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

In [2]:
```
df=pd.read_csv(r"C:\Users\91903\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 [3]:
```
df.head()
```

Out[3]:

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

In [4]:
```
df.tail()
```

Out[4]:

	Gender	Age	Income($)
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 3 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   Gender     200 non-null    object
 1   Age        200 non-null    int64 
 2   Income($)  200 non-null    int64 
dtypes: int64(2), object(1)
memory usage: 4.8+ KB
```

In [6]:
```
df.describe()
```

Out[6]:

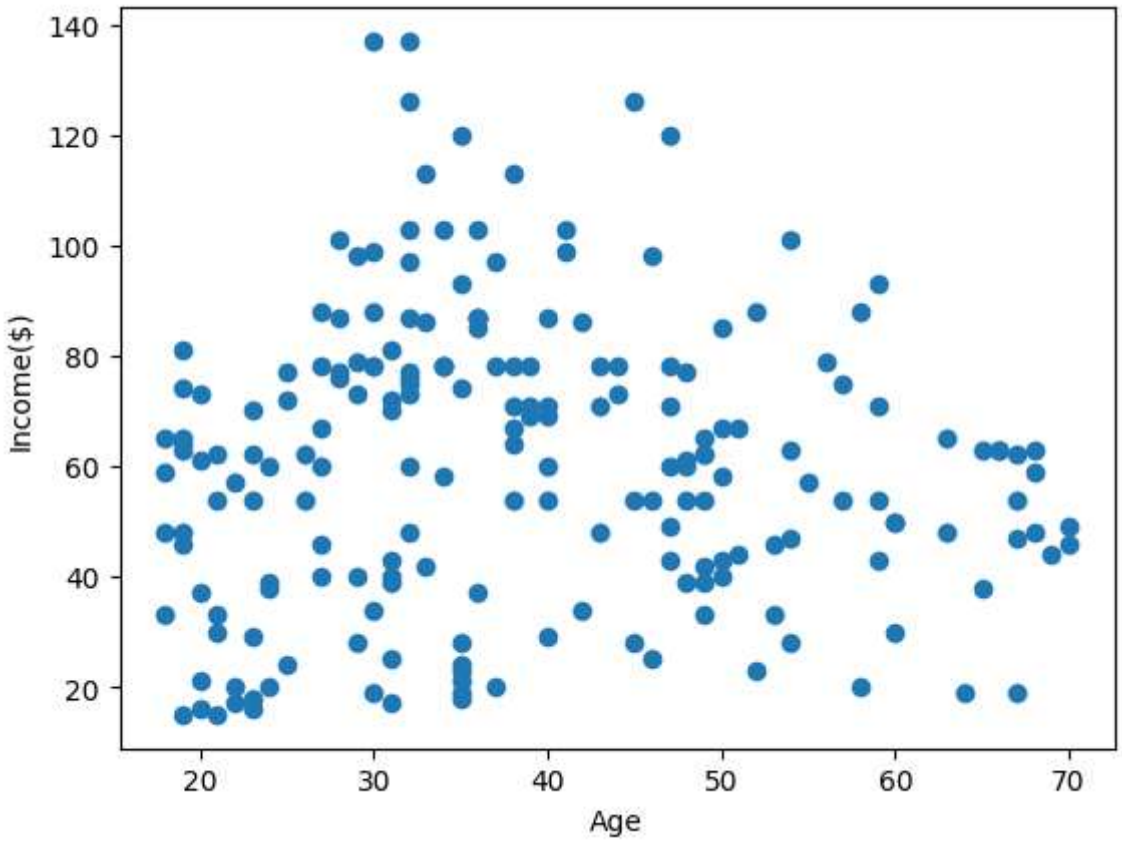
	Age	Income($)
count	200.000000	200.000000
mean	38.850000	60.560000
std	13.969007	26.264721
min	18.000000	15.000000
25%	28.750000	41.500000
50%	36.000000	61.500000
75%	49.000000	78.000000
max	70.000000	137.000000

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
Gender       0
Age          0
Income($)    0
dtype: int64
```

In [8]:
```
plt.scatter(df["Age"],df["Income($)"])
plt.xlabel("Age")
plt.ylabel("Income($)")
```

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



In [9]:
```
from sklearn.cluster import KMeans
km=KMeans()
km
```

Out[9]:
```
KMeans()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [10]:
```
y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

