AN ANALYSIS OF POTENTIAL RELATIONSHIPS BETWEEN SOLAR MARKETPLACE SHOPPING AND SOLAR EQUIPMENT QUALITY

by
Emily Walker

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Environmental Sciences and Policy

Baltimore, Maryland
December 2022

Executive Summary

This research examines the impact of comparison shopping through the United States' leading solar quote marketplace, EnergySage, on the quality of installed solar equipment. As a Senior Writer at EnergySage, the author of this analysis is responsible for leading research and content strategy pertaining to solar and solar adjacent products. While previous research has demonstrated that EnergySage shoppers receive lower priced quotes than those who do not use EnergySage, the author wanted to better understand if installations completed through EnergySage's quote platform also include higher quality equipment. This research ties together the author's passion for and professional knowledge of clean energy with the knowledge and skills she has gained in completing her Master of Science in Environmental Sciences and Policy through Johns Hopkins University. Specifically, it highlights her policy experience, earned through completing courses such as *Environmental Policymaking and Policy Analysis* and *Introduction to Energy Law and Policy*. It also utilizes research and statistical analysis skills she developed through coursework in *Understanding Public Attitudes for the Communication of Climate Energy and Policy*. In addition to her Master of Science, she is pursuing a Certificate in Geographic Information Systems (GIS) at Johns Hopkins University. Her capstone employs technical skills she has gained through her GIS coursework. For example, through the course *Spatial Databases and Data Interoperability*, she learned database management skills, including how to clean and manipulate large datasets using FME workbench. Her other GIS courses have helped her enhance her mapmaking skills using ArcPro. Overall, the author's capstone synthesizes much of what she has learned during her time at Johns Hopkins, while showcasing timely and relevant data from her own company.

Abstract

The current Biden administration has aggressive goals to dramatically increase solar deployment across the United States. While solar is highly sustainable compared to fossil fuel electricity sources, solar modules (colloquially known as solar panels) can contain hazardous waste and solar module recycling is still in its infancy. High quality solar equipment results in less waste overall, making it pivotal to the future of solar. This study uses solar installation data from the Lawrence Berkeley National Laboratory to examine the impact of comparison shopping completed through EnergySage, the United States' leading solar quote marketplace, on the quality of installed solar equipment. It finds that EnergySage installations include modules of higher efficiency, modules of greater wattage capacity, and more advanced module and inverter technology, compared to installations completed external to EnergySage. It also finds that over the past three years, despite supply chain constraints severely impacting the solar industry, modules have significantly increased in efficiency and capacity, at roughly the same rate for installations completed through and external to EnergySage. However, over these same three years, EnergySage installations consistently contained higher quality equipment, suggesting that competition may drive installers to quote and solar consumers to request and choose higher quality solar equipment.

Primary Reader and Advisor: Dr. Jennifer da Rosa

Secondary Reader: Dr. Michael Schwebel

Acknowledgments

First, I would like to thank the team (and especially Galen Barbose) at the Lawrence Berkeley National Laboratory (LBNL) for creating the *Tracking the Sun* dataset and helping me parse which installations were completed through EnergySage's quote comparison platform. I would also like to thank my colleague Spencer Fields for helping ignite the idea for this research through his work updating EnergySage's equipment rating system. Additionally, my fiancé, Rocco Donohue, has been pivotal to this project by being the first reader of my paper and my biggest support system throughout the research process. Finally, I would like to thank both Dr. Jennifer da Rosa for being my advisor and primary reader and Dr. Michael Schwebel for being my capstone instructor and secondary reader.

Table of Contents

List of Tables

List of Figures

I. Introduction

According to a 2021 study conducted by the Department of Energy (DOE), solar has the potential to power 40% of electricity in the United States by 2035 and 45% by 2050 (Solar Energy Technologies Office, 2021). However, solar currently represents just 2.8% of electricity generation (EIA, 2022). For the United States to achieve this level of solar deployment, it needs to double current growth from 15 gigawatts (GW) annually to 30 GW until 2025, and again to 60 GW annually between 2025 and 2030 (Solar Energy Technologies Office, 2021). On August 16, 2022, President Biden signed into law the Inflation Reduction Act, which represents the United States' largest investment in climate change to date, increasing and extending crucial incentives to promote solar adoption. The bill also aims to enhance domestic manufacturing of solar modules and supports new technological advancements in the solar industry. Ultimately, the Inflation Reduction Act will help the United States rapidly increase its solar deployment and become closer to its goal of achieving a carbon-free electric grid.

In order for solar capacity to quickly scale in an affordable and efficient manor, technology improvements are required (Solar Energy Technologies Office, 2021). Improvements in solar module (colloquially known as solar panel) efficiency, output capacity, and lifespan will all help decrease the cost of solar and make it more sustainable (Solar Energy Technologies Office, 2021). The high costs associated with recycling solar modules currently far outweigh the potential revenue derived from extracted materials (Tao et al., 2020). However, The International Renewable Energy Agency (IRENA) predicts that the waste of modules could reach 78 million tons cumulatively by 2050, or 6 million tons annually across the world (IRENA, 2016).

Solar modules with higher capacity (measured in Watts, W) and greater efficiency produce more power, helping to reduce the number of modules needed and thus the amount of waste created; ultimately, better performing equipment makes solar an even more sustainable solution to the clean energy transition. This project will assess the quality of solar equipment installed between 2019 and 2021 using publicly available installation data from the annual *Tracking the Sun* report produced by the Department of Energy (DOE)'s Lawrence Berkeley National Laboratory (LBNL).

Some residential installations in LBNL's dataset were completed through a quote comparison platform called the EnergySage Marketplace, the leading solar marketplace in the United States. Founded in 2012, EnergySage enables consumers to compare quotes from its network of solar installers at no cost. Installers are able to compete for consumers' business by paying EnergySage a small fee to participate. However, EnergySage also creates regulation in its marketplace by requiring that solar installers meet certain criteria to participate, including:

- "At least 3 years of experience installing solar
- Licensed and insured for solar installations
- NABCEP certified
- Reputation for excellent customer service and quality solar installations
- Installs high quality solar equipment" (EnergySage, 2021, para. 3).

In this list, "NABCEP certified" refers to installers who have earned certification from the North American Board of Certified Energy Practitioners, a nonprofit organization that ensures installers meet certain skill levels.

The purpose of this research is to evaluate relationships between solar marketplace shopping, using the EnergySage Marketplace as a proxy, and the quality of residential solar equipment, based on module efficiency, module capacity, module cell type, and inverter type. Three alternative hypotheses are proposed:

1. *The mean efficiency and nameplate capacity of solar modules are higher in residential installations completed through EnergySage compared to residential installations completed external to EnergySage.*
2. *Over the past three years, the mean efficiency and mean nameplate capacity of solar modules have increased at a faster rate for residential installations completed through EnergySage compared to residential installations completed external to EnergySage.*
3. *Residential installations completed through EnergySage are more likely to include monocrystalline solar cells in modules and inverters with module-level power electronics, compared to residential installations completed external to EnergySage.*

II. Literature Review

2.1 Marketplaces

Comparison shopping is a type of online shopping in which consumers compare a type of product based on relevant information such as price and quality (Wan, 2009). Instead of searching through multiple websites, consumers can use comparison shopping platforms, also called marketplaces, which aggregate product information and make the comparison process simpler and faster (Wan, 2009).

Since the rise of the Internet, marketplaces have grown in popularity due to their convenience (Ong, 2011). Popular marketplaces span many different product categories, from travel sites like Expedia or Kayak to loan comparison platforms like Bankrate (Wan, 2009). In *Nudge: The Final Edition*, Sunstein and Thaler (2021) discuss the importance of comparison shopping in creating more product standardization and increasing the quality and affordability of products for consumers. However, despite greater use of marketplaces, research in the field of comparison shopping is still relatively limited.

Every six months, EnergySage releases its *Marketplace Intel Report*, which covers pricing and equipment trends in the United States solar and storage industries based on quote data from its marketplace. The cost of solar, reported on a dollar per watt ($/W) basis, has historically been lower on EnergySage compared to the national average (Barbose et al., 2022). However, according to EnergySage's *Marketplace Intel Report 15*, in the second half of 2021, 60% of consumers on the platform did not select the lowest priced quote that they received, suggesting the importance of equipment quality to solar comparison shoppers (EnergySage, 2022).

2.2 Solar Equipment Quality

According to data from the most recent EnergySage *Intel Report 15*, higher capacity solar modules are more frequently included in quotes from solar installers (EnergySage, 2022). Similarly, the most recent LBNL *Tracking the Sun* report found that the median efficiency of modules in residential solar systems has increased from 13.4% in 2002 to 20.1% in 2021, which reflects an increasing market share of higher efficiency monocrystalline solar modules as well as other recent technological advancements in solar (Barbose et al., 2022). However, installed solar equipment is still quite diverse; residential systems installed in 2021 ranged in efficiency from 16% to over 22%, with the vast majority falling between 19% and 21% (Barbose et al., 2022). Modules with efficiencies below that range were primarily made of polysilicon while modules above that range were mostly premium-efficiency modules only offered by select manufacturers (Barbose et al., 2022).

In addition to solar modules, it is also vital to examine the quality of solar inverters, which are necessary to convert the direct current (DC) electricity generated by solar modules to usable alternating current (AC) electricity. According to LBNL, almost all (94%) residential systems now include module-level power electronics (MLPEs), which enhance the performance and power output of solar energy systems by providing module-level optimization, either in the form of central inverters with DC optimizers, which centrally convert electricity but optimize it by module, or microinverters, which convert and optimize electricity at the module level (Barbose et al., 2022).

2.3 Supply Chain Constraints

The timing of this research is significant because it covers a period of time in which the solar industry has been heavily impacted by supply chain constraints. China controls the majority of the solar manufacturing components, possessing 72% of global polysilicon manufacturing capacity, 98% of ingots, 97% of wafers, 81% of cells, and 77% of modules (Solar Energy Technologies Office, 2022). The United States relies on Chinese subsidiaries operating in Vietnam, Malaysia, and Thailand, which produce 75% of silicon solar cells that are incorporated into solar modules installed in the United States (Solar Energy Technologies Office, 2022).

In 2018, the United States imposed tariffs on imported solar cells and modules, which disrupted the supply chain (Solar Energy Technologies Office, 2022). United States policies aimed at addressing forced labor concerns in Xinjiang, China have also impacted the solar supply chain (Solar Energy Technologies Office, 2022). Finally, COVID-19 pandemic-related supply chain constraints, including shortages of shipping containers that ship cargo including solar equipment, limited availability of workers, and rising gas prices, have all impacted the solar industry over the past few years (Solar Energy Technologies Office, 2022). Especially as the United States shifts towards more domestic manufacturing of solar equipment, it is important to understand how recent supply chain disruptions have impacted equipment quality in the industry.

2.4 Importance of Research

This research contributes to the nascent field of research surrounding comparison shopping; specifically, it covers the impact of marketplace shopping on the quality of solar equipment. It also examines solar equipment trends across a crucial period of time during which the industry grew and advanced while also being limited by supply chain constraints.

III. Methods

3.1 Data Collection

Solar installation data were obtained from LBNL's publicly available *Tracking the Sun* dataset published in September 2022. The full dataset includes approximately 2.5 million solar energy systems from 2000 through 2021, representing 77% of the United States market (Barbose et al., 2022). The data from 2021 include about 340,000 systems, encompassing 68% of the United States market (Barbose et al., 2022). However, only a subset of these solar installations was included in the analysis, as detailed in Section 3.3.

As part of its *Tracking the Sun* report, LBNL compares its dataset to matched EnergySage quotes. According to LBNL's report, "[f]or a subset of EnergySage price quotes culminating in an installed system, we can identify the corresponding record from the Tracking the Sun dataset"

(Barbose et al., 2022, p. 32). However, these installations completed through EnergySage are not indicated in LBNL's public dataset. Thus, the team at LBNL was contacted for the list of matched EnergySage records, which they provided. Due to privacy restrictions, the total sample size of installations completed through EnergySage is not disclosed; however, the dataset includes thousands of records and is considered robust.

3.2 Dependent Variable Selection

Due to the large volume of data and limited time, dependent variables to assess equipment quality were largely determined based on their presence in LBNL's dataset; no outside resources were used to discern other variables that could serve as proxy for equipment quality (i.e., warranty length, temperature coefficient, etc.). Some systems included multiple module and inverter models; only the primary equipment (labeled by "_1" in the dataset) was included in the analysis. Variables were selected that directly impact the performance and power output of solar energy systems, as shown in Table 1.

Table 1. Summary of dependent variables used as a proxy for solar equipment quality.

Variable	Description	Variable type
Module capacity (Watts)	The capacity of each solar module represents its "theoretical power production under ideal sunlight and temperature conditions" (Aggarwal, 2022, para. 3).	Continuous
Module efficiency (%)	The efficiency of a solar module measures its "ability to convert sunlight into usable electricity" (Marsh, 2022b, para. 3).	Continuous
Module cell type (monocrystalline or polycrystalline)	A monocrystalline solar cell is composed of a single crystal, whereas a polycrystalline cell includes multiple fragments of silicon; with monocrystalline solar cells, electrons have more room to move across the single crystals and thus generate electricity more efficiently (Marsh, 2022a).	Nominal
Inverter type (microinverter, optimized string	Both microinverters and optimized string inverters include MLPEs that optimize the power output of each solar module independently; string inverters	Nominal

inverter, or string inverter)	without power optimizers can only optimize power at the string level, so potential shading impacts the performance of the entire string of modules (Thoubboron, 2022).	

3.3 Data Cleaning

Data cleaning was primarily performed using Safe Software's Feature Manipulation Engine (FME) workbench, as summarized in Table 2. The *Tracking the Sun* dataset and the dataset of matched EnergySage records were both loaded into FME workbench. Then, using the AttributeManager transformer, the attributes in both datasets were pared to only include those needed for analysis (Appendix I).

The *Tracking the Sun* dataset included one unique attribute ("data provider") that could be matched to a unique attribute ("data record ID") in the EnergySage records dataset to determine which installations in the *Tracking the Sun* dataset were completed through EnergySage. However, there were some duplicate records based on the unique keys in both datasets, which were removed using the DuplicateFilter transformer.

Next, using the FeatureJoiner transformer, a left join was performed between the two datasets based on the unique attributes. Using the SubstringExtractor transformer, a new attribute was created to just show the years the system was installed based on the last four characters of the "installation date" attribute. Records were then pared to only include installations that included a model for the "module model" attribute, were residential, and were installed between 2019 and 2021. The final dataset was output as a CSV file. The full workflow in FME workbench is included as Appendix IIa.

Table 2. Summary of data cleaning performed using FME workbench.

Step	FME transformer used	Description
Process the *Tracking the Sun* dataset	AttributeManager	Limit attributes to what is needed for analysis (see Appendix I)
	DuplicateFilter	Remove records that do not contain unique IDs based on the data provider and data record ID attributes
Process the matching EnergySage records dataset	AttributeManager	Limit attributes to what is needed for analysis (see Appendix I)
	DuplicateFilter	Remove records that do not contain unique IDs based on the data provider and data record ID attributes
Data joining and processing	FeatureJoiner	Join both datasets together based on the data provider and data record ID attributes
	SubstringExtractor	Create a new attribute for the year the solar energy system was installed using a string of the installation date attribute
	Tester	Limit records to only include those that contain the module model, are residential installations, and were installed between 2019 and 2021
Data output	Writer	Create a new CSV file based on the transformations performed in FME workbench

Data processing of the CSV file was then performed using Microsoft Excel. First, a "-1" value was assigned to missing records in the "quote_ID" attribute to represent installations that were not completed through EnergySage. Then, a new attribute was created called "ES"; if "quote_ID" was equal to "-1", a value of "0" was assigned (a non-EnergySage installation) and if "quote_ID" was not equal to "-1", it was assigned a value of "1" (an EnergySage installation). Next, an attribute was created called "total_installs"; a value of "1" was added for each record to indicate that it represents one installation.

Finally, a separate CSV file was created and processed in Microsoft Excel for each dependent variable being assessed, as shown in Table 3.

Table 3. Summary of separate CSV file processing for each dependent variable.

Variable/ CSV file	Data processing in Microsoft Excel
"Module_capacity"	Data were pared to only include records in which the nameplate capacity is at least 100; any installations with capacities below 100 would likely represent outliers.
"Module_efficiency"	Data were pared to only include records in which efficiency data exist.
"Module_tech"	Data were pared to only include records in which the module cell technology is polycrystalline or monocrystalline.
"Inverter_type"	Data were pared to only include records in which inverter data exist; a new attribute ("inverter_type") was created based on the existing "microinverter" and "power optimizer" attributes to assign a value if the inverter type is microinverter ("Micro"), optimized string inverter ("Optimized") or non-optimized string inverter ("String").

3.4 Assessing Data Location

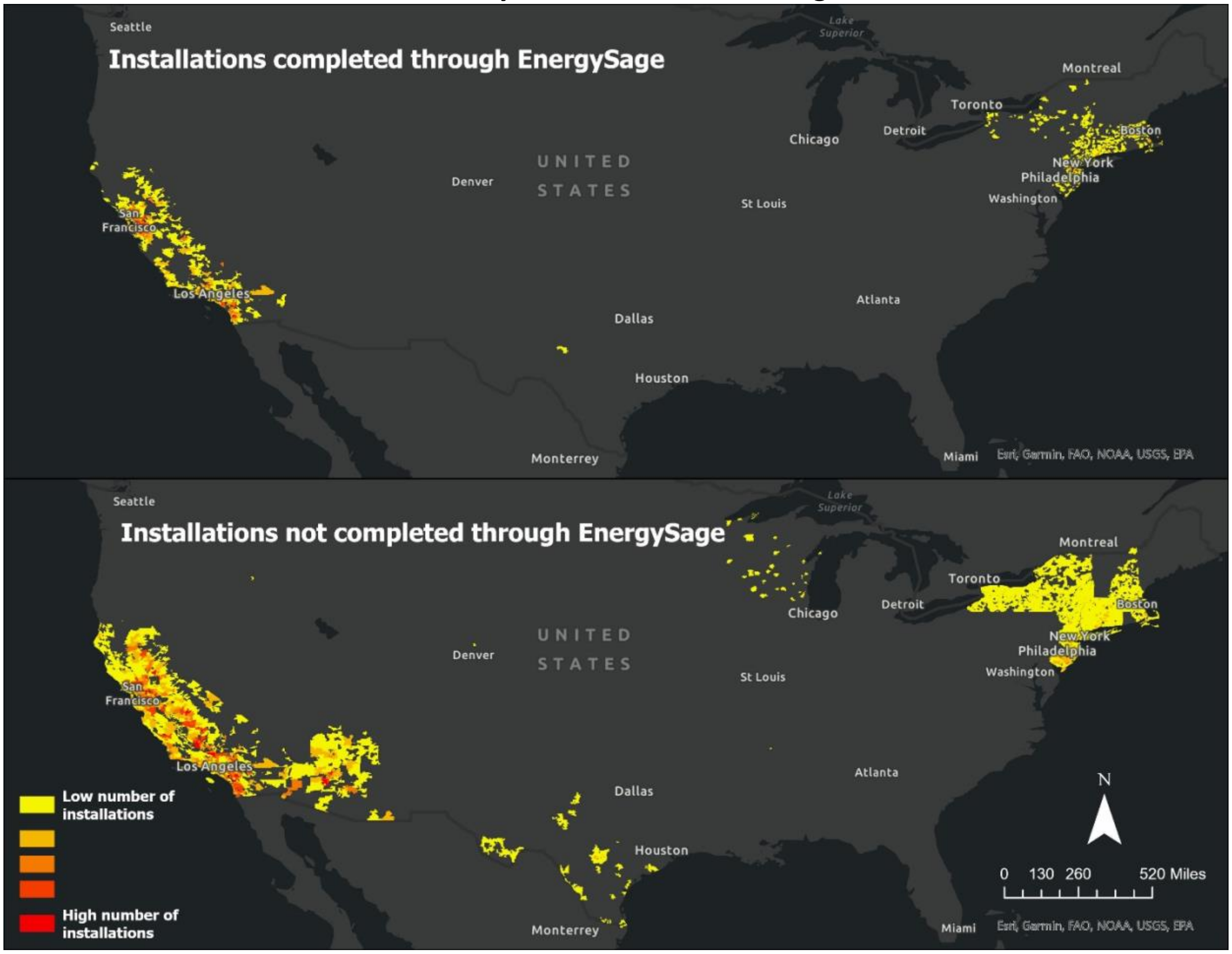
To understand the geographic distribution of installations, FME workbench was again used to create shapefiles that could be loaded into Esri's ArcPro, as summarized in Table 4. The newly created CSV file containing all records and a shapefile from the U.S. Census Bureau containing zip code areas called "TIGER/Line Shapefile" were both loaded into FME workbench. Unnecessary attributes were removed from both datasets using the AttributeManager transformer. The Tester transformer was used to separate EnergySage and non-EnergySage installations contained in the CSV file into two separate datasets. Both datasets were aggregated using the Aggregator transformer; the total number of installations per zip code was calculated by summing the "total_installs" attribute (as described in section 3.3) based on the "zip_code" attribute. Using the FeatureJoiner transformer, both datasets were joined to the U.S. Census Bureau shapefile based on the "zip_code" attribute. Finally, the EnergySage and non-EnergySage datasets were exported as separate shapefiles. The full workflow is included as Appendix IIb.

Table 4. Summary of location data processing performed using FME workbench.

Step	FME transformer used	Description
Process the newly created CSV file	AttributeManager	Remove unnecessary attributes to only include: zip_code, city, state, ES, total_installs
	Tester	Separate EnergySage and non-EnergySage installations based on the "ES" attribute
Process the U.S. Census Bureau shapefile	AttributeManager	Remove unnecessary attributes to only include: zip_code, lat, long
Create shapefiles of summarized installations	Aggregator	Summarize the data by calculating the total number of installations by zip code for EnergySage and non-EnergySage installations
	FeatureJoiner	Join the U.S. Census Bureau shapefile with both the EnergySage and non-EnergySage installation datasets based on zip code
	Writer	Create two shapefiles: one for EnergySage installations and one for non-EnergySage installations

The two shapefiles were then loaded into ArcPro. Each shapefile layer was symbolized by Graduated Colors using Natural Breaks (Jenks) to show zip code areas of high and low number of installations relative to that layer's own dataset; thus "high number of installations" and "low number of installations" for each map shown in Figure 1 do not necessarily correlate to the same numbers. It was observed that the EnergySage and non-EnergySage datasets were geographically similar in that most installations occurred on the West Coast and in the Northeast; however, installations not completed through EnergySage had wider geographic coverage, especially near New York, Wisconsin, Texas, and Arizona.

Figure 1. Density of EnergySage and non-EnergySage installations by zip code. Natural breaks for each dataset determine the density breaks shown in the legend.



3.5 Test Selection & Data Analysis

To test the three alternative hypotheses, all four CSV files were loaded into IBM's SPSS statistical software platform. Distributions for continuous variables were considered normal (Appendix III) and each sample had thousands of records. Three statistical tests were selected based on the three alternative research hypotheses, including t-test, ANCOVA, and chi-square test, as summarized in Table 5.

Table 5. Description of statistical tests used to test research hypotheses.

Alternative research hypothesis	**Test performed**	**Test description**
H1	T-Test	"A t-test is an inferential statistic used to determine if there is a significant difference between the means of two groups and how they are related" (Hayes, 2022b, para. 1).
H2	ANCOVA	"In basic terms, the ANCOVA examines the influence of an independent variable on a dependent variable while removing the effect of the covariate factor" (Statistics Solutions, 2021, para. 2)
H3	Chi-Square Test	"A chi-square (χ^2) statistic is a test that measures how a model compares to actual observed data" (Hayes, 2022a, para. 1)

Each test was performed twice based on null sub-hypotheses; thus, each alternative research hypothesis included two null sub-hypotheses, as shown in Table 6.

Table 6. Hypotheses tested using SPSS.

Alternative research hypothesis	**Null sub-hypothesis**	**Dependent variable**	**Independent variable(s)**	**Test performed**
H1	There is no difference in mean module capacity between installations completed through EnergySage and installations not completed through EnergySage.	Module capacity (Watts)	Non-EnergySage vs. EnergySage	T-Test
	There is no difference in mean module efficiency between installations completed through EnergySage and installations not completed through EnergySage.	Module efficiency (%)		
H2	There is no relationship between the year and the mean module capacity, controlling for if the installation was completed through EnergySage or not.	Module capacity (Watts)	Year, Non-EnergySage vs. EnergySage	ANCOVA
	There is no relationship	Module		

	between the year and the mean module efficiency, controlling for if the installation was completed through EnergySage or not.	efficiency (%)		
H3	There is not a relationship between if installations were completed through EnergySage or not and the type of solar module cell.	Module cell type (monocrystalline or polycrystalline)	Non-EnergySage vs. EnergySage	Chi-Square Test
	There is not a relationship between if installations were completed through EnergySage or not and the type of inverter.	Module inverter type (microinverter, optimized string inverter, or string inverter)		

IV. Results

4.1 Module Performance (H1)

Both null hypotheses relating to module performance were rejected, meaning the alternative research hypothesis one (H1) was accepted. Installations completed through EnergySage and installations that were completed external to EnergySage had a statistically significant difference in mean module capacity, $t(585878) = -42.106$, $p = <0.05$; $d = 0.513$ (Appendix IVa). The effect size for this analysis ($d = 0.513$) was moderate ($d = 0.51\text{-}1.00$). Installations completed through EnergySage tended to have a higher module capacity, with a mean of 348 Watts, compared to those completed external to EnergySage, with a mean of 333 Watts (Table 7).

Installations completed through EnergySage and installations completed external to EnergySage also had a statistically significant difference in mean module efficiency, $t(579300) = -54.610$, $p = <0.05$; $d = 0.666$ (Appendix IVb). The effect size for this analysis ($d = 0.666$) was moderate ($d = 0.51\text{-}1.00$). Installations completed through EnergySage tended to have a higher module efficiency, with a mean of 20.7%, compared to those completed external to EnergySage, with a mean of 19.8% (Table 7).

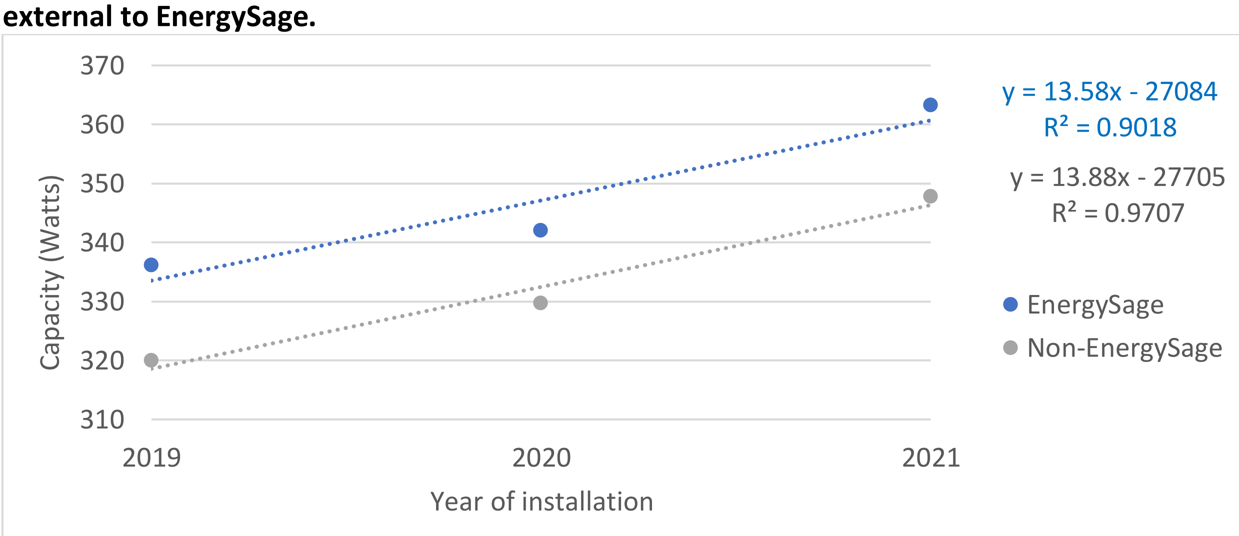
Table 7. Summary of module performance (H1) results.

Dependent variable	EnergySage mean	Non-EnergySage mean	P-value
Module capacity	348 Watts	333 Watts	<0.05
Module efficiency	20.7%	19.8%	<0.05

4.2 Module Performance Over Time (H2)

Both null hypotheses relating to module performance over time were not rejected, meaning the alternative research hypothesis two (H2) was rejected. There was not a statistically significant effect of year on the mean module capacity after controlling for if the installation was completed through EnergySage or not, F (1, 4) = 0.003, p = 0.958 (Appendix IVc). However, there was a statistically significant effect of year on the mean module capacity without controlling for if the installation was completed through EnergySage or not, F (1, 5) = 29.118, p= <0.05; overall, the capacity of modules in installations completed both through and external to EnergySage increased about 14 W annually over the past three years (Figure 2, Table 8).

Figure 2. Module capacity from 2019-2021 for installations completed through EnergySage vs. external to EnergySage.



There also was not a statistically significant effect of year on the mean module efficiency after controlling for if the installation was completed through EnergySage or not, *F* (1, 4) = 0.000, *p* = 1.000 (Appendix IVd). However, there was a statistically significant effect of year on the mean module efficiency without controlling for if the installation was completed through EnergySage or not, *F* (1, 5) = 6.400E-5 *p*= <0.05; overall, the efficiency of modules in installations completed both through and external to EnergySage increased 0.4 percentage points annually over the past three years (Figure 3, Table 8).

Figure 3. Module efficiency from 2019-2021 for installations completed through EnergySage vs. external to EnergySage.

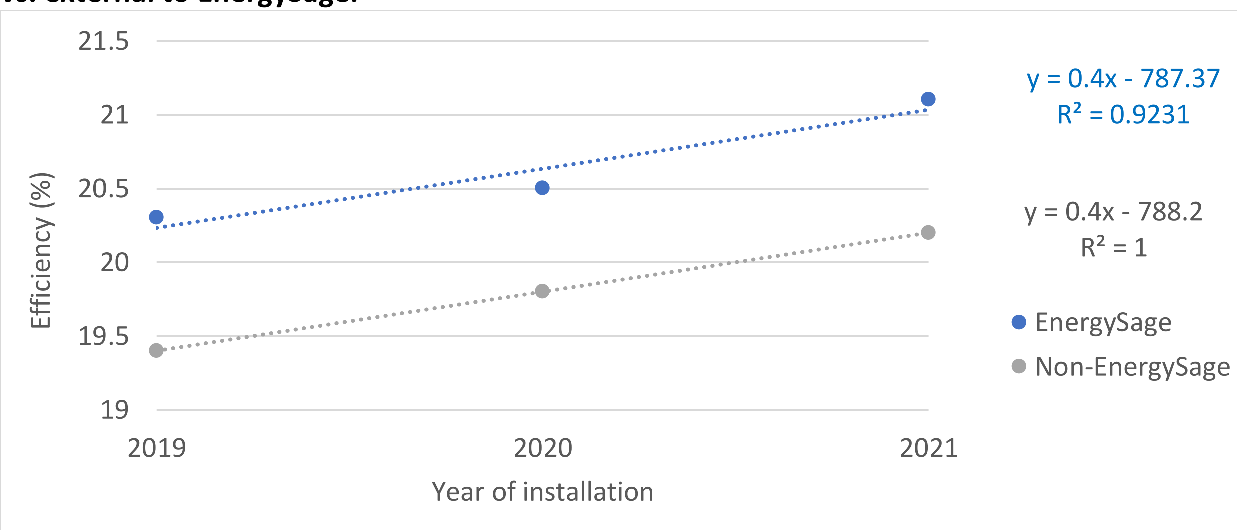


Table 8. Summary of module performance over time (H2) results.

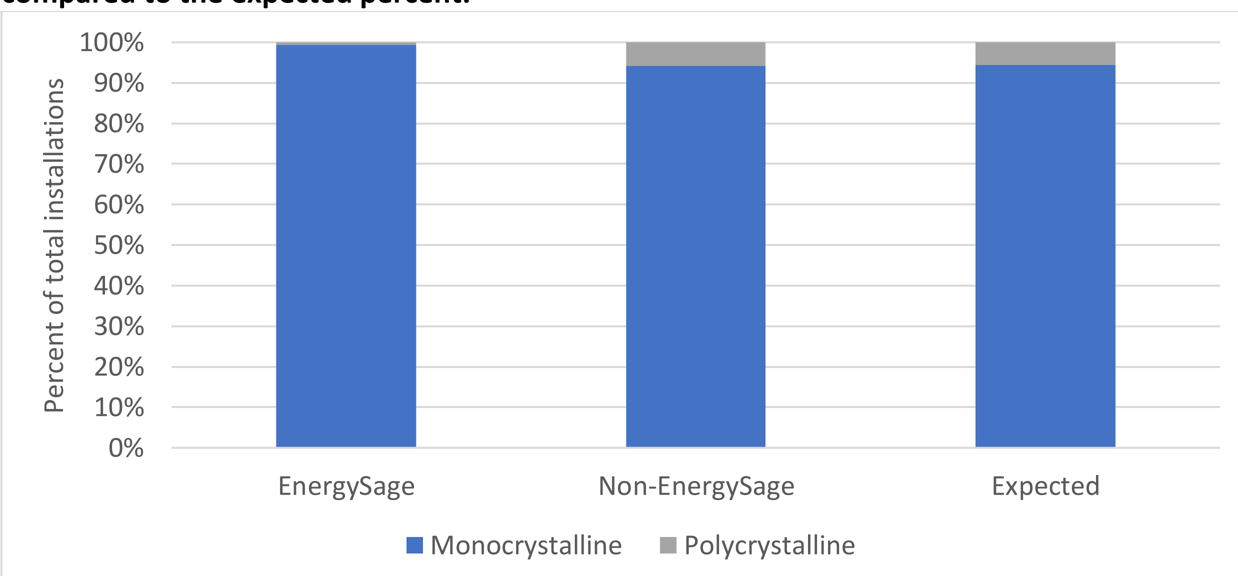
Dependent variable	**EnergySage mean rate of annual change**	**Non-EnergySage mean rate of annual change**	**P-value**
Module capacity	14 Watts	14 Watts	0.958
Module efficiency	0.4 percentage points	0.4 percentage points	1.000

4.3 Solar Equipment Type (H3)

Both null hypotheses relating to equipment type were rejected, meaning the alternative research hypothesis three (H3) was accepted. There was a relationship in if installations were completed through EnergySage or not and module cell type, *χ2* (1, *N*=583144) = 326.172, *p* <

0.05; ϕ = 0.024 (Appendix IVe). The effect size for this analysis (ϕ = 0.024) was weak ($\phi < 0.1$). The count of installations with monocrystalline solar cell modules (99.4%) was higher than the expected count (94.4%) for those completed through EnergySage, while the count of installations with polycrystalline solar cell modules (0.63%) was lower than the expected count (5.6%). The counts for installations with monocrystalline (94.3%) and polycrystalline (5.7%) solar cell modules were comparable to the expected counts for those completed outside of EnergySage (Figure 4, Table 9).

Figure 4. Percent of total installations with monocrystalline and polycrystalline solar cell modules for installations completed through EnergySage vs. external to EnergySage, compared to the expected percent.



There was a relationship in if installations were completed through EnergySage or not and inverter type, $\chi 2$ (2, N=537357) = 1129.743, $p < 0.05$; ϕ = 0.046 (Appendix IVf). The effect size for this analysis (ϕ = 0.046) was weak ($\phi < 0.1$). The count of installations with microinverters (64.3%) was higher than the expected count (43.7%) for those completed through EnergySage, while the count of installations with optimized string inverters (33.5%) and string inverters (2.2%) was lower than the expected count (49.8% and 6.5%, respectively). The counts for installations with microinverters (43.5%), optimized string inverters (49.9%), and string inverters (6.6%) were similar to the expected counts for those completed outside of EnergySage (Figure 5, Table 9).

Figure 5. Percent of total installations with microinverters, optimized string inverters, and string inverters for installations completed through EnergySage vs. external to EnergySage, compared to the expected percent.

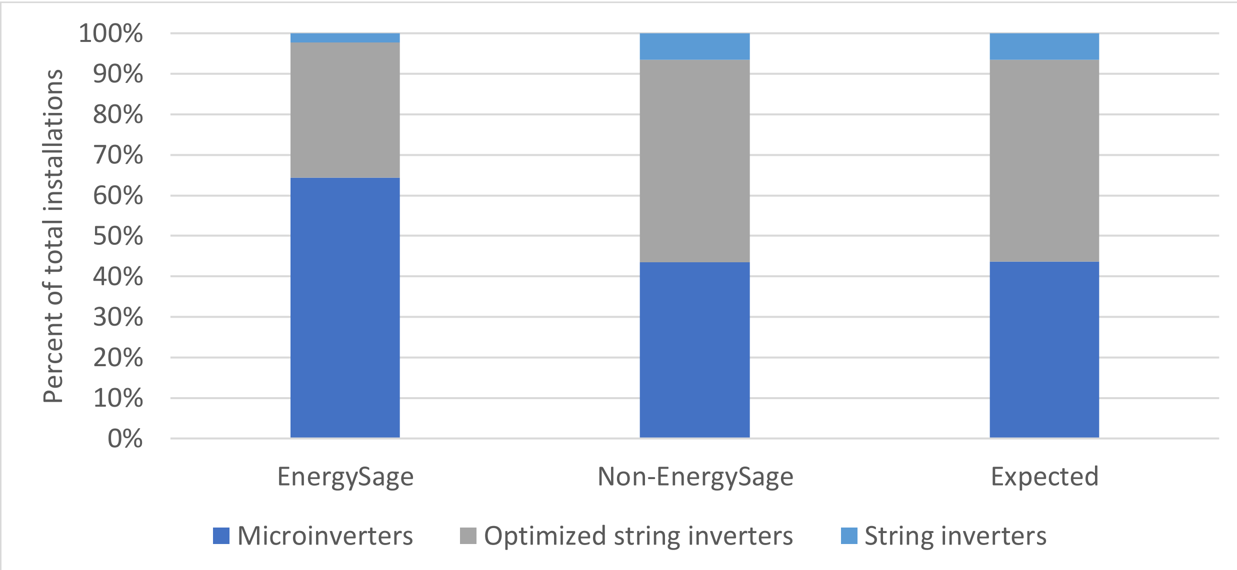


Table 9. Summary of solar equipment type (H3) results.

Equipment category	Equipment type	Total percentage EnergySage	Total percentage non-EnergySage	Expected percentage	P-value
Solar modules	Monocrystalline	99.4%	94.3%	94.4%	<0.05
	Polycrystalline	0.63%	5.7%	5.6%	
Inverters	Microinverters	64.3%	43.5%	43.7%	<0.05
	Optimized string inverters	33.5%	49.9%	49.8%	
	String inverters	2.2%	6.6%	6.5%	

V. Discussion

Based on efficiency, capacity, and cell type, solar modules installed through EnergySage were of significantly higher quality compared to the country as a whole. These results align with quote-level data included in EnergySage's most recently published biannual *Intel Report 15*, covering the first half of 2022, which found that the most quoted solar modules in the

EnergySage Marketplace are offered by REC (25%), Panasonic (16%), and Q CELLS (16%). These three top module brands accounted for 66% of all quotes in the first half of 2022 and all offer high-quality equipment (EnergySage, 2022). The current module series by each of these companies range in efficiency from 19.1% to 22.3% (Table 10). Across country-wide installations completed in 2021, the module efficiencies were lower; according to LBNL's *Tracking the Sun* report, most modules ranged in efficiency from 19% to 21% (Barbose et al., 2022). However, while current series can be an indicator of the company's overall equipment quality, they do not necessarily reflect the equipment included in previous and even current installations because installers often carry older equipment.

Installations completed through EnergySage almost exclusively included monocrystalline solar modules, while non-EnergySage installations primarily included monocrystalline modules, but included some polycrystalline as well. All top EnergySage solar module companies (REC, Panasonic, and Q CELLS) currently only offer monocrystalline solar modules (Table 10). Similarly, LBNL's 2022 *Tracking the Sun* report found that, depending on the industry segment, monocrystalline modules have increased share to about 89% to 98% in 2021 alone (Barbose et al., 2022). According to EnergySage's *Intel Report 15*, high-capacity modules are also increasing in EnergySage Marketplace quotes; in the first half of 2022, 390+ W modules accounted for about 50% of all quotes, increasing from just 16% in the second half of 2021 (EnergySage, 2022).

Table 10. Module efficiency ranges and cell type of current series by brand. Data were obtained directly through company websites.

Brand	Efficiency range	Cell type
REC	20.6%	Monocrystalline
Panasonic	19.7%-22.2%	Monocrystalline
Q CELLS	19.1%-22.3%	Monocrystalline

However, in the residential solar market segment, SunPower solar modules are the most efficient at 22.8% efficiency (Marsh, 2022b). SunPower is among the top 10 solar installers in the United States, with 2.7% market share in 2021, and it offers its own line of solar module equipment (Connelly, 2022). While SunPower solar modules can be quoted by EnergySage installers, they comprise a small percentage of marketplace share, at only 5% in the first half of 2022 (EnergySage, 2022). Thus, if SunPower gains more share in the larger solar market, while remaining a less quoted module brand on EnergySage, it could drive up solar equipment quality outside of the EnergySage Marketplace.

Almost all installations completed through EnergySage included inverters with MLPEs (97.8%), while installations external to EnergySage were slightly less likely to include MLPEs (93.4%). EnergySage installations included more MLPEs in the form of microinverters, while external installations included more optimized string inverters. Data from EnergySage's *Intel Report 15* support these findings, showing that Enphase, a microinverter company, was the most quoted inverter brand on the EnergySage Marketplace in the first half of 2022 (59%), followed by SolarEdge, an optimized string inverter company (32%); thus, these two MLPE inverter companies represent 91% of total quotes. Similarly, LBNL's 2022 *Tracking the Sun* report found that optimized string inverters have led overall in terms of inverter technology growth since 2013, but in recent years, microinverters have gained share (Barbose et al., 2022).

The rate of module quality increase, based on efficiency and output capacity, was not statistically different between EnergySage and non-EnergySage installations. Across both datasets, the average efficiency of modules has increased by 0.4 percentage points annually while the average capacity of modules has increased by about 14 W annually over the past three years, indicating that solar technology is advancing and equipment quality is increasing. LBNL's 2022 *Tracking the Sun* report found that between 2002 and 2021, the average module efficiency increased by 48% across the United States (6.5 percentage points), supporting this trend (Barbose et al., 2022). Similarly, according to EnergySage's *Intel Report 15*, in the first half of 2018 the most quoted solar modules were between 320 and 330 W (35%), while in the first half of 2022, the most quoted modules were between 390 and 400 W (32%). EnergySage equipment consistently remained of higher quality each year from 2019 through 2021, suggesting that comparison shopping results in higher solar equipment quality.

As previously discussed, installers are pre-vetted by EnergySage based on strict criteria before being welcomed into its network (see Introduction), which results in regulation of the marketplace. Furthermore, EnergySage shoppers receive access to free expert Energy Advisors that can answer questions and guide them through the decision-process when choosing an installer and solar equipment. Ultimately, consumers visiting EnergySage's marketplace can compare quotes and equipment, and benefit from increased installer competition on the platform.

A 2017 study by the National Renewable Energy Laboratory (NREL) similarly found that consumers can benefit from obtaining multiple quotes and increased installer competition when installing solar (O'Shaughnessy & Margolis, 2017). According to NREL's study, EnergySage's installer network provided lower priced quotes compared to external installers (O'Shaughnessy & Margolis, 2017). NREL posited that both increased competition and price transparency might result in lower prices (O'Shaughnessy & Margolis, 2017); similarly, this competition and equipment transparency and choice could be driving the higher quality of solar equipment in installations completed through EnergySage.

VI. Limitations of the Study

This study was limited by the number of equipment quality metrics it could assess. It only examined metrics included in LBNL's *Tracking the Sun* dataset because researching other parameters for the equipment models would take time beyond what was allocated for this study; while modules with high efficiency and high capacity do tend to also be favorable in terms of other quality parameters, more research is required to confirm the correlation. Some data were excluded in the analysis due to gaps in equipment quality variables. Some installations also included multiple module and inverter models; however, only the primary equipment was included in this analysis. Additionally, the study was not able to control for quote comparison that consumers perform outside of EnergySage; thus, it should be thought of as preliminary research.

Finally, the EnergySage and non-EnergySage datasets were geographically similar in that most installations occurred on the West Coast and in the Northeast, but not geographically identical, likely in part because EnergySage does not have installer networks in every state (Figure 1). Previous research by EnergySage has indicated that solar equipment can vary substantially by state, which could impact quality metrics (EnergySage, 2022). For example, EnergySage's *Intel Report 15* found there were 11 different solar module brands that were the most quoted in at least one state in the United States (EnergySage, 2022).

VII. Future Study

Comparison shopping research is extremely limited in general, and is even further limited in the solar industry. Additional research should be conducted to assess how other solar equipment quality metrics, including temperature coefficient, power warranty, and performance warranty, are impacted by shopping through quote comparison platforms. Similar analysis performed in this study could also be applied to other technology, such as solar batteries. Finally, more research should be performed to understand the driving force behind higher equipment quality on EnergySage; for example, a consumer survey could be run to better understand why consumers choose certain equipment, such as because it was simply what was included in the lowest-priced quote, their Energy Advisor helped them choose, or they requested and/or selected specific equipment based on their own research.

VIII. Conclusion

According to the U.S. Energy Information Administration, renewable energy represents about 20% of total electricity generation in the United States (EIA, 2022). The Biden administration has a goal of reaching 100% clean electricity by 2035, which means that solar and other renewable energy sources need to rapidly scale over the next decade (The White House, 2021). While solar is significantly more sustainable than fossil-fuel generation sources, some modules do contain metal, including lead and cadmium, that are considered hazardous waste (U.S. EPA, 2022). Therefore, as the rate of solar deployment increases, the quality solar equipment is crucial to reduce waste. Ultimately, high quality solar energy systems should provide high power output (reducing the number of modules needed at once) and last a long time (reducing the number of modules needed over time). The results of this study indicate that while the quality of solar equipment is increasing in installations across the country, it is highest in those completed through EnergySage. The competition driven by the EnergySage Marketplace may incentivize installers to quote higher quality equipment and allow consumers to request and choose certain equipment based on performance factors, which could ultimately result in less solar module waste in the future.

IX. References

Aggarwal, V. (2022, November 9). *Solar panel output: How much energy does a solar panel produce?* EnergySage. https://news.energysage.com/what-is-the-power-output-of-a-solar-panel/

Barbose, G., Darghouth, N., O'Shaughnessy, E., & Forrester, S. (2022, September). *Tracking the Sun*. Lawrence Berkeley National Laboratory. https://emp.lbl.gov/tracking-the-sun

Connelly, C. (2022, March 29). *Competition heats up for the top three US residential solar installers*. Wood Mackenzie. https://www.woodmac.com/news/opinion/competition-heats-up-for-the-top-three-us-residential-solar-installers/

EIA. (2022, November). What is U.S. electricity generation by energy source? https://www.eia.gov/tools/faqs/faq.php?id=427EnergySage. (2021, June 9). *Choosing a Solar Installer*. https://www.energysage.com/solar/how-to-go-solar/choosing-solar-installer/

EnergySage. (2022, March). *Solar Marketplace Intel Report: Data from H1 2021 to H2 2021*. https://www.energysage.com/data/

Hayes. (2022a, July 20). *T-Test: What It Is With Multiple Formulas and When To Use Them*. Investopedia. https://www.investopedia.com/terms/t/t-test.asp

Hayes. (2022b, October 23). *Chi-Square (χ2) Statistic: What It Is, Examples, How and When to Use the Test*. Investopedia. https://www.investopedia.com/terms/c/chi-square-statistic.asp

IRENA. (2016, June). *End-of-life management: Solar Photovoltaic Panels*. https://www.irena.org/publications/2016/Jun/End-of-life-management-Solar-Photovoltaic-Panels

Marsh, J. (2022a, October 3). *Monocrystalline and polycrystalline solar panels: which is the best for you?* EnergySage. https://news.energysage.com/monocrystalline-vs-polycrystalline-solar/

Marsh, J. (2022b, November 29). *What are the most efficient solar panels?* EnergySage. https://news.energysage.com/what-are-the-most-efficient-solar-panels-on-the-market/

Ong, B. S. (2011). Online Shoppers' Perceptions and Use of Comparison-Shopping Sites: An Exploratory Study. *Journal of Promotion Management*, *17*(2), 207–227. https://doi.org/10.1080/10496491.2011.553789

O'Shaughnessy, & Margolis. (2017, April). *Using Residential Solar PV Quote Data to Analyze the Relationship between Installer Pricing and Firm Size*. National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy17osti/68010.pdf

Solar Energy Technologies Office. (2021, August 9). *Solar Futures Study*. Energy.Gov. https://www.energy.gov/eere/solar/solar-futures-study

Solar Energy Technologies Office. (2022). *Solar Photovoltaics Supply Chain Review Report*. Energy.Gov. https://www.energy.gov/eere/solar/solar-photovoltaics-supply-chain-review-report

Statistics Solutions. (2021, August 10). *Conduct and Interpret a One-Way ANCOVA*. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/one-way-ancova/

Tao, M., Fthenakis, V., Ebin, B., Butler, E., Sinha, P., Corkish, R., Wambach, K., & Simon, E. (2020). Major Challenges and Opportunities in Silicon Solar Panel Recycling. *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. https://doi.org/10.1109/pvsc45281.2020.9300650

Thaler, R. H., & Sunstein, C. R. (2021). *Nudge: The final edition*. Penguin Books.

Thoubboron, K. (2022, November 29). *String inverters vs. power optimizers vs. microinverters: what you need to know*. EnergySage. https://news.energysage.com/string-inverters-power-optimizers-microinverters-compared/

The White House. (2021, April 22). *FACT SHEET: President Biden Sets 2030 Greenhouse Gas Pollution Reduction Target Aimed at Creating Good-Paying Union Jobs and Securing U.S. Leadership on Clean Energy Technologies*. https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/22/fact-sheet-president-biden-sets-2030-greenhouse-gas-pollution-reduction-target-aimed-at-creating-good-paying-union-jobs-and-securing-u-s-leadership-on-clean-energy-technologies/

US Census Bureau. (2022, September 1). *TIGER/Line Shapefile*. Census.gov. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

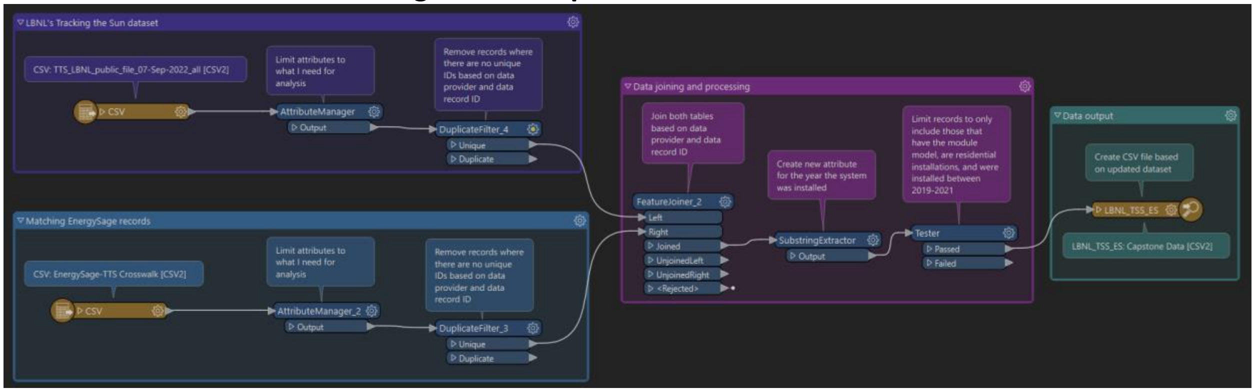
U.S. EPA. (2022, August 28). *End-of-Life Solar Panels: Regulations and Management*. https://www.epa.gov/hw/end-life-solar-panels-regulations-and-management

Wan, Y. (2009). Comparison-Shopping Services and Agent Design. *Comparison-Shopping Services and Agent Designs*, 1–17. https://doi.org/10.4018/978-1-59904-978-6.ch001

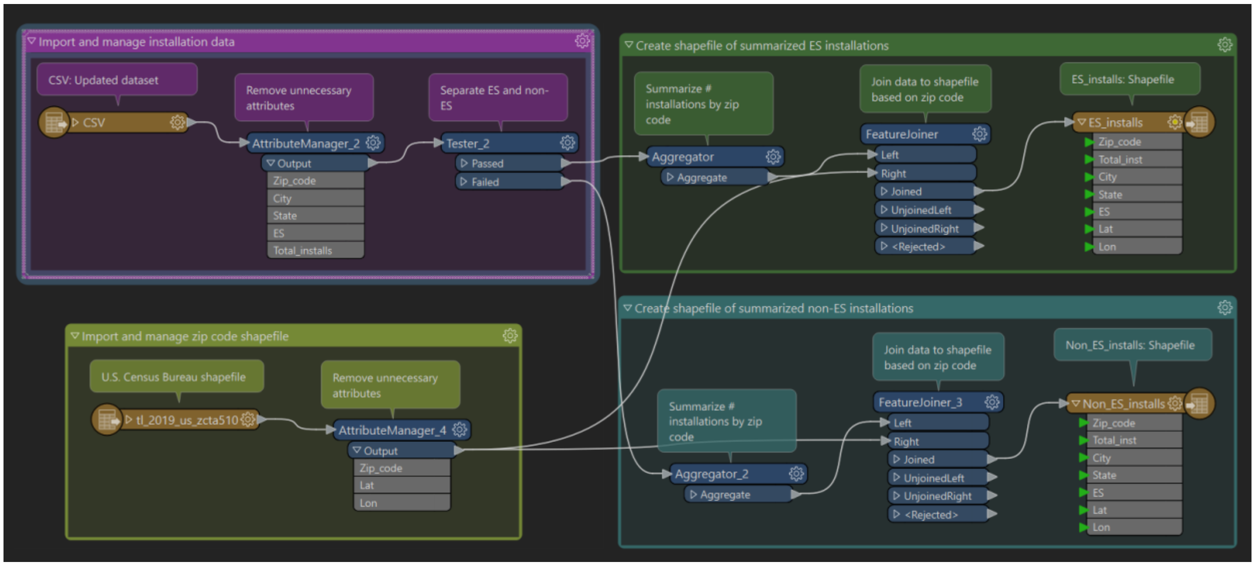
APPENDIX I. Pared attributes included in analysis.

Dataset	Variable	Description
Tracking the Sun & matching EnergySage records	data_provider_1	Agency that reported the data.
	system_ID_1	The LBNL unique ID used to identify the system.
Tracking the Sun	installation_date	The date of the installation.
	zip_code	The zip code in which the solar energy system was installed.
	module_manufacturer_1	The manufacturer of the primary solar module.
	module_model_1	The model of the primary solar module.
	technology_module_1	The type of solar cell technology used in the primary module.
	nameplate_capacity_module_1	The rated capacity of the primary module, measured in W.
	efficiency_module_1	The efficiency of the primary module.
	inverter_manufacturer_1	The manufacturer of the primary solar inverter.
	inverter_model_1	The model of the primary solar inverter.
	micro_inverter_1	Is the primary inverter a microinverter?
	DC_optimizer	Does the primary inverter include a DC optimizer?
Matching EnergySage records	quote_id	The EnergySage unique ID used to identify the system and match it with the LBNL unique ID.

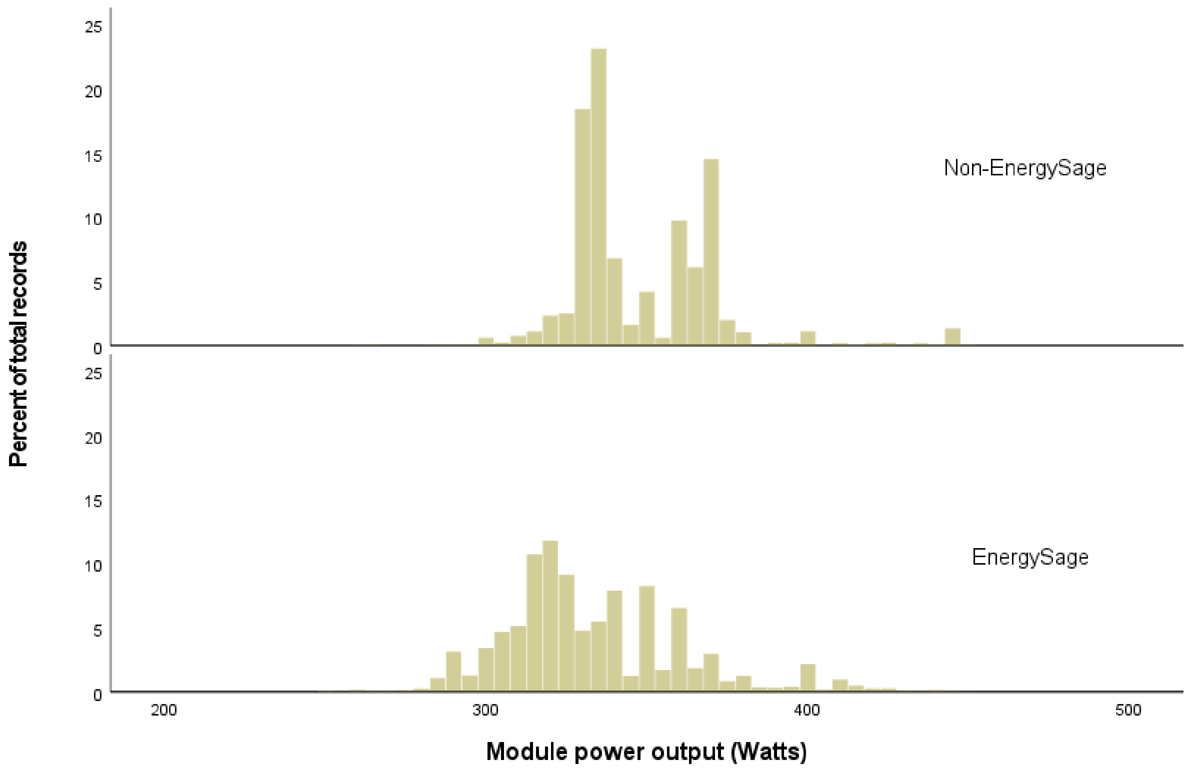
APPENDIX IIa. Full data cleaning workflow performed in Safe Software's FME workbench.

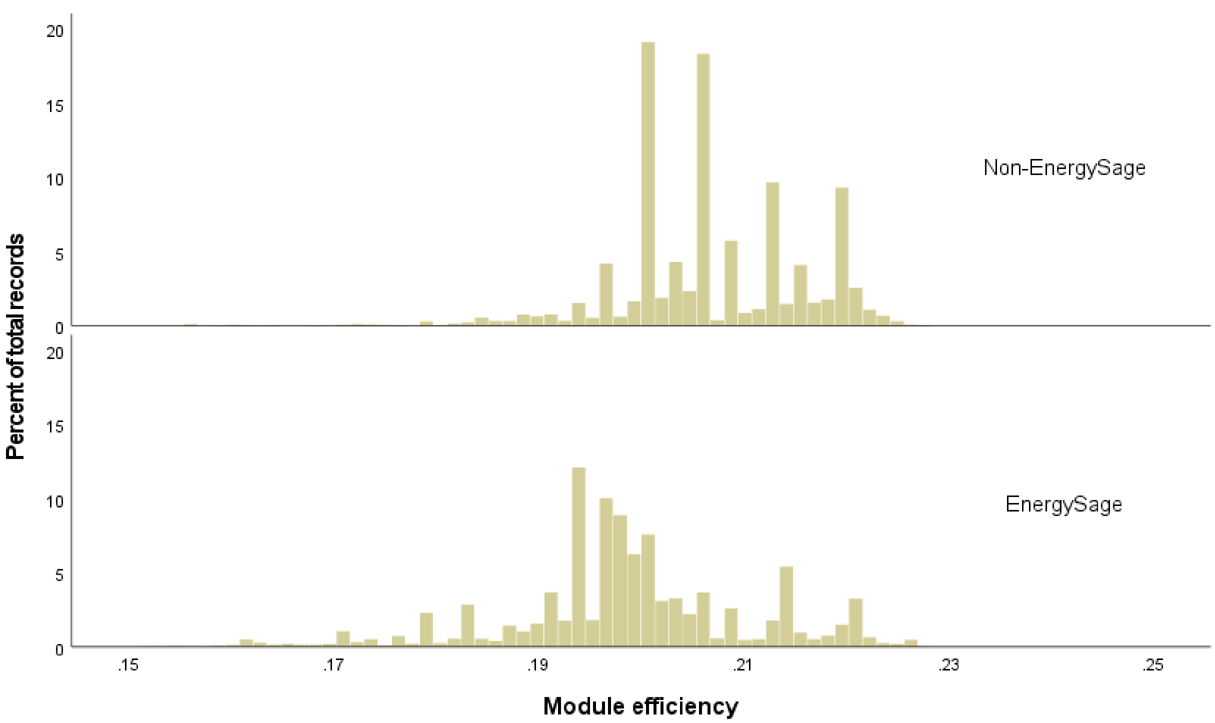


APPENDIX IIb. Full location analysis workflow performed in Safe Software's FME workbench.



APPENDIX III. Distributions of module power output and module efficiency for EnergySage and non-EnergySage installations.





APPENDIX IVa. SPPS results from t-test of module capacity.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
nameplate_capacity_module_1	Equal variances assumed	256.358	<.001	-42.106	585878	.000	.000	-14.713	.349	-15.398	-14.028
	Equal variances not assumed			-51.340	7061.165	.000	.000	-14.713	.287	-15.274	-14.151

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
nameplate_capacity_module_1	Cohen's d	28.686	-.513	-.537	-.489
	Hedges' correction	28.686	-.513	-.537	-.489
	Glass's delta	23.458	-.627	-.653	-.601

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

APPENDIX IVb. SPPS results from t-test of module efficiency.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
efficiency_module_1	Equal variances assumed	229.409	<.001	-54.610	579300	.000	.000	-.008362372	.0001531279	-.008662498	-.008062247
	Equal variances not assumed			-73.065	7095.569	.000	.000	-.008362372	.0001144504	-.008586729	-.008138015

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
efficiency_module_1	Cohen's d	.0125529174	-.666	-.690	-.642
	Hedges' correction	.0125529336	-.666	-.690	-.642
	Glass's delta	.0093376061	-.896	-.924	-.867

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

APPENDIX IVc. SPPS results from ANCOVA test of module capacity.

Tests of Between-Subjects Effects

Dependent Variable: Avg_capacity

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	735.688	1	735.688	7792.806	.007
	Error	.094	1	.094[a]		
ES	Hypothesis	.094	1	.094	.004	.957
	Error	51.793	2	25.897[b]		
Year	Hypothesis	754.052	1	754.052	29.118	.033
	Error	51.793	2	25.897[b]		
ES * Year	Hypothesis	.090	1	.090	.003	.958
	Error	51.793	2	25.897[b]		

a. MS(ES)

b. MS(Error)

APPENDIX IVd. SPPS results from ANCOVA test of module efficiency.

Tests of Between-Subjects Effects

Dependent Variable: Avg_efficiency

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	6.084E-5	1	6.084E-5	3574670.234	<.001
	Error	1.702E-11	1	1.702E-11[a]		
ES	Hypothesis	1.702E-11	1	1.702E-11	.000	.997
	Error	2.667E-6	2	1.333E-6[b]		
Year	Hypothesis	6.400E-5	1	6.400E-5	48.000	.020
	Error	2.667E-6	2	1.333E-6[b]		
ES * Year	Hypothesis	.000	1	.000	.000	1.000
	Error	2.667E-6	2	1.333E-6[b]		

a. MS(ES)

b. MS(Error)

APPENDIX IVe. SPPS results from chi-square test of module type.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	326.172[a]	1	<.001		
Continuity Correction[b]	325.220	1	<.001		
Likelihood Ratio	517.805	1	<.001		
Fisher's Exact Test				<.001	<.001
N of Valid Cases	583144				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 385.35.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	-.024	<.001
	Cramer's V	.024	<.001
N of Valid Cases		583144	

APPENDIX IVf. SPPS results from chi-square test of inverter type.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1129.743[a]	2	<.001
Likelihood Ratio	1158.607	2	<.001
N of Valid Cases	537357		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 407.60.

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.046	<.001
	Cramer's V	.046	<.001
N of Valid Cases		537357	