## Time Limit:

4 hours

## Dataset:

The dataset is provided as a .json file containing the **data** key. This is data collected across different types of voltage sensors put inside of a battery used in an Electric Vehicle (EV). To ensure the smooth functioning of the EV, it is important to collect and analyze EV battery readings across different voltage sensors. In the dataset provided, **tid** represents tag\_id of the voltage sensor, **timestamps** is a list/array that contains the consecutive timestamps at which the readings from the voltage sensor were recorded, **values** is a list/array that represents the readings of the voltage sensor at a given timestamp. The value read at a timestamp is linked by the array index. Hence if you need to find what value was read at the timestamp **timestamps[idx]**, it would be simply **values[idx]**.

## Assessment Requirements:

For this assessment, an exploratory data analysis (EDA) is required. Each voltage sensor’s time series should be plotted. The following are also required :

|  |  |  |
| --- | --- | --- |
| Question Number | Question | Answer Format |
| 1 | How many Outliers were removed? | 335,200 |
| 2 | How many Empty values were removed? | 710,215 |
| 3 | What trends did you observe (e.g. minutely/hourly/daily) in each | [*https://drive.google.com/file/d/1zUshNbdyLyfw-7OdLFnwb5rju-rEQrO7/view?usp=sharing*](https://drive.google.com/file/d/1zUshNbdyLyfw-7OdLFnwb5rju-rEQrO7/view?usp=sharing)  *Something else that I also found interesting was that majority of the time around 70 percent. The volatage data was on average slightly lower in the night time compared to the day. I found this out by setting a range of time for sunrise vs sunset during the month of may which I kept in mind.* |
| 4 | What is the mean value for each sensor tid alongwith the spread (standard deviation) | *https://drive.google.com/file/d/170EuIg-HIlowzmN\_5eYihw-ajK1wlI2q/view?usp=sharing* |
| 5 | Peak voltage times for each sensor tid | *https://drive.google.com/file/d/1SWoLTJU6iki9RjnNjZe4E3lXE0UJx21k/view?usp=sharing* |
| 6 | Low voltage times for each sensor tid | *https://drive.google.com/file/d/1jFjFST2liCAZ70LNZZ6\_VeDkmBYeVKlR/view?usp=sharing* |
| 7 | How Many potential anomalies were found? Describe your thought process | This is the part I struggled with the most with this assessment. I wasn’t able to find any anomalies. I tried a couple methods to find anomalies one of which was to check if the voltage varied a lot from the moving average. Another method I tried to do was to detect anomalies with a threshold of 3 standard deviations. I thought this would detect voltage values that were far off from the mean. I assumed that around 99.7 percent of normally distributed data are withing 3 standard deviations. Both methods however didn’t work out in my favor and I couldn’t find anomalies. |

Apart from the above, you are open to exploring the dataset and adding in any interesting insights you find as you develop an intuition about the dataset.

## 

## Submission Instructions:

A Github handle is mandatory for attempting this activity.

All submissions should be provided as a Jupyter Notebook with a requirements.txt file if one is required.  
Please be mindful of the 4-hour deadline. Submissions made beyond won’t be accepted.