DS250: Data Analytics and Visualization

Assignment 3c: Cluster Time Series and show in dashboard.

Nikhil Chaudhary (11940740), Anjana Kannan (11940130)

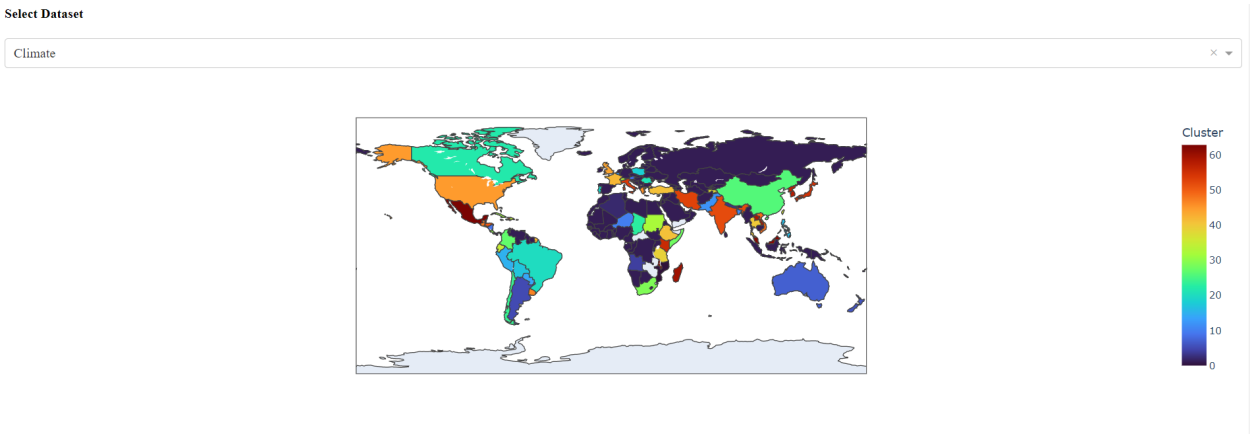
We have done the **manual implementation** of the **Hierarchical Clustering** algorithm. We found out that the clustering is optimum for k between 50 to 80. There are two types of distance metrics used which are DTW and Euclidean distance. We have used MIN and MAX for the inner distance matrix.

Flow of Code:

1. Load the dataset into pandas dataframe.
2. Normalize the data using the min max normalization
3. Compute proximity matrix by using either DTW or Euclidean distance
4. Create a heap which contains distance between every point. The heap should be min-heap if MIN is being used and max-heap if MAX is being used.
5. Initialize all the points as separate clusters
6. Use the hierarchical algorithm to merge the clusters
7. Stop when the number of clusters reaches K.

Two types of graphs have been plotted:

1. This graph lets the user choose the dataset that they want to visualize. The value of K is chosen randomly between 50-80 because this range gives the optimal answer. DTW and MIN are used in this graph.



2. This graph visualizes a single dataset but lets the user choose the parameters like the values of K, distance and inner distance.