# Cleantech Solar – Internship Assignment

**Objective:** Replicate the below graph using the data provided.

Chart

Description automatically generated

**Dataset:** The dataset needed for the creation of this graph has been provided to you. It consists of three columns:

1. Date (July 2019 to present)
2. PR (Performance Ratio) – This parameter is used to track the daily performance of the PV plant. A high value indicates that the plant is performing well and there are no issues.
3. GHI (Irradiation) – This parameter tracks the total irradiation for a particular day. A high value indicates a sunny day.

***Important Notes:***

* The **red line** on the graph represents the 30-d moving average of the PR (Performance Evolution) whereas the scatter points depict the GHI.
  + - * The **dark green line** represents the budget line. The value begins from 73.9 and should reduce by 0.8% every year (Do not hardcode the values). As you can see the, the values are:
* 73.9 for the first year (July 2019 to June 2020)
* 73.3 for the second year (July 2020 to June 2021)
* 72.7 for the third year (July 2021 to present)
* The points for the scatter plot are **colour coded** (as per the legend above). That is: if the GHI [Daily Irradiation] is:
  + Less than 2: Navy blue
  + 2-4: Light blue
  + 4-6: Orange
  + >6: Brown
* The points above **Target Budget PR** represent the number of PR points above the Budget PR for that particular year.
* The **bottom right section** of the graph simply shows the average PR for the last 7 days, the last 30 days, last 60 days and so on.
* Please note that the values and the trends will not match the graph exactly since we have changed the data slightly.

**Files that you have to submit:**

* The completed output graph generated by you.
* The code you used to generate the graph (preferably in Python)