**Experiment No.: 12**

**Name: Jobin T J**

**Roll No:7**

**Batch:MCA B**

**Date:14/11/22**

**Aim**

Implement KMeans Algorithm .

**Program and Output**

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()

**Output**

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

scaler = MinMaxScaler()

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

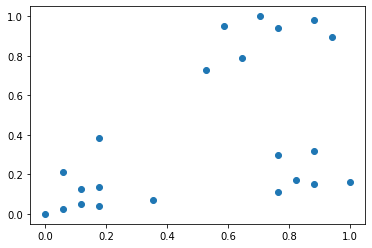
df['Income($)'] = scaler.transform(df[['Income($)']])

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

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

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

Output



**Program**

km = KMeans(n\_clusters=3)

y\_predicted = km.fit\_predict(df[['Age','Income($)']])

y\_predicted

**Output**

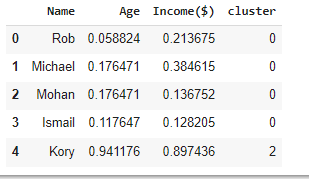
array([0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 2, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1], dtype=int32)

**Program**

df['cluster']=y\_predicted

df.head()

**Output**



km.cluster\_centers\_

**Output**

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

**Program**

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**

