```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
        %matplotlib inline
```

```
In [2]: df=pd.read_csv(r"C:\Users\pappu\Downloads\Income.csv")
        df
```
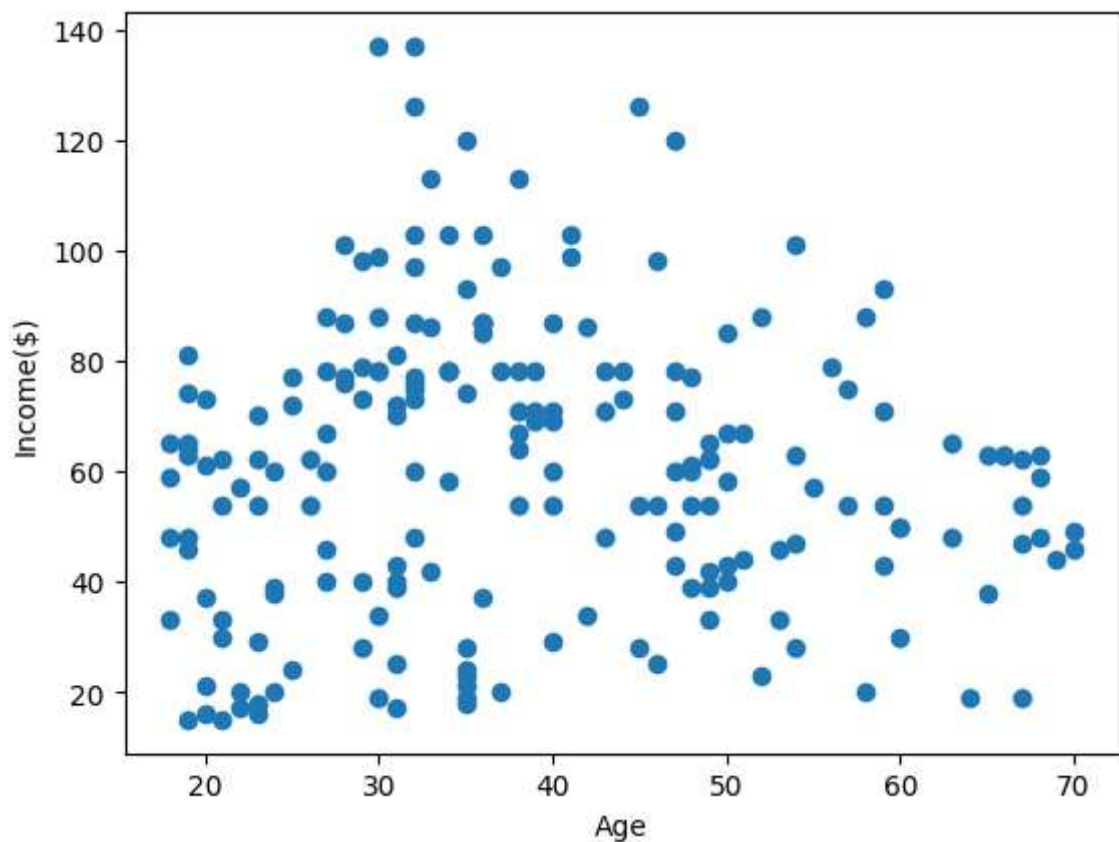
Out[2]:

	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 [4]:
```python
plt.scatter(df['Age'],df['Income($)'])
plt.xlabel('Age')
plt.ylabel('Income($)')
```

Out[4]: Text(0, 0.5, 'Income($)')

In [6]:
```python
from sklearn.cluster import KMeans
km=KMeans()
```

In [7]:
```python
y_pred=km.fit_predict(df[['Age','Income($)']])
y_pred
```

C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(

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

In [8]:
```python
df['cluster']=y_pred
df.head()
```

Out[8]:

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

In [9]:
```python
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='red')
plt.scatter(df3['Age'],df3['Income($)'],color='red')
plt.xlabel('Age')
plt.ylabel('Income($)')
```

Out[9]: Text(0, 0.5, 'Income($)')
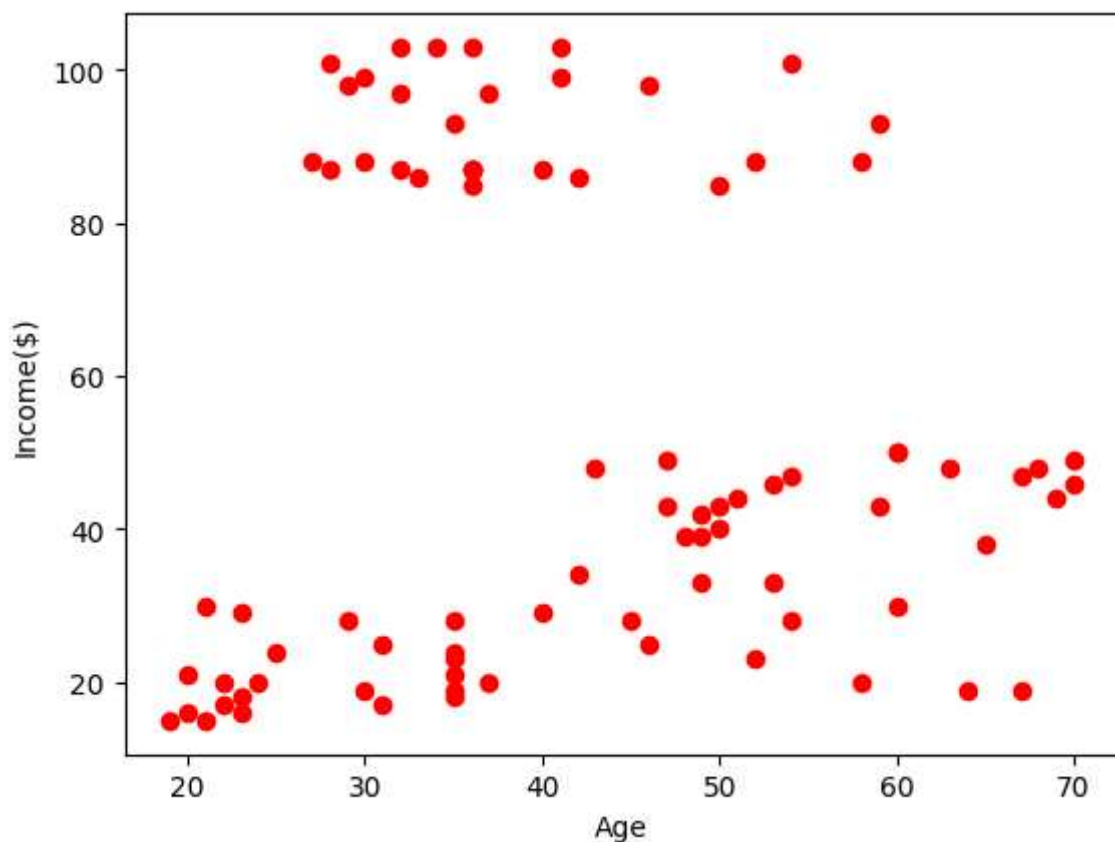
In [10]:
```python
from sklearn.preprocessing import MinMaxScaler
```

In [11]:
```python
scaler=MinMaxScaler()
```

In [12]:
```python
scaler.fit(df[["Income($)"]])
df['Income($)']=scaler.transform(df[['Income($)']])
df.head()
```

Out[12]:

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

In [13]:
```python
scaler.fit(df[['Age']])
df['Age']=scaler.transform(df[['Age']])
df.head()
```

Out[13]:

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

In [14]:
```python
km=KMeans()
```

In [15]:
```python
y_predicted=km.fit_predict(df[['Age','Income($)']])
y_predicted
```

```
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
```

Out[15]:
```
array([4, 4, 4, 4, 4, 4, 4, 4, 0, 4, 0, 4, 2, 4, 4, 4, 4, 4, 2, 4, 4, 4,
       2, 4, 2, 4, 2, 4, 2, 4, 0, 4, 2, 4, 2, 4, 2, 4, 4, 4, 0, 4, 2, 4,
       2, 4, 2, 4, 4, 4, 2, 4, 4, 0, 2, 2, 2, 0, 4, 2, 0, 1, 0, 2, 0, 1,
       2, 0, 1, 4, 0, 2, 0, 0, 0, 1, 2, 3, 1, 2, 0, 3, 0, 2, 1, 2, 5, 1,
       3, 5, 0, 1, 5, 3, 3, 1, 5, 1, 5, 1, 1, 5, 0, 1, 5, 1, 0, 5, 0, 0,
       0, 1, 3, 1, 1, 1, 0, 5, 5, 5, 1, 3, 3, 3, 1, 7, 3, 3, 5, 3, 5, 3,
       1, 7, 1, 7, 3, 7, 1, 3, 5, 7, 7, 7, 7, 7, 5, 7, 7, 7, 3, 3, 3, 3,
       5, 7, 3, 7, 7, 7, 5, 7, 1, 7, 5, 7, 3, 7, 7, 7, 3, 7, 7, 7, 5, 7,
       5, 7, 5, 7, 6, 7, 6, 7, 6, 7, 5, 7, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
       6, 6])
```

In [18]:
```python
df['New cluster']=y_predicted
df.head()
```

Out[18]:

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

In [19]:
```python
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($)'],color='blue')
plt.xlabel('Age')
plt.ylabel('Income($)')
```
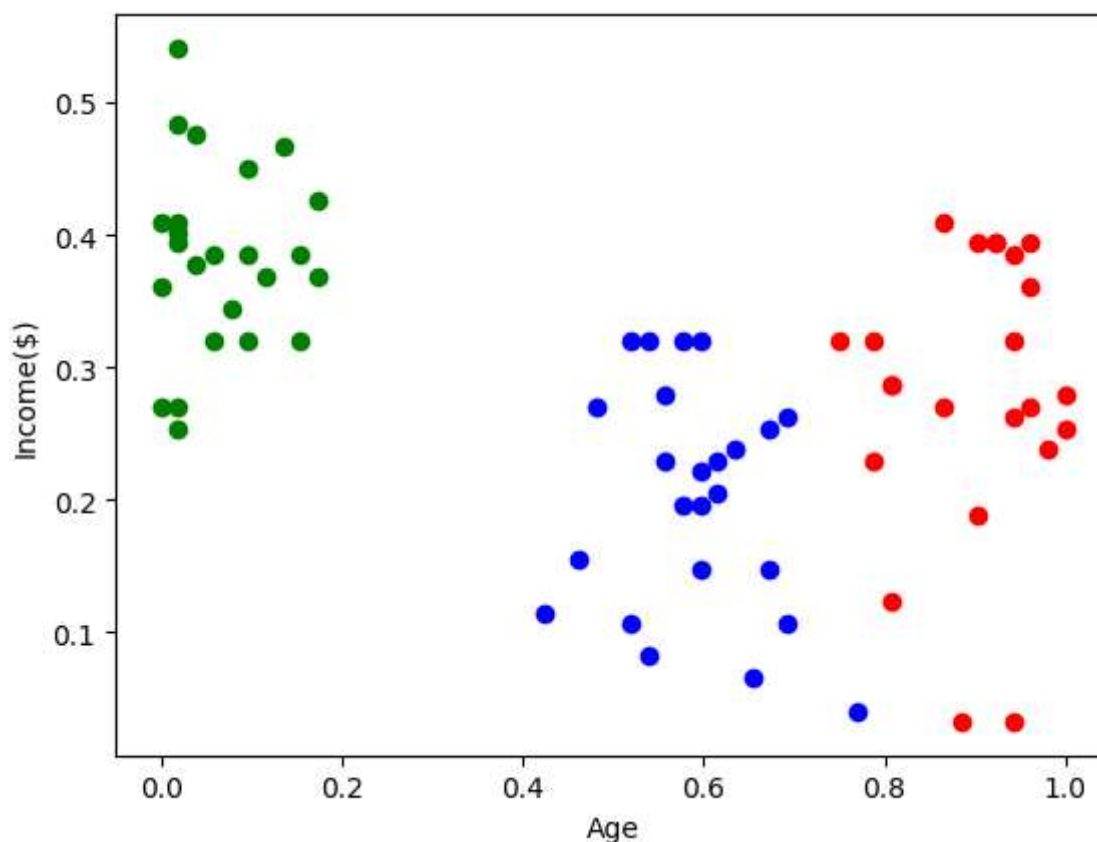
Out[19]: Text(0, 0.5, 'Income($)')
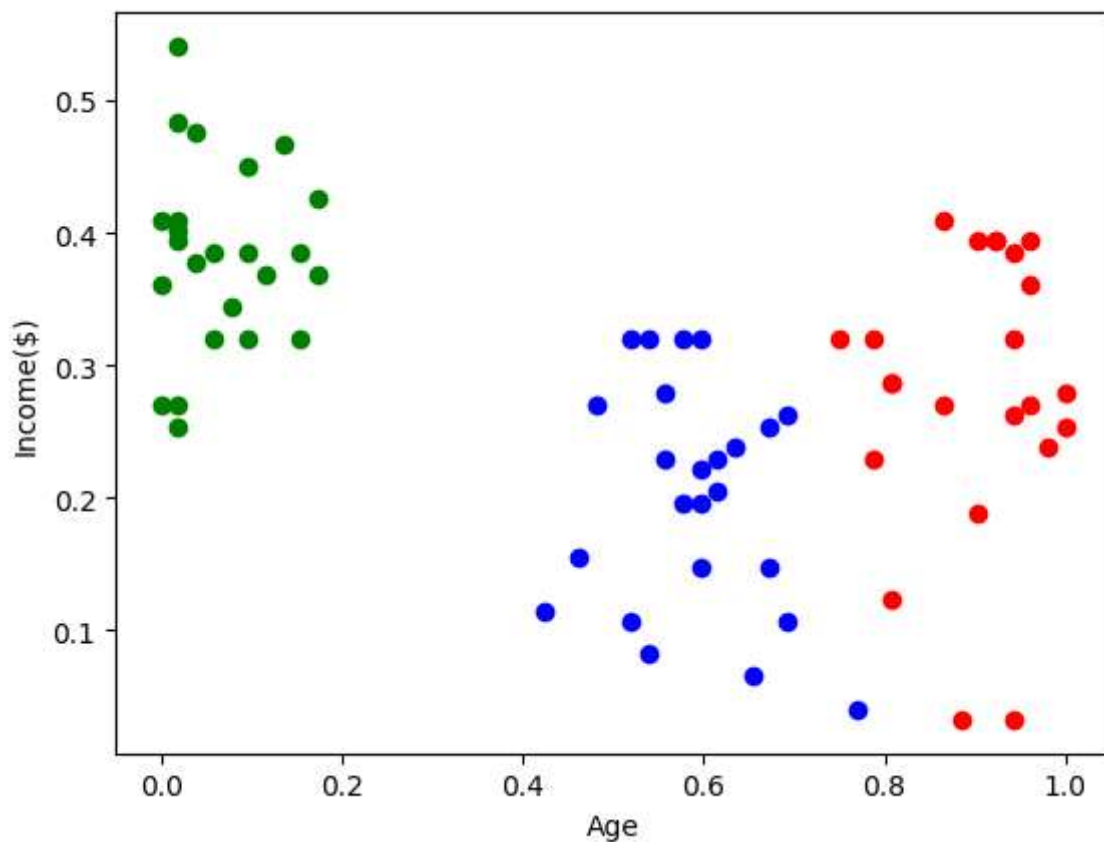
In [21]: `km.cluster_centers_`

Out[21]:
```
array([[0.89799331, 0.28011404],
       [0.06650641, 0.38285519],
       [0.58974359, 0.2011612 ],
       [0.40646853, 0.45417288],
       [0.18269231, 0.108283  ],
       [0.64772727, 0.46870343],
       [0.37051282, 0.80491803],
       [0.25300481, 0.55430328]])
```

In [22]:
```python
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[22]: `Text(0, 0.5, 'Income($)')`

In [23]:
```python
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color='orange')
plt.xlabel('Age')
plt.ylabel('Income($)')
```

Out[23]: Text(0, 0.5, 'Income($)')
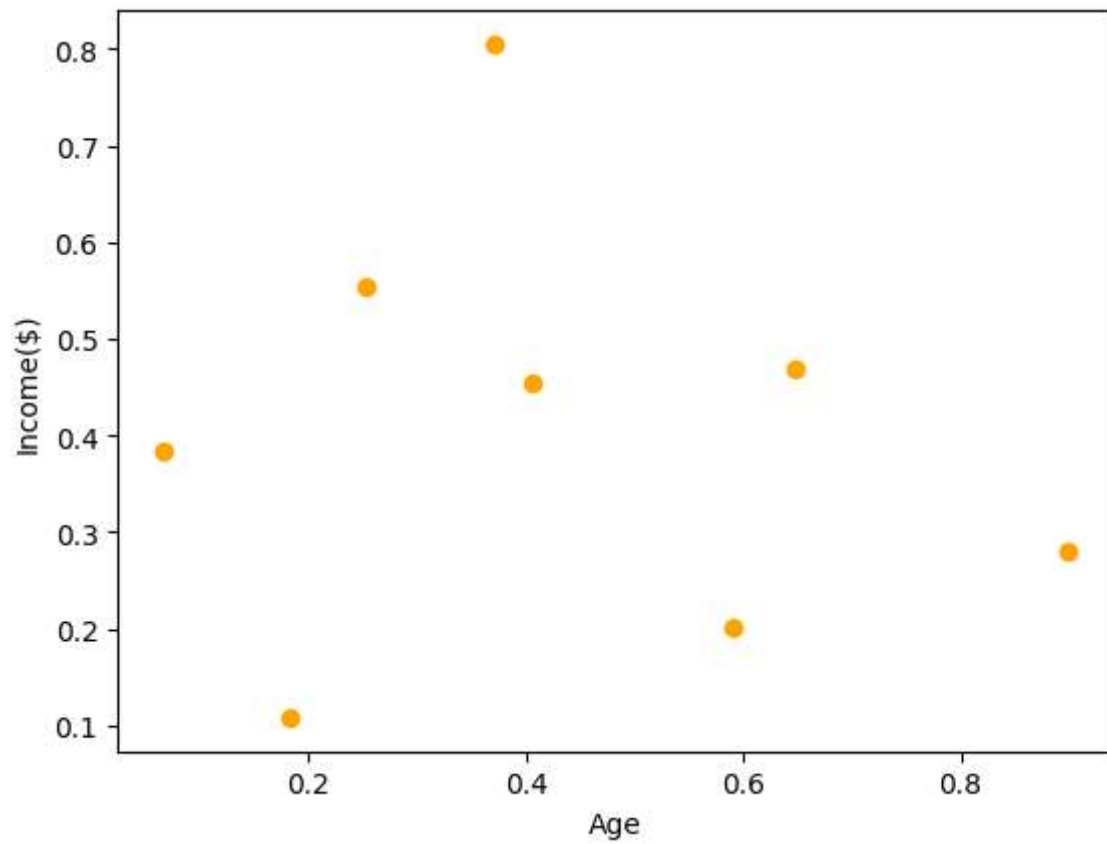
In [24]:
```python
k_rng=range(1,10)
sse=[]
for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[['Age','Income($)']])
    sse.append(km.inertia_)
sse
```

C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(
C:\Users\pappu\AppData\Local\Programs\Python\Python310\lib\site-packages\skle
arn\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 supp
ress the warning
  warnings.warn(

Out[24]: [23.583906150363603,
         13.02893842801829,
         7.49302484330499,
         6.0728847287425545,
         4.713811834695168,
         3.868236374381552,
         3.0559862119202013,
         2.642520043536072,
         2.334810712216323]

In [25]:
```python
plt.plot(k_rng,sse)
plt.xlabel('k')
plt.ylabel("Sum of squared error")
```

Out[25]: Text(0, 0.5, 'Sum of squared error')