In [1]:

```
from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
%matplotlib inline
```

In [2]:

```
df = pd.read_csv("income.csv")
df.head()
```

Out[2]:

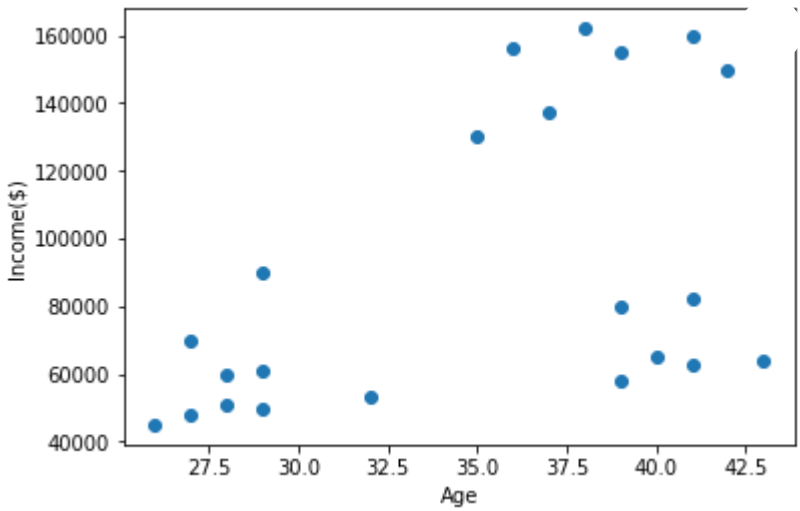
| | Name | Age | Income($) |
|---|---|---|---|
| **0** | Rob | 27 | 70000 |
| **1** | Michael | 29 | 90000 |
| **2** | Mohan | 29 | 61000 |
| **3** | Ismail | 28 | 60000 |
| **4** | Kory | 42 | 150000 |

In [3]:

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

Out[3]:

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



In [4]:

```
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Age','Income($)']])
y_predicted
```

Out[4]:

```
array([1, 1, 2, 2, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 2, 1, 1, 2])
```

In [5]:

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

Out[5]:

| | Name | Age | Income($) | cluster |
|---|---|---|---|---|
| **0** | Rob | 27 | 70000 | 1 |
| **1** | Michael | 29 | 90000 | 1 |
| **2** | Mohan | 29 | 61000 | 2 |
| **3** | Ismail | 28 | 60000 | 2 |
| **4** | Kory | 42 | 150000 | 0 |

In [6]:

```
km.cluster_centers_
```

Out[6]:

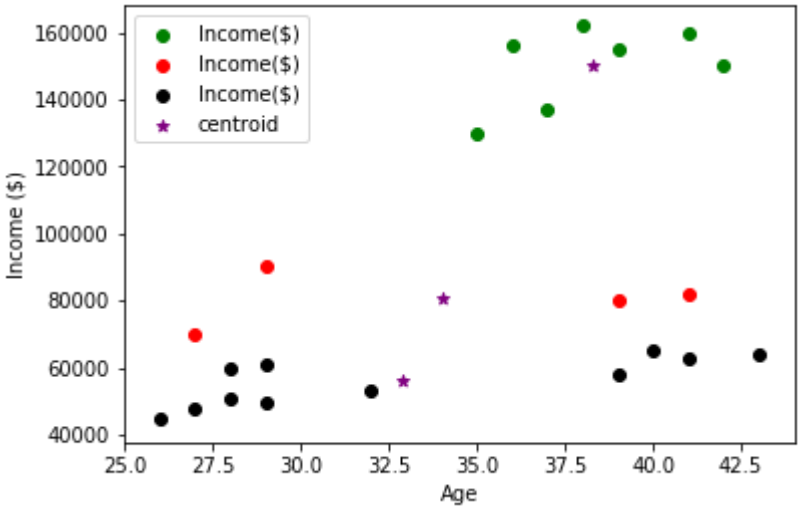
```
array([[3.82857143e+01, 1.50000000e+05],
       [3.40000000e+01, 8.05000000e+04],
       [3.29090909e+01, 5.61363636e+04]])
```

In [7]:

```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1.Age,df1['Income($)'],color='green')
plt.scatter(df2.Age,df2['Income($)'],color='red')
plt.scatter(df3.Age,df3['Income($)'],color='black')
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color='purple',marker='*',lab
plt.xlabel('Age')
plt.ylabel('Income ($)')
plt.legend()
```

Out[7]:

```
<matplotlib.legend.Legend at 0x2c1471ddc88>
```




In [9]:

```
# Preprocessing using min max scaler
scaler = MinMaxScaler()

scaler.fit(df[['Income($)']])
df['Income($)'] = scaler.transform(df[['Income($)']])

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

In [10]:

