In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```python
df = pd.read_csv('Mall_Customers.csv')
print(df.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
```

In [3]:

```python
# clean the data
df = df.drop(['CustomerID', 'Age', 'Gender'], axis=1)
print(df.head())
```

```
   Annual Income (k$)  Spending Score (1-100)
0                  15                      39
1                  15                      81
2                  16                       6
3                  16                      77
4                  17                      40
```

In [4]:

```python
### create the clusters
from sklearn.cluster import AgglomerativeClustering

# create the hiearchical model
hiearchical = AgglomerativeClustering(n_clusters=5)
```

In [5]:

```python
# fit the data into the clusters
clusters = hiearchical.fit_predict(df)
print(clusters)
```

```
[4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4 3 4
 3 4 3 4 3 4 1 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 0 2 0 2 1 2 0 2 0 2 0 2 0 2 1 2 0 2 1 2
 0 2 0 2 0 2 0 2 0 2 0 2 1 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2 0
 2 0 2 0 2 0 2 0 2 0 2 0 2 0 2]
```

In [6]:

```python
print(hiearchical.children_)
```

```
[[ 65  68]
 [ 48  49]
 [156 158]
 [129 131]
 [ 21  23]
 [ 51  53]
 [ 60  61]
 [ 67  69]
 [ 64  66]
 [ 79  82]
 [ 74  85]
 [ 78  80]
 [ 75  81]
 [ 92  97]
 [101 109]
 [107 113]
 [ 94  98]
 [103 104]
 [100 105]
 [150 154]
 [119 120]
 [170 174]
 [151 155]
 [ 63 200]
 [ 99 213]
 [  3   5]
 [ 26  28]
 [ 40  42]
 [ 10  14]
 [ 54  57]
 [ 58  62]
 [ 70  73]
 [ 91  93]
 [108 112]
 [126 132]
 [137 139]
 [147 159]
 [ 77 210]
 [ 76 212]
 [149 222]
 [ 12 228]
 [116 233]
 [ 13  15]
 [  2   6]
 [ 87  90]
 [106 110]
 [114 115]
 [128 130]
 [134 136]
 [148 152]
 [153 157]
 [165 171]
 [111 217]
 [  0   4]
 [ 50  52]
 [ 86  88]
 [ 89  96]
 [146 160]
 [216 218]
 [164 168]
 [181 183]
 [187 191]
 [186 190]
 [176 221]
 [133 235]
 [208 230]
 [140 248]
 [125 203]
 [ 38  44]
 [ 83 209]
 [ 59 207]
 [122 220]
 [ 16  20]
 [ 30  32]
 [118 121]
 [102 117]
 [ 34  36]
 [214 215]
 [135 143]
 [167 173]
 [ 95 224]
 [138 247]
 [232 258]
 [ 37  39]
 [ 55 201]
 [ 84 238]
 [ 46 254]
 [ 43  45]
 [245 246]
 [ 72 270]
 [ 11  19]
 [  8 243]
 [ 56 265]
 [ 47 229]
 [123 127]
 [141 145]
 [172 263]
 [206 231]
 [241 274]
 [ 33  41]
 [236 250]
 [ 25  29]
 [178 182]
 [211 285]
 [144 219]
 [225 242]
 [142 257]
 [244 255]
 [162 202]
 [239 278]
 [256 280]
 [169 177]
 [ 71 269]
 [163 295]
 [  7 290]
 [179 260]
 [  1 305]
 [277 288]
 [  9  17]
 [ 31 204]
 [252 275]
 [223 297]
 [205 287]
 [ 35 301]
 [124 234]
 [ 24 276]
 [ 18 272]
 [180 184]
 [166 259]
 [264 267]
 [289 292]
 [ 22 273]
 [188 262]
 [161 309]
 [195 197]
 [284 293]
 [266 281]
 [175 279]
 [237 312]
 [249 304]
 [189 315]
 [226 326]
 [271 320]
 [310 317]
 [192 194]
 [282 298]
 [303 307]
 [318 319]
 [227 268]
 [294 313]
 [185 340]
 [251 311]
 [306 324]
 [300 329]
 [296 302]
 [286 321]
 [ 27 347]
 [196 198]
 [333 349]
 [299 323]
 [240 291]
 [308 336]
 [199 334]
 [322 355]
 [325 331]
 [253 341]
 [193 350]
 [283 356]
 [328 339]
 [330 335]
 [327 332]
 [337 358]
 [343 345]
 [261 351]
 [344 357]
 [338 369]
 [342 346]
 [360 364]
 [316 367]
 [354 368]
 [314 359]
 [348 365]
 [372 376]
 [363 375]
 [361 379]
 [353 373]
 [370 374]
 [378 380]
 [362 366]
 [352 382]
 [371 385]
 [377 381]
 [383 389]
 [388 390]
 [384 386]
 [391 392]
 [387 395]
 [393 394]
 [396 397]]
```

visualization using KMeans method

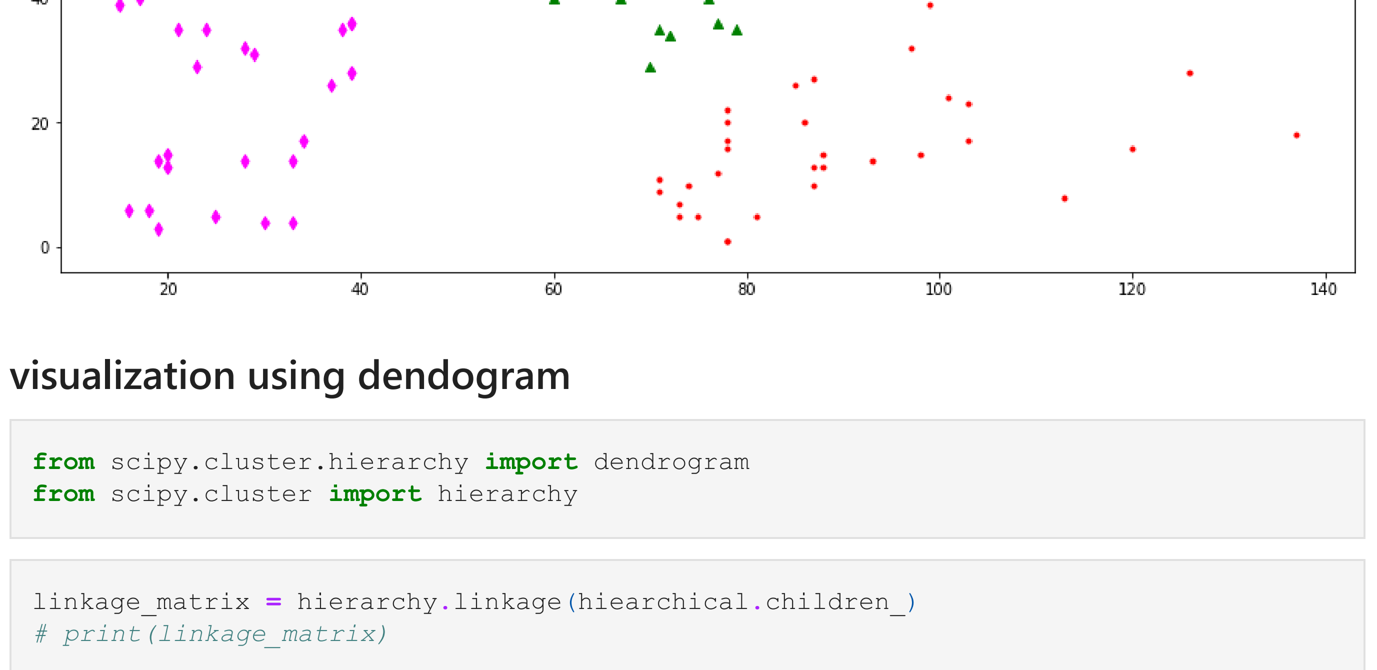
In [7]:

```python
plt.figure(figsize=(15, 8))
colors = ['red', 'green', 'blue', 'brown', 'magenta']
markers = ['.', '^', '1', 's', 'd']

# create clusters
for k in range(5):
    plt.scatter(
        df['Annual Income (k$)'][clusters == k],
        df['Spending Score (1-100)'][clusters == k],
        label=f"cluster {k}",
        color=colors[k],
        marker=markers[k])

plt.legend()
```

Out[7]: <matplotlib.legend.Legend at 0x1469637a520>



visualization using dendogram

In [8]:

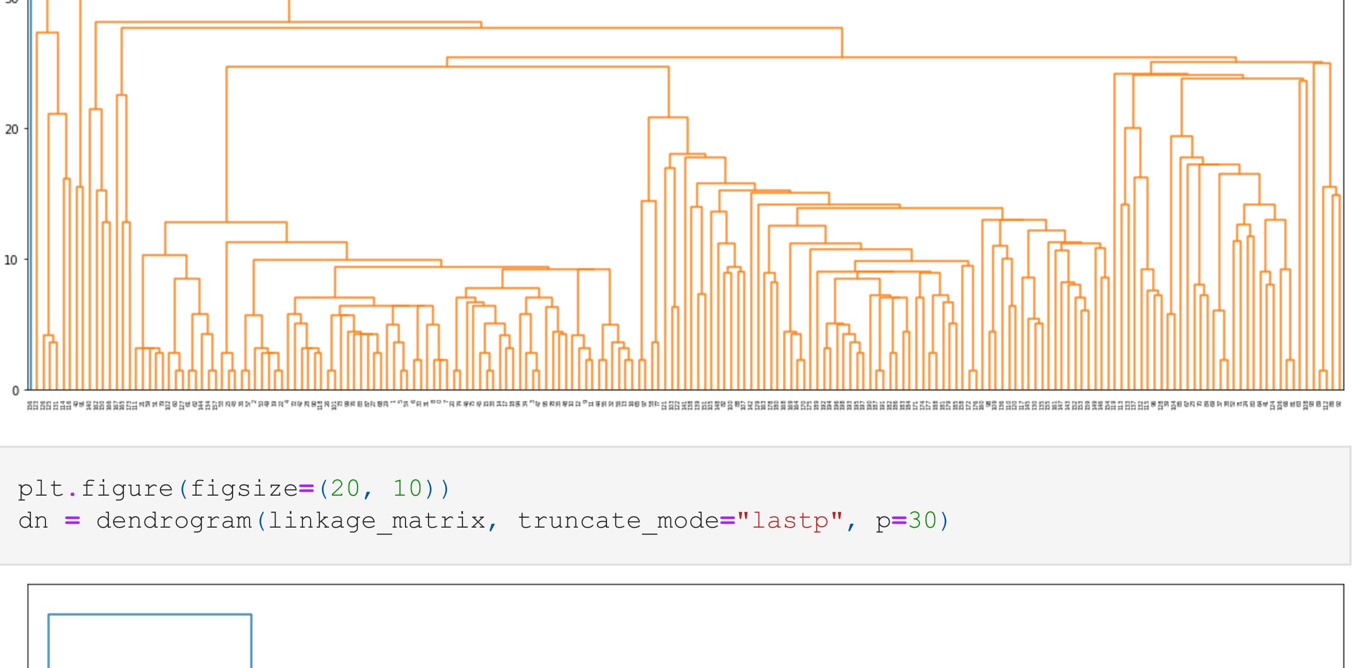
```python
from scipy.cluster.hierarchy import dendrogram
from scipy.cluster import hierarchy
```

In [9]:

```python
linkage_matrix = hierarchy.linkage(hiearchical.children_)
# print(linkage_matrix)
```

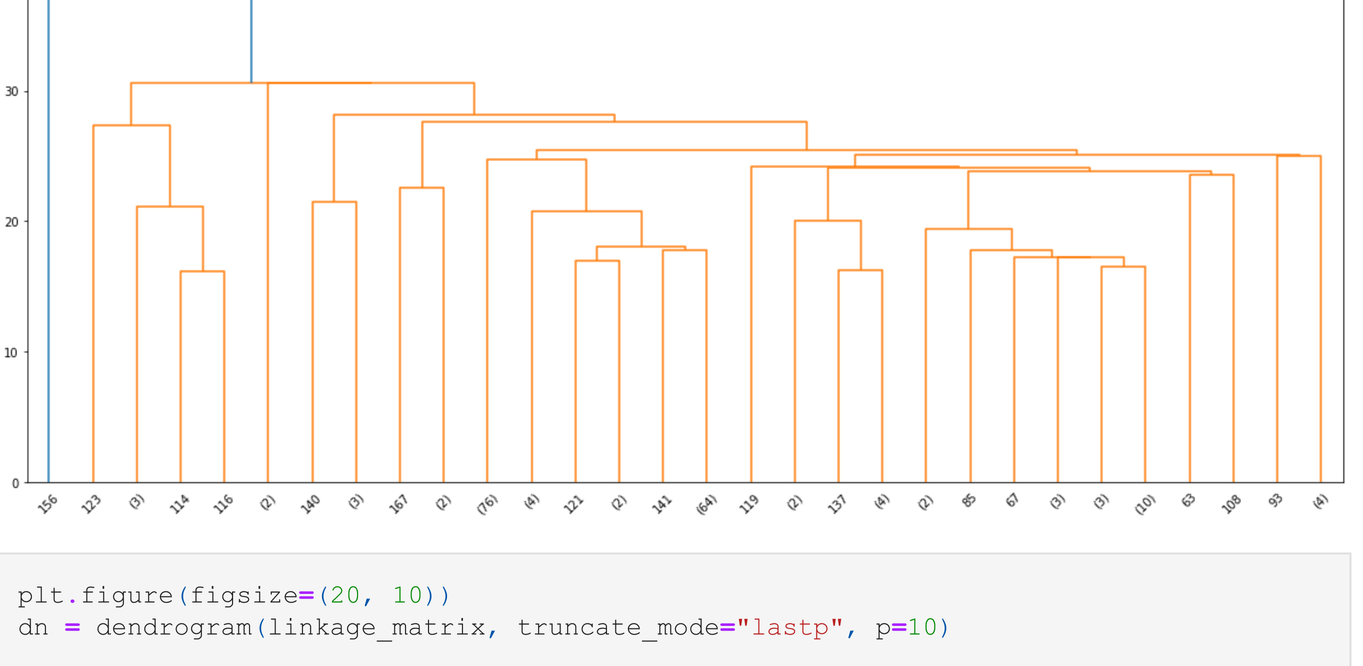
In [10]:

```python
plt.figure(figsize=(20, 10))
dn = dendrogram(linkage_matrix)
```



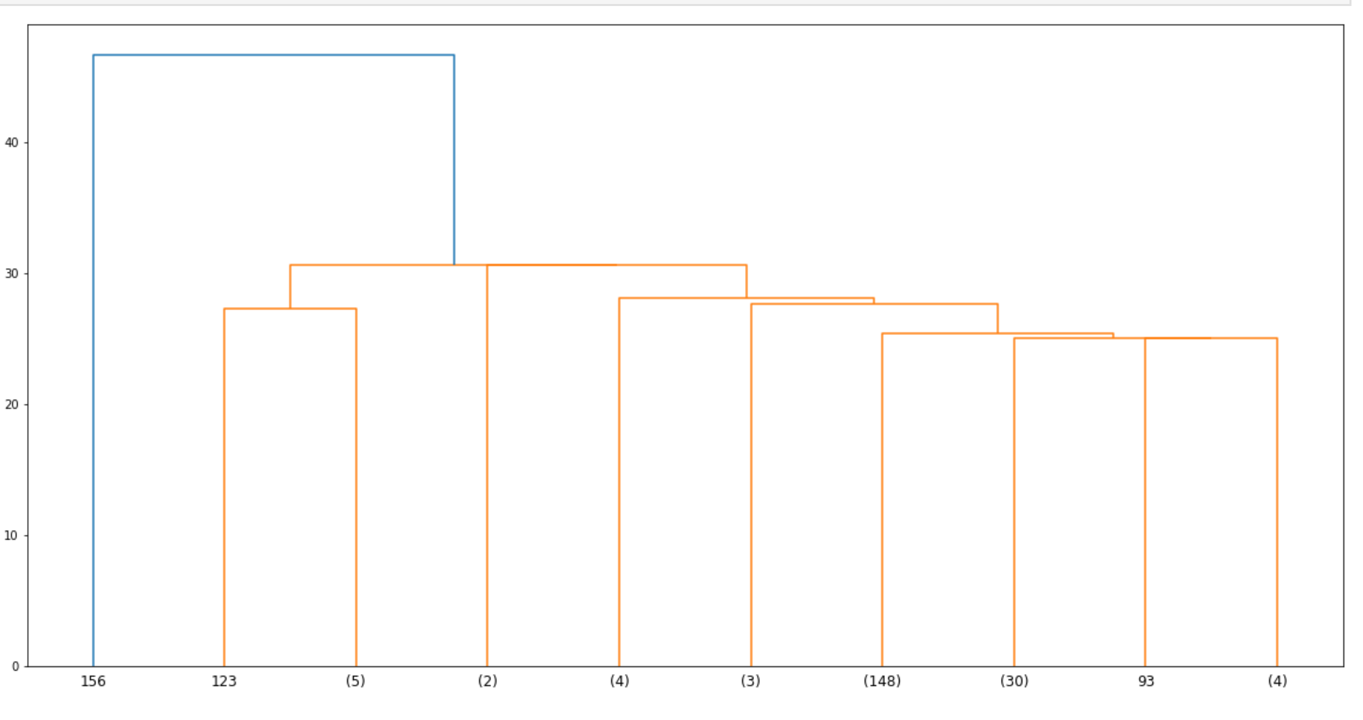
In [11]:

```python
plt.figure(figsize=(20, 10))
dn = dendrogram(linkage_matrix, truncate_mode="lastp", p=30)
```



In [12]:

```python
plt.figure(figsize=(20, 10))
dn = dendrogram(linkage_matrix, truncate_mode="lastp", p=10)
```



In [ ]: