The Data Open : Team 29

Hemanth, Dastagiri, Shreyas, Raymond

Topic Question

Where should a solar power plant be built to have the maximum impact on the surrounding community?

Motivation and Definitions

Since the turn of the century, the topic du jour for environmentalists and politicians alike has been climate change--and it is easy to see why. According to the National Oceanic and Atmospheric Administration, 16 of the 17 hottest years on record have occurred since 2000[1], and it's widely known that sea levels are on the rise. With more and more intense heatwaves, droughts, and storms resulting from these changes each year, it's an open question as to what humanity can do to address climate change.

It's been long hypothesized that a major driver of climate change is human pollution; the fossil fuels that burn in our cars, heat our homes, and power our cities emit harmful pollutants that affect our health and the environment in profound ways. Until now, few alternatives to using such pollutants existed that were both cost-efficient and technologically feasible.

Today, however, solar, wind, nuclear, and other renewable energy technologies have matured and are on track to overtake fossil fuels as the core power sources of our world. Solar energy has shown a great deal of promise in particular--with businesses such as Elon Musk's SolarCity leading the way. With this backdrop in mind, our team thought it best to explore an interesting question about the future of solar power plants: **Where should a solar power plant be built to have the maximum impact on the surrounding community?**

To address this question, we offer the following definitions of our project's scope. Although several types of solar power exist, we focus on the most common solar technology--photovoltaic (PV) cells. We discuss our system for determining maximum impact more formally in the methodology section below, but it can be summarized as measuring where building a PV plant will cause energy prices to shift such that gasoline consumption drops the most. Based on these results, we assess community impact at the city level--meaning that we offer one simple deliverable: a recommendation for the best city to build a PV plant in.

Relevant Datasets

Provided

[1] https://www.nasa.gov/press-release/nasa-noaa-data-show-2016-warmest-year-on-record-globally

- *solar_radiation*: used in calculating the subset of 10 states which have the highest average GHI
- *photovoltaic*: used in calculating the cost/watt of a PV station in a given zipcode
- *fossil_fuel_consumption*: used in visualizing the trend of fossil fuel consumption in the subset of 10 selected states

Additional datasets

- Average gasoline prices in major U.S. regions -- https://www.eia.gov/dnav/pet/pet_pri_gnd_a_epm0_pte_dpgal_w.htm
- U.S. zipcode database -- http://federalgovernmentzipcodes.us/
- Gasoline energy conversion -- https://www.eia.gov/energyexplained/index.cfm?page=about_energy_conversion_calculator

Non-Technical Executive Summary

Elon Musk, build your next solar plant in San Bernardino, California! This is the city that had the best differential between solar energy prices and fossil fuel prices. According to a 2016 Gallup Poll, 73% of Americans would prefer to use alternative energy sources to fossil fuels.[2] For the longest time, the biggest drawback has been the price. We argue that the location most likely to make an impact on solar energy is indeed this (not so) little town in Southern California as the difference between paying for solar energy and paying for gasoline is the lowest. The argument goes that because people naturally want alternative energy, when the price differential is lowest between fossil fuels and solar power, people are more likely to use solar power. Even if solar power is more expensive than fossil fuels (and it is— we find that solar power is almost always more expensive), an area with the lowest differential is most likely to have people willing to pay the premium for solar power. The devil's in the details, of course, and stringing together the various datasets provided in order to determine the price of solar versus the price of gas for any given city was quite the challenge. If we had more time, we'd finishing conducting some predictions we were working on about how this new plant would affect solar prices versus gas prices and explore the relationship between the two more. If you're curious, we break down our results even further and provide the best city to build a plant in for several individual states in the *Technical Executive Summary* below.

[2] http://www.gallup.com/poll/190268/prioritize-alternative-energy-oil-gas.aspx

Technical Executive Summary

Methodology

Assumptions

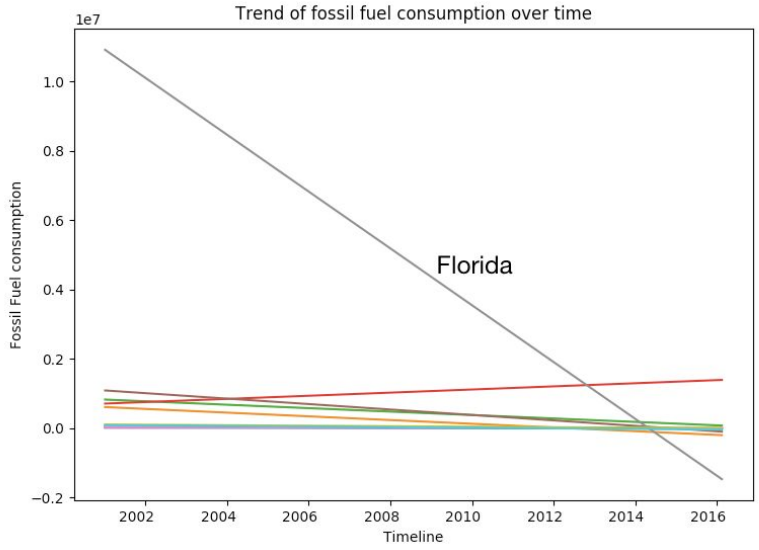
In the process of developing our approach to solving the problem we were required to make a few assumptions. This is both a byproduct of the limited time in this competition and also due to the difficulty in obtaining real-world data at the required level of granularity. The assumptions are detailed below:

1. The cost/watt of a given PV station is indicative of the cost/watt of any future PV stations built in the same area. Our problem statement is not to predict future cost/watt in an area and therefore in the interest of maintaining a reasonable scope, we make this assumption.
2. GHI is a good indicator of viability for successful implementation of solar energy. We are applying our analysis to a subset of 10 states, which are the ones with the highest average GHI over the last 15 years. Though there are other indicators (temperature, DHI, and DNI) we believe this to be the best single indicator.
    a. The chosen states are: AZ, NV, CA, TX, FL, OK, LA, MS, AL, UT
3. Our proposed metric for determining the "best" city to build a PV station, where best is defined as the city which will experience the greatest positive impact, is the city with the most negative difference between cost/watt of solar energy and cost/watt of traditional fossil fuel.
4. Fossil fuel consumption trends may be indicative of likelihood to adopt solar energy successfully. That is, of the locations we propose, we believe that if they are truly the best locations to build PV stations, the region should have a negative trend of fossil fuel consumption.

- Oil prices are similar across a state and furthermore, are an indicator of the cost of fossil fuels as a category. Comparing the cost of solar energy and fossil fuels is a key part of our analysis and we would have liked to use more granular data at the city-level but were unable to find a satisfactory dataset in the constraints of this competition.
- Energy use in a household is constant throughout the year

Analytic Process

1. As mentioned in *Assumptions*, because of the incompleteness of data available, we narrowed our analysis to the states that have the highest average Global Horizontal Irradiance (GHI). This metric can be thought of as the total amount of radiation that reaches the surface of each region. The states with the highest GHI are our most likely candidates for the best city to be a part of as these are the states where solar power is the most viable. Notably, we found that all but one of these states are decreasing their fossil fuel usage over time.



2. Within our subset of states, we determined the average cost per watt for any given zip-code in each state by averaging cost-per-watt across the dataset.
3. Next, we determined the average cost of gasoline per gallon for all zip-codes in our list of states. To do this, we used the EIA gasoline price database as the oil_price table given has no location information. Unfortunately, this data was not granular at the city or state level, but rather gave prices by region. We assumed that all zip-codes in a state had the same gas prices and that all states in a given region had the same gasoline prices as well.
4. At this point, we had the average cost per watt for solar power and the average cost per gallon of gasoline in any given region. We annualized these figures to find the amount of money an average american household in each of these zip codes would spend on electricity total using solar power and the amount of money every household spends on gasoline. We calculated the difference between these figures to determine which zip-codes have the greatest incentive to use solar power--that is areas with the greatest difference between solar prices and gas prices.
5. Using an additional zipcode database we connected each of these zip codes to a city and averaged differences by city. We then selected the cities in each state with the most negative difference between annualized solar prices and annualized gas prices.

Top Cities by Zipcode for each State for Building a PV Station

('ENTERPRISE', 'UT', 84725),
('DUNCAN', 'AZ', 85534),
('VIRDEN', 'AZ', 85534),
('LAS VEGAS', 'NV', 89177),
('MGM PROPERTIES', 'NV', 89177),
('SAN BERNARDINO', 'CA', 92415),
('S B COUNTY OFFICES', 'CA', 92415),
('JACKSONVILLE', 'FL', 32226),
('JAX', 'FL', 32226),

('HAZEL GREEN', 'AL', 35750),
('HATTIESBURG', 'MS', 39402),
('BATON ROUGE', 'LA', 70813),
('SOUTHERN UNIVERSITY', 'LA', 70813),
('VINITA', 'OK', 74301),
('CENTRALIA', 'OK', 74301),
('PENSACOLA', 'OK', 74301),
('WHITE OAK', 'OK', 74301),
('HOUSTON', 'TX', 77055),
('HILSHIRE VILLAGE', 'TX', 77055)

Top Zipcodes Across All 10 States for Building a PV Station

('DUNCAN', 'AZ', 85534),
('VIRDEN', 'AZ', 85534),
('POSEY', 'CA', 93260),
('BALANCE ROCK', 'CA', 93260),
('BRYN MAWR', 'CA', 92318),
('PARKER DAM', 'CA', 92267),
('BLACK MEADOW LANDING', 'CA', 92267),
('SAN BERNARDINO', 'CA', 92415),
('SN BERNRDNO', 'CA', 92415),
('S B COUNTY OFFICES', 'CA', 92415),
('CHICAGO PARK', 'CA', 95712),
('TRAVER', 'CA', 93673),
('YETTEM', 'CA', 93670),
('IRVING', 'TX', 75014),
('HOUSTON', 'TX', 77055),
('HILSHIRE VILLAGE', 'TX', 77055)