In [1]:

```
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline

```

In [3]:

```
df=pd.read_csv(r"C:\Users\kunam\Downloads\Income.csv")
df
```

Out[3]:

	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 [5]:

```
df.isnull().sum()
```

Out[5]:

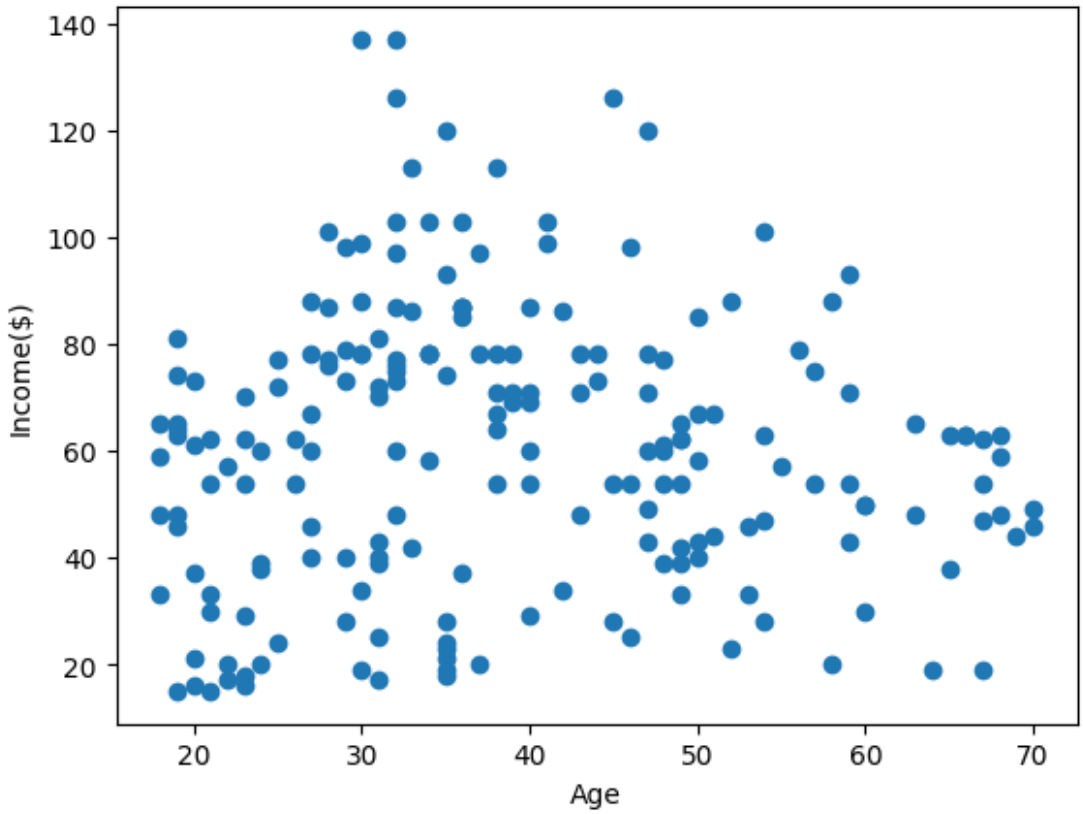
```
Gender       0
Age          0
Income($)    0
dtype: int64
```

In [4]:

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

Out[4]:

```
Text(0, 0.5, 'Income($)')
```



In [6]:

```
from sklearn.cluster import KMeans
km=KMeans()
km
```

Out[6]:

KMeans

KMeans()

In [11]:

```
y_predict=km.fit_predict(df[["Age","Income($)"]])
y_predict
```

```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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[11]:

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

In [12]:

```
df["cluster"]=y_predict
df.head()
```

Out[12]:

	Gender	Age	Income($)	cluster
0	Male	19	15	5
1	Male	21	15	5
2	Female	20	16	5
3	Female	23	16	5
4	Female	31	17	5

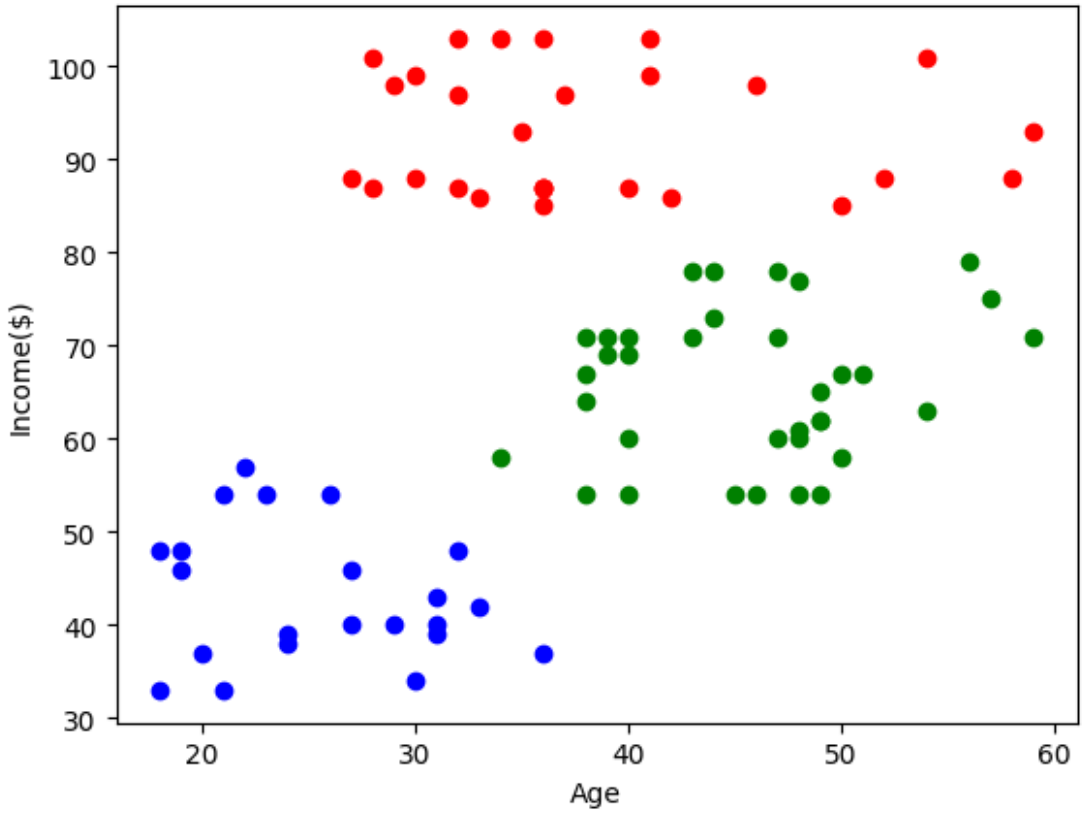
In [13]:

```
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="green")
plt.scatter(df3["Age"],df3["Income($)"],color="blue")
plt.xlabel("Age")
plt.ylabel("Income($)")

```

Out[13]:

```
Text(0, 0.5, 'Income($)')
```



In [14]:

```
from sklearn.preprocessing import MinMaxScaler
scaler=MinMaxScaler()
scaler.fit(df[["Income($)"]])
df["Income($)"]=scaler.transform(df[["Income($)"]])
df.head()
```

Out[14]:

	Gender	Age	Income($)	cluster
0	Male	19	0.000000	5
1	Male	21	0.000000	5
2	Female	20	0.008197	5
3	Female	23	0.008197	5
4	Female	31	0.016393	5

In [15]:

```
scaler.fit(df[["Age"]])
df["Age"]=scaler.transform(df[["Age"]])
df.head()
```

Out[15]:

	Gender	Age	Income($)	cluster
0	Male	0.019231	0.000000	5
1	Male	0.057692	0.000000	5
2	Female	0.038462	0.008197	5
3	Female	0.096154	0.008197	5
4	Female	0.250000	0.016393	5

In [17]:

```
y_predict=km.fit_predict(df[["Age","Income($)"]])
y_predict
```

```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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[17]:

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

In [18]:

```
df["New Cluster"]=y_predicted
df.head()
```

Out[18]:

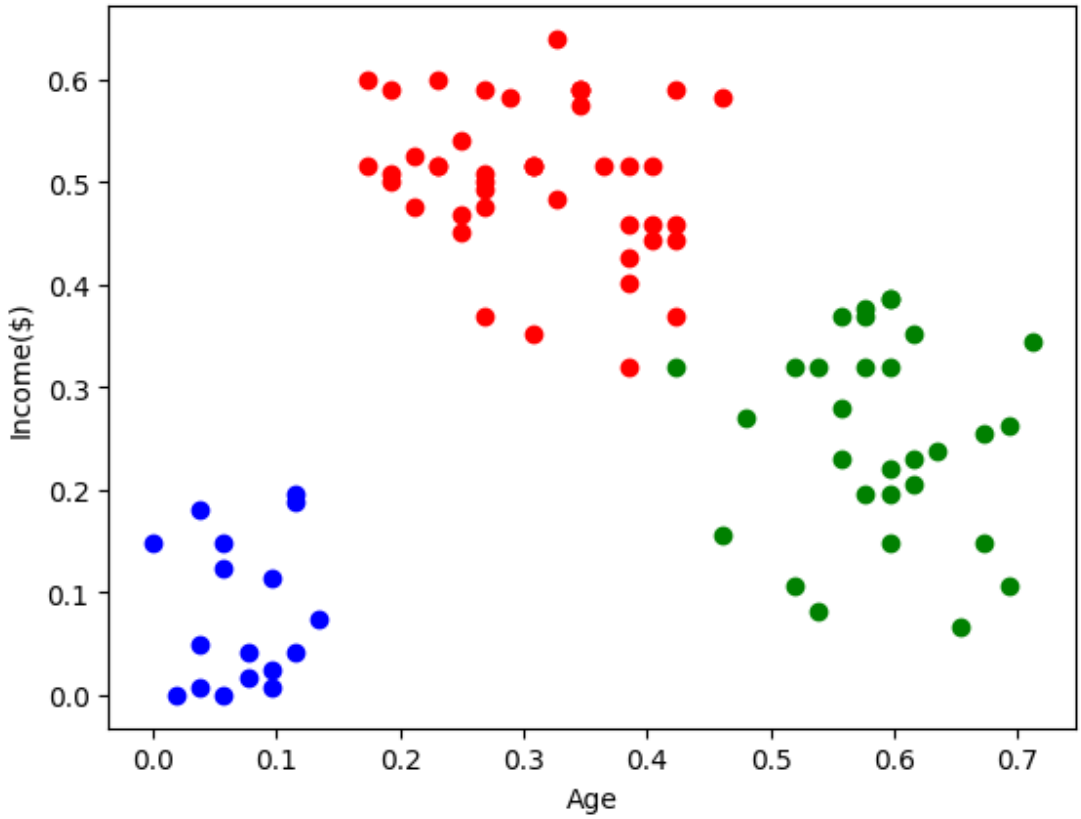
	Gender	Age	Income($)	cluster	New Cluster
0	Male	0.019231	0.000000	5	2
1	Male	0.057692	0.000000	5	2
2	Female	0.038462	0.008197	5	2
3	Female	0.096154	0.008197	5	2
4	Female	0.250000	0.016393	5	7

In [19]:

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

Out[19]:

```
Text(0, 0.5, 'Income($)')
```



In [20]:

```
km.cluster_centers_
```

Out[20]:

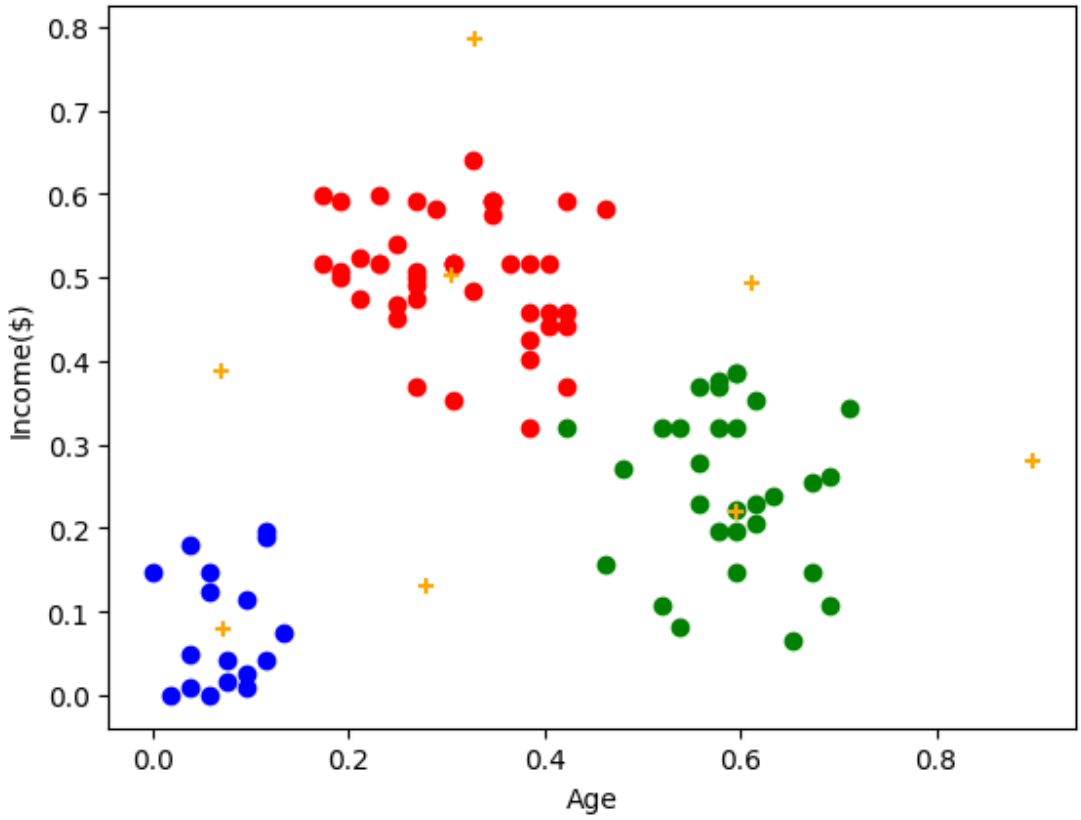
```
array([[0.27884615, 0.13040238],
       [0.89799331, 0.28011404],
       [0.3059034 , 0.50247808],
       [0.5954142 , 0.2203657 ],
       [0.07239819, 0.08003857],
       [0.61094675, 0.49401009],
       [0.06923077, 0.38786885],
       [0.32905983, 0.78551913]])
```

In [21]:

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

Out[21]:

```
Text(0, 0.5, 'Income($)')
```



In [22]:

```
k_rng=range(1,10)
sse=[]
```

In [23]:

```
for k in k_rng:
 km=KMeans(n_clusters=k)
 km.fit(df[["Age","Income($)"]])
 sse.append(km.inertia_)
print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

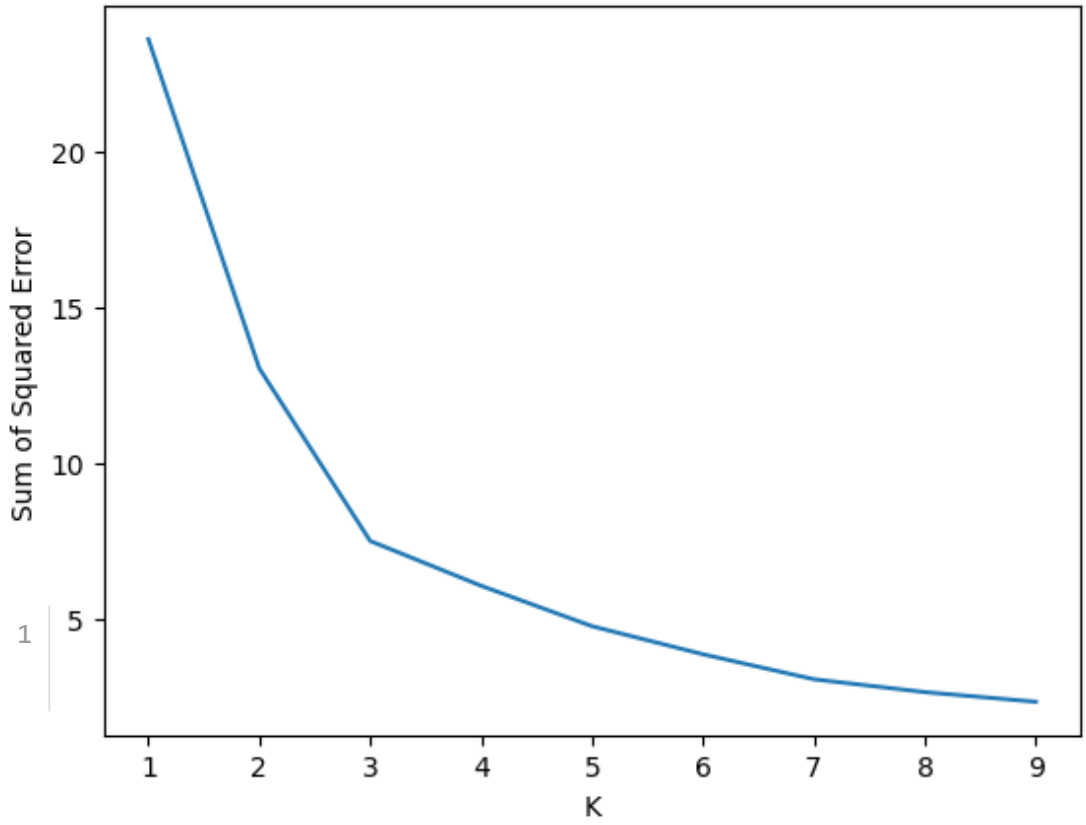
```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `
n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init`
explicitly to suppress the warning
  warnings.warn(
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package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

[23.583906150363603, 13.02893842801829, 7.492107868586013, 6.0558246675
99624, 4.7586923799663365, 3.8591087368887136, 3.058061107078988, 2.646
0609774305146, 2.335983809814013]
```

```
C:\Users\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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\magam\AppData\Local\Programs\Python\Python311\Lib\site-package
s\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[23]:

```
Text(0, 0.5, 'Sum of Squared Error')
```




```
1
```