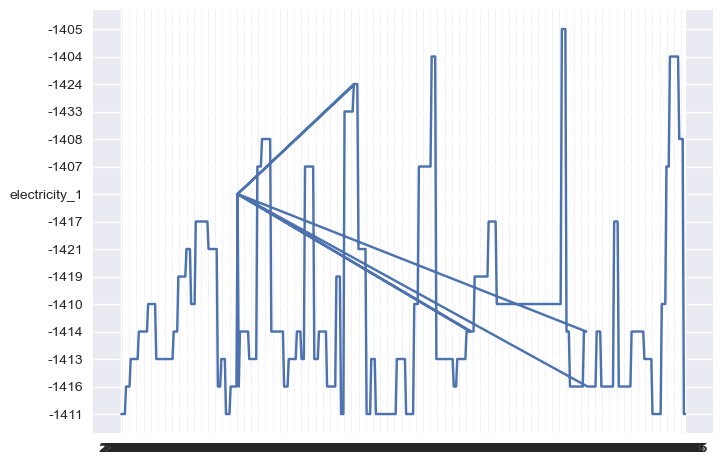
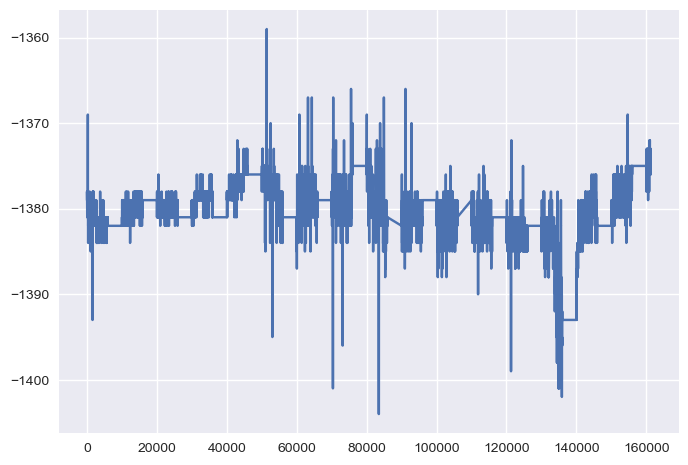
**CNCITY Dataset (exploratory data analysis)**

**Agenda:**

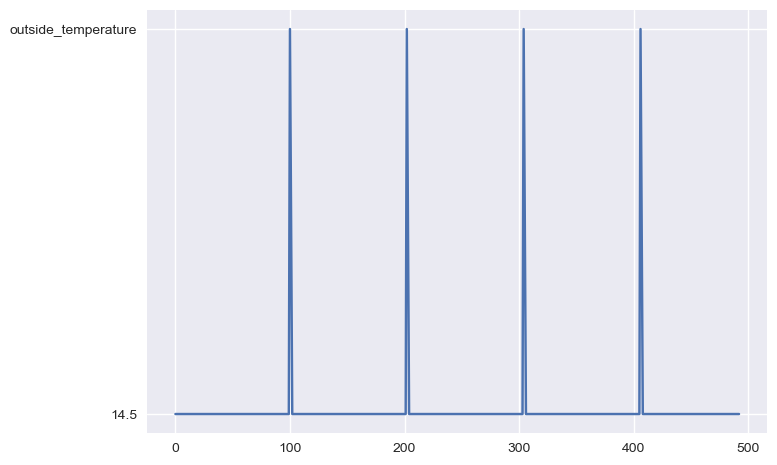
1. **Tabulate and clean the dataset, make it readable & computer-codable**
2. **Create a time series plot for each variable**
3. Tabulate and clean the dataset - import necessary python packages
   1. Tqdm: displays the progress bar: useful when working with large datasets
   2. Plt.plot → plots the variable
   3. Created a variable called ‘table\_df’: pandas dataframe with the first 500 rows
4. Need to convert each variable into a float or an integer before plotting them



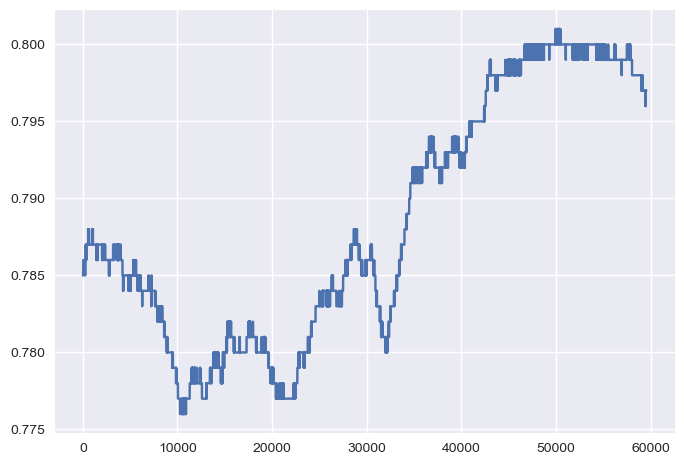
**X-axis = time, Y-axis = electricity\_1**



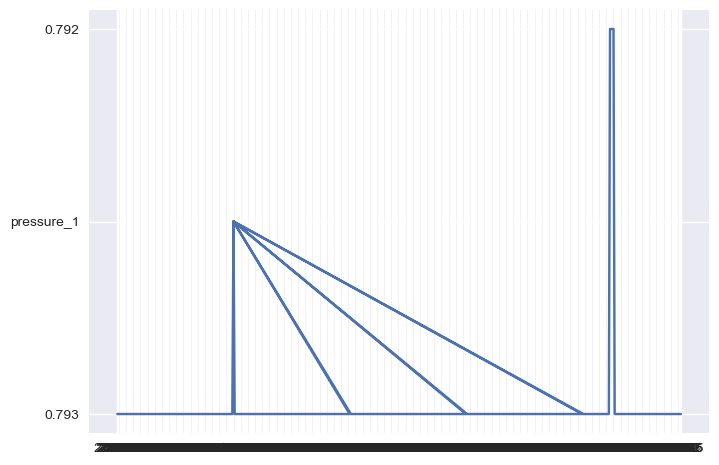
**<Outside\_temperature>**



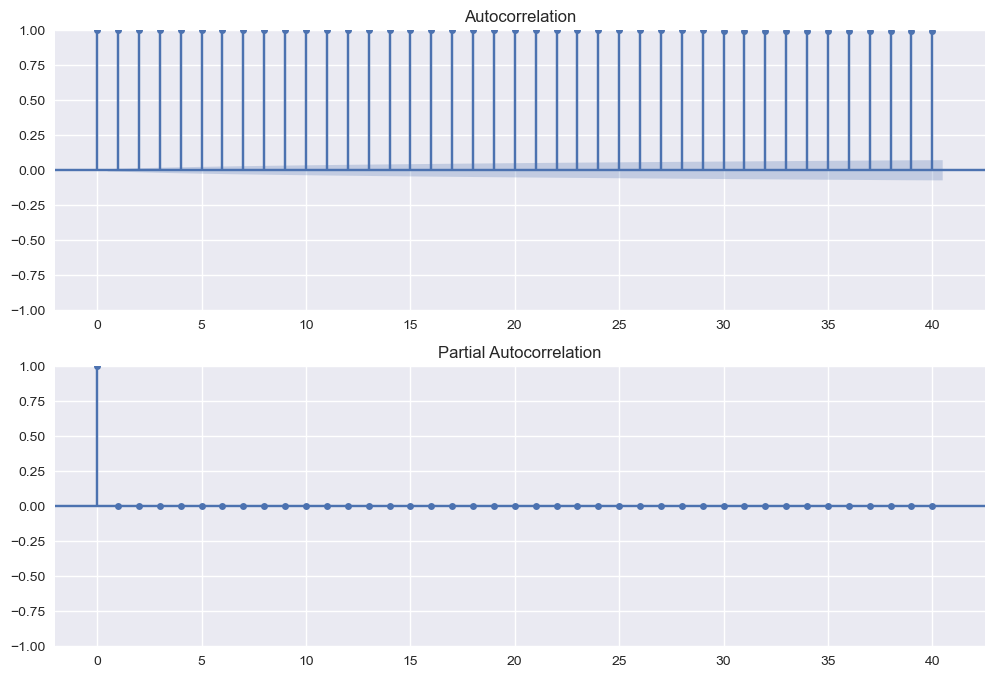
**<pressure\_1>**

****

**<x axis: time, y axis: pressure\_1>**



**<autocorrelation plot for pressure\_1 and partial autocorrelation plot for gas\_leak>**

****

**Autocorrelation:**

* The degree of correlation of a variable’s values over time
  + The correlation between a time series and the delayed version of itself
  + **Partial autocorrelation:** measures the correlation between a time series and the delayed version of itself at a specific lag, gives the partial correlation of a stationary time series

**Autocorrelation vs. Partial autocorrelation**

* **Partial autocorrelation excludes the contributions from intermediate lags**