

In [5]:

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

In [6]:

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

Out[6]:

	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

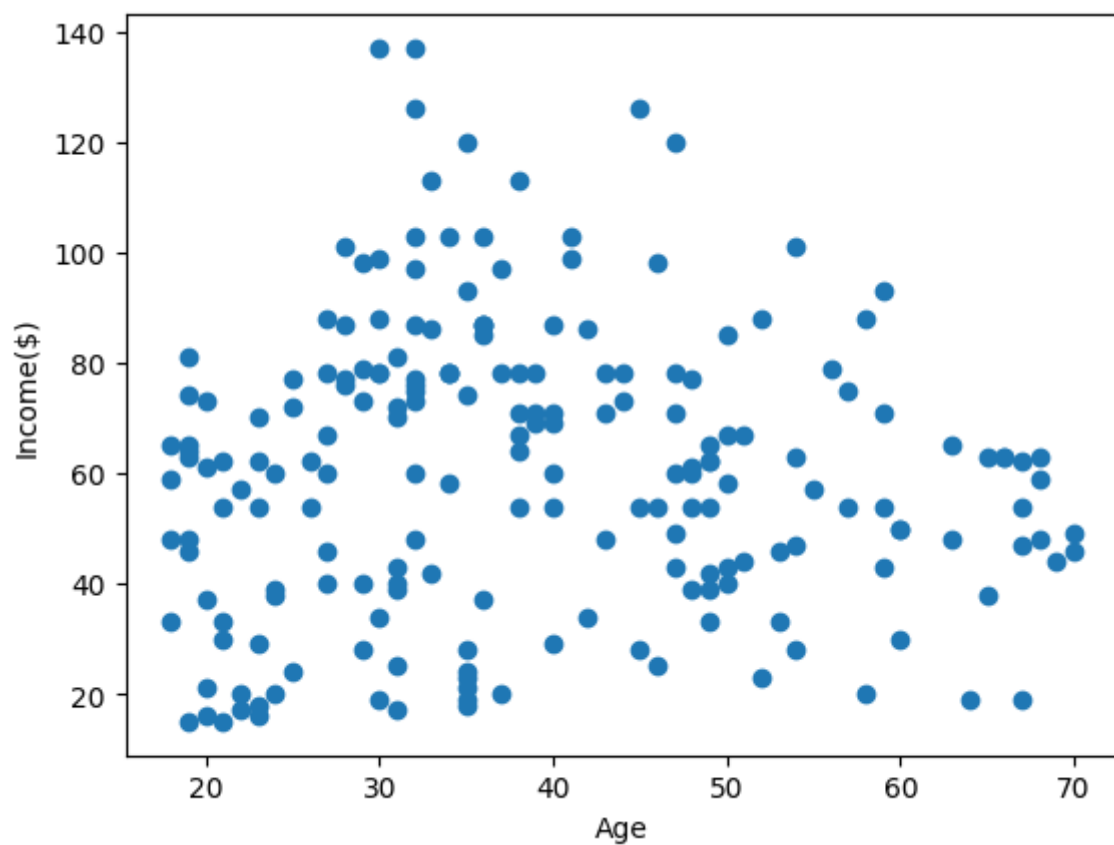
200 rows × 3 columns

In [7]:

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

Out[7]:

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



In [8]:

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

In [9]:

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

Out[9]:

▼ KMeans  
KMeans()

In [10]:

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

C:\Users\DELL\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, 1, 2, 1, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2,
       1, 2, 1, 2, 1, 2, 1, 2, 1, 2, 1, 6, 1, 6, 1, 6, 6, 6, 1, 6, 1, 6,
       1, 6, 1, 6, 6, 6, 1, 6, 6, 1, 1, 1, 1, 7, 6, 1, 7, 6, 7, 1, 7, 6,
       1, 7, 6, 6, 7, 1, 7, 7, 7, 6, 3, 3, 6, 3, 7, 3, 7, 3, 6, 3, 7, 4,
       3, 3, 7, 4, 3, 3, 4, 4, 3, 4, 3, 4, 4, 3, 7, 4, 3, 4, 7, 3, 7, 7,
       7, 4, 3, 4, 4, 4, 7, 3, 3, 3, 4, 3, 3, 3, 4, 4, 3, 3, 3, 3, 3, 3,
       4, 4, 4, 4, 3, 4, 4, 4, 3, 4, 4, 4, 4, 4, 3, 4, 4, 4, 3, 3, 3, 4,
       3, 4, 4, 4, 4, 4, 3, 4, 4, 4, 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 [11]:

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

Out[11]:

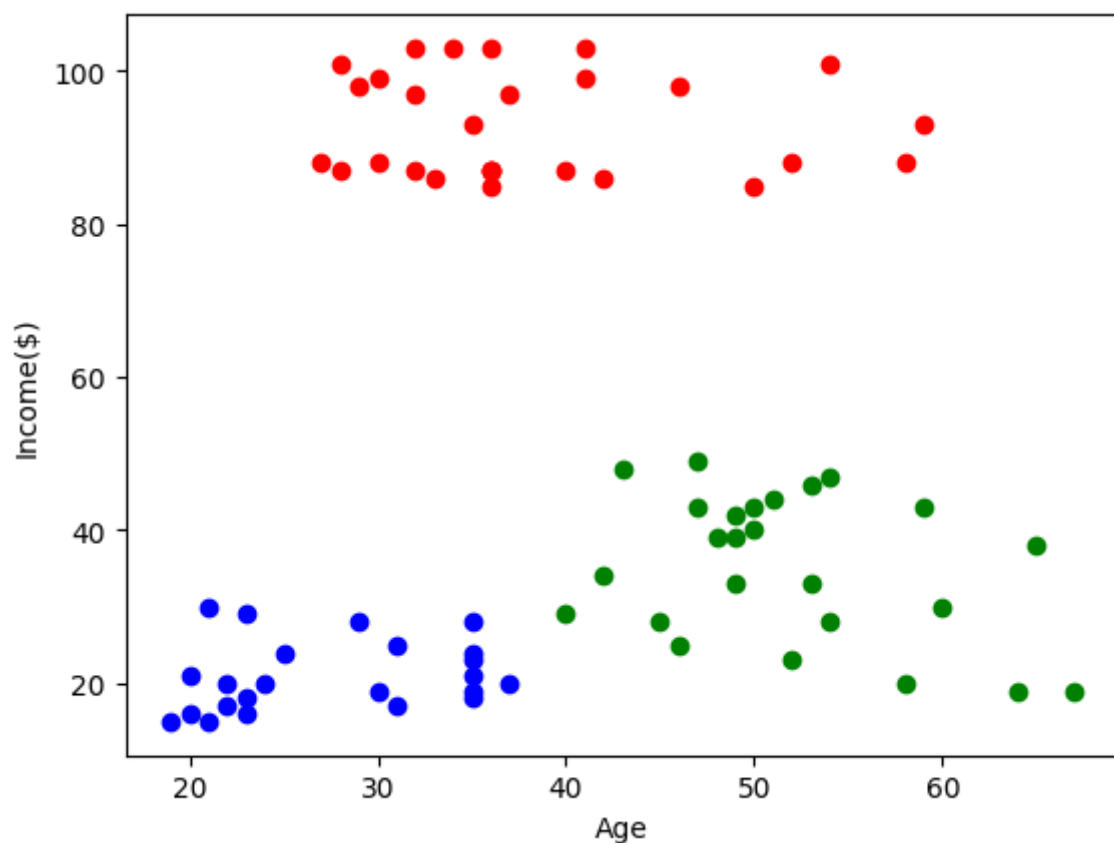
	Gender	Age	Income(\$)	cluster
0	Male	19	15	2
1	Male	21	15	2
2	Female	20	16	2
3	Female	23	16	2
4	Female	31	17	2

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
```

In [14]:

```
scaler=MinMaxScaler()
```

In [15]:



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

Out[15]:

	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 [16]:



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

Out[16]:

	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 [17]:



```
km=KMeans()
```

In [18]:

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

C:\Users\DELL\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[18]:

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

In [19]:

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

Out[19]:

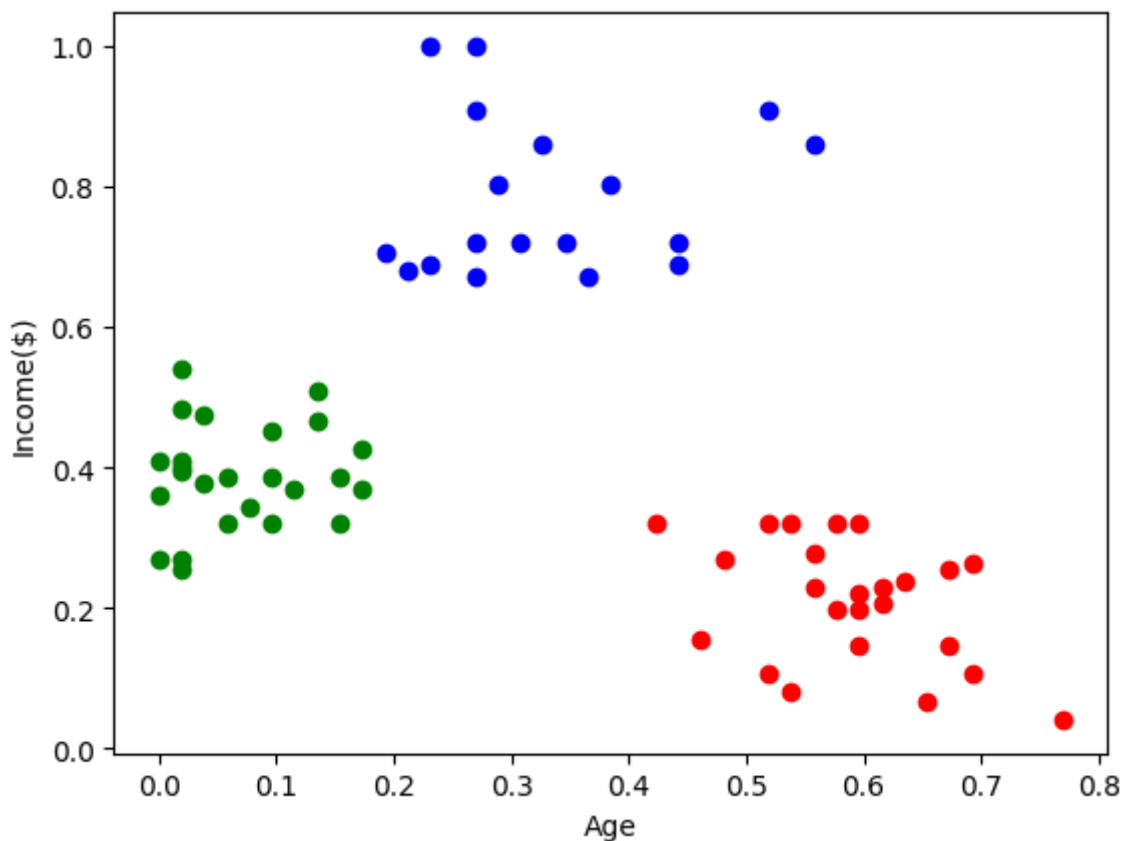
	Gender	Age	Income(\$)	cluster	New Cluster
0	Male	0.019231	0.000000	2	4
1	Male	0.057692	0.000000	2	4
2	Female	0.038462	0.008197	2	4
3	Female	0.096154	0.008197	2	4
4	Female	0.250000	0.016393	2	6

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($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[20]:

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



In [21]:

km.cluster\_centers\_

Out[21]:

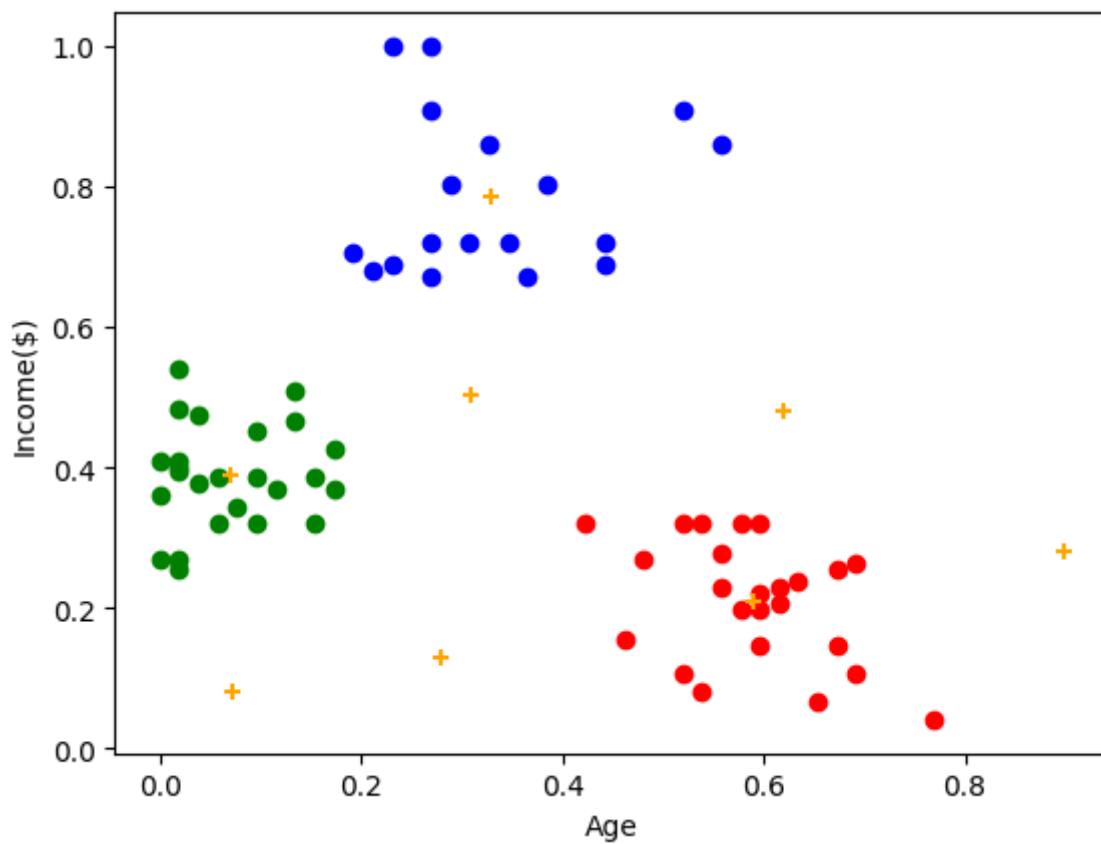
```
array([[0.58974359, 0.20969945],
       [0.06923077, 0.38786885],
       [0.32905983, 0.78551913],
       [0.30944056, 0.50428465],
       [0.07239819, 0.08003857],
       [0.89799331, 0.28011404],
       [0.27884615, 0.13040238],
       [0.62037037, 0.47996357]])
```

In [22]:

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

Out[22]:

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



In [23]:

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



In [24]:



```

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 errorprint(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")

```

C:\Users\DELL\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(
```

C:\Users\DELL\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(
```

C:\Users\DELL\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(
```

C:\Users\DELL\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(
```

C:\Users\DELL\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(
```

C:\Users\DELL\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(
```

C:\Users\DELL\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(
```

C:\Users\DELL\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

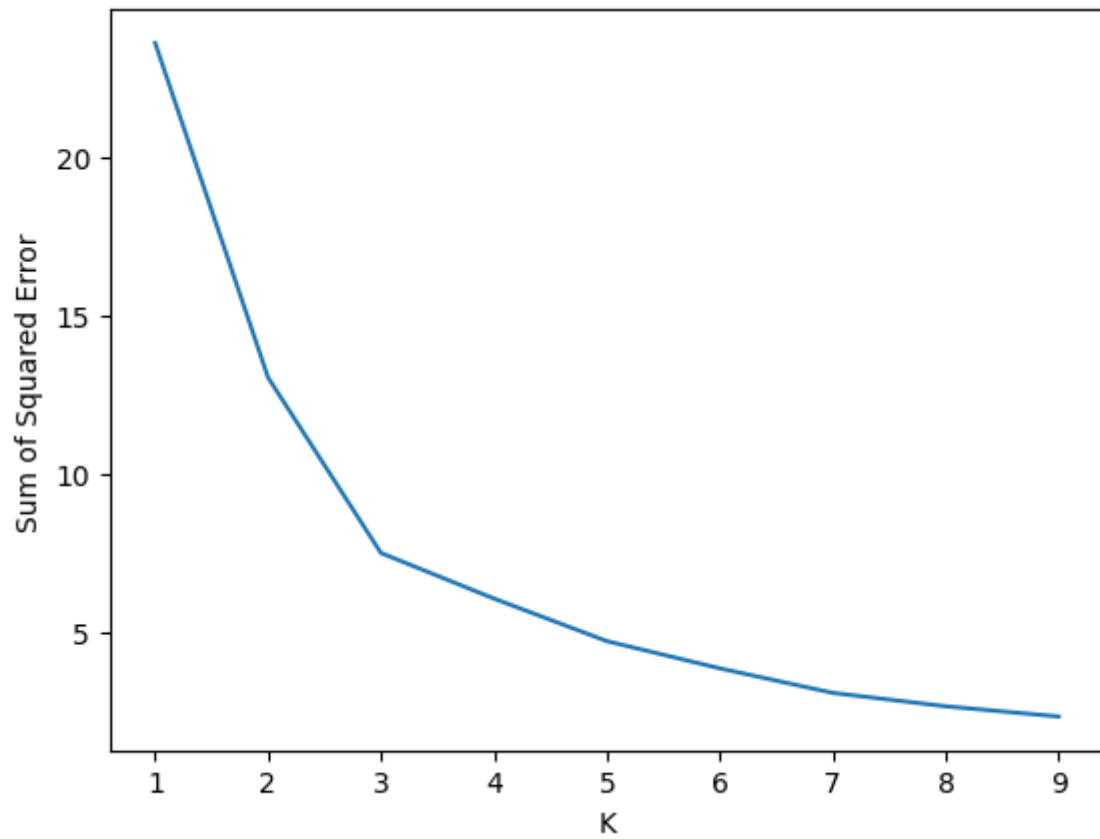
```
warnings.warn(
```

C:\Users\DELL\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[24]:

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



In [ ]:

