

In [22]:

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

In [23]:

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

Out[23]:

	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 [24]:

df.head

Out[24]:

<bound method NDFrame.head of

	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 x 3 columns]>

In [25]:

```
df.tail
```

Out[25]:

```
<bound method NDFrame.tail of      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
```

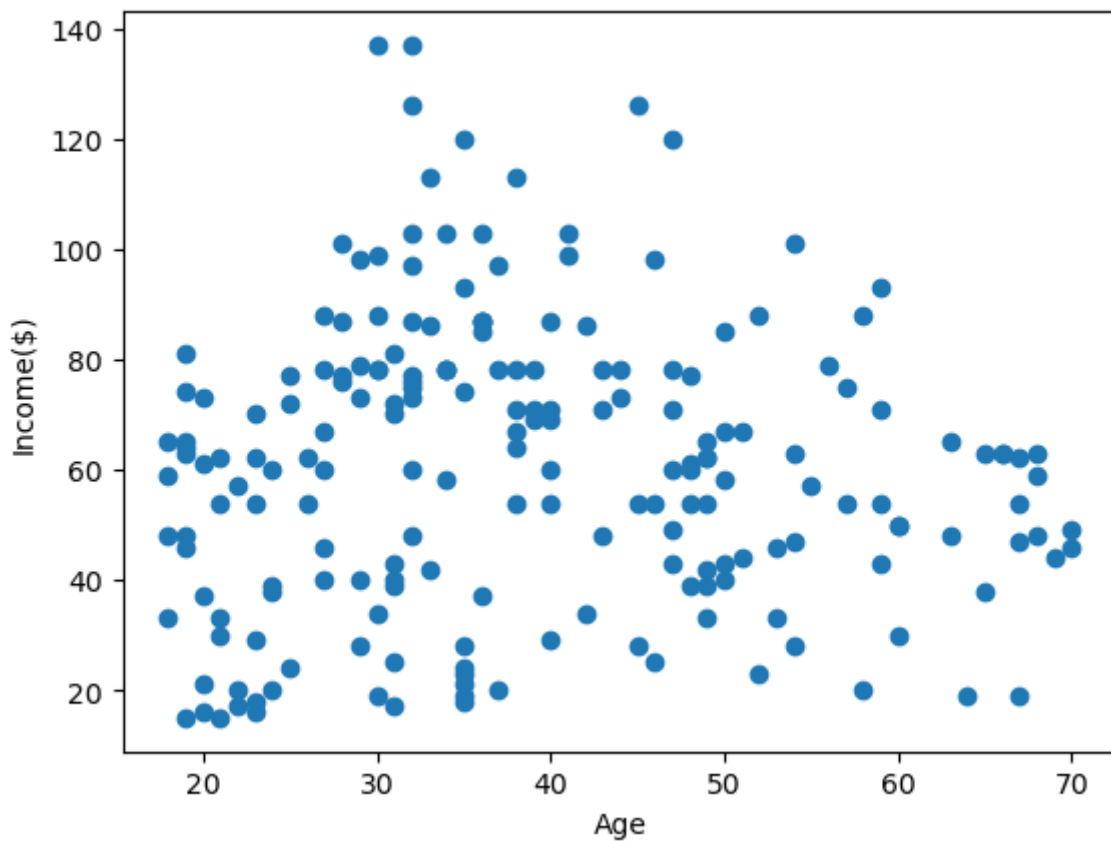
[200 rows x 3 columns]>

In [26]:

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

Out[26]:

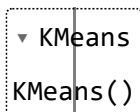
Text(0, 0.5, 'Income(\$)')



In [27]:

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

Out[27]:



In [28]:

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

C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OMP_NUM_THREADS=1.
warnings.warn(

Out[28]:

```
array([3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 3, 4, 3, 4, 3, 3, 3, 3, 3, 4, 3, 3, 3,  

       4, 3, 4, 3, 4, 3, 3, 3, 4, 3, 4, 3, 4, 3, 4, 3, 3, 3, 4, 3, 4, 3,  

       4, 3, 4, 3, 3, 3, 4, 3, 0, 4, 4, 4, 4, 6, 0, 4, 6, 0, 6, 4, 6, 0,  

       4, 6, 0, 0, 6, 4, 6, 6, 6, 0, 1, 1, 0, 1, 6, 1, 6, 1, 0, 1, 6, 0,  

       0, 1, 6, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 6, 0, 1, 0, 6, 1, 6, 6,  

       6, 0, 1, 0, 0, 0, 6, 1, 1, 1, 0, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1,  

       2, 2, 2, 2, 1, 2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2,  

       1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 5, 2, 5, 2, 2, 2, 5, 2, 2, 2, 5, 2,  

       5, 2, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 7, 7, 7, 7, 7, 7,  

       7, 7])
```

In [29]:

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

Out[29]:

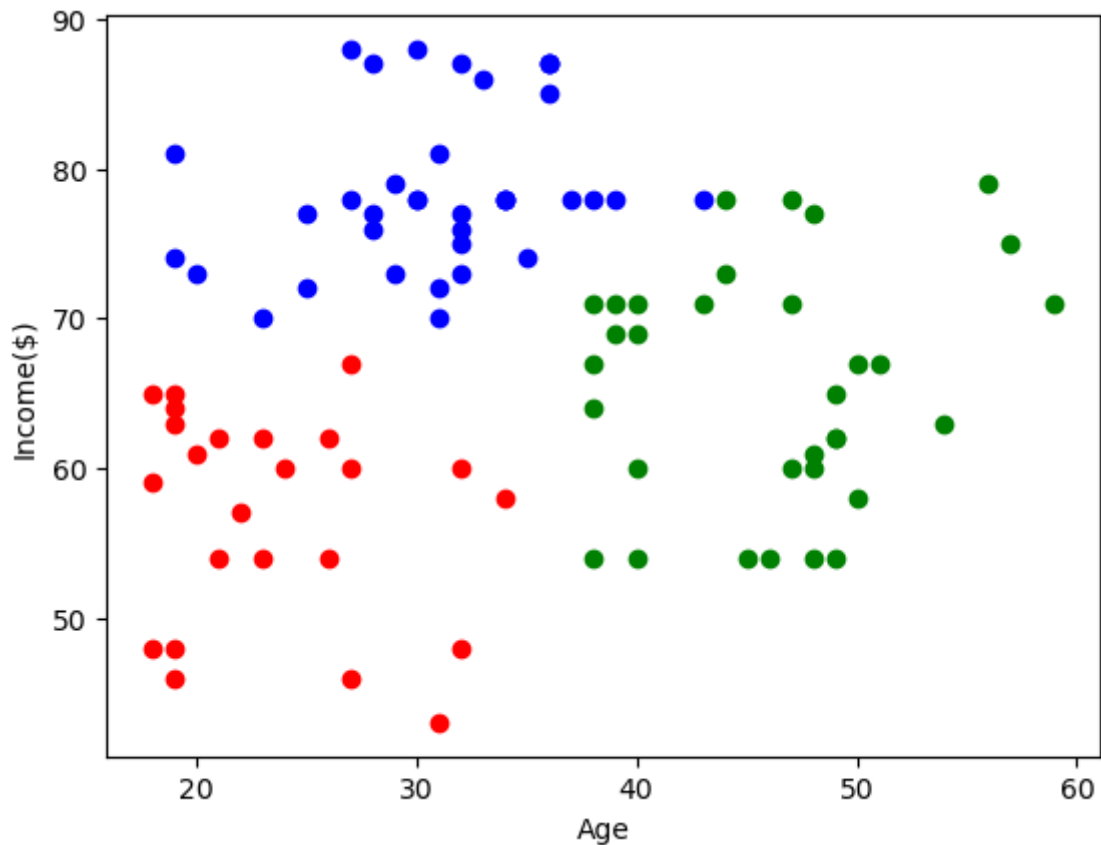
	Gender	Age	Income(\$)	cluster
0	Male	19	15	3
1	Male	21	15	3
2	Female	20	16	3
3	Female	23	16	3
4	Female	31	17	3

In [30]:

```
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[30]:

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



In [31]:

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

Out[31]:

	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 [32]:

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

Out[32]:

	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 [33]:

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

C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OMP_NUM_THREADS=1.
warnings.warn(

Out[33]:

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

In [34]:

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

Out[34]:

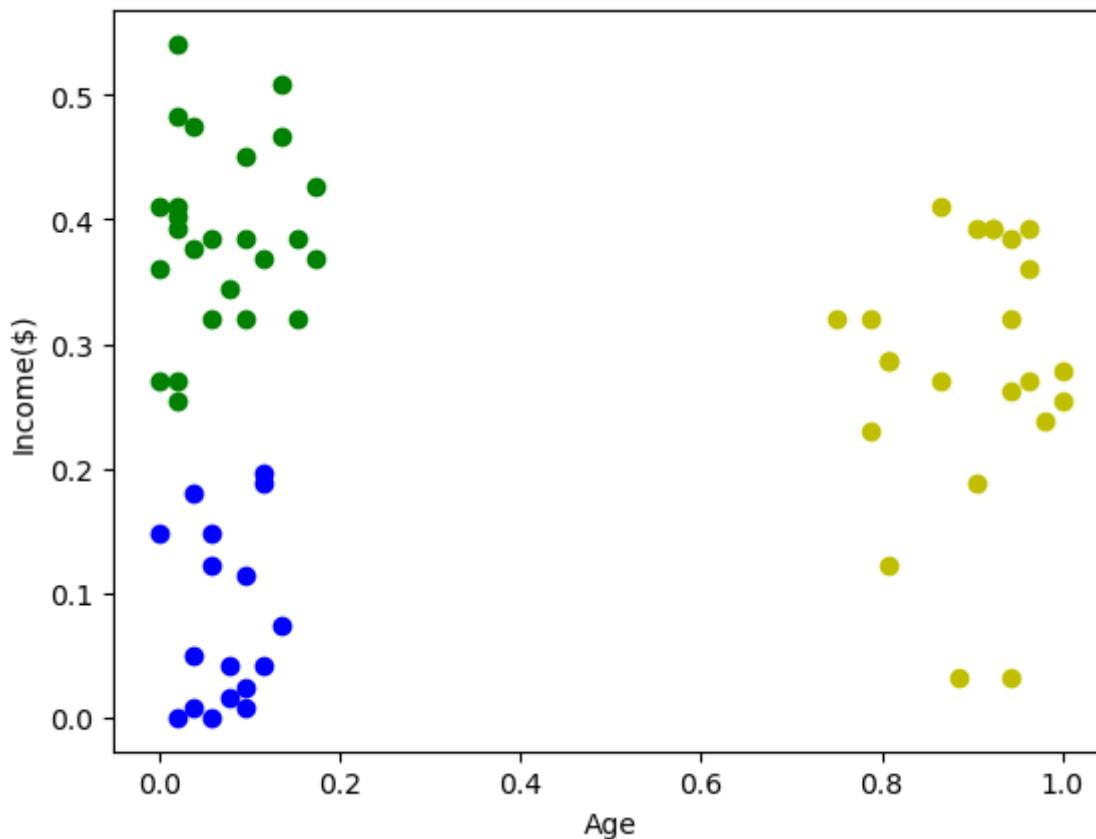
	Gender	Age	Income(\$)	cluster	New Cluster
0	Male	0.019231	0.000000	3	2
1	Male	0.057692	0.000000	3	2
2	Female	0.038462	0.008197	3	2
3	Female	0.096154	0.008197	3	2
4	Female	0.250000	0.016393	3	6

In [35]:

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

Out[35]:

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



In [36]:

```
km.cluster_centers_
```

Out[36]:

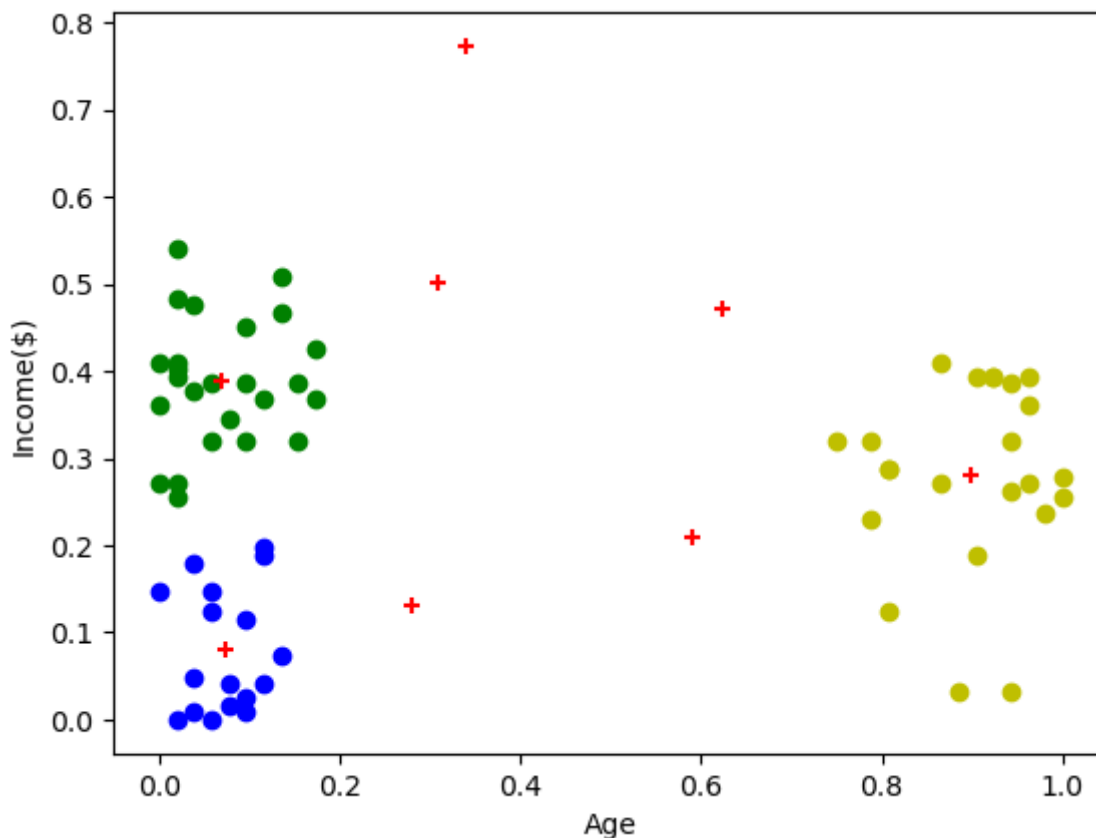
```
array([[0.89799331, 0.28011404],
       [0.06923077, 0.38786885],
       [0.07239819, 0.08003857],
       [0.30903399, 0.50114373],
       [0.58974359, 0.20969945],
       [0.33942308, 0.77295082],
       [0.27884615, 0.13040238],
       [0.62352071, 0.47225725]])
```

In [40]:

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

Out[40]:

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



In [38]:

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


In [45]:

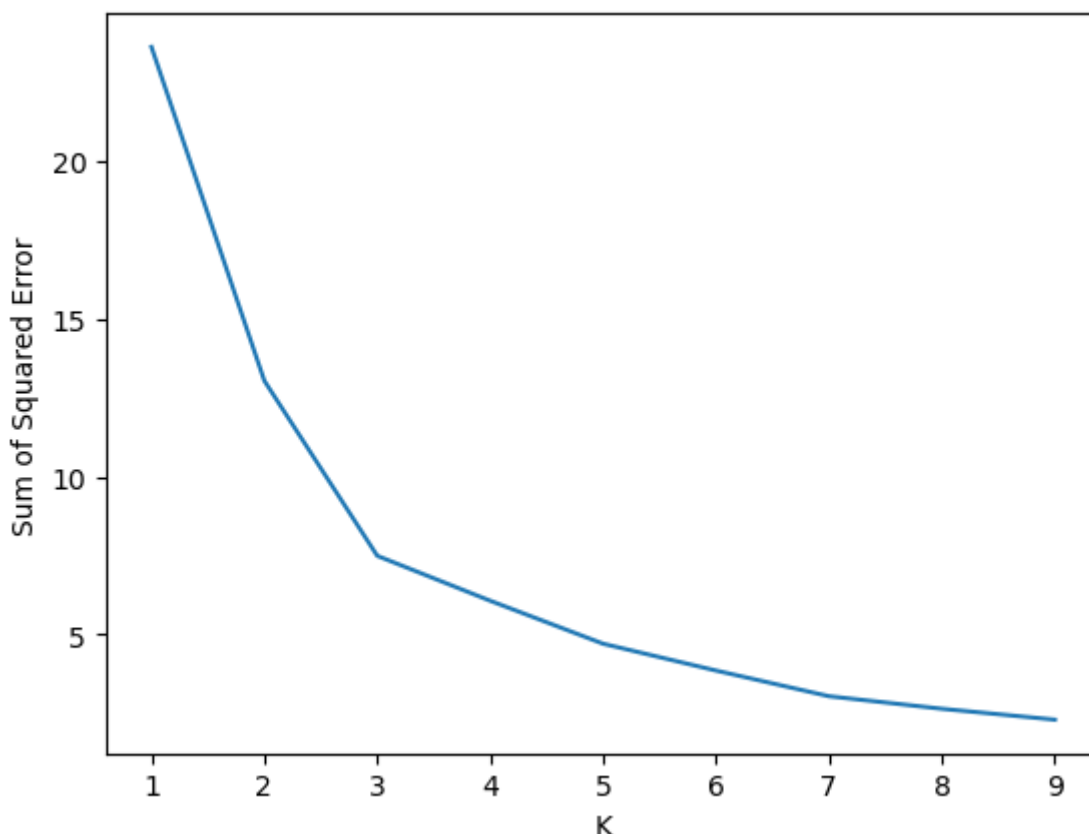
```
for k in k_rng:
    km=KMeans(n_clusters=k)
    km.fit(df[["Age","Income($)"]])
    sse.append(km.inertia_)
#km.inertia_ will give you the value of sum of square error
print(sse)
plt.plot(k_rng,sse)
plt.xlabel("K")
plt.ylabel("Sum of Squared Error")
```

```
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
    warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
    warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
    warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
    warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
    warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
    warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
```

```
ting the environment variable OMP_NUM_THREADS=1.
warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
warnings.warn(
C:\Users\jas_m\anaconda3\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\jas_m\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:138
2: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
warnings.warn(
[23.583906150363603, 13.028938428018286, 7.493024843304991, 6.072884728742
5545, 4.713811834695168, 3.865956392529676, 3.054717436369358, 2.657386592
1973026, 2.3135720353543285]
```

Out[45]:

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



In []:

In []:

In []:

In []: