

```
In [18]: import numpy as np
import pandas as pd
```

```
In [19]: df=pd.read_csv("E:\income.csv")
df.head()
```

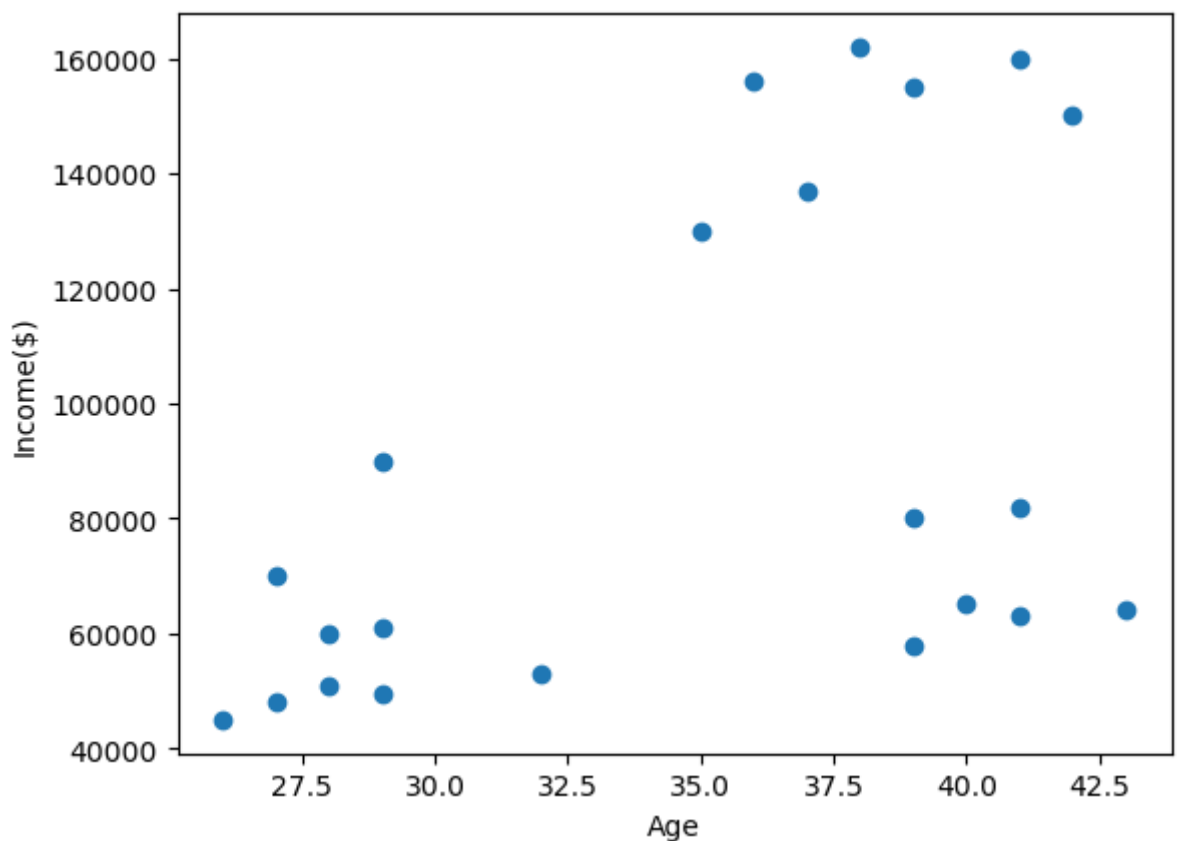
```
Out[19]:
```

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

```
In [20]: import matplotlib.pyplot as plt
```

```
In [22]: plt.scatter(df.Age,df['Income($)'])
plt.xlabel('Age')
plt.ylabel('Income($)')
```

```
Out[22]: Text(0, 0.5, 'Income($)')
```



```
In [25]: from sklearn.cluster import KMeans
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Age', 'Income($)']])
y_predicted
```

```
Out[25]: array([0, 0, 2, 2, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 0, 0, 2])
```

```
In [26]: df['cluster']=y_predicted
df.head()
```

```
Out[26]:
```