```
C:\Users\91903\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[10]:
```
array([2, 2, 2, 2, 2, 2, 2, 2, 4, 2, 4, 2, 4, 2, 2, 2, 2, 2, 4, 2, 2, 2,
       4, 2, 4, 2, 4, 2, 2, 2, 4, 2, 4, 2, 4, 2, 4, 2, 2, 2, 4, 2, 4, 2,
       4, 2, 4, 2, 2, 2, 4, 5, 5, 4, 4, 4, 4, 0, 5, 4, 0, 5, 0, 4, 0, 5,
       4, 0, 5, 5, 0, 4, 0, 0, 0, 5, 6, 6, 5, 6, 0, 5, 0, 6, 5, 6, 0, 5,
       5, 6, 0, 5, 6, 6, 5, 5, 6, 5, 6, 5, 5, 6, 0, 5, 6, 5, 0, 6, 0, 0,
       0, 5, 6, 5, 5, 5, 0, 6, 6, 6, 5, 6, 6, 6, 1, 1, 6, 6, 6, 6, 6, 6,
       1, 1, 1, 1, 6, 1, 1, 1, 6, 1, 1, 1, 1, 1, 6, 1, 1, 1, 6, 1, 6, 1,
       6, 1, 1, 1, 1, 1, 6, 1, 1, 1, 3, 1, 3, 1, 1, 1, 3, 1, 1, 1, 3, 1,
       3, 1, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 7, 7, 7, 7, 7, 7,
       7, 7])
```

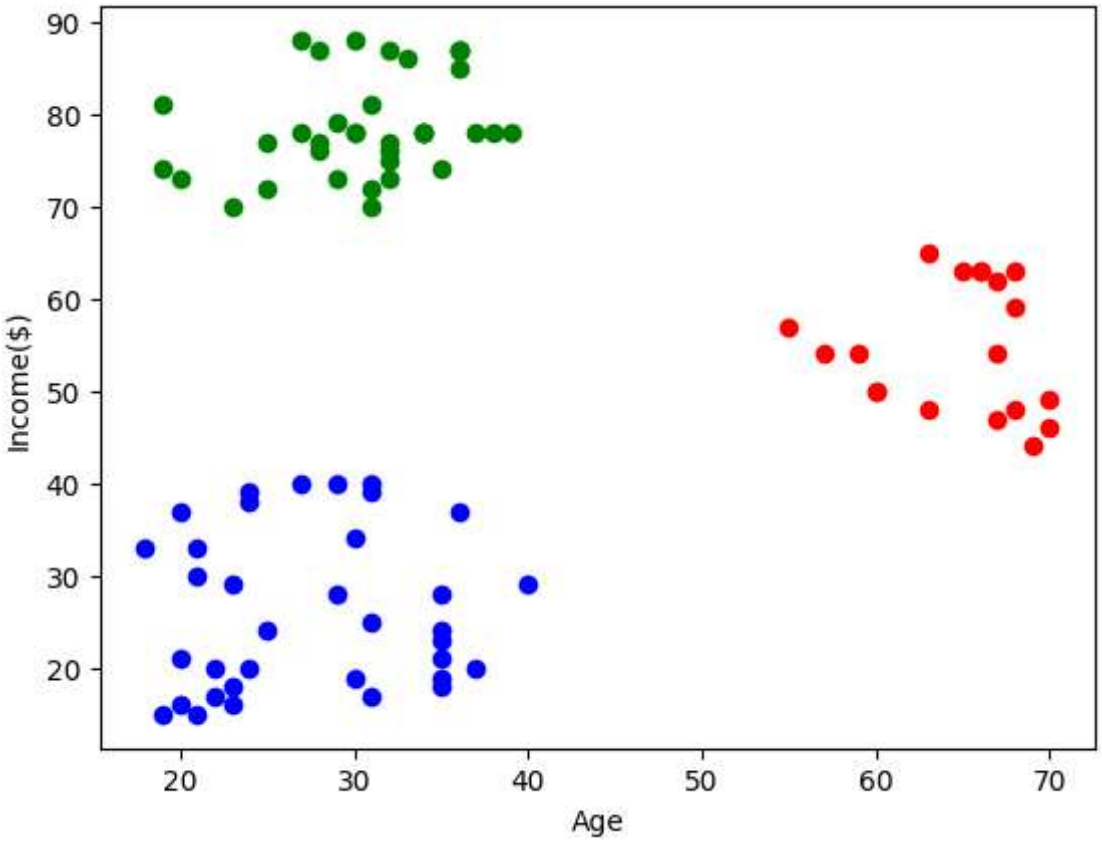
In [11]:
```
df["cluster"]=y_predicted
df.head()
```

Out[11]:

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

In [12]:
```
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="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[12]:
```
Text(0, 0.5, 'Income($)')
```



In [13]:
```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["Income($)"]])
df["Income($)"]=scaler.transform(df[["Income($)"]])
df.head()
```

Out[13]:

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

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

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

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

```
C:\Users\91903\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, 7, 2, 7, 2, 3, 7, 3, 7, 1, 2, 7, 2, 7, 2, 1, 7, 7, 2,
       1, 7, 1, 7, 1, 7, 7, 2, 3, 2, 1, 2, 1, 2, 1, 7, 7, 2, 3, 2, 1, 7,
       1, 2, 1, 7, 7, 7, 1, 7, 7, 3, 1, 1, 1, 3, 4, 1, 3, 4, 3, 1, 3, 4,
       1, 3, 4, 7, 3, 1, 3, 3, 3, 4, 1, 1, 4, 1, 3, 0, 3, 1, 4, 1, 5, 4,
       0, 5, 3, 4, 5, 0, 0, 4, 5, 4, 5, 4, 4, 5, 3, 4, 5, 4, 3, 5, 3, 3,
       3, 4, 0, 4, 4, 4, 3, 5, 5, 5, 4, 0, 0, 0, 4, 0, 5, 0, 5, 0, 5, 0,
       4, 0, 4, 0, 5, 0, 4, 0, 5, 0, 0, 0, 4, 0, 5, 0, 0, 0, 5, 0, 5, 0,
       5, 0, 0, 0, 0, 0, 5, 0, 4, 0, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5, 0,
       5, 0, 5, 0, 6, 6, 5, 6, 6, 6, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
       6, 6])
```

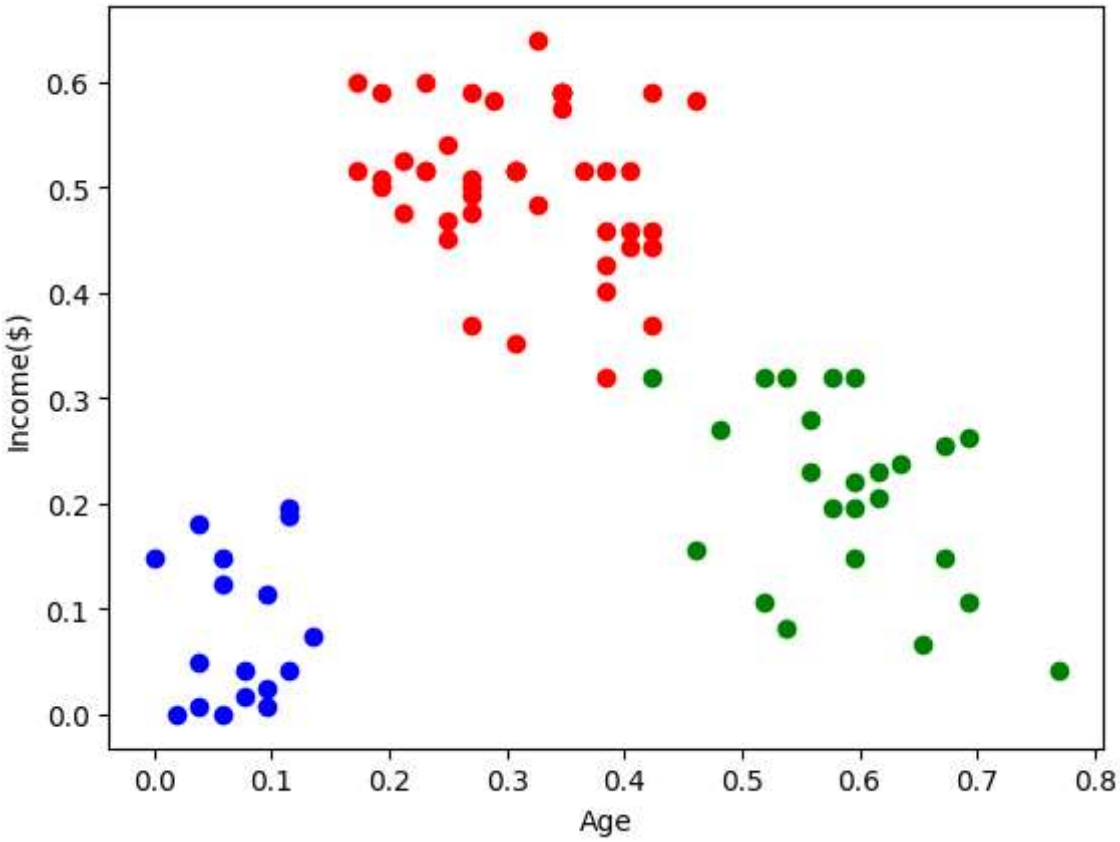
In [17]:
```
df["New Cluster"]=y_predicted
df.head()
```

Out[17]:

	Gender	Age	Income($)	cluster	New Cluster
0	Male	0.019231	0.000000	2	2
1	Male	0.057692	0.000000	2	2
2	Female	0.038462	0.008197	2	2
3	Female	0.096154	0.008197	2	2
4	Female	0.250000	0.016393	2	7

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

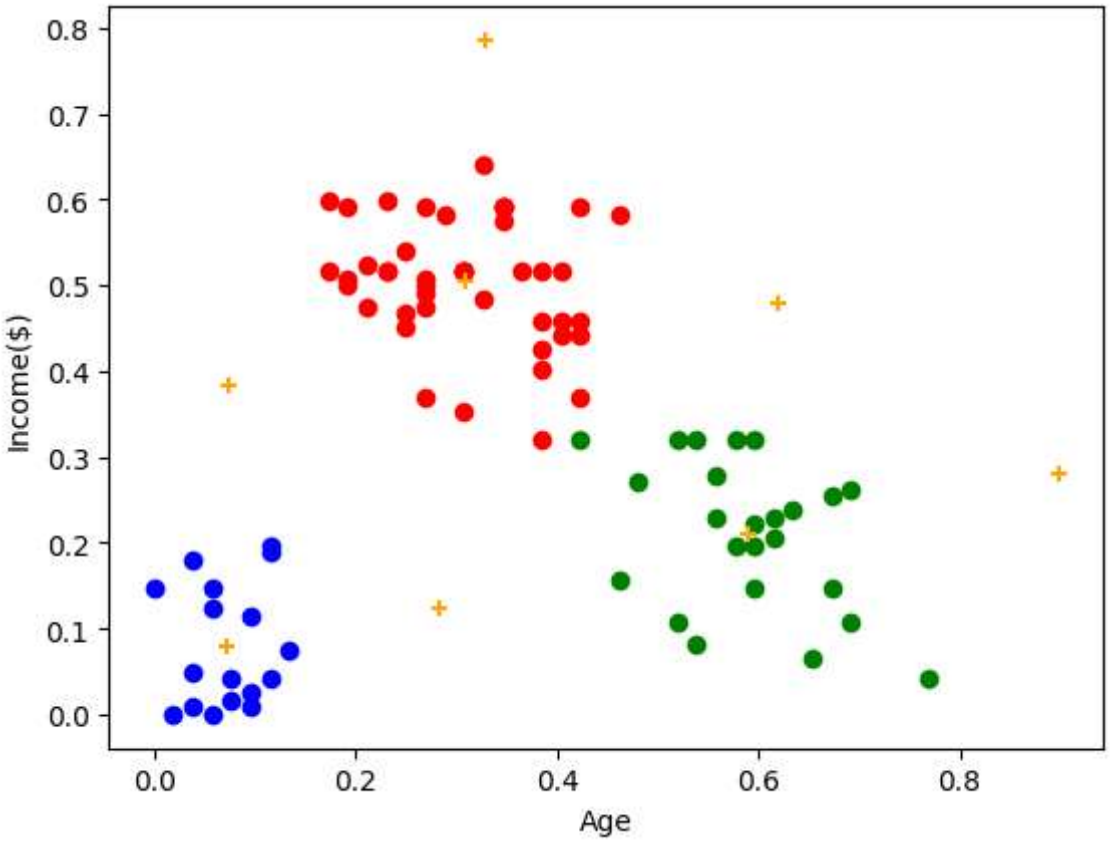


In [19]:
```
km.cluster_centers_
```

Out[19]:
```
array([[0.30944056, 0.50428465],
       [0.58974359, 0.20969945],
       [0.07239819, 0.08003857],
       [0.89799331, 0.28011404],
       [0.07322485, 0.38272383],
       [0.62037037, 0.47996357],
       [0.32905983, 0.78551913],
       [0.28388278, 0.1245121 ]])
```

In [20]:
```
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($)"],color="blue")
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color="orange",m
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[20]:
```
Text(0, 0.5, 'Income($)')
```



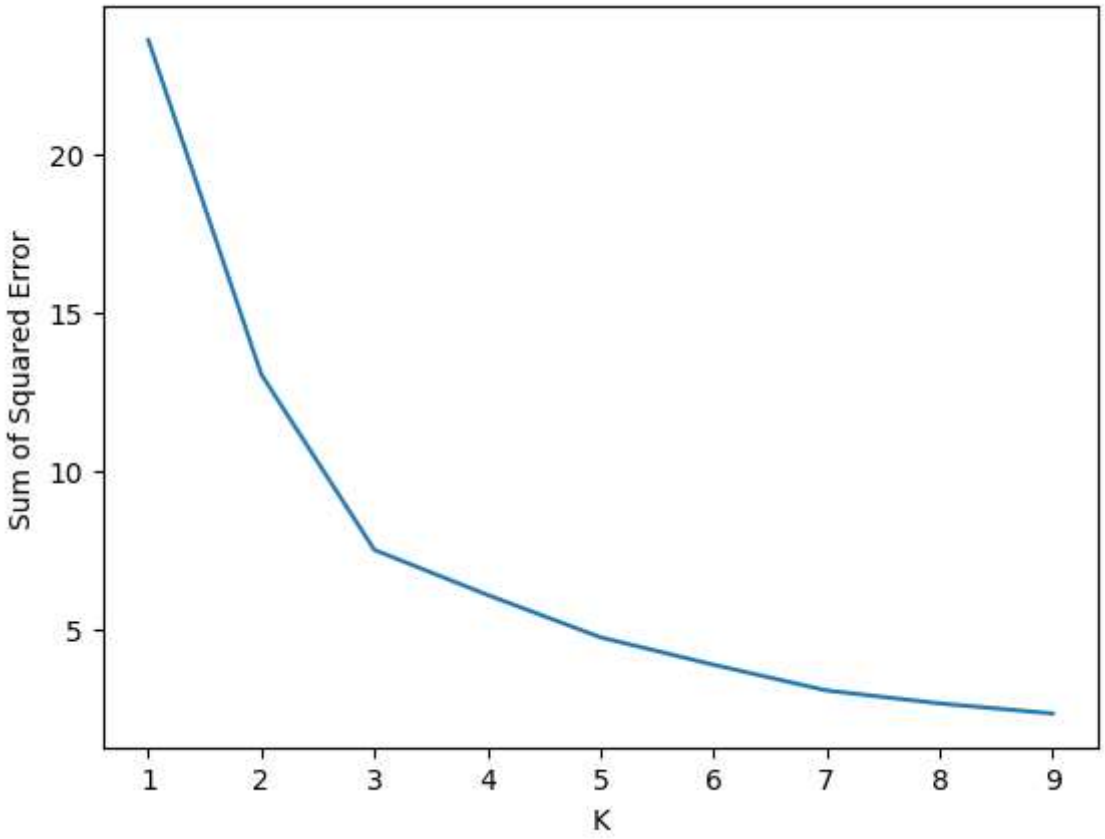
In [21]:
```
k_rng=range(1,10)
sse=[]
```

In [22]:
```
for k in k_rng:
 km=KMeans(n_clusters=k)
 km.fit(df[["Age","Income($)"]])
 sse.append(km.inertia_)
#km.inertia_ will give you the value of sum of square error
print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

```
[23.583906150363603, 13.028938428018286, 7.492107868586012, 6.072884728742554
5, 4.733776701093291, 3.8685816788341514, 3.055986211920202, 2.64865939458645
7, 2.329176554466517]
```

```
C:\Users\91903\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\91903\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\91903\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\91903\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\91903\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\91903\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\91903\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\91903\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\91903\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[22]:
```
Text(0, 0.5, 'Sum of Squared Error')
```



In [ ]: