**PROBLEM STATEMENT**

**Overview:**

Dream Vidyut is a modern age Electricity Generation company that dreams about electrifying the entire India in a sustainable manner. As part of their efforts, currently they monitor the electricity consumption of more than a dozen of Corporate Buildings in Delhi-NCR to approximately forecast the demands of these buildings so that Over-Production of electricity can be reduced and the resources be used elsewhere. Their current algorithms are only approximate and have high variability leading them to not trust the predictions completely.

This is where you as a Data Scientist comes in. Dream Vidyut has made this data available to you for creating a model better than theirs to forecast the electricity demand. Below are the details of the competition.

**Files available:**

1. train.csv - contains the historical electricity consumption of **5 buildings** measured across **three electrical meters for a period of 2017**

2. test.csv - contains the timestamps from year 2018 and the building ids for which predictions need to be made.

3. sample\_submission.csv - contains a sample submission file. The predictions that you submit should be in the same format as this file.

**Data Distribution:**

Training Dataset - Contains data for the year 2017 (April – Dec)

Test Dataset - Contains data for the year 2018 (Jan – 18th April)

**Evaluation Metric:**

Simple Average across errors of all 5 buildings where error for a single building is calculated as -

**C:\Users\jashpatel\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\8ADC799.tmp**

Where A, B and C represent Main meter, Sub-meter 1 and Sub-meter 2 respectively,

* T is the total number of timestamps in the test.csv for a particular building,
* t is the timestamp for which prediction is being made,
* k  = (ln 2) / 100
* d(t) is the value of the day in which timestamp t falls
* m(it) is the actual value of meter i at time step t
* Caret(m(it)) is the predicted value of meter i at time step t
* Bar(m(i)) is the mean value of meter i - This value is used for normalization

\*(e-kd(t), is basically added to down weight predictions farther in future)

An example excel workbook to understand the metric calculation can be accessed here. Also, its attached in mail case the link doesn't work.

**Rules:**

Future data cannot be used to make past predictions.

For example – you cannot use data of February 2018 to make predictions for January 2018 and so on.