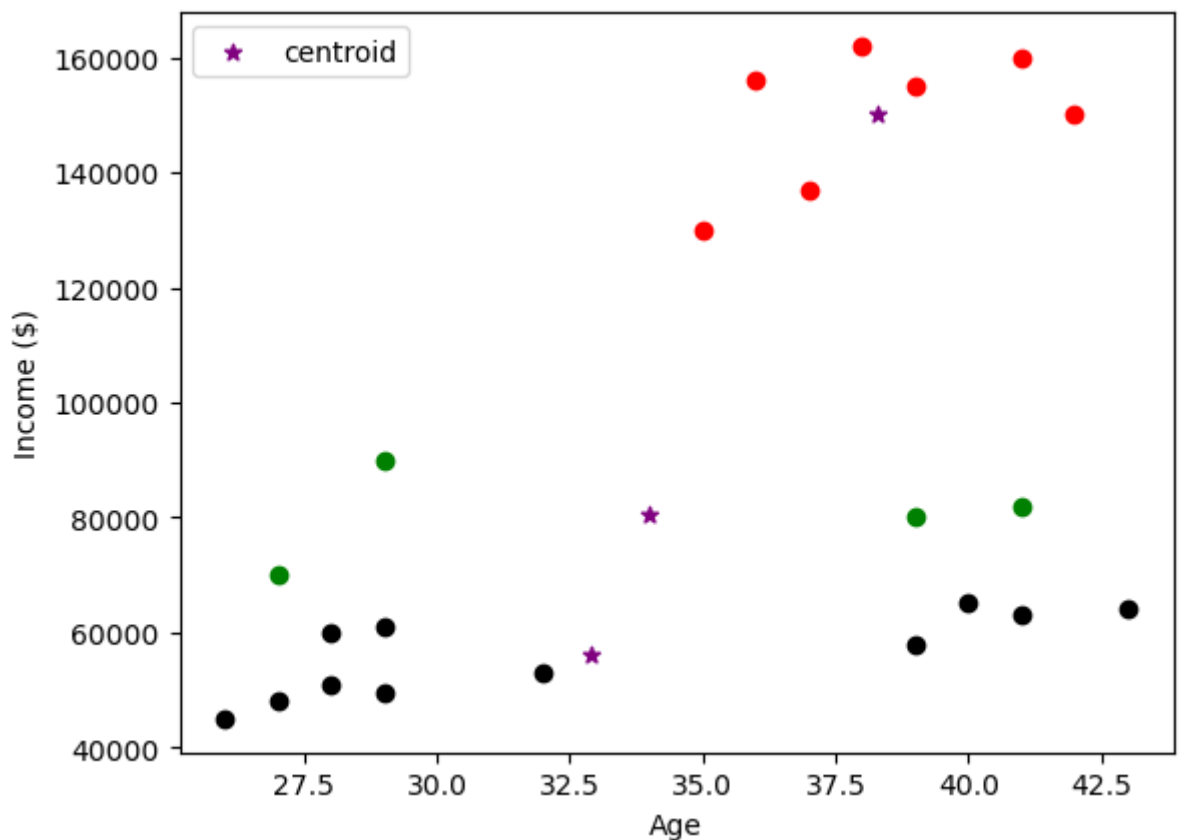
	Name	Age	Income(\$)	cluster
0	Rob	27	70000	0
1	Michael	29	90000	0
2	Mohan	29	61000	2
3	Ismail	28	60000	2
4	Kory	42	150000	1

```
In [28]: km.cluster_centers_
```

```
Out[28]: array([[3.40000000e+01, 8.05000000e+04],
        [3.82857143e+01, 1.50000000e+05],
        [3.29090909e+01, 5.61363636e+04]])
```

```
In [29]: 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($)'], color='black')
plt.scatter(km.cluster_centers_[0], km.cluster_centers_[1], color='purple')
plt.xlabel('Age')
plt.ylabel('Income ($)')
plt.legend()
```

```
Out[29]: <matplotlib.legend.Legend at 0x238a1aac400>
```



```
In [35]: from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

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

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

```
In [36]: from sklearn.preprocessing import LabelEncoder
Le=LabelEncoder()

df['Income($)']=Le.fit_transform(df['Income($)'])

df['Age']=Le.fit_transform(df['Age'])
```

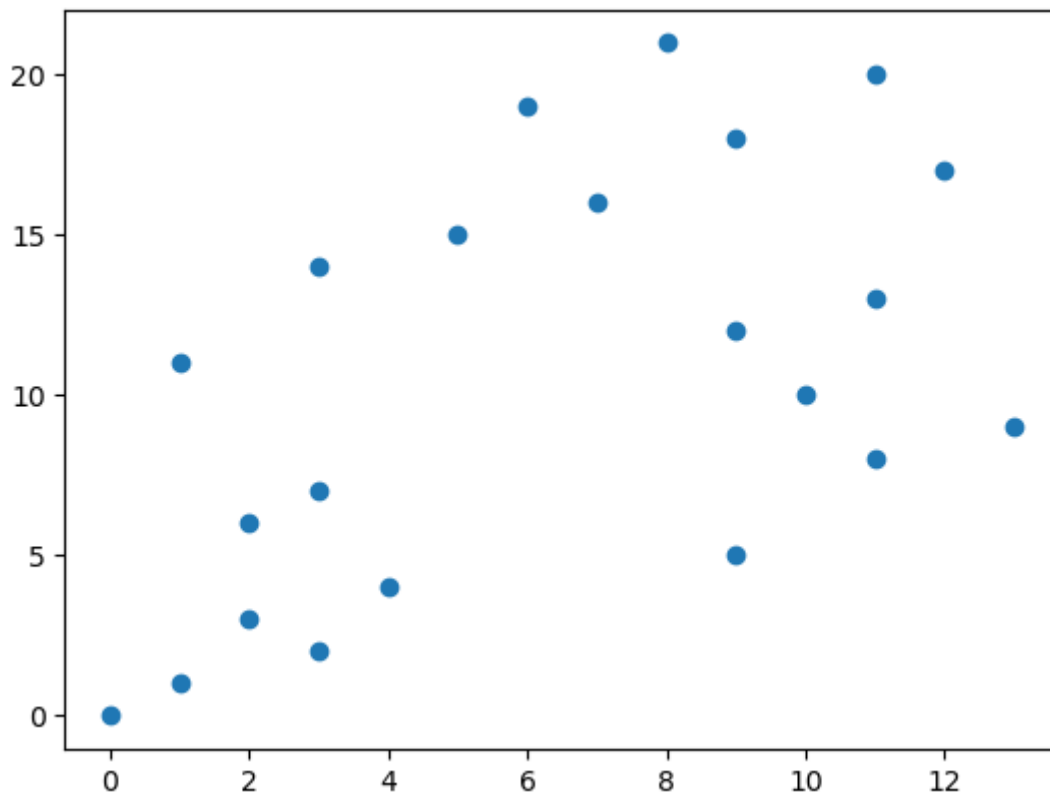
```
In [34]: df.head()
```

```
Out[34]:
```

	Name	Age	Income(\$)	cluster
0	Rob	1	11	0
1	Michael	3	14	0
2	Mohan	3	7	2
3	Ismail	2	6	2
4	Kory	12	17	1

```
In [38]: plt.scatter(df.Age,df['Income($)'])
```

```
Out[38]: <matplotlib.collections.PathCollection at 0x238a1b33f10>
```



```
In [39]: km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Age', 'Income($)']])
y_predicted
```

```
Out[39]: array([1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2])
```

```
In [40]: df['cluster']=y_predicted
df.head()
```

```
Out[40]:
```

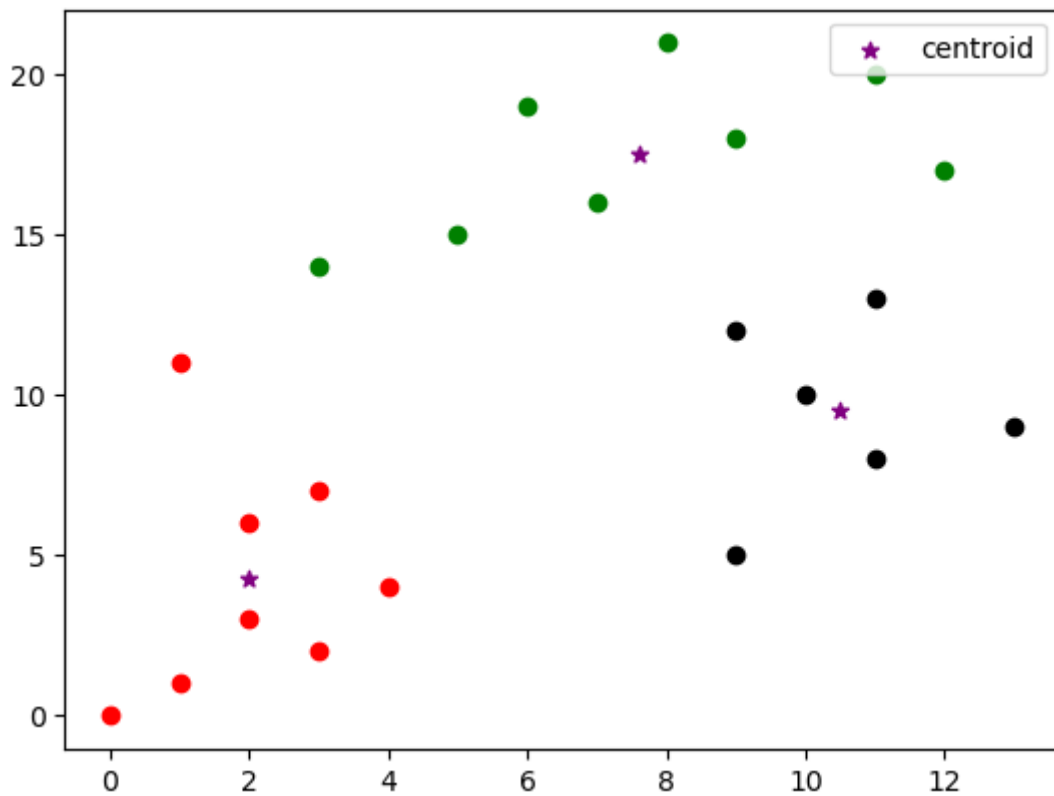
	Name	Age	Income(\$)	cluster
0	Rob	1	11	1
1	Michael	3	14	0
2	Mohan	3	7	1
3	Ismail	2	6	1
4	Kory	12	17	0

```
In [41]: km.cluster_centers_
```

```
Out[41]: array([[ 7.625, 17.5  ],
               [ 2.   ,  4.25 ],
               [10.5  ,  9.5  ]])
```

```
In [42]: 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')
plt.legend()
```

```
Out[42]: <matplotlib.legend.Legend at 0x238a1bb0730>
```



```
In [43]: df
```

Out[43]:

	Name	Age	Income(\$)	cluster
0	Rob	1	11	1
1	Michael	3	14	0
2	Mohan	3	7	1
3	Ismail	2	6	1
4	Kory	12	17	0
5	Gautam	9	18	0
6	David	11	20	0
7	Andrea	8	21	0
8	Brad	6	19	0
9	Angelina	5	15	0
10	Donald	7	16	0
11	Tom	0	0	1
12	Arnold	1	1	1
13	Jared	2	3	1
14	Stark	3	2	1
15	Ranbir	4	4	1
16	Dipika	10	10	2
17	Priyanka	11	8	2
18	Nick	13	9	2
19	Alia	9	12	2
20	Sid	11	13	2
21	Abdul	9	5	2

In []:

In []:

```
import numpy as np
import pandas as pd

df=pd.read_csv("E:\income.csv")
df.head()

import matplotlib.pyplot as plt

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

from sklearn.cluster import KMeans
km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['Age', 'Income ($)']])
y_predicted

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

km.cluster_centers_

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 ($)'], color='black')
plt.scatter(km.cluster_centers_[0], km.cluster_centers_[1], color='purple')
plt.xlabel('Age')
plt.ylabel('Income ($)')
plt.legend()

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

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

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

from sklearn.preprocessing import LabelEncoder
Le=LabelEncoder()

df['Income ($)']=Le.fit_transform(df['Income ($)'])

df['Age']=Le.fit_transform(df['Age'])

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

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

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

km.cluster_centers_

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')
plt.legend()

df

```