Customer Segmentation Using Python

Importing Modules

```
In [1]:  import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
         import seaborn as sns
         import scipy.cluster.hierarchy as sch

         import warnings
         warnings.filterwarnings('ignore')
```

```
In [2]:  data = pd.read_csv('Mall_Customers.csv')
         data.head()
```

Out[2]:

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

Use Dendrogram to find optimal number of clusters

```
In [4]:  dendrogram = sch.dendrogram(sch.linkage(X, method='ward'))
         plt.title('Dendrogram')
         plt.xlabel('Customers')
         plt.ylabel('Euclidean Distance')
         plt.show()
```

Dendrogram
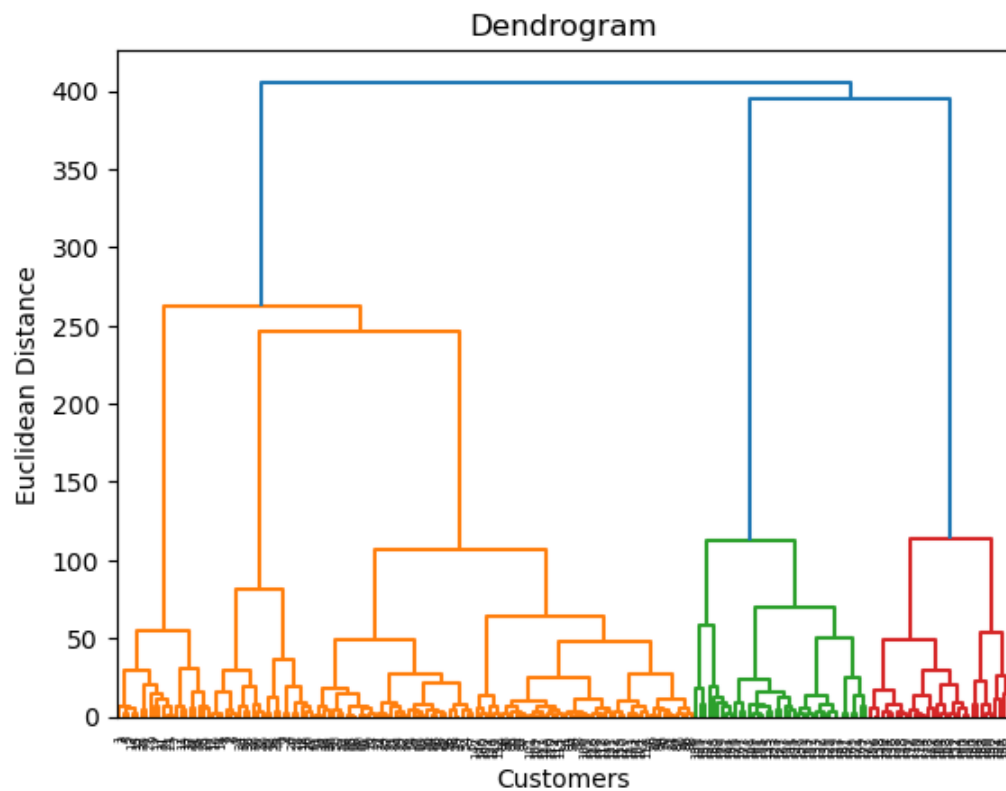
Perform Hierarchical Clustering

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

```
In [6]: plt.scatter(X[y_hc == 0, 0], X[y_hc == 0, 1], s=100, c='red', label='Cluster 1')
        plt.scatter(X[y_hc == 1, 0], X[y_hc == 1, 1], s=100, c='blue', label='Cluster 2')
        plt.scatter(X[y_hc == 2, 0], X[y_hc == 2, 1], s=100, c='green', label='Cluster 3')
        plt.scatter(X[y_hc == 3, 0], X[y_hc == 3, 1], s=100, c='cyan', label='Cluster 4')
        plt.scatter(X[y_hc == 4, 0], X[y_hc == 4, 1], s=100, c='magenta', label='Cluster 5')
        plt.xlabel('Annual Income (k$)')
        plt.ylabel('Spending Score (1-100)')
        plt.legend()
        plt.show()
```

You can Find Project on **GitHub.**