K-Means Clustring

In [12]:

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

In [13]:

```
df=pd.read_csv(r"C:\Users\RAMADEVI SURIPAKA\Downloads\Income.csv")
df.head()
```

Out[13]:

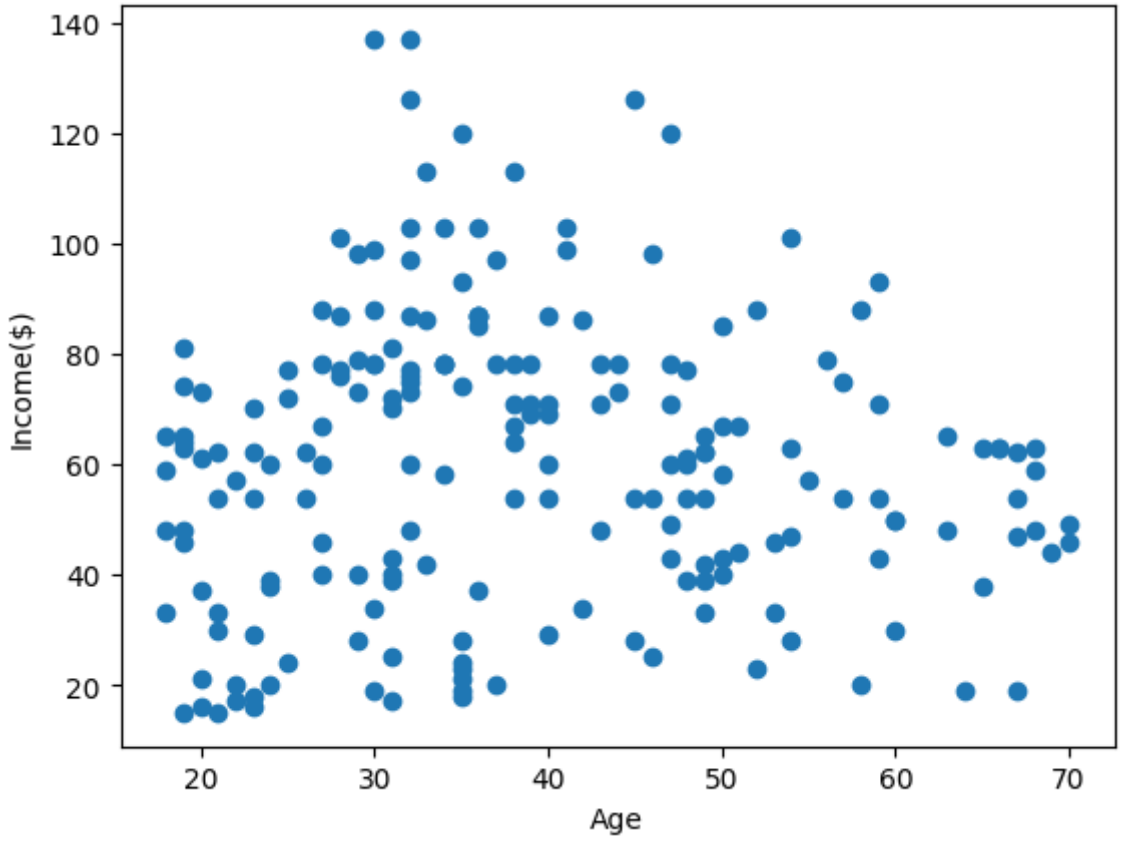
	Gender	Age	Income($)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17

In [14]:

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

Out[14]:

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



In [20]:

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

In [21]:

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

Out[21]:

```
KMeans()
```

In [22]:

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

```
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
```

Out[22]:

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

In [23]:

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

Out[23]:

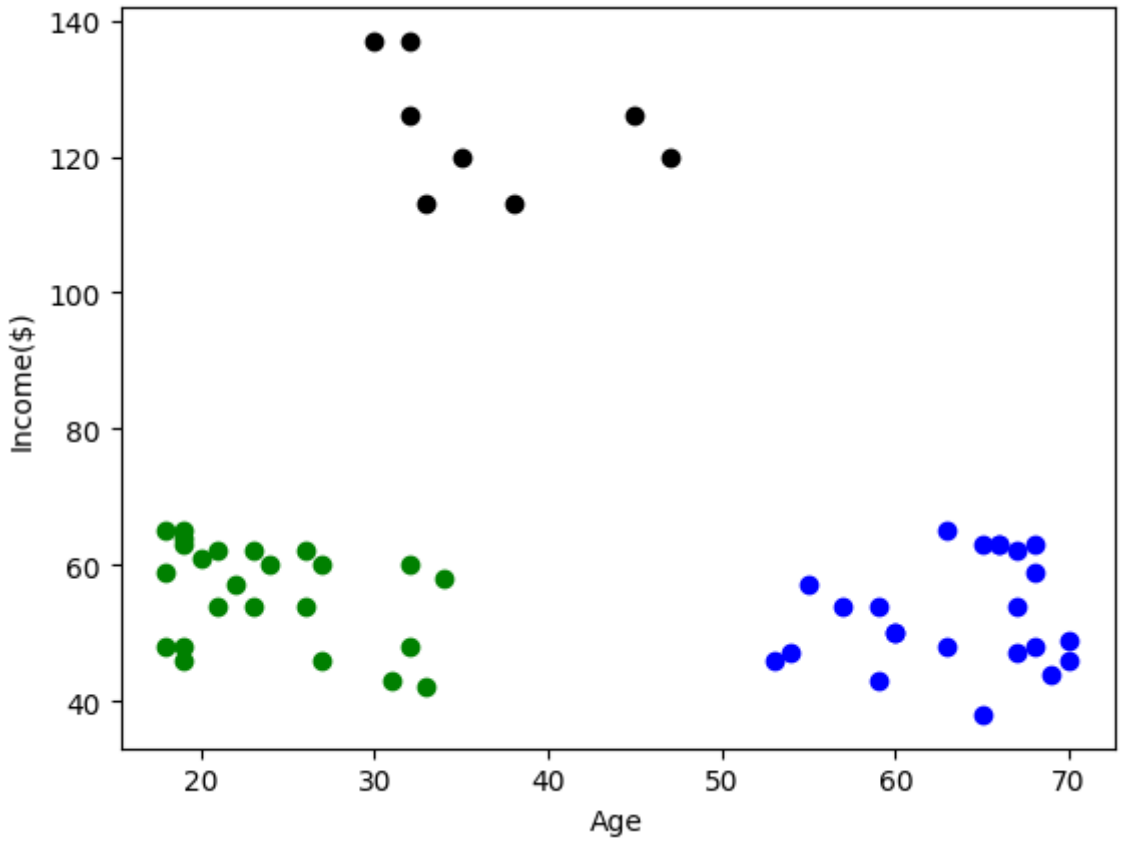
	Gender	Age	Income($)	cluster
0	Male	19	15	6
1	Male	21	15	6
2	Female	20	16	6
3	Female	23	16	6
4	Female	31	17	6

In [27]:

```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="black")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[27]:

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



In [28]:

```
from sklearn.preprocessing import MinMaxScaler
```

In [29]:

```
Scaler=MinMaxScaler()
```

In [32]:

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

Out[32]:

	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 [33]:

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

Out[33]:

	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 [36]:

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

Out[36]:

```
KMeans
KMeans()
```

In [37]:

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

```
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
```

Out[37]:

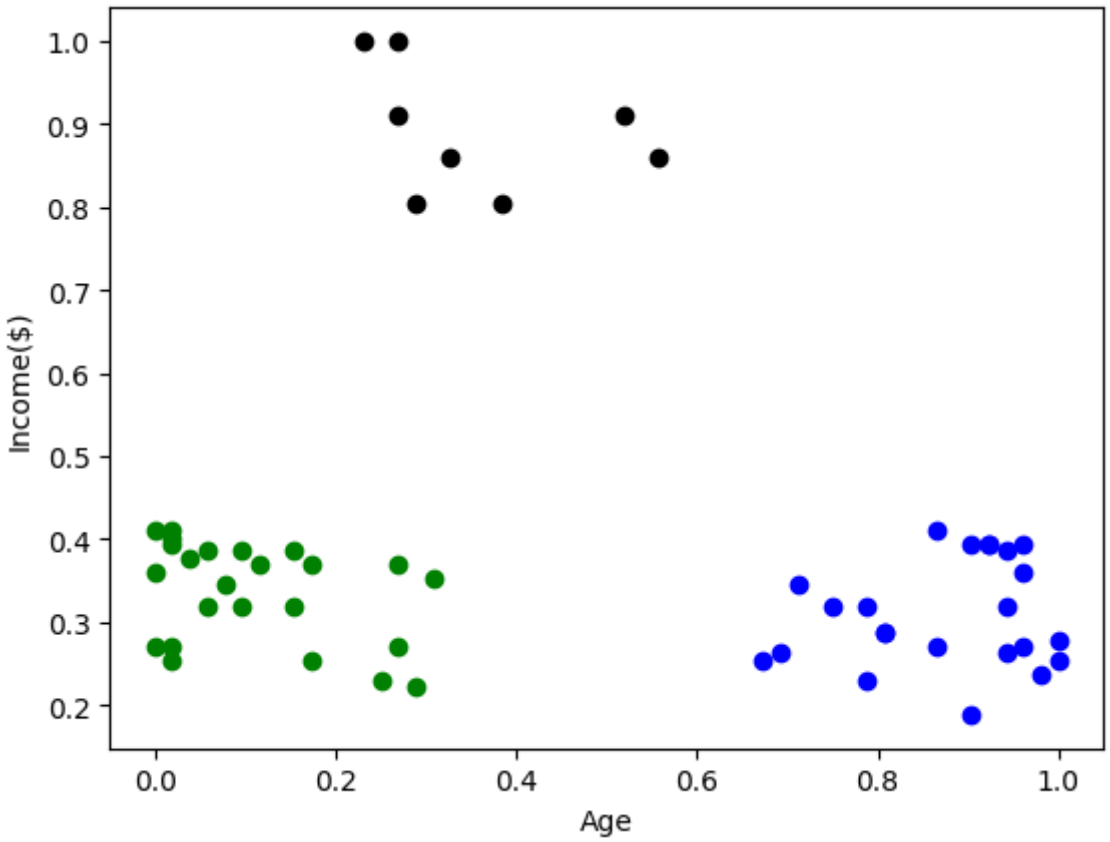
```
array([2, 2, 2, 2, 7, 2, 7, 2, 3, 7, 3, 7, 0, 2, 7, 2, 7, 2, 0, 7, 7, 2,
       0, 7, 0, 7, 0, 7, 7, 2, 3, 2, 0, 2, 0, 2, 0, 7, 7, 2, 3, 2, 0, 7,
       0, 2, 0, 7, 7, 7, 0, 7, 7, 3, 0, 0, 0, 3, 7, 0, 3, 1, 3, 0, 3, 1,
       0, 3, 1, 7, 3, 0, 3, 3, 3, 1, 0, 0, 1, 0, 3, 6, 3, 0, 1, 0, 5, 1,
    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

In [38]:

```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="black")
plt.scatter(df2["Age"],df2["Income($)"],color="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[38]:

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



In [39]:

```
km.cluster_centers_
```

Out[39]:

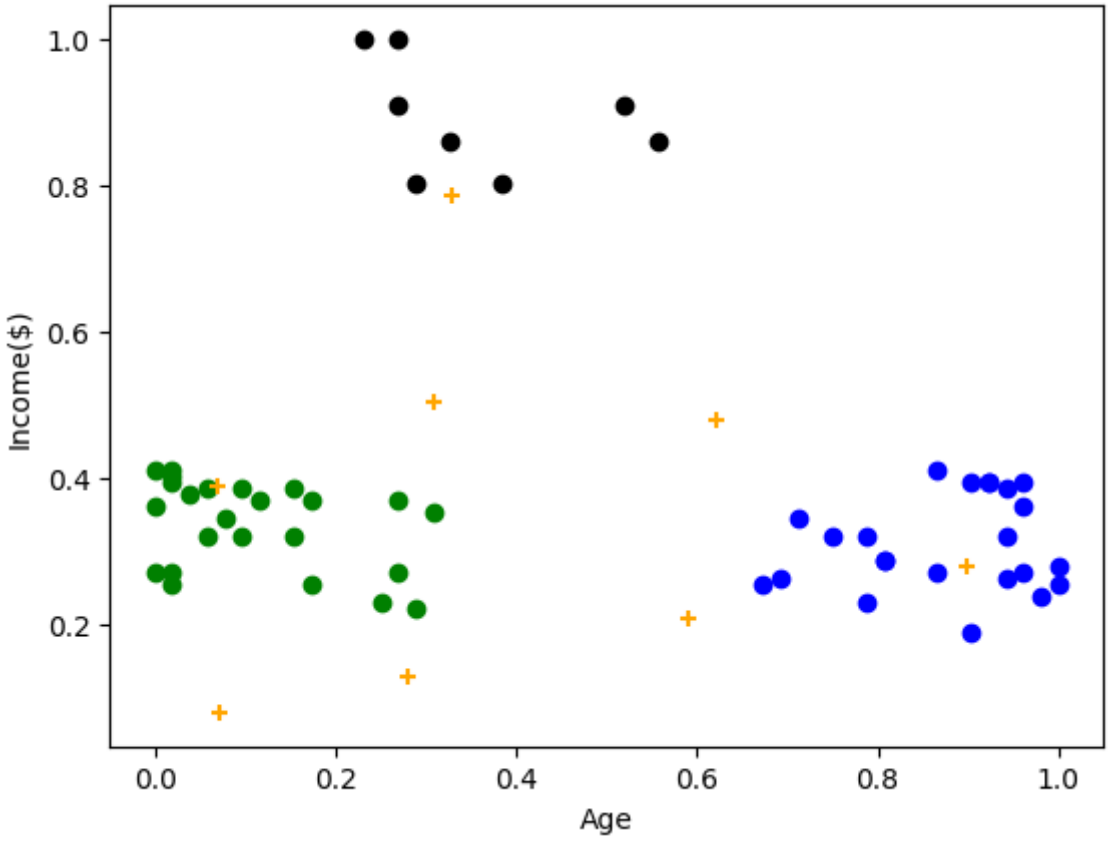
```
array([[0.58974359, 0.20969945],
       [0.06923077, 0.38786885],
       [0.07239819, 0.08003857],
       [0.89799331, 0.28011404],
       [0.32905983, 0.78551913],
       [0.62037037, 0.47996357],
       [0.30944056, 0.50428465],
       [0.27884615, 0.13040238]])
```

In [40]:

```
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="black")
plt.scatter(df2["Age"],df2["Income($)"],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[40]:

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



In [44]:

```
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
```

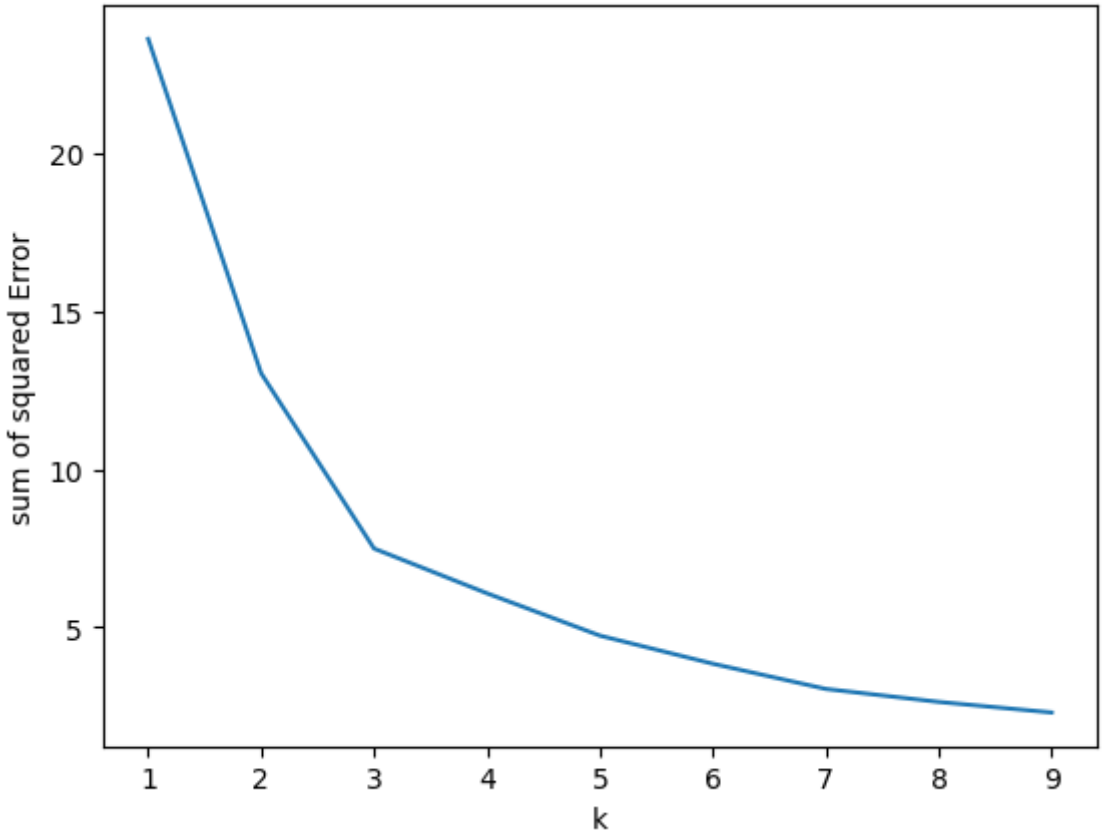
```
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
C:\Users\RAMADEVI SURIPAKA\AppData\Local\Programs\Python\Python310\lib\sit
e-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default valu
e of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_in
it` explicitly to suppress the warning
  warnings.warn(
```

In [46]:

```
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("sum of squared Error")
```

Out[46]:

```
Text(0, 0.5, 'sum of squared Error')
```



In [ ]: