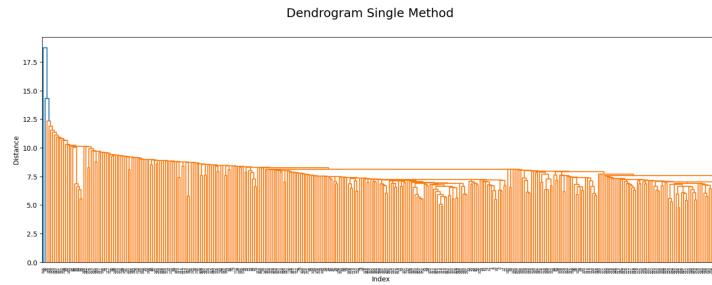


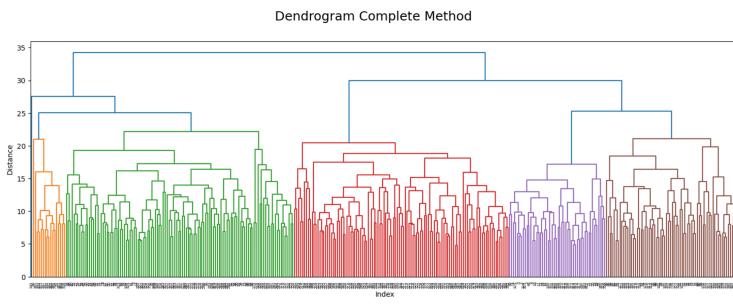
Initial Dendrograms

The following initial dendrograms were

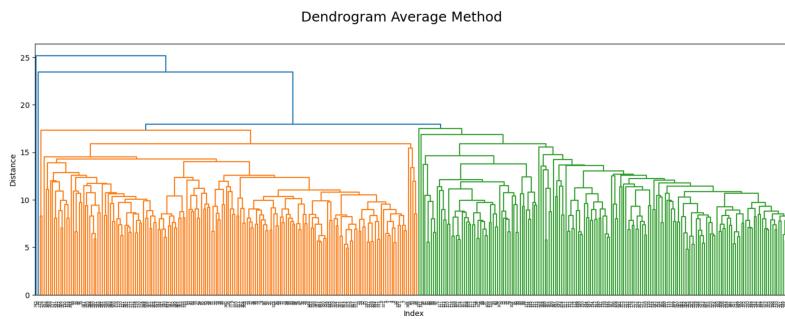
1. Performed on the weather data merged with the pleasant ratings
2. Filtered for 2020
3. Scaled



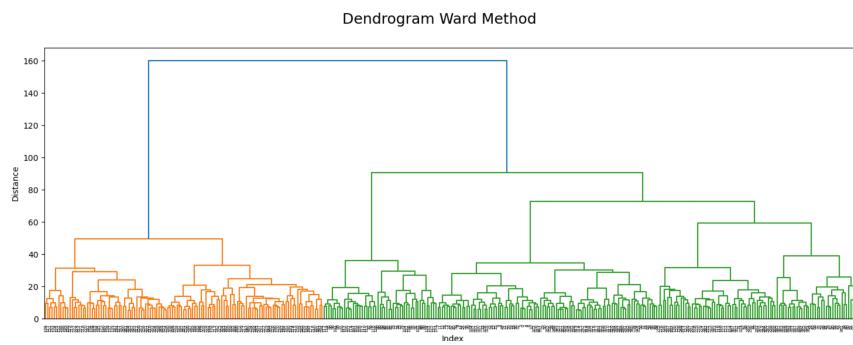
The single method finds the distance between two closest points within a cluster to determine which group they belong to. Using the single method on the merged dataset, we get only one group.



The complete method looks for the longest distance between two points in a cluster to determine their group. Here we see 5 groups on the merged dataset.

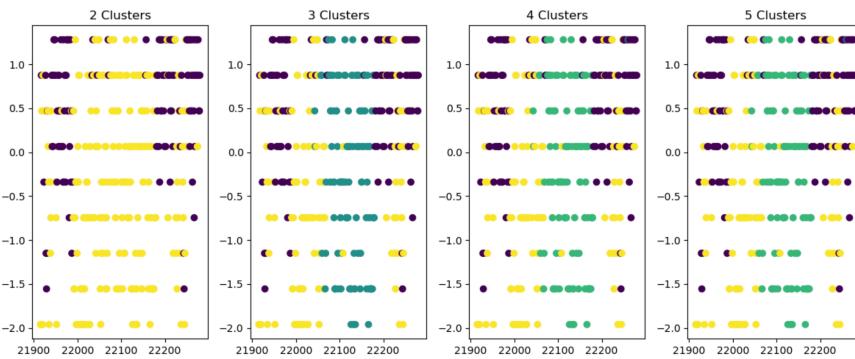


The Average method looks for the average distance between two points in a cluster to determine their group. Here the data has been grouped into two clusters.



The Ward method uses something called the minimum distance of the sum of squares and tries to minimize the variance between the two clusters. Two distinct groups are presented with the merged data using the Ward method.

Clusters from the Complete Linkage method were plotted out, and silhouette scores performed.



Silhouette score for 2 clusters: 0.132

Silhouette score for 3 clusters: 0.141

Silhouette score for 4 clusters: 0.141

Silhouette score for 5 clusters: 0.141

Silhouette methods are ranked -1 to 1 with 1 being a great fit. These are all low—2,4, and 5 clusters have the slightly better score of 0.141.

After PCA was performed, it seemed clear that 2 clusters was the apparent choice, though in later conclusions, this could be because of the inclusion of the weather ratings within the dataset.

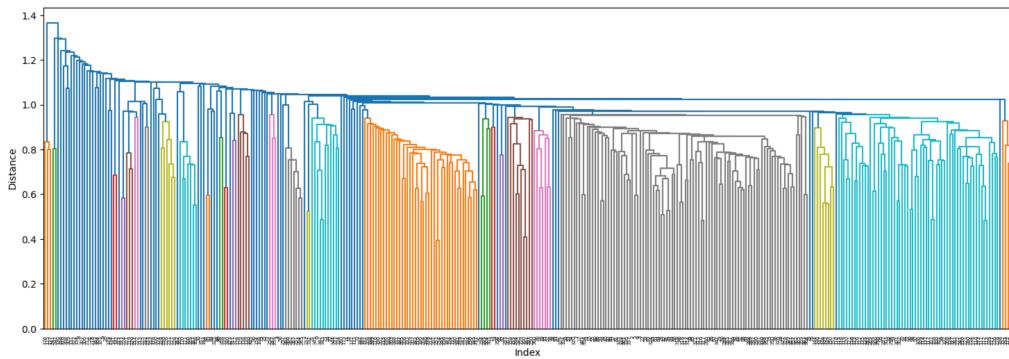
The following reductions were made for further exploration:

1. Mean Temps for 15 Weather Stations, 2020
2. Oslo / Valentia ALL Data, 2020 (for comparison of more climate extremes)
3. Oslo / Stockholm ALL data, 2020 (for comparison of more subtle data)

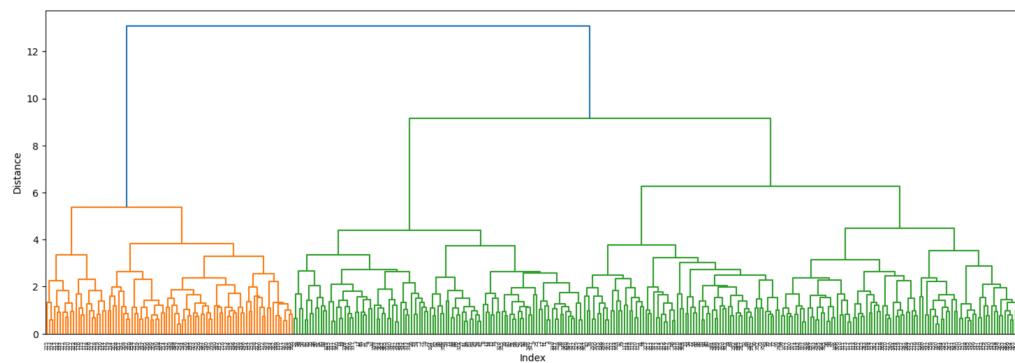
Reduced Dendrograms

- All Stations - Mean Temps

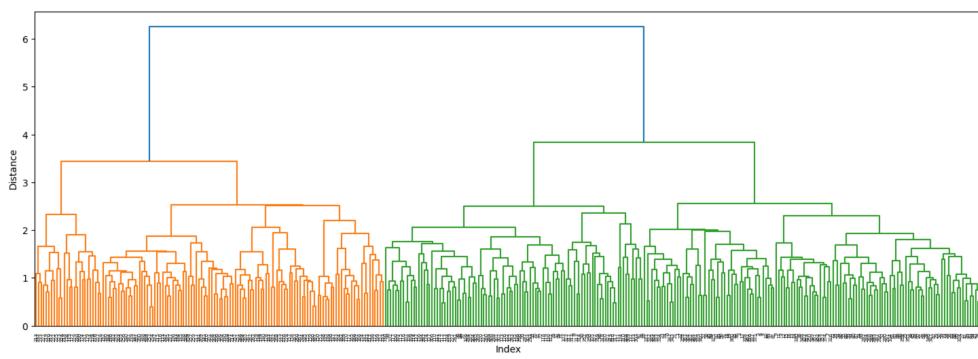
Dendrogram Single Method | All Stations Mean Temps



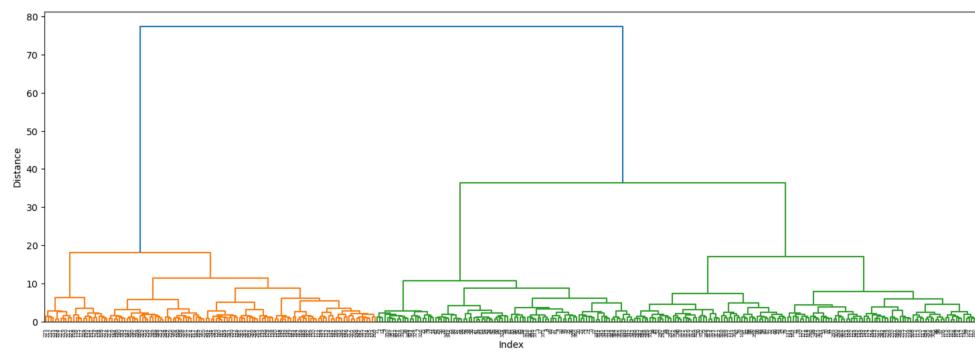
Dendrogram Complete Method | All Stations Mean Temps



Dendrogram Average Method | All Stations Mean Temps

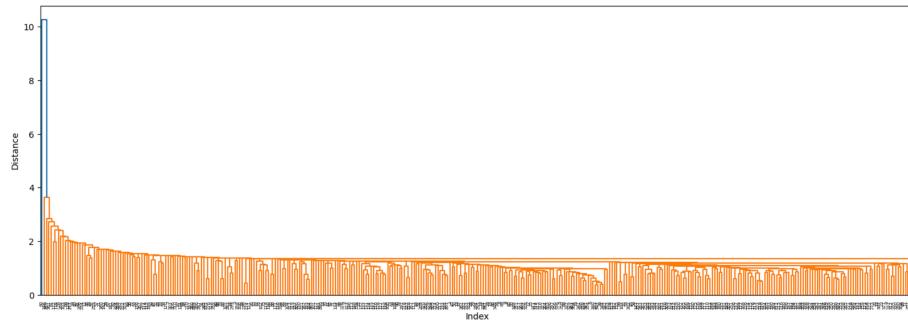


Dendrogram Ward Method | All Stations Mean Temps

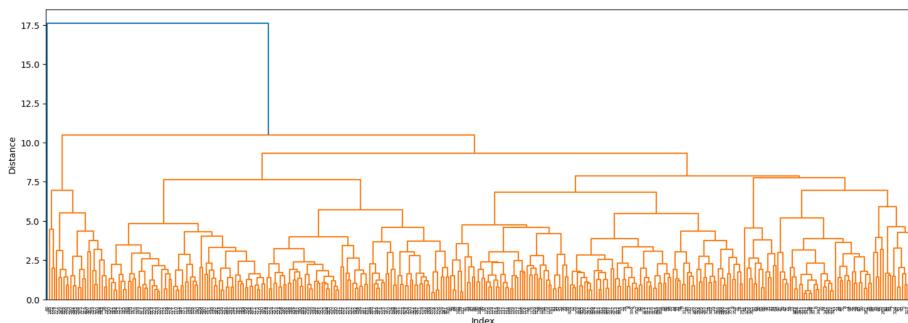


- Oslo / Valentia

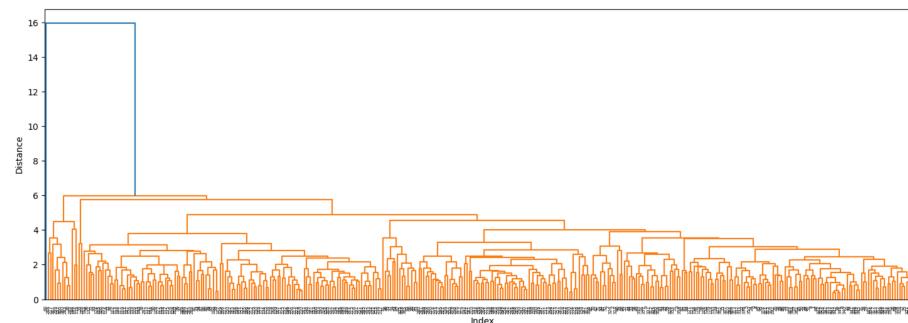
Dendrogram Single Method | Oslo Valentia



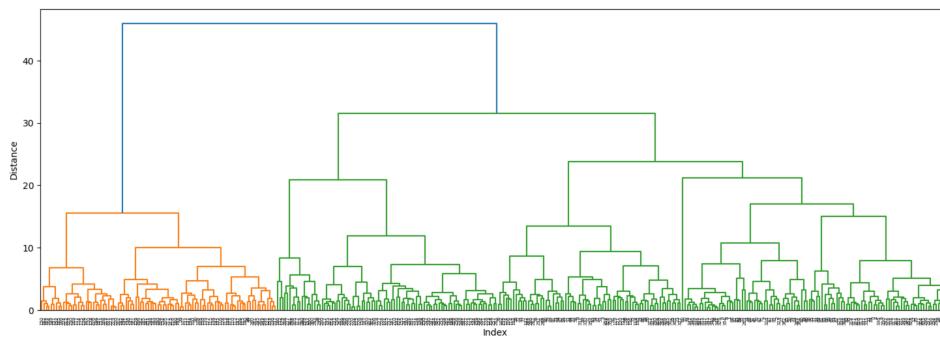
Dendrogram Complete Method | Oslo Valentia



Dendrogram Average Method | Oslo Valentia

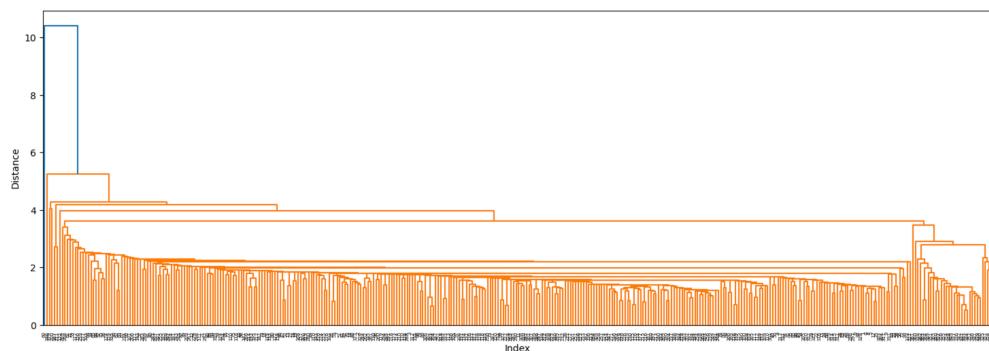


Dendrogram Ward Method | Oslo Valentia

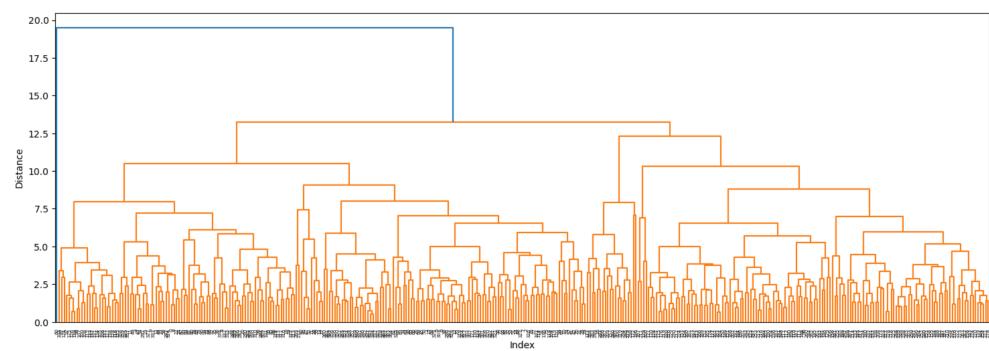


- Oslo / Stockholm

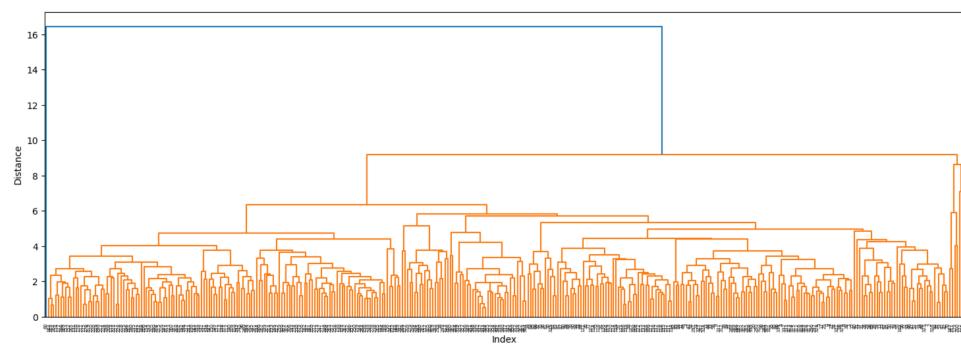
Dendrogram Single Method | Oslo Stockholm



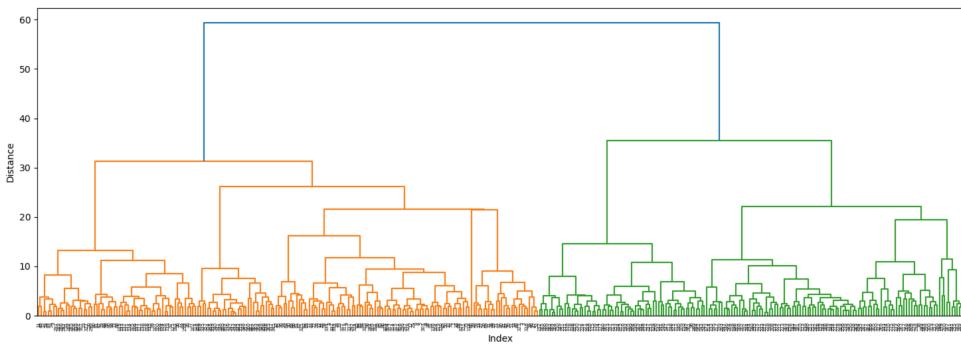
Dendrogram Complete Method | Oslo Stockholm



Dendrogram Average Method | Oslo Stockholm

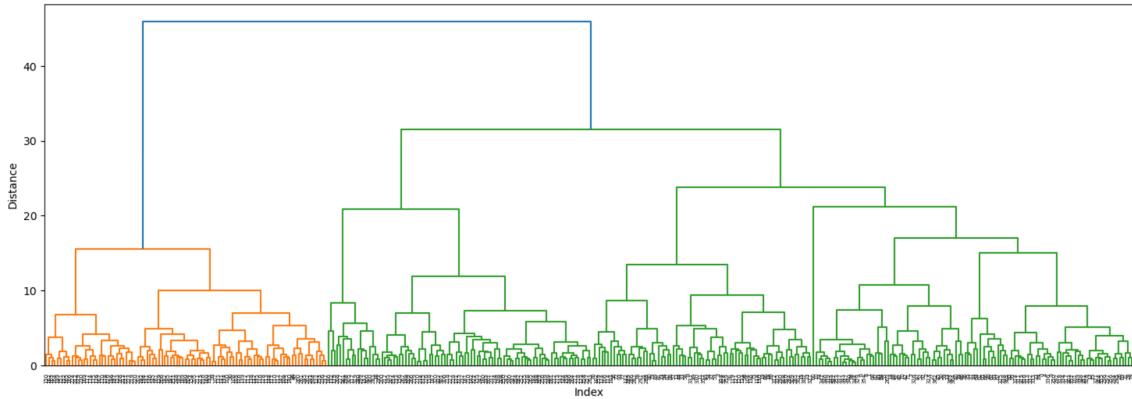


Dendrogram Ward Method | Oslo Stockholm



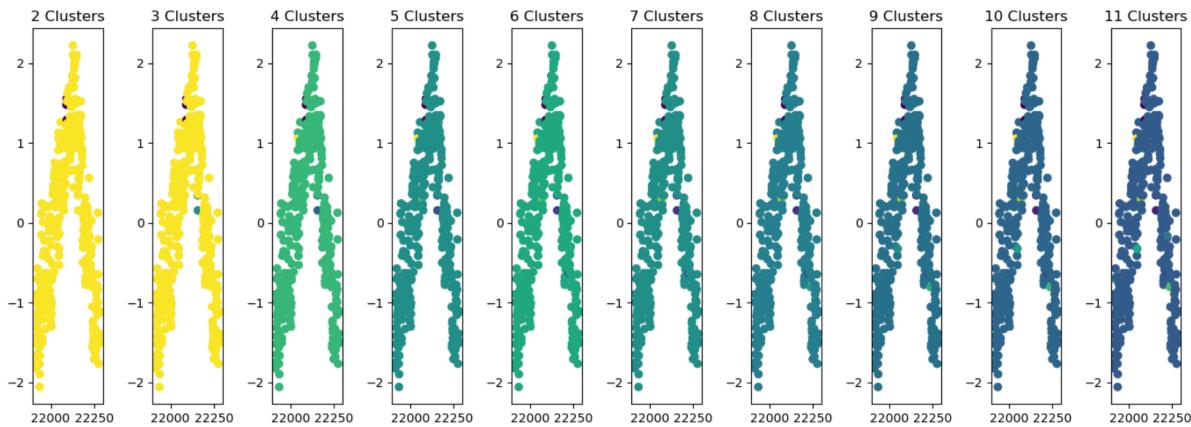
- Oslo

Dendrogram Ward Method | Oslo



Clustering / Silhouette Scores

- All Stations - Mean Temps; Single Method



Silhouette score for 2 clusters: 0.206

Silhouette score for 3 clusters: -0.044

Silhouette score for 4 clusters: -0.172

Silhouette score for 5 clusters: -0.271

Silhouette score for 6 clusters: -0.437

Silhouette score for 7 clusters: -0.443

Silhouette score for 8 clusters: -0.455

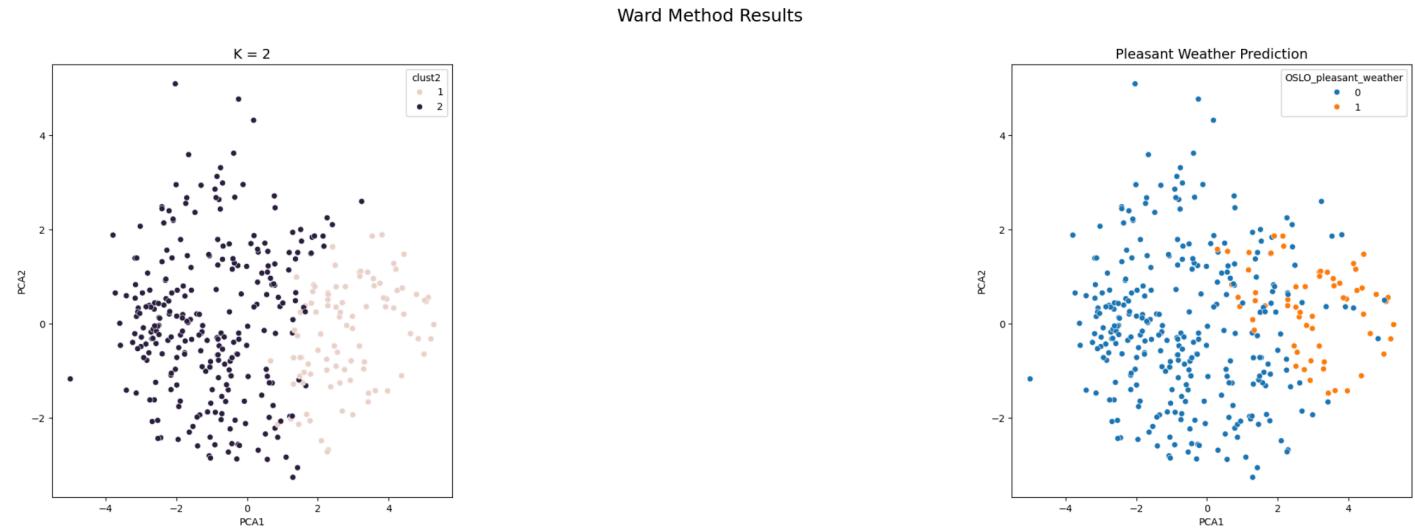
Silhouette score for 9 clusters: -0.455

Silhouette score for 10 clusters: -0.476

Silhouette score for 11 clusters: -0.480

PCA on Reduced Oslo Data

- 2 Components



Ward Method Clusters and Pleasant Rating clusters were quite similar!

Conclusions:

When comparing the clusters against the pleasant weather ratings, you see a similar pattern. The “unpleasant” days seem to appear slightly more than the cluster “2” (purple) from Ward Method, but still fairly similar.