

**DATA SCIENCE LAB****Experiment No.: 16****Aim**

Implement KMeans algorithm using python.

**Procedure**

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

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

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

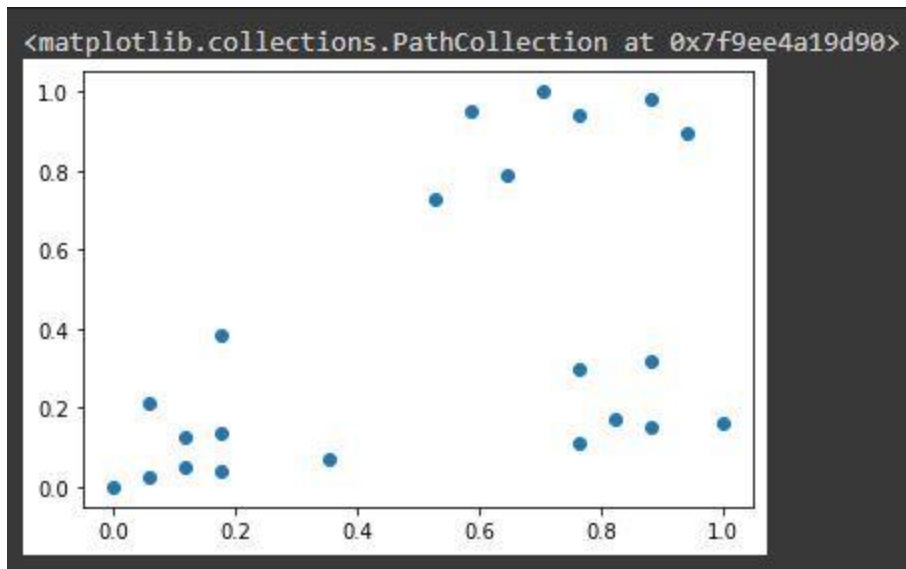
```
scaler = MinMaxScaler()
scaler.fit(df[['Income($)']])
df['Income($)'] = scaler.transform(df[['Income($)']])
scaler.fit(df[['Age']])
df['Age'] = scaler.transform(df[['Age']])
plt.scatter(df.Age,df['Income($)'])
```

**Name: SWETHA PRAKASH**

**Roll No: 46**

**Batch: B**

**Date: 14/11/2022**



```
km = KMeans(n_clusters=3)
```

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

```
y_predicted
```

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

```
df['cluster']=y_predicted
```

```
df.head()
```

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

```
km.cluster_centers_
```

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

```
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='*',label='centroid')
plt.legend()
```

## Output

