How much will your Solar Installation cost?

Objective

In an effort to tackle a major issue like climate change, more investment needs to be focused on renewable energy technologies such as Solar energy. A major challenge in implementing renewable energy technologies is the cost. During the nascent stages of developing Photovoltaic technology scaling it has been a major challenge as adoption was difficult due to cost concerns.

Over time, the efficiency in implementing the technology has improved and there has been strong economic support from the United States Federal government in terms of policies and incentives. Thanks to this, there has been a steady growth in the adoption of solar installations both in residential and commercial sectors.

As investment efforts in these fields have grown, there is a need to be able to identify opportunities where costs can be reduced further. This is where a data driven solution comes into the picture. An analysis of historical PV installations can give an insight into what worked and what did not.

The data driven solution described in this report carries out an analysis to predict the cost of an installation based on the factors related to Solar energy technology alone. The focus of this report and this project is on residential systems alone and the cost calculated is prior to any deductions, or tax incentives.

Who Benefits?

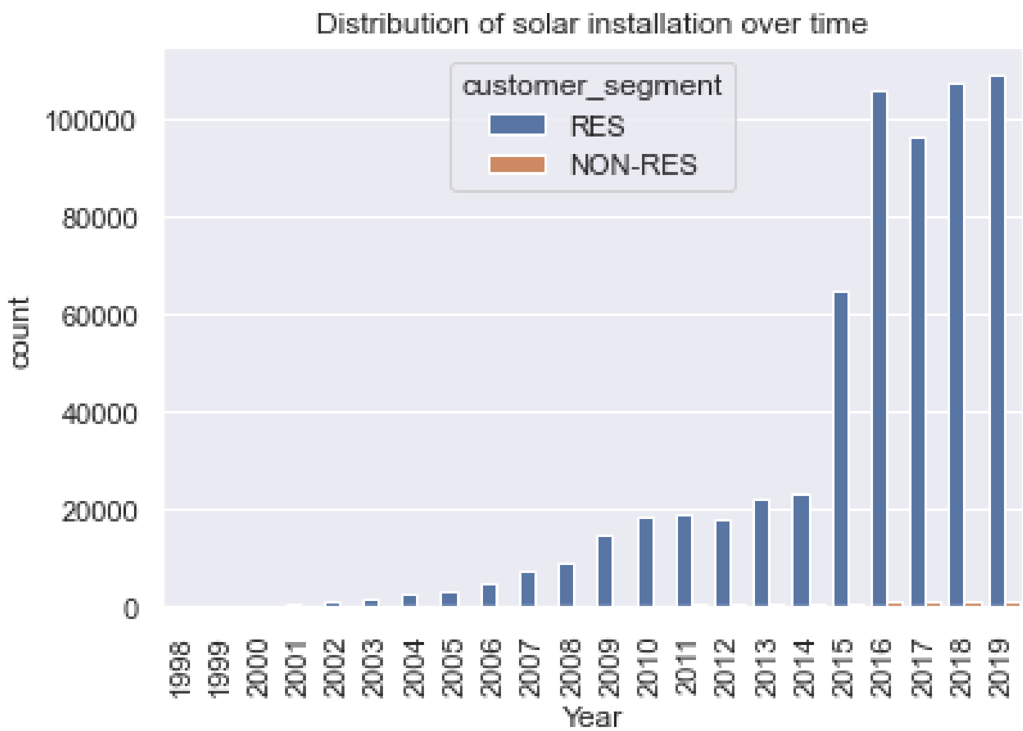
Any institute or organization that aims at implementing solar energy installations can be a beneficiary of this kind of research. Some of the primary stakeholders that can be identified are :

- Photovoltaic system installers : They are consultants who implement the technology for the end consumer
- Policy makers - Organizations such as the United States Federal government can use this to identify key areas for further investment in the form of subsidies and incentives.

Data Source

A team of researchers at Lawrence Berkeley National Laboratory as a part of their initiative - Tracking The Sun, have been developing a PV installation data set that they have been kind enough to make open source. The analysis for this project has been done using that dataset. The data set has about 1 million data points with 78 feature columns. Through a systematic process we consider points only for residential systems and select 13 essential features for our prediction. Additionally some features used in the regression analysis have been calculated as recommended in the regression analysis conducted by Lawrence Berkeley National Laboratory.

While the system has historic data from 2000, we consider installations for the years 2018 and 2019. Additionally, the focus has been on predicting cost for residential systems only. These are systems that have a system size of between 5kW to 100kW.



As you can see a majority of the data in the data set seems to be for residential systems, hence a pricing model only for that specific segment seems more relevant here.

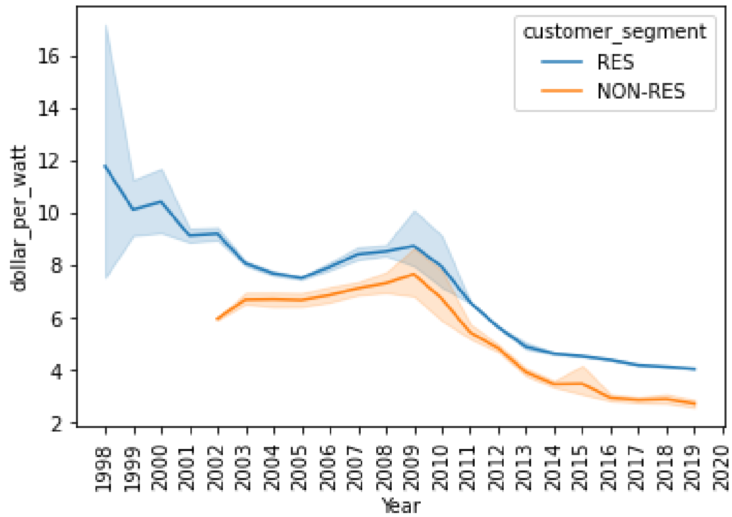
Lets get some insights!

While the data for the prediction model focuses on installations for the years of 2018 and 2019 only, the original reference data set has installations from the year 2000. We use this data to visualize some trends.

Let’s start by listing down the features we are using in this model and then we will also analyze the trends for those features.

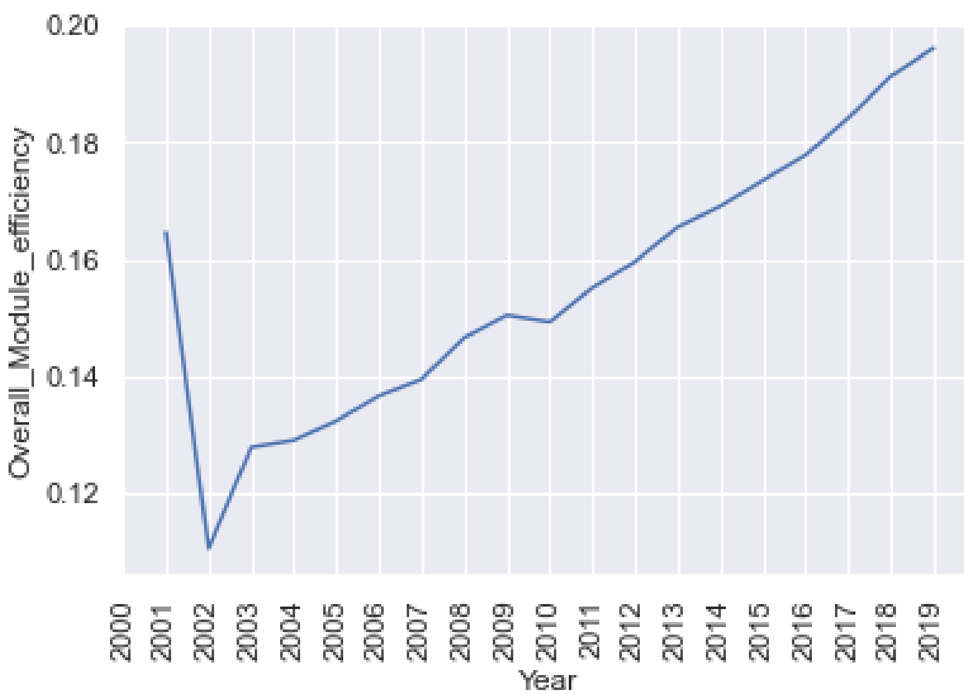
- Year - The year of installation
- System Size in kW - DC system size of the installation in kW
- New Construction - Dummy Variable indicating if the system installation is a new construction or an upgrade of an existing installation
- Ground Mounted - Dummy Variable indicating if the system is installed on the ground or the rooftop
- State - Location of the installation in the United State. States covered are CA, MA, NH, FL
- Installer Name - Name of the Solar system installer
- DC Optimizer - A device that helps improve efficiency of the system. This feature is a dummy variable indicating if the system has a DC Optimizer or not.
- Dollar per watt - The dependent variable that is of interest in this problem. Calculated values are in $ / kiloWatt
- Overall Module Efficiency - Numerical value of the overall efficiency of the PV modules being installed
- Micro Inverter - Dummy variable indicating if the installation includes a micro-inverter technology
- Installer experience - Uses the date of installation and Installer name to calculate the total number of systems installed to that date.
- Premium Module - Dummy Variable indicating if the system uses a Premium Module i.e a module with efficiency >20%.

Pricing

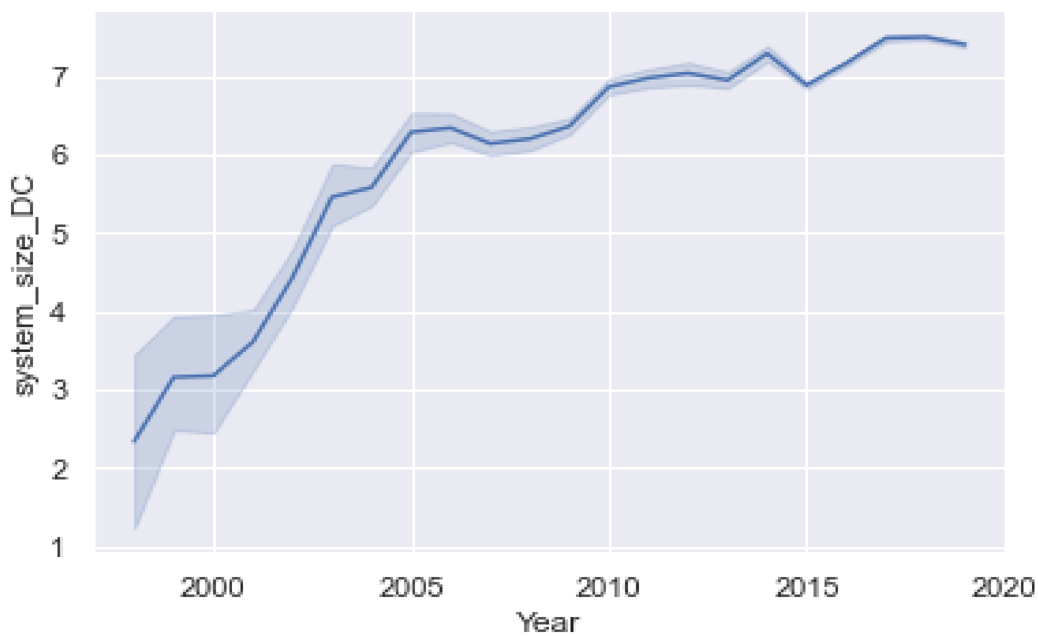


We start with the question we are looking to answer through this system. The cost of installation in terms of $ per kiloWatt energy installed. The good news is there has been a decrease in price over the years!

System size and module efficiency



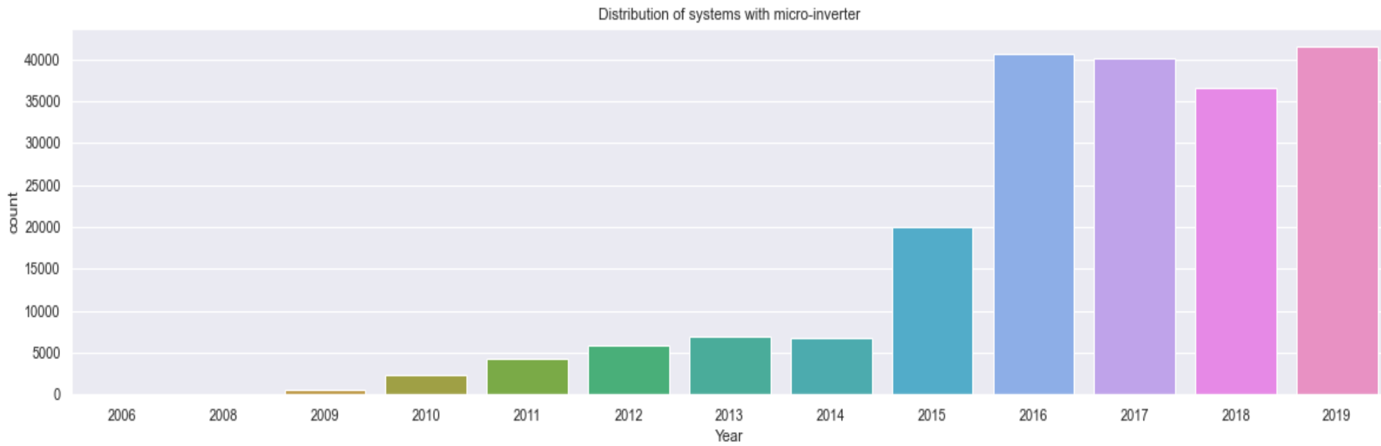
PV modules are the heart of solar energy technology. They are the devices that convert sunlight into electricity. The big plates you would typically see on a rooftop, that's them. Module efficiency is basically telling us how good the conversion process of sunlight to electricity is. Over time there has been an increase in the efficiency of modules being used. During the initial stages we see that there was a drop in efficiency before it picked up again. This can be due to different reasons. If we look at the number of installations in the years 2000 - 2002, there aren't too many. Hence it could be that at the time the technology was still experimental. From the year 2002 onwards, we note a good increase in module efficiency.

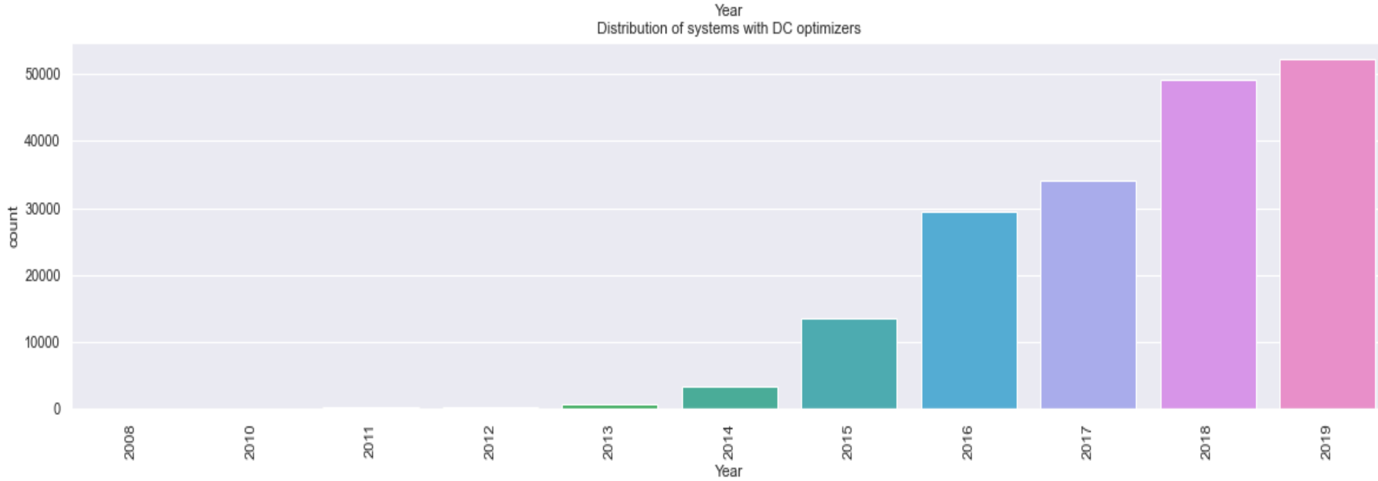


As the module efficiency increased this increased the output per panel and hence that could be one of the reasons for the increase in system size as we see in the image above.

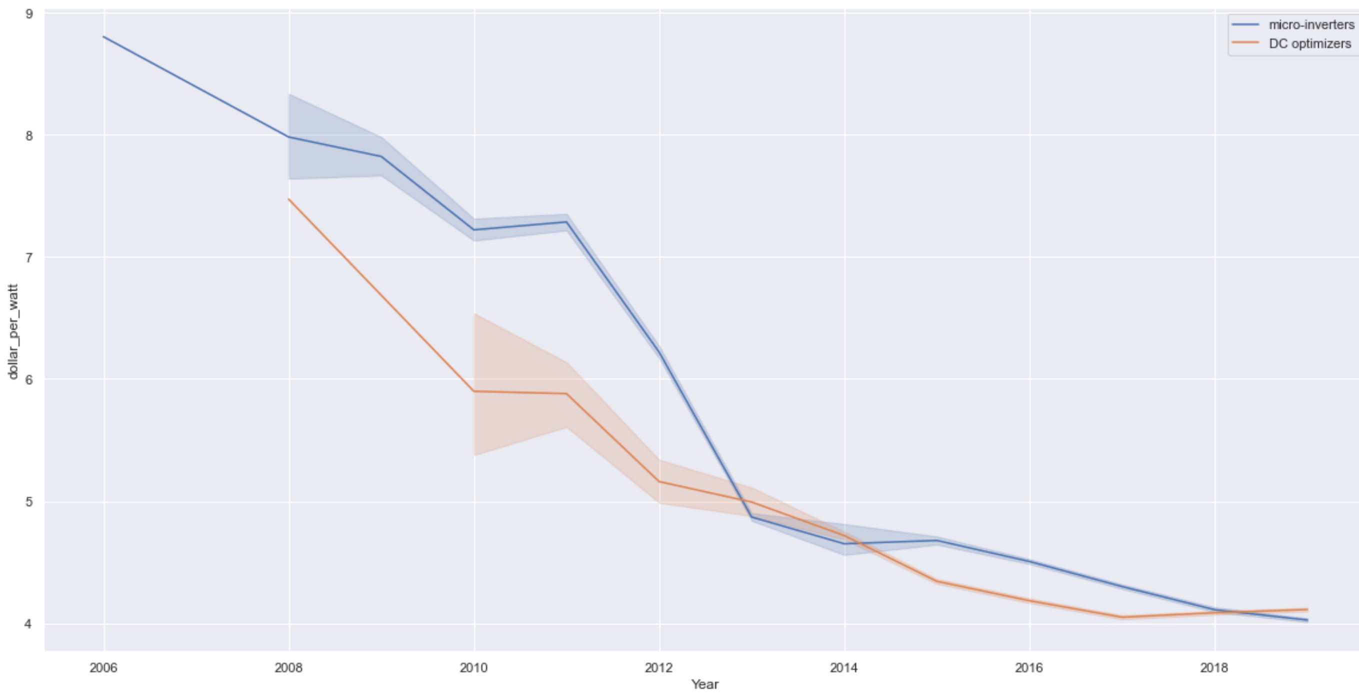
Efficiency boosters

Along with PV panels, there come supplementary devices that help in improving the efficiency of the system thus allowing for a greater power output. This data focuses on two such devices - Micro-inverters and DC optimizers. Collectively these devices can be referred to as Module Level Power Electronics or MLPE. We see a steady increase in adoption of MLPE over the years.





In the trend below we also see that with the increased adoption of MLPEs there has been a decrease in cost. This is interesting, as one might think that more sophisticated technology costs more in terms of hardware and labor costs. However it might be possible that the long term benefits probably offset the initial investment costs. This opens the door for looking into other factors (probably non-technical?) that affects the system pricing. This is certainly an idea for future research.



micro-inverters
DC optimizers
dollar_per_watt
Year
9
8
7
6
5
4
2006
2008
2010
2012
2014
2016
2018

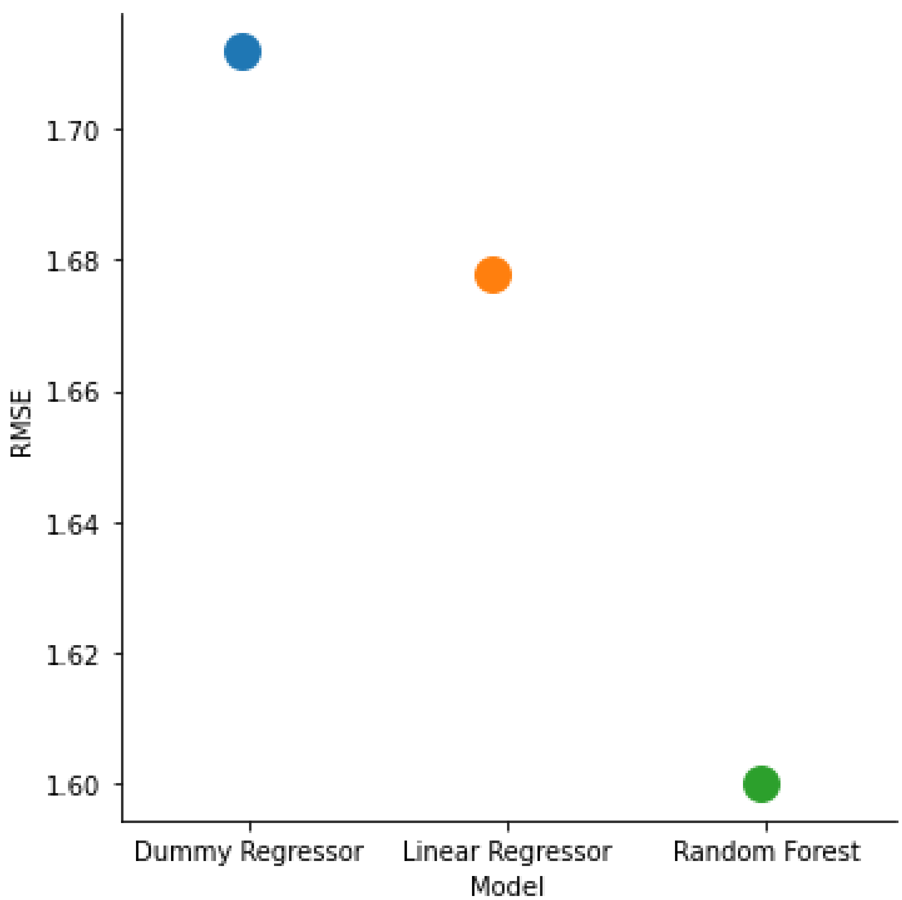
Model Selection

As a part of the regression analysis three separate models were evaluated based on the Root Mean Square Error metric. The end goal is to be able to predict the price of an installation within a certain margin of error. The Root Mean Square Error metric inherently penalizes predictions that have a higher error margin, thereby giving a more reliable or accurate price prediction.

The models created were:

1) Dummy Regressor - Uses the Mean of the data set as the predicted price
2) Linear Regressor - Each feature is given a weight or coefficient and the price is calculated as a weighted sum of the product of the coefficients and the feature values. Performs better than the Dummy Regressor Model.
3) Random Forest Regressor - The most accurate model and also computationally the most expensive. The model predicts the value based on decision trees.

As mentioned before we evaluate the performance based on the Root Mean Square Error and we can see that visually based on the plot below.



We can see from the model evaluation chart above that the Random Forest Model gives the lowest Root Mean Square Error. What does this mean? It basically implies that if the model comes up with a price prediction of $5 per kiloWatt (kW) for an installation system, we can say that the variation in that price on an average will be within $1.6 per kW of power produced.

Recommendations and Further Research

A few recommendations for key stakeholders

For PV system installers - This analysis can tell PV system installers what factors to consider when estimating the cost and what the cost could be.

For policy and law makers - The model developed here can help lawmakers identify opportunities for improved tax incentives or economic reforms to help PV system installers reduce costs thus increasing the rate adoption of this technology by end consumers.

Further Research

1) Consider the impact of tax incentives on system installation costs
2) Include more data for installations in segments outside of the residential sector - such as commercial and industrial.
3) Consider location based factors such as cost of living, or typical household income on the cost of installations. Since the cost of living varies across states, this can impact the price of goods, services and commodities in that state.