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

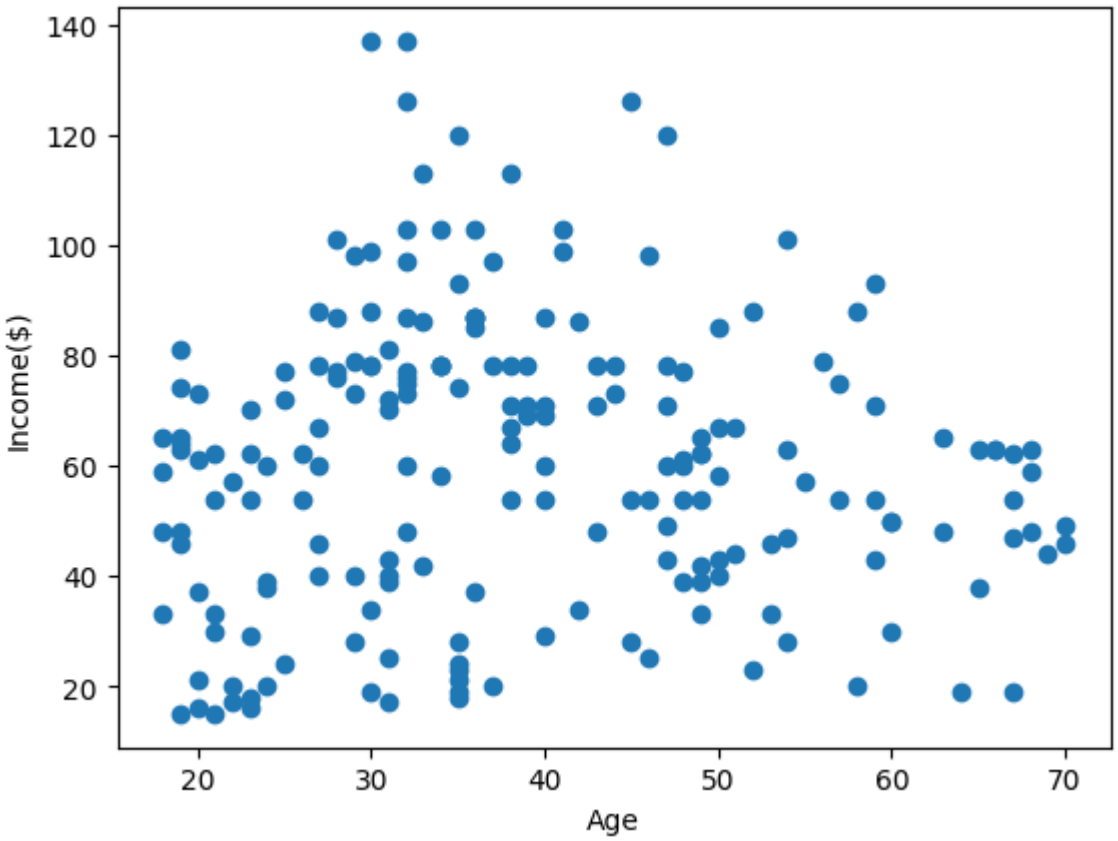
In [2]:
```
df=pd.read_csv(r"C:\Users\rubin\Downloads\Income.csv")
df.head()
```

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

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\rubin\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[6]:
```
array([6, 6, 6, 6, 6, 6, 6, 6, 3, 6, 3, 6, 3, 6, 6, 6, 6, 6, 3, 6, 6, 6,
       3, 6, 3, 6, 3, 6, 3, 6, 3, 6, 3, 0, 3, 0, 3, 0, 0, 0, 3, 0, 3, 0,
       3, 0, 3, 0, 0, 0, 3, 0, 0, 3, 3, 3, 3, 2, 0, 3, 2, 0, 2, 3, 2, 0,
       3, 2, 0, 0, 2, 3, 2, 2, 2, 0, 7, 7, 0, 7, 2, 7, 2, 7, 0, 7, 2, 0,
       7, 7, 2, 4, 7, 7, 4, 4, 7, 4, 7, 4, 4, 7, 2, 4, 7, 4, 2, 7, 2, 2,
       2, 4, 7, 4, 4, 4, 2, 7, 7, 7, 4, 7, 7, 7, 4, 4, 7, 7, 7, 7, 7, 7,
       4, 4, 4, 4, 7, 4, 4, 4, 7, 4, 4, 4, 4, 4, 7, 4, 4, 4, 7, 4, 7, 4,
       7, 4, 4, 4, 4, 4, 7, 4, 4, 4, 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, 5, 5, 5, 5, 5, 5,
       5, 5])
```

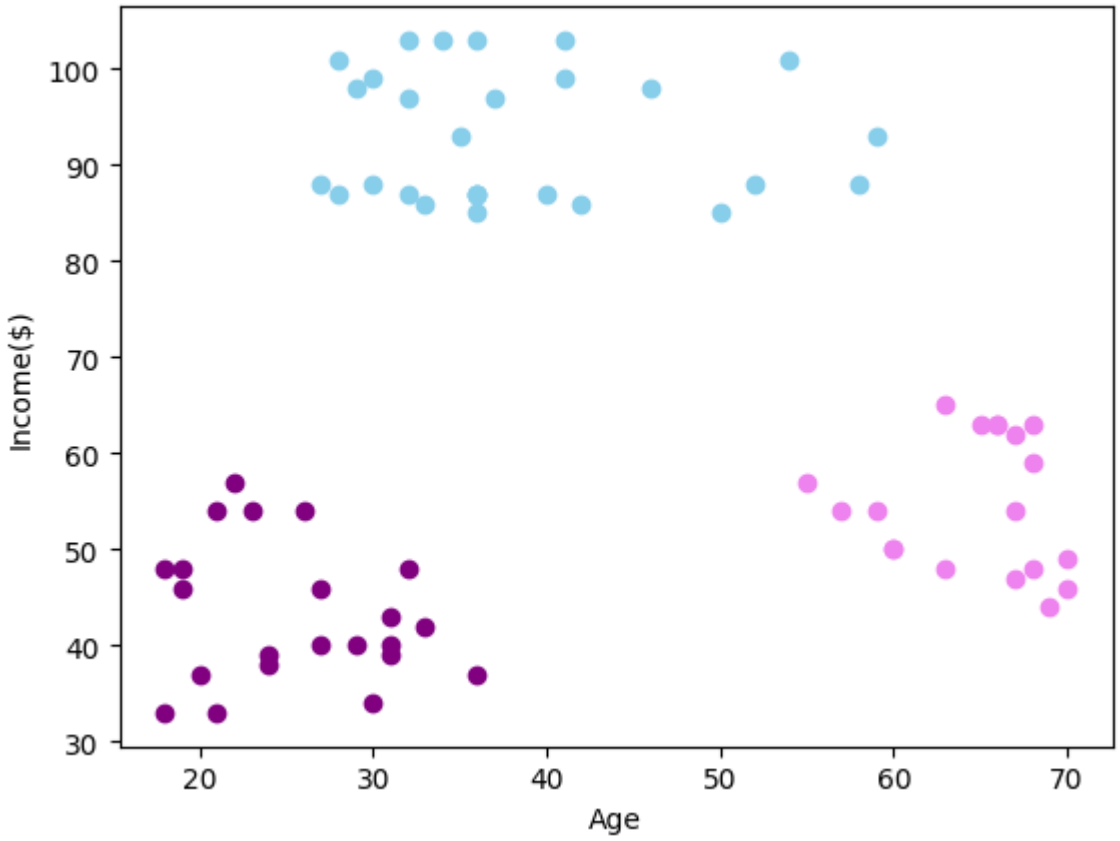
In [7]:
```
df["cluster"]=y_predicted
df.head()
```

Out[7]:

	Gender	Age	Income($)	cluster
0	Male	19	15	6
1	Male	21	15	6
2	Female	20	16	6
3	Female	23	16	6
4	Female	31	17	6

In [10]:
```
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[10]:
```
Text(0, 0.5, 'Income($)')
```



In [11]:
```
from sklearn.preprocessing import MinMaxScaler
```

In [12]:
```
Scaler=MinMaxScaler()
```

In [13]:
```
Scaler.fit(df[["Income($)"]])
df["Income($)"]=Scaler.transform(df[["Income($)"]])
df.head()
```

Out[13]:

	Gender	Age	Income($)	cluster
0	Male	19	0.000000	6
1	Male	21	0.000000	6
2	Female	20	0.008197	6
3	Female	23	0.008197	6
4	Female	31	0.016393	6

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	6
1	Male	0.057692	0.000000	6
2	Female	0.038462	0.008197	6
3	Female	0.096154	0.008197	6
4	Female	0.250000	0.016393	6

In [15]:
```
km=KMeans()
km
```

Out[15]:

▾ KMeans

KMeans()

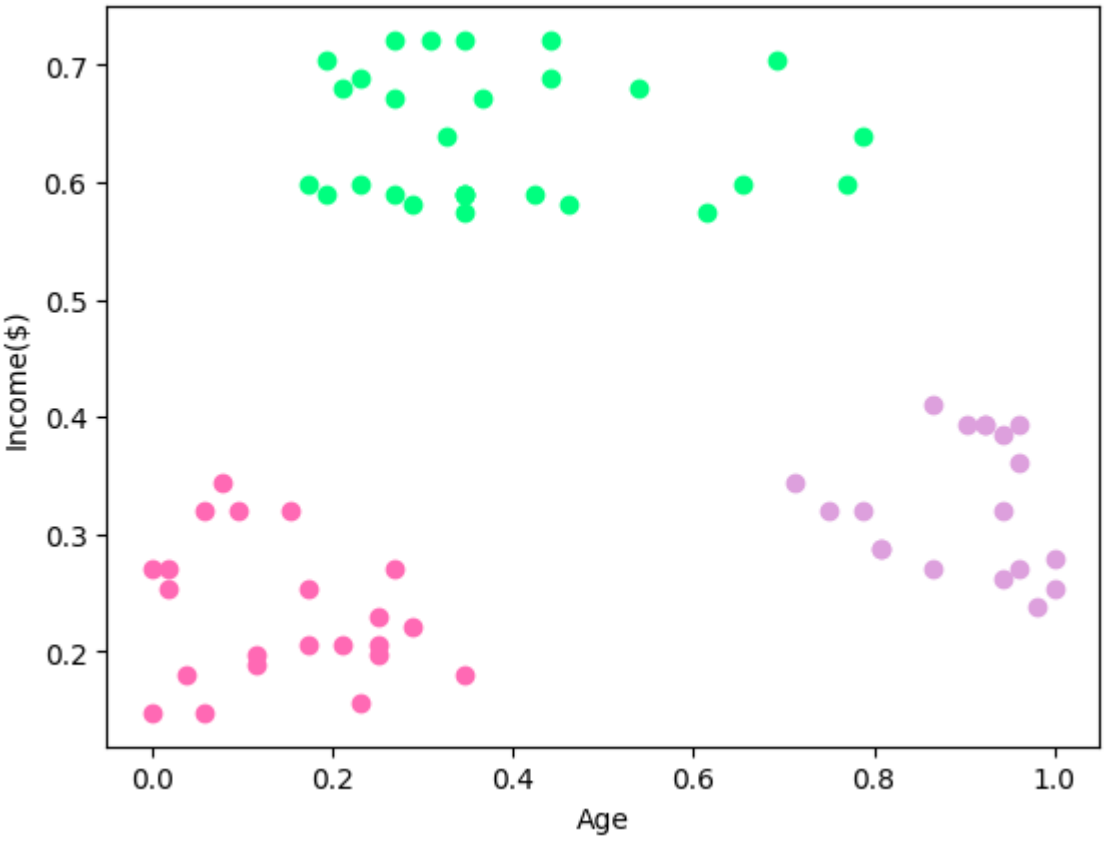
In [16]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\rubin\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[16]:
```
array([2, 2, 2, 2, 5, 2, 5, 2, 6, 5, 6, 5, 0, 2, 5, 2, 5, 2, 0, 5, 5, 2,
       0, 5, 0, 5, 0, 5, 5, 2, 6, 2, 0, 2, 0, 2, 0, 5, 5, 2, 6, 2, 0, 5,
       0, 2, 0, 5, 5, 5, 0, 5, 5, 6, 0, 0, 0, 6, 4, 0, 6, 4, 6, 0, 6, 4,
       0, 6, 4, 5, 6, 0, 6, 6, 6, 4, 0, 0, 4, 0, 6, 3, 6, 0, 4, 0, 7, 4,
       3, 7, 6, 4, 7, 3, 3, 4, 7, 4, 7, 4, 4, 7, 6, 4, 7, 4, 6, 7, 6, 6,
       6, 4, 3, 4, 4, 4, 6, 7, 7, 7, 4, 3, 3, 3, 4, 3, 7, 3, 7, 3, 7, 3,
       4, 3, 4, 3, 7, 3, 4, 3, 7, 3, 3, 3, 4, 3, 7, 3, 3, 3, 7, 3, 7, 3,
       7, 3, 3, 3, 3, 3, 7, 3, 4, 3, 7, 3, 3, 3, 3, 3, 3, 3, 3, 3, 7, 3,
       7, 3, 7, 1, 1, 1, 1, 1, 1, 1, 7, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1])
```

In [18]:
```
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[18]:
```
Text(0, 0.5, 'Income($)')
```

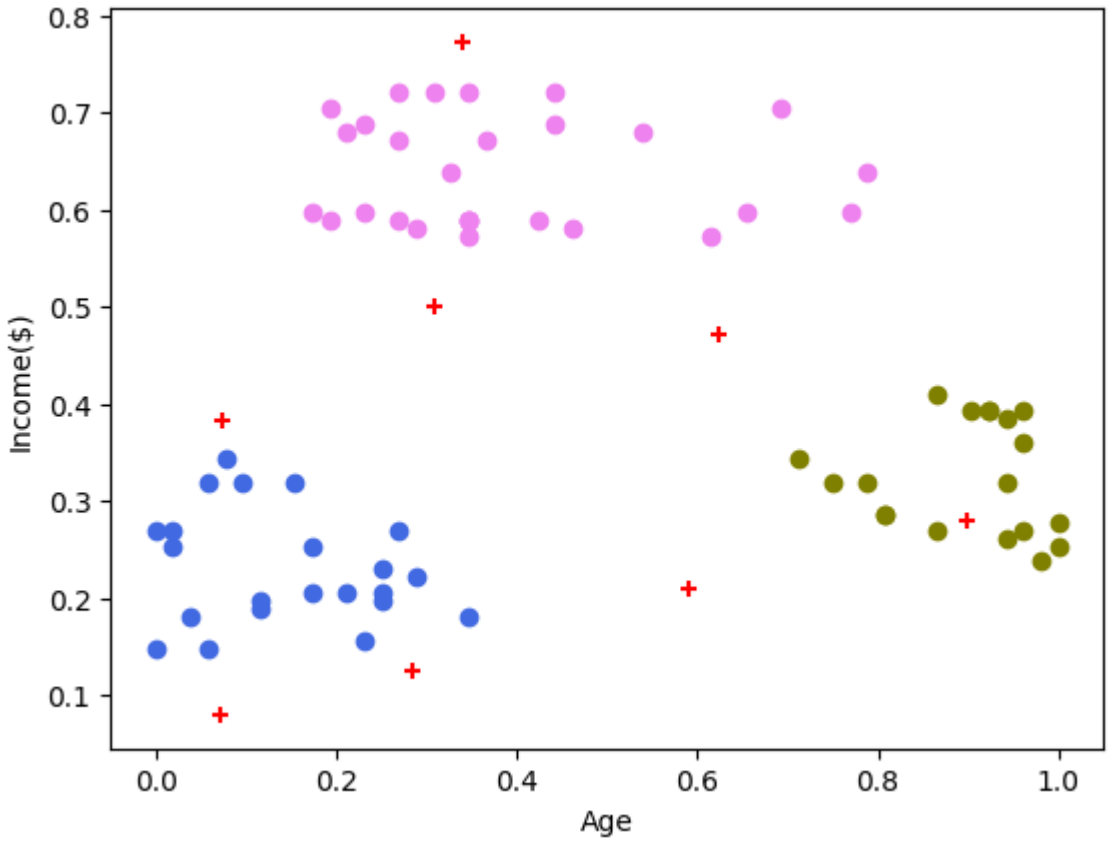


In [19]:
```
km.cluster_centers_
```

Out[19]:
```
array([[0.58974359, 0.20969945],
       [0.33942308, 0.77295082],
       [0.07239819, 0.08003857],
       [0.30903399, 0.50114373],
       [0.07322485, 0.38272383],
       [0.28388278, 0.1245121 ],
       [0.89799331, 0.28011404],
       [0.62352071, 0.47225725]])
```

In [25]:
```
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($)"],color="olive")
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="red",mark
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[25]:
```
Text(0, 0.5, 'Income($)')
```



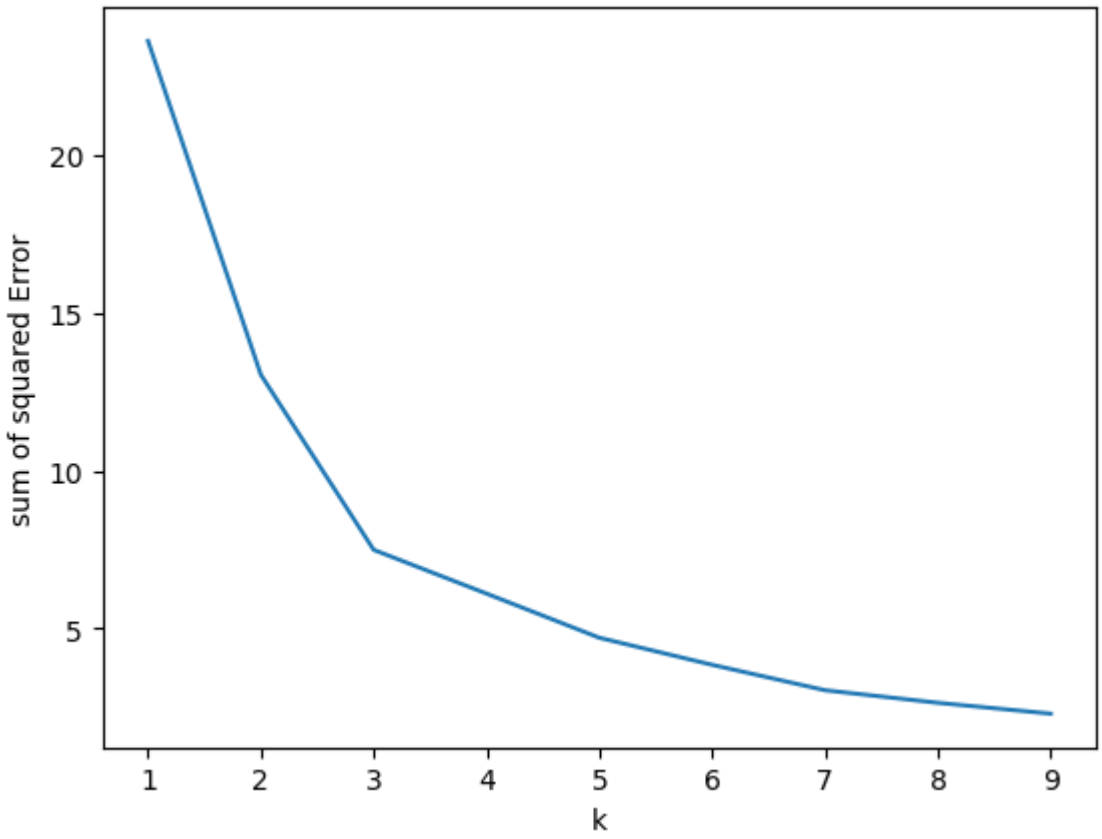
In [26]:
```
k_rng=range(1,10)
sse=[]
for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[["Age","Income($)"]])
    sse.append(km.inertia_)
sse
```

```
C:\Users\rubin\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\rubin\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\rubin\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\rubin\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\rubin\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\rubin\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\rubin\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\rubin\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\rubin\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[26]:
```
[23.583906150363607,
 13.028938428018291,
 7.493024843304991,
 6.109631252714937,
 4.71335402850073,
 3.857891822164646,
 3.0547174363693586,
 2.6570812587207415,
 2.313572035354328]
```

In [27]:
```
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("sum of squared Error")
```

Out[27]:
```
Text(0, 0.5, 'sum of squared Error')
```



In [ ]: