

Exercise 6

Given a dataset that contains customer information (such as Age, Income, and Spending Score), perform K-means clustering to group customers into clusters. Use visualization chart, plot the data before and after grouping. Also, use the Elbow Method to determine the optimal number of clusters.

```
In [23]: import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import pandas as pd
```

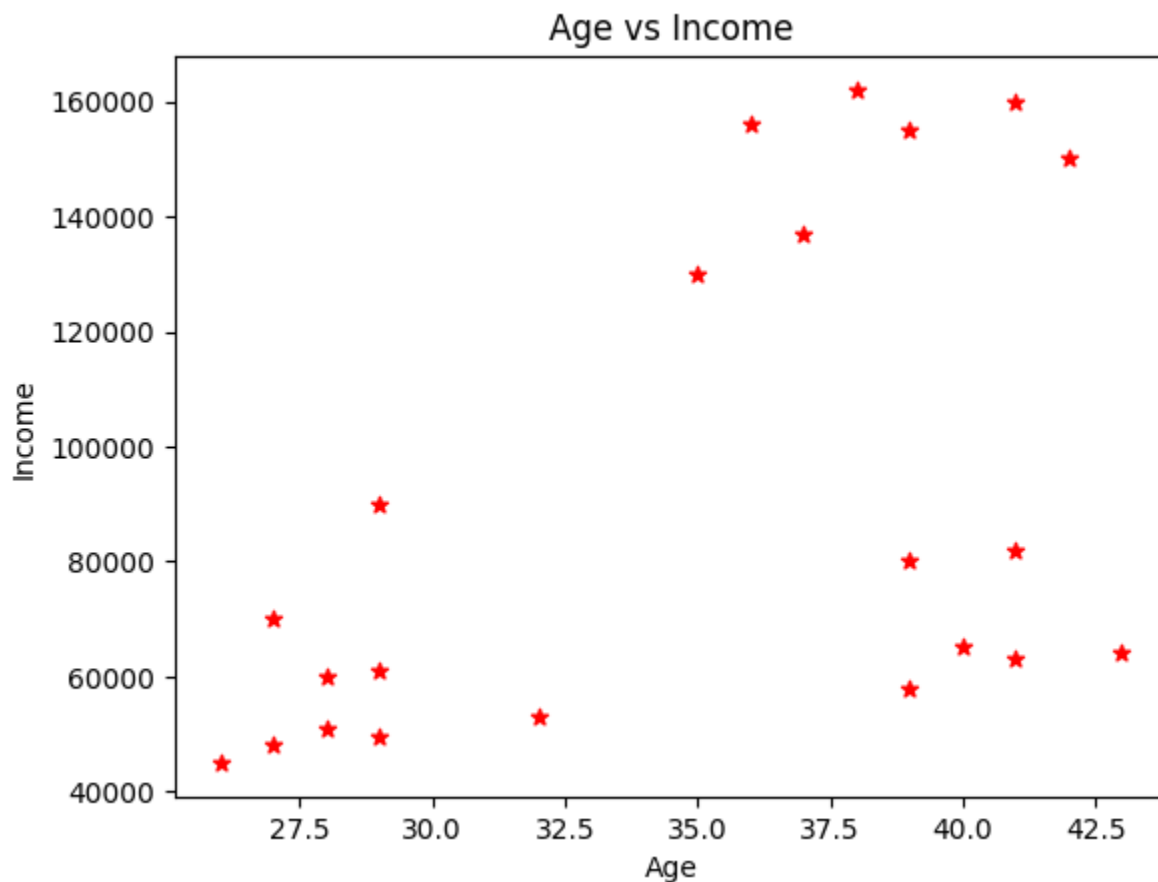
```
In [24]: df = pd.read_csv("income_clustering.csv")

df = df[["Age", "Income($)"]]

scaler = StandardScaler()
sc_df = scaler.fit_transform(df)
```

```
In [25]: plt.scatter(df["Age"], df["Income($)"], color="r", marker="*")
plt.title("Age vs Income")
plt.xlabel("Age")
plt.ylabel("Income")
```

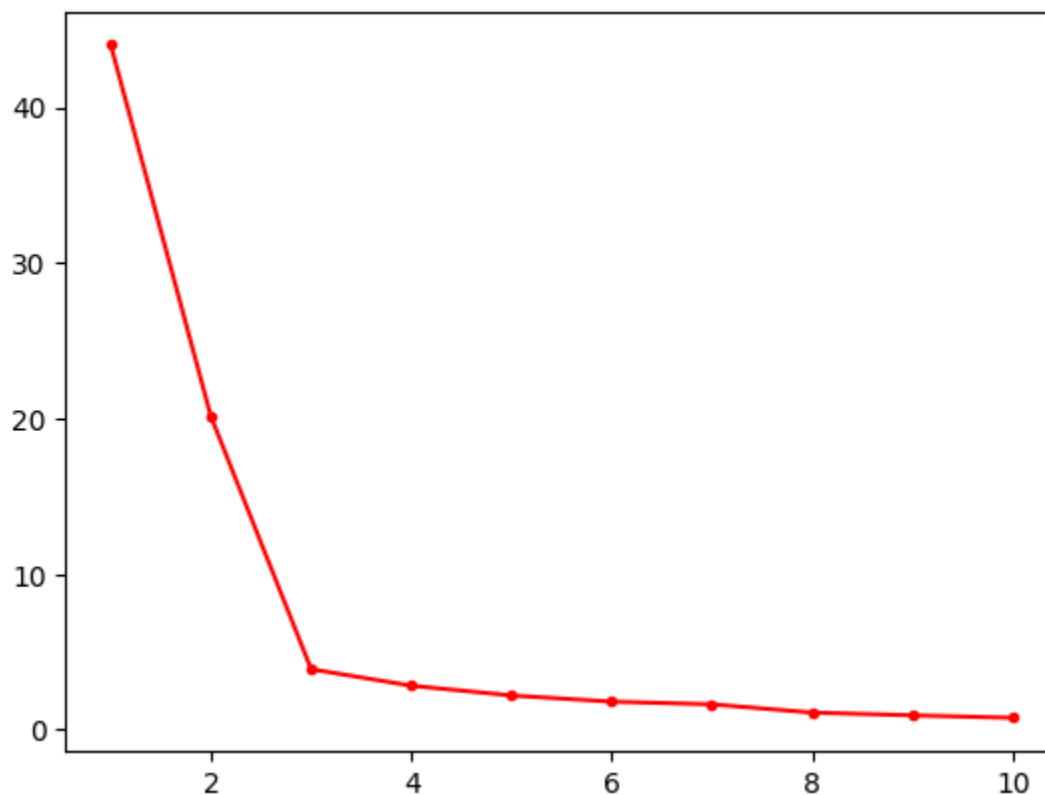
```
Out[25]: Text(0, 0.5, 'Income')
```



```
In [26]: k_range = range(1, 11)
sse = []
for k in k_range:
    kmn = KMeans(n_clusters=k)
    kmn.fit(sc_df)
    sse.append(kmn.inertia_)
```

```
In [27]: plt.plot(k_range, sse, color="r", marker=".")
```

```
Out[27]: [matplotlib.lines.Line2D at 0x71dd15d8b890]
```



```
In [28]: kmn = KMeans(n_clusters=3)
clusters = kmn.fit_predict(sc_df)
```

```
In [29]: df['clusters'] = clusters
df.head()
```

```
Out[29]:
```

	Age	Income(\$)	clusters
0	27	70000	0
1	29	90000	0
2	29	61000	0
3	28	60000	0
4	42	150000	1

```
In [30]: cl1 = df[df['clusters'] == 0]
cl2 = df[df['clusters'] == 1]
```

```
cl3 = df[df['clusters'] == 2]

centroids = scaler.inverse_transform(kmn.cluster_centers_)
```

```
In [31]: plt.title("K-means clustering of Income and Age data.")
plt.xlabel("Age")
plt.ylabel("Income($)")
plt.scatter(cl1['Age'], cl1['Income($)'], color="r", marker="s", label="Cluster 1")
plt.scatter(cl2['Age'], cl2['Income($)'], color="b", marker="+", label="Cluster 2")
plt.scatter(cl3['Age'], cl3['Income($)'], color="g", marker="v", label="Cluster 3")
plt.scatter(centroids[:, 0], centroids[:, 1], label="Centroids", s=200, marker="*",
color="black")
plt.legend()
```

Out[31]: <matplotlib.legend.Legend at 0x71dd15c24cd0>

