

In [1]:

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

In [2]:

```
df=pd.read_csv(r"C:\Users\poorn\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.isnull().sum()
```

Out[3]:

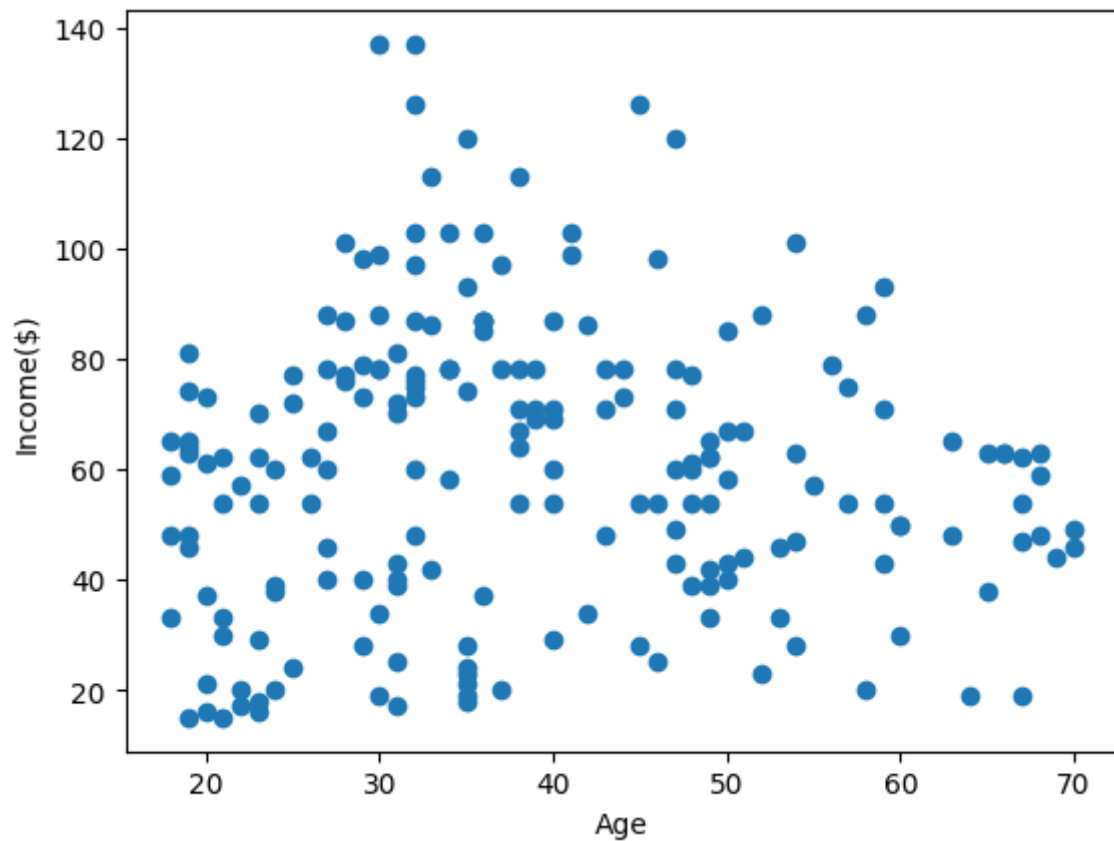
```
Gender      0
Age          0
Income($)   0
dtype: int64
```

In [4]:

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

Out[4]:

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



In [9]:

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

In [10]:

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

Out[10]:

```
▼ KMeans  
KMeans()
```

In [11]:

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

C:\Users\poorn\AppData\Local\Programs\Python\Python311\Lib\site-packages\s  
klearn\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` explicit  
ly to suppress the warning  
warnings.warn(

Out[11]:

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

In [12]:

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

Out[12]:

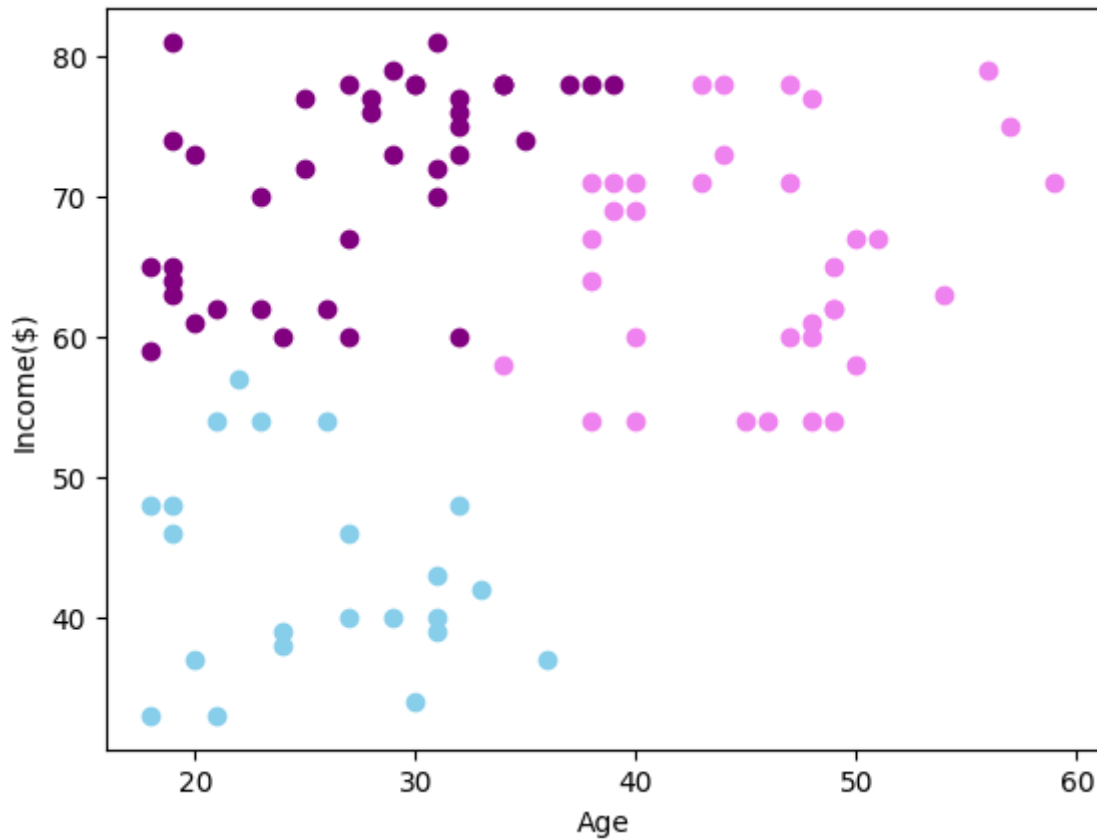
	Gender	Age	Income(\$)	cluster
0	Male	19	15	5
1	Male	21	15	5
2	Female	20	16	5
3	Female	23	16	5
4	Female	31	17	5

In [13]:

```
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[13]:

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



In [14]:

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

In [15]:

```
Scaler=MinMaxScaler()
```

In [16]:

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

Out[16]:

	Gender	Age	Income(\$)	cluster
0	Male	19	0.000000	5
1	Male	21	0.000000	5
2	Female	20	0.008197	5
3	Female	23	0.008197	5
4	Female	31	0.016393	5

In [17]:

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

Out[17]:

	Gender	Age	Income(\$)	cluster
0	Male	0.019231	0.000000	5
1	Male	0.057692	0.000000	5
2	Female	0.038462	0.008197	5
3	Female	0.096154	0.008197	5
4	Female	0.250000	0.016393	5

In [18]:

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

Out[18]:

▼ KMeans

KMeans()

In [19]:

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

```
C:\Users\poorn\AppData\Local\Programs\Python\Python311\Lib\site-packages\s
klearn\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` explicit
ly to suppress the warning
    warnings.warn(
```

Out[19]:

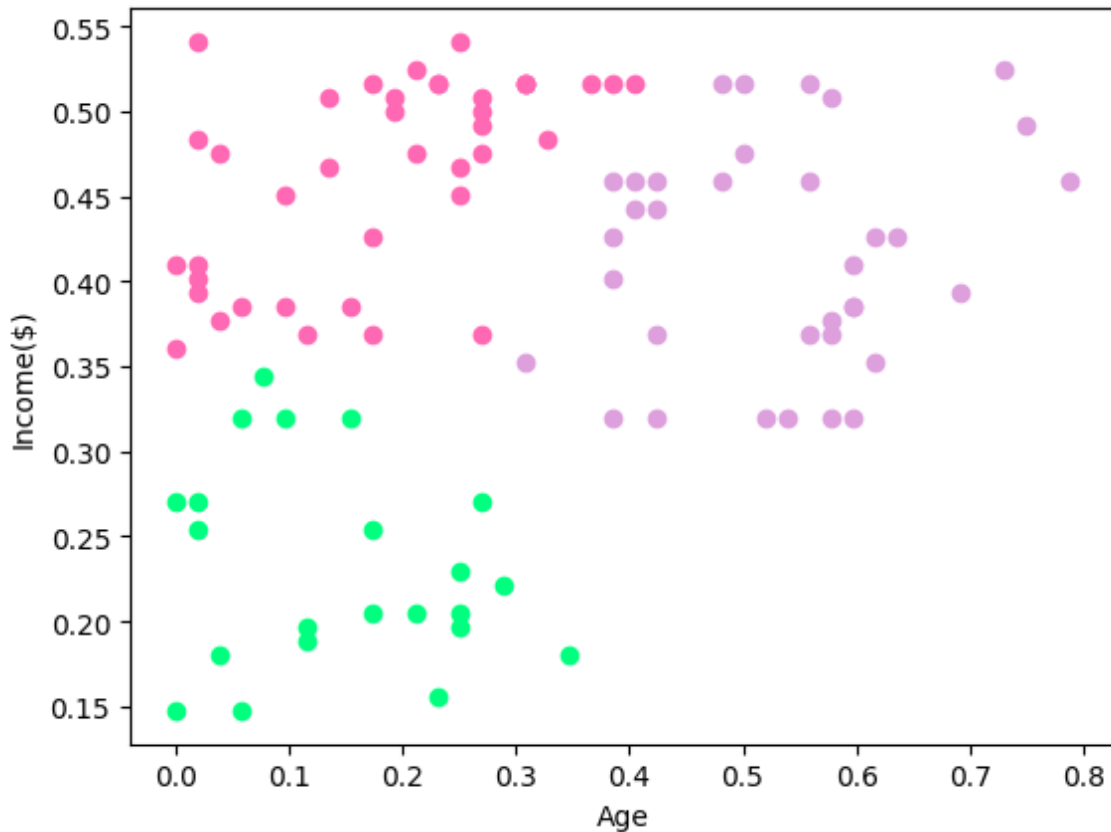
```
array([5, 5, 5, 5, 7, 5, 7, 5, 4, 7, 4, 7, 3, 5, 7, 5, 7, 5, 3, 7, 7, 5,
       3, 7, 3, 7, 3, 7, 7, 5, 4, 5, 3, 5, 3, 5, 3, 7, 7, 5, 4, 5, 3, 7,
       3, 5, 3, 7, 7, 7, 3, 7, 7, 4, 3, 3, 3, 4, 2, 3, 4, 2, 4, 3, 4, 2,
       3, 4, 2, 7, 4, 3, 4, 4, 4, 2, 3, 3, 2, 3, 4, 6, 4, 3, 2, 3, 0, 2,
    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

In [20]:

```
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[20]:

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



In [21]:

```
km.cluster_centers_
```

Out[21]:

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

In [27]:

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

Cell In[27], line 7

```
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color="red",marks)
```

^

**SyntaxError:** positional argument follows keyword argument



In [24]:

```
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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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]:

```
[23.583906150363603,  
13.028938428018286,  
7.493024843304991,  
6.0728847287425545,  
4.713416604872824,  
3.8651257592912613,  
3.055986211920202,  
2.646037617631439,  
2.3374563204198613]
```

In [25]:

```
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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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\poorn\AppData\Local\Programs\Python\Python311\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(

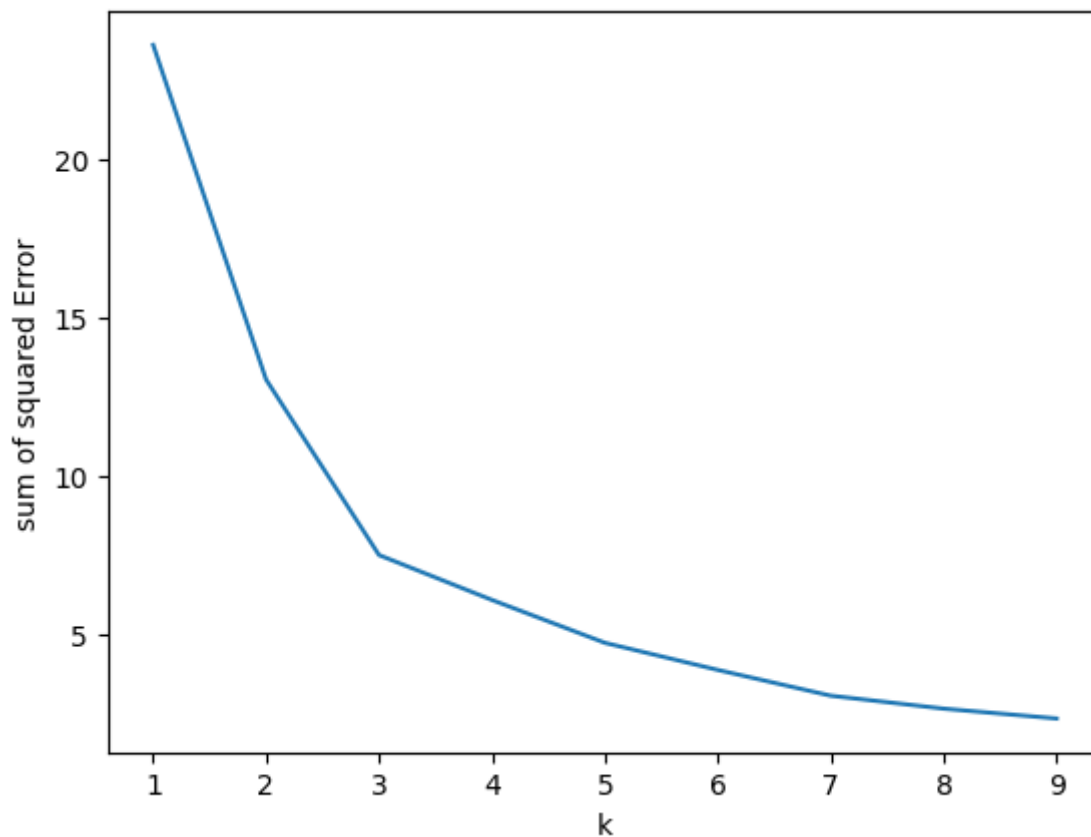
Out[25]:

```
[23.583906150363603,  
13.028938428018286,  
7.492113413237458,  
6.075958471031436,  
4.722729718973683,  
3.8675723784151086,  
3.054717436369358,
```

```
2.6460609774305146,  
2.33533953687878]  
plt.plot(k, rns, sse)  
plt.xlabel("k")  
plt.ylabel("sum of squared Error")
```

Out[26]:

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



In [ ]: