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

In [2]:
```
df=pd.read_csv(r"C:\Users\jangidi veena\OneDrive\Documents\jupyter\Income.csv")
df
```

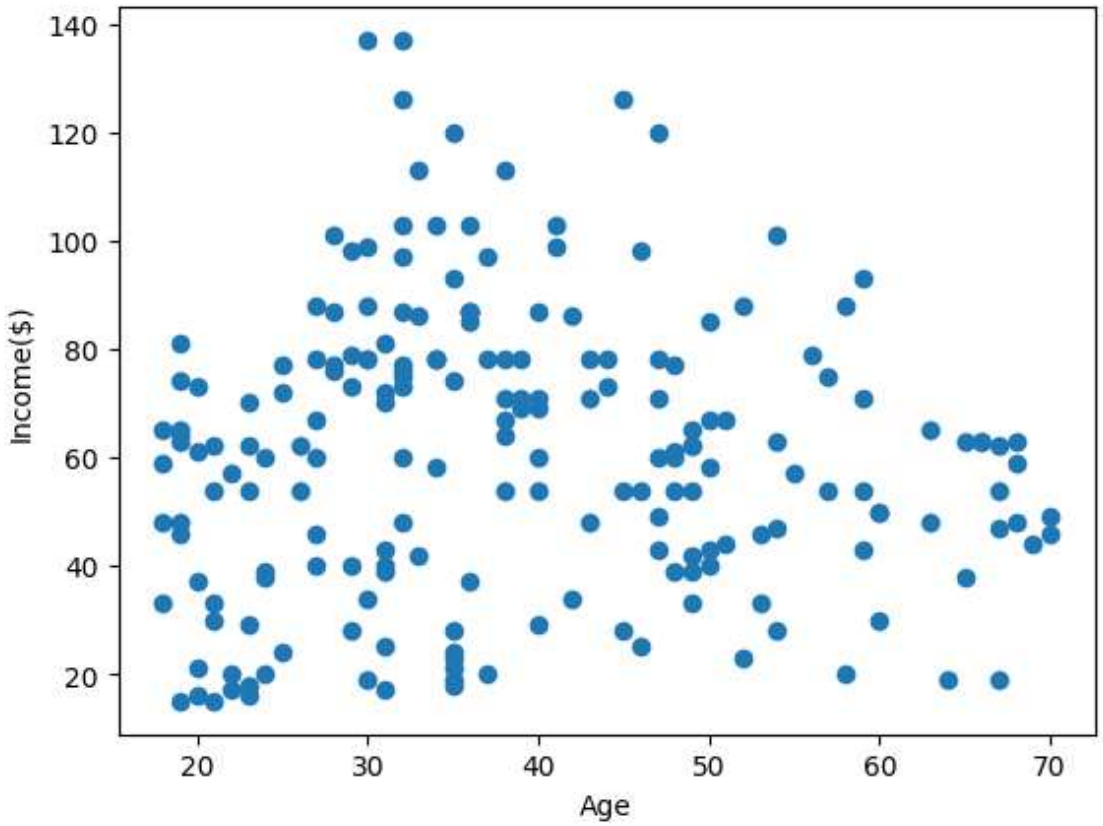
Out[2]:

