

```

1 import matplotlib.pyplot as plt
2 import numpy as np
3 import pandas as pd
4 %matplotlib inline

```

```

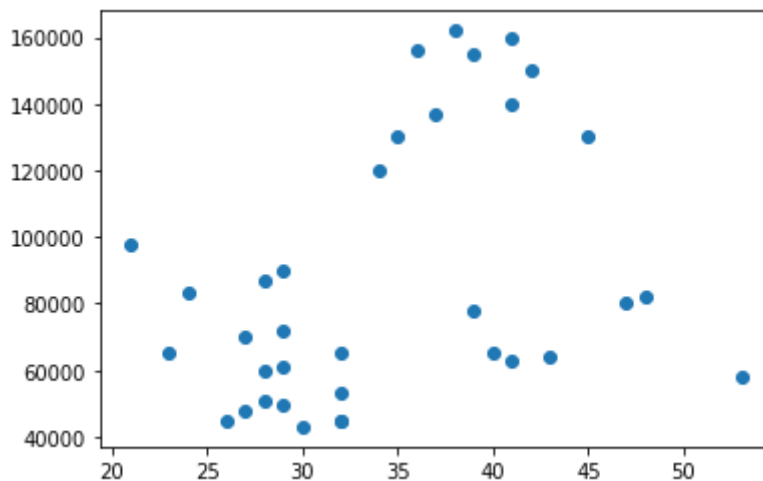
1 df = pd.read_csv("income.csv")
2 df.head()

```

	Name	Age	Income(\$)	Gender	
0	Rob	27	70000	male	
1	Bob	29	90000	male	
2	Luci	29	61000	female	
3	Joy	28	60000	male	
4	Sky	42	150000	female	

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

<matplotlib.collections.PathCollection at 0x7fc539a12fd0>



```

1 from sklearn.cluster import KMeans
2 km = KMeans(n_clusters = 3)
3 km

```

KMeans(n_clusters=3)

```

1 y_predict = km.fit_predict(df[['Age', 'Income($)']])
2 y_predict

```

array([1, 1, 2, 2, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 1, 1, 2,
2, 2, 1, 1, 1, 0, 0, 0, 2, 1, 1, 2, 2], dtype=int32)

```

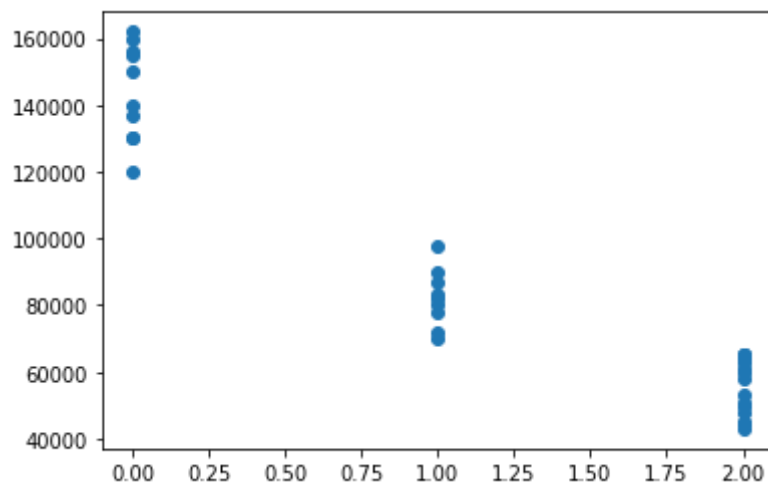
1 df['cluster']=y_predict
2 df.head()

```

	Name	Age	Income(\$)	Gender	cluster
0	Rob	27	70000	male	1
1	Bob	29	90000	male	1
2	Luci	29	61000	female	2
3	Joy	28	60000	male	2
4	Sky	42	150000	female	0

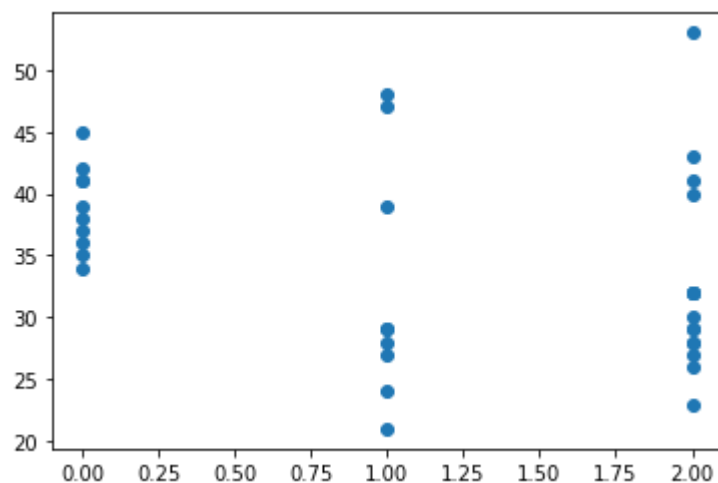
```
1 plt.scatter(df['cluster'], df['Income($)'])
```

<matplotlib.collections.PathCollection at 0x7fc539988e10>



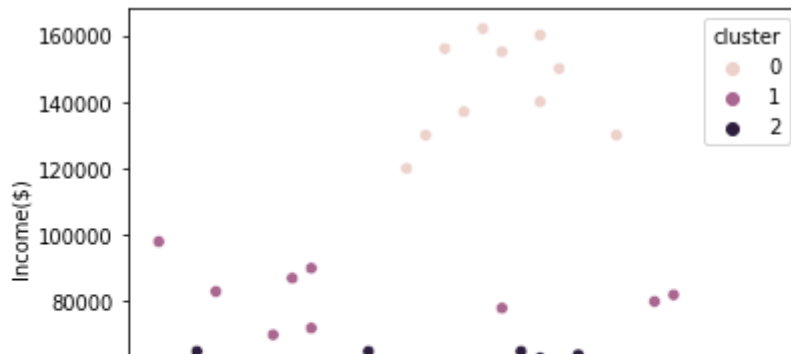
```
1 plt.scatter(df['cluster'], df['Age'])
```

<matplotlib.collections.PathCollection at 0x7fc539908210>



```
1 import seaborn as sns
2 sns.scatterplot(x=df['Age'], y=df['Income($)'], hue=df['cluster'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fc5398bde0>

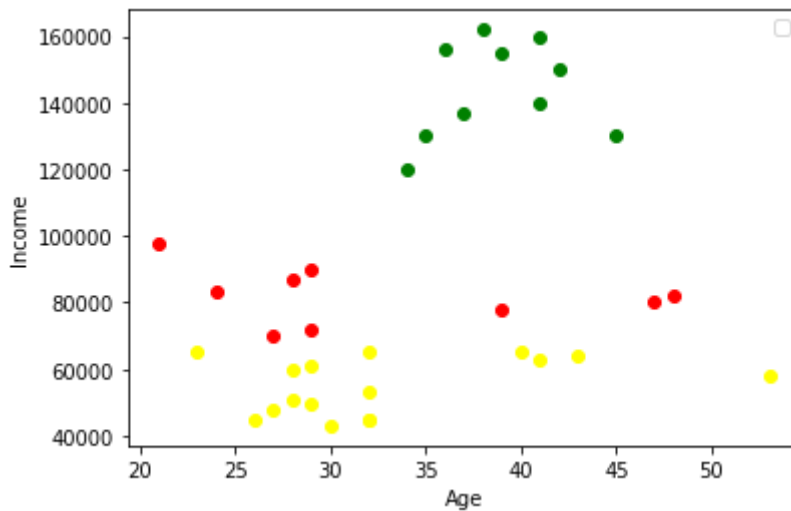


```
1 df1 = df[df.cluster ==0]
2 df2 = df[df.cluster ==1]
3 df3 = df[df.cluster ==2]
```


```
1 plt.scatter(df1['Age'], df1['Income($)', color = 'green')
2 plt.scatter(df2['Age'], df2['Income($)', color = 'red')
3 plt.scatter(df3['Age'], df3['Income($)', color = 'yellow')
4
5 plt.xlabel('Age')
6 plt.ylabel('Income')
7 plt.legend()
```

No handles with labels found to put in legend.

<matplotlib.legend.Legend at 0x7fc5398085d0>



```
1 from sklearn.preprocessing import MinMaxScaler
2 scaler = MinMaxScaler()
3 features = ['Income($)']
4 scaler.fit(df[features])
5 df['Income($)'] =scaler.transform(df[features])
6 df.head()
```

	Name	Age	Income(\$)	Gender	cluster	
0	Rob	27	0.226891	male	1	

```
1 scaler = MinMaxScaler()
2 features = ['Age']
3 scaler.fit(df[features])
4 df['Age'] =scaler.transform(df[features])
5 df
```

3	Joy	0.21875	0.142857	male	2
4	Sky	0.65625	0.899160	female	0
5	Gautam	0.56250	0.941176	male	0
6	Raj	0.62500	0.983193	male	0
7	Roja	0.53125	1.000000	female	0
8	Ranbir	0.46875	0.949580	male	0
9	Tom	0.43750	0.731092	male	0
10	Kory	0.50000	0.789916	female	0
11	Javed	0.15625	0.016807	male	2
12	Mohamod	0.18750	0.042017	male	2

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

```
array([0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2,
       0, 0, 2, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0], dtype=int32)
```

17	Monica	0.62500	0.168067	female	2
----	--------	---------	----------	--------	---

```
1 df['cluster']=y_predict
2 df.head()
```

	Name	Age	Income(\$)	Gender	cluster
0	Rob	0.18750	0.226891	male	0
1	Bob	0.25000	0.394958	male	0
2	Luci	0.25000	0.151261	female	0
3	Joy	0.21875	0.142857	male	0
4	Sky	0.65625	0.899160	female	1
26	Geetha	0.00000	0.462185	female	1

```
1 import seaborn as sns
2 sns.scatterplot(x=df['Age'],y=df['Income($)'],hue=df['cluster'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7fc5397e0910>



```
1 km.cluster_centers_
```

```
array([[0.22222222, 0.16643324],
       [0.55625    , 0.8487395 ],
       [0.73214286, 0.22689076]])
```



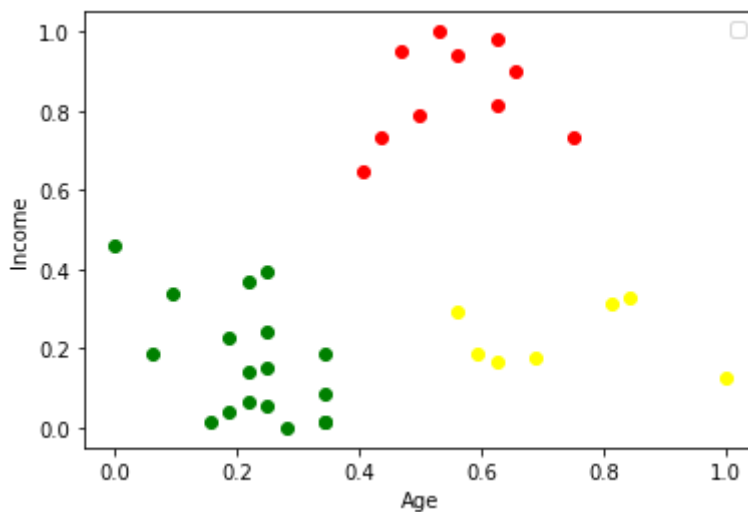
```
1 df1 = df[df.cluster ==0]
2 df2 = df[df.cluster ==1]
3 df3 = df[df.cluster ==2]
```



```
1 plt.scatter(df1['Age'], df1['Income($)', color = 'green')
2 plt.scatter(df2['Age'], df2['Income($)', color = 'red')
3 plt.scatter(df3['Age'], df3['Income($)', color = 'yellow')
4
5 plt.xlabel('Age')
6 plt.ylabel('Income')
7 plt.legend()
```

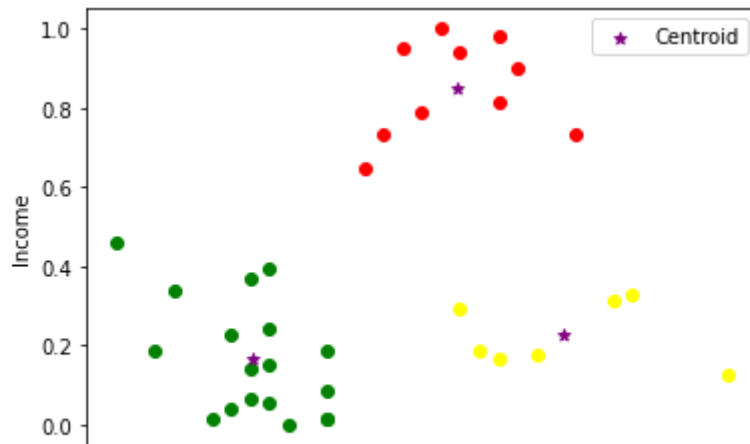
No handles with labels found to put in legend.

<matplotlib.legend.Legend at 0x7fc53974fed0>



```
1 df1 = df[df.cluster ==0]
2 df2 = df[df.cluster ==1]
3 df3 = df[df.cluster ==2]
4
5
6 plt.scatter(df1['Age'], df1['Income($)', color = 'green')
7 plt.scatter(df2['Age'], df2['Income($)', color = 'red')
8 plt.scatter(df3['Age'], df3['Income($)', color = 'yellow')
9 plt.scatter(km.cluster_centers_[0], km.cluster_centers_[1], color = 'purple', mark
10 plt.xlabel('Age')
11 plt.ylabel('Income')
12 plt.legend()
```

<matplotlib.legend.Legend at 0x7fc539681310>



```

1 k_rng = range(1,10)
2 sse = []
3 for k in k_rng:
4     km = KMeans(n_clusters = k)
5     km.fit(df[['Age', 'Income($)']])
6     sse.append(km.inertia_)
7 sse

```

```

[5.708753283351154,
 2.2851486642451095,
 0.9562308716378707,
 0.6399754025225444,
 0.5125185891446198,
 0.39412650542037386,
 0.3308234032517945,
 0.27402053135132737,
 0.21684146429711595]

```

```

1 plt.xlabel('Number of K')
2 plt.ylabel("Sum of squared errors")
3 plt.plot(k_rng,sse)

```

☞ [<matplotlib.lines.Line2D at 0x7fc534d8e6d0>]

