**Feature Engineering:**

Our approach to feature engineering was representing variables differently for experimenting with different model algorithms – turning the inputs in the dataset into things that the algorithm can understand.

* So, we used the datetime variable in the dataset (utc\_time) to generate three components of the variable – ***hour, month and date***.
  + Hour variable seemed to have significant relation with the air quality variables – Understanding: time of the day matters due to factory/manufacturing industries open times and traffic patterns
  + Month variables also mattered to predict air quality variables – Understanding: Weather patterns do have impact on the air quality
* Additionally we added couple of variables – ***weekend and holiday*** based on a basic understanding of what could affect air quality around the stations.
  + If it was a weekend, there could be high traffic which affects air quality
  + If it was a holiday, closures in manufacturing industries/holiday traffic patterns affect air quality
    - Used an external data source suggested in ‘Discussion board’ to identify dates considered as holidays - <https://www.timeanddate.com/holidays/china/2017> , <https://www.timeanddate.com/holidays/china/2018> , <https://www.gov.uk/bank-holidays>