Aim: To forecast ground-level O3 and NO2 for 24-hour (i.e., 1-day) at hourly interval for Delhi

* Data (July 2019 - June 2024: 5 years) at hourly interval are provided for 7 sites in CSV format along with headers. (First develop your model for any one site, and later try with other sites)
* Input parameters: year, month, day, hour, O3\_forecast, NO2\_forecast, T\_forecast, q\_forecast, u\_forecast, v\_forecast, w\_forecast, NO2\_satellite, HCHO\_satellite, ratio\_satellite
* Parameters to be estimated: O3\_target, NO2\_target (These are true measured values in ugm-3)
* Use “site\_X\_train\_data.csv” (X is a site number) to train and test your model. You can divide this data into 75%-25% to develop your model. Data are shuffled for different days, but in order over 24 hours.
* Use “site\_X\_unseen\_input\_data.csv” as input data for the final validation of the model. True value of O3\_target, NO2\_target corresponding to this input data will be provided during the hackathon. That will decide the accuracy of your model.
* Note:
  + Parameters with “forecast” label are from the reanalysis model
  + Satellite observations (i.e., NO2\_satellite, HCHO\_satellite, ratio\_satellite) are once in a day (not hourly interval). It is challenging to effectively utilise these data in your model. If you do not find solution how to utilise it, drop the ‘satellite’ data and go ahead with ‘forecast’ data.
  + If you want to estimate O3\_target and NO2\_target for a given day, the input parameters (forecast as well as satellite) of the same day should be used. Additionally, if you wish to use past data, past 7 days of “O3\_target”, and “NO2\_target” could be used.
  + Matrices to evaluate your model: RMSE (Root mean square error, unit: ugm-3), R2, and RIA (Refined Index of Agreement)