

In [1]:

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

In [2]:

```
df=pd.read_csv(r"C:\Users\samit\OneDrive\Desktop\jupyter\Income.csv")
df.head()
```

Out[2]:

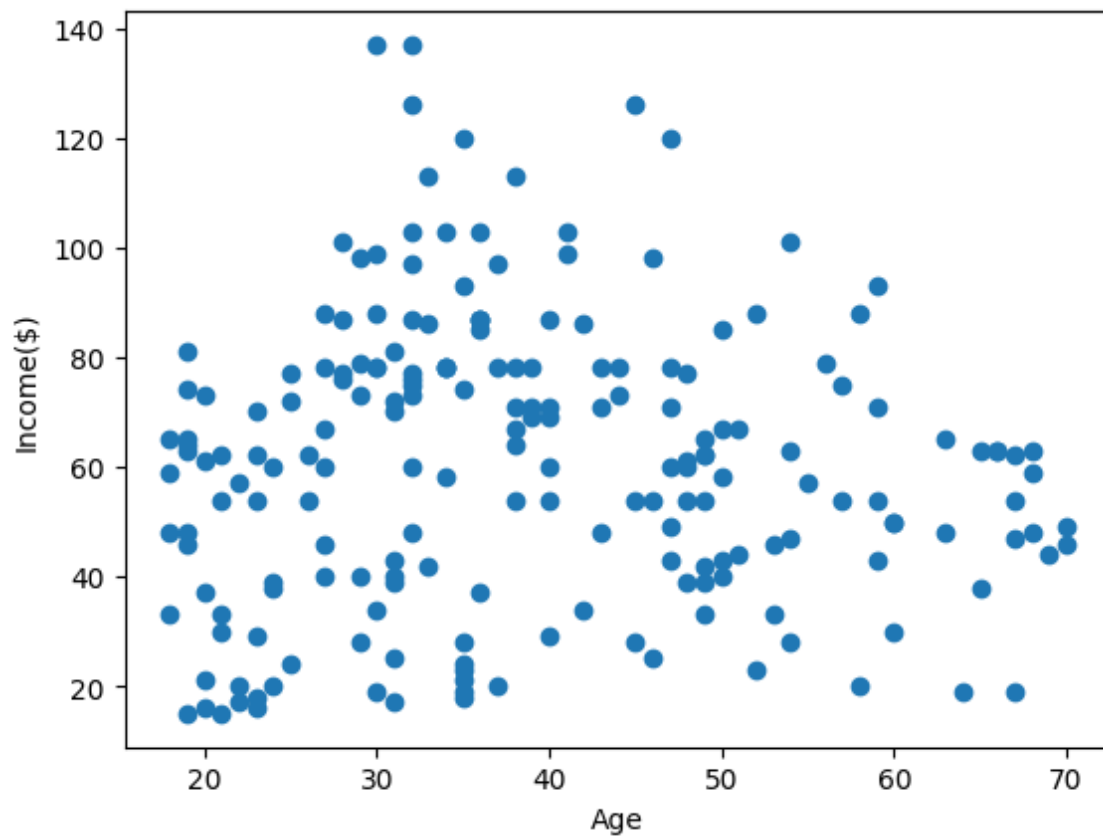
	Gender	Age	Income(\$)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17

In [3]:

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

Out[3]:

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



In [4]:

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

In [5]:

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

Out[5]:

▼ KMeans  
KMeans()

In [6]:

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

C:\Users\samit\AppData\Local\Programs\Python\Python311\Lib\site-packages  
 \sklearn\cluster\\_kmeans.py:870: FutureWarning: The default value of `n\_i  
 nit` will change from 10 to 'auto' in 1.4. Set the value of `n\_init` expl  
 icitly to suppress the warning  
 warnings.warn(

Out[6]:

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

In [7]:

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

Out[7]:

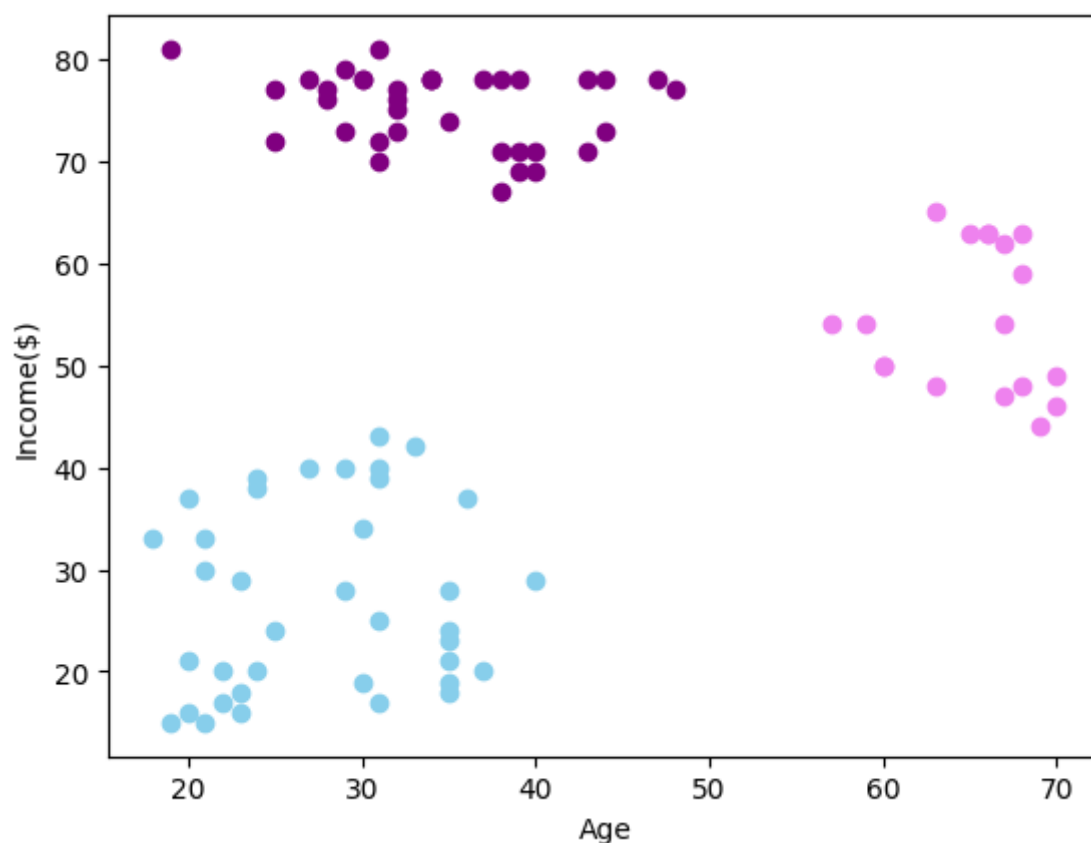
	Gender	Age	Income(\$)	cluster
0	Male	19	15	1
1	Male	21	15	1
2	Female	20	16	1
3	Female	23	16	1
4	Female	31	17	1

In [8]:

```
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[8]:

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



In [9]:

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

In [10]:

```
Scaler=MinMaxScaler()
```

In [11]:

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

Out[11]:

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

In [12]:

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

Out[12]:

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

In [13]:

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

Out[13]:

▼ KMeans

KMeans()

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

Out[14]:

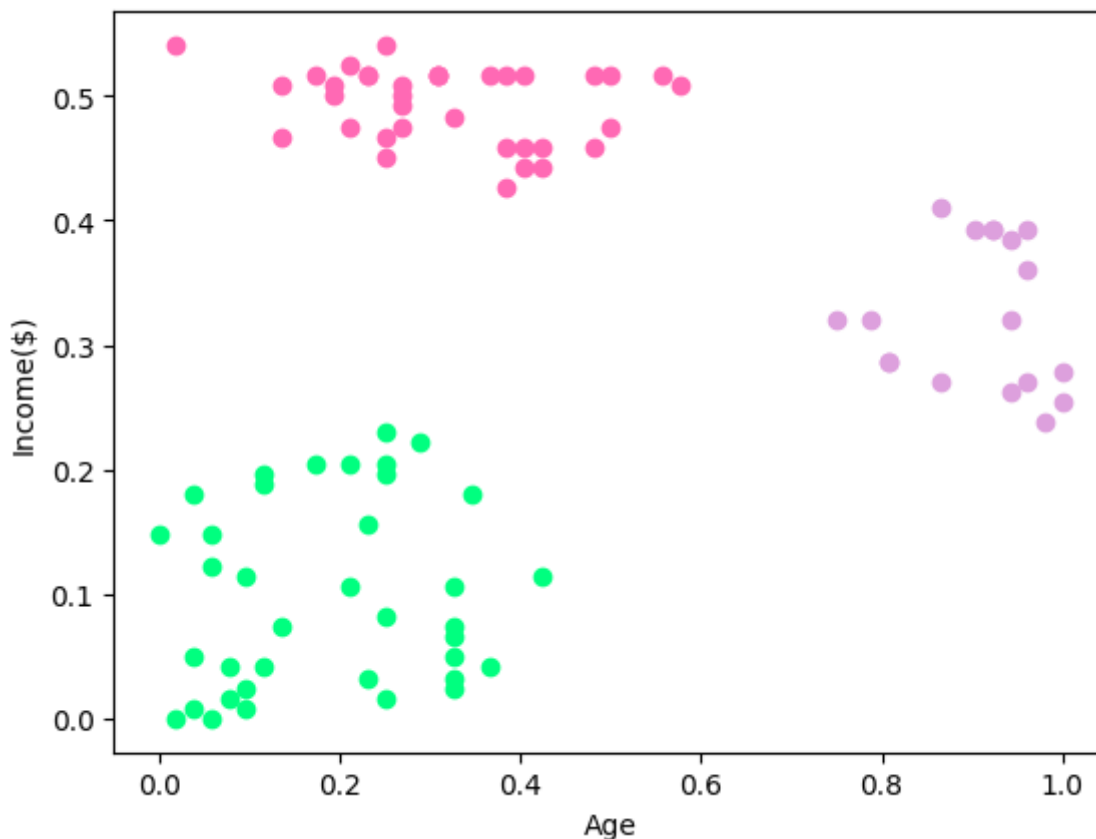
```
array([3, 3, 3, 3, 0, 3, 0, 3, 5, 0, 5, 0, 2, 3, 0, 3, 0, 3, 2, 0, 0, 3,
       2, 0, 2, 0, 2, 0, 0, 3, 5, 3, 2, 3, 2, 3, 2, 0, 0, 3, 5, 3, 2, 0,
       2, 3, 2, 0, 0, 0, 2, 0, 0, 5, 2, 2, 2, 5, 6, 2, 5, 6, 5, 2, 5, 6,
       2, 5, 6, 0, 5, 2, 5, 5, 5, 6, 2, 2, 6, 2, 5, 4, 5, 2, 6, 2, 2, 6,
    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

In [15]:

```
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[15]:

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



In [16]:

```
km.cluster_centers_
```

Out[16]:

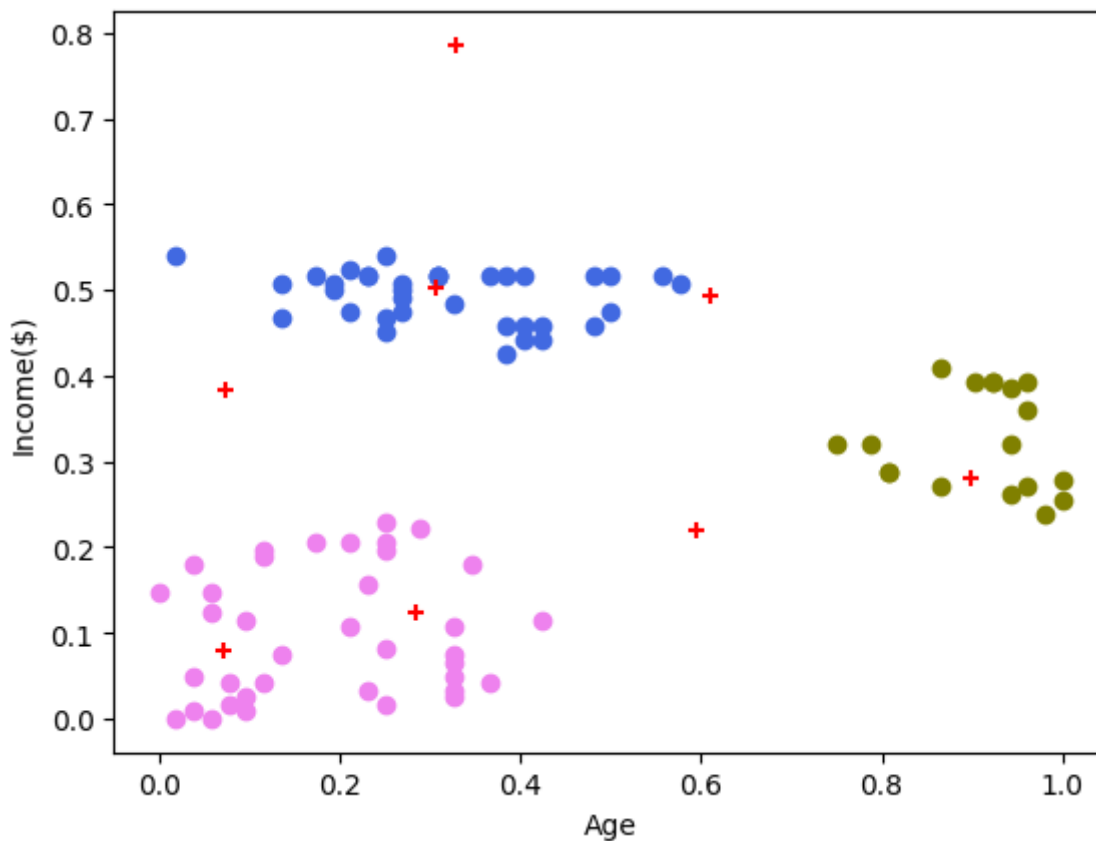
```
array([[0.28388278, 0.1245121 ],
       [0.61094675, 0.49401009],
       [0.5954142 , 0.2203657 ],
       [0.07239819, 0.08003857],
       [0.3059034 , 0.50247808],
       [0.89799331, 0.28011404],
       [0.07322485, 0.38272383],
       [0.32905983, 0.78551913]])
```

In [18]:

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

Out[18]:

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





In [19]:

```

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\samit\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\samit\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\samit\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\samit\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\samit\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\samit\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\samit\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\samit\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\samit\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(
```

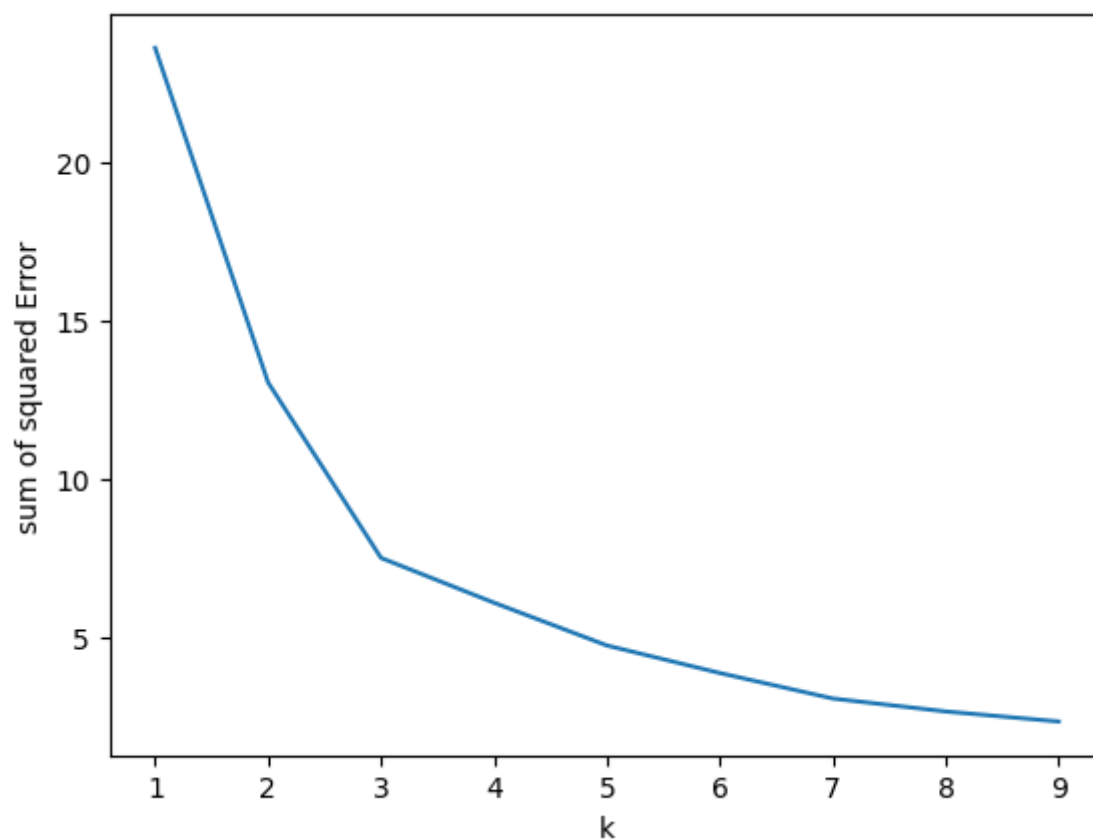
Out[19]:

```
[23.583906150363603,  
 13.02893842801829,  
 7.492113413237459,  
 6.079203145994377,  
 4.729046373331373,  
 3.85068122608658,  
 3.054717436369359,  
 2.616065774805146,  
 2.3291765544605166]
```

```
plt.xlabel("k")  
plt.ylabel("sum of squared Error")
```

Out[20]:

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



In [ ]: