

```
In [11]: import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [12]: df=pd.read_csv(r"C:\Users\USER\Downloads\Income.csv")
df
```

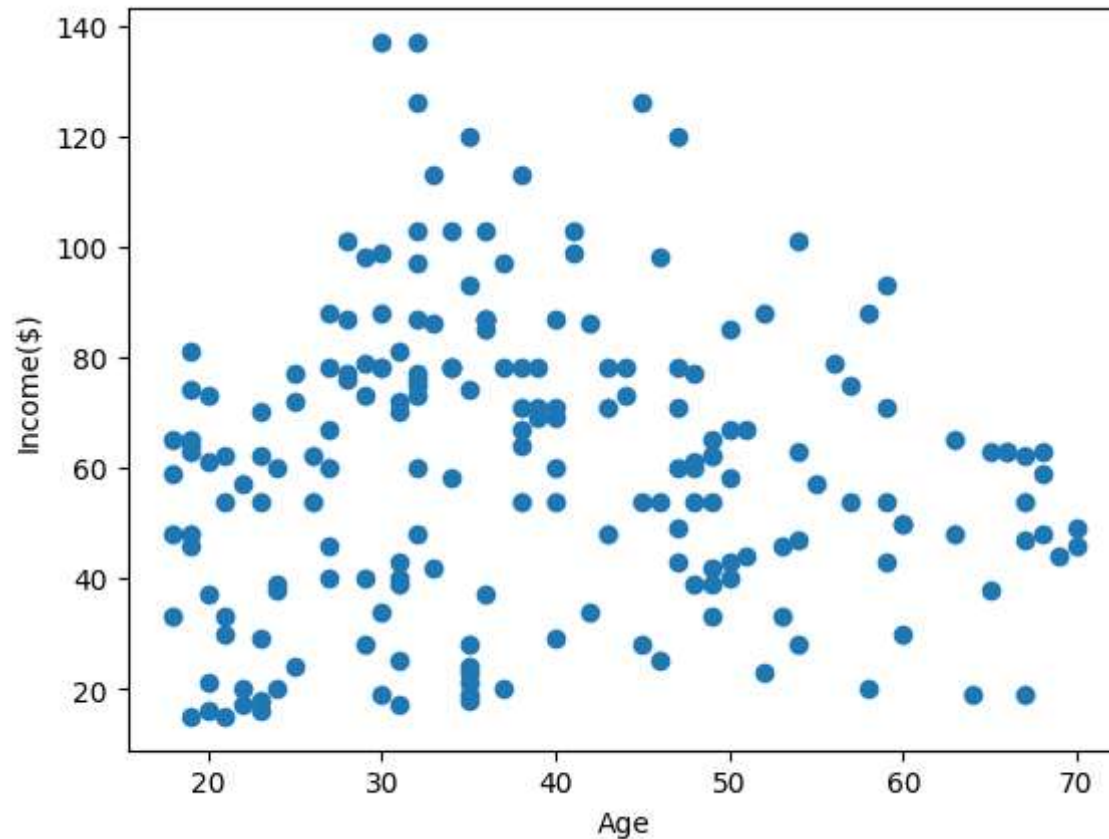
Out[12]:

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

```
Out[13]: Text(0, 0.5, 'Income($)')
```



```
In [14]: from sklearn.cluster import KMeans  
km=KMeans()  
km
```

```
Out[14]: 

▼ KMeans  
KMeans()


```

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

C:\Users\USER\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([7, 7, 7, 7, 7, 7, 7, 7, 1, 7, 1, 7, 1, 7, 7, 7, 7, 1, 7, 7, 7,
 1, 7, 1, 7, 1, 7, 1, 7, 1, 3, 1, 3, 1, 3, 3, 3, 1, 3, 1, 3,
 1, 3, 1, 3, 3, 3, 1, 3, 3, 1, 1, 1, 1, 6, 3, 1, 6, 3, 6, 1, 6, 3,
 1, 6, 3, 3, 6, 1, 6, 6, 6, 3, 2, 2, 3, 2, 6, 2, 6, 2, 3, 2, 6, 3,
 2, 2, 6, 0, 2, 2, 0, 0, 2, 0, 2, 0, 0, 2, 6, 0, 2, 0, 6, 2, 6, 6,
 6, 0, 2, 0, 0, 0, 6, 2, 2, 2, 0, 2, 2, 2, 0, 0, 2, 2, 2, 2, 2, 2,
 0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 2, 0,
 2, 0, 0, 0, 0, 0, 2, 0, 0, 0, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 4, 4, 4, 4,
 4, 4])
```

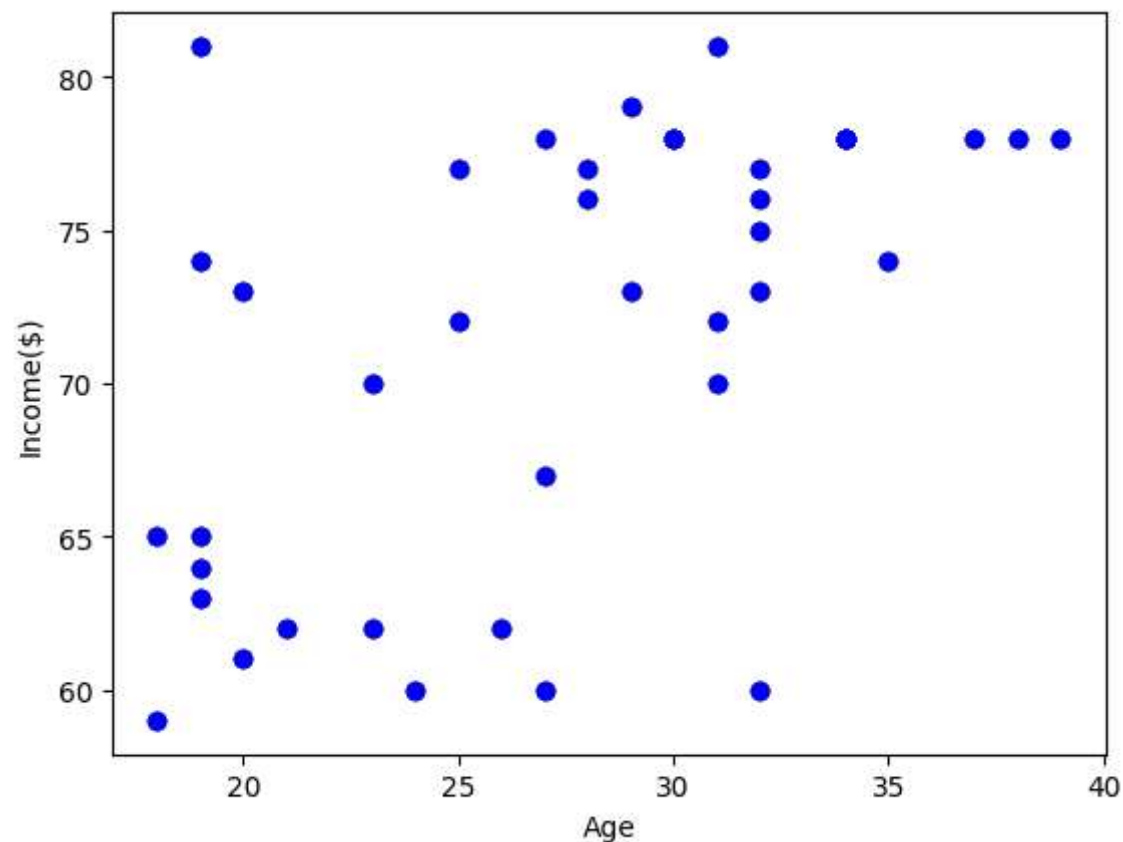
```
In [16]: df["cluster"]=y_predicted
df.head()
```

```
Out[16]:
```

	Gender	Age	Income(\$)	cluster
0	Male	19	15	7
1	Male	21	15	7
2	Female	20	16	7
3	Female	23	16	7
4	Female	31	17	7

```
In [19]: df1=df[df.cluster==0]
df2=df[df.cluster==1]
df3=df[df.cluster==2]
plt.scatter(df1['Age'],df1['Income($)',color='red')
plt.scatter(df1['Age'],df1['Income($)',color='green')
plt.scatter(df1['Age'],df1['Income($)',color='blue')
plt.xlabel('Age')
plt.ylabel('Income($))'
```

Out[19]: Text(0, 0.5, 'Income(\$))')



```
In [20]: from sklearn.preprocessing import MinMaxScaler
```

```
In [21]: scaler=MinMaxScaler()  
scaler.fit(df[['Income($)']])  
df['Income($)']=scaler.transform(df[['Income($)']])
```

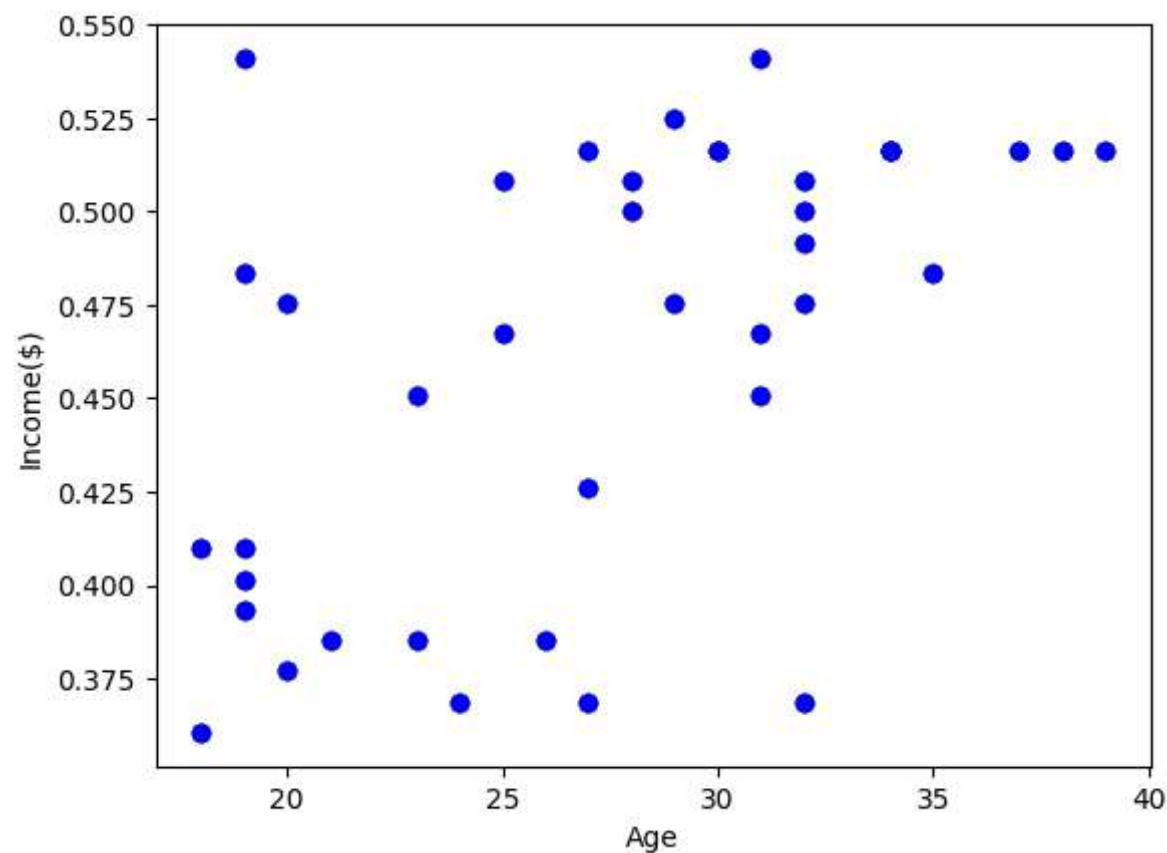
```
In [22]: df['New Cluster']=y_predicted  
df.head()
```

Out[22]:

	Gender	Age	Income(\$)	cluster	New Cluster
0	Male	19	0.000000	7	7
1	Male	21	0.000000	7	7
2	Female	20	0.008197	7	7
3	Female	23	0.008197	7	7
4	Female	31	0.016393	7	7

```
In [24]: df1=df[df['New Cluster']==0]
df2=df[df['New Cluster']==1]
df3=df[df['New Cluster']==2]
plt.scatter(df1['Age'],df1['Income($)'],color='red')
plt.scatter(df1['Age'],df1['Income($)'],color='green')
plt.scatter(df1['Age'],df1['Income($)'],color='blue')
plt.xlabel('Age')
plt.ylabel('Income($)')
```

Out[24]: Text(0, 0.5, 'Income(\$))')

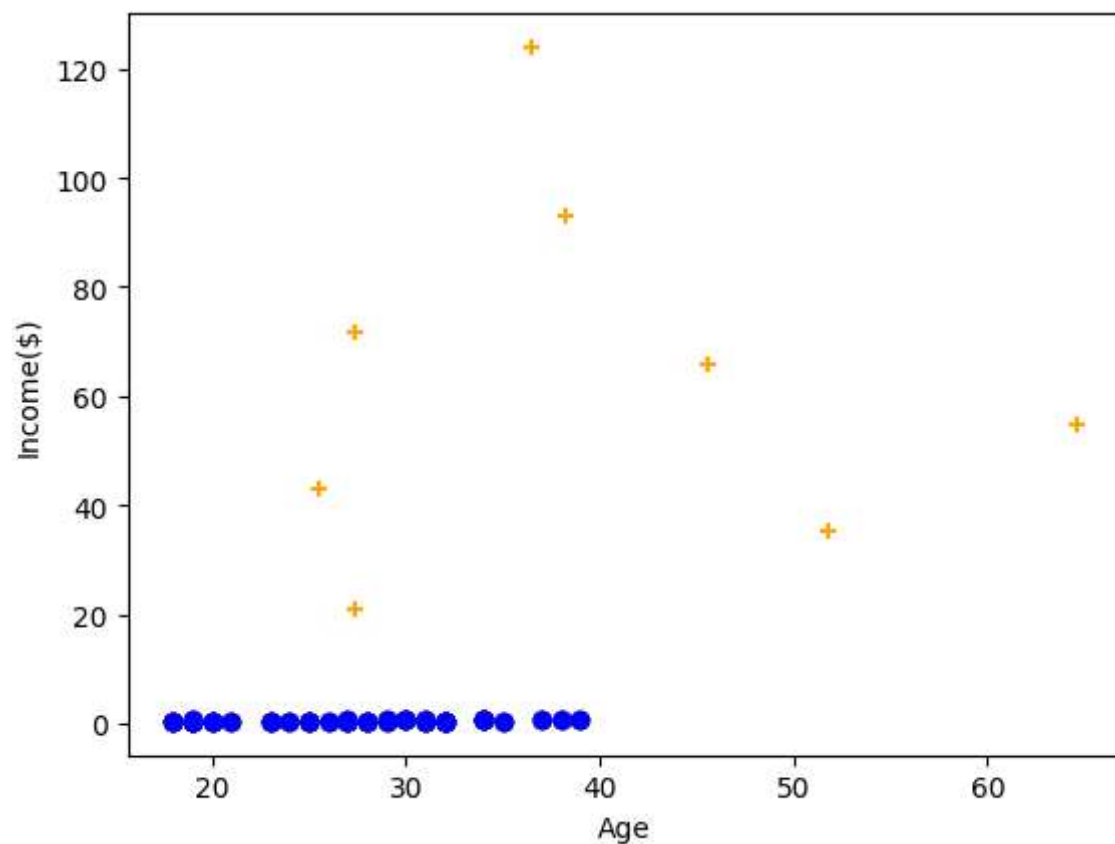


```
In [30]: km.cluster_centers_
```

```
Out[30]: array([[ 27.425      ,  71.55      ],  
                [ 51.8       ,  35.28      ],  
                [ 45.57142857,  65.71428571],  
                [ 25.54545455,  43.18181818],  
                [ 36.5       , 124.        ],  
                [ 38.21428571,  93.         ],  
                [ 64.63157895,  54.68421053],  
                [ 27.43478261,  21.         ]])
```

```
In [34]: df1=df[df['New Cluster']==0]
df2=df[df['New Cluster']==1]
df3=df[df['New Cluster']==2]
plt.scatter(df1['Age'],df1['Income($)'],color='red')
plt.scatter(df1['Age'],df1['Income($)'],color='green')
plt.scatter(df1['Age'],df1['Income($)'],color='blue')
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color='orange',marker='+')
plt.xlabel('Age')
plt.ylabel('Income($)')
```

Out[34]: Text(0, 0.5, 'Income(\$)')





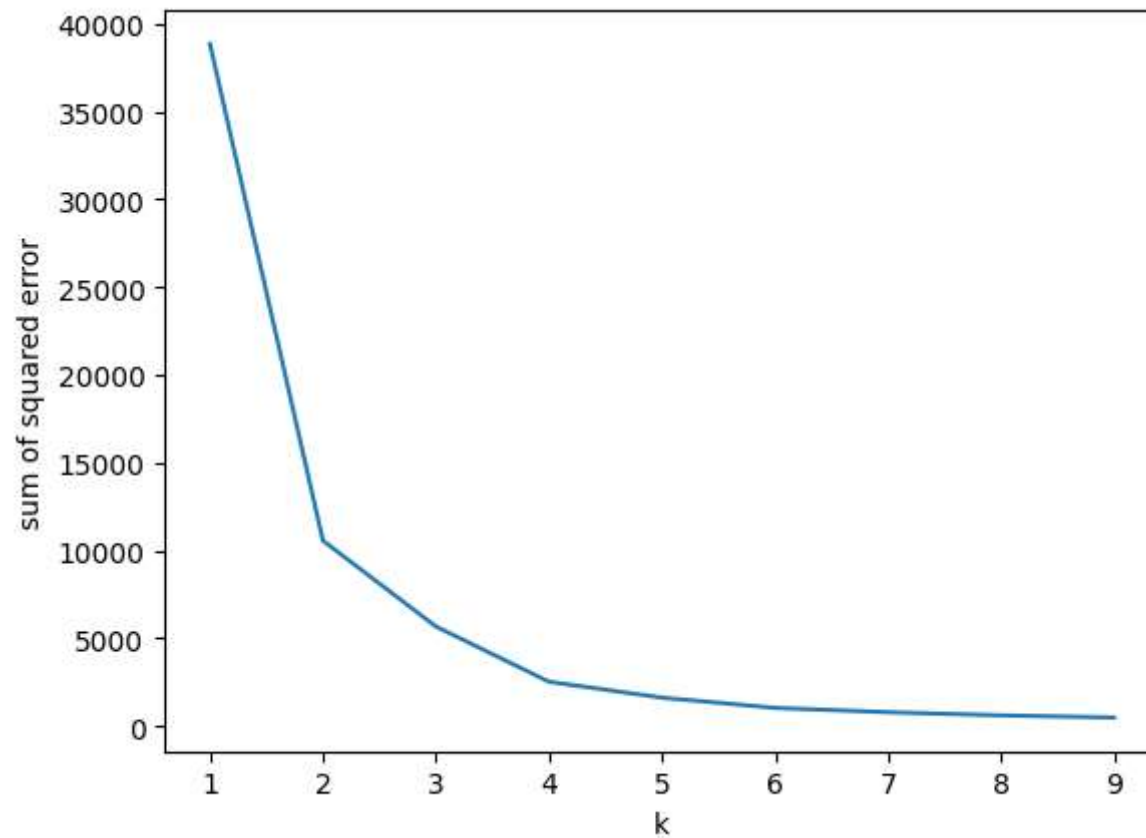
```
In [38]: K_rng=range(1,10)
sse=[]
for k in K_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[['Age', 'Income($)']])
    sse.append(km.inertia_)
sse
```

```
C:\Users\USER\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\USER\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\USER\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\USER\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\USER\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\USER\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\USER\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\USER\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\USER\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[38]: [38840.723144316034,  
          10558.82532963463,  
          5664.019792836167,  
          2521.859262993656,  
          1614.5937582632425,  
          1033.4314152361626,  
          786.4533149505711,  
          605.7951335581947,  
          484.85907027256576]
```

```
In [39]: plt.plot(K_rng,sse)  
plt.xlabel('k')  
plt.ylabel("sum of squared error")
```

```
Out[39]: Text(0, 0.5, 'sum of squared error')
```



In [ ]: