In [1]:

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

In [2]:

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

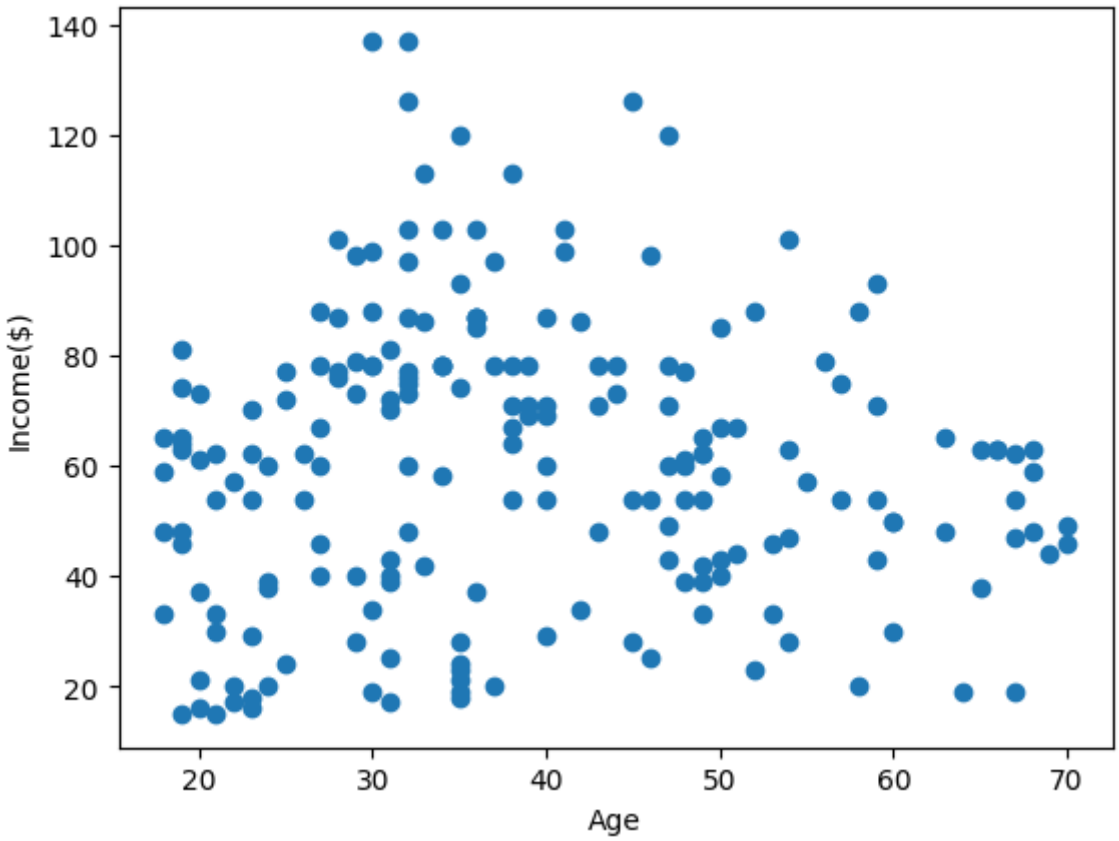
	Gender	Age	Income($)
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

In [5]:

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

Out[5]:

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



In [6]:

```
from sklearn.cluster import KMeans
km=KMeans()
km
```

Out[6]:

KMeans

KMeans()

In [7]:

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

```
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
```

Out[7]:

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

In [8]:

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

Out[8]:

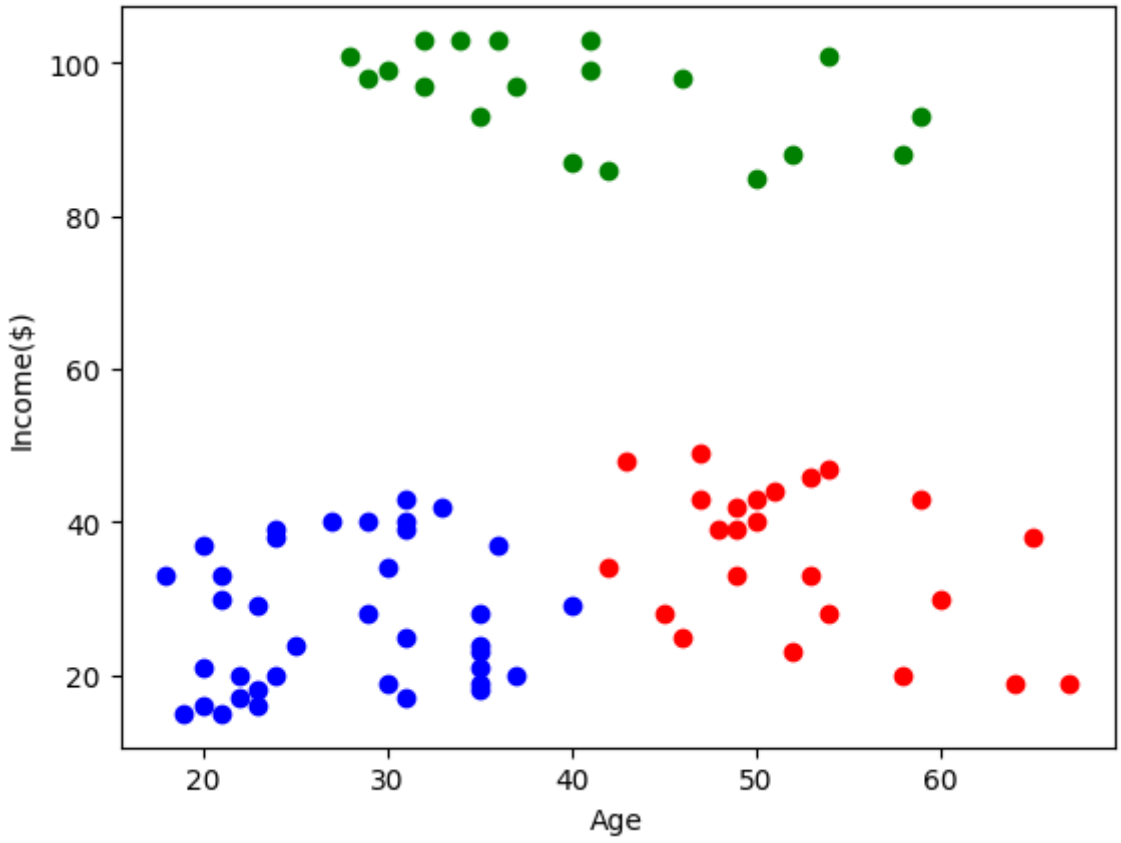
	Gender	Age	Income($)	Cluster
0	Male	19	15	2
1	Male	21	15	2
2	Female	20	16	2
3	Female	23	16	2
4	Female	31	17	2

In [10]:

```
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[10]:

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



In [12]:

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

In [13]:

```
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 [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	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 [16]:

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

```
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
```

Out[16]:

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

In [17]:

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

Out[17]:

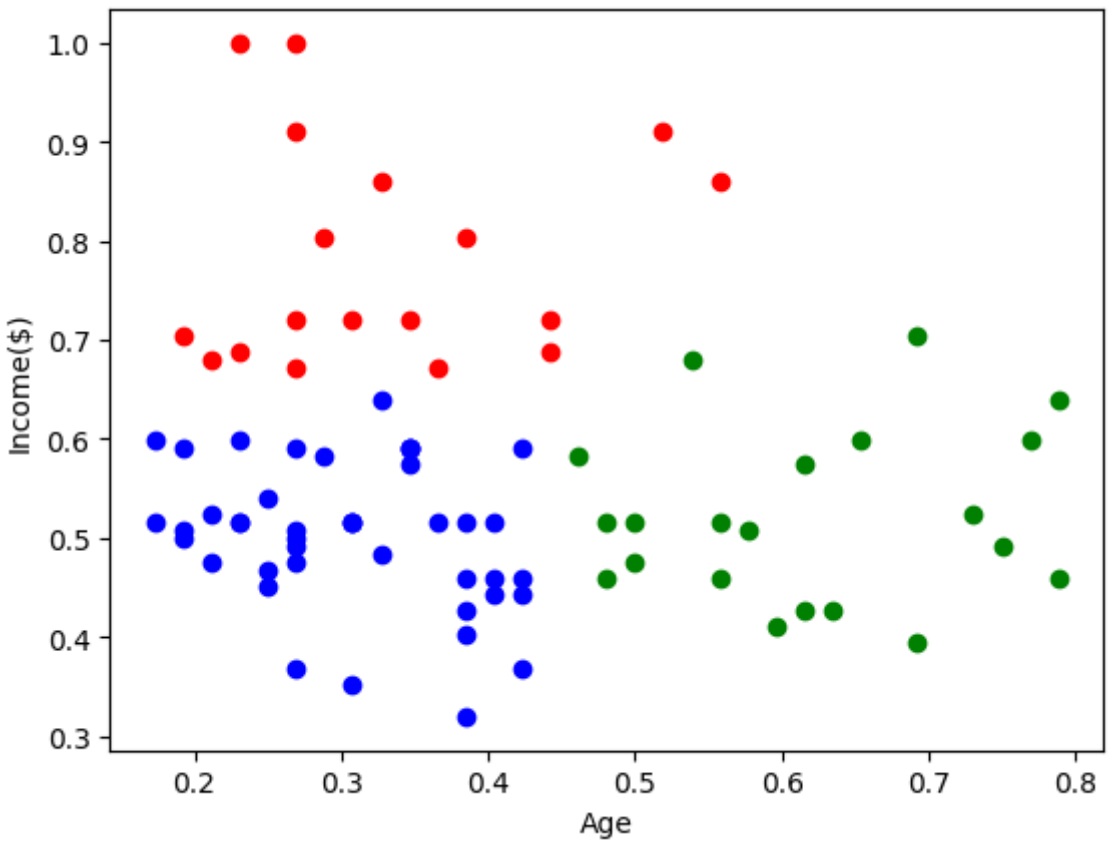
	Gender	Age	Income($)	Cluster	New Cluster
0	Male	0.019231	0.000000	2	4
1	Male	0.057692	0.000000	2	4
2	Female	0.038462	0.008197	2	4
3	Female	0.096154	0.008197	2	4
4	Female	0.250000	0.016393	2	7

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($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[18]:

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



In [19]:

```
km.cluster_centers_
```

Out[19]:

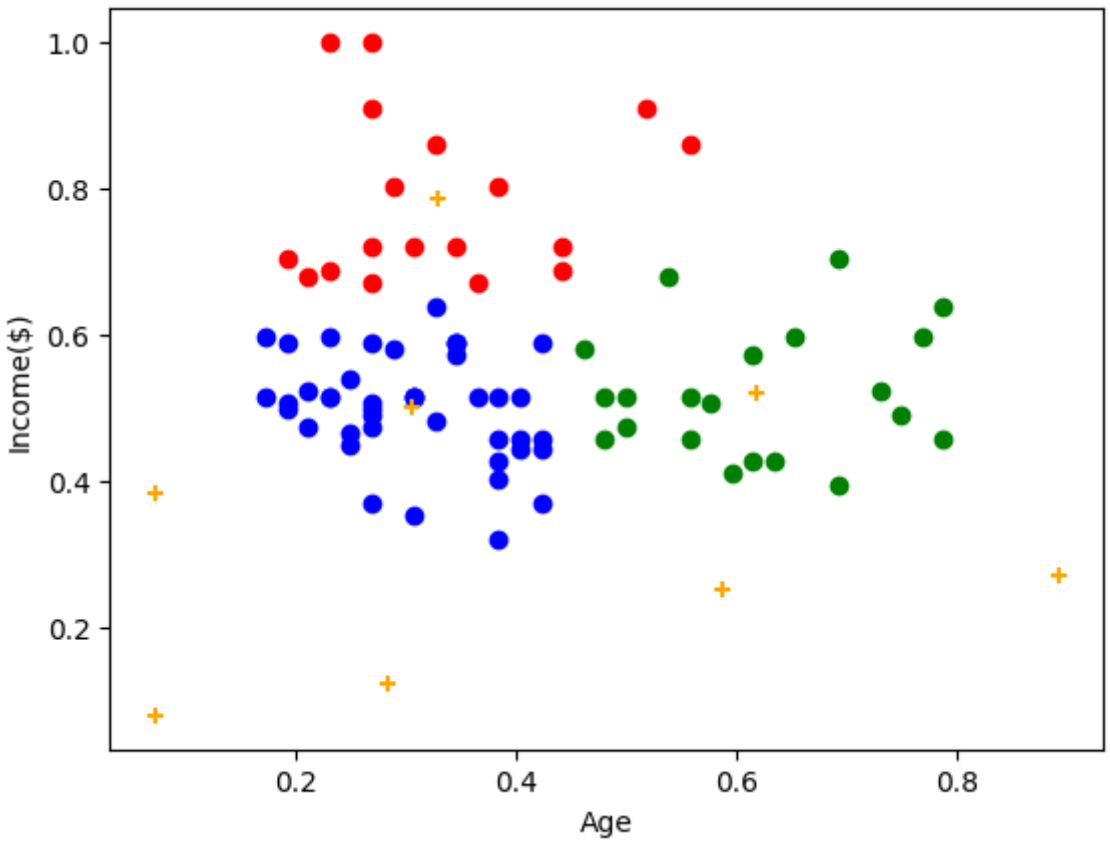
```
array([[0.32905983, 0.78551913],
       [0.61813187, 0.52185792],
       [0.3059034 , 0.50247808],
       [0.89262821, 0.27015027],
       [0.07239819, 0.08003857],
       [0.07322485, 0.38272383],
       [0.58717949, 0.25245902],
       [0.28388278, 0.1245121 ]])
```

In [20]:

```
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[20]:

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



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)
```

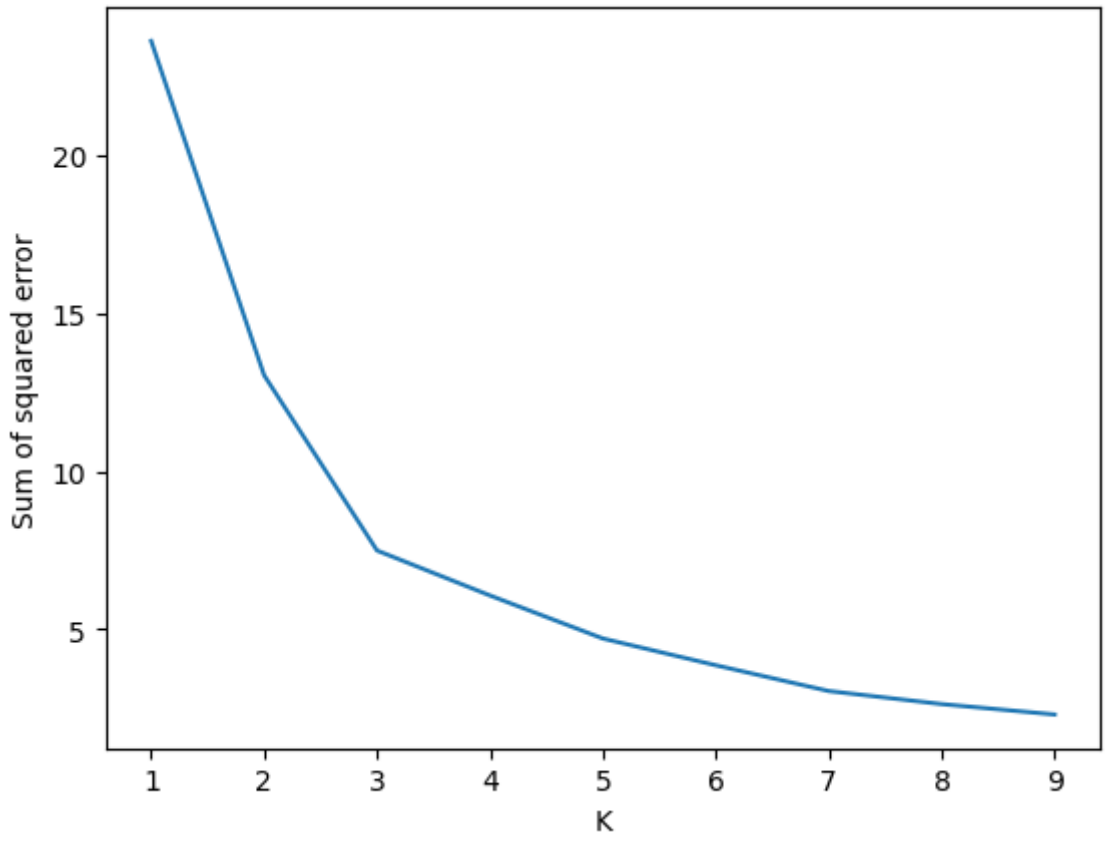
```
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(

[23.583906150363603, 13.028938428018286, 7.492107868586012, 6.072884728742
5545, 4.713025598595381, 3.868901807239719, 3.054717436369358, 2.642520343
536072, 2.3135720353543285]

C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
C:\Users\lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_in
it` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explic
itly to suppress the warning
  warnings.warn(
```

In [23]:

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

Out[23]:

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



In [ ]: