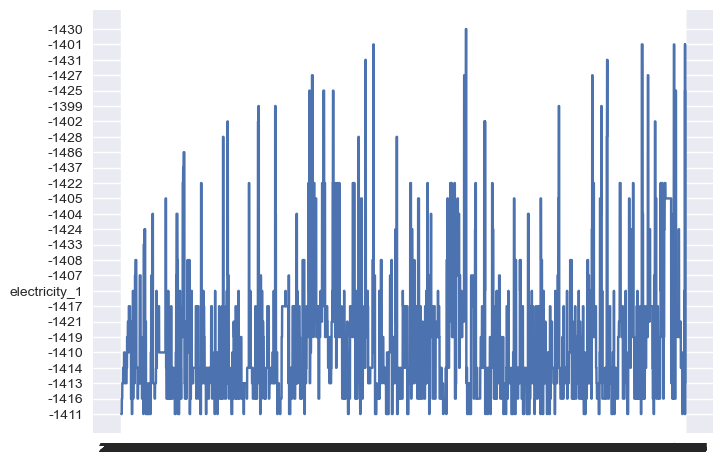
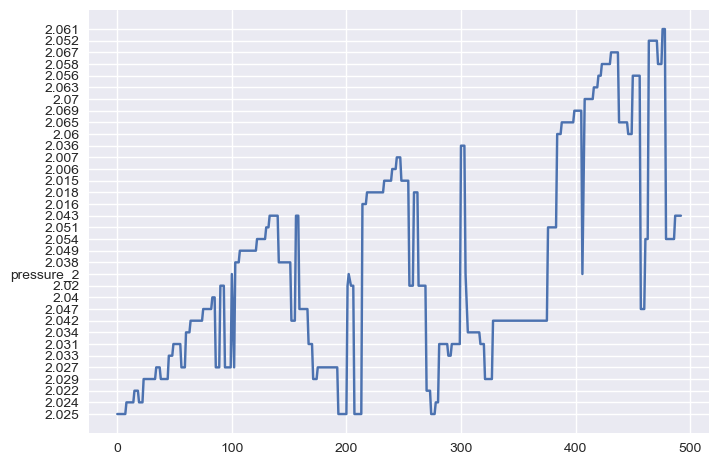
Exploratory Data Analysis (CNCITY) - (the actual report starts from page 8)

**Gas Pressure Regulator 1:**

**Electricity 1: (5000 lines)**

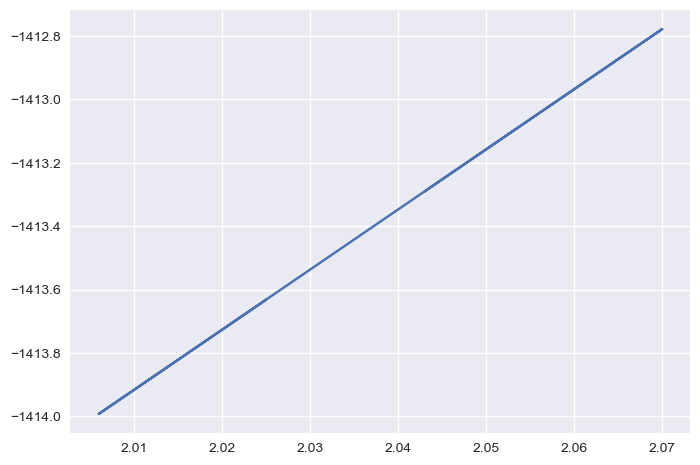
****

**Linear Regression: (pressure\_2, is\_maintenance) - epic failure**

****

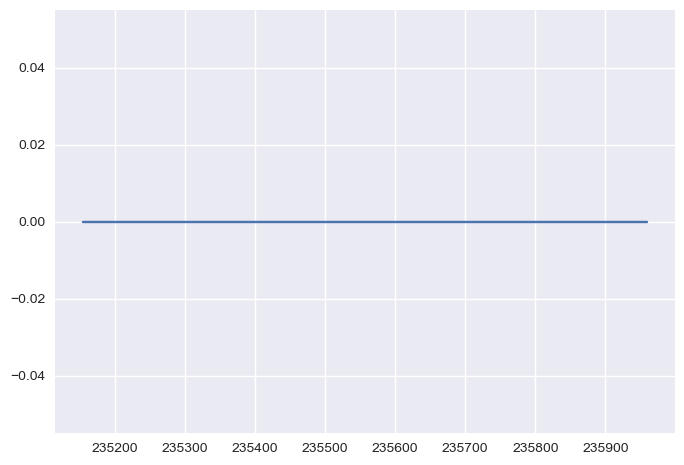
**Key takeaways from this graph:**

* **Since linear regression gives us a float between 0 and 1, logistic regression seems to be a better model when predicting is\_maintenance. (is\_maintenance is a binomial variable)**
* **It turns out that the ‘is\_maintenance’ column is just 0’s**

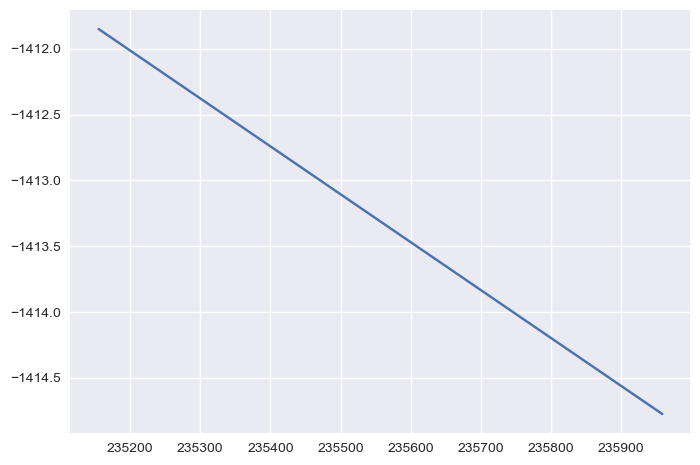
****

**Ran linear regression with ‘electricity\_1’ and ‘pressure\_2’**

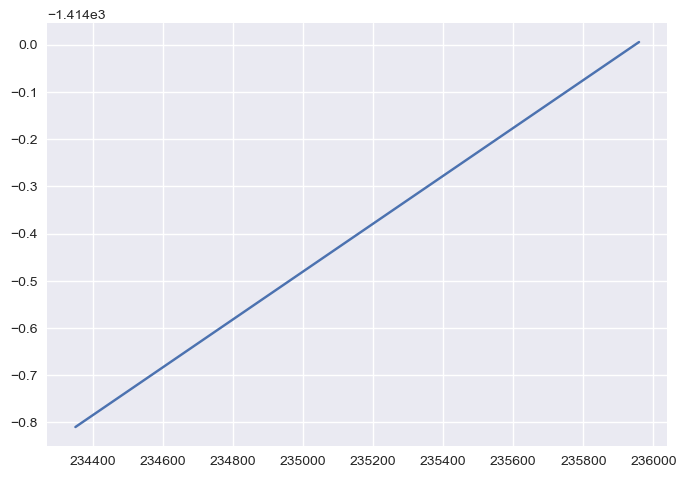
* **‘Electricity\_1’ and ‘pressure\_2’ are positively/linearly correlated**

****

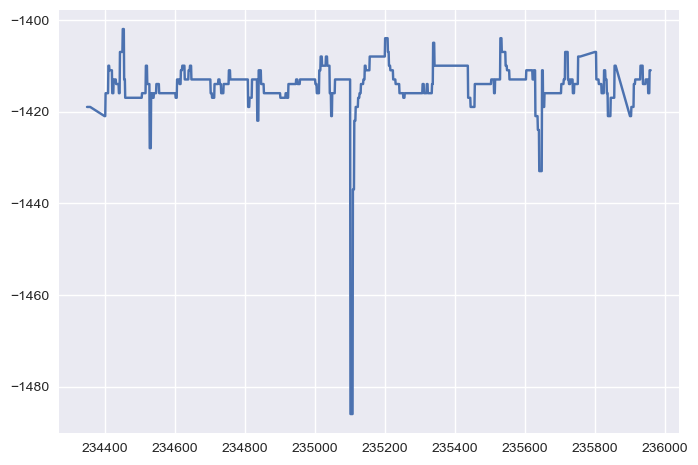
**X label: time, y label: electricity 2**

****

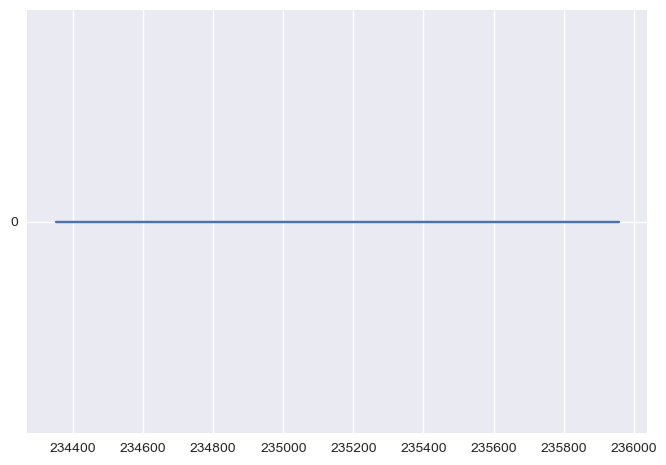
**Linear Regression: time, electricity\_1**

****

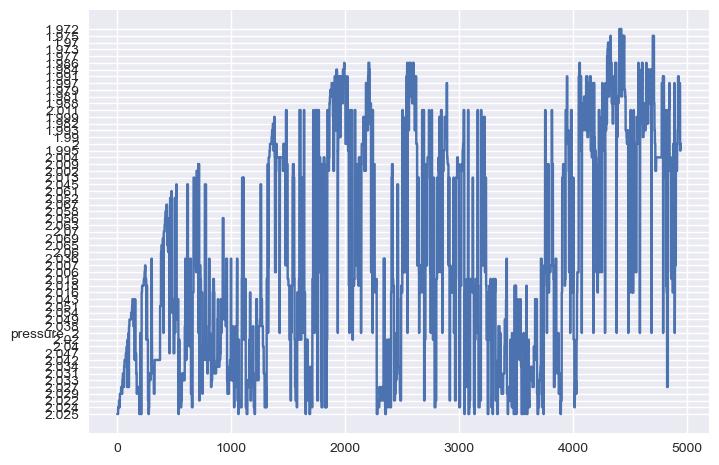
**The graph below is far more informative and intuitive.**

****

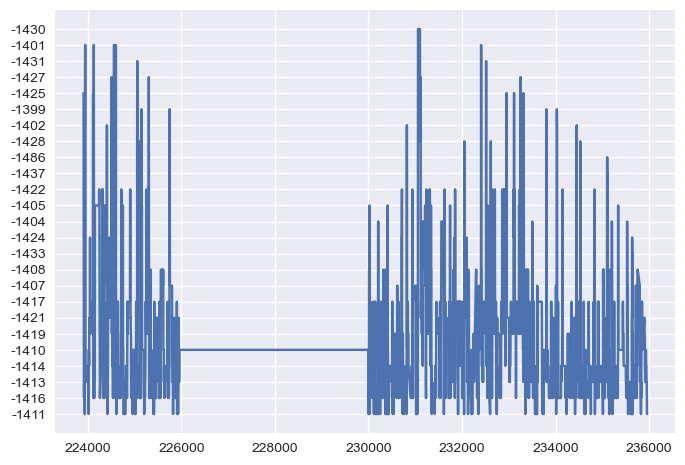
**Pressure\_1\_status and time**

****

**Pressure\_2:**

****

**X axis: time, Y axis: status\_electricity\_1**

****

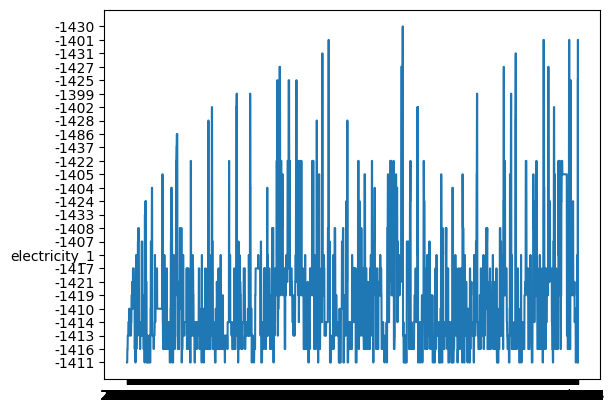
**X axis: time, Y axis: electricity\_1**

[**https://www.w3schools.com/python/python\_ml\_knn.asp**](https://www.w3schools.com/python/python_ml_knn.asp)

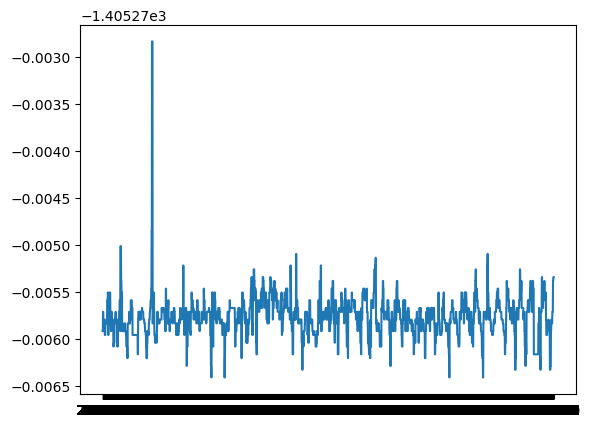
**6/11/2024:**

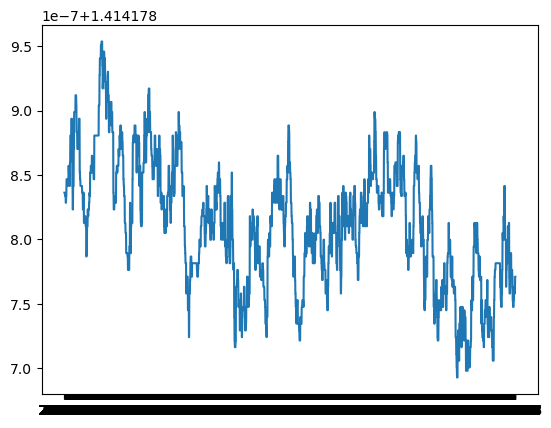
* **Query + connection manager**
* **Table\_df (5000+ lines)**
* **Stanford lecture videos**

**Gas pressure regulator (1) - EDA**

1. Visualization
   1. (graphs can be found in pages 1-7)  
      
      1. Electricity 1: no seasonality, data seems to be stationary
2. Split the dataset into train and test sets
3. Linear regression and logistic regression
4. Decision trees
5. Stationarity tests

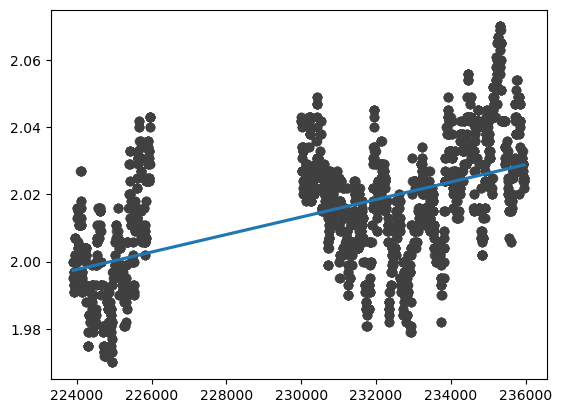
Plotted y\_pred\_new (used linear regression to predict the values of electricity\_1 over time)





‘Pressure\_2’ and ‘time’: model\_pressure\_2.intercept\_: array([1.41417356])

Model\_pressure\_2.coef\_: array([[2.60484877e-06]])

Linear regression plot (pressure\_2 and time)

Key takeaways:

row/sec : must read at least 86,400 rows