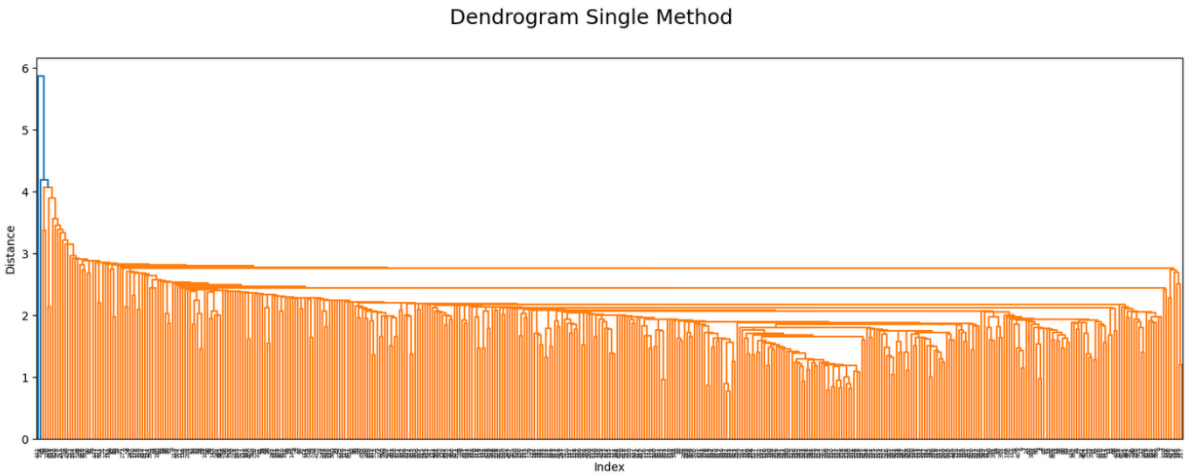
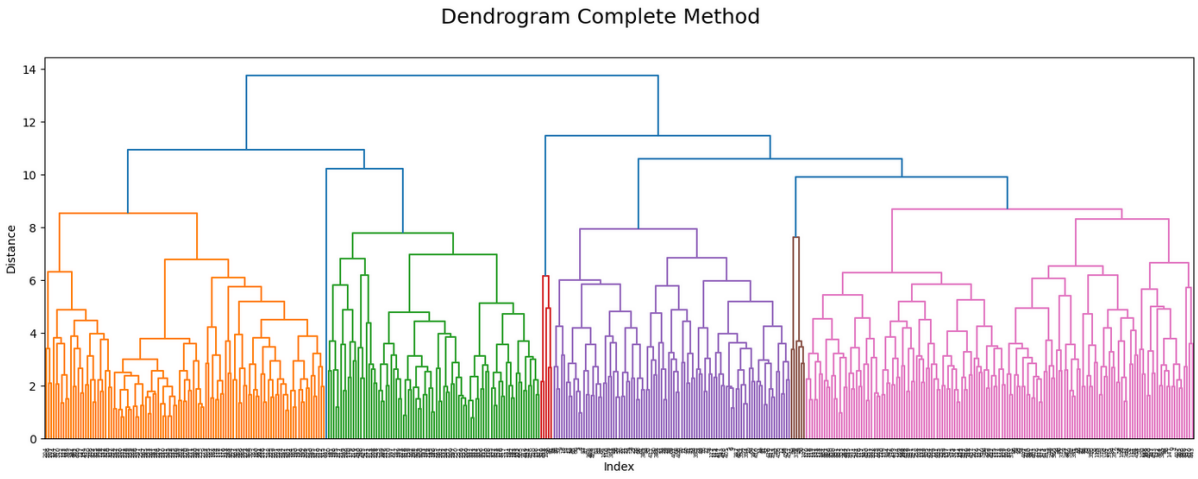
Dendograms

1. Madrid vs Munich 2010

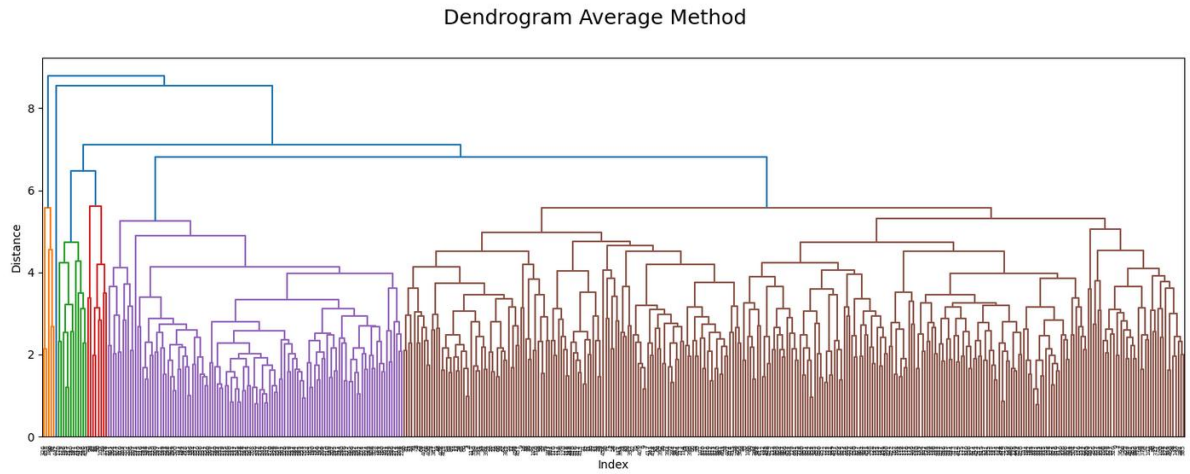
Single Method: The distance between the two closest members of each cluster to determine which group they should belong to



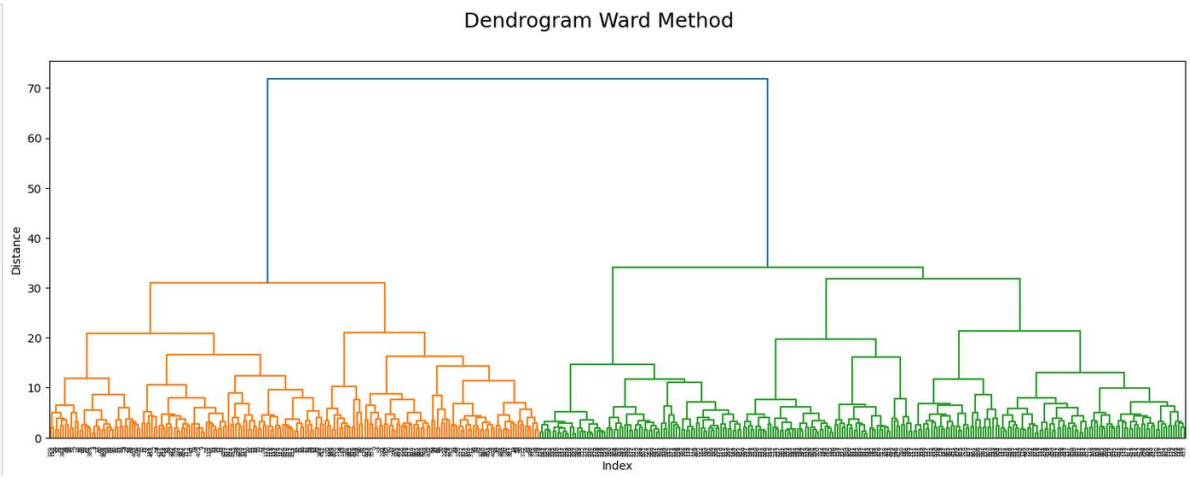
Complete Method: The distance between the two farthest members of each cluster to determine which group they should belong to



Average Method: The distance between the average of the members of each cluster to determine which group they should belong to



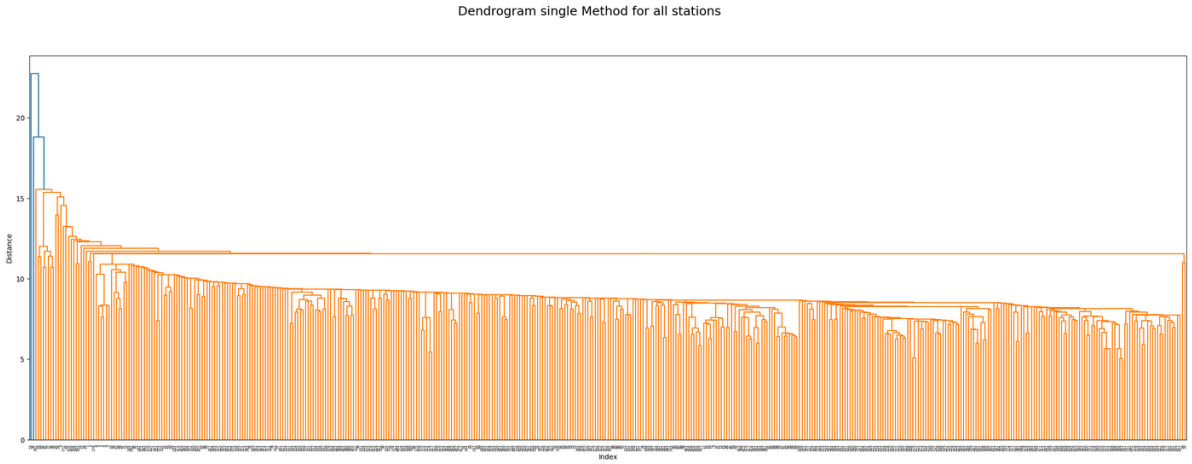
Ward Method: This uses a metric called the minimum increase of sum of squares (MISSQ) to find the distance between two clusters. It attempts to minimize the variance between the two clusters



Ward tends to produce a larger distance scale than the other methods. Single linkage places most days into one category, which suggests poor separation between clusters. Complete linkage, on the other hand, forms the most well-balanced clusters, resulting in around six distinct color-coded groups that may correspond to the four seasons or other periodic trends. The average linkage approach creates two large clusters and two smaller ones, offering some distinction but less detail than the complete method. Ward ultimately produces two large clusters at a higher distance, possibly reflecting broad geographical differences (such as Madrid versus Munich) or seasonal splits (like winter versus summer).

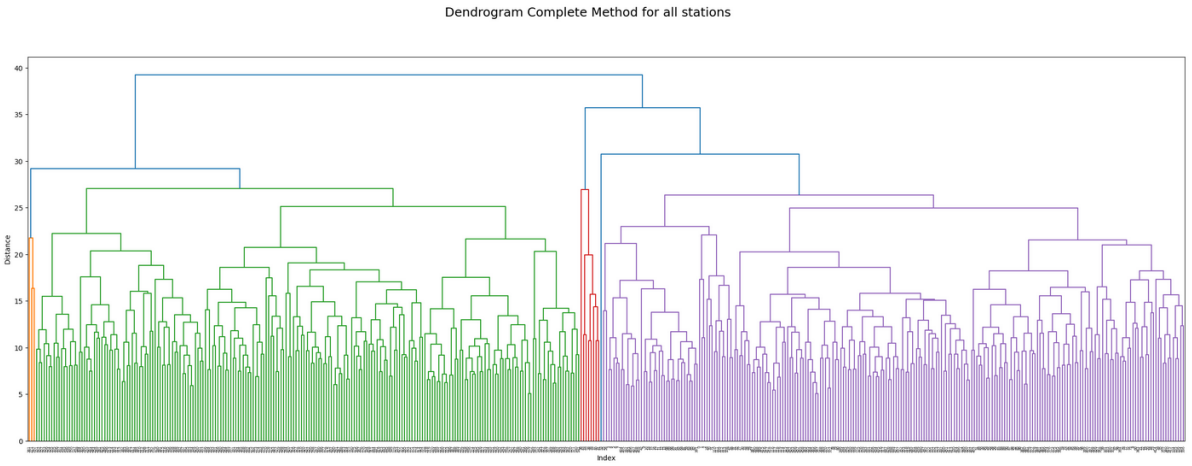
2. All Stations 2010

Single Method



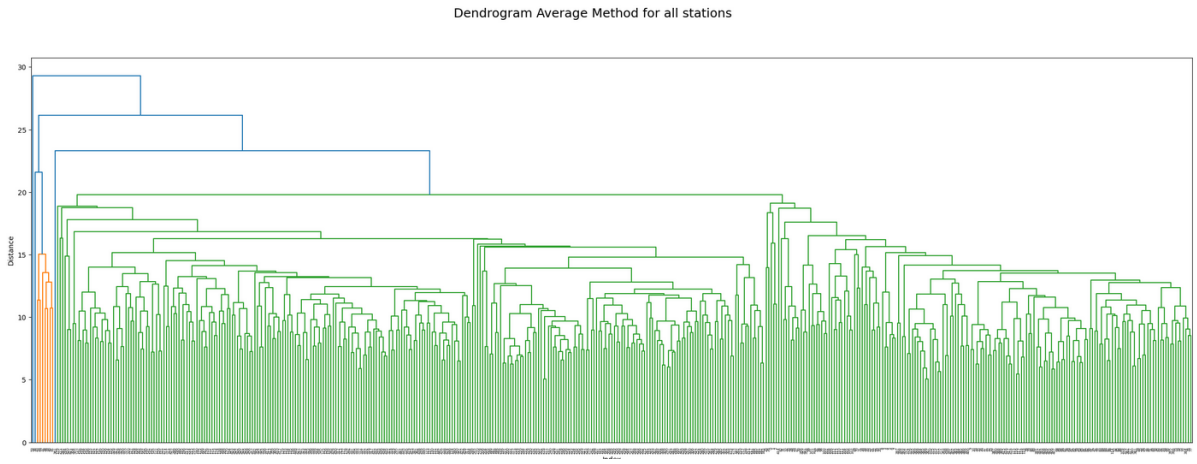
Insight: The method over-generalizes the data, grouping everything together and failing to provide any actionable or interpretable insights.

Complete Method



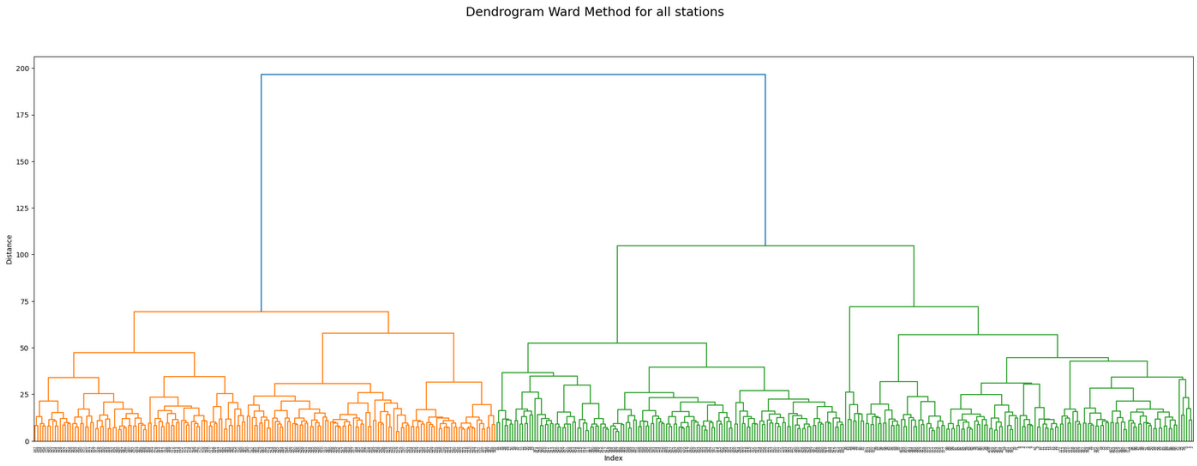
Insight: It separates into 4 categories: 2 large and 2 small, almost evenly split into two main clusters with two smaller ones

Average Method



Insight: It's clumped into 2 groups, but the green category seems overfitted and lacks meaning, possibly due to temperature differences between hotter and cooler locations

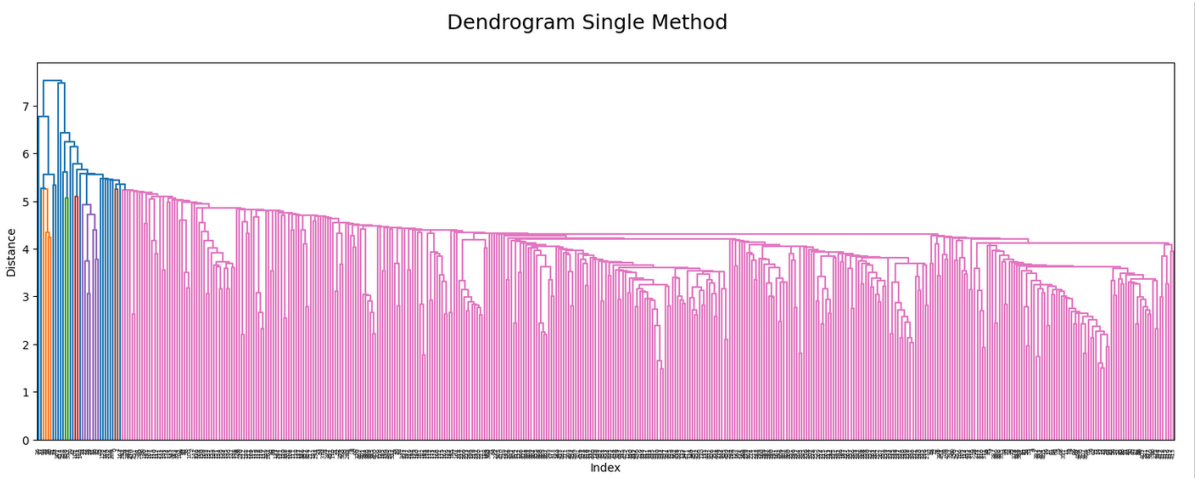
Ward Method



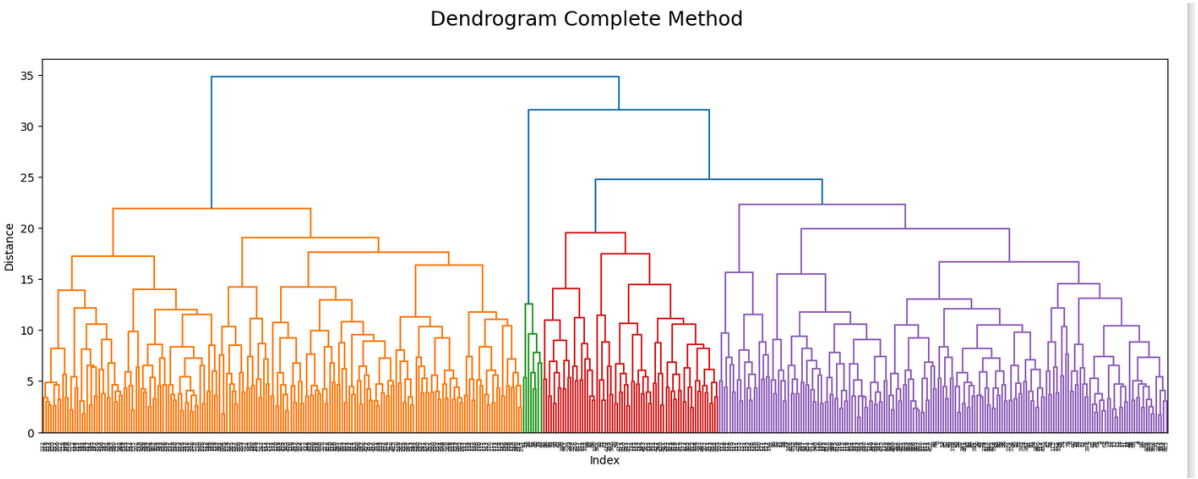
Insight: It split all locations into two clusters for 2010, likely due to location-based temperature differences.

3. Reduced Data 2010

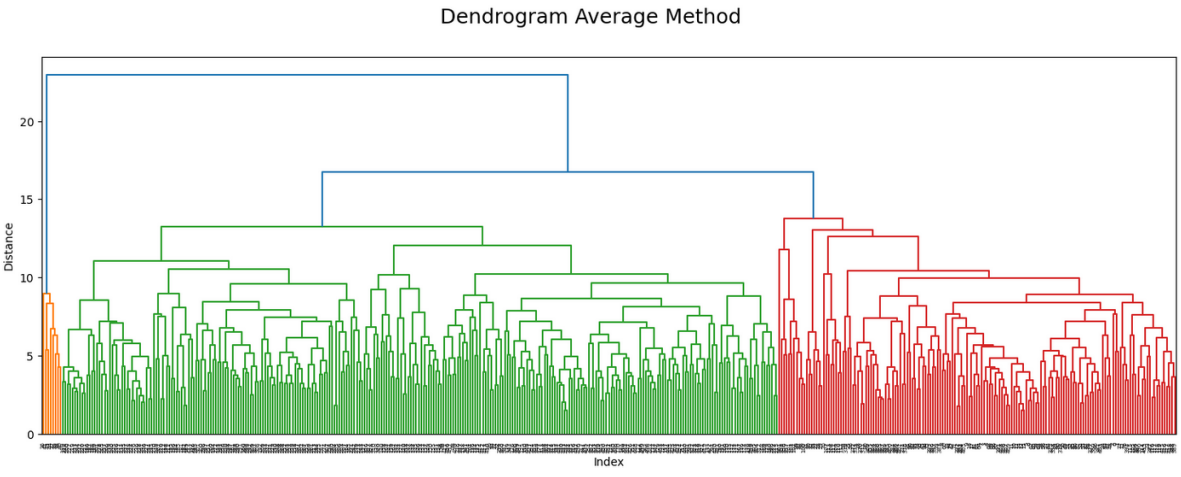
Single Method



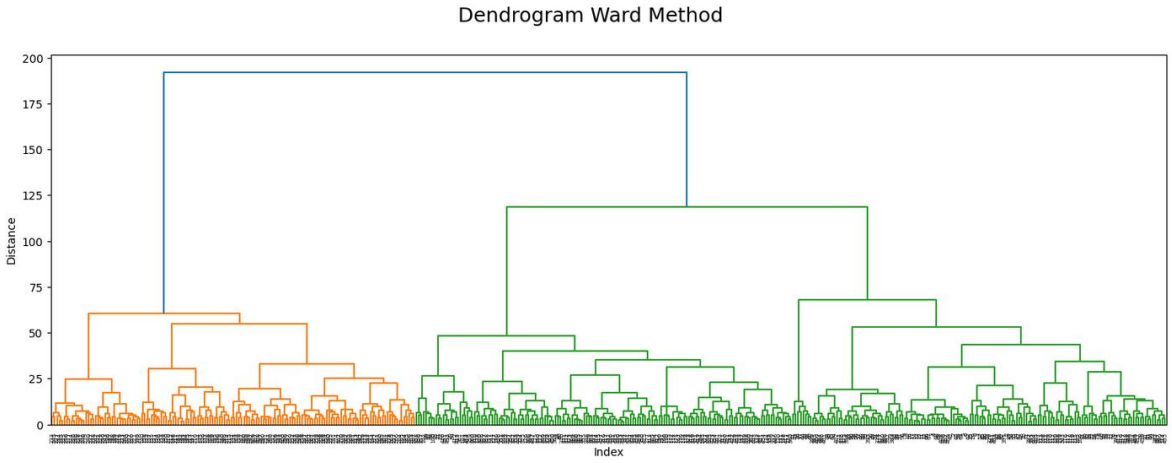
Complete Method



Average Method



Ward Method



Summary

1. The single linkage method displays more categories overall, but mostly results in one main cluster, making it less useful for meaningful separation.
2. The complete linkage method creates a good mix of clusters, likely highlighting differences based on location or seasonal trends.
3. The average method produces three distinct clusters, which could reflect differences in location, seasonal patterns, or a combination of both, resulting in a more meaningful division of the data.
4. The Ward method yields two clear clusters, which may correspond to broad distinctions such as climate zones or seasonal patterns, offering a simplified but potentially insightful split in the data.