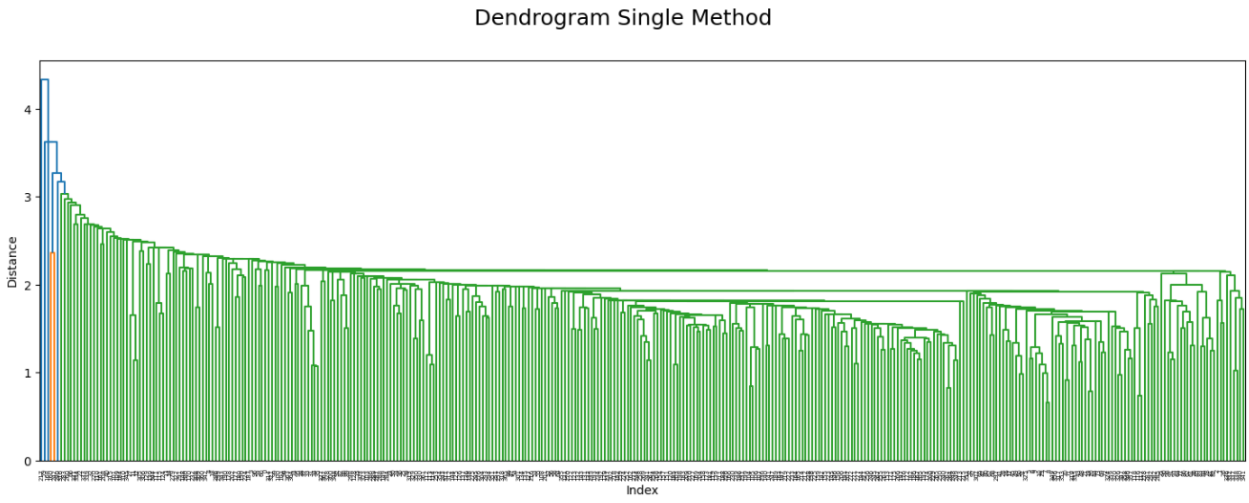
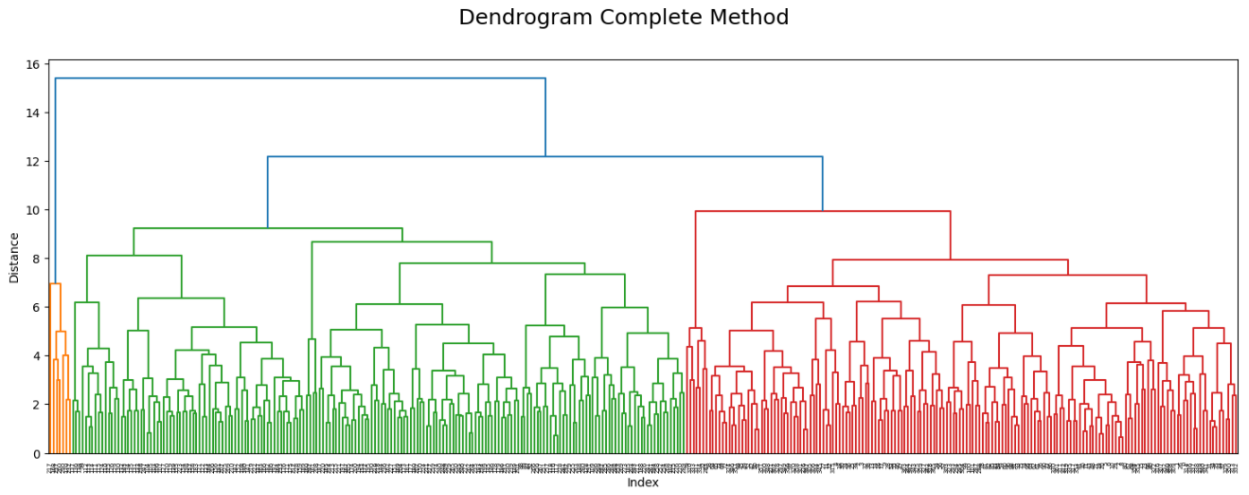
Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

Note that the following Dendrograms were created with scaled data for the year 1980.

BUDAPEST v OSLO – SINGLE METHOD

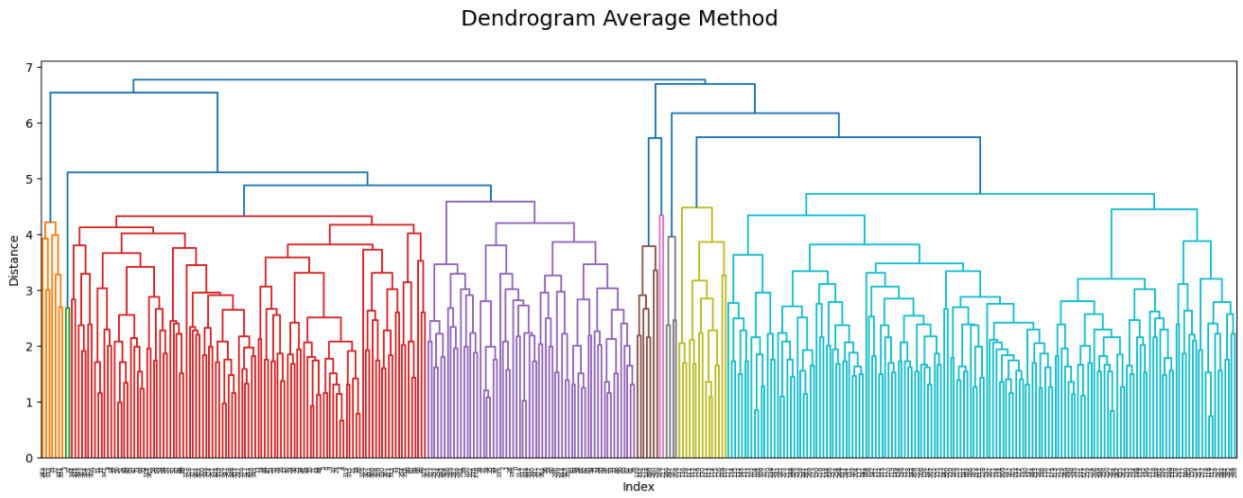


BUDAPEST v OSLO – COMPLETE METHOD

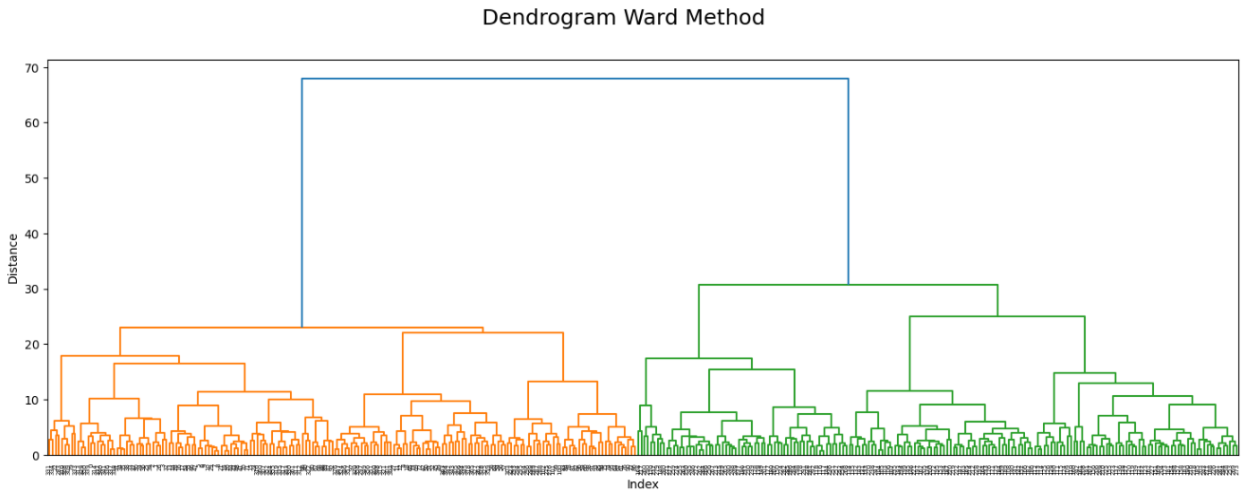


Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

BUDAPEST v OSLO – AVERAGE METHOD



BUDAPEST v OSLO – WARD METHOD

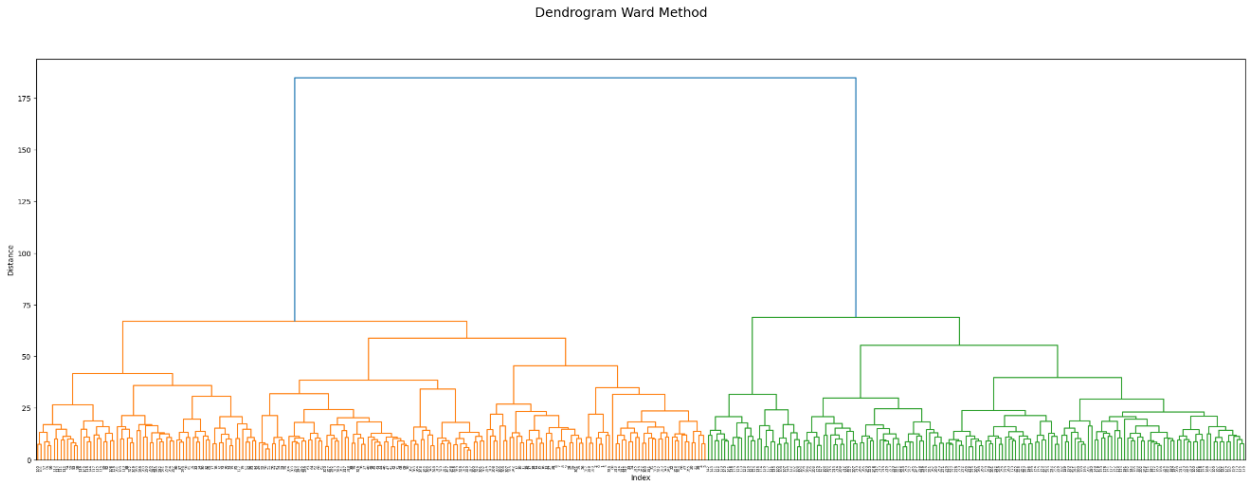


The Single method groups most days into one category, with a few outliers in the beginning.

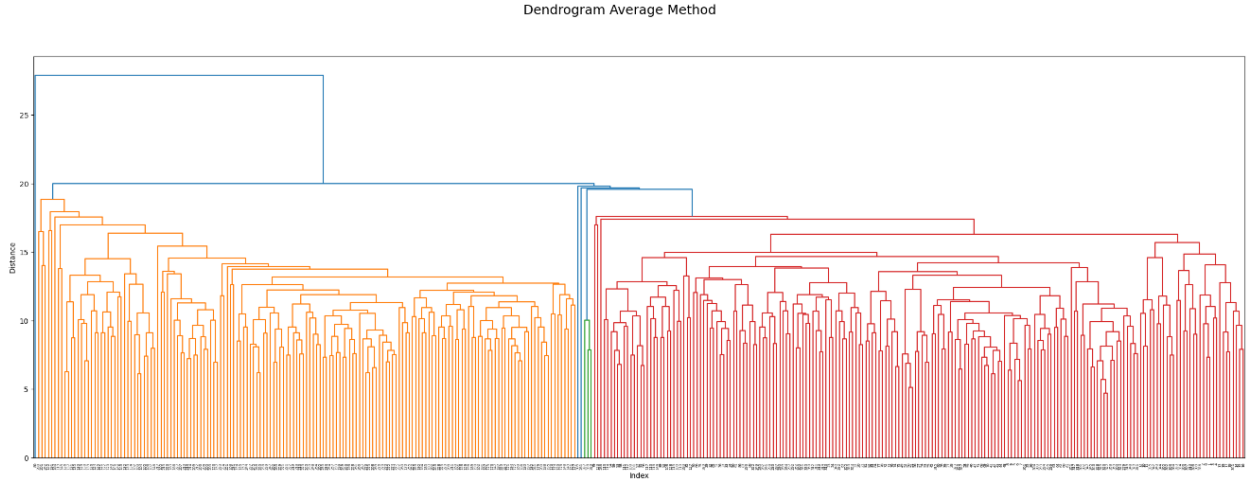
The Complete method more evenly distributes days, possibly representing summer, winter, and a spring and fall group, with three noticeable groups. The Average method forms a nice distribution of at least five identifiable groups with some overlapping in the middle. Finally, Ward's method forms two main clusters, which might indicate differences between locations or could represent two different ranges of temperatures. The Ward method also shows larger distances compared to the other three methods (different scale along the y-axis).

Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

ALL WEATHER STATIONS – WARD METHOD



ALL WEATHER STATIONS – AVERAGE METHOD

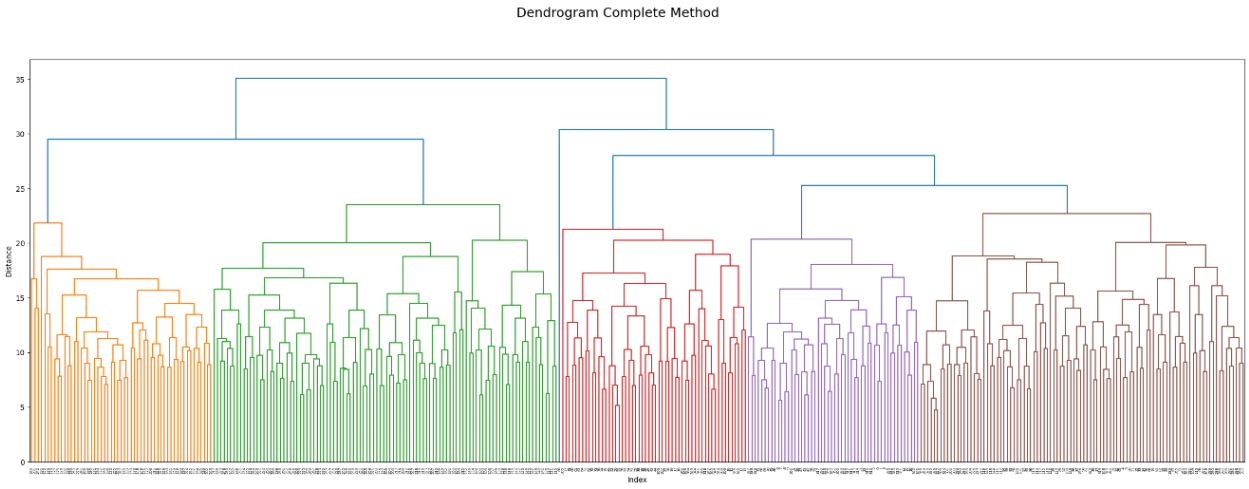


The Ward method again shows something strikingly similar to the previous ward method applied to Budapest and Oslo. Given that this time it is receiving inputs from all weather stations, I believe we can eliminate geography as a basis for the two clusters.

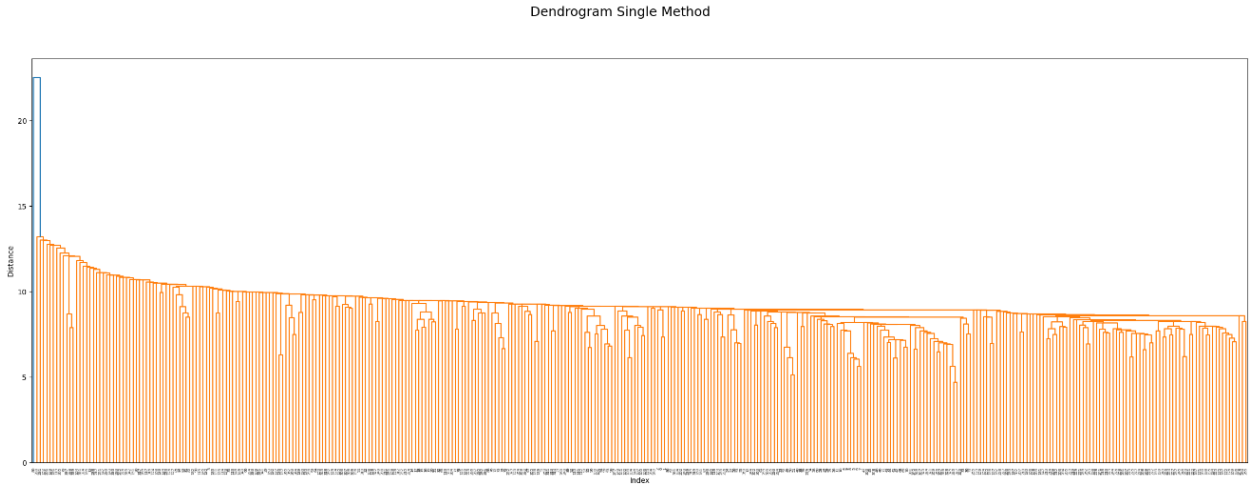
The Average method here is not as diverse as the average method previously applied to Budapest and Oslo. This is interesting given the additional 16 cities which are seemingly resulting in fewer clusters, perhaps due to more similarities in each of the cities data.

Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

ALL WEATHER STATIONS – COMPLETE METHOD



ALL WEATHER STATIONS – SINGLE METHOD

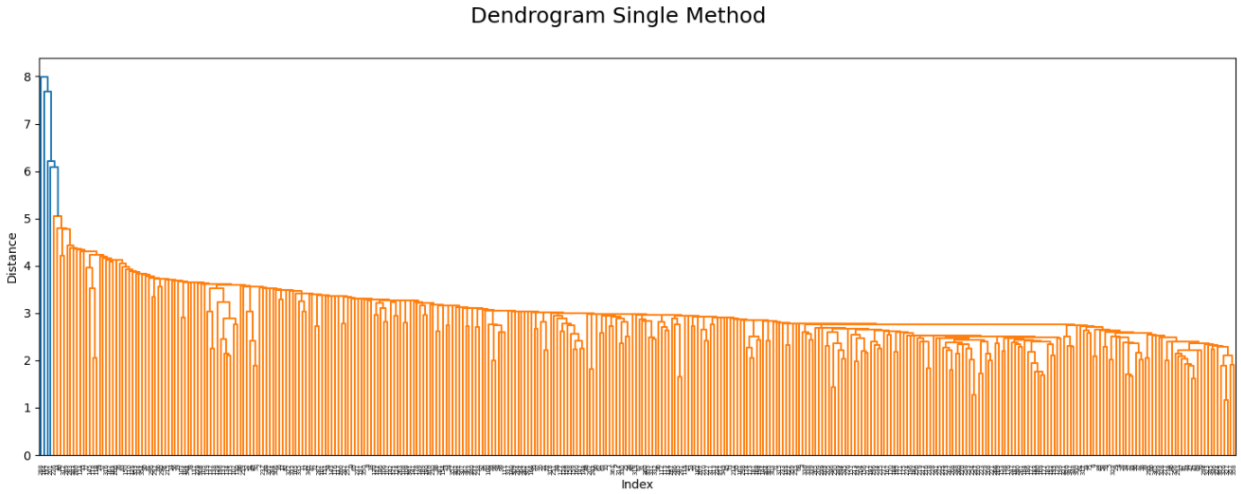


The Complete method above is the best result thus far from a grouping standpoint. There are five (5) nearly perfect groups clearly defined with little-to-no overlapping. This is excellent and should be investigated further with supervised learning to find out which variables are being used to group these colors.

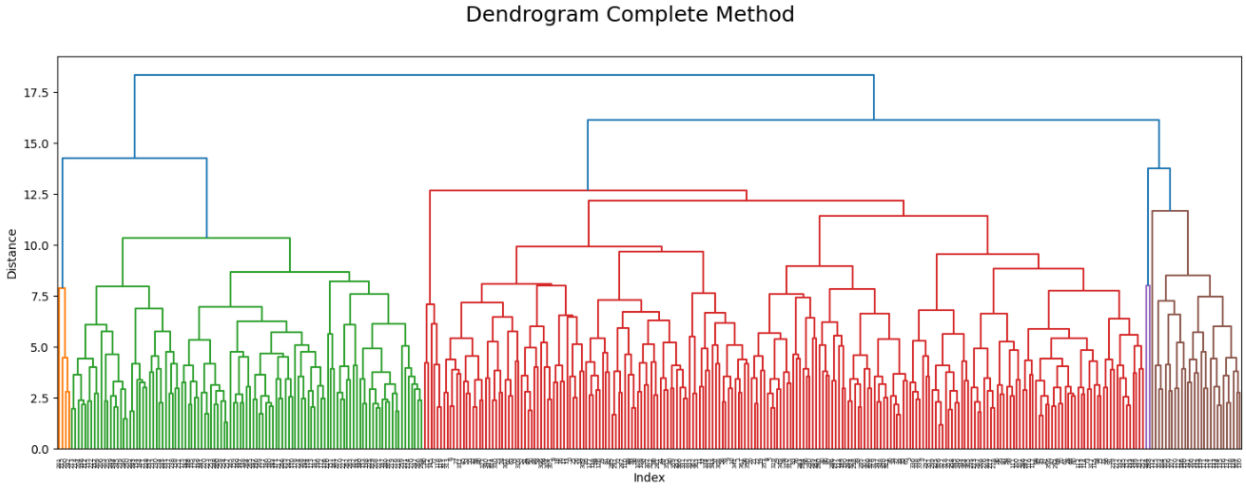
The Single method is almost identical to the previous single method for Budapest and Oslo and is not really demonstrating any use given the homogenous group for the entire period of time.

Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

MADRID v HEATHROW v MAASTRICHT – SINGLE METHOD



MADRID v HEATHROW v MAASTRICHT – COMPLETE METHOD

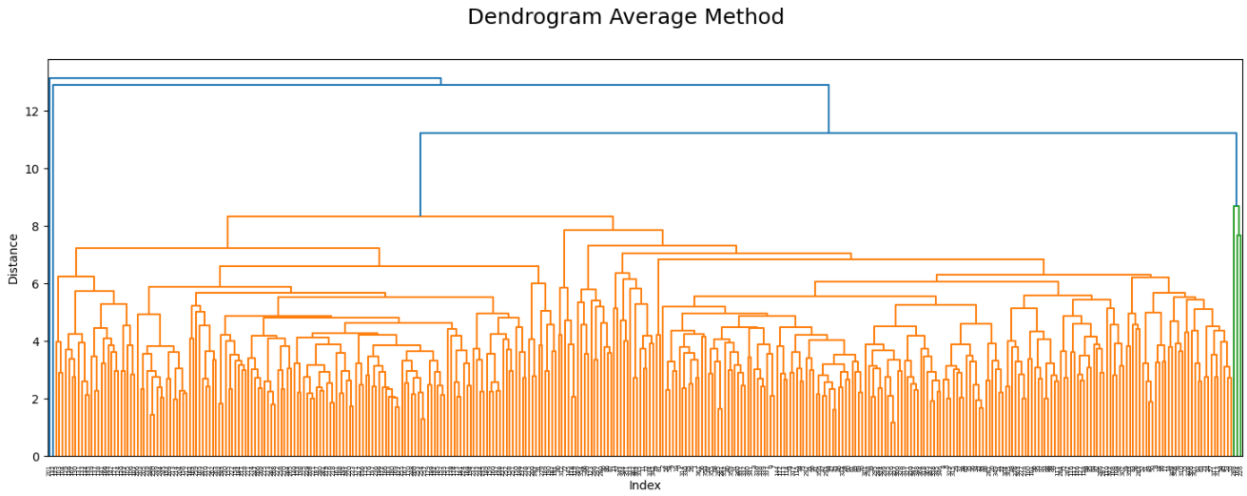


The single method continues to disappoint with one main group and few outliers.

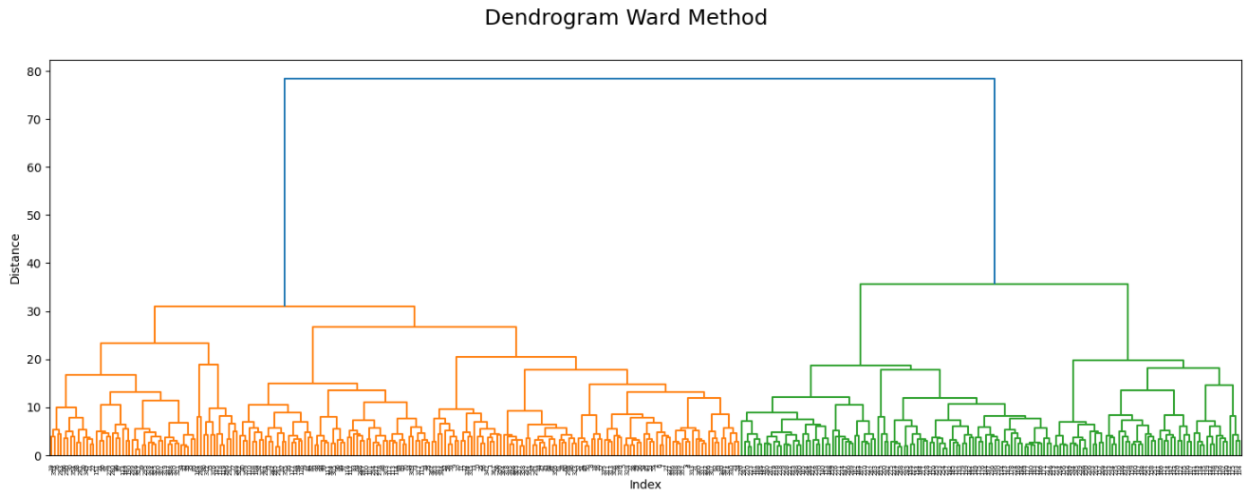
The complete method here is nice with three clear groups and perhaps some outliers on the left-most range of the x-axis. So far the complete method is outperforming the others.

Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

MADRID v HEATHROW v MAASTRICHT – AVERAGE METHOD



MADRID v HEATHROW v MAASTRICHT – WARD METHOD



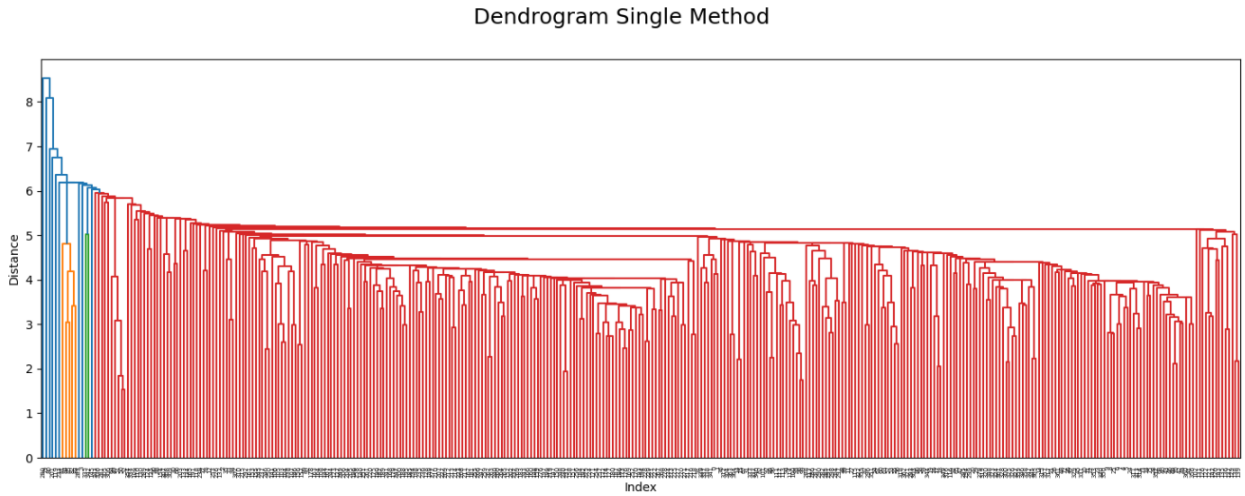
This average method result surprised me because it is so different from the previous results, but perhaps that is because the averages for these three stations is so similar that they are not differentiated.

The ward method continues to present two groups time and time again with our data. Even when we change the amount and location of weather stations.

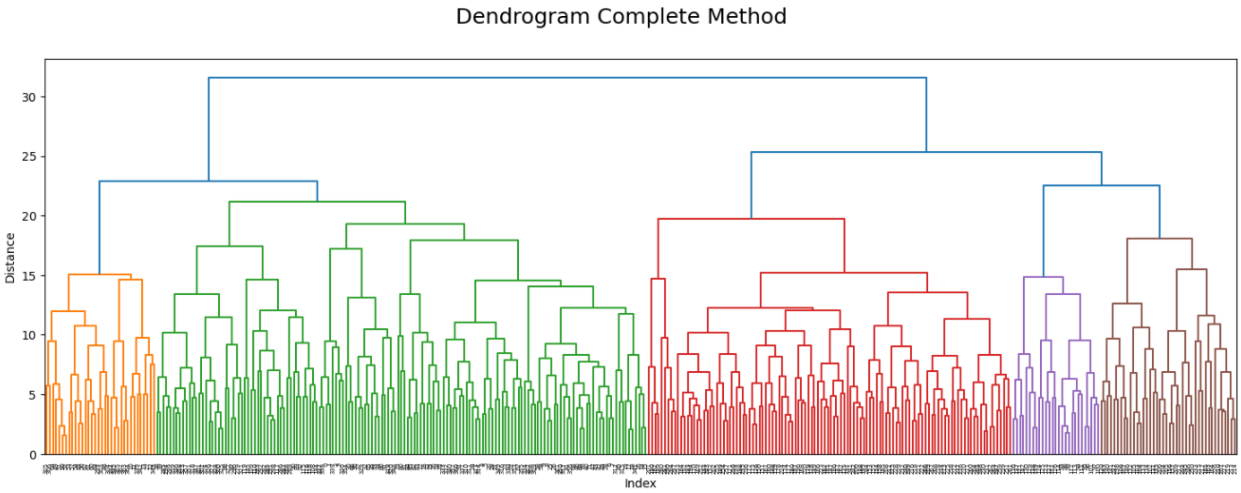
Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

PRINCIPAL COMPONENT ANALYSIS (PCA)

1980 PCA – SINGLE METHOD (9 COMPONENTS)



1980 PCA – COMPLETE METHOD (9 COMPONENTS)

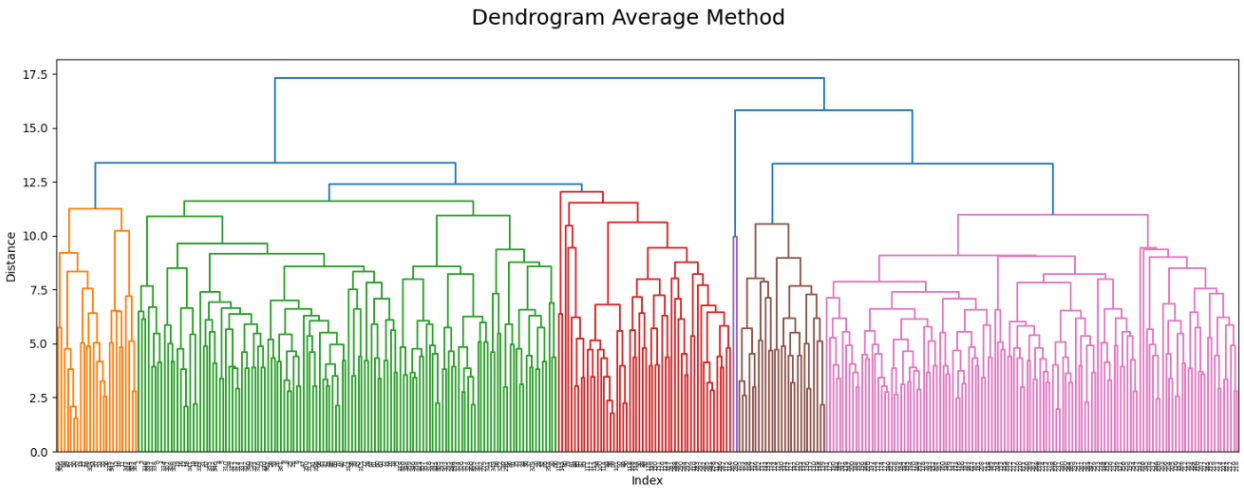


The Single method for our PCA actually has the most clusters, however the first three are extremely small relative to the large red grouping.

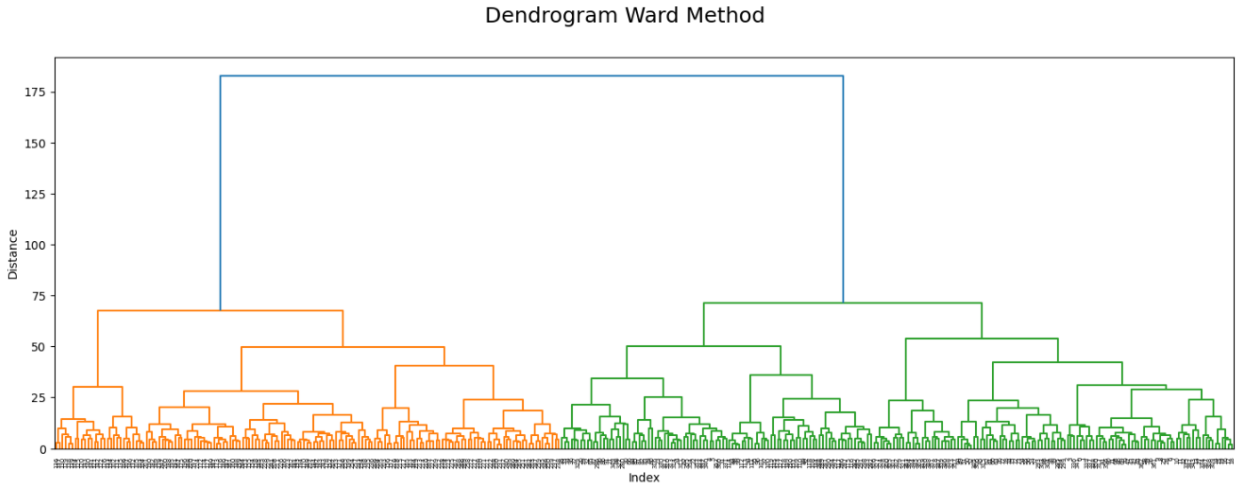
The Complete method lives up to its reputation achieved from the results above, with five (5) beautifully identified groupings and little-to-no overlap.

Josh Wattay
Machine Learning with Python
Task 2.1
CareerFoundry

1980 PCA – AVERAGE METHOD (9 COMPONENTS)



1980 PCA – WARD METHOD (9 COMPONENTS)



The PCA Average method is excellent with five (maybe even 6) groups that are well-defined.

The Ward method continues to provide two large groups.

Given the results of these Dendrograms, I would recommend using the Complete method, or in the alternative the Average method, as these have performed the best so far.