In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:

```
df=pd.read_csv('D:/Users/Janhavi/TE/TE1/ML/Mall_Customers.csv')
```

In [3]:

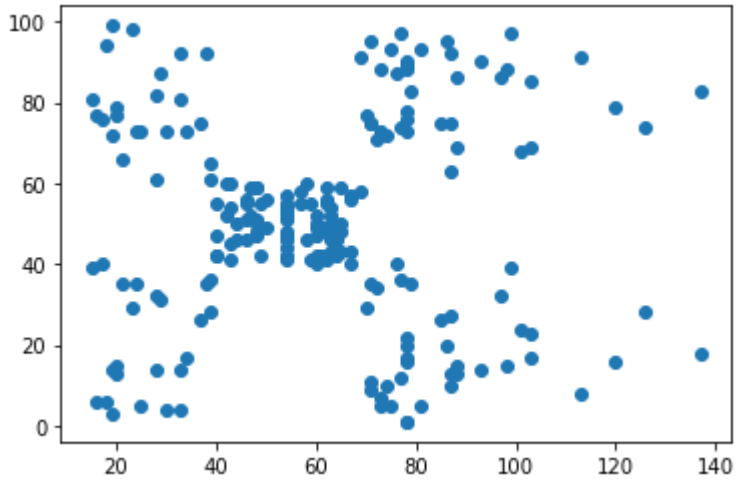
```
df.shape   # stores the number of rows and columns as a tuple (number of rows, number of columns)
```

Out[3]:

```
(200, 5)
```

In [4]:

```
import matplotlib.pyplot as plt
plt.scatter(df['Annual Income (k$)'],df['Spending Score (1-100)'])
plt.show()
```




In [5]:

```
X=df[['Annual Income (k$)','Spending Score (1-100)']].values
```

In [6]:

```
X
```

Out[6]:

```
array([[ 15,  39],
       [ 15,  81],
       [ 16,   6],
       [ 16,  77],
       [ 17,  40],
       [ 17,  76],
       [ 18,   6],
       [ 18,  94],
       [ 19,   3],
       [ 19,  72],
       [ 19,  14],
       [ 19,  99],
       [ 20,  15],
       [ 20,  77],
       [ 20,  13],
       [ 20,  79],
       [ 21,  35],
       [ 21,  66],
       [ 23,  29],
       [ 23,  98],
       [ 24,  35],
       [ 24,  73],
       [ 25,   5],
       [ 25,  73],
       [ 28,  14],
       [ 28,  82],
       [ 28,  32],
       [ 28,  61],
       [ 29,  31],
       [ 29,  87],
       [ 30,   4],
       [ 30,  73],
       [ 33,   4],
       [ 33,  92],
       [ 33,  14],
       [ 33,  81],
       [ 34,  17],
       [ 34,  73],
       [ 37,  26],
       [ 37,  75],
       [ 38,  35],
       [ 38,  92],
       [ 39,  36],
       [ 39,  61],
       [ 39,  28],
       [ 39,  65],
       [ 40,  55],
       [ 40,  47],
       [ 40,  42],
       [ 40,  42],
       [ 42,  52],
       [ 42,  60],
       [ 43,  54],
       [ 43,  60],
       [ 43,  45],
       [ 43,  41],
       [ 44,  50],
       [ 44,  46],
       [ 46,  51],
```

```
       [ 46,  46],
       [ 46,  56],
       [ 46,  55],
       [ 47,  52],
       [ 47,  59],
       [ 48,  51],
       [ 48,  59],
       [ 48,  50],
       [ 48,  48],
       [ 48,  59],
       [ 48,  47],
       [ 49,  55],
       [ 49,  42],
       [ 50,  49],
       [ 50,  56],
       [ 54,  47],
       [ 54,  54],
       [ 54,  53],
       [ 54,  48],
       [ 54,  52],
       [ 54,  42],
       [ 54,  51],
       [ 54,  55],
       [ 54,  41],
       [ 54,  44],
       [ 54,  57],
       [ 54,  46],
       [ 57,  58],
       [ 57,  55],
       [ 58,  60],
       [ 58,  46],
       [ 59,  55],
       [ 59,  41],
       [ 60,  49],
       [ 60,  40],
       [ 60,  42],
       [ 60,  52],
       [ 60,  47],
       [ 60,  50],
       [ 61,  42],
       [ 61,  49],
       [ 62,  41],
       [ 62,  48],
       [ 62,  59],
       [ 62,  55],
       [ 62,  56],
       [ 62,  42],
       [ 63,  50],
       [ 63,  46],
       [ 63,  43],
       [ 63,  48],
       [ 63,  52],
       [ 63,  54],
       [ 64,  42],
       [ 64,  46],
       [ 65,  48],
       [ 65,  50],
       [ 65,  43],
       [ 65,  59],
       [ 67,  43],
       [ 67,  57],
```

```
        [ 67,  56],
        [ 67,  40],
        [ 69,  58],
        [ 69,  91],
        [ 70,  29],
        [ 70,  77],
        [ 71,  35],
        [ 71,  95],
        [ 71,  11],
        [ 71,  75],
        [ 71,   9],
        [ 71,  75],
        [ 72,  34],
        [ 72,  71],
        [ 73,   5],
        [ 73,  88],
        [ 73,   7],
        [ 73,  73],
        [ 74,  10],
        [ 74,  72],
        [ 75,   5],
        [ 75,  93],
        [ 76,  40],
        [ 76,  87],
        [ 77,  12],
        [ 77,  97],
        [ 77,  36],
        [ 77,  74],
        [ 78,  22],
        [ 78,  90],
        [ 78,  17],
        [ 78,  88],
        [ 78,  20],
        [ 78,  76],
        [ 78,  16],
        [ 78,  89],
        [ 78,   1],
        [ 78,  78],
        [ 78,   1],
        [ 78,  73],
        [ 79,  35],
        [ 79,  83],
        [ 81,   5],
        [ 81,  93],
        [ 85,  26],
        [ 85,  75],
        [ 86,  20],
        [ 86,  95],
        [ 87,  27],
        [ 87,  63],
        [ 87,  13],
        [ 87,  75],
        [ 87,  10],
        [ 87,  92],
        [ 88,  13],
        [ 88,  86],
        [ 88,  15],
        [ 88,  69],
        [ 93,  14],
        [ 93,  90],
        [ 97,  32],
```

```
       [ 97,  86],
       [ 98,  15],
       [ 98,  88],
       [ 99,  39],
       [ 99,  97],
       [101,  24],
       [101,  68],
       [103,  17],
       [103,  85],
       [103,  23],
       [103,  69],
       [113,   8],
       [113,  91],
       [120,  16],
       [120,  79],
       [126,  28],
       [126,  74],
       [137,  18],
       [137,  83]], dtype=int64)
```

In [7]:

```
import numpy as np
np.sqrt(200)
```

Out[7]:

```
14.142135623730951
```

In [8]:

```
from sklearn.cluster import KMeans
k=range(1,15)
sse=[]
for i in k:
    model_demo=KMeans(n_clusters=i,random_state=0)
    model_demo.fit(X)
    sse.append(model_demo.inertia_)
```

In [9]:

```
k=5
model=KMeans(n_clusters=k,random_state=0)
model.fit(X)
Y=model.predict(X)
Y
```

Out[9]:

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

In [10]:

```
model.cluster_centers_
```

Out[10]:

```
array([[55.2962963 , 49.51851852],
       [25.72727273, 79.36363636],
       [86.53846154, 82.12820513],
       [26.30434783, 20.91304348],
       [88.2       , 17.11428571]])
```

In [11]:

```
model.cluster_centers_[0]
```

Out[11]:

```
array([55.2962963 , 49.51851852])
```

In [12]:

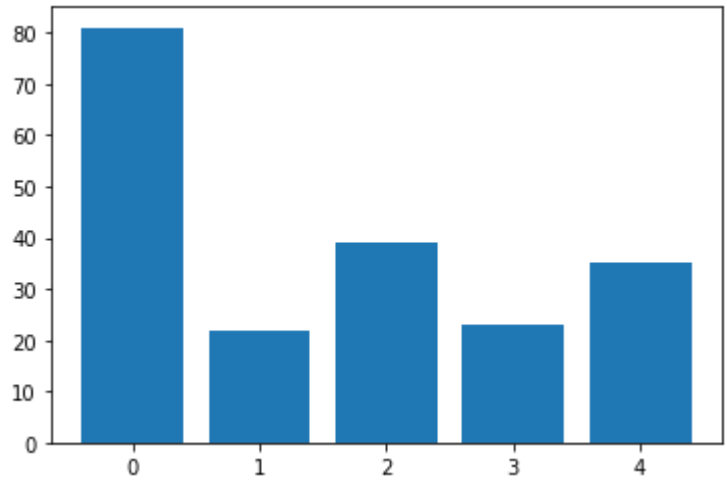
```
import numpy as np
u, c = np.unique(Y, return_counts = True)
```

In [13]:

```
plt.bar(u,c)
```

Out[13]:

```
<BarContainer object of 5 artists>
```




In [14]:

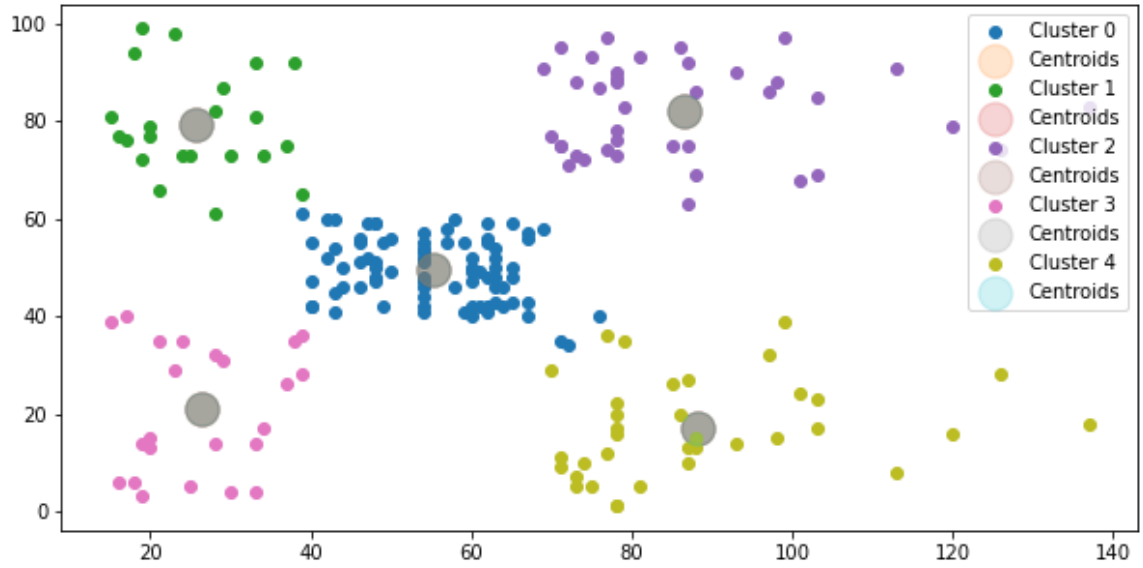
```
np.sum(c)
```

Out[14]:

```
200
```

In [15]:

```
plt.figure(figsize = (10, 5))
for i in range (k):
  plt.scatter(X[Y==i, 0], X[Y==i, 1], label=f'Cluster {i}')
  plt.scatter(model.cluster_centers_[:,0], model.cluster_centers_[:,1], s=300, alpha=0.
2, label= 'Centroids')
  plt.legend()
```



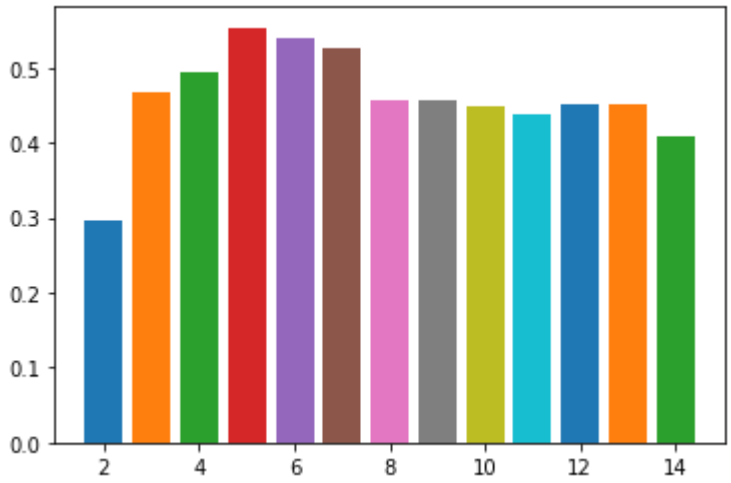

In [16]:

```
from sklearn.metrics import silhouette_score
```

In [17]:

```
k = range(2, 15)
for i in k:
  model_demo = KMeans(n_clusters=i, random_state=0)
  model_demo.fit(X)
  Y = model_demo.predict(X)
  print(f'{i} Clusters = {silhouette_score(X, Y) : 2f}')
  plt.bar(i,silhouette_score(X, Y))
  # plt.show()
```

```
2 Clusters =  0.296897
3 Clusters =  0.467614
4 Clusters =  0.493196
5 Clusters =  0.553932
6 Clusters =  0.539392
7 Clusters =  0.527029
8 Clusters =  0.457569
9 Clusters =  0.456508
10 Clusters =  0.449795
11 Clusters =  0.437842
12 Clusters =  0.450331
13 Clusters =  0.451642
14 Clusters =  0.408016
```




In [ ]: