Analysis of Solar Power Cost over Time

Milestone Report, Capstone Project 1

Springboard, DSCT

09Apr2018

Mark Sausville

saus@ieee.org

Analysis of Solar Power Cost over Time

Capstone 1 Milestone Report

The Problem

The goal of this project is to produce a model that will allow prediction of the Fair Market Value cost of installing a solar electrical system on a typical home in the 4-8 quarters beyond the end of the data.

Having built that model, we will predict the cost/watt of a solar installation over the 18 months past the end of the data, so the customer can make an informed decision about whether to install now or defer to enjoy more favorable pricing;

The Customer

The customer is a householder considering the installation of a solar electrical system. The question to be answered is "Will it be more cost-effective to install now, or will I save money by waiting a year or two?". This is a reasonable question given the market context of rapidly falling costs for solar power installation over the past decade.

Project Structure

This project breaks down into the following steps:

1. Acquire the data.
2. Identify and investigate the relevant data in the dataset and wrangle into a format that can be used for the following steps.
3. Perform a statistical learning analysis to find a function that expresses solar installation cost as a function of time (and other relevant features) and use that function to predict the cost of installation in the near term. Compare the performance of these various models and determine how they will be used to generate predictions. Also, study the effect of geographic location in the modeling.
4. Validate the methodology by back-testing on current and prior years.

Data Acquisition

The data for this project were acquired at the site: https://openpv.nrel.gov/search. This site contains data sets from both the National Renewable Energy Laboratory (NREL) and Lawrence Berkeley National Laboratory.

Both datasets are of similar size and nature (around 1 million rows representing distinct installations, approximately 80 features). The documentation on the site implies that the NREL data is more comprehensive, so we began with that data.

After an initial cleaning of the NREL data (detailed in the report: https://github.com/leonkato/springboard/blob/master/solar-cap-1/docs/solar_data_wrangling.pdf), we discovered that there was no mechanism in this data to prevent duplicate entries. In fact, many suspected duplicate entries were found. As there was no differentiating field available for these records, this dataset was set aside as unreliable.

Fortunately, the LBNL data was available and better structured including unique identifiers and the source of the data for each record. The LBNL data also had coverage for 2016, while the NREL data ended with 2015.

The LBNL dataset also included documentation including a short description of each feature. The details of the data wrangling for this dataset are available here.

Cost Model for Solar Installation

In market cycles we have come to expect that as the number of adopters grows, the cost of adoption drops as economies of scale come into play (producer cost/unit falls) and competition forces producers and installers to operate more efficiently.

In technology cycles, these effects are often exaggerated by innovation and discovery. Moore's Law is a well-known example: for many years processor power/cost has roughly doubled every 18 months.

In solar energy, there is an analogous formulation, Swanson's law, i.e., the price of solar photovoltaic modules tends to drop 20 percent for every doubling of cumulative shipped volume. While not as dramatic a decrease as Moore's Law, drops in the cost of solar panels have been sufficient to have cut the price by 50% in the last decade.

For the customer, if the installation price is falling rapidly, waiting for the price to drop may be beneficial. On the other hand, if price is flat or increasing, buying now may be the best option. This study aims to educate that decision by predicting near term price by extrapolating from a model learned from the data.

The cost of a solar installation has many components. The costliest single component has historically been the photovoltaic modules, but non-module costs currently amount to perhaps 2-3 times the module cost.

The following components of the price could be considered in an analysis:

- Solar Panels
- Other hardware
  - Inverter
  - Racking
  - Cables, conduit, interconnection hardware, etc
- Installation labor
- Installer/Integrator costs and profit

Unfortunately, the data did not support a detailed hardware/overhead breakdown and separate modeling of each component. While features existed in the data that seemed to allow this kind of approach, on examination these fields were largely unpopulated (or completely null).

LBNL explained that while the fields were documented, they had been unable to acquire reliable data for these features and had dropped them from the public dataset.

While this was disappointing, there is a strong basis for modeling the cost of solar installations. This will be apparent from an exploration of the data.

Description of the data

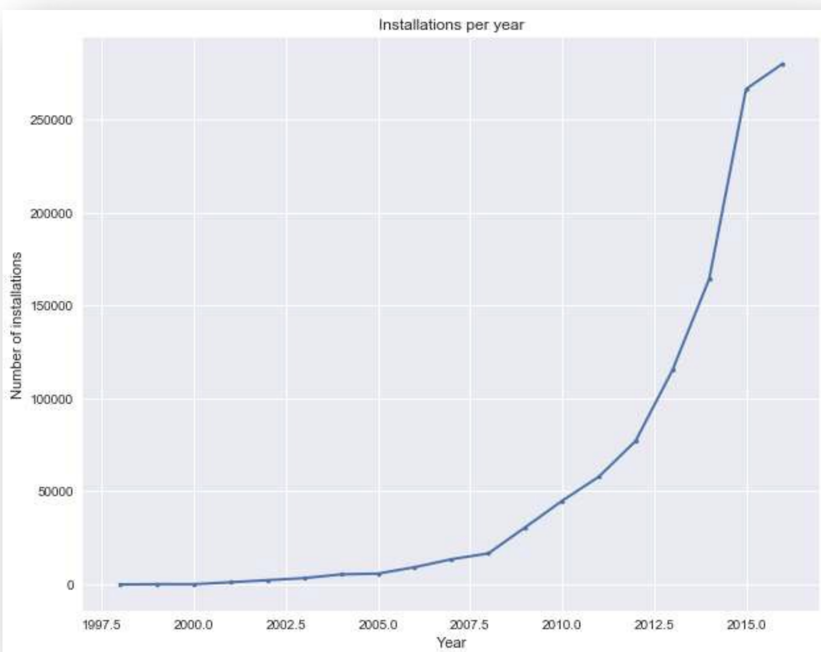
The growth of solar power

We often hear that the number of solar installations has grown dramatically. The first question is: "Does the dataset support that claim?".

While this dataset by no means describes every solar installation in the US (only 19 states are present in the data), all the largest state markets are included. The data are sourced primarily from state agencies and utilities that administer PV incentive programs, solar renewable energy credit registration systems, or interconnection processes.

The growth shown in Figure 1. is exponential through 2015, with an average annual growth rate about 50%. In 2016, the data show that growth rate may be beginning to decline. In any case, it is evident that US installations can be expected to continue in the hundreds of thousands for the near future as solar becomes more price competitive with traditional power sources.

Figure 1. Number of Solar Installations per year



What is the behavior of cost/watt over time?

The cost of a solar power installation is generally measured in dollars per watt and calculated by dividing the total size of the installation in watts by the total cost in dollars. As the media tell us, the price of solar power has decreased dramatically over the past few years.

Figure 2. Median cost/watt in residential and commercial installations

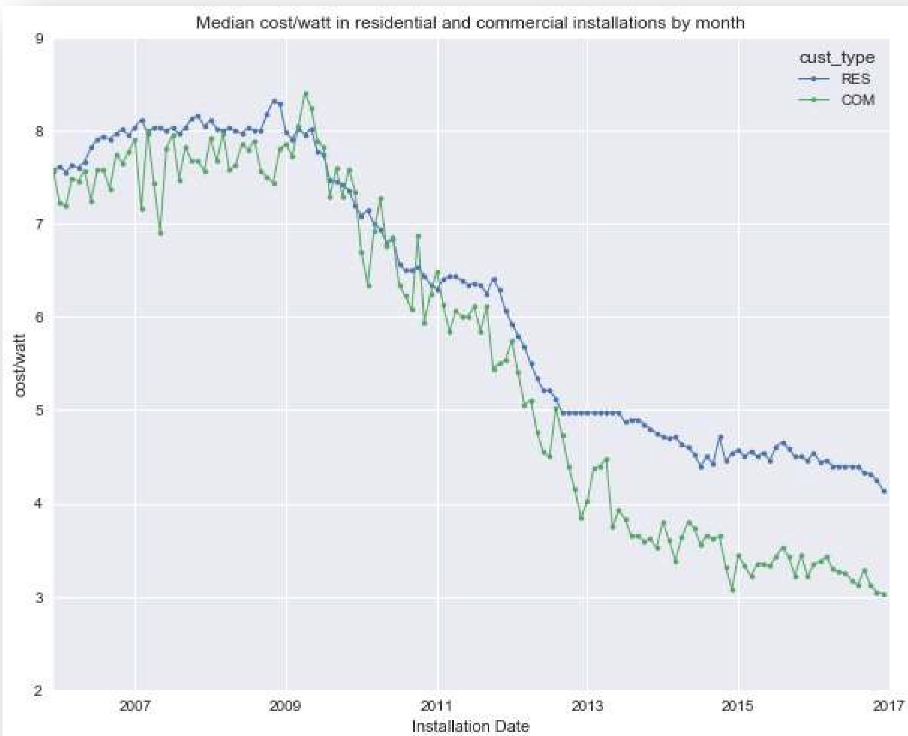


Figure 2 shows that cost/watt has fallen dramatically over the past decade. From the beginning of 2007 to the end of 2016, the median cost/watt for residential customers fell from about $8/watt to near $4/watt. Commercial customers have consistently paid less over this period showing a more pronounced reduction in price.

Does cost/watt vary by customer type

We see in the graph above that commercial customers pay less per watt of solar capacity than residential users. Does that apply to the other customer types? There are seven types of customer in the dataset. Residential installations comprise 96% of the data. Commercial installations are 2%. The other five types represent only 2% of the data. Table 1 shows the number of each customer type and percentage of the total.

Table 1. Number of customers of each type

Customer Type	RES	COM	NON-RES	GOV	NON-PROFIT	SCHOOL	TAX-EXEMPT
Count	745688	15199	7042	3996	1951	1748	70
Percentage	96%	2%	1%	1%	0%	0%	0%

In Figure 3, we display the median cost/watt by month for four installation types. One thing to keep in mind while considering the graph: the number of nonresidential installations is quite small compared to the residential installations (30k vs. 750k).

Further, those 30k nonresidential installations are split among 6 types, the largest of which is commercial (business) installations. The smaller customer types show much more variability than the larger groups. This explains the peaks seen below. Only the residential and commercial customer types have enough participants for their aggregate to be well-behaved.