```
df.head()
```

Out[10]:

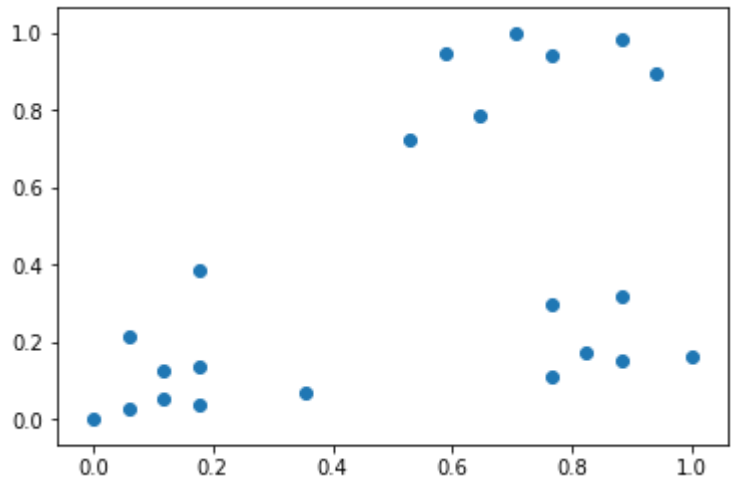
| | Name | Age | Income($) | cluster |
|---|---|---|---|---|
| **0** | Rob | 0.058824 | 0.213675 | 1 |
| **1** | Michael | 0.176471 | 0.384615 | 1 |
| **2** | Mohan | 0.176471 | 0.136752 | 2 |
| **3** | Ismail | 0.117647 | 0.128205 | 2 |
| **4** | Kory | 0.941176 | 0.897436 | 0 |

In [11]:

```
plt.scatter(df.Age,df['Income($)'])
```

Out[11]:

```
<matplotlib.collections.PathCollection at 0x2c14732ada0>
```




In [12]:

```
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Age','Income($)']])
y_predicted
```

Out[12]:

```
array([0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2])
```

In [13]:

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

Out[13]:

| | Name | Age | Income($) | cluster |
|---|---|---|---|---|
| **0** | Rob | 0.058824 | 0.213675 | 0 |
| **1** | Michael | 0.176471 | 0.384615 | 0 |
| **2** | Mohan | 0.176471 | 0.136752 | 0 |
| **3** | Ismail | 0.117647 | 0.128205 | 0 |
| **4** | Kory | 0.941176 | 0.897436 | 1 |

In [14]:

```
km.cluster_centers_
```

Out[14]:

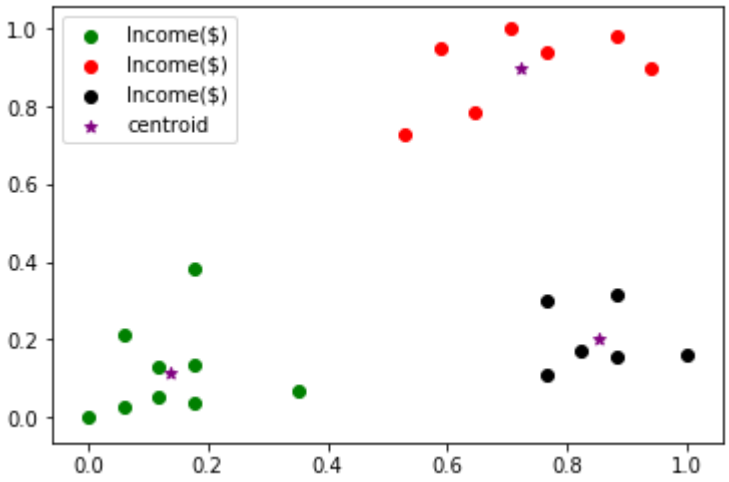
```
array([[0.1372549 , 0.11633428],
       [0.72268908, 0.8974359 ],
       [0.85294118, 0.2022792 ]])
```

In [15]:

```
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1.Age,df1['Income($)'],color='green')
plt.scatter(df2.Age,df2['Income($)'],color='red')
plt.scatter(df3.Age,df3['Income($)'],color='black')
plt.scatter(km.cluster_centers_[:,0],km.cluster_centers_[:,1],color='purple',marker='*',lab
plt.legend()
```

Out[15]:

```
<matplotlib.legend.Legend at 0x2c14738db00>
```




In [16]:

```
sse = []
k_rng = range(1,10)
for k in k_rng:
    km = KMeans(n_clusters=k)
    km.fit(df[['Age','Income($)']])
    sse.append(km.inertia_)
```

In [18]:

```
sse
```

Out[18]:

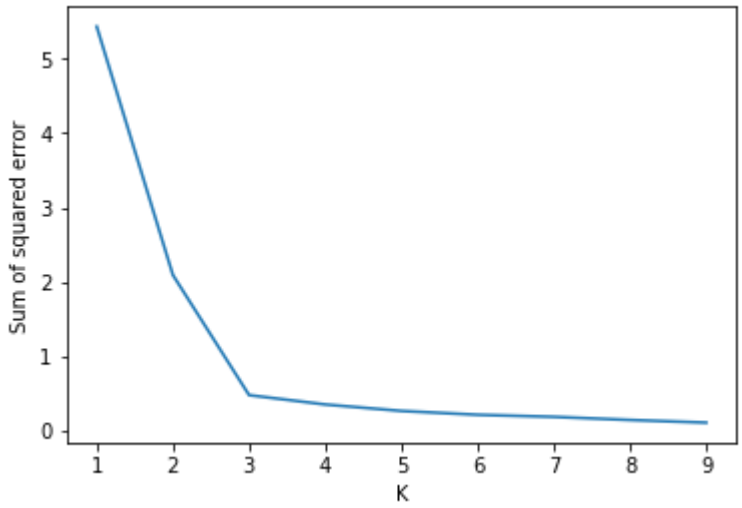
```
[5.434011511988176,
 2.0911363886990766,
 0.4750783498553095,
 0.3491047094419565,
 0.2621792762345213,
 0.2105547899547249,
 0.18281784627753633,
 0.14083648477890331,
 0.10497488680620906]
```

In [17]:

```
plt.xlabel('K')
plt.ylabel('Sum of squared error')
plt.plot(k_rng,sse)
```

Out[17]:

```
[<matplotlib.lines.Line2D at 0x2c147420be0>]
```




In [ ]:

```

```