Pre-processing

* Started by checking for missing values in rl-kpis and met-real datasets. Substituted the missing values using KNN imputer for float or int type values, and for object type, substituted with the value having maximum occurrence. Dropped few columns in met-real such as measured\_date, measured\_hour as it was not significant and columns such as pressure and pressure\_sea\_level as it had excessive missing values.
* Then checked for duplicates in both the datasets.
* Started resampling of rl-kpis and met-real datasets at 1D on datetime using group by as site-id and station-no respectively, and then dropped null rows.
* For establishing relation between site-ids and station-nos, we used the distances dataset. Using this dataset, we found for every site-id which is the closest station-no.
* Using the above relation, we merged rl-kpis and met-real.
* Then converted columns like type, tip, card\_type, adaptive\_modulation, and freq\_band into int datatype.
* Converted datetime column into datetime datatype and set the index as datetime. Then resampled again at 1D.
* Dropped the columns like station-no, site-id, tip and freq\_band as it had no unique values.
* Resampled the merged dataset on 4D.