Analyzing Solar Power Utilization and Adoption Trends

MGT 6203 Final Project Report

Kirpa Singh, Chetna Kewalramani, Lyle Sweet, Karthik Madhur, Nai Ning Chi

Contents

Introduction

Over the past decade, environmental concerns have been rising to the top of regional and global agendas. In the McKinsey Global Energy Perspective 2022 report, global power consumption is projected to triple by 2050 as living standards and electrification grow [1]. Among the many different decarbonization efforts, electrification with renewable energies such as solar and wind power is the most economic and easiest to implement in most sectors. For solar power, it is predicted that the global solar PV (photovoltaic) capacity will grow 30 times by 2030. As countries aim for ambitious decarbonization targets, renewable energy, led by solar power is set to become the cornerstone of the worlds' power supply.

In the US, according to the solar market insight and research data from SEIA (Solar Energy Industries Association)[2], solar power has grown at a rate of 33% annually in the last decade. The cost of solar energy installation has dropped by more than over 60% over the past decade. The average sized residential solar system has dropped from $5.79 per system price ($/Watt) in 2010 to roughly $1.38 per system price ($/Watt) in 2021. Thanks to the support from federal policies such as the solar tax credit, declining costs of solar technology and increasing demand for clean energy for both private and public sectors, more than 135 gigawatts (GW) of solar capacity installed as of 2021, enough to power 23 million homes in the U.S. and the number continues to grow.

A recent study (Araújo et al., 2019)[3] investigated the trends with socio-demographic, political and geographic profiles in New York to identify the adoption of solar energy by time series clustering and regression. The results indicated clustering analysis as statistically significant on the differences between the less active group and active solar adopter with factors such as medium income, medium home value, % undeclared party affiliation, population etc. In the regression analysis on solar installed capacity (kW) with demographics, political affiliation and other attributes, political orientation is not predictive as Democratic and Republican tendencies appeared to be insignificant to solar installed capacity. However, the study identified a weak relationship for undeclared voters with the adoption of solar energy, which suggested that undeclared voters might be more open to innovative technologies. To expand from the findings of this study, it would be interesting to analyze the effect of political orientation on solar energy adoption from a different perspective of how people identify their political positions, such as conservatives and liberals.

While California has been the dominating market for the U.S. solar industry, markets in other states are expanding rapidly. In 2021, states other than California made up the largest share of the market, led by Florida and Texas. As demand for solar energy continues to grow nationwide, identifying potential markets and customer demographics will help the solar industry target more customers and ensure the logistic and supply deployment to the markets.

Problem Statement

The primary question we plan to address over the course of this project is: How has solar energy grown in recent years and what factors influence customer adoption of solar energy? According to data from the Energy Information Administration (EIA)[4], net generation of solar has been steadily increasing over the past decade. As more money gets poured into this rapidly growing industry, it is necessary to do further research to better understand solar markets, the target customer, and their motivations. Our survey dataset[5] contains several potential factors that we plan to analyze for solar adopters vs. non-adopters. These include demographic factors, such as education level and income, as well as other factors like square footage, type of home, and proximity to others using solar energy. Additionally, the survey dataset includes questions on political beliefs (conservative, liberal, moderate, etc.) and climate consciousness that we also hope to address in our analyses.

We are also interested in the differences between residential homeowners who lease their solar panels and those who purchase their solar panels. Leasing vs. owning provides flexibility to customers and presents two different business models for installation providers. If there is a noticeable difference between leasing and buying, then installation providers can adjust their marketing strategy when expanding into new markets. We plan to look at customer demographics, reasons for choosing solar, and other factors to understand the difference between customers.

Building off our primary question, we will also be answering the question: is electricity cost a primary factor in the regions of solar energy utilization? Our datasets include a wide variety of valuable information such as how important saving on electricity is for survey responders, how much adopters are saving post-solar installation, as well as time-series data on average retail price of electricity over the past decade. We plan to use these datasets to reach a conclusion regarding electricity cost as a potential driving factor for solar adoption.

Our initial hypothesis is that more affluent households are likely to adopt solar, particularly those with liberal ideas who may care more about climate change and environmental impact. Though this hypothesis may seem obvious, as solar becomes cheaper and traditional energy costs rise, solar will be more appealing to larger audiences for purely economic reasons. We also hypothesize that cost will be a primary factor when choosing solar as very few people, even those with environmental concerns, are unwilling to pay extra for their ideals. Regarding solar adoption via leasing vs. buying, we predict that leasing will be more popular with lower income households, since they may be unable to afford the upfront cost of installation.

Our analyses will help us better understand solar markets by understanding the thought process behind solar adoption. We hope the results will give us insight into some of the factors that significantly affect solar adoption, which can help companies when identifying areas for growth.

Data Overview

The data used in this analysis comes from three main sources. Our Solar Generation dataset[4] provides data on net generation from solar in thousand megawatt hours at the national and state level from January 2014-December 2022. It distinguishes between residential, commercial, and industrial sectors. The Survey dataset[5] contains 3 different surveys - “Adopter”, “General Population”, and “Considerer”. All survey responses combined provide data for understanding how those who do not have rooftop PV differ from those who do. Survey questions include state of residence, home type, household income, age, education level, purchasing habits, etc. The Electricity Price dataset[6] provides data on retail prices of electricity in cents per kilowatt-hour from January 2014-December 2022 in residential, commercial, industrial, and transportation sectors. For the purpose of this analysis, we will be focusing on the residential sector.

Cleaning the survey data consisted of three main steps. The first step was to combine the survey data into one dataset that we can use to create our models. This would easily allow us to compare the responses of the three groups of respondents. The second step was to create dummy variables for the factors we were analyzing. The survey response data was categorical and had to be transformed into something measurable. Additionally, some responses were missing or withheld. To determine how to best deal with these values, three different versions of the data were created to build our models: one where those values were replaced with NA, one where they were replaced with the same value, like 99, to mean “missing”, and one where they were replaced with the mode of that category.

The final step of the data cleaning was extracting a subset of data from the Solar Generation[4] and Electricity Price[6] datasets. The survey data only covered four states – NJ, NY, AZ, and CA. To relate all three datasets, we focused on data relating to these four states from the Solar Generation[4] and Electricity Price [6] datasets. For the analysis, we merged the solar generation and electricity price by month into one dataset. Since the Solar Generation dataset has units in Thousand Megawatt-hour, we modified the electricity price from Cents per Kilowatt-hour to dollar per Megawatt-hour so that the units are consistent between these two parameters. To visually represent the data collected and cleaned in this study, a couple of plots as an example have been included below:

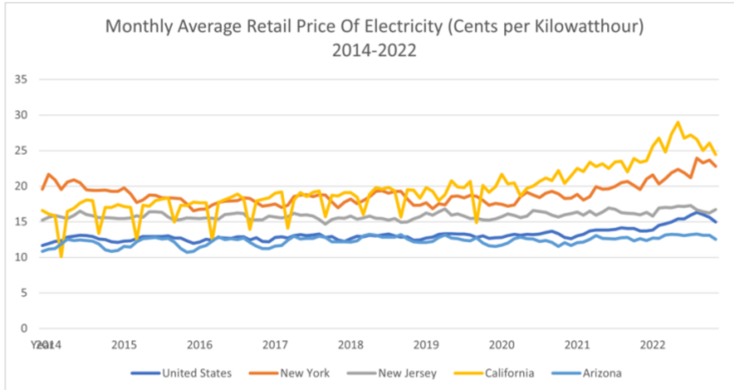


Figure 1: Monthly Average Retail Price of Electricity (cents per kWh) 2014-2022

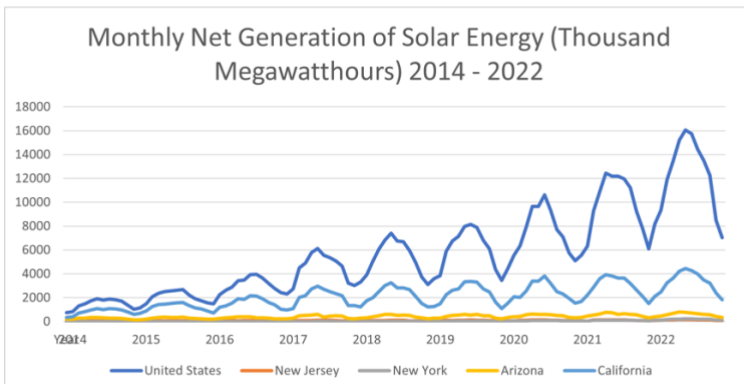


Figure 2: Monthly Net Generation of Solar Energy (Thousand MWh) 2014-2022

Modeling

We will primarily rely on regression analysis to answer our research questions. We will be using two regression models for the analysis. First, we will use logistic regression to determine the factors that influence customer adoption of solar energy. Equation 1 presents the logistic model.

$$\text{Logit(p)} = \beta_0 + \beta_1 Educ_i + \beta_2 Income_i + \gamma OtherFactors_i + \epsilon_i \qquad \text{...(1)}$$

p is Probability that an individual i adopted solar energy i.e. $Y_i = 1$, $Educ_i$ is the education level of individual i , $Income_i$ is the income level for the individuals. $OtherFactors_i$ is the vector of variables that include other factors such as square footage of house, political stance, hearing about solar panels from peers, and others. Using the above model, we will estimate marginal effects of different factors on probability of customer adoption of solar power. In our case, the marginal effect of a predictor variable (for example Education level) is the change in the predicted probability of the outcome variable (Individual adopted solar energy) associated with a one-unit increase in that predictor, while holding all other predictors constant.

To determine the differential effect of factors for leasing vs. buying solar panels, we'll run the above model with two different dependent variables. Let's define $Z_{1i} = 1$ if the customer bought the solar panels and 0 otherwise. Similarly, let $Z_{2i} = 1$ if the customer leased the solar panels and 0 otherwise. We will estimate equation 1 using p=Pr($Z_{1i} = 1$) to estimate the effect of factors on buying solar panels and p=Pr($Z_{2i} = 1$) to estimate the effect of factors on leasing solar panels.

In the second model, we will employ linear regression analysis on time series data on monthly residential generation of solar and monthly average retail price of electricity. Equation 2 represents the linear regression model.

$$Generation_t = \beta_0 + \beta_1 \cdot price_t + \epsilon_t \quad \text{...(2)}$$

$Generation_t$ is the residential net generation of solar power and $price_t$ is the average retail price of electricity during month t. One of the challenges faced in the linear regression model above is that we have found a cyclic seasonality in the retail price per month of electricity based on the demand. The retail price of electricity always seems to increase during the summer months. Prices are usually highest in the summer when total demand is high because more expensive generation sources are added to meet the increased demand. To control the cyclical variation in the magnitude of monthly residential solar power generation, in equation (3), we will add dummy variables for 11 months with January as the base month. We plan to identify the trends in solar adoption from this analysis.

$$Generation_t \; = \; \beta_0 + \beta_1 \cdot price_t + \beta_2 \cdot Feb + \beta_3 \cdot March + .. + \beta_{12} \cdot Dec + \epsilon_t \quad \text{...(3)}$$

Results

What factors affect solar adoption the most?

The first model we created was a logistic regression model to predict whether a customer would adopt solar panels based on the following factors: square footage of their house, education, income, climate consciousness, political stance, and which state they lived in. Three different versions of this model exist, differentiated only by how the data used for each dealt with missing values.

For the model that replaced all missing and withheld values with NA and then disregarded those data points in the model, the most significant factors were whether the customer lived in California, if they had an income above $150k, if the square footage of their home was between

1500-2000, and if they had any higher education level. These factors have a positive impact on the prediction for a customer adopting solar.

For the next model that replaced all those NA values with 99, the most significant factors were whether the customer lived in California, if they had an income above $100k or was unknown, if the square footage of their home was between 1500-2000 or was unknown, if they had any higher education level, if they agreed with having a personal obligation to prevent climate change or did not disclose that information at all, and if they identified as conservative or did not disclose their political affiliation. These factors have a positive impact on the prediction for a customer adopting solar.

For the last model, the missing values were replaced with the mode response to that survey question. The most significant factors were whether the customer lived in California, if they had an income above $100k, if the square footage of their home was between 1500-2000, if they had any higher education level, if they strongly agreed with having a personal obligation to prevent climate change, and if they identified as moderate or conservative. These factors have a positive impact on the prediction for a customer adopting solar.

Looking at all three of these models, Californians, those with higher incomes and higher education are more likely to adopt solar. The model with the lowest AIC is the first one, the one that disregarded the missing/withheld data points. Using this model to evaluate our initial hypotheses, we can say that our initial hypothesis that more affluent households are more likely to adopt solar is accurate. The model indicates that having a liberal stance is insignificant so that hypothesis appears incorrect. Finally, strongly agreeing and agreeing with having a personal obligation to stop climate change are significant factors as well. Strongly agreeing does have a positive impact on likelihood to adopt, but agreeing has a negative impact.

What role does social influence play in solar adoption?

To better understand how social influence may play a role in solar adoption, we ran a logistic regression on a factor from the survey data – INFLNC_PPLTALK. The prompt is: "While considering solar, I frequently heard people I know talking about solar panels."

There are five degrees of responses ranging from strongly disagree (1) to strongly agree (5). The only statistically significant factor for predicting solar adoption in this model is for people who Strongly Agree (5) with the prompt. The relationship is strongly negative and seems to suggest that if someone considering solar panels is in a setting where their peers frequently discuss solar, that solar could have a negative connotation in the peer group. This result is unexpected as we had suspected that someone who had heard more about solar from their peer group would be more likely to adopt.

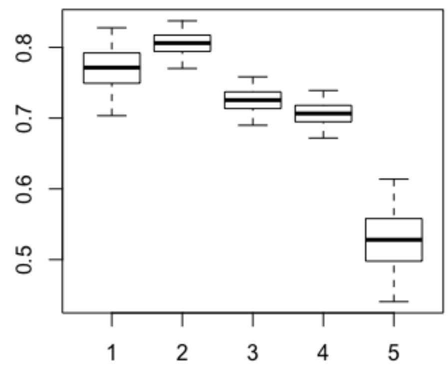


Figure 3: Distributions of Predicted Responses

Is there a difference in factors for leasers vs. buyers?

The third model we created was a logistic regression model for which the response variable was whether the customer bought or leased their solar panels. The BUY_LEASE factor is a Boolean factor that indicates whether the responder bought (0) or leased (1). For this model we had to exclude the "General Population" survey responses because the question of buying vs. leasing was only given in the "Adopters" and "Considerers" surveys. In our first model where we used different methods of handling missing data, the model that kept all missing values as "NA" had the lowest AIC score, so we decided to use the "NA" data for this model as well.

The most significant factor for the buy vs. lease model was whether the responder lived in Arizona. This factor had a negative effect on the probability of leasing, meaning Arizonans are more likely to buy. Some other significant factors included whether the responder lived in California, whether the responder has a master's/doctoral degree, and whether the responder has a home living space of >2500 square feet. For all these factors the coefficient was negative, meaning they increase the probability of the responder being a buyer rather than a leaser. We had originally predicted that leasing would be more popular for responders with a lower income, but the model shows that the correlation for income is insignificant.

Is electricity cost a factor in solar energy generation?

To answer this question, we conducted linear regression analysis on time series data of monthly residential generation of solar power and monthly average retail price of electricity. We employed four linear regression models for each state i.e., New York, New Jersey, California, and Arizona.

All the models showed a positive and significant correlation between price and net generation, which means that there is a positive association between retail price of electricity and solar power. The coefficients in Table 1 provide suggestive evidence in support of our initial hypotheses. Out of the four states, California has the largest coefficient. From Table 1 column 2, we can see that in California, $1 increase in retail price per Megawatt-hour is associated with 8.6

thousand Megawatt-hour increase in residential generation of solar power. New York has smallest coefficient on retail price. In New York, $1 increase in retail price per Megawatt-hour is associated with 0.76 thousand Megawatt-hour increase in residential generation of solar power. Coefficients for all the four states are significant at 99% confidence level.

