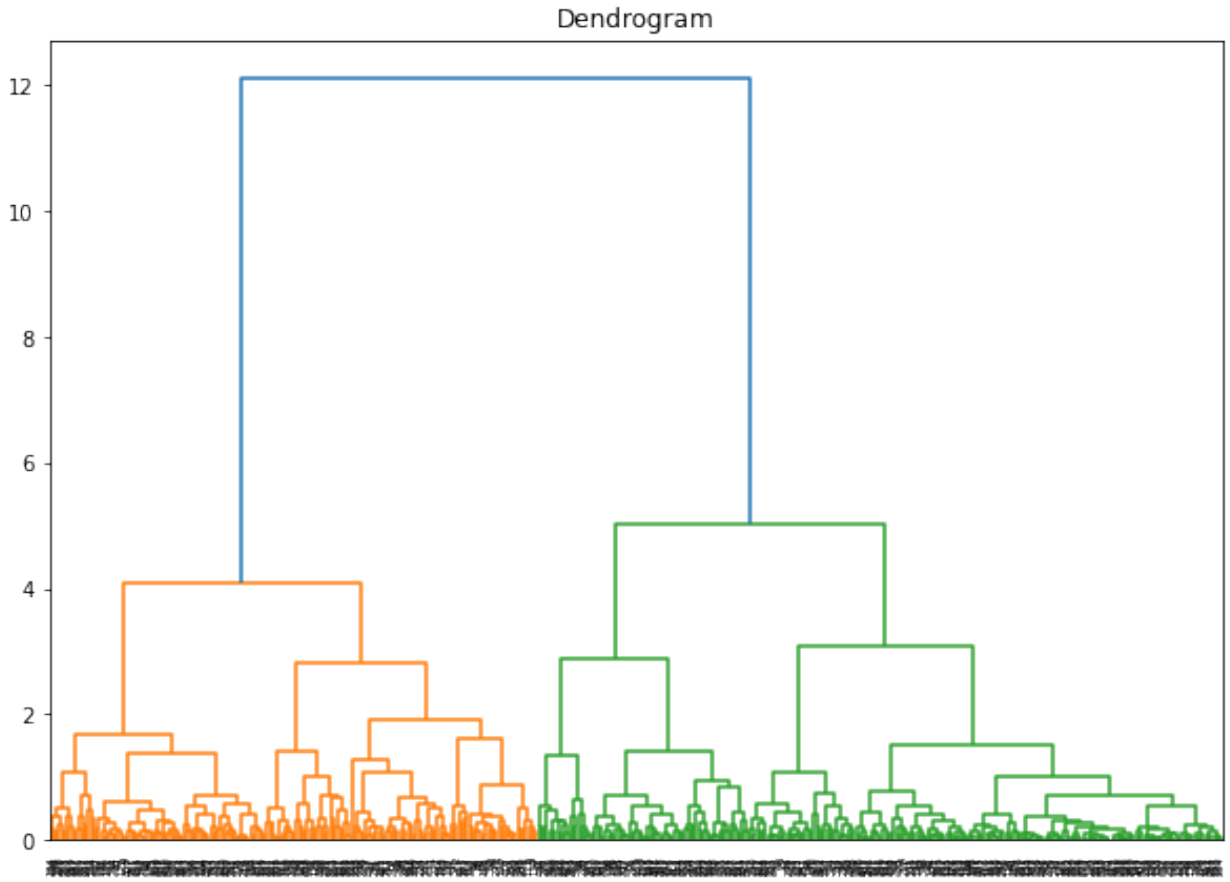
1. Why do you think it is important to normalise the data so that the scale of each variable is the same?

We are using p-norm distance measures like L1/L2 distances that rely heavily on the scale of the data. If some of the features have higher values, the distance will completely depend on those features. So, it is necessary to normalize the data to avoid that kind of skewness and make sure each feature is taken into account. That is accomplished with normalization.

2. Include an image of your hierarchical clustering dendrogram.



3. Looking at the plot generated by cell 8, where a threshold is taken on the dendrogram, how many clusters do we have?

We have 2 clusters, which is determined by the number of vertical lines intersected by the horizontal threshold line.

4. Why can we only see the values of 0s and 1s in the output of cell 10?

There are two clusters, and each data point is assigned a cluster 0-1 so we only see those predictions for each training data. If we had more clusters(>2) we would see the results being 0,1,2,3 etc.