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

Out[3]:
```
Text(0, 0.5, 'Income($)')
```



In [4]:
```
from sklearn.cluster import KMeans
```

In [5]:
```
km=KMeans()
km
```

Out[5]:
```
KMeans()
```

In [6]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
```

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

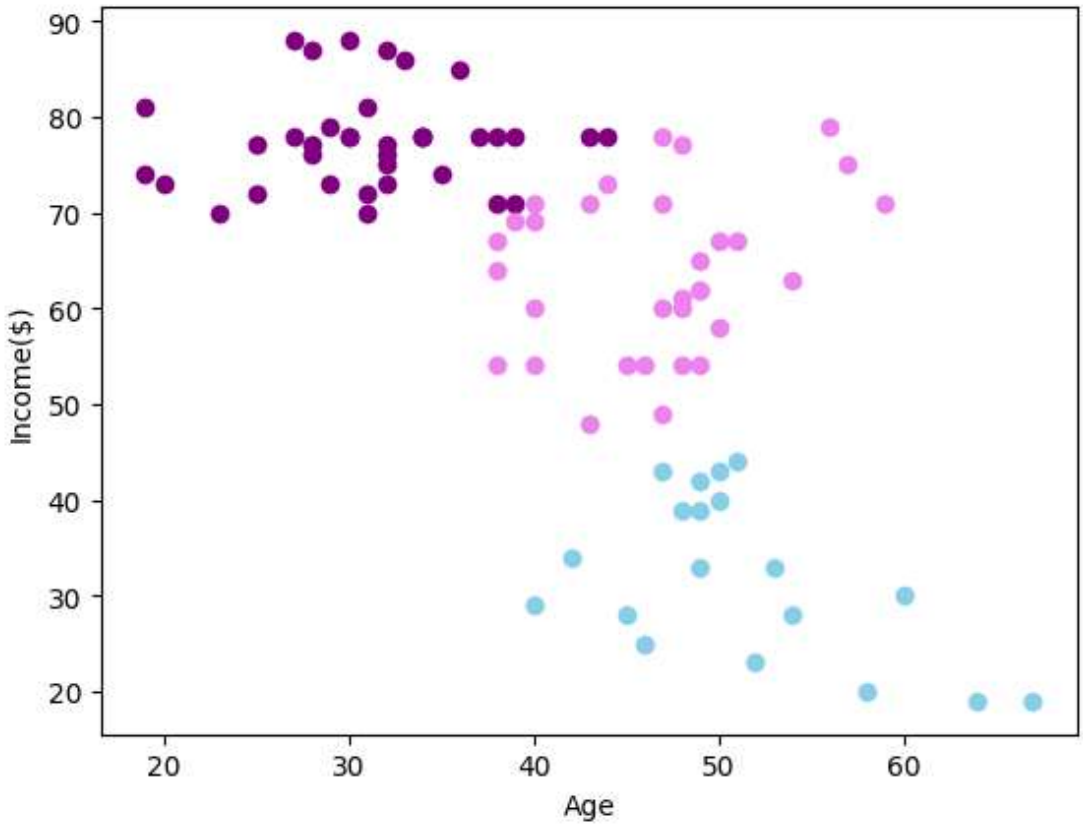
In [7]:
```
df["cluster"]=y_predicted
df.head()
```

Out[7]:

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

In [8]:
```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="purple")
plt.scatter(df2["Age"],df2["Income($)"],color="skyblue")
plt.scatter(df3["Age"],df3["Income($)"],color="violet")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[8]:
```
Text(0, 0.5, 'Income($)')
```



In [9]:
```
from sklearn.preprocessing import MinMaxScaler
```

In [10]:
```
Scaler=MinMaxScaler()
```

In [11]:
```
Scaler.fit(df[["Income($)"]])
df["Income($)"]=Scaler.transform(df[["Income($)"]])
df.head()
```

Out[11]:

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

In [12]:
```
Scaler.fit(df[["Age"]])
df["Age"]=Scaler.transform(df[["Age"]])
df.head()
```

Out[12]:

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

In [13]:
```
km=KMeans()
km
```

Out[13]:

▾ KMeans

KMeans()

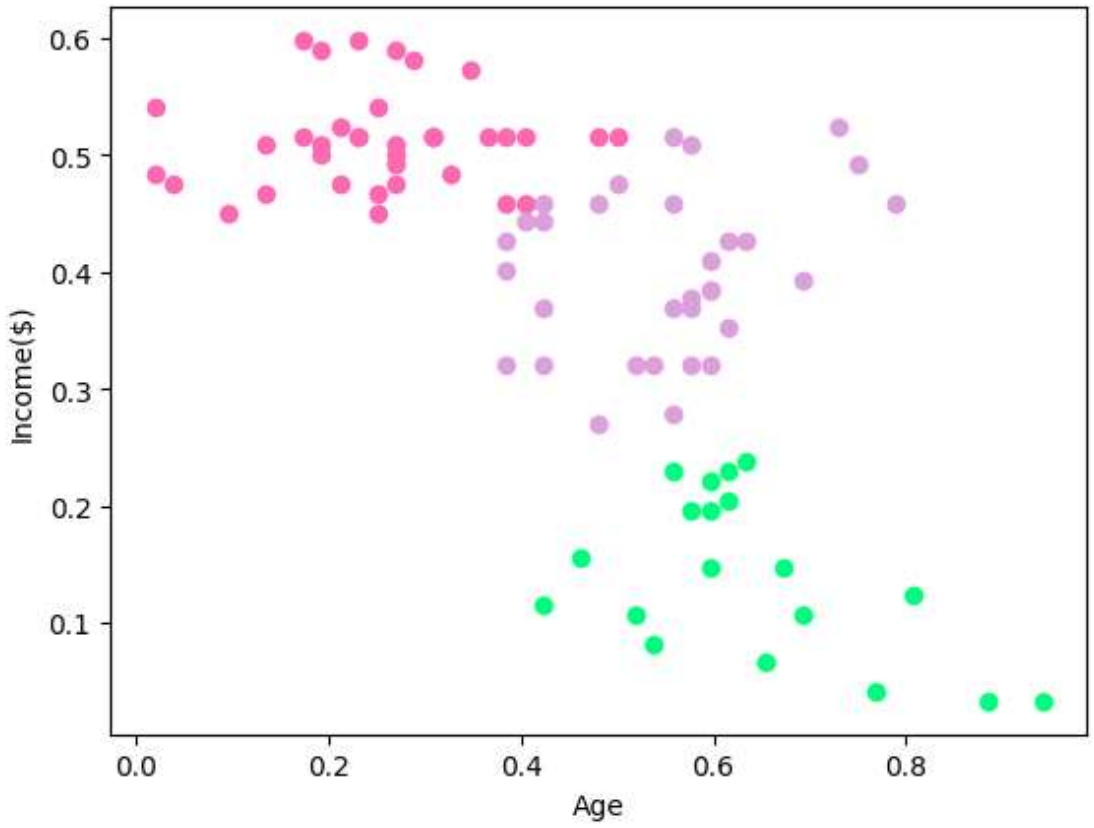
In [14]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
```

Out[14]:
```
array([1, 1, 1, 1, 3, 1, 3, 1, 5, 3, 5, 3, 0, 1, 3, 1, 3, 1, 0, 3, 3, 1,
       0, 3, 0, 3, 0, 3, 3, 1, 5, 1, 0, 1, 0, 1, 0, 3, 3, 1, 5, 1, 0, 3,
       0, 1, 0, 3, 3, 3, 0, 3, 3, 5, 0, 0, 0, 5, 6, 0, 5, 6, 5, 0, 5, 6,
       0, 5, 6, 3, 5, 0, 5, 5, 5, 6, 0, 0, 6, 0, 5, 7, 5, 0, 6, 0, 0, 6,
       7, 0, 5, 6, 2, 7, 7, 6, 2, 6, 2, 6, 6, 2, 5, 6, 2, 6, 5, 2, 5, 5,
       5, 6, 7, 6, 6, 6, 5, 2, 2, 2, 6, 7, 7, 7, 6, 7, 2, 7, 2, 7, 2, 7,
       6, 7, 6, 7, 2, 7, 6, 7, 2, 7, 7, 7, 6, 7, 2, 7, 7, 7, 2, 7, 2, 7,
       2, 7, 7, 7, 7, 7, 2, 7, 6, 7, 2, 7, 2, 7, 7, 7, 7, 7, 7, 7, 2, 7,
       2, 7, 2, 7, 4, 4, 2, 4, 4, 4, 2, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
       4, 4])
```

In [15]:
```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="hotpink")
plt.scatter(df2["Age"],df2["Income($)"],color="SpringGreen")
plt.scatter(df3["Age"],df3["Income($)"],color="plum")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[15]:
```
Text(0, 0.5, 'Income($)')
```

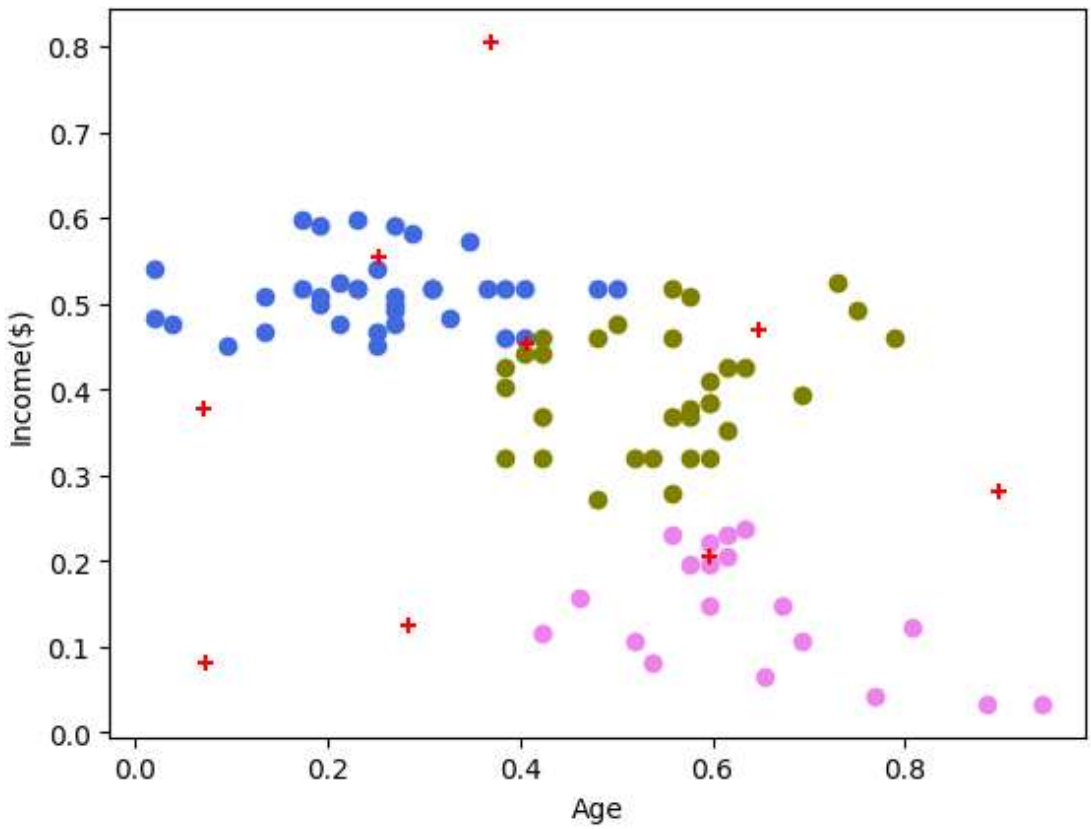


In [16]:
```
km.cluster_centers_
```

Out[16]:
```
array([[0.5954142 , 0.2203657 ],
       [0.07239819, 0.08003857],
       [0.61094675, 0.49401009],
       [0.28388278, 0.1245121 ],
       [0.32905983, 0.78551913],
       [0.89799331, 0.28011404],
       [0.07322485, 0.38272383],
       [0.3059034 , 0.50247808]])
```

In [30]:
```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="royalblue")
plt.scatter(df2["Age"],df2["Income($)"],color="violet")
plt.scatter(df3["Age"],df3["Income($)"],color="olive")
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="red",marker="
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[30]:
```
Text(0, 0.5, 'Income($)')
```



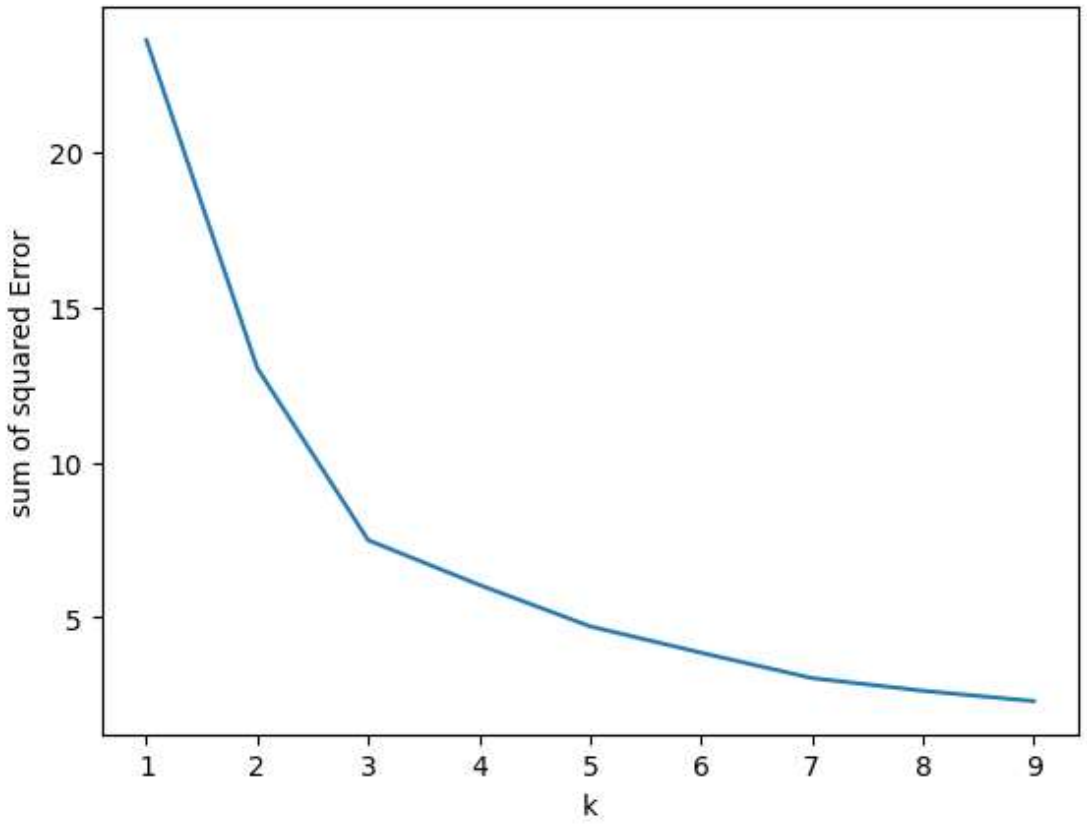
In [20]:
```
k_rng=range(1,10)
sse=[]
for k in k_rng:
 km=KMeans(n_clusters=k)
 km.fit(df[["Age","Income($)"]])
 sse.append(km.inertia_)
sse
```

```
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
C:\Users\jangidi veena\AppData\Local\Programs\Python\Python311\Lib\site-packages
\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` wil
l change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppre
ss the warning
  warnings.warn(
```

Out[20]:
```
[23.583906150363607,
 13.028938428018291,
 7.492107868586011,
 6.055858644812546,
 4.718644501279133,
 3.870455202564072,
 3.0547174363693586,
 2.646037617631439,
 2.313572035354328]
```

In [21]:
```
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("sum of squared Error")
```

Out[21]:
```
Text(0, 0.5, 'sum of squared Error')
```



In [ ]: