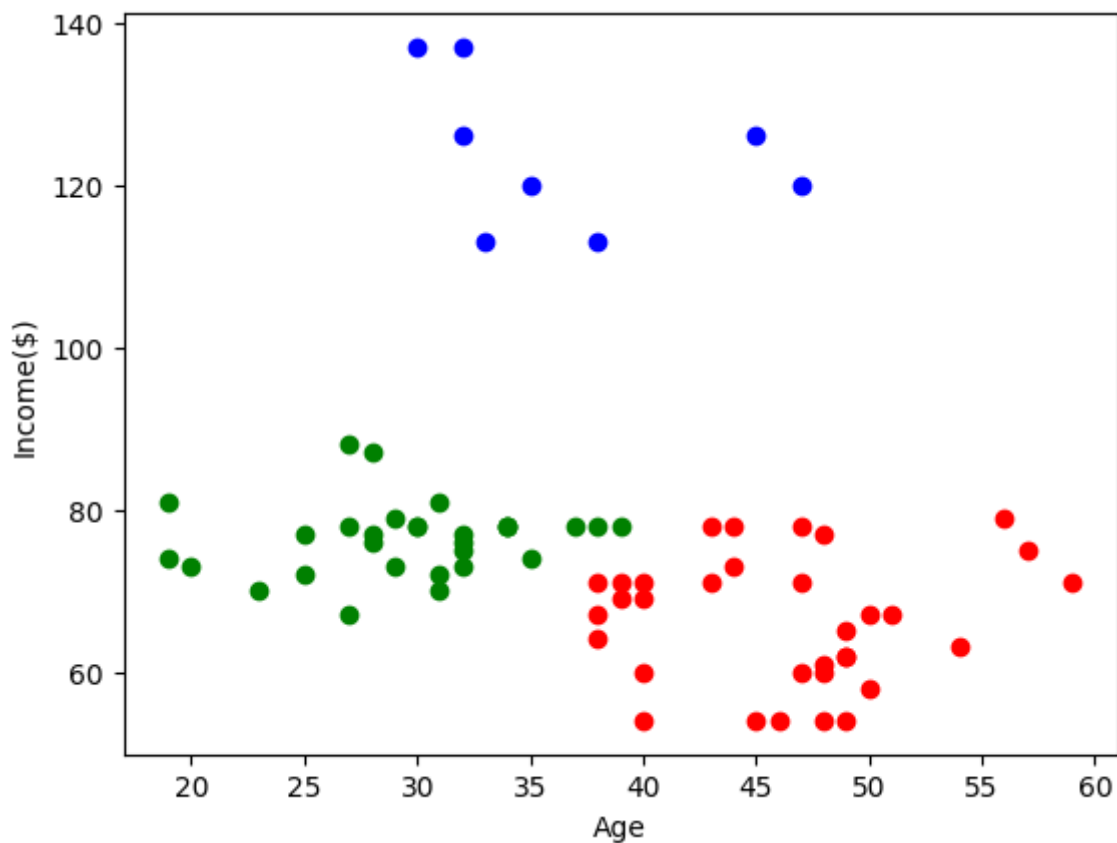
	Gender	Age	Income(\$)	cluster
0	Male	19	15	5
1	Male	21	15	5
2	Female	20	16	5
3	Female	23	16	5
4	Female	31	17	5

In [38]:

```
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[38]:

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



In [40]:

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

Out[40]:

	Gender	Age	Income(\$)	cluster
0	Male	19	0.000000	5
1	Male	21	0.000000	5
2	Female	20	0.008197	5
3	Female	23	0.008197	5
4	Female	31	0.016393	5

In [41]:

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

Out[41]:

	Gender	Age	Income(\$)	cluster
0	Male	0.019231	0.000000	5
1	Male	0.057692	0.000000	5
2	Female	0.038462	0.008197	5
3	Female	0.096154	0.008197	5
4	Female	0.250000	0.016393	5

In [42]:

```
km=KMeans()
```

In [43]:

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

C:\Users\shaha\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[43]:

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

In [47]:

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

Out[47]:

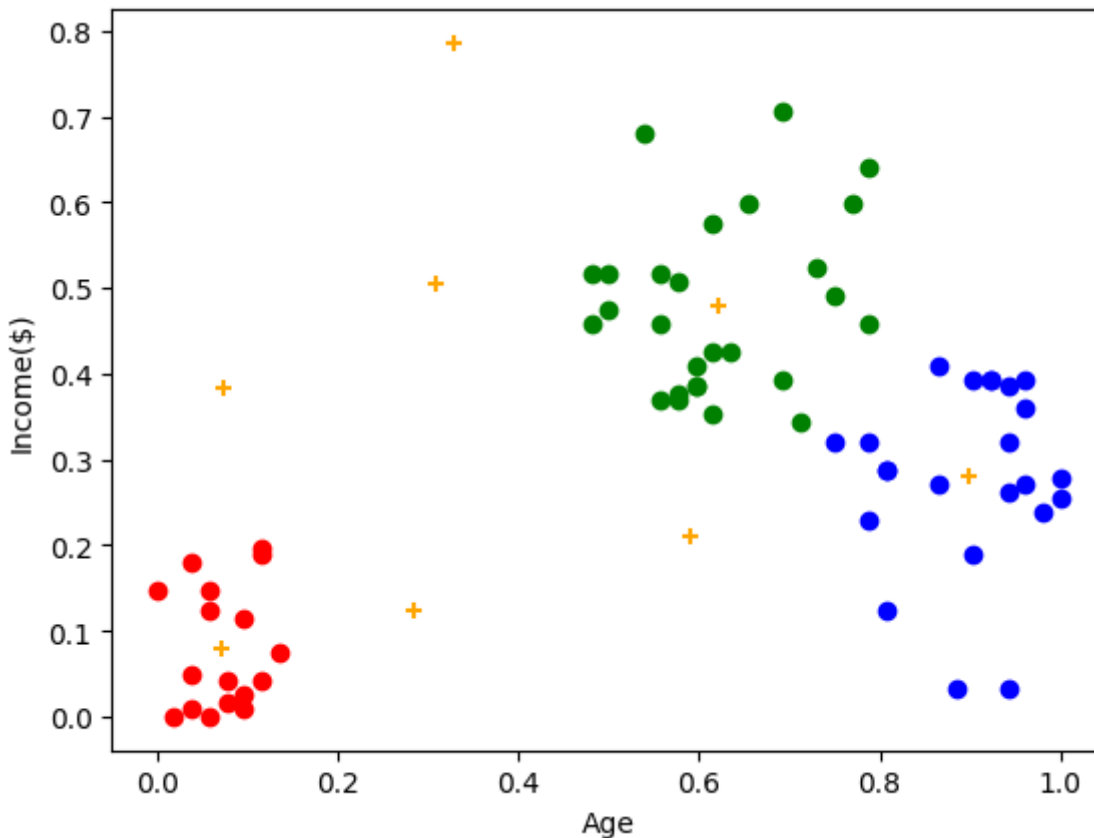
	Gender	Age	Income(\$)	cluster	New cluster	New Cluster
0	Male	0.019231	0.000000	5	0	0
1	Male	0.057692	0.000000	5	0	0
2	Female	0.038462	0.008197	5	0	0
3	Female	0.096154	0.008197	5	0	0
4	Female	0.250000	0.016393	5	4	4

In [48]:

```
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[48]:

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



In [51]:

```
km.cluster_centers_
```

Out[51]:

```
array([[0.07076923, 0.37770492],
       [0.40646853, 0.45417288],
       [0.89799331, 0.28011404],
       [0.64772727, 0.46870343],
       [0.25300481, 0.55430328],
       [0.28388278, 0.1245121 ],
       [0.37051282, 0.80491803],
       [0.59698997, 0.20491803],
       [0.07239819, 0.08003857]])
```

In [52]:

```

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\shaha\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\shaha\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\shaha\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\shaha\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\shaha\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\shaha\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\shaha\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\shaha\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(

```

[23.583906150363603, 13.02893842801829, 7.492107868586013, 6.0583724533531
55, 4.734128359733764, 3.871058219814432, 3.054717436369359, 2.65318124013
31285, 2.335983809814013]

```

C:\Users\shaha\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

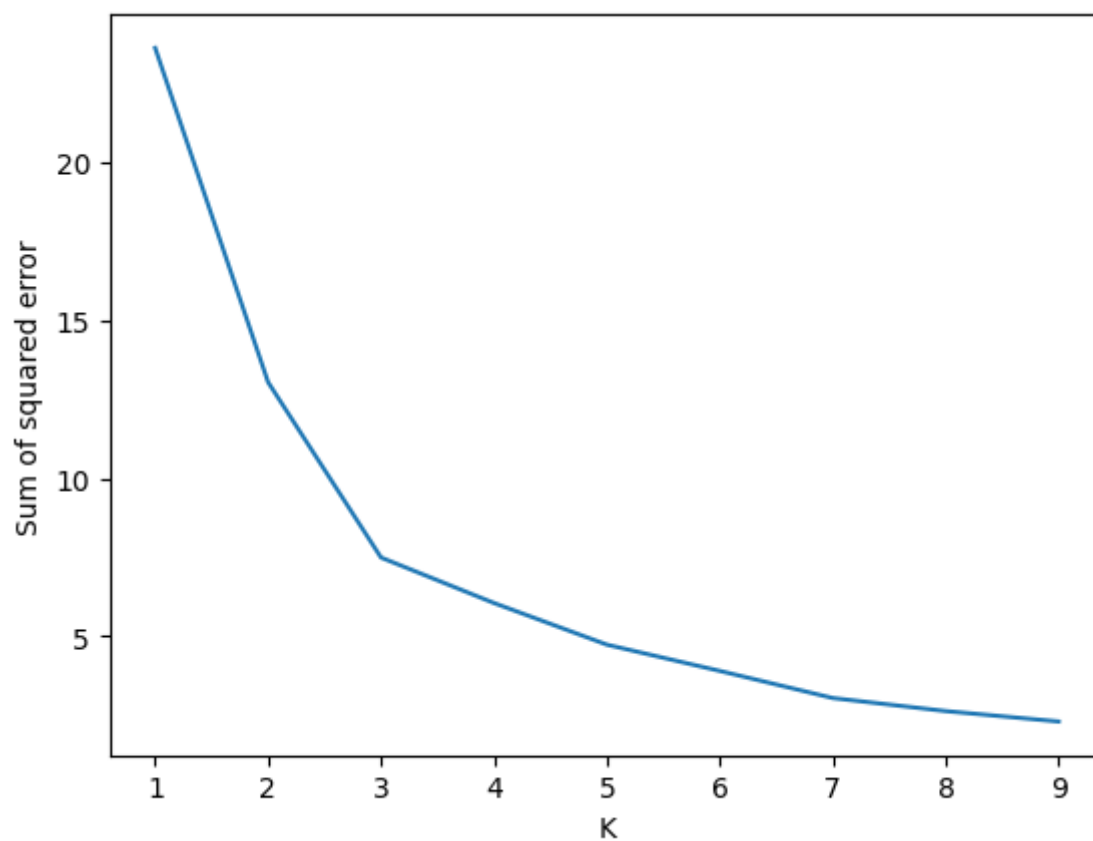
warnings.warn(

In [50]:

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

Out[50]:

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



In []: