**DATASCI 200: Introduction to Data Science Programming**

**Project 2 Proposal**

Zejia Jing ([jzejia1@berkeley.edu](mailto:jzejia1@berkeley.edu))

Jianyi Teng (jianyiubc@berkeley.edu)

Minghung Cho(ming.cho@berkeley.edu)

Team's GitHub repository:

<https://github.com/UC-Berkeley-I-School/SEC6_ZEJIA_JIANYI_MING>

**1) Overview**

The goal of this analysis is to research the relationship of location data, time data, and weather data including cloud type to real-time solar radiation (evaluated in GHI).

This analysis will also research the relationship between real-time solar radiation and Photovoltaic (PV) Panel Power output and estimate the efficiency of different types of PV panels.

**2) Data**

The solar radiation data is downloaded from the National Solar Radiation Database (NSRDB) developed by the National Renewable Energy Laboratory (NREL) [1].

The PV module power output data being researched is measured and validated in the NREL project [2]. The picture of experiment PV Modules is shown in Fig. 1.

Solar panels on a roof

Description automatically generatedA person standing next to a solar panel

Description automatically generatedA yellow container with solar panels

Description automatically generated

Fig. 1. PV modules and solar radiation measurement devices deployed on the Golden, Colorado(left); Cocoa, Florida(middle) and Eugene, Oregon(right) 2

**3) Plot (sampled)**

A close-up of a number

Description automatically generatedA graph of blue dots

Description automatically generated

**4) Sample Size**

The research sample will include the NSRDB solar radiation data in three locations (Cocoa, Florida; Eugene, Oregon; and Golden, Colorado) from 2012 to 2013. Besides, the sample will also include the real-world PV panel power output data from 2012 to 2013 in these three locations with a solar power output of 11 types of solar panels in each location.

**5) Experiment**

The solar radiation data includes these variables we intend to explore:

* Weather data: temperature, cloud type
* Time data: year, month, day, and hour
* Location data: latitude and longitude
* Solar radiation data: GHI

We expect to find a strong relationship between weather data, time data, location data to solar radiation data. Besides, we will summarize the data with the average estimated solar radiation based on different weather, time, and location. This analysis will give users a good understanding and estimate of solar radiation based on their location, time, and weather conditions.

The PV panel data includes these variables we intend to explore:

* Solar radiation data: GHI
* Solar panel material type: 11 types
* PV power output

We expect to find a strong relationship between solar radiation and types of solar panel material to PV power output. Besides, we will compare the estimated solar panel efficiency of different types of solar panels with their manufacturer specifications. This analysis will give users a good understanding of the real-world efficiency of different types of solar panels.

**6) Deliverables**

First, the relationship and estimate value of solar radiation in different weather, time, and location parameters will be summarized in tables and plots.

Secondly, based on the PV solar energy output and solar radiation, the solar panel efficiency of different types of panel materials will be estimated and compared to manufacturer specifications.

**7) Reference**

[1] National Solar Radiation Database (NSRDB): https://nsrdb.nrel.gov/

[2] W. Marion et al. , “User’s Manual for Data for Validating Models for PV Module Performance,” 2014, Accessed: Jul. 31, 2022. [Online]. Available: https://www.nrel.gov/docs/fy14osti/61610.pdf.