Jackson Curtis

Solar EDA

We are interested in analyzing the amount of money customers save by switching to solar, so solar companies can use that information as a sales tactic. If solar companies can accurately demonstrate the savings that customers will have, they can better market their products to willing customers. Specifically, we would like to estimate the monthly savings by this customer (and by extension, people similar to him) after switching to solar, as well as estimate how long it will take him to make up the initial $15,000 investment. The first question concerns the mean of the distribution (how much do you save on average), and the second concerns the distribution of savings (will some months involve huge savings and some months few or is it consistent?).

The plot below shows that the data has several interesting features. One is that it appears unlikely that any normal assumptions apply, there is high variation month to month in the data. Also, we are dealing with points in time, so we could face the possibility of seasonality (where certain months of the year are consistently higher than others) as well as a moving average (where the price of electricity trends up and down over time). The indicator of solar or not also turns this into two samples, that we may or may not treat independently.

I would probably do a time series analysis on this problem. It would allow us to deal with the non-independent data and the seasonality. Using solar as a covariate will allow us to make inference on the difference in “Solar yes” and “Solar no.”

I don’t know how to account for the seasonality, especially how to diagnose if it’s there and if it is significant. I also don’t know how to account for uncertainty when our normal assumptions don’t apply.

