Customer_Segmentation_Using_Python

In [26]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.cluster.hierarchy as sch
```

Import and Understand the data

In [27]:
```
data = pd.read_csv('Mall_Customers.csv')
```

In [28]:
```
data.head()
```

Out[28]:

	CustomerID	Genre	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

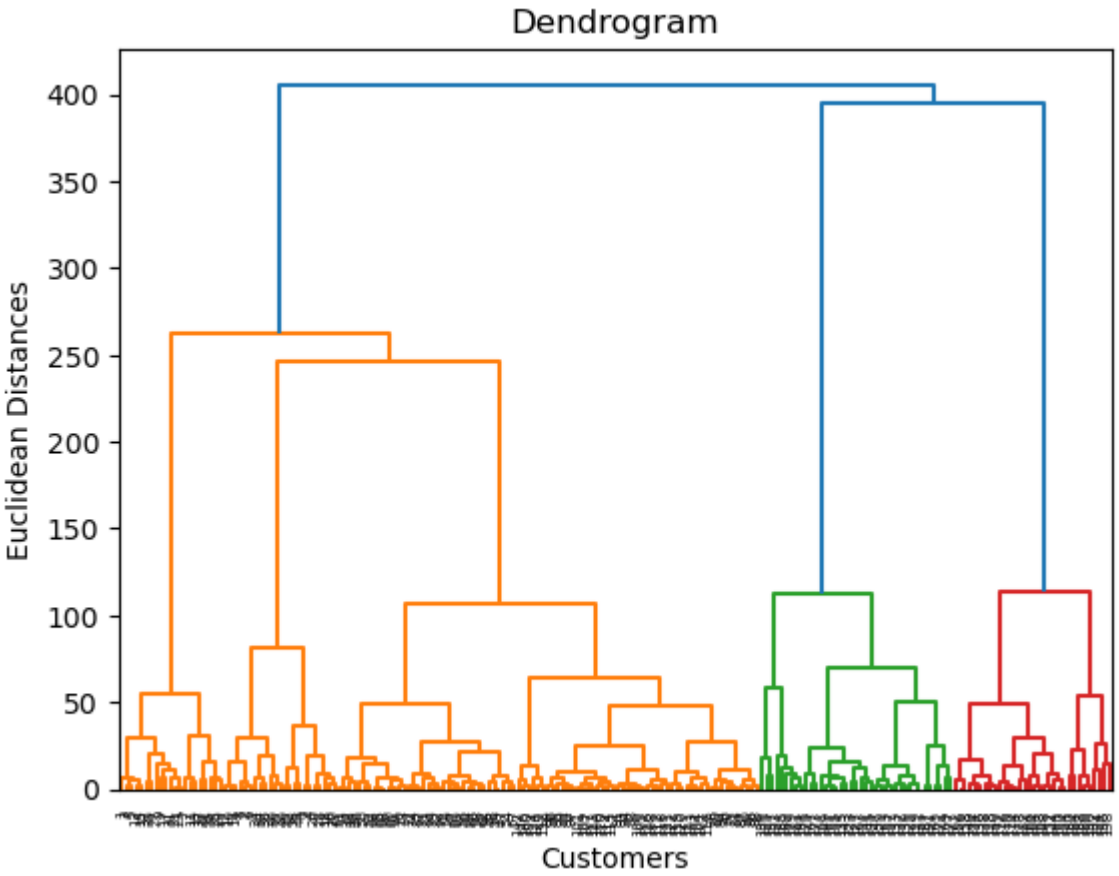
Extract relevant features for segmentation

In [33]:
```
X = data.iloc[:, [3, 4]].values
```

In [35]:
```
# X
```

Use Dendrogram to find optimal number of clusters

In [36]:
```
dendrogram = sch.dendrogram(sch.linkage(X, method='ward'))
plt.title('Dendrogram')
plt.xlabel('Customers')
plt.ylabel('Euclidean Distances')
plt.show()
```



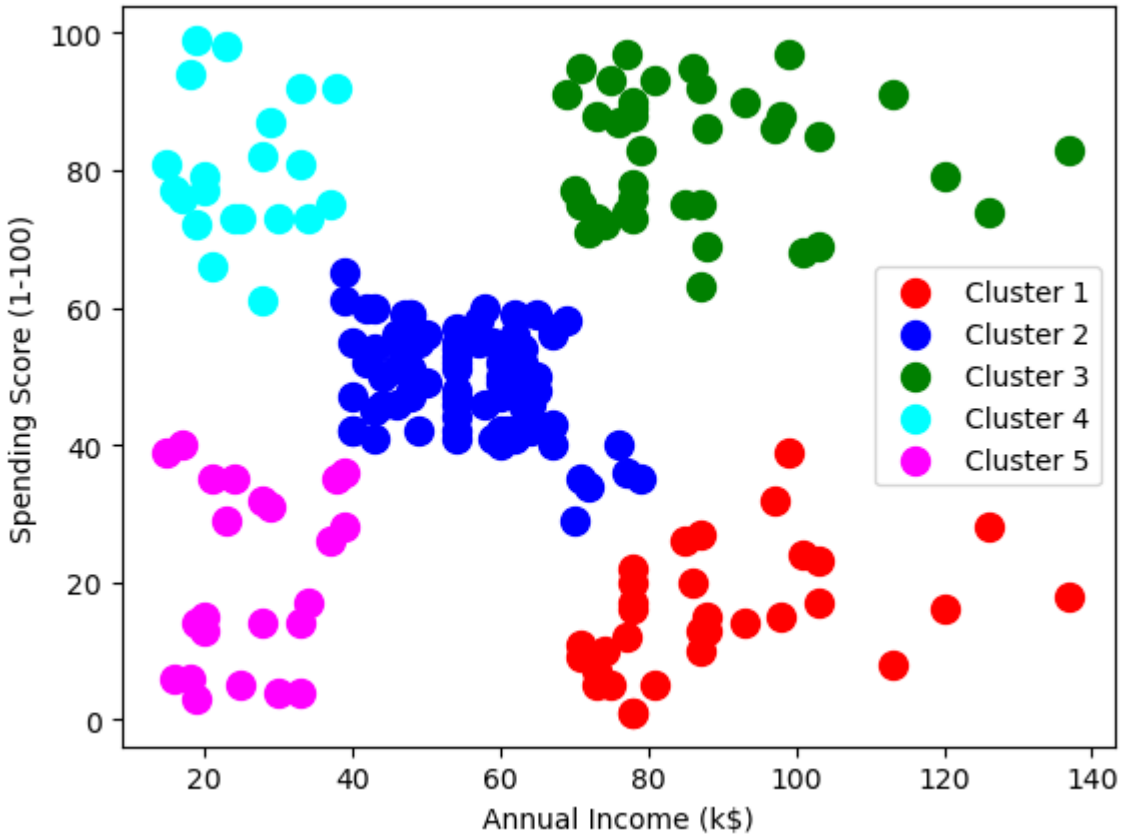
Perform Hierarchical Clustering

In [37]:
```
from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters=5, affinity='euclidean', linkage='wa
y_hc = hc.fit_predict(X)
```

```
C:\Users\baps\anaconda3\lib\site-packages\sklearn\cluster\_agglomerative.p
y:1005: FutureWarning: Attribute `affinity` was deprecated in version 1.2
and will be removed in 1.4. Use `metric` instead
  warnings.warn(
```

Visualize the clusters

In [38]:
```
plt.scatter(X[y_hc == 0, 0], X[y_hc == 0, 1], s=100, c='red', label='Cluster
plt.scatter(X[y_hc == 1, 0], X[y_hc == 1, 1], s=100, c='blue', label='Cluste
plt.scatter(X[y_hc == 2, 0], X[y_hc == 2, 1], s=100, c='green', label='Clust
plt.scatter(X[y_hc == 3, 0], X[y_hc == 3, 1], s=100, c='cyan', label='Cluste
plt.scatter(X[y_hc == 4, 0], X[y_hc == 4, 1], s=100, c='magenta', label='Clu
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.show()
```



THANK YOU!!!

GitHub: https://github.com/anujtiwari21?tab=repositories
[(https://github.com/anujtiwari21?tab=repositories)](https://github.com/anujtiwari21?tab=repositories)