In [1]:

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

In [2]:

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

Out[2]:

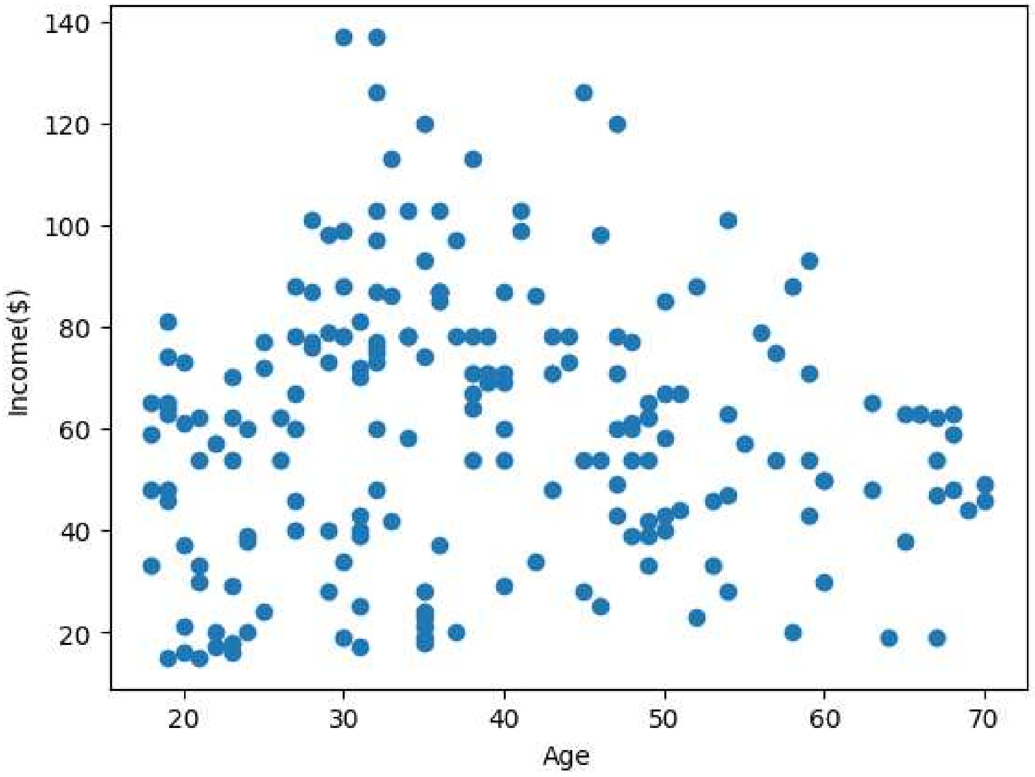
```
     Gender  Age  Income($)
  0    Male   19         15
  1    Male   21         15
  2  Female   20         16
  3  Female   23         16
  4  Female   31         17
...     ...  ...        ...
195  Female   35        120
196  Female   45        126
197    Male   32        126
198    Male   32        137
199    Male   30        137

200 rows × 3 columns
```

In [3]:

```
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\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```

Out[6]:

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

In [7]:

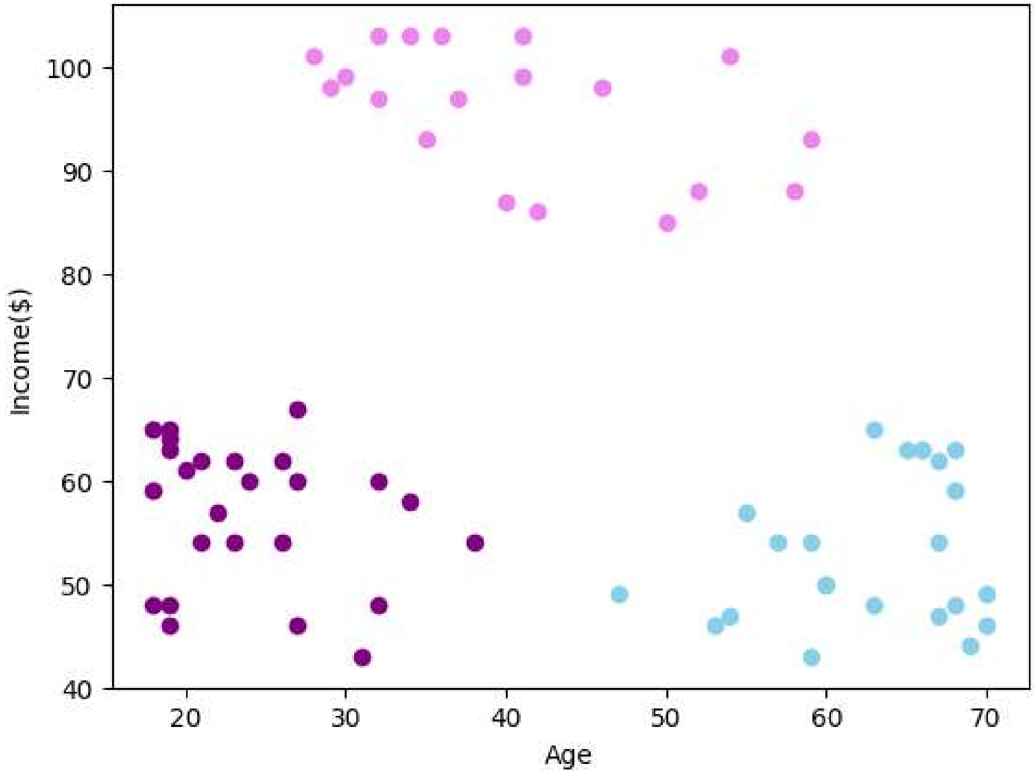
```
df["cluster"]=y_predicted
df.head()
```

Out[7]:

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

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	3
1	Male	21	0.000000	3
2	Female	20	0.008197	3
3	Female	23	0.008197	3
4	Female	31	0.016393	3

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

In [13]:

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

Out[13]:

```
▾ KMeans
KMeans()
```

In [14]:

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

```
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```

Out[14]:

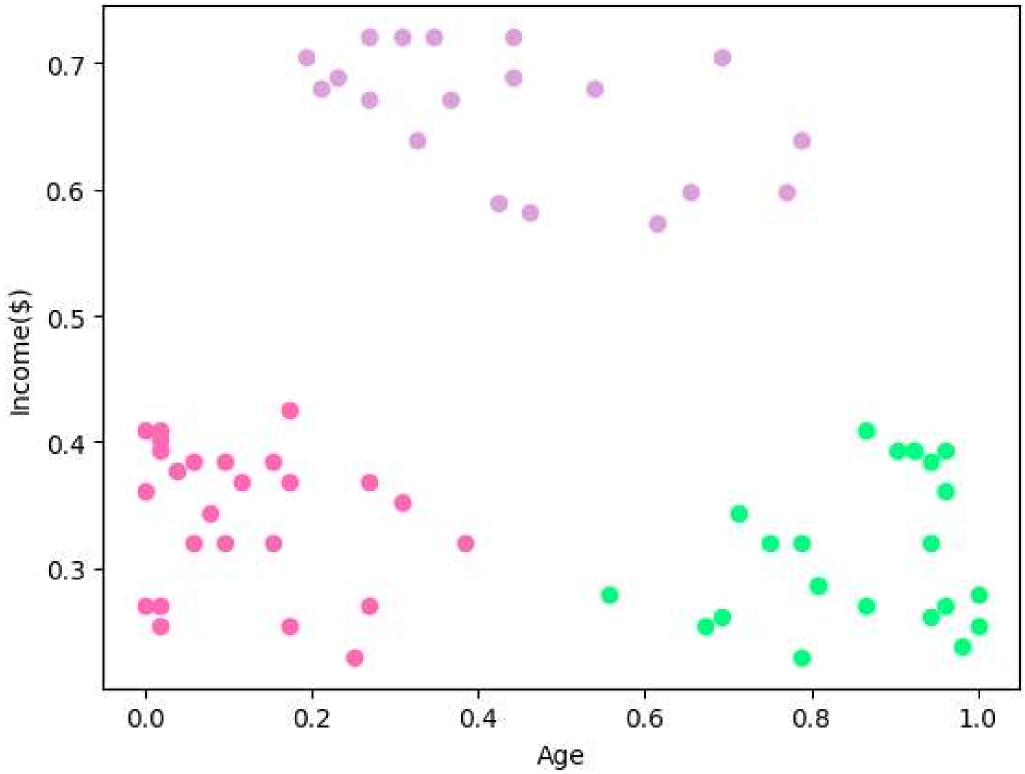
```
array([6, 6, 6, 6, 4, 6, 4, 6, 3, 4, 3, 4, 3, 6, 4, 6, 4, 6, 2, 4, 4, 6,
       2, 4, 2, 4, 2, 4, 4, 6, 3, 6, 2, 6, 2, 6, 2, 4, 4, 6, 3, 6, 2, 4,
       2, 6, 2, 4, 4, 4, 2, 4, 4, 3, 2, 2, 2, 3, 1, 2, 3, 1, 3, 2, 3, 1,
       2, 3, 1, 4, 3, 2, 3, 3, 3, 1, 2, 2, 1, 2, 3, 5, 3, 2, 1, 2, 2, 1,
    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

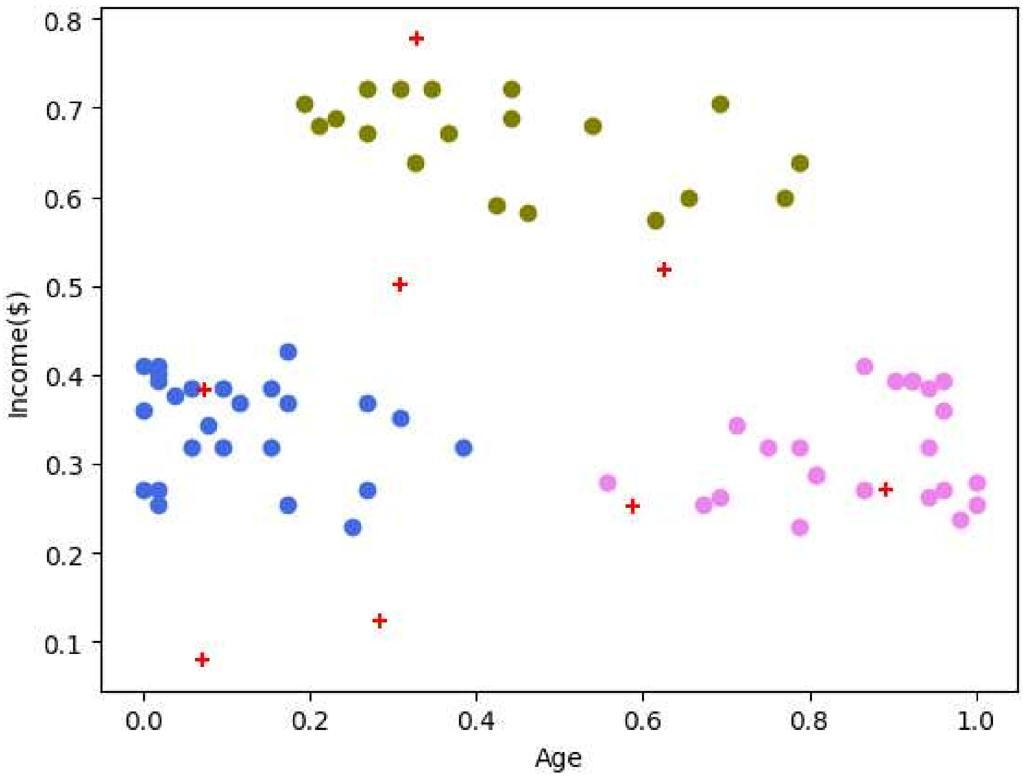
```
array([[0.62596154, 0.51885246],
       [0.07322485, 0.38272383],
       [0.58717949, 0.25245902],
       [0.89262821, 0.27015027],
       [0.28388278, 0.1245121 ],
       [0.30903399, 0.50114373],
       [0.07239819, 0.08003857],
       [0.32894737, 0.77782571]])
```

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\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
C:\Users\Niranjan\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\clu
ster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 t
o 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  warnings.warn(
```

Out[19]:

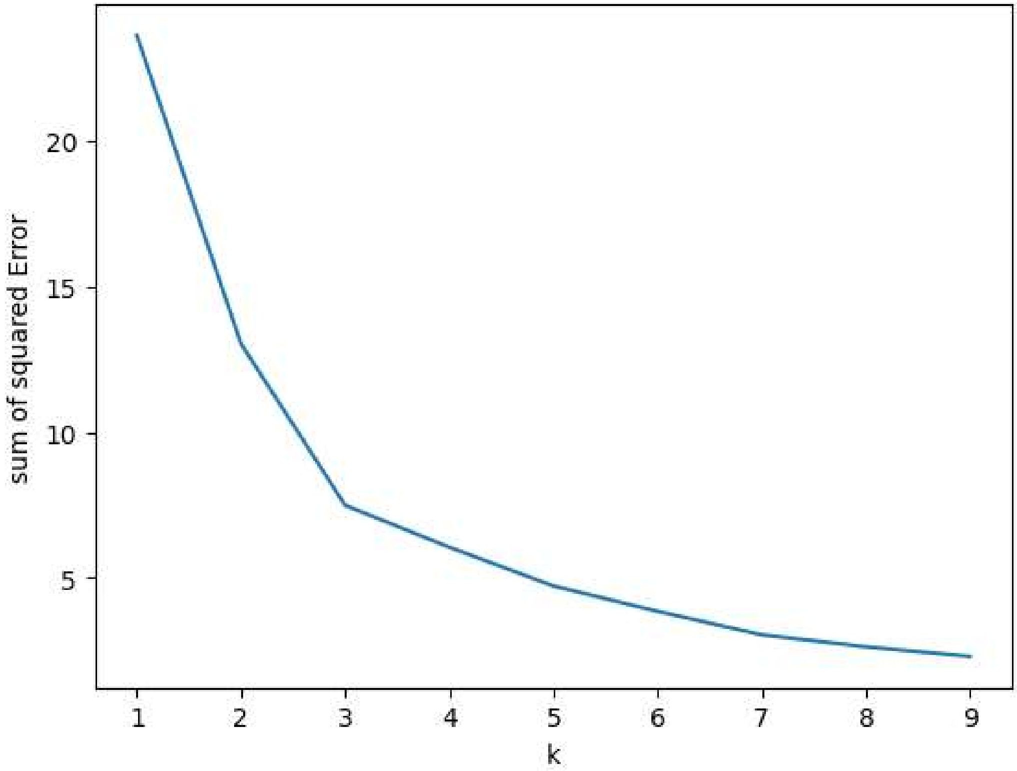
```
[23.583906150363603,
 13.028938428018286,
 7.492113413237458,
 6.0556979561778626,
 4.729046373331373,
 3.859108736888713,
 3.054717436369358,
 2.645887374044777,
 2.3135720353543285]
```

In [20]:

```
plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("sum of squared Error")
```

Out[20]:

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



In [ ]:

```

```