In [1]:

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

In [2]:

```
data=pd.read_csv(r"C:\Users\Prathyusha\Downloads\Income.csv")
data
```

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

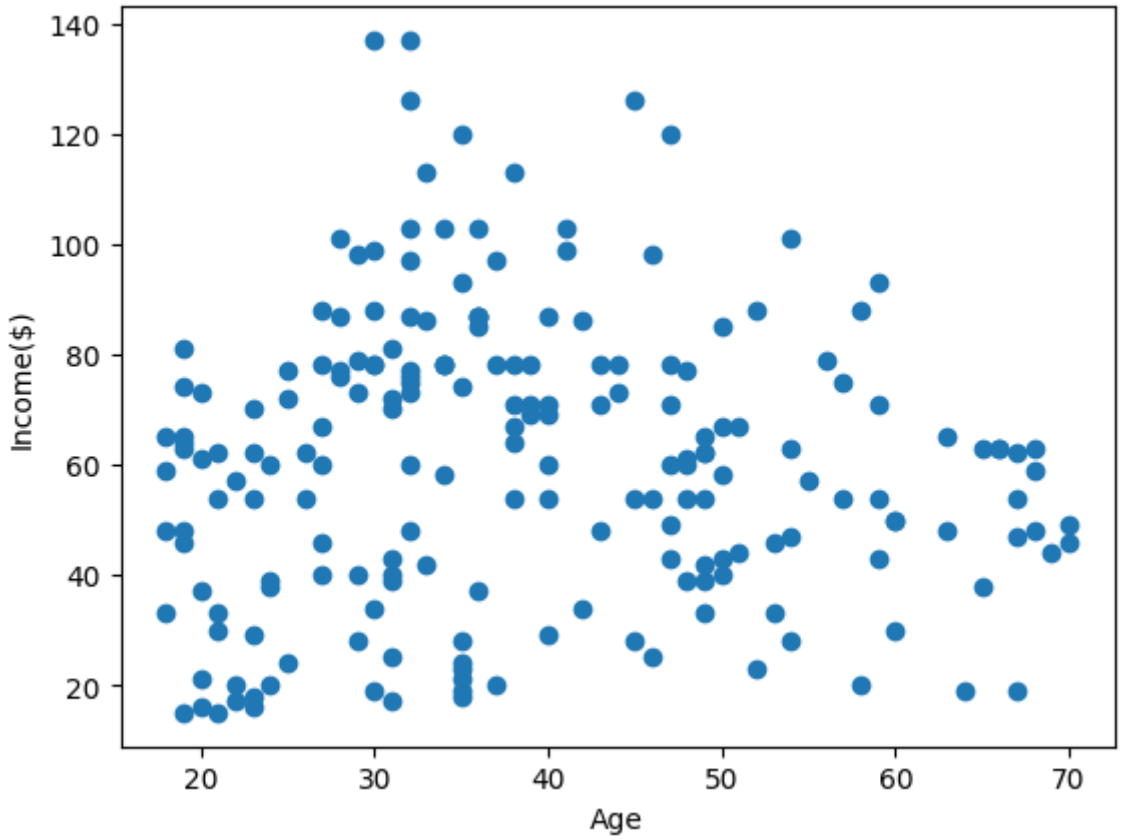
200 rows × 3 columns

In [3]:

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

Out[3]:

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



In [4]:

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

Out[4]:

```
KMeans()
```

In [5]:

```
y_predicted=Km.fit_predict(data[["Age","Income($)"]])
y_predicted
```

```
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
```

Out[5]:

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

In [6]:

```
data['cluster']=y_predicted
data.head()
```

Out[6]:

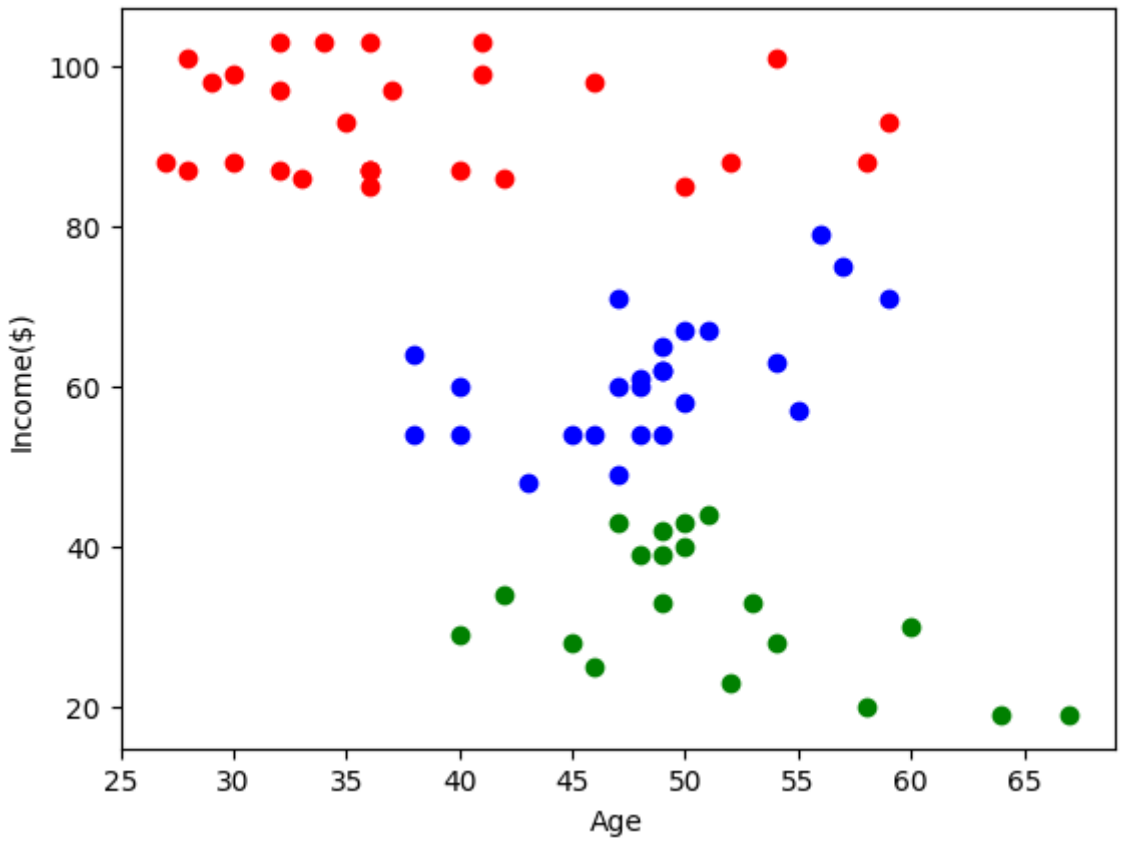
	Gender	Age	Income($)	cluster
0	Male	19	15	7
1	Male	21	15	7
2	Female	20	16	7
3	Female	23	16	7
4	Female	31	17	7

In [7]:

```
data1=data[data.cluster==0]
data2=data[data.cluster==1]
data3=data[data.cluster==2]
plt.scatter(data1["Age"],data1["Income($)"],color="red")
plt.scatter(data2["Age"],data2["Income($)"],color="green")
plt.scatter(data3["Age"],data3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[7]:

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



In [8]:

```
from sklearn.preprocessing import MinMaxScaler
Scaler=MinMaxScaler()
Scaler.fit(data[["Income($)"]])
data["Income($)"]=Scaler.transform(data[["Income($)"]])
data.head()
```

Out[8]:

	Gender	Age	Income($)	cluster
0	Male	19	0.000000	7
1	Male	21	0.000000	7
2	Female	20	0.008197	7
3	Female	23	0.008197	7
4	Female	31	0.016393	7

In [9]:

```
Scaler.fit(data[["Age"]])
Km=KMeans()
y_predicted=Km.fit_predict(data[["Age","Income($)"]])
y_predicted
```

```
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
```

Out[9]:

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

In [10]:

```
data["New cluster"]=y_predicted
data
```

Out[10]:

	Gender	Age	Income($)	cluster	New cluster
0	Male	19	0.000000	7	6
1	Male	21	0.000000	7	6
2	Female	20	0.008197	7	6
3	Female	23	0.008197	7	0
4	Female	31	0.016393	7	5
...	...	...	...	...	...
195	Female	35	0.860656	5	4
196	Female	45	0.909836	5	2
197	Male	32	0.909836	5	5
198	Male	32	1.000000	5	5
199	Male	30	1.000000	5	5

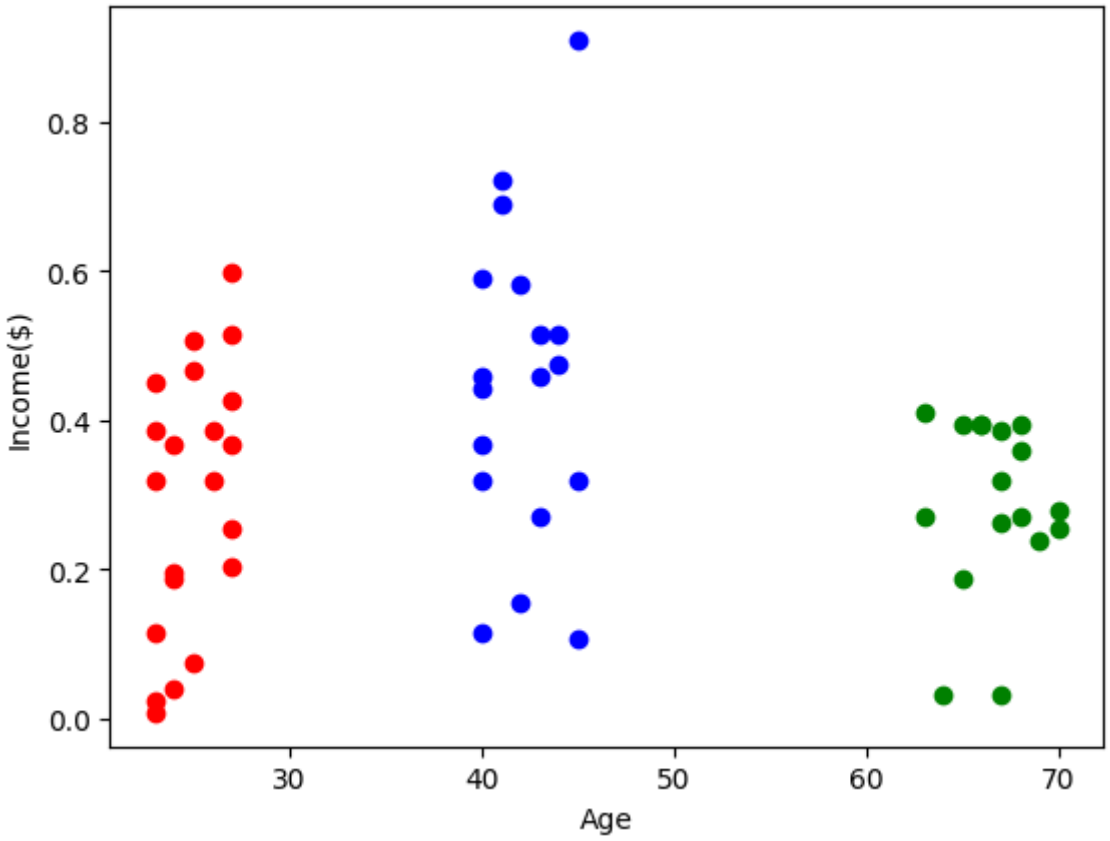
200 rows × 5 columns

In [11]:

```
data1=data[data["New cluster"]==0]
data2=data[data["New cluster"]==1]
data3=data[data["New cluster"]==2]
plt.scatter(data1["Age"],data1["Income($)"],color="red")
plt.scatter(data2["Age"],data2["Income($)"],color="green")
plt.scatter(data3["Age"],data3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[11]:

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



In [12]:

```
Km.cluster_centers_
```

Out[12]:

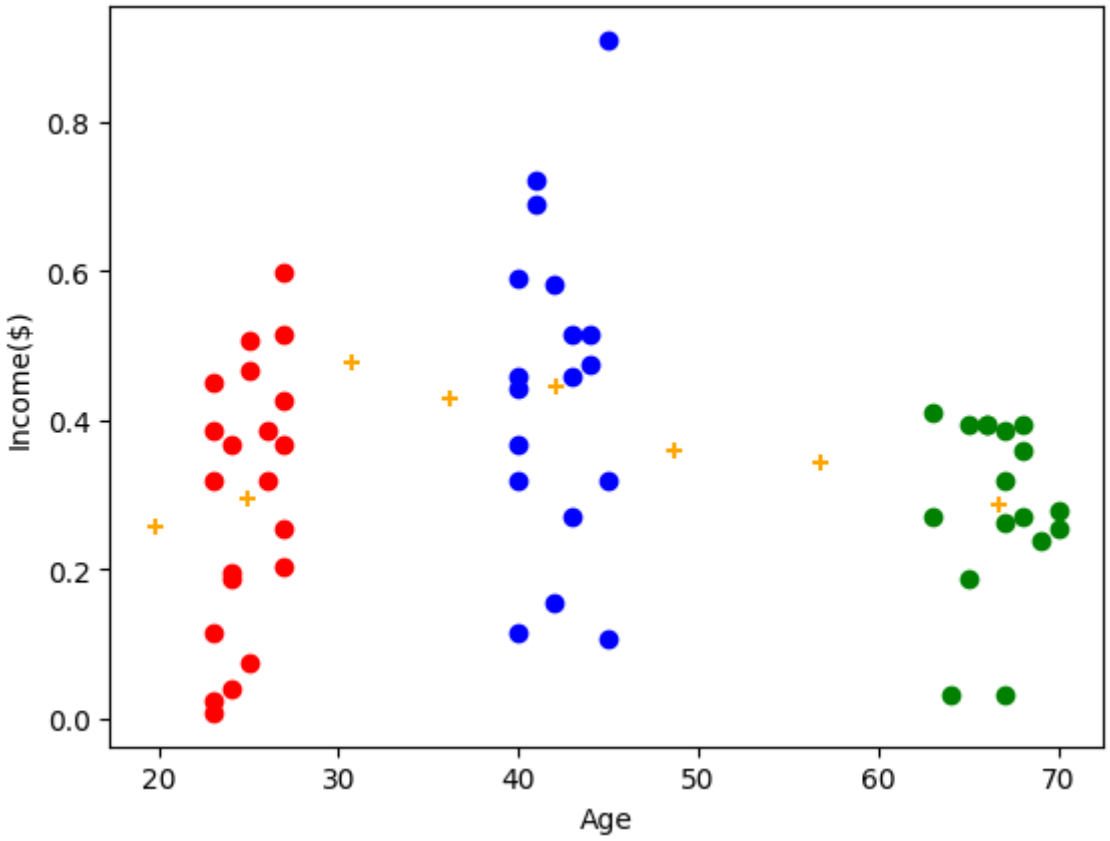
```
array([[24.9047619 ,  0.29625293],
       [66.64705882,  0.28688525],
       [42.11111111,  0.44535519],
       [48.63333333,  0.35819672],
       [36.15625   ,  0.43058402],
       [30.68421053,  0.47648835],
       [19.8       ,  0.25639344],
       [56.78947368,  0.34383089]])
```

