Solar power plant data analysis summary

Analysing performance data from a photovoltaic solar power plant in India

Context

- A renewable energy supplier owns solar power plants in India.

- Two of those power plants present power production anomalies affecting the total power generation and thus reducing business income.

- Before sending the technical team to fix the problem the data science team has been required to analyse power plant sensors' data to make a first approximation to the root cause of failure of the equipment.

- What should the data science team do? How should they approach the problem?

Methodology

1. Set the project objective

2. Establish the variables that affect the business objective

3. Define the KPIs of those variables

4. Enumerate and classify the relevant entities that are involved and the data that can retrieved from them

5. Determine the first questions that need to be answered in order to start finding relevant insights about each of the variables affecting the business objective

6. Analyse the data and look for insights

7. Suggest an action plan

For the complete methodology description please refer to [this](#) link.

PROJECT INSIGHTS

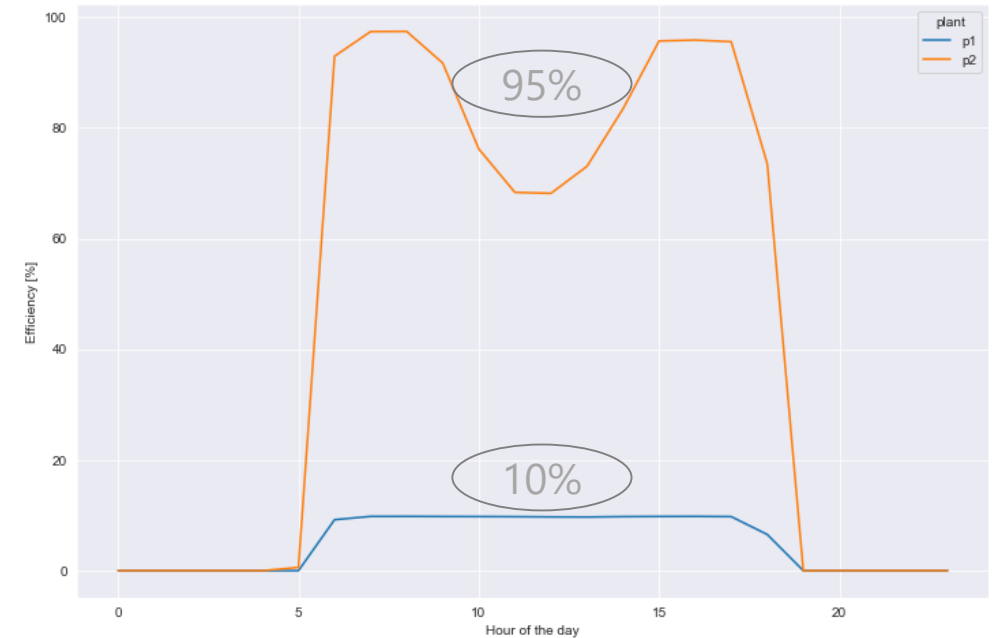# There are important data quality issues that suggest treating the following statements as not conclusive

DC power generated by plant [kW]
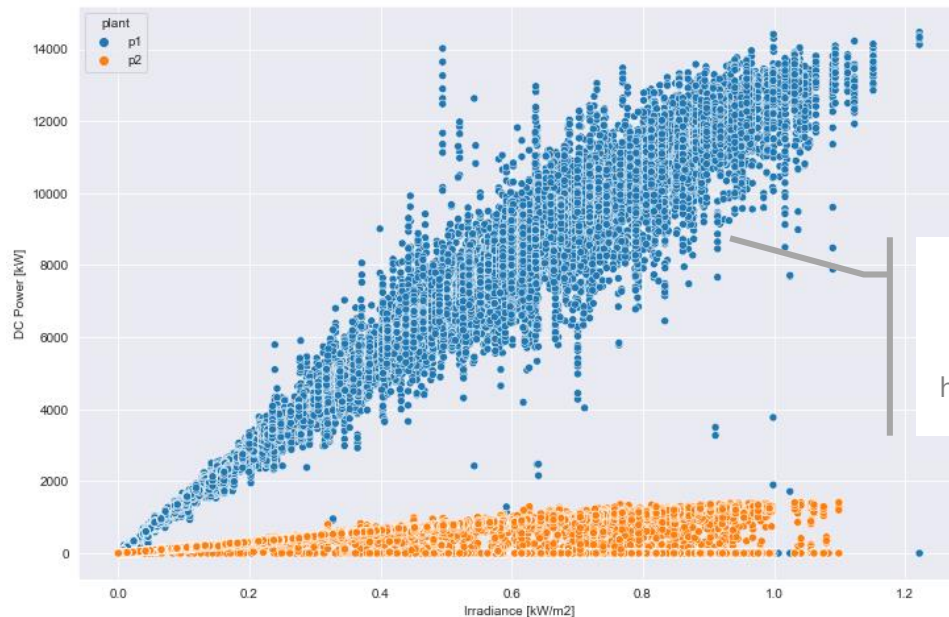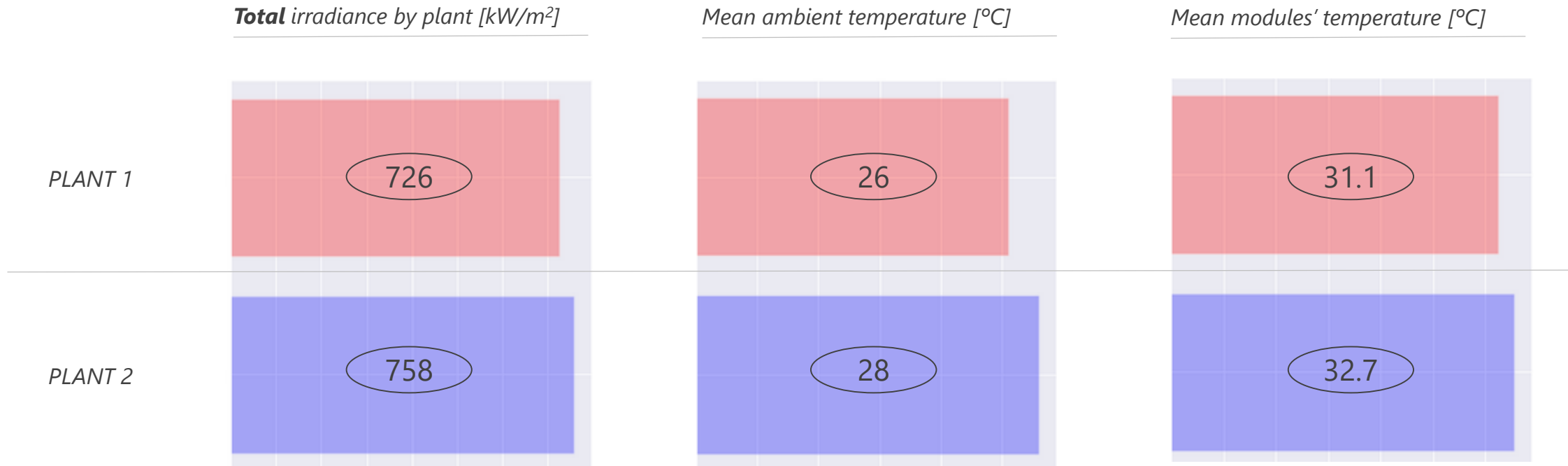

PENDING A SECOND REVIEW

DC generation in power plant 1 is around 10 times higher than in plant 2

Plant efficiency throughout the day [%]

It has been required to review the data acquisition system to check if it has been a data quality issue.

Both power plants receive high and similar amount of irradiance, with no sign of power availability problems

	Total *irradiance by plant [kW/m²]*	*Mean ambient temperature [ºC]*	*Mean modules' temperature [ºC]*
PLANT 1	726	26	31.1
PLANT 2	758	28	32.7

The small difference found between both power plants suggest that they should have very similar DC power generation capabilities.

# DC power generation in plant 1 is very high, on the scale of 10 MW daily generation

PLANT 1

Mean hourly DC power generation by day [kW]
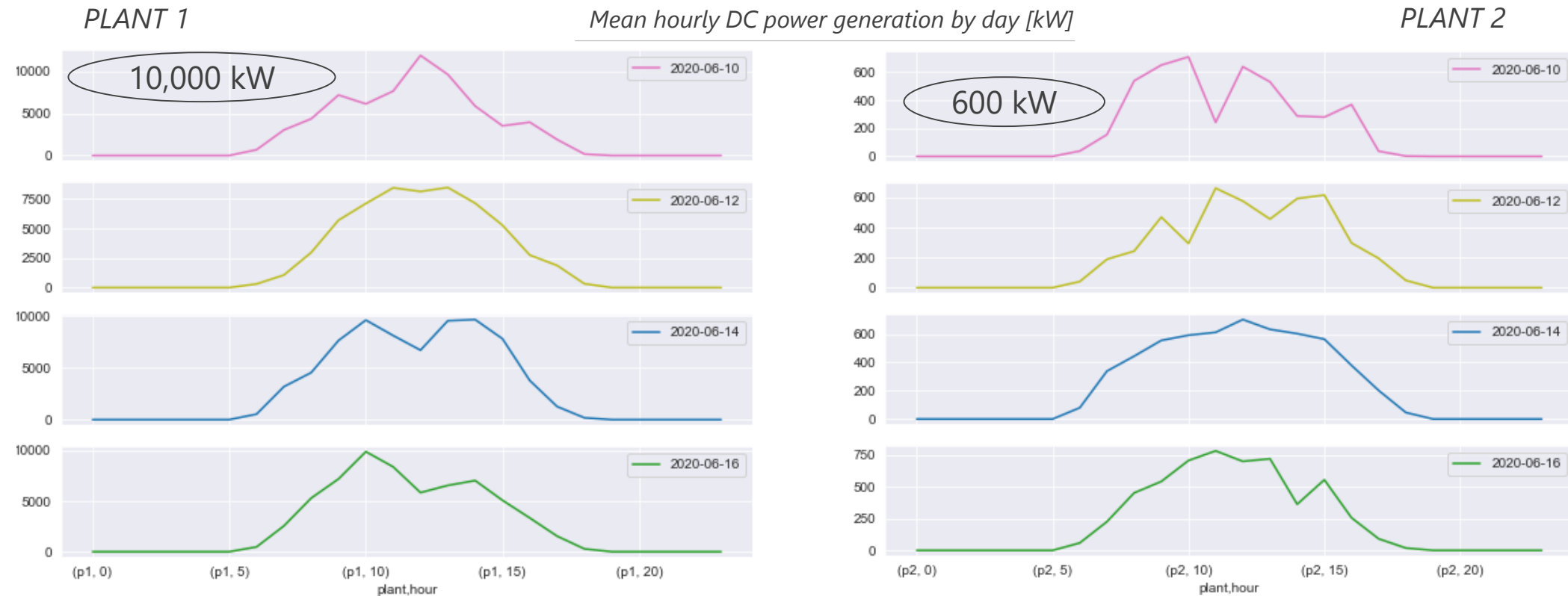
PLANT 2

Plant 1 and plant 2 have the same number of inverters but the DC power produced is around 10 times higher for plant 1. It is suspected that DC power generation in plant 1 is not measured correctly or artificially scaled up.

There are some inverters in plant 2 that do not produce DC power 30% of the time during irradiance peak hours

Mean inverters' efficiency [%] distribution between 09:00 and 15:00
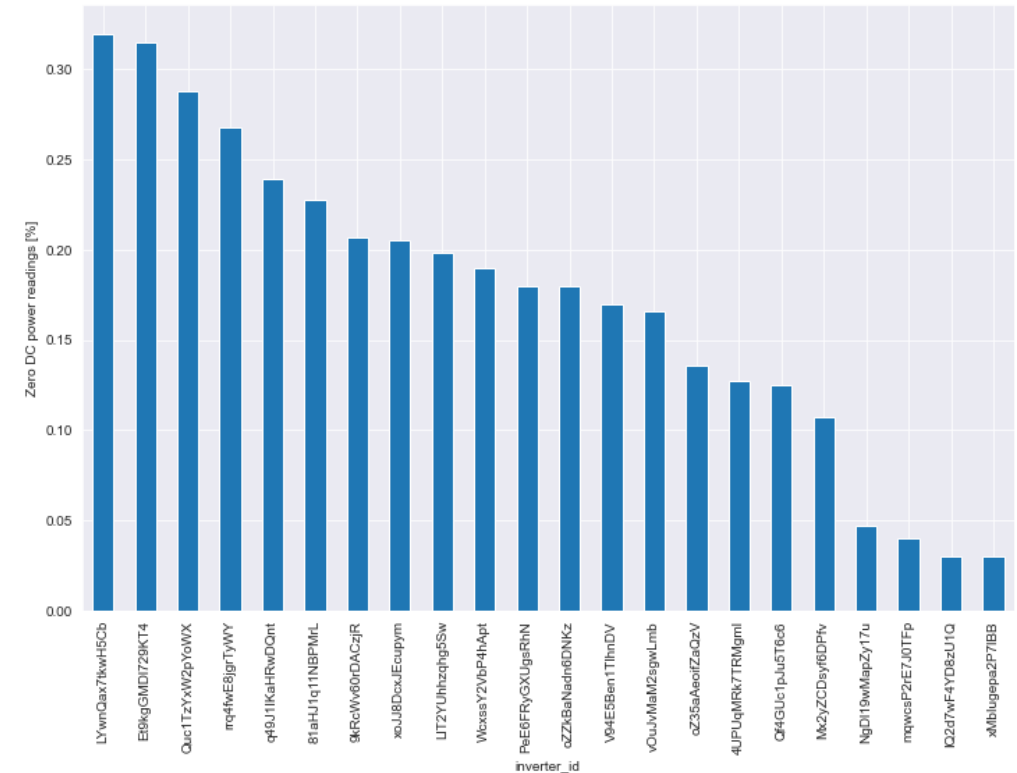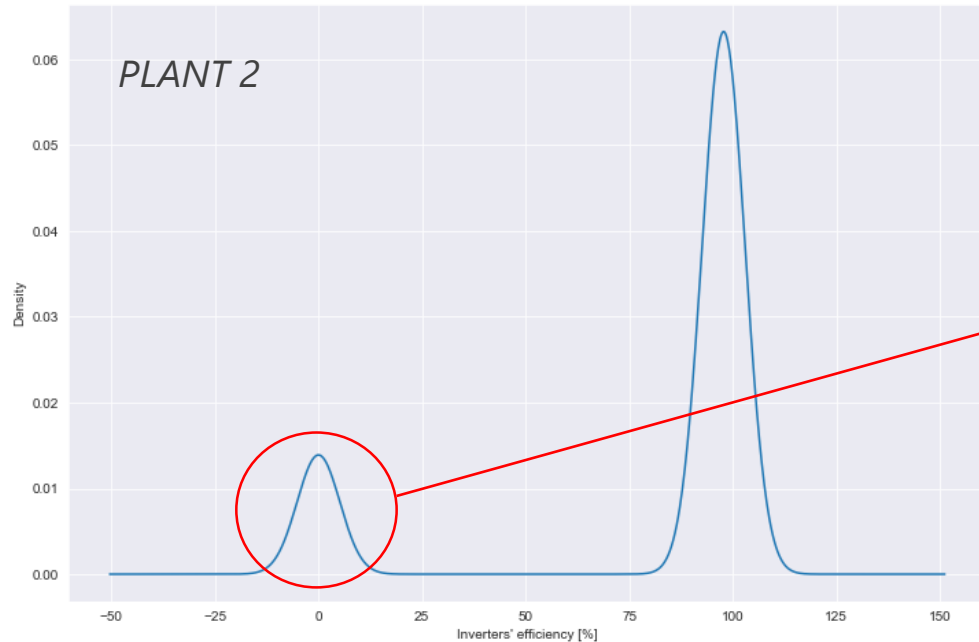
It is highly suggested to check those inverters from plant 2 in the next maintenance check.

After carrying out the data analysis, the following action plan is suggested:

	Data quality	Irradiance	DC Production	Inverters' efficiency
PLANT 1	❌	✅	❌	❌
PLANT 2	❌	✅	✅	❌

ACTION #1: Overall data acquisition process check.

ACTION #2: Make sure there is no artificial scaling when measuring DC production in plant 1.

ACTION #3: Maintenance of all inverters from plant 1 and the selected inverters from plant 2.