In [1]:
```
import sklearn
import pandas as pd
import numpy as np
import scipy.cluster.hierarchy as sch
from sklearn import metrics
from scipy.cluster.hierarchy import linkage, fcluster
from sklearn.metrics import silhouette_score
from sklearn.preprocessing import StandardScaler
from matplotlib import pyplot as plt
```

In [2]:
```
dataset = pd.read_csv("churn_clean.csv", sep=",")
```

In [3]:
```
X = dataset[['Age', 'Contacts', 'Children', 'Tenure']]
scaler = StandardScaler()
scaled = scaler.fit_transform(X)
print(scaled)
```

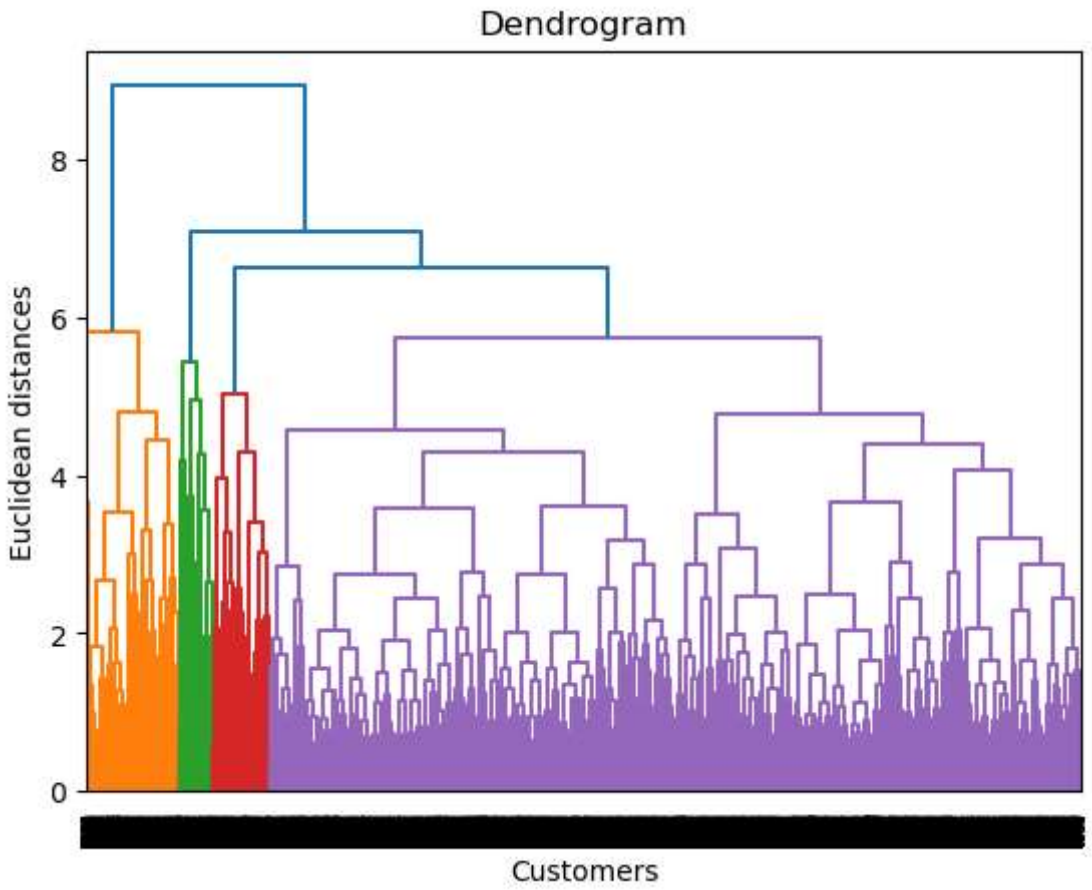
```
[[ 0.72092524 -1.0058517  -0.97233791 -1.04874621]
 [-1.25995716 -1.0058517  -0.50659192 -1.26200116]
 [-0.14873045 -1.0058517   0.89064606 -0.7099398 ]
 ...
 [-0.24535886 -1.0058517  -0.50659192  0.48751337]
 [-0.6801867   0.00586797 -0.50659192  1.38301834]
 [-1.21164295  0.00586797 -0.50659192  1.09012007]]
```

In [4]:
```
#check for any missing values
percent_missing =round(100*(X.isnull().sum())/len(X),2)
percent_missing
```

Out[4]:
```
Age         0.0
Contacts    0.0
Children    0.0
Tenure      0.0
dtype: float64
```

In [5]:
```
pd.DataFrame(scaled).to_csv("churn_clean_scaled.csv")
```

In [6]:
```
dendro = sch.dendrogram(sch.linkage(scaled, method = 'complete'))
plt.title('Dendrogram')
plt.xlabel('Customers')
plt.ylabel('Euclidean distances')
plt.show()
```



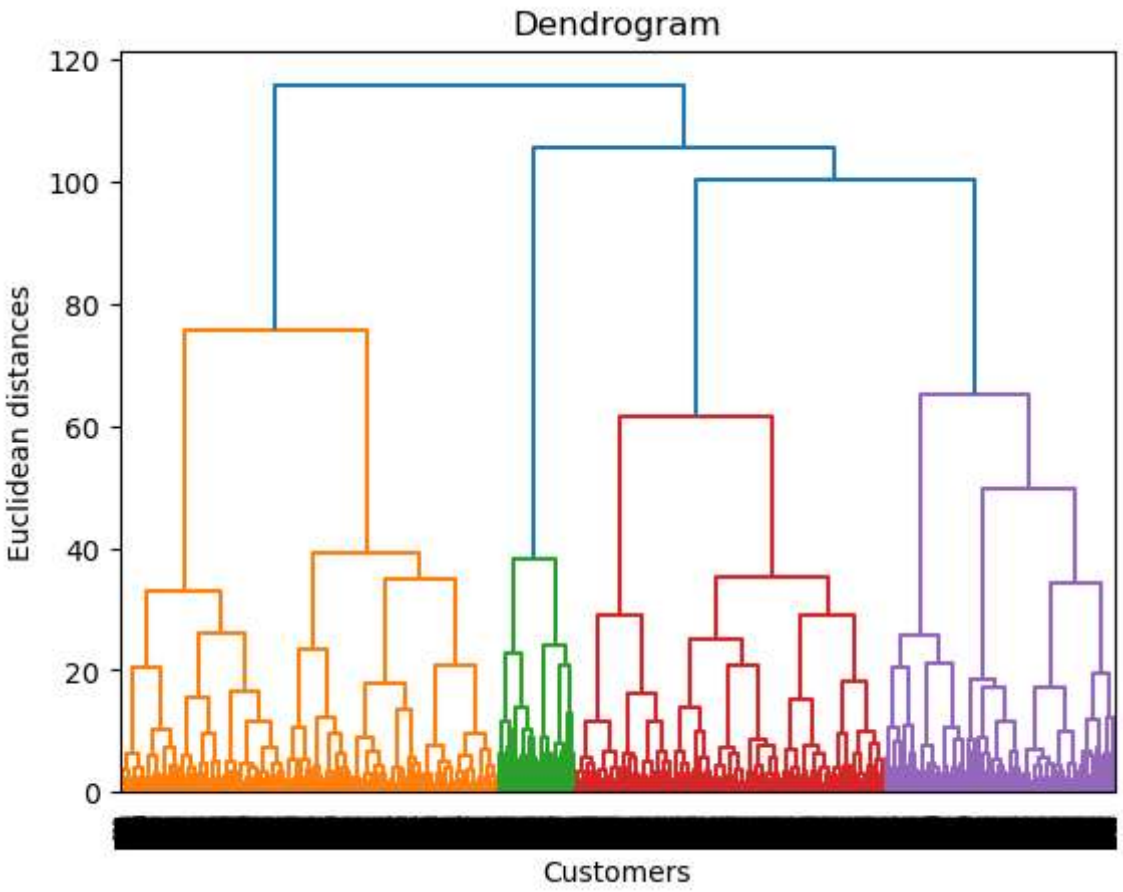
In [7]:
```
mergings = linkage(X, method='complete')
labels = fcluster(mergings, 7, criterion='distance')
print(labels)
```

```
[ 66  95 147 ... 329 240 225]
```

In [8]:
```
metrics.silhouette_score(X, labels, metric='euclidean')
```

Out[8]:
```
0.1691114042628263
```

In [9]:
```
#Using 'ward' method
dendro = sch.dendrogram(sch.linkage(scaled, method = 'ward'))
plt.title('Dendrogram')
plt.xlabel('Customers')
plt.ylabel('Euclidean distances')
plt.show()
```



In [10]:
```
mergings = linkage(X, method='ward')
labels = fcluster(mergings, 110, criterion='distance')
print(labels)
```

```
[27 28 34 ... 18 17 12]
```

In [11]:
```
metrics.silhouette_score(X, labels, metric='euclidean')
```

Out[11]:
```
0.2245383331363044
```

In [12]:
```
# https://scikit-learn.org/stable/modules/clustering.html#silhouette-coefficient

# https://campus.datacamp.com/courses/unsupervised-learning-in-python/visualization-wi
```