Figure 3. Median monthly cost by type

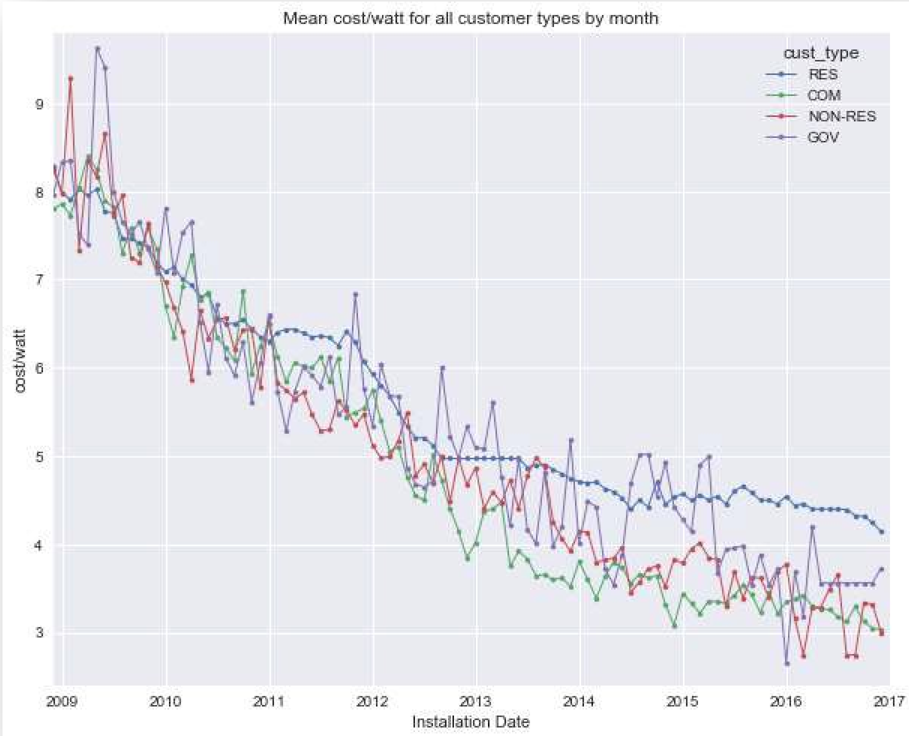


Figure 3 shows that residential customers generally pay more at any given time than non-residential customers (see also Figure 2). One possible explanation for this difference is that nonresidential customers typically buy larger installations and therefore get more favorable pricing. We examine the relationship between cost/watt and size of installation below.

Does cost/watt decrease with the size of the installation?

We see above that residential installations tend to have a higher cost/watt than all other types. Table 2. shows that residential installations median size is 5.75kw, while commercial installations are generally larger with a median of 27.72kw.

Table 2. Mean and median size and cost/watt for residential and commercial solar installations

	Com size (kw)	Com cost/watt	Res size (kw)	Res cost/watt
median	25.2	4.7	5.75	4.94
mean	131.3	5.21	6.45	5.14
count	15199	15199	745688.	745688.

In Table 2, the mean (and median) cost does not appear to differ drastically over the full dataset between commercial and residential installation. That is because the bulk of residential installations come in recent years at lower prices, pulling down the mean and median cost. While the number commercial installations grew quickly starting in 2015, the residential installations were 1000 times greater. This is clearly visible in Figure 4 where the installation counts are shown on a logarithmic scale.

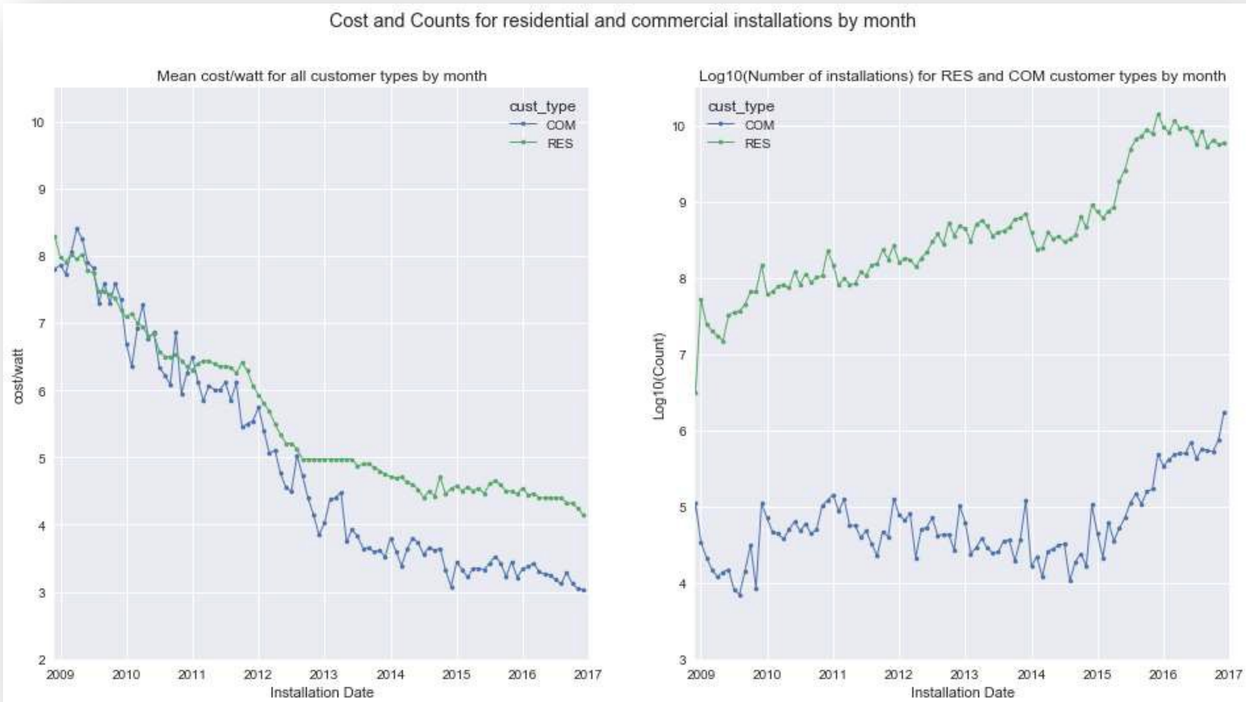


Figure 4. Monthly median cost/watt and counts for residential and commercial installations

It is reasonable to expect that commercial installations have less expensive per watt costs due to quantity discounts. We examine this assertion below.

To see if cost/watt varies with size, it is important to compare installations where only the size differs, so we restrict the set to residential installations. Residential installations comprise 96% of the data. We begin by looking at the distribution of sizes

Figure 5 Distribution of size (kw) for residential installations.

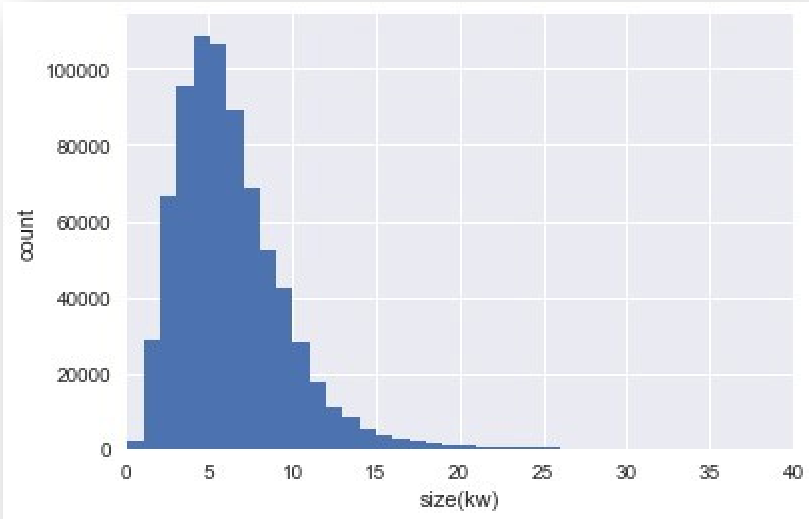


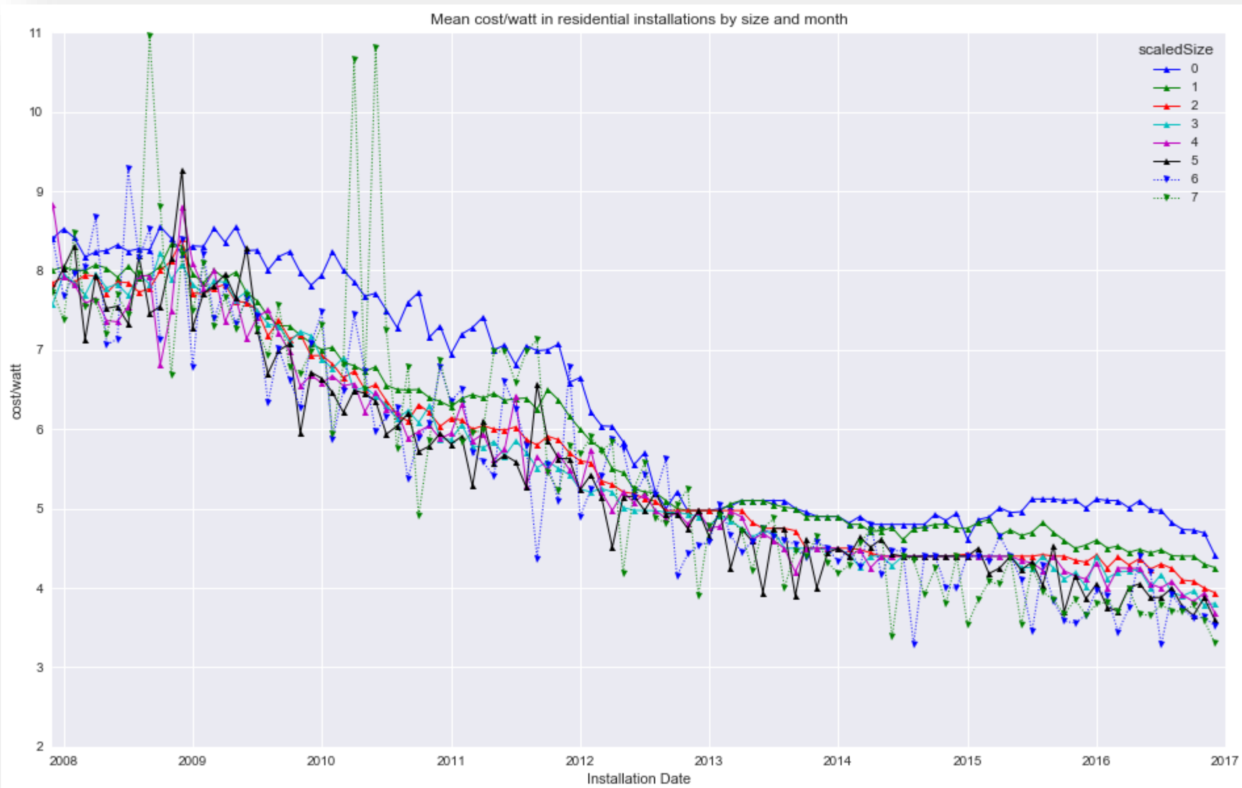
Table 1 Summary statistics for size (kw) in commercial and residential installations

	count	mean	std	min	25%	50%	75%	max
com size(kw)	15199	131.3	373.8	0.33	10	25.2	80.6	5999
res size(kw)	745688	6.45	7.1	0.11	4.03	5.75	7.95	1989

Figure 6 shows the distribution of size in kilowatts for 746k residential installations. As the summary statistics show, 50% of the residential installations are between 4 and 8 kilowatts.

In the following figures, we plot cost/watt vs. time, distinguishing different size groups.

Figure 6 Cost/watt vs. installation month by size



The three graphs, Figs. 6, 7 and 8 all show the same data, depicted in three different ways. In each, the size of the installation binned into 2.5kw-wide bins (i.e. 0-2.5 kw, 2.5-5 kw, etc). Anything larger than 25kw is grouped in the bin representing the largest installations.

The first graph (Figure 6) is perhaps the clearest in that the groups are visually separable, and it is quite apparent that within most month-long periods, the smallest installations pay the highest price, with price decreasing as size increases.

The second two graphs add visual intuition about the shape of the surface of cost/watt as a function of both time and size, showing a surface with generally negative slope as both time and size increase. The 'cliffs' in the surface plot are due to filling in missing data in the size-group space. In the earlier data, few size groups are populated as the number of installations is smaller. At later dates, as the number of installations increases, all size groups are represented, and the surface becomes better behaved.

From smallest to largest, all groups show decrease in cost over time. The decrease in price as size increases while time is held constant is not as steep. Note that in the surface plot (Figure 9), the darkest blue area (corresponding to least expensive cost/watt) is found at the latest time and the largest size group.

Figure 7 Cost/watt vs. time and size

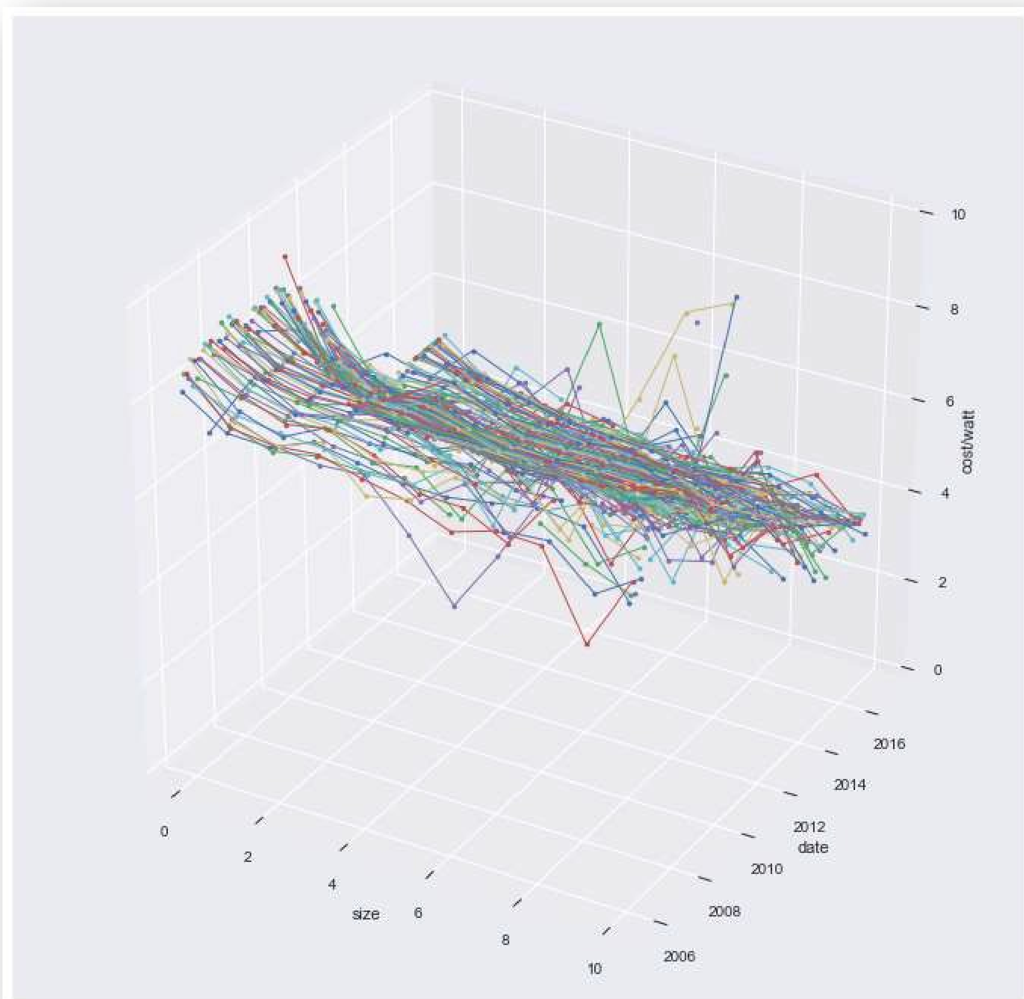
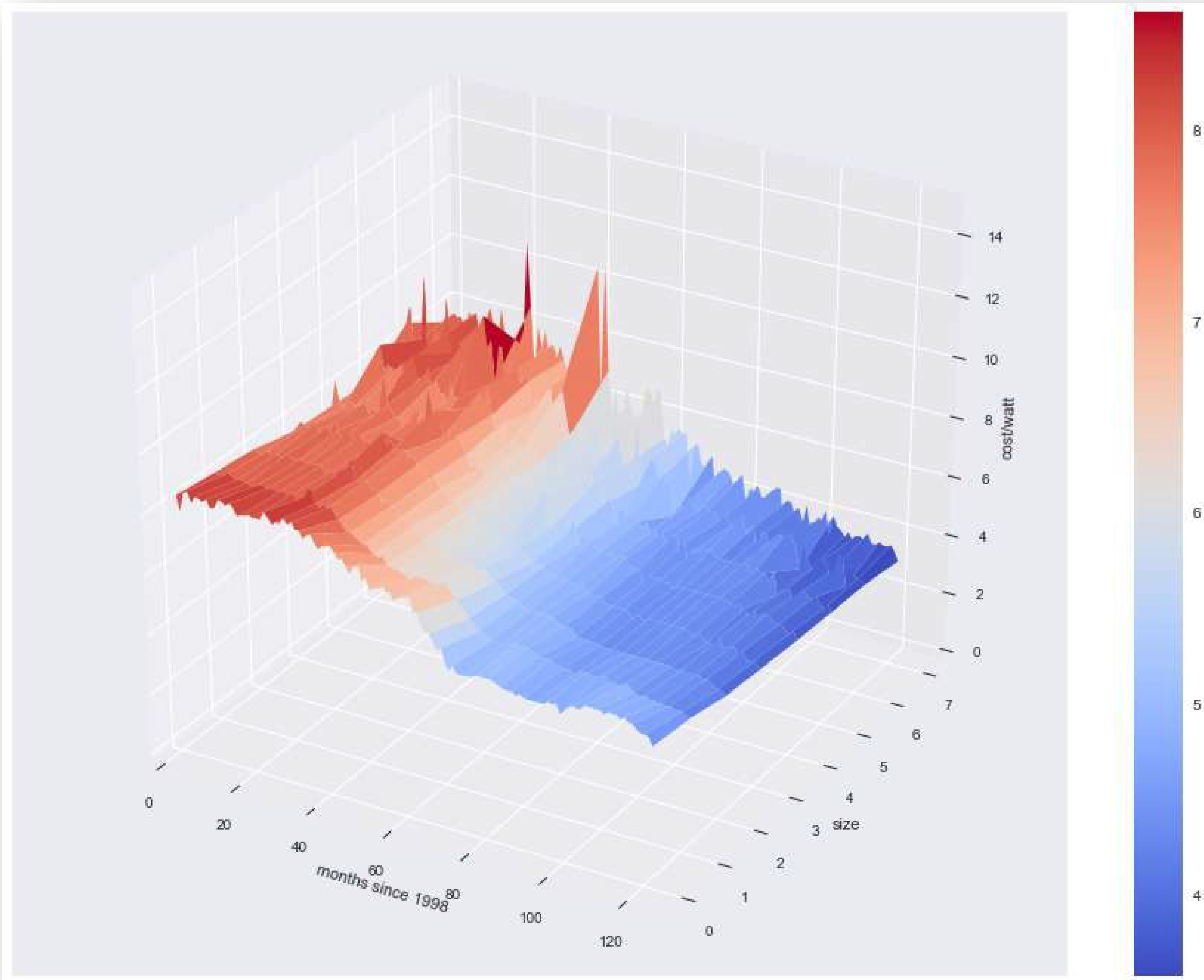


Figure 8 Cost/watt vs time and size



Does cost/watt vary by region?

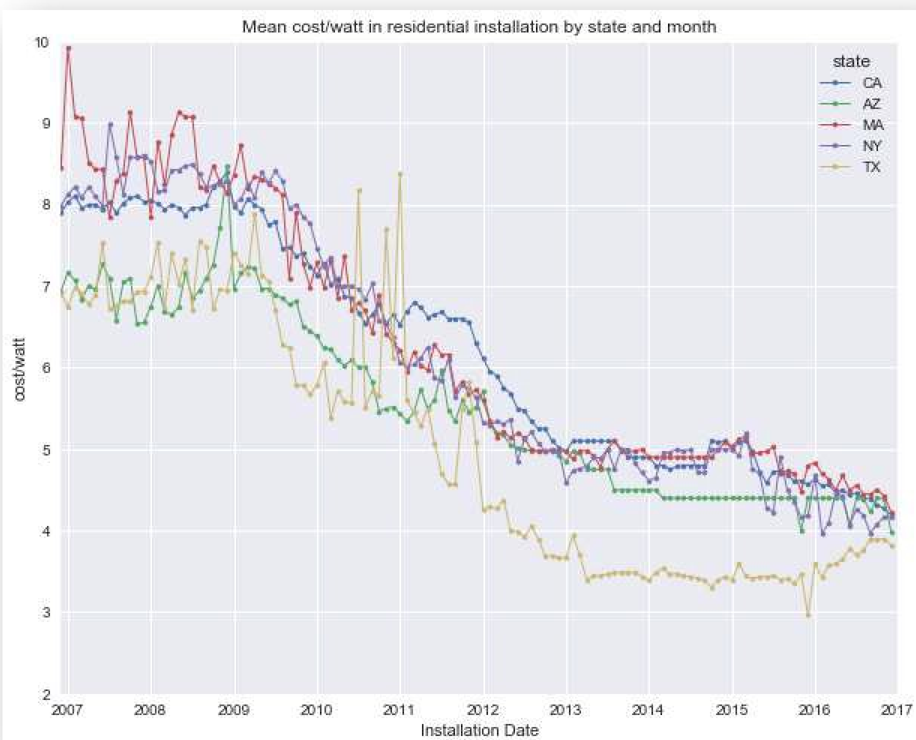
It is natural to ask if the cost of solar power varies by location. Figure 9 below says that location has a significant impact on the market price of solar installations.

The differences in market price for solar installations can perhaps be explained by 'cost of doing business'. Texas is consistently the least expensive state. California, where higher costs are offset by strong incentives is by far the largest market. For clarity not all states are show in the image.

Figure 9 has two interesting features:

1) Historically the cost of solar power has varied widely from state to state. For example, in 2014, the median cost/watt in New York was over $5/watt, while in Texas the median cost was about $3.50. In earlier years, prices diverged by even greater amounts.
2) Over time, the differential by state seems to be diminishing at the same time as costs decrease in almost every state.

Figure 9 Median cost/watt vs. time by state



What is the impact of the Boolean features on cost/watt?

The following features are Boolean variables in the dataset. There are two different kinds of features: those having to do with price characteristics and those associated with hardware in the installation.

- third_party - is the installation is owned by a third party?
- appraised_value - is the cost/watt a market price or an appraised price?
- new_const - is the installation part of the initial construction of the residence?

These features describe how the price in the dataset can be interpreted.

Third-party installations are owned by the vendor (i.e. neither the consumer or utility) and typically installed at no upfront coat to the consumer. The installation is funded by the owner of the system in conjunction with a long-term power purchase contract with the consumer.

An appraised price is not as trustworthy as a market price because it is possible to manipulate. There are possible motives to report an inflated or reduced cost appealing to various parties (installers, consumer, system owner, incentive program participants).

New construction may provide an opportunity to save on installation costs.

The features below also take the form of Boolean variables but represent physical aspects of the installation.

- ground mounted
- battery
- tracking
- uinverter (microinverter)
- DC optimizer

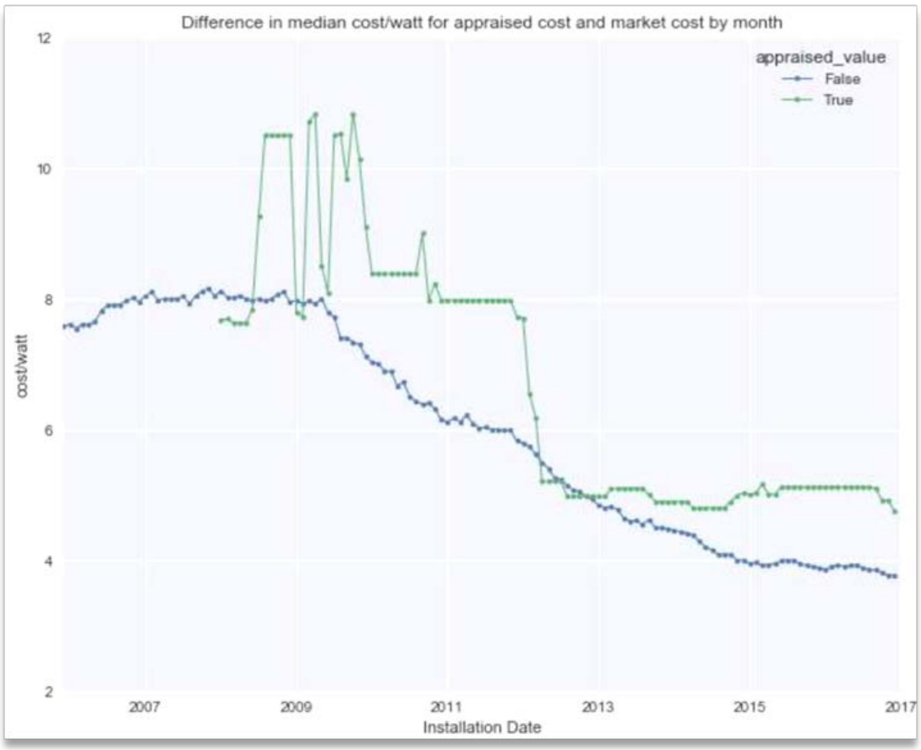
Ground mounted systems may be better located with respect to the amount of sun and may offer a simpler (and less expensive) installation. Battery equipped systems store power for use at night but are costlier. Tracking systems incorporate machinery to aim the solar array for higher power production but are costlier.

Microinverter systems are simpler to install and configure but cost more in aggregate than a single large inverter. DC optimizers provide advantages in flexibility of system design and management at additional hardware cost.

Appraised Prices

It is clear from the data that appraised prices tend to exceed market prices for solar installations. One implication for an analysis is that since there are so many (approximately 230k) it may be worthwhile to exclude appraised prices from a model training set.

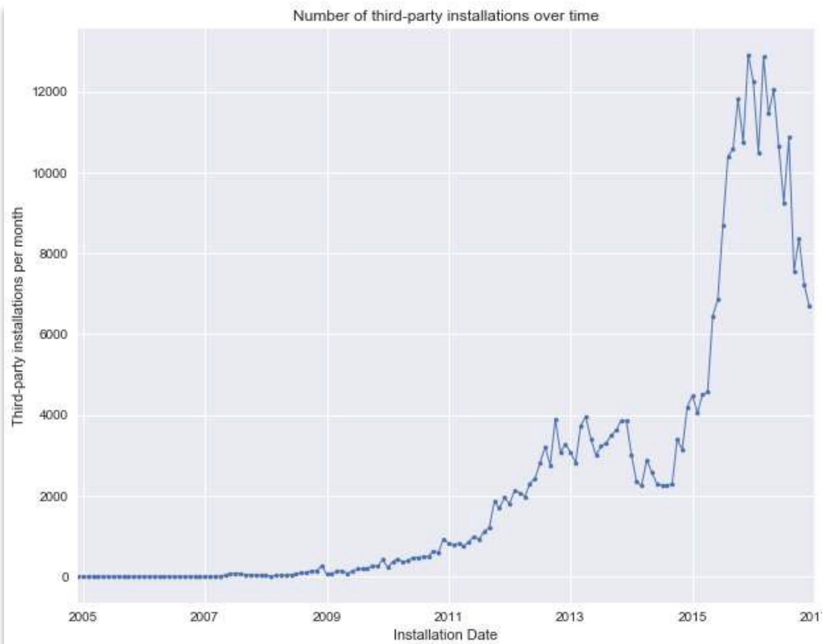
Figure 10 Compare appriased and market prices



Third-Party Ownership

It's interesting to see that in the 2006-2007 time frame (see Figure 11), third-party ownership becomes popular as a result of tax incentives and the wide availability of credit to third-party owners.

Figure 11 Number of third-party owned systems installed vs. time



There seems to be a consistent trend for third-party pricing to exceed direct ownership pricing though the difference tends to be less in recent years (Figure 12).

There are many reasons for the differential (e.g. price may be inflated since third-party owner often controls the supply chain and installation process). We will not examine this in detail in this project.

It should be noted that third-party owned systems comprised a very substantial subset. For accurate analysis of market price, it may be necessary to exclude these from cost/watt modeling.

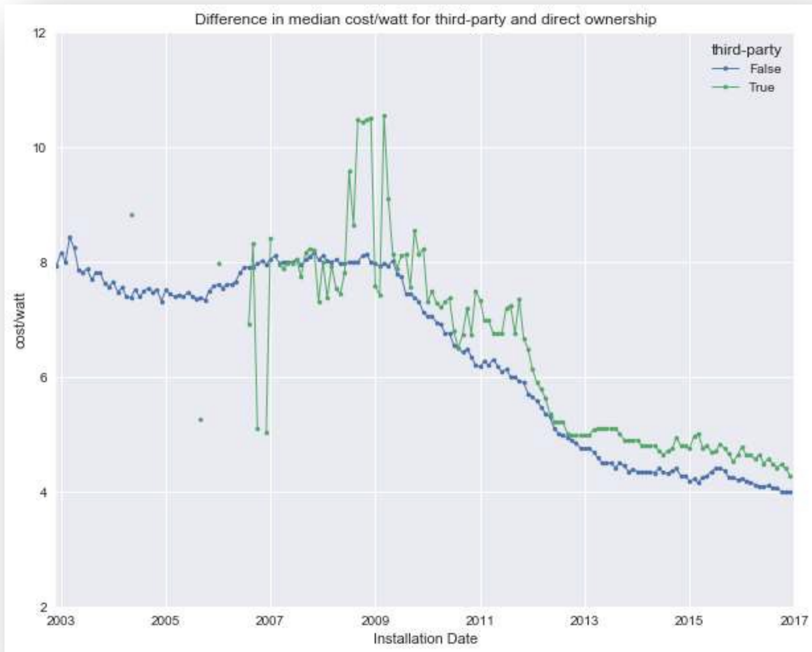


Figure 12 Compare third-party/market prices

Other Boolean features

The other Boolean features in the data were examined and found to be not relevant to the thrust of the project because a) they are not well represented (e.g. battery systems) or b) have little impact on price (e.g. DC Optimizer). For those interested, the examination of these features (as well as the rest of the EDA code) is presented in https://github.com/leonkato/springboard/blob/master/solar-cap-1/story/tts_story_01.ipynb.