## Observations from Data:

- Data consist of 4 columns namely time, operation_hours, fuel_consumption and engine_id. We set time as an index of dataset.
- As mentioned in problem statement package contains a historical dataset of 20 engines for past 100 days. Hence following analysis are done by grouping data of each engine by its engine_id.
- Empty_col_reduct(In[3]) function checks if there's any missing data or not. We don't have any missing data in our dataset.
- In[4] plots box plot as well as violin plot to analyse spread of the data and spot outliers. In operating_hours data points above 14000 are considered as outliers. And fuel_consuption data ranges from 2450 to 5200 litters.
- To verify removal of outliers we plot box plot for each engine. There are no outliers in operating_hours.
- To get to know relation of data with time, we draw a line plot for every engine. As we can observe each engine has unique way of interaction with time but one thing in observation is it has linear pattern.
- Visualizing the data based on its ETS(Error-Trend-Seasonality) is a good way to build an understanding of its behaviour. That's what we did in In[13]. From that we can observe that data is following a linear trend for individual engine. Also it does have seasonal component. And very high residue.
- We apply additive model when it seems that the trend is linear and seasonality and trend components seems to be constant over time.(Here we used 10 day window size).
- Applying this model we are not getting results as we expected because data sometimes changing its trend very abruptly. And holding that trend for a time. From which we can conclude that data gives more weightage towards recent values than older once.
- Which leads to our next model EWMA(Exponentially weighted moving average).EWMA gives weightage to new data more than older data. Also from previous observation we know that both trend and seasonality are important aspects of data hence simple EWMA (which only considers trend(alpha component)) is not efficient.
- Hence we are choosing model which accounts both trend as well as seasonality which is Holt-winters EWMA. We are getting 108.5 of RMSE in an average.
- I applied same model to fuel_consuption data but as per observation fuel_consumption data is stationary. i.e. it does not exhibits trend or seasonality i.e best model for it could be ARIMA.(I didn't implement it due to time constraints)
- And for forcasting I trained previous model on full data set and by keeping window of 10 days, Forecasted 15 days of data for each engine. And stored it in respective excel file.As each column represents engine by its engine ID.