In [19]:

```
data1=data[data["New cluster"]==0]
data2=data[data["New cluster"]==1]
data3=data[data["New cluster"]==2]
plt.scatter(data1["Age"],data1["Income($)"],color="red")
plt.scatter(data2["Age"],data2["Income($)"],color="green")
plt.scatter(data3["Age"],data3["Income($)"],color="blue")
plt.scatter(Km.cluster_centers_[:,0],Km.cluster_centers_[:,1],color="orange",marker="+")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[19]:

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



In [23]:

```
K_rng=range(1,10)
sse=[]
for K in K_rng:
    Km=KMeans(n_clusters=K)
    Km.fit(data[["Age","Income($)"]])
    sse.append(Km.inertia_)
    sse
```

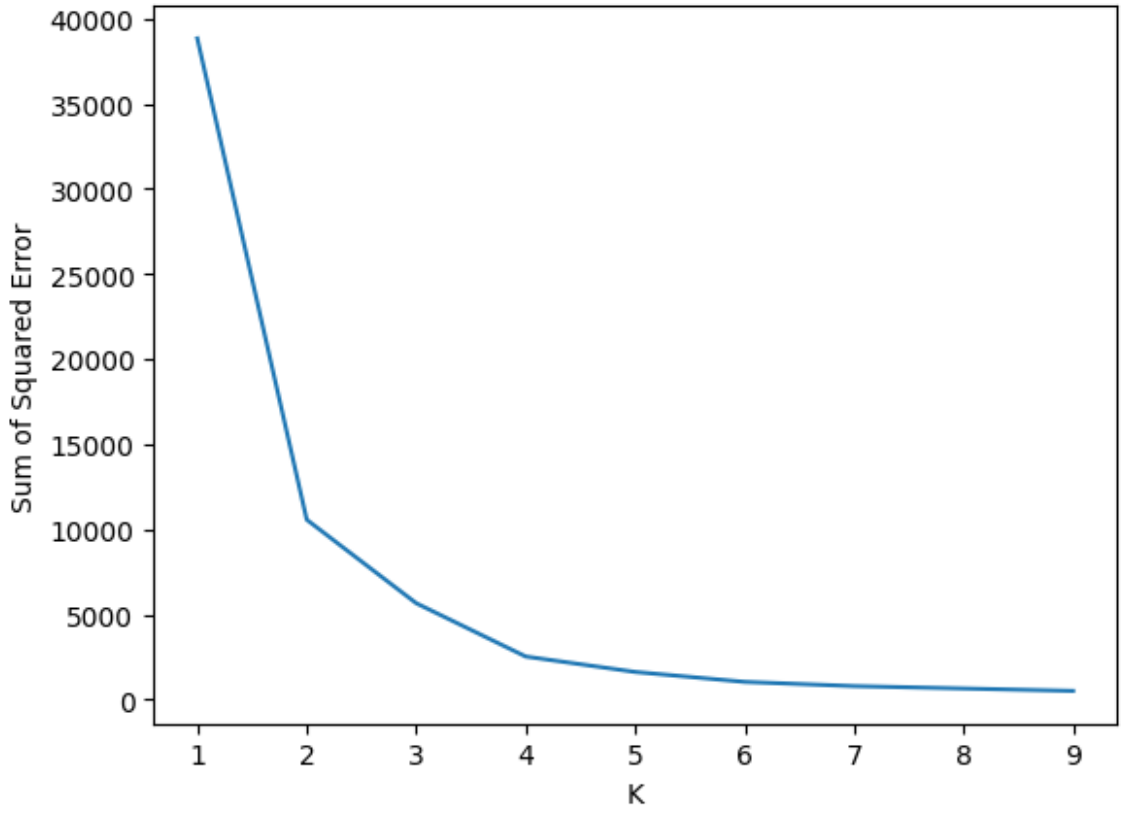
```
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n
_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
C:\Users\Prathyusha\AppData\Local\Programs\Python\Python310\lib\site-packa
ges\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` exp
licitly to suppress the warning
  warnings.warn(
```

In [25]:

```
plt.plot(K_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

Out[25]:

```
Text(0, 0.5, 'Sum of Squared Error')
```



In [ ]: