Final Capstone Proposal

Predicting Untapped Solar Panel Install Markets

The solar industry is a rapidly growing sector of the economy in labor as well as finance. The market for these products can in some respect be considered the entirety of the United States. However, for companies in this business, finding the most cost-effective places to invest resources to promote growth is advantageous. Not all markets are going to be equally accessible and there are an variety of factors that could make one market easier or more difficult to stimulate growth in.

Using a collection of data sources ranging from Solar radiation and existing installation data from Google’s project sunroof on Kaggle, Zillow’s median home prices, voting and party affiliation data, cost of electricity in a given region as well as other socio-economic data gathered from csv downloads and web scraping will be useful in properly constructing a model for identifying locations more amenable to residential solar installation growth.

To complete this project I intend to use various techniques including web scraping, feature engineering, data visualization,and supervised learning to identify which features together mix to create health markets for this type of business growth. From outside of the course some mapping techniques to generate visuals to aid in the understanding of the results.

The list of factors that can influence such a model is likely expansive, and finding data at a localized level that would aid in providing the most detailed insights map prove to be difficult. In my initial efforts to gather data and build a model I found it difficult to collect the data that I wanted to gather with similar localization parameters. Such an example would be the difficult I had in locating voter affiliation, and electric price data at the zip code level and as of now am settling for the state level of the data. Getting as much data at as localized a level as I would like could prove to be challenging.

Included in the following link are my initial efforts to gather and model data for this project:

<https://github.com/gurkpet/Final-Capstone>