

```
In [1]: import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
```

```
In [2]: df=pd.read_csv(r"C:\Users\Lenovo\OneDrive\Desktop\Data Sets\income.csv")
df
```

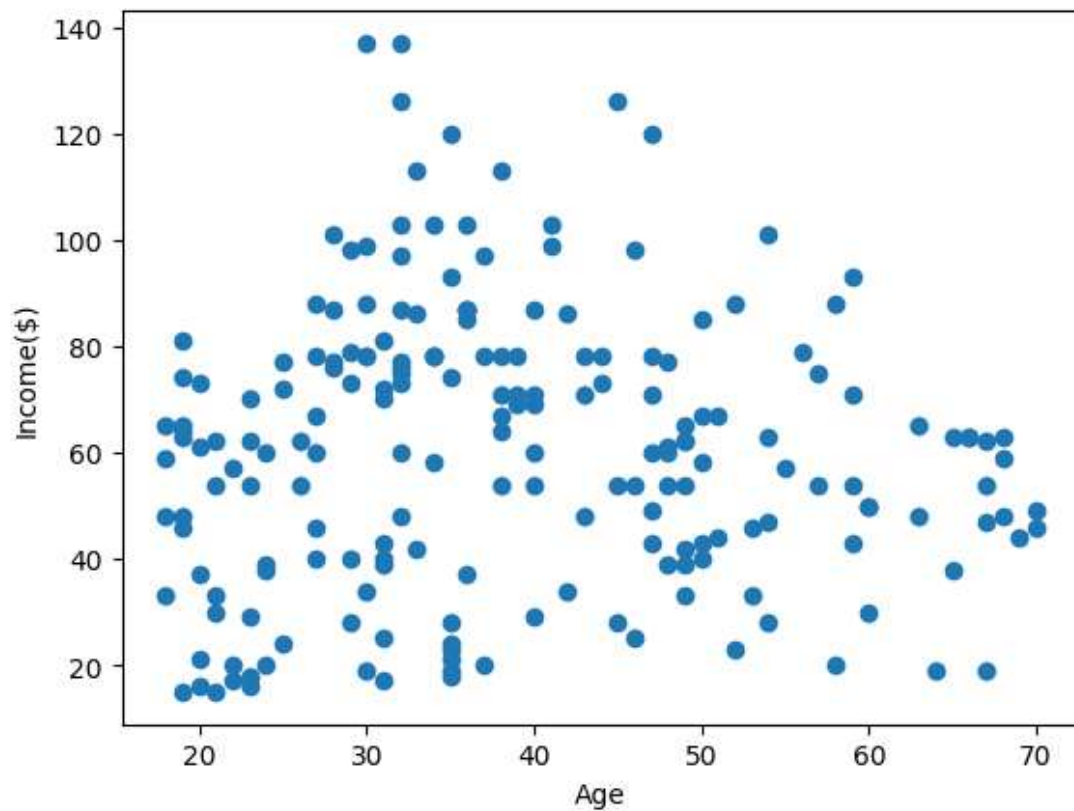
Out[2]:

	Gender	Age	Income(\$)
0	Male	19	15
1	Male	21	15
2	Female	20	16
3	Female	23	16
4	Female	31	17
...
195	Female	35	120
196	Female	45	126
197	Male	32	126
198	Male	32	137
199	Male	30	137

200 rows × 3 columns

```
In [3]: plt.scatter(df["Age"],df["Income($)"])
plt.xlabel("Age")
plt.ylabel("Income($)")
```

```
Out[3]: Text(0, 0.5, 'Income($)')
```



```
In [4]: from sklearn.cluster import KMeans
```

```
In [5]: km=KMeans()
km
```

```
Out[5]: 

▼ KMeans


KMeans()
```

```
In [6]: y_predicted=km.fit_predict(df[["Age", "Income($)"]])
y_predicted
```

C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(

```
Out[6]: array([3, 3, 3, 3, 3, 3, 3, 3, 1, 3, 1, 3, 1, 3, 3, 3, 3, 3, 1, 3, 3, 3,
        1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 1, 3, 3, 3, 6, 3, 1, 3,
        1, 3, 1, 3, 3, 3, 1, 3, 5, 6, 1, 1, 1, 6, 5, 6, 6, 5, 6, 6, 6, 5,
        4, 6, 5, 5, 6, 4, 6, 6, 6, 5, 4, 4, 5, 4, 6, 4, 6, 4, 5, 4, 6, 5,
        5, 4, 6, 5, 4, 4, 5, 5, 4, 5, 4, 5, 5, 4, 6, 5, 4, 5, 6, 4, 6, 6,
        6, 5, 4, 5, 5, 5, 6, 4, 4, 4, 5, 4, 4, 4, 2, 2, 4, 4, 4, 2, 4, 4,
        2, 2, 2, 2, 4, 2, 2, 2, 4, 2, 2, 2, 2, 2, 4, 2, 2, 2, 2, 2, 2, 2,
        4, 2, 2, 2, 2, 2, 4, 2, 2, 2, 0, 2, 0, 2, 2, 2, 0, 2, 2, 2, 0, 2,
        0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 7, 7, 7, 7, 7,
        7, 7])
```

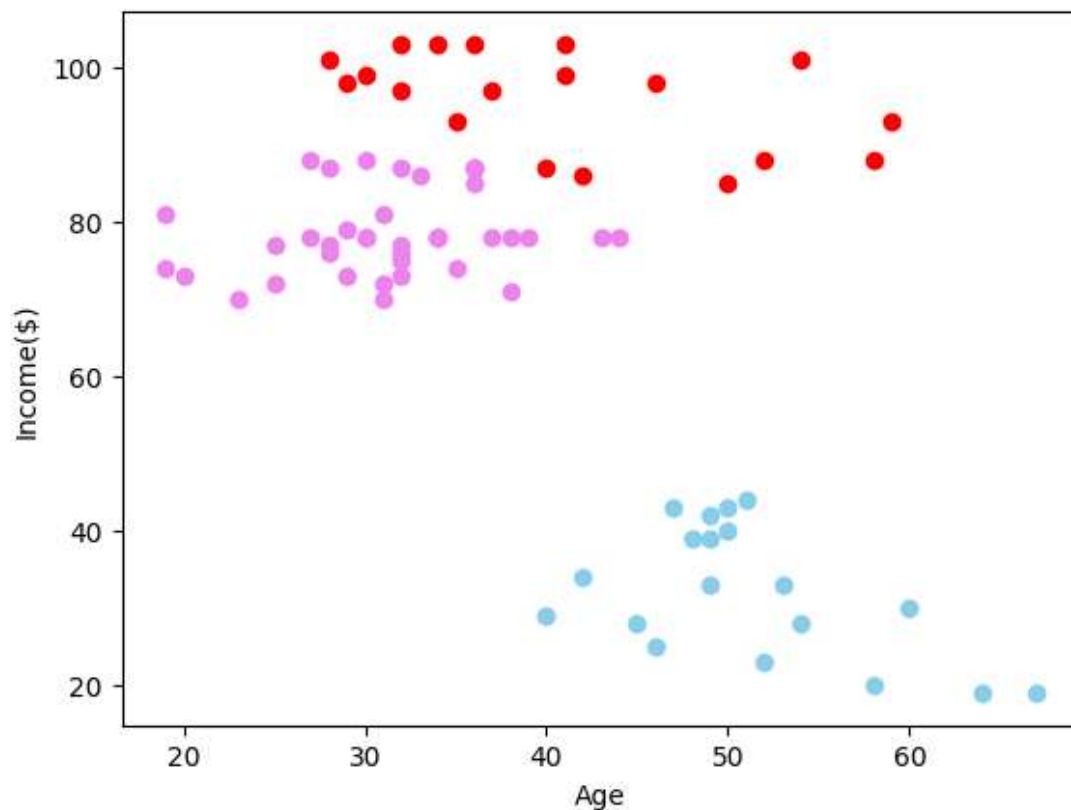
```
In [7]: df["cluster"]=y_predicted
df.head()
```

Out[7]:

	Gender	Age	Income(\$)	cluster
0	Male	19	15	3
1	Male	21	15	3
2	Female	20	16	3
3	Female	23	16	3
4	Female	31	17	3

```
In [9]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="red")
plt.scatter(df2["Age"],df2["Income($)"],color="skyblue")
plt.scatter(df3["Age"],df3["Income($)"],color="violet")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[9]: Text(0, 0.5, 'Income(\$))')



```
In [10]: from sklearn.preprocessing import MinMaxScaler
```

```
In [11]: Scaler=MinMaxScaler()
```

```
In [12]: Scaler.fit(df[["Income($)"]])
df["Income($)"]=Scaler.transform(df[["Income($)"]])
df.head()
```

Out[12]:

	Gender	Age	Income(\$)	cluster
0	Male	19	0.000000	3
1	Male	21	0.000000	3
2	Female	20	0.008197	3
3	Female	23	0.008197	3
4	Female	31	0.016393	3

```
In [13]: Scaler.fit(df[["Age"]])
df["Age"]=Scaler.transform(df[["Age"]])
df.head()
```

Out[13]:

	Gender	Age	Income(\$)	cluster
0	Male	0.019231	0.000000	3
1	Male	0.057692	0.000000	3
2	Female	0.038462	0.008197	3
3	Female	0.096154	0.008197	3
4	Female	0.250000	0.016393	3

```
In [14]: km=KMeans()
km
```

Out[14]:

```
▼ KMeans
KMeans()
```

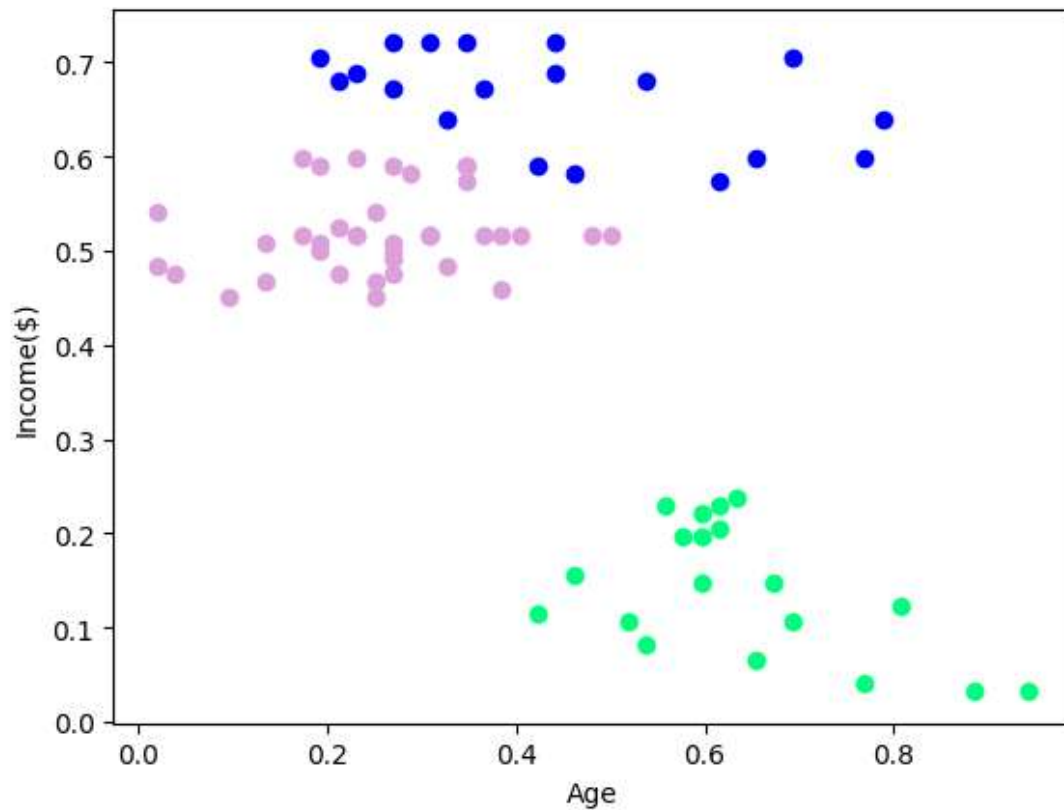
```
In [15]: y_predicted=km.fit_predict(df[["Age","Income($)"]])
y_predicted
```

C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(

```
Out[15]: array([2, 2, 2, 2, 5, 2, 5, 2, 7, 5, 7, 5, 1, 2, 5, 2, 5, 2, 1, 5, 5, 2,
 1, 5, 1, 5, 1, 5, 5, 2, 7, 2, 1, 2, 1, 2, 1, 5, 5, 2, 7, 2, 1, 5,
 1, 2, 1, 5, 5, 5, 1, 5, 5, 7, 1, 1, 1, 7, 6, 1, 7, 6, 7, 1, 7, 6,
 1, 7, 6, 5, 7, 1, 7, 7, 7, 6, 1, 1, 6, 1, 7, 0, 7, 1, 6, 1, 3, 6,
 0, 3, 7, 6, 3, 0, 0, 6, 3, 6, 3, 6, 6, 3, 7, 6, 3, 6, 7, 3, 7, 7,
 7, 6, 0, 6, 6, 6, 7, 3, 3, 3, 6, 0, 0, 0, 6, 0, 3, 0, 3, 0, 3, 0,
 6, 0, 6, 0, 3, 0, 6, 0, 3, 0, 0, 0, 6, 0, 3, 0, 0, 0, 3, 0, 3, 0,
 3, 0, 0, 0, 0, 0, 3, 0, 6, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0,
 3, 0, 3, 0, 4, 4, 4, 4, 4, 4, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
 4, 4])
```

```
In [18]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="blue")
plt.scatter(df2["Age"],df2["Income($)"],color="SpringGreen")
plt.scatter(df3["Age"],df3["Income($)"],color="plum")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

```
Out[18]: Text(0, 0.5, 'Income($)')
```

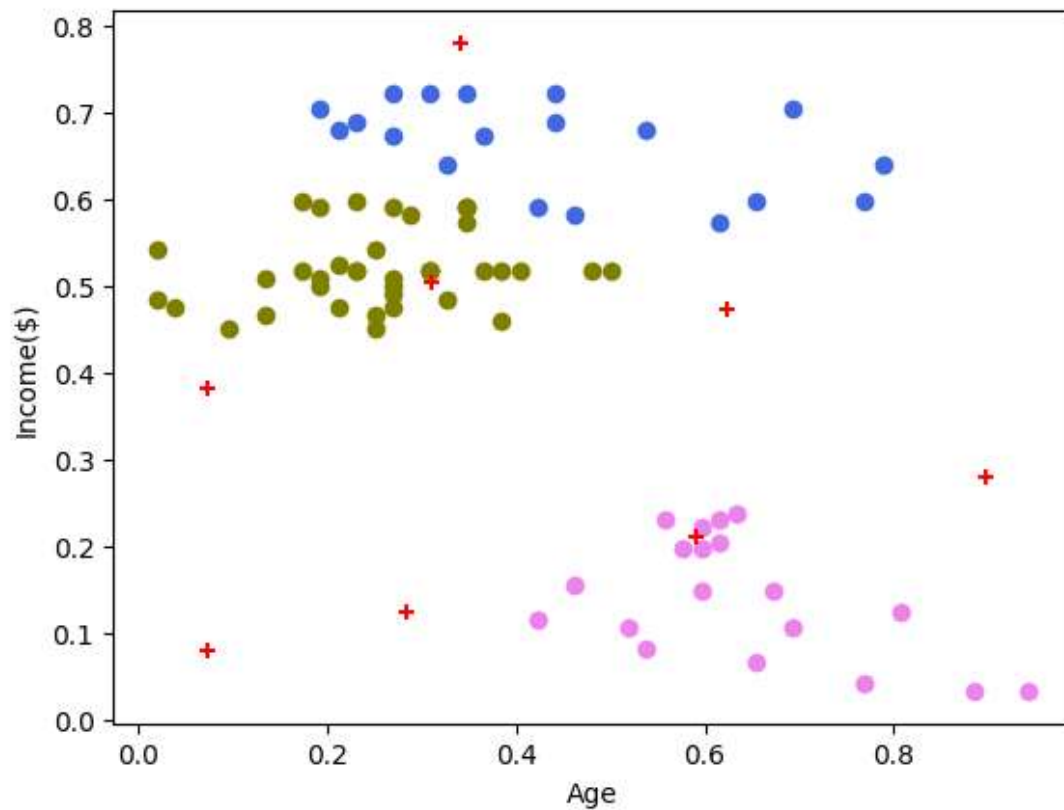


```
In [19]: km.cluster_centers_
```

```
Out[19]: array([[0.30944056, 0.50428465],
 [0.58974359, 0.20969945],
 [0.07239819, 0.08003857],
 [0.62352071, 0.47225725],
 [0.34008097, 0.77998274],
 [0.28388278, 0.1245121 ],
 [0.07322485, 0.38272383],
 [0.89799331, 0.28011404]])
```

```
In [21]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1["Age"],df1["Income($)"],color="royalblue")
plt.scatter(df2["Age"],df2["Income($)"],color="violet")
plt.scatter(df3["Age"],df3["Income($)"],color="olive")
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color="red",marker="+")
plt.xlabel("Age")
plt.ylabel("Income($)")
```

Out[21]: Text(0, 0.5, 'Income(\$)')



```
In [22]: k_rng=range(1,10)
          sse=[]
          for k in k_rng:
              km=KMeans(n_clusters=k)
              km.fit(df[["Age", "Income($)"]])
              sse.append(km.inertia_)
          sse
```

```
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
C:\Users\Lenovo\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
warnings.warn(
```

```
Out[22]: [23.583906150363607,
          13.028938428018286,
          7.49210786858601,
          6.072884728742553,
          4.725579405694896,
          3.8653915853151117,
          3.0547174363693586,
          2.642520343536072,
          2.3135720353543285]
```



```
In [23]: plt.plot(k_rng,sse)
plt.xlabel("k")
plt.ylabel("sum of squared Error")
```

Out[23]: Text(0, 0.5, 'sum of squared Error')

