Description of algorithm

1. For each column impute missing values using cubic splines. Any missing values coming before the first non-missing value or after the last non-missing value are not imputed since the behavior of extrapolating cubic splines before the first spline or after the last spline can sometimes be erratic.

2. For each column calculate DFFT (Discrete Fourier Transform) of the non-missing values (after imputation). Because frequency domain is symmetrical, take only positive frequencies.

3. Do agglomerative hierarchical clustering in the frequency domain (using the DFFT-transformed. The distance metric between individual columns x and y is
   $1 - abs(corr(x, y))$
   1 - absolute value of correlation of column x and column y.
   For groups u and v of multiple columns the distance is calculated as the max of all distances between individual columns of u and v.
   $$d(u, v) = max\big(dist(u[i], v[j])\big).$$

4. The selected criterion for stopping the agglomeration of different groups is that in order for two groups to be joined the distance must be less than half of the maximum of distances between any two columns in the data.