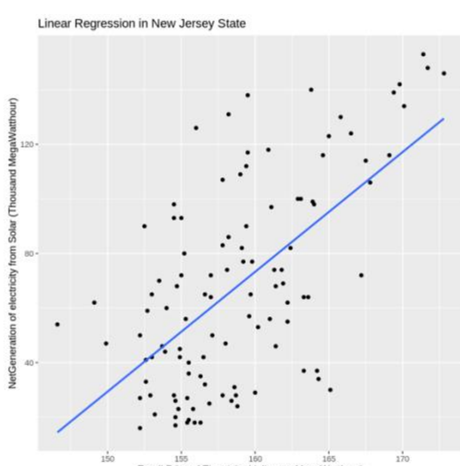
During our analysis we observed that there is a seasonal pattern in retail prices for electricity where the prices fluctuate a lot in a fixed period. For example, in California, the price of electricity dropped every 6 months – there was a steep decline in March and September between 2014 to 2019. Similarly, we also find seasonal patterns in generation of solar power.

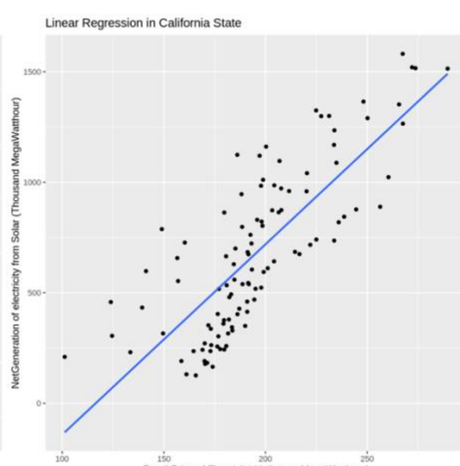
To isolate the effect of seasonality from the effect of retail prices on generation of solar power, we employed a second linear regression model where we created dummy variables for each month as mentioned in equation 3 above. We noticed that after the inclusion of these dummy variables, there were changes in the magnitude of coefficients on retail price. For California and Arizona, we see an increase in the coefficients whereas for New York and New Jersey, there is a decrease in the coefficients.

State	**Linear Regression Model 1**	**Linear Regression Model 2**
New York	0.7636**	0.6675**
New Jersey	4.3903***	3.8052***
California	8.609***	9.5968***
Arizona	6.6518***	9.306***

Significant codes: '***' 0.001 '**' 0.01 '*' 0.05

Table 1: Comparison of coefficients before and after inclusion of dummy variables. Model 1 is before and Model 2 is after inclusion of dummy variables.





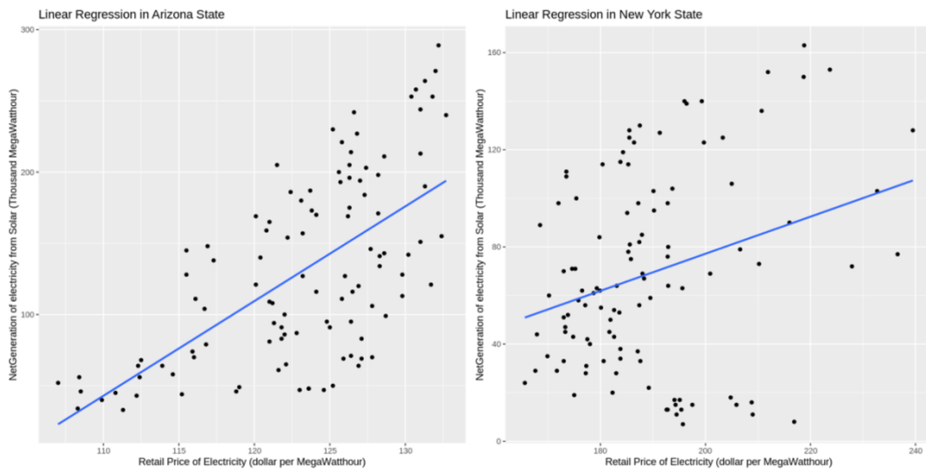


Figure 4: Plots for Linear Regression model in each state
Retail Price of Electricity vs Generation of Solar Power

Conclusion

The first logistic regression model confirmed our hypothesis that having a household with higher income and education level, and being more conscious about climate increases the probability of solar power adoption. According to the results, the political orientation (being liberal or conservative) is not a strong factor, so our hypothesis is incorrect. However, there's a slight significant effect for the middle, indicating that people who identified themselves as moderate in political opinions might be more open to renewable technologies like solar power.

The second logistic regression of solar power adoption by social influence (i.e. whether they frequently heard people they know talking about solar panels) indicated a negative effect on considering solar power application when frequently heard about the topic from one's peers. This might be helpful for an installation provider to rely less on word of mouth for their marketing efforts (referral discounts, viral campaigns etc.) and to seek other means of encouraging adoption.

In our third logistic regression model, it is suggested that living in Arizona and California, along with higher education level and larger household space are significant factors for the tendency towards buying rather than leasing solar power applications. This information is beneficial for a business evaluating the solar panel market in the four states, identifying and categorizing its potential customers, and deploying resources to different locations according to local preferences.

Our linear regression analysis on time series data suggested that the monthly average retail price of electricity is positively correlated with monthly residential generation of solar power in all four states. The result confirmed our hypothesis that retail price of electricity is a driving factor for solar power generation and may influence customer motivation in adopter.

With the logistic and linear regression models we built, we were able to test our hypotheses and answer some interesting questions about the potential determining factors for solar power adoption. The models can be useful when there is more survey data available for different states in the future and will help solar power businesses gain more comprehensive insights on the solar panel industry.

References

McKinsey & Company (2022). Global Energy Perspective 2022 Executive Summary. https://www.mckinsey.com/~/media/McKinsey/Industries/Oil%20and%20Gas/Our%20Insights/Global%20Energy%20Perspective%202022/Global-Energy-Perspective-2022-Executive-Summary.pdf.

Solar Energy Industries Association. Solar Industry Research Data. https://www.seia.org/solar-industry-research-data

Araújo, Kathleen & Boucher, Jean & Aphale, Omkar. (2019). A Clean Energy Assessment of Early Adopters in Electric Vehicle and Solar Photovoltaic Technology: Geospatial, Political and Socio-demographic Trends in New York. Journal of Cleaner Production. 216. 10.1016/j.jclepro.2018.12.208.https://www.sciencedirect.com/science/article/abs/pii/S0959652618339283

U.S. Energy Information Administration - EIA - Independent Statistics and Analysis. https://www.eia.gov/electricity/data/browser/#/topic/0?agg=1,0,2&fuel=004&geo=qnifi05c03j78&sec=o3g&linechart=ELEC.GEN.SUN-US-99.M~ELEC.GEN.SUN-NV-99.M~ELEC.GEN.SUN-CA-99.M&columnchart=ELEC.GEN.SUN-US-99.M~ELEC.GEN.SUN-NV-99.M~ELEC.GEN.SUN-CA-99.M&map=ELEC.GEN.SUN-US-99.M&freq=M&start=200101&end=201611&ctype=linechart&ltype=pin&rtype=s&maptype=0&rse=0&pin=

Sigrin, Ben; Dietz, Tom; Henry, Adam; Ingle, Aaron; Lutzenhiser, Loren; Moezzi, Mithra; Spielman, Seth; Stern, Paul; Todd, Annika; Tong, James; Wolske, Kim (2017): Understanding the Evolution of Customer Motivations and Adoption Barriers in Residential Solar Markets: Survey Data. National Renewable Energy Laboratory. 10.7799/1362095. https://data.nrel.gov/submissions/68

U.S. Energy Information Administration - EIA - Independent Statistics and Analysis. https://www.eia.gov/electricity/data/browser/#/topic/7?agg=0,1&geo=g&linechart=ELEC.PRICE.US-ALL.M~ELEC.PRICE.US-RES.M~ELEC.PRICE.US-COM.M~ELEC.PRICE.US-IND.M&columnchart=ELEC.PRICE.US-ALL.M~ELEC.PRICE.US-RES.M~ELEC.PRICE.US-COM.M~ELEC.PRICE.US-IND.M&map=ELEC.PRICE.US-ALL.M&freq=M&start=200101&end=201611&ctype=linechart&ltype=pin&rtype=s&maptype=0&rse=0&pin=&endsec=vg