# Capstone 2 - Getting the most solar power for your dollar

Prepared by Zachary Brown

**Problem Statement:**
The goal of this project is to determine what features influence the cost of a residential solar panel installation and then use this model to identify the parameters needed to optimize cost efficiency in Austin, Texas.

**Context:**
Electric grid reliability is becoming increasingly concerning across the US as temperatures rise and climate events like the winter storm in Texas pressure the electrical system throughout the country. One strategy to continue meeting electricity demands is to install solar panels on buildings that will feed electricity back into the grid when the building has excess. The Tracking the Sun dataset from Lawrence Berkeley National Laboratory compiles data on photovoltaic installations throughout the US from 2017 through 2021 including system size, cost, location, manufacturer, installer, and whether the system was installed with any grants or subsidies. This project will delve into this dataset to identify what factors contribute to the cost efficiency of these solar panel systems and will then forecast the cost efficiency of a solar panel installation throughout the US.

**Criteria for Success:**
This project will be deemed successful when a model has been generated that predicts cost efficiency of a solar panel installation. This model will then predict the optimal equipment, installers, project size, and time of year for an installation in Austin, Texas.

**Scope of Solution Space:**
This study will look primarily at price, equipment, installers, and location to identify any key features that dictate cost efficiency. This solution will apply to solar panel installations throughout the continental United States and will be valid for the remainder of 2022. Once the Tracking the Sun 2022 data has been released the model can be updated for 2023 installations.

**Constraints:**
This dataset does not contain data on how much energy is generated from these solar panels, so the cost efficiency measure will be based on the cost divided by the system size (in kW). This will not take into account how much actual energy is produced by the panels and even if weather data were incorporated into the study there would still be assumptions in terms of sun exposure translating to power generation, so that is one inherent weakness of the dataset. This project will also focus only on residential installations and will not apply to commercial installations.

**Stakeholders:**
The stakeholders in this project are residents of Austin Texas, like myself, who might be interested in a solar panel installation. This project will help those residents get the best deal possible when installing solar panels.

**Data Sources:**
The data used for this project comes from the Tracking the Sun project at the Lawrence Berkeley National Laboratory. This project collects data from over 2 million solar panel installations across the continental United States in 2021. The data includes the installation date, system size, total price, value of any rebates or grants, the type of customer, whether the system is a new build or expansion, the location, tilt and azimuth angles, utility servicer, module manufacturer, installer name, and module information.
https://emp.lbl.gov/tracking-sun-tool - 2021 data in .csv format
https://github.com/openEDI/documentation/blob/main/TrackingtheSun.md - project documentation with examples of how to connect to the Parquet data
https://data.openei.org/s3_viewer?bucket=oedi-data-lake&prefix=tracking-the-sun%2F - data from 2018 through 2021 in Parquet format

**Strategy:**
To determine cost efficiency I must first calculate the cost of an installation while factoring in any rebates or grants. This value must then be divided by the system size to determine the cost per kW for each installation. This will be the key metric which I will seek to minimize with our model. I'll then develop models with optimized variables, cross validate, and then determine the minimum cost per kW possible.

**Deliverables:**
This project will generate the following:

- A model
- A report summarizing the findings of the project
- A slide deck presenting the findings of the project