```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
```

```python
df=pd.read_csv('/content/kmeans.csv')
```

```python
df
```

	Name	Age	Income($)
0	Rob	27	70000
1	Michael	29	90000
2	Mohan	29	61000
3	Ismail	28	60000
4	Kory	42	150000
5	Gautam	39	155000
6	David	41	160000
7	Andrea	38	162000
8	Brad	36	156000
9	Angelina	35	130000
10	Donald	37	137000
11	Tom	26	45000
12	Arnold	27	48000
13	Jared	28	51000
14	Stark	29	49500
15	Ranbir	32	53000
16	Dipika	40	65000
17	Priyanka	41	63000
18	Nick	43	64000
19	Alia	39	80000
20	Sid	41	82000
21	Abdul	39	58000

```python
krange=range(1,10)
sse=[]

for i in krange:
  m=KMeans(i)
  m.fit(df[['Age','Income($)']])
  sse.append(m.inertia_)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
```

```
sse
```

```
[39748148458.545456,
 3318233941.161905,
 1577046058.883117,
 814746044.2545455,
 296500418.0952381,
 239333671.3333333,
 177967003.79523814,
 119633613.78333333,
 95133611.78333333]
```

```
plt.plot(krange,sse)
```

```
[<matplotlib.lines.Line2D at 0x7d8df26bb8b0>]
```



```
model=KMeans(n_clusters=3)
model.fit(df[['Age','Income($)']])
```
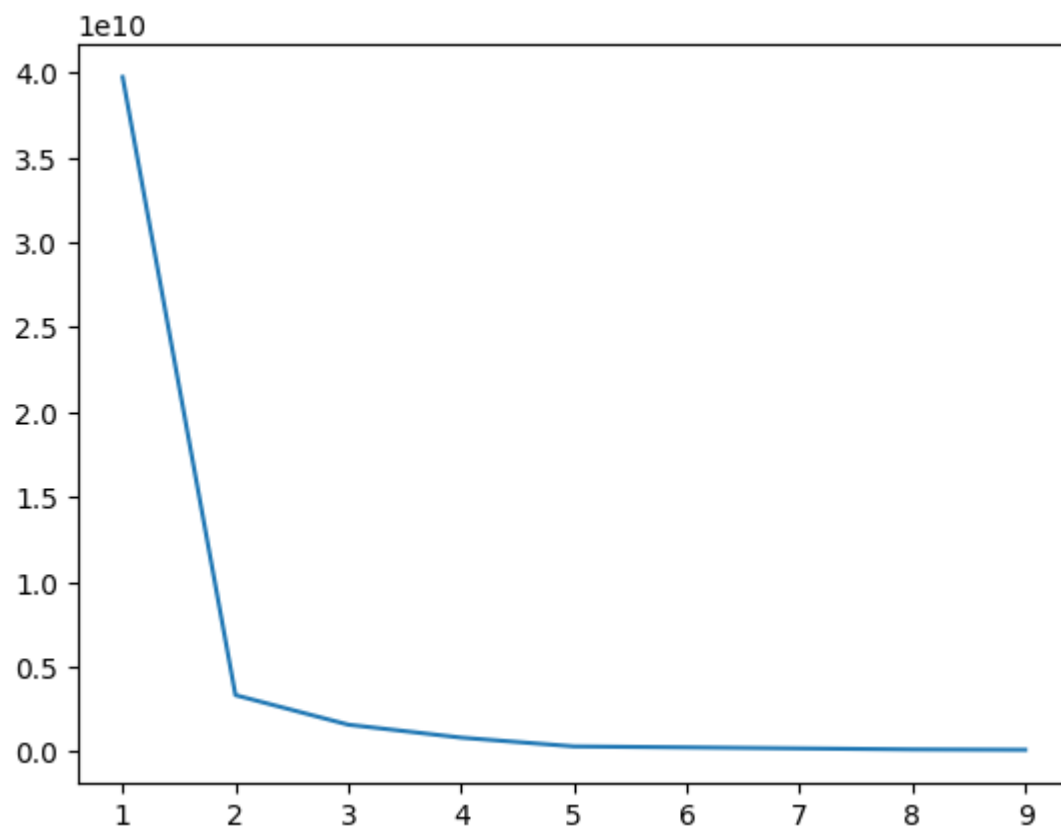
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
```

```
▼        KMeans
KMeans(n_clusters=3)
```

```
y=model.predict(df[['Age','Income($)']])
```

```
df['cluster']=y
```

```
df
```

	Name	Age	Income($)	cluster
0	Rob	27	70000	2
1	Michael	29	90000	2
2	Mohan	29	61000	1
3	Ismail	28	60000	1
4	Kory	42	150000	0
5	Gautam	39	155000	0
6	David	41	160000	0
7	Andrea	38	162000	0
8	Brad	36	156000	0
9	Angelina	35	130000	0
10	Donald	37	137000	0

```python
df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1['Age'],df1['Income($)'],color='black',label='cluster1')
plt.scatter(df2['Age'],df2['Income($)'],color='yellow',label='cluster2')
plt.scatter(df3['Age'],df3['Income($)'],color='red',label='cluster3')
plt.xlabel('Age')
plt.ylabel('Income($)')
plt.legend()
plt.show()
```
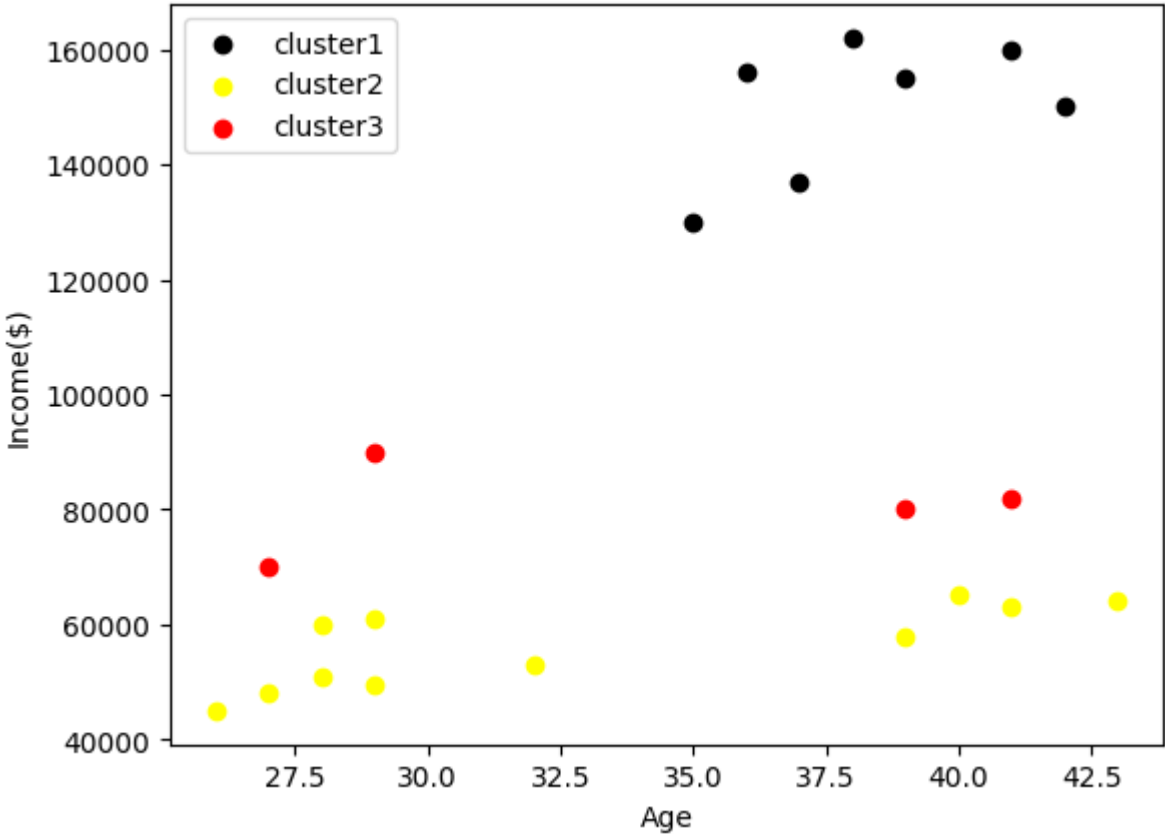


```python
scaler=MinMaxScaler()
scaler.fit(df[['Age']])
df['Age']=scaler.transform(df[['Age']])

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

```python
df
```

	Name	Age	Income($)	cluster
0	Rob	0.058824	0.213675	2
1	Michael	0.176471	0.384615	2
2	Mohan	0.176471	0.136752	1
3	Ismail	0.117647	0.128205	1
4	Kory	0.941176	0.897436	0
5	Gautam	0.764706	0.940171	0
6	David	0.882353	0.982906	0
7	Andrea	0.705882	1.000000	0
8	Brad	0.588235	0.948718	0
9	Angelina	0.529412	0.726496	0
10	Donald	0.647059	0.786325	0
11	Tom	0.000000	0.000000	1
12	Arnold	0.058824	0.025641	1
13	Jared	0.117647	0.051282	1
14	Stark	0.176471	0.038462	1
15	Ranbir	0.352941	0.068376	1

```
model1=KMeans(n_clusters=3)
model1.fit(df[['Age','Income($)']])
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change
  warnings.warn(
```

```
▼        KMeans
KMeans(n_clusters=3)
```

```
t=model1.predict(df[['Age','Income($)']])
```

```
df['clusters']=t
df.drop('cluster',axis=1,inplace=True)
```

```
df
```

	Name	Age	Income($)	clusters
0	Rob	0.058824	0.213675	0
1	Michael	0.176471	0.384615	0
2	Mohan	0.176471	0.136752	0

```
model1.cluster_centers_
```

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

| **6** | David | 0.882353 | 0.982906 | 1 |

```
df1=df[df.clusters==0]
df2=df[df.clusters==1]
df3=df[df.clusters==2]
plt.scatter(df1['Age'],df1['Income($)'],color='black',label='cluster1')
plt.scatter(df2['Age'],df2['Income($)'],color='yellow',label='cluster2')
plt.scatter(df3['Age'],df3['Income($)'],color='red',label='cluster3')
plt.scatter(model1.cluster_centers_[:,0],model1.cluster_centers_[:,1],marker='*',label='centroids')
plt.xlabel('Age')
plt.ylabel('Income($)')
plt.legend()
plt.show()
```