

## Amperon - Data Science Candidate Assignment

Thank you for taking the time to apply for Amperon for a data science position.

In this assignment you will have to make predictions regarding the peak load hours given some datasets.

Please use a Jupyter notebook for your solution using Python3.

### Data sets:

- 4 years of weather data in the file “weather\_data.csv”
  - There are 28 stations with hourly weather for each
  - Each row has date, hour of the day (1-24), Station ID, temperature (Fahrenheit)
- 3 years of load history in the file “load\_hist\_data.csv”
  - Times are matching weather data
  - Each row has date, hour of the day (1-24), load for that hour
- You can use US holiday information

### Task:

Compute the **probability** of each hour of the following year (2008) being the peak load hour of the day.

A file named “probability\_estimates.csv” is provided with each row having a Date, Hour and an empty column for the daily peak probability, which you will have to fill in.

For simplicity no day has two peak hours.

Please note that hours here are (1-24) and not (0-23).

### Error measurement:

The following scoring rule will be used to evaluate your forecast for each given hour:

$$S = (p-a)^2$$

where:

- p is the forecast probability
- a is the actual outcome (0 if not the peak hour, 1 if it's the daily peak hour)
- S is the score for that hour

For example,

- If the forecast is 1 (100%), and that hour is the daily peak hour, you get a score of 0, which is the best achievable score;
- If the forecast is 1 (e.g., 100%), and that hour is not the daily peak hour, you get a score of 1, which is the worst achievable score;
- If the forecast is 0.8 (e.g., 80%), and that hour is the daily peak hour, you get a score of  $(0.8 - 1)^2 = 0.04$ ;
- If the forecast is 0.2 (e.g., 20%), and that hour is the daily peak hour, you get a score of  $(0.2 - 1)^2 = 0.64$ ;
- If the forecast is 0.5 (e.g., 50%), you get a score of 0.25, no matter the hour is daily peak hour or not.

The final score the sum of scores for all hours in that year. The lower the score, the better the solution is.

Good luck! ☺