

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
```

```
path = 'https://raw.githubusercontent.com/NayemMustakim/AiQuest-ML-Batch-18-All-Assignments/main/Assignment%20n%20Cluster%20analysis/Mall_Customers.csv'
```

```
data = pd.read_csv(path)
```

```
data.head()
```

	CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

```
data = data.drop('CustomerID' , axis=1)
```

```
data.head()
```

	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
0	Male	19	15	39
1	Male	21	15	81
2	Female	20	16	6
3	Female	23	16	77
4	Female	31	17	40

```
data.rename(columns={'Annual Income (k$)': 'Incom', 'Spending Score (1-100)': 'Score'}, inplace=True)
```

```
data.head()
```

	Gender	Age	Incom	Score
0	Male	19	15	39
1	Male	21	15	81
2	Female	20	16	6
3	Female	23	16	77
4	Female	31	17	40

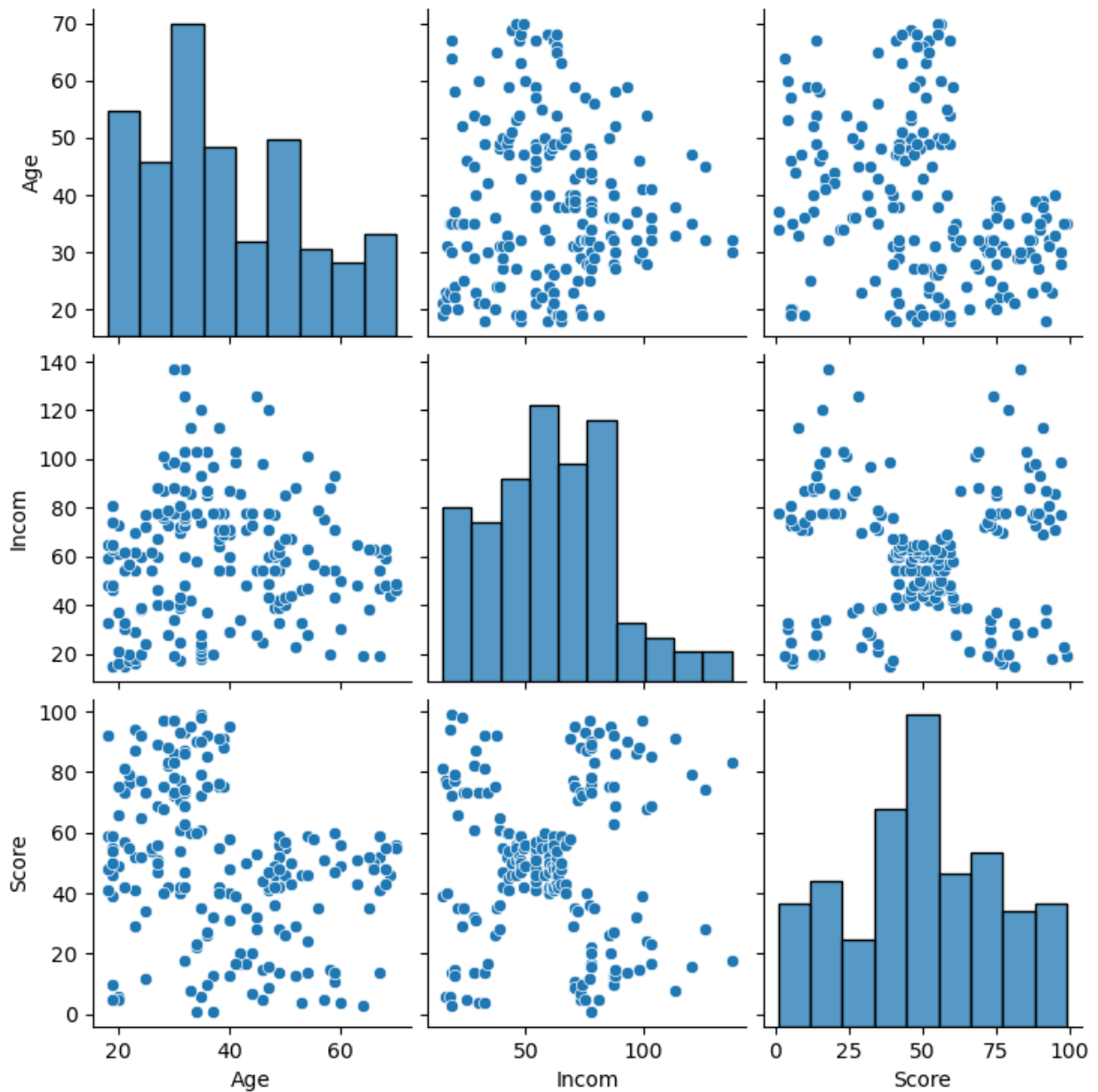
```
import seaborn as sns
```

```
import warnings
```

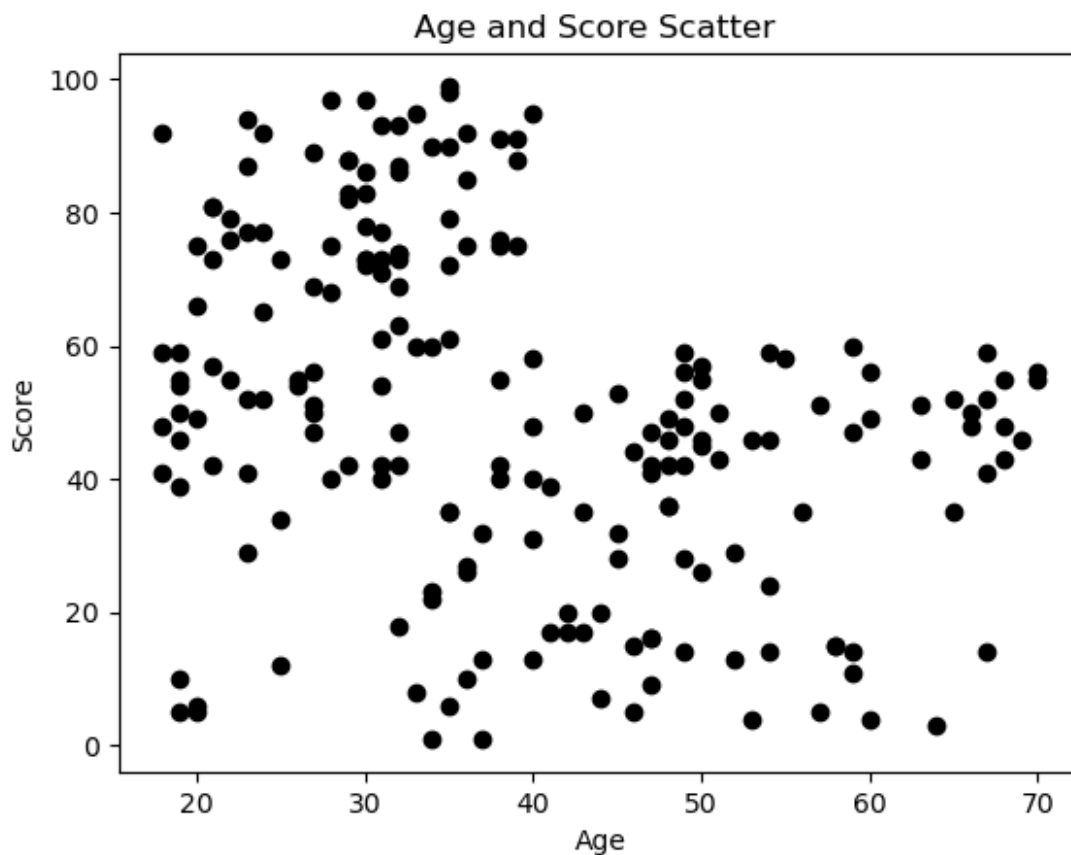
```
warnings.simplefilter('ignore')
```

```
sns.pairplot(data[['Age', 'Incom', 'Score']])
```

```
<seaborn.axisgrid.PairGrid at 0x26fe45f7a10>
```



```
plt.scatter(data.Age , data.Score,color='black')
plt.title('Age and Score Scatter ')
plt.xlabel('Age')
plt.ylabel('Score')
Text(0, 0.5, 'Score')
```



```
from kneed import KneeLocator
from sklearn.cluster import KMeans

wcss = []
for k in range(1,15):
    kms = KMeans(n_clusters=k)
    kms.fit(data[['Age', 'Score']])
    wcss.append(kms.inertia_)
```

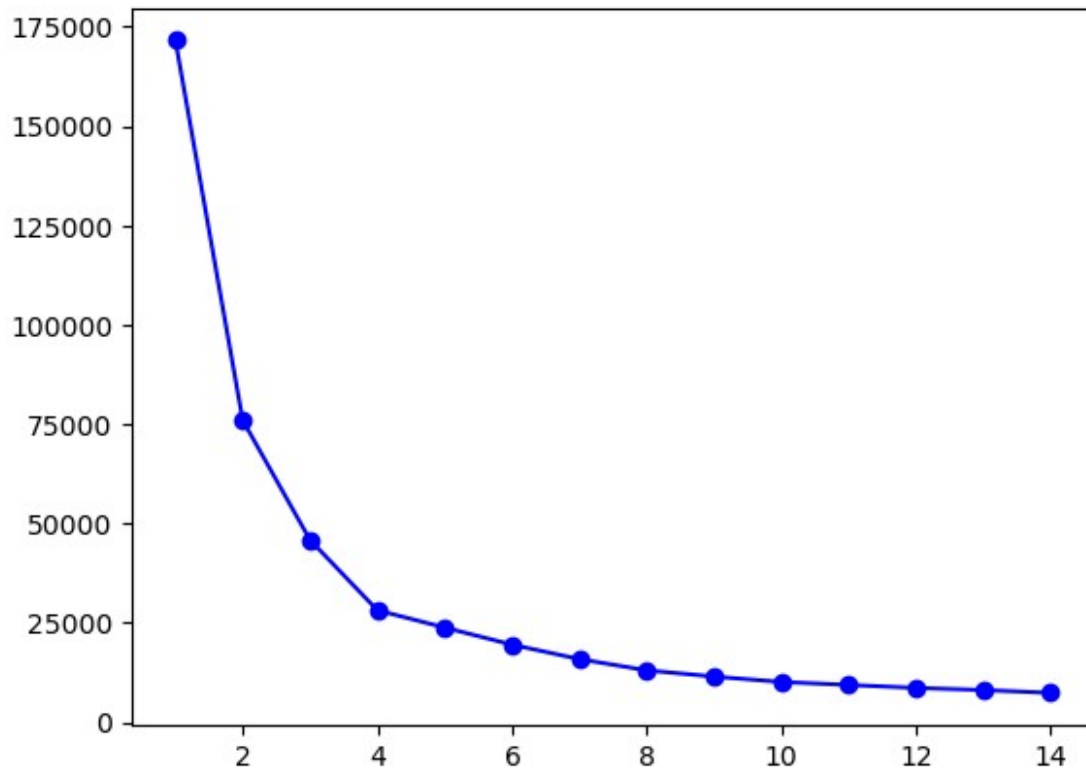
wcss

```
[171535.5,
 75949.15601023019,
 45840.67661610867,
 28165.58356662934,
 23810.46223307136,
 19506.941015125227,
 15877.41462011807,
 13050.131029357115,
 11484.814123841843,
 10131.183985072232,
 9383.343082009007,
 8631.648919553847,
```

```
8096.062582025817,  
7380.651206663706]
```

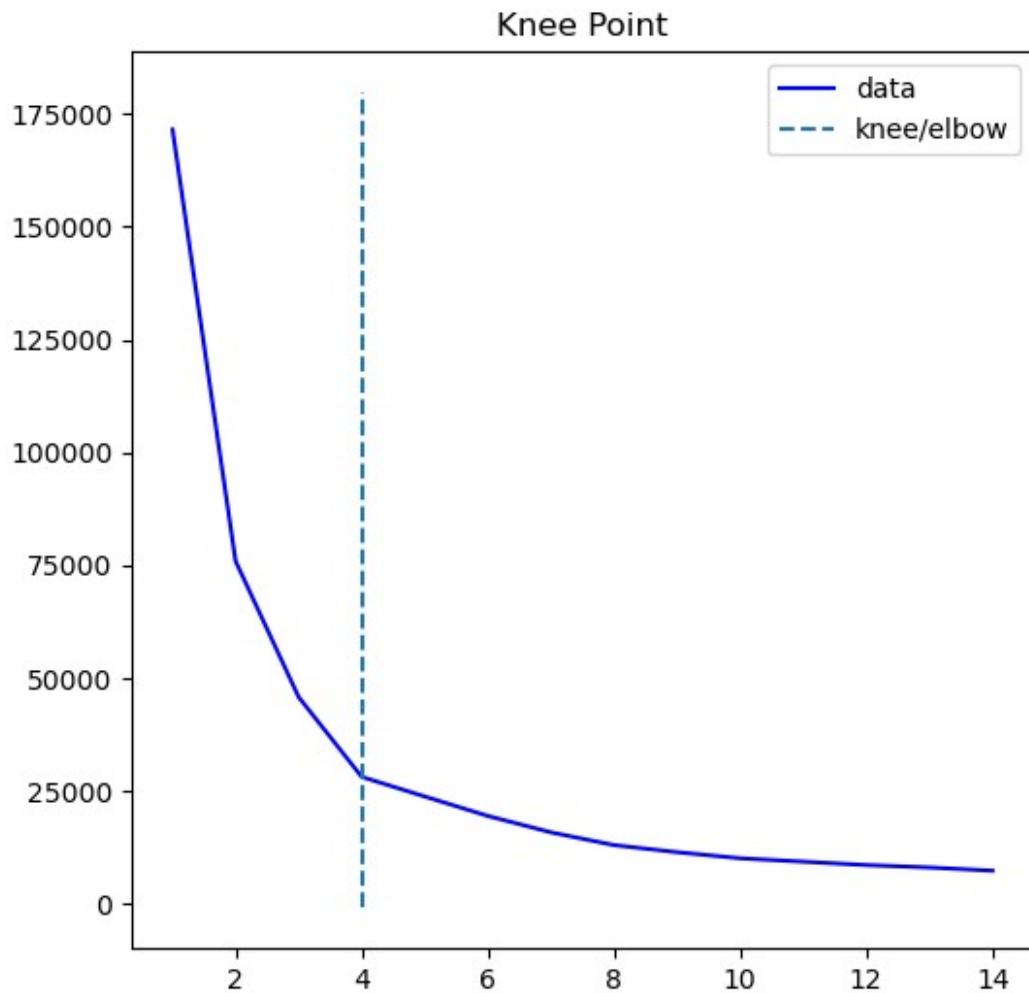
```
plt.plot(range(1,15),wcss,marker='o',color='b')
```

```
[<matplotlib.lines.Line2D at 0x26fedb3ef10>]
```



```
kn = KneeLocator(range(1,15), wcss, direction='decreasing',  
curve='convex')
```

```
kn.plot_knee()
```



```
km = KMeans(n_clusters=4)
km.fit(data[['Age', 'Score']])
KMeans(n_clusters=4)
data['clusters'] = km.predict(data[['Age', 'Score']])
data.head()
```

	Gender	Age	Incom	Score	clusters
0	Male	19	15	39	3
1	Male	21	15	81	1
2	Female	20	16	6	0
3	Female	23	16	77	1
4	Female	31	17	40	3

```
c_center = km.cluster_centers_
c_center
```

```

array([[43.29166667, 15.02083333],
       [30.1754386 , 82.35087719],
       [55.70833333, 48.22916667],
       [27.61702128, 49.14893617]])

d0 = data[data['clusters']==0]
d1 = data[data['clusters']==1]
d2 = data[data['clusters']==2]
d3 = data[data['clusters']==3]

plt.title('Age and score Cluster visualization')
plt.xlabel('Age')
plt.ylabel('Score')
plt.scatter(d0.Age , d0.Score,color = 'blue',label='Cluster0')
plt.scatter(d1.Age , d1.Score,color = 'red',label='Cluster1')
plt.scatter(d2.Age , d2.Score,color = 'orange',label='Cluster2')
plt.scatter(d3.Age , d3.Score,color = 'magenta',label='Cluster3')
plt.scatter(c_center[:,0] , c_center[:,1], color='black',
label='centroids')
plt.legend()

<matplotlib.legend.Legend at 0x26fef16c